HYPONUM

User manual

Desktop edition

Analysis, interpretation and reproducible examples

Reference edition

September 2026

Using this manual

Use this book to choose an analysis, prepare data, run it and interpret the result. Start with the family that matches your task. The opening part covers files, tables and the working environment.

Part	Use it to
Working with HypoNum	Open and prepare data, work with tables and understand the analysis population.
Explore	Examine distributions, patterns and relationships graphically.
Statistics	Estimate, compare and model; study associations and survival.
Survey	Summarize a survey design and examine item consistency.
Time Series	Study ordered observations, cross-correlation and forecasting.
Quality	Monitor a process, assess measurement and capability, and study reliability and acceptance sampling.
Design	Plan experiments, sample sizes and prospective power.
Predict	Apply retained models and examine their predictions.
Results and support	Save and present results, extend the application and recover work.

The book follows the current task-oriented menus: File, Edit, Data, Explore, Statistics, Survey, Time Series, Quality, Design, Predict, Results, Plug-ins and Help. Each procedure gives its exact menu path. A command offered in two places has one primary explanation with the alternative location identified.

This edition describes HypoNum. Save packaged projects with the `.hnum` extension and use `hyponum` for command-line examples. Pre-launch `.tabelz` project files are not supported or converted. Python imports and registered operation identifiers retain the technical name `tabelz`. The companion validation report identifies its own historical assessment baseline; updated product branding does not extend that assessment to new methods or a new release.

Read the question, required data, assumptions, settings and interpretation together. A result concerns the population and design that produced it. Follow each example's stated settings: changing a baseline, exclusion or rule selection can change the answer.

Illustrations explain the stated calculations. Compare data, settings and numerical values rather than on-screen styling; tables state the rounding used.

The companion *HypoNum Statistical Validation Report* presents selected comparisons with published references and states their scope. A method's availability does not imply independent assessment in that report.

Reading conventions

Menu paths use the form **Data > Merge tabs...**; control labels use bold sans serif, as in **Create tab**. Literal entries and formulas use a fixed-width face, as in `split([code],"+",2)`. *Included rows* are rows not currently excluded from analysis. *Missing* includes nulls and floating-point NaN values unless the procedure gives a more specific definition.

Contents

Using this manual i

I Working with HypoNum 1

1 Your first project 2

1.1 Start with the observational unit . . . 2
1.2 Import, inspect and save . . . 2
1.3 Selection, exclusion and deletion . . . 3
1.4 What a preview represents . . . 3
1.5 Preparation history and saved results . . . 3
1.6 Missing values need a reason . . . 3

2 Finding your way around the workspace 5

2.1 The active dataset . . . 5
2.2 Selecting, editing and inspecting . . . 5
2.3 Keyboard navigation . . . 6
2.4 Display themes and printed reports . . . 6
2.5 Preparing a change . . . 6
2.6 Operations, history and saved results . . . 6

3 Importing, saving and exporting data 8

3.1 Import a text file or workbook . . . 8
3.2 Import a database result . . . 9
3.3 Import from a database server . . . 10
3.4 Find a source or a value . . . 11
3.5 Save and reopen a project . . . 11
3.6 Export a workbook . . . 12
3.7 Import statistical data files . . . 12
3.8 Import open interchange formats . . . 13
3.9 Exchange files without promising lossless round trips . . . 14

4 Editing values and columns **16**
4.1 Undo, redo and row selection . . . 16
4.2 Create and name a column . . . 16
4.3 Calculate a column with a formula . . . 17
4.4 Recode categories . . . 17
4.5 Column order and row order . . . 17
4.6 Direct edits and filling downward . . . 18

5 Column meaning, case weights and analysis groups **19**
5.1 Describe a column . . . 19
5.2 Declare frequency weights . . . 20
5.3 Run the same settings by group . . . 20
5.4 Define a multiple-response question . . . 20

6 Preparing and combining tables **22**
6.1 Choose the condition before the action . . . 22
6.2 Decide what a duplicate means . . . 23
6.3 Handle missing values deliberately . . . 24
6.4 Inspect potential outliers . . . 27
6.5 Split a table . . . 28
6.6 Concatenate tables . . . 29
6.7 Choose a merge relationship . . . 29
6.8 Stack and pivot measurements . . . 29
6.9 General joins in Merge tabs . . . 31

7 Create derived columns with a preview **33**
7.1 Recode values explicitly . . . 33
7.2 Ranks and rank-derived scores . . . 33
7.3 Standardize with a stated denominator . . . 34
7.4 Count matching cells . . . 34
7.5 Choose and retain bin boundaries . . . 34

8 From a question to an analysis **35**
8.1 Write the question in measurable terms . . . 35
8.2 Choose the analysis population . . . 35
8.3 Inspect the data before fitting . . . 36
8.4 Understand an estimate and its interval . . . 36
8.5 Hypotheses, direction and multiplicity . . . 36
8.6 Missing data and prediction . . . 37
8.7 Read the complete result . . . 37

II Explore **38**

9 Exploring distributions and relationships **39**

9.1 Choose a plot for the question . . . 39
9.2 Scatterplots and descriptive overlays . . . 40
9.3 Colour, facets and linked selection . . . 41
9.4 Compare several variables . . . 41
9.5 Variation between groups and occasions . . . 42
9.6 Prioritize categories with a Pareto chart . . . 42
9.7 Investigate more than one mode . . . 42
9.8 Save, reuse and export a plot . . . 43

10 Counts, percentages and cumulative distributions **44**

10.1 Set the roles and the summary . . . 44
10.2 State the percentage denominator . . . 44
10.3 Use a pie for a simple part-to-whole summary . . . 45
10.4 Read a two-way heatmap . . . 45
10.5 Read an empirical cumulative distribution . . . 45
10.6 Limits and retained results . . . 46
10.7 Matrix, probability and spatial displays . . . 46
10.8 Editing a plot without changing the analysis . . . 47

III Statistics **48**

11 Comparing groups and testing claims **49**

11.1 A small dataset for following the calculations . . . 49
11.2 Select the design before the test . . . 49
11.3 One mean: calculate the effect and its uncertainty . . . 50
11.4 Independent groups and paired observations . . . 51
11.5 ANOVA and multiple comparisons . . . 52
11.6 Proportions, tables and variances . . . 52
11.7 Rank and sign procedures . . . 53
11.8 Read and report the result . . . 54
11.9 Plan a future sample . . . 54
11.10 Establishing a sufficiently small difference . . . 55

12 Regression for a numeric response **56**

12.1 Set up the response and predictors . . . 56
12.2 Worked example: a fitted line . . . 57
12.3 Model scale and assumptions . . . 57
12.4 Read diagnostics before drawing conclusions . . . 58
12.5 Choose a covariance model . . . 58
12.6 Understand this dialog's validation . . . 59
12.7 Search for a predictor subset . . . 60
12.8 Whole-factor linear effects . . . 60
12.9 General linear models . . . 61
12.10 Fitting a nonlinear curve . . . 62
12.11 Comparing two measurement methods . . . 65

13 Models for categories and counts **68**

13.1 Binary logistic regression . . . 68
13.2 Ordinal logistic regression . . . 69
13.3 Nominal logistic regression . . . 69
13.4 Poisson regression and exposure . . . 70
13.5 Negative binomial regression . . . 71

14 Repeated observations and dependent data **72**

14.1 One-way panel fixed-effects regression . . . 72
14.2 A Gaussian model with a random intercept . . . 74
14.3 Nested, crossed and random-slope mixed models . . . 76
14.4 Binomial and Poisson mixed models . . . 78
14.5 Population-average GEE for clustered responses . . . 80
14.6 Repeated measures with a residual covariance . . . 81

15 Relationships and multivariate structure **85**

15.1 Correlation . . . 85
15.2 Principal component analysis . . . 86
15.3 Exploratory factor analysis . . . 88
15.4 Hierarchical clustering . . . 89
15.5 K-means clustering . . . 89
15.6 Simple correspondence analysis . . . 90
15.7 Multivariate analysis of variance . . . 92
15.8 Further reading . . . 94

16 Building and evaluating predictive models **95**
16.1 A training, tuning and test workflow . . . 95
16.2 Regression trees . . . 96
16.3 Classification trees . . . 97

17 Survival curves and event-time regression **100**
17.1 Kaplan–Meier survival curves . . . 100
17.2 Cox proportional-hazards regression . . . 102

18 Multiple imputation and missing-data uncertainty **103**
18.1 Multiple imputation with pooled regression . . . 103
18.2 Missing value analysis before imputation . . . 105

19 Fitting probability distributions **108**
19.1 Choose measurements and families . . . 108
19.2 Decide the origin before fitting . . . 108
19.3 Worked example: a normal likelihood fit . . . 108
19.4 Parameter estimates and model comparison . . . 109
19.5 Why some familiar candidates are unavailable . . . 109
19.6 Goodness of fit and simulation uncertainty . . . 110
19.7 Multimodality and graphical review . . . 110

20 Extended regression and count-table models **111**
20.1 Enter predictors in planned blocks . . . 111
20.2 Know which scale a curve fits . . . 111
20.3 Estimate a variance power deliberately . . . 112
20.4 Two-stage least squares requires valid instruments . . . 112
20.5 Generalized linear models . . . 112
20.6 Loglinear models for count tables . . . 113

21 CHAID, nearest neighbours and finite mixtures **115**
21.1 CHAID trees . . . 115
21.2 Nearest neighbours . . . 116
21.3 Model-based clustering . . . 116

22 Distances, canonical relationships and category quantification **118**
22.1 Multidimensional scaling . . . 118
22.2 Canonical correlation . . . 118
22.3 Categorical principal components . . . 119
22.4 Multiple correspondence analysis . . . 119

23 Grouped survival and changing covariates **120**
23.1 Actuarial life tables . . . 120
23.2 Cox regression with time-dependent covariates . . . 121

24 Grouped summaries and cross-tabulation **122**
24.1 Summarize a numeric response . . . 122
24.2 Understand frequency counts . . . 122
24.3 Build a categorical table . . . 123
24.4 Display, export and save . . . 123

25 Descriptive reports and categorical association **124**
25.1 Choose a report for the question . . . 124
25.2 Build a custom table with explicit denominators . . . 124
25.3 Crosstab statistics depend on scale and direction . . . 125
25.4 Kendall and partial correlation . . . 125

26 Distribution diagnostics **126**
26.1 Normality . . . 126
26.2 One suspected outlier: Grubbs . . . 127
26.3 A Poisson count model . . . 127
26.4 Data use and retained output . . . 127

27 Probability calculations and random data **128**
27.1 Specify the distribution . . . 128
27.2 Density, probability and quantiles . . . 128
27.3 Worked probability calculations . . . 129
27.4 Read the plot and retain the calculation . . . 129
27.5 Generate reproducible random values . . . 129
27.6 Simulating inputs through a model . . . 130

28 Known-sigma, variance and event-rate inference **133**
28.1 Choose raw rows or an entered summary . . . 133
28.2 A mean with known sigma . . . 133
28.3 A single population variance . . . 134
28.4 One Poisson event rate . . . 134
28.5 Compare two Poisson rates . . . 135
28.6 Units, limits and saved results . . . 135

29 Blocked ranks and ordered runs **136**
29.1 Friedman: one observation per treatment and block . . . 136
29.2 Runs: preserve the observation order . . . 137
29.3 Exact and approximate inference . . . 137
29.4 Record population and limits . . . 137

30 Permutation tests **138**
30.1 Independent and paired designs . . . 138
30.2 Worked example: enumerate every assignment . . . 138
30.3 Monte Carlo inference . . . 139
30.4 Read and save . . . 139

31 Comparisons with a control and between unequal groups **140**
31.1 Dunnett: several treatments against one control . . . 140
31.2 Games–Howell: all pairs with unequal variances . . . 140
31.3 Specify eligible groups and inspect the results . . . 141

32 One-sample and paired equivalence **142**
32.1 Choose the contrast and margins . . . 142
32.2 Worked example: paired differences . . . 142
32.3 The matching interval . . . 143
32.4 Report the declared decision . . . 143

33 Repeated-measures ANOVA and comparison families **144**
33.1 Define the population and the reference values . . . 144
33.2 Read both within-subject inferential approaches . . . 144
33.3 Contrasts, simple effects and interaction contrasts . . . 145
33.4 Independent-group post hoc procedures . . . 145
33.5 Analysis of means . . . 146

34 Distribution, ordered and paired categorical tests **148**
34.1 Distributional hypotheses . . . 148
34.2 Ordered groups and rank comparisons . . . 149
34.3 Paired and stratified categories . . . 149
34.4 A variance confidence set can be disconnected . . . 149

35 Forests, penalized classification and conditional quantiles **151**
35.1 A regression random forest . . . 151
35.2 Penalized logistic regression . . . 152
35.3 Quantile regression . . . 153

36 Bootstrap uncertainty for an estimate **154**
36.1 Select the statistic and the resampling unit . . . 154
36.2 Separate sampling uncertainty from simulation error . . . 154
36.3 Use BCa only where supported . . . 154
36.4 Read failed-draw and endpoint warnings . . . 155

37 Combining evidence across studies **156**
37.1 Prepare comparable study effects . . . 156
37.2 Choose the pooling and uncertainty model . . . 156
37.3 Read heterogeneity and prediction separately . . . 157
37.4 Moderators, sensitivity and small-study checks . . . 157

38 Bayesian estimation and model comparison **158**
38.1 Declare the scientific parameter and its prior . . . 158
38.2 Work through a beta–binomial update . . . 159
38.3 Interpret posterior summaries and Bayes factors . . . 159
38.4 Keep model selection uncertainty visible . . . 159

39 Guided statistical studies **160**
39.1 Comparing groups . . . 160
39.2 Association and prediction . . . 160
39.3 Missing-data analysis . . . 161
39.4 Time-series studies . . . 161
39.5 Reliability and experimental studies . . . 162

IV Survey **164**

40 Survey summaries and scale consistency **165**
40.1 Survey means and proportions . . . 165
40.2 Item consistency and summative scores . . . 167

41 Inference from a complex sample **169**
41.1 Save the table's sampling design . . . 169
41.2 Estimate a domain without discarding its design context . . . 169
41.3 Tables and survey models . . . 170
41.4 Supplementary proportion intervals and design effects . . . 170
41.5 Count, nominal and interacting survey models . . . 171
41.6 Generate replicate weights from a declared sampling plan . . . 172
41.7 Calibrate to fixed known population totals . . . 173

42 Scale reliability and rater agreement **174**
42.1 Item consistency and declared halves . . . 174
42.2 Intraclass correlations . . . 174
42.3 Categorical agreement . . . 175

V Time Series **177**

43 Understanding an ordered series **178**
43.1 Declare time, calendar and missingness . . . 178
43.2 A small series used throughout these chapters . . . 178
43.3 Autocorrelation and partial autocorrelation . . . 178
43.4 Stationarity and differencing . . . 179
43.5 Seasonal decomposition . . . 180

44 Fitting and interpreting forecasts **181**
44.1 Begin with meaningful benchmarks . . . 181
44.2 Worked example: three different assumptions . . . 181
44.3 What the prediction intervals assume . . . 182
44.4 Fit ARMA or ARIMA explicitly . . . 182
44.5 Automatic selection and diagnostics . . . 183

45 Evaluating and preserving a forecast **184**
45.1 Plan the evaluation before comparing models . . . 184
45.2 Read accuracy in the response's units . . . 184
45.3 Follow a small rolling evaluation . . . 185
45.4 Fit transformations inside each training history . . . 185
45.5 Use information that was available at the origin . . . 186
45.6 Review warnings, then save the complete study . . . 186

46 Predictors, multiple series and changing volatility **187**
46.1 Conditional forecasts with predictors . . . 187
46.2 Joint dynamics: VAR and VECM . . . 187
46.3 Volatility: GARCH . . . 188
46.4 A unit root with one possible break . . . 188
46.5 Sources and further study . . . 189
46.6 Deterministic trend analysis . . . 189

47 Cross-correlation **191**

47.1 Choose the series and their time axis . . . 191
47.2 Interpret the sign and overlap . . . 191
47.3 Reproduce a one-period delay . . . 191
47.4 Treat reference bands as a limited guide . . . 192
47.5 Keep gaps in their original positions . . . 192

VI Quality **193**

48 Control charts **194**

48.1 Choose the question and the chart . . . 194
48.2 Prepare the table and set up the analysis . . . 195
48.3 Individuals and moving range . . . 196
48.4 Subgroup means and dispersion . . . 199
48.5 Counts, proportions and rates . . . 201
48.6 Small sustained changes: CUSUM and EWMA . . . 203
48.7 Recognize and interpret sequence signals . . . 206
48.8 From a chart to a documented decision . . . 207
48.9 Short-run standardized Z–MR monitoring . . . 208
48.10 Finite-window moving-average charts . . . 210
48.11 References and further reading . . . 212

49 Between/within and zone monitoring **213**

49.1 Between/within I–MR–R/S charts . . . 213
49.2 Accumulate zone scores . . . 214
49.3 Monitor against an approved baseline . . . 214

50 Multivariate subgroup monitoring **215**

50.1 Declare complete reference and monitoring groups . . . 215
50.2 Locate a change in the multivariate mean . . . 215
50.3 Locate a change in joint spread . . . 216
50.4 Separate a centre shift from a spread shift . . . 216
50.5 Review and retain the baseline . . . 217

51 Multivariate EWMA **218**

51.1 Set a weight and a documented decision limit . . . 218
51.2 Freeze the reference population . . . 218
51.3 Follow the recursion and its changing covariance . . . 218

51.4 Calculate the first two monitoring points . . . 219
51.5 Use stages for an intentional restart . . . 219
51.6 Retain the signal and source context . . . 219

52 Planning a measurement-system study 221
52.1 Choose parts, appraisers and trials deliberately . . . 221
52.2 Create a randomized collection worksheet . . . 222
52.3 Inspect individual readings first . . . 222

53 Repeatability and reproducibility 223
53.1 Set up a crossed study . . . 223
53.2 Keep variance, spread and percentages distinct . . . 224
53.3 Work through a complete crossed study . . . 224
53.4 HypoNum extended estimation and uncertainty . . . 225
53.5 Read and retain the whole result . . . 225
53.6 Choose and retain a measurement standard . . . 225
53.7 Methods in the AIAG profile . . . 228

54 Other measurement designs 234
54.1 An additional site or fixture factor . . . 234
54.2 Nested and destructive studies . . . 234
54.3 Linearity and bias . . . 235
54.4 Type-1 study of one standard . . . 235
54.5 Check stability in acquisition order . . . 236
54.6 A preliminary repeatability screen . . . 236

55 Agreement and binary gage response 237
55.1 Repeatability of categorical decisions . . . 237
55.2 Read exact agreement and kappa together . . . 237
55.3 A binary gage's switching curve . . . 238
55.4 Why a threshold interval may be unbounded . . . 239
55.5 Fit a transition curve from counted trials . . . 239
55.6 Check support and model adequacy . . . 240

56 Specifications and capability 241
56.1 Define the requirement . . . 241
56.2 Choose the population and variation estimate . . . 242
56.3 Understand the normal indices . . . 242
56.4 Work through thirty measurements . . . 242

56.5 Keep estimation uncertainty visible . . . 243
56.6 Separate conformance from a capability claim . . . 244
56.7 Counts: defectives and defects . . . 244
56.8 Capability for nonnormal measurements . . . 245
56.9 Within and between batch capability . . . 247
56.10 Review capability with stability and model fit . . . 249
56.11 Sources and interpretation . . . 250

57 Reliability engineering 251
57.1 Right-censored life distributions . . . 251
57.2 Mixed exact and interval-censored life data . . . 254
57.3 Accelerated life at declared stress conditions . . . 256
57.4 Reliability demonstration and precision plans . . . 259
57.5 Broader life families and nonparametric support . . . 261
57.6 Regression with life data . . . 262

58 Recurring failures, degradation and warranty 263
58.1 Repairable systems . . . 263
58.2 Degradation-based lifetime . . . 264
58.3 Warranty triangles and predictions . . . 264

59 Attribute acceptance and tolerance intervals 266
59.1 Plan acceptance from defective-unit counts . . . 266
59.2 Estimate an interval covering a population fraction . . . 267
59.3 The compact censored-Weibull panel . . . 268
59.4 Save the selected study . . . 269

60 Acceptance sampling by variables 270
60.1 State the two quality points . . . 270
60.2 How the acceptance rule is constructed . . . 271
60.3 Worked example: an upper dimensional limit . . . 271
60.4 Read and retain the operating characteristic . . . 272

61 Guided quality studies 273
61.1 Review methods and preserve the reasoning . . . 273
61.2 A measurement-system study . . . 273
61.3 From measurement to stability and capability . . . 275

VII Design **279**

62 Designing experiments **280**

62.1 Choose a design for the question . . . 280
62.2 Create a reproducible run sheet . . . 281
62.3 Analyze responses using the design's model . . . 281
62.4 Respect the whole-plot error in split-plot experiments . . . 283
62.5 Binary responses and variability . . . 283
62.6 Optimize several responses without hiding an unacceptable one . . . 284
62.7 Extend an executed design . . . 284
62.8 Related planning workflows . . . 285

63 Robust settings and Taguchi designs **286**

63.1 Robust parameter design . . . 286
63.2 Taguchi orthogonal arrays . . . 287
63.3 Power for the planned or actual design . . . 287

64 Planning for estimation precision **289**

64.1 Match the quantity to the intended interval . . . 289
64.2 Mean precision and a complete planning example . . . 289
64.3 Proportion precision uses the Wilson convention . . . 290
64.4 Tolerance content is different from confidence . . . 290
64.5 Distribution-free planning specifies content, not width . . . 291
64.6 Read the candidate check and preserve assumptions . . . 291

65 Prospective power and sample size **292**

65.1 Choose a family and count the units correctly . . . 292
65.2 Understand the initial settings . . . 293
65.3 Follow a prospective mean calculation . . . 293
65.4 Proportions, ANOVA, variance and event rates . . . 293
65.5 Equivalence asks for an interval inside the margins . . . 294
65.6 Choose a solve or curve with supported meaning . . . 294
65.7 Save the design assumptions . . . 295
65.8 Correlation, regression and repeated-measures planning . . . 295
65.9 Exact binary planning and search claims . . . 296
65.10 Additional correlation and allocation choices . . . 297
65.11 A distinct paired-proportion approximation . . . 297

VIII Predict **298**

66 Fit once and score another table **299**

66.1 Fit a model with explicit response meaning . . . 299
66.2 A complete training example . . . 300
66.3 Why the two OLS intervals differ . . . 300
66.4 Inspect influence without automatically deleting observations . . . 300
66.5 Score a new table using the saved fit . . . 301
66.6 Retain the model as well as its displayed rows . . . 302

67 Profiles and multiresponse desirability **303**

67.1 Select compatible models and an operating region . . . 303
67.2 Give the response goals their own units . . . 303
67.3 Follow the saved five-row model . . . 304
67.4 Read a profile as a held-setting comparison . . . 304
67.5 Understand the search and its limits . . . 304
67.6 Save the scenarios with their retained models . . . 305

68 Scoring discriminant and PLS models **306**

68.1 Missing predictors preserve the target population . . . 306
68.2 Interpret what the saved fit predicts . . . 306

IX Results and support **307**

69 Build and share an analysis report **308**

69.1 Assemble a readable report . . . 308
69.2 Distinguish a saved snapshot from current data . . . 308
69.3 Export for reading or further typesetting . . . 309

70 Plug-ins, help and recovery **310**

70.1 Run a trusted local analysis plug-in . . . 310
70.2 Find version, validation and licence information . . . 310
70.3 Preserve the project before attempting recovery . . . 311
70.4 Inspect, export or reproduce the recorded workflow . . . 311
70.5 Work with scaling and keyboard navigation . . . 311
70.6 Prepare a useful support request . . . 311

X Reference **313**

A Column formula reference **314**

A.1 Common rules . . . 314
A.2 Numeric formulas . . . 314
A.3 Text formulas . . . 315
A.4 Follow a calculation through missing values . . . 315
A.5 Term counting and validation . . . 316
A.6 Conditions and missing values . . . 316
A.7 Additional numeric and series functions . . . 316
A.8 Dates, text and conversion . . . 317

B Defaults and limits at a glance **318**

C Menu command index **326**

C.1 File . . . 326
C.2 Edit . . . 327
C.3 Data . . . 327
C.4 Explore . . . 328
C.5 Statistics . . . 329
C.6 Survey . . . 333
C.7 Time Series . . . 334
C.8 Quality . . . 334
C.9 Design . . . 336
C.10 Predict . . . 337
C.11 Results . . . 337
C.12 Plug-ins . . . 337
C.13 Help . . . 337

D Statistical references and example data **338**

D.1 General methods, design and reliability . . . 338
D.2 Time series . . . 338
D.3 Process monitoring and capability . . . 339
D.4 Measurement systems and method comparison . . . 339
D.5 Additional inference and modelling references . . . 340
D.6 Reproduce an example with its own data . . . 341

E Glossary **342**

List of Figures

48.1 Individuals and moving-range charts for the filling-mass example 198
48.2 One-sided CUSUM for the filling-mass example 205

Part I

Working with HypoNum

CHAPTER 1

Your first project

HypoNum brings data preparation, statistical analysis and reporting into one desktop workspace. A project holds the tables you are using, the preparation steps you have applied, and the results you choose to save. Begin with a scientific or practical question: the table is evidence for that question, not a substitute for defining it.

1.1 Start with the observational unit

Decide what one row represents before importing or combining data. It may be one component, one respondent, one production subgroup or one visit by a patient. These are different designs. Ten measurements of the same component do not provide ten independent components.

Keep an identifier when rows belong to the same person, item, batch or experimental run. For measurements over time, retain a time or occasion column as well. A column of consecutive row numbers can help locate data; it does not establish that observations are independent or equally spaced.

Write column names that describe the measurement and keep its units consistent. A temperature column should not mix Celsius and Fahrenheit. Record a group as a categorical column when its values name groups. Numeric codes such as 1, 2 and 3 do not by themselves make the differences between categories meaningful.

1.2 Import, inspect and save

A useful first session follows this sequence:

1. Choose **File > Import data...**, select the source and check the preview. Confirm column names, types, decimal marks and missing-value codes before importing.
2. Select the new dataset tab. Check the row count and inspect a few records, especially identifiers, dates and categorical labels.
3. Prepare the data needed for the question. Keep the original meaning of the observations when sorting, stacking or joining tables.
4. Run an appropriate analysis, inspect its population counts and diagnostics, and interpret the estimates in their original units.
5. Use **Add to results** when offered to retain the analysis, and **File > Save as...** to save the project.

A project file and an exported workbook serve different purposes. The project preserves your working analysis; a workbook exports table data. New-format projects embed captured original inputs by default, so reopening does not require the original files at their recorded paths. Those paths identify the sources. See section 3.5 for what a saved project contains and any separately required trusted plug-ins before moving it to another computer.

1.3 Selection, exclusion and deletion

A *selection* highlights rows for inspection or a subsequent action. Selecting a point or row does not change its value and does not, by itself, restrict every analysis to that selection. Check the population shown by the tool you are about to run.

An *exclusion* retains the row in the table while omitting it from built-in analyses and plots. It is useful when a row should remain visible in the record but is unsuitable for the current analysis. A deletion removes the row from the prepared table. Review the relevant command and its undo behavior before applying it.

Worked example

Choose **File > Import data...** and select `sample-data/manual/maintenance-four.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. The **record** column documents a maintenance fault for batch B, whose **pressure** is −999. Click the second row's header and choose **Edit > Select > Exclude selected rows**. Check that this row is marked excluded while all four rows remain in the table. The included pressure values are 10, 10.2 and 13. Record the reason in your study or report. An unusually high but valid pressure during ordinary operation is a different case: removing it merely because it is unusual could conceal the problem the analysis is intended to find.

1.4 What a preview represents

The grid, import previews and some plots display a bounded subset of a large table. A display limit does not normally limit the calculation: the operation uses the full eligible data unless its setup says otherwise. Read both the analyzed count and the displayed count when they differ.

This distinction matters when looking for rare events. A scatter plot showing a sample of points may omit an unusual observation that remains part of the fitted model. A selected subset, an excluded subset and a display sample are three different things.

1.5 Preparation history and saved results

Formula columns retain their expressions and recalculate when their inputs change. Most data-changing operations can be reversed through **Edit > Undo**. Random splits retain their seed; reproducing the same assignment also requires the same data and settings.

Analyses refer to the data as they existed when the analysis ran. If a value changes afterwards, a saved result remains the earlier result and can be marked out of date. Review and rerun the analysis deliberately when the new data should replace its evidence.

For example, an imported value of 2 may be edited to 5 before a formula doubles it to 10. Reproduction begins with the starting data and applies the edit before the formula. It does not double an already doubled column. Closing a table does not discard a prerequisite still needed by a surviving saved result.

1.6 Missing values need a reason

A blank cell means that the value is unavailable; it does not ordinarily mean zero. A code such as 999 must be declared as missing if it represents *not recorded* rather than a real measurement.

Many analyses omit rows missing one of the required measurements. Correlation can use a different complete pair of rows for each pair of columns. Other procedures refuse incomplete data or offer an explicit missing-data method. Each chapter explains the relevant rule.

Filling blanks can make a table complete without making the analysis valid. A single mean or nearest-neighbor fill does not express the uncertainty about the unobserved value. Keep the original data and consider the missing-data methods when inference depends on that uncertainty.

CHAPTER 2

Finding your way around the workspace

2.1 The active dataset

A dataset tab is a working table. Click its tab before choosing a command: most commands use that table, even when several datasets are open. A command becomes available when its requirements are met. For example, **Merge tabs...** needs two datasets and **Outlier values...** needs a numeric column. Split, concatenate, reshape and join workflows create a new tab; inspect the result there before continuing your analysis.

The grid displays a bounded part of a large table. Paging changes what you see, not which rows belong to the dataset. A calculation that uses the whole eligible table can involve more rows than appear on screen. Always read the population count in the analysis result.

2.2 Selecting, editing and inspecting

Click row-number cells to select rows. Linked plots help you locate the same observations in another view. A selection is an aid to inspection: it does not, by itself, remove observations from an analysis. Exclusion is a separate command, described in chapter 4.

Right-click a column header to create, rename, recode, reorder or calculate a column. Edit ordinary cells directly. A formula column is read-only because its values are determined by the stored formula; remove the formula before editing individual cells.

Worked example

Choose **File > Import data...** and select `sample-data/manual/maintenance-four.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Inspect the **batch**, **pressure** and **record** columns. Batch B has pressure −999 and a documented maintenance fault; batch D has pressure 13 and is recorded as a valid high pressure. Click batch B's row header and choose **Edit > Select > Exclude selected rows**. Verify that the row remains visible but excluded and that batch D remains included. To reverse the decision, select batch B's row header and choose **Edit > Select > Include selected rows**. A row that is simply unusual may be the most informative observation in the study.

2.3 Keyboard navigation

Use Tab and Shift+Tab to move between controls, Space to activate buttons or check boxes, and Escape to close most dialogs. Common Windows shortcuts are:

Action	Shortcut
Open project	Ctrl+O
Save	Ctrl+S
Save as	Ctrl+Shift+S
Undo	Ctrl+Z
Redo	Ctrl+Y

The application follows platform-standard bindings; additional bindings, including Ctrl+Shift+Z for Redo, may also be available. A shortcut follows the availability of its menu command. In particular, choose **Save as...** for a project's first save; **Save** requires an existing project path.

2.4 Display themes and printed reports

The button at the upper right of the menu bar switches between Light and Dark themes. The choice affects windows, dialogs and plots immediately and is remembered on this machine. A new user starts with Light.

Plot backgrounds, axes and series colours follow the selected theme. Saved report documents and HTML, PDF and LaTeX report exports use the light print palette. The printed manual uses black and white; distinguish plotted series by their labels and line or point styles as well as by colour on screen.

2.5 Preparing a change

For a tool that removes or fills data, first choose the method and all its options. Use its selection or preview action to inspect the affected rows, then compare the reported count with what you expected. Apply the change after resolving warnings, and review the resulting table before saving.

A preview has its own display limits. Its row count may describe the full operation even when only a few example rows are shown. Changing a setting can invalidate the preview and require a new one.

2.6 Operations, history and saved results

Choose **Operations** or **Browse operations** to inspect the recorded workflow. Select a method, complete its labelled fields, refresh the preview and choose **Apply**. The history supports keyboard selection. **Edit selected operation** changes an earlier step and preserves its relationship to the steps that depend on it.

Applying an operation pins the input table revision it consumed, including the relevant missing-value declarations and recorded formula date. Later edits to the table do not silently rewrite that completed operation's input. To use the current table, edit and apply the operation again. The resulting new calculation can then be reviewed and reproduced with the project. Older operations that lack the necessary recorded input cannot acquire an immutable reproduction claim merely by reopening them.

Changing an upstream step can make a later operation or a saved analysis stale. A stale result remains a record of the earlier calculation; it is not automatically a result for the new data. Recompute deliberately after reviewing the change.

Difference Finder and Selection Explainer require two explicitly declared populations. Their results describe differences or associations between those groups; a selected pattern does not establish causation. For a random split, retain the seed. For repeated observations or time-ordered data, check whether the chosen split keeps entities together or respects chronology. A reproducible row split alone does not prevent information from crossing the intended training and evaluation boundary.

CHAPTER 3

Importing, saving and exporting data

A reliable analysis begins with a faithful import. Before fitting a model, check a few rows against the source, confirm units and column types, and count missing values. Save the project after this check so that the analysis starts from an identifiable version of the data.

3.1 Import a text file or workbook

Choose **File > Import data...**. Supported files are Excel `.xls` and `.xlsx`, OpenDocument `.ods`, and delimited `.csv`, `.tsv` and `.txt`. Statistical and open interchange formats are described later in this chapter. Inspect the preview before choosing **Import**.

3.1.1 Delimited text

A delimiter separates fields; a decimal mark separates the integer and fractional parts of a number. They are different settings. For a file with semicolon-separated fields and decimal commas, set the delimiter to semicolon and the decimal mark to comma.

Control	Default	What to check
Delimiter	Comma	Exactly one character. Use a tab for TSV when the file type has not supplied it.
Encoding	`utf-8`	Accents and other text must display correctly. Resolve decoding errors before importing.
Decimal	Period	Exactly one character. Check a value with a fractional part.
Header row	1	One-based row containing names; range 1–100,000.
First data row	2	One-based start of observations, after the header; range 2–100,000. Intervening rows are skipped.
Missing codes	Empty	Comma-separated source tokens to treat as missing, in addition to native missing values.
Type overrides	`{}`	A JSON object assigning requested types, for example `{"amount": "float"}`.

Choose **Refresh preview** after changing options. A valid preview is required before import. A preview may stop at its byte limit, while **Import** reads the complete source. **Cancel** interrupts pending work.

Worked example

Choose **File > Import data...** and select `sample-data/manual/specimens-two.csv`. Set delimiter to comma, encoding to `utf-8`, decimal to period, header row to 1 and first data row to 2. Enter `-999` in **Missing codes** and `{"specimen":"string","concentration":"float"}` in **Type overrides**. Click **Refresh preview**, verify specimen labels 0017 and 0018,

concentration 2.5 in the first row and a blank in the second, then click **Import**. These two rows should appear in the new table. A laboratory exports specimen identifiers `0017` and `0018` alongside a concentration. Treat identifiers as text to preserve leading zeros and concentration as numeric. If `-999` means *not measured*, declare it as a missing code after confirming that it cannot be a genuine observation. Otherwise a numerical analysis could treat it as an extreme concentration.

3.1.2 Excel worksheets

Tick a worksheet's checkbox to import it. Click its name to preview it and edit its settings. These are separate actions. Each selected worksheet becomes a dataset tab, within the remaining tab capacity.

Header row, first data row and type overrides are stored separately for each worksheet. Defaults are header row 1, first data row 2 and no overrides. The first data row must follow the header. Inspect every selected sheet when a workbook uses different layouts.

Formula cells supply the values cached in the workbook. HypoNum does not recalculate Excel formulas. Recalculate and save the workbook in its spreadsheet application first when those cached values may be out of date.

3.1.3 Several files with the same layout

Choose **File > Import multiple files...** for delimited files sharing one set of options. All selected files are validated before any new tab opens. The number of files must fit within the remaining tab capacity. Use **Import data...** for Excel workbooks.

3.2 Import a database result

Choose **File > Import database...** to read a local SQLite (`.db`, `.sqlite`, `.sqlite3`) or DuckDB (`.duckdb`, `.ddb`) database. The connection is read-only.

Select a table or view, or enter one read-only query beginning with `SELECT` or `WITH`. The catalog distinguishes tables from views and shows declared columns, nullability, primary-key order and suggested types. A selected table's declared primary key is used for row matching when available. A custom query can filter or combine source tables, but must return uniquely named, non-reserved columns.

3.2.1 Parameters and types

Named parameters let you change a selection without rewriting its query. SQLite uses placeholders such as `:minimum`; DuckDB uses `$minimum`. Supply values as a JSON object, for example `{"minimum": 10}`. Values may be strings, numbers, booleans or null. DuckDB positional question-mark parameters are not supported. A missing required parameter prevents execution; unused supplied values are ignored.

Database storage and analysis types need not be identical. A date stored as text may need a conversion in **Column types**. Supported type names are `string`, `int`, `float`, `bool`, `date`, `datetime`, `time` and `binary`; use only aliases offered by the dialog. A failed conversion stops the import and leaves its settings available for correction.

3.2.2 Stable keys and refreshing

A stable key identifies the same real-world row across refreshes. Its columns must appear in the query result and be non-missing and unique over the complete result, not merely the preview. A composite key is useful when, for example, specimen and occasion together identify a row.

Without a stable key, selections and exclusions belong to the current snapshot and may reset after refresh. Tables without a primary key and views are still importable; heed the corresponding warning.

The project keeps the query settings and saved result snapshot. Opening the project uses that snapshot without contacting the database. **File > Refresh database dataset** explicitly reruns the saved query. The new data replace the snapshot only after a successful refresh; failure leaves the previous snapshot available. A successful refresh can mark saved analyses *Source changed*.

Queries have row, byte and execution-time limits. Commands that write data, attach databases, load extensions or read external files are refused. DuckDB may refuse access while another process holds a write lock. Close or coordinate that writer before retrying; a failed refresh does not establish that the source is unchanged.

3.3 Import from a database server

Choose **File > Import from database server...** for PostgreSQL or Microsoft SQL Server. This imports a snapshot of one query result. Opening a saved project reads its snapshot without reconnecting to the server.

3.3.1 Connect with the intended identity and encryption

Select the server type and enter host, port, database and user name. Ports default to 5432 for PostgreSQL and 1433 for SQL Server. Supply the connection password. **Remember the password in the system keychain** is off by default. When selected and available, it uses the operating system's credential store; otherwise the password remains only for the session and the window explains the limitation.

The connection password is excluded from projects, recipes, exported scripts, receipts and application logs. Query text itself is retained, so do not put credentials or other unintended secrets into a query.

Required (recommended) encryption encrypts the connection but does not verify the server certificate or host name. Choose **Required, and verify the server certificate** to request those identity checks as well. Your database administrator must configure the driver's trust roots. PostgreSQL uses its client root-certificate configuration; SQL Server uses Microsoft's ODBC driver's trust configuration. Turning encryption off is recorded with the dataset.

SQL Server additionally needs an installed Microsoft ODBC driver; leaving its name blank selects **ODBC Driver 18 for SQL Server**. HypoNum does not bundle that driver. After **Connect**, choose a table/view or enter a query in the database import window, preview it, review types and stable key columns, then import.

3.3.2 Understand the limits of read-only querying

Use a login granted only the reads required for your work. The application accepts one `SELECT` or `WITH…SELECT` statement and rejects write/session-changing constructs and SQL Server sequence draws. These checks restrict query syntax; they are not a database permission system.

PostgreSQL transactions are read-only, but a function or extension can still have effects outside the database. SQL Server queries run in transactions that are rolled back; effects outside rollback are not thereby prevented. Neither arrangement proves that every accepted query is free of side effects. Stored procedures, multiple result sets and server-query parameters are not supported. Use an explicit `ORDERBY` when row order matters.

The default query time limit is 30 seconds. The preview shows at most 25 rows. Row and size limits apply to the full import: an oversized result is refused, not silently truncated to a different population.

3.3.3 Preserve exact decimals deliberately

Whole numbers, floating-point values, text, logical values and supported dates keep corresponding column types. PostgreSQL zoned date-times retain their instant in UTC. Exact decimal types, including SQL Server money values, initially become text to retain every digit.

For numeric analysis, explicitly request `float` in the import types. A floating-point value carries about 15–17 significant decimal digits: `12345678901234.5678` becomes `12345678901234.568`. Keep text when every digit is essential, such as an identifier or an exact monetary record. This is a representation choice, not an inferential adjustment.

3.3.4 Save a snapshot and refresh explicitly

The dataset retains the server address, query, snapshot time, server/driver versions and encryption choice, without the password. A same-session refresh runs the query again using session or remembered credentials, or requests a new connection. A reopened database snapshot is not refreshed in place; import the query again to obtain newer rows. Compare row counts, keys and exclusions before treating a new snapshot as the same analysis population.

3.4 Find a source or a value

Import recent lists available recent sources. Selecting a text file or workbook reopens its preview so that you can check the settings. Missing sources and sources that cannot fit in the current tab capacity are disabled.

File > Search... searches the complete active dataset. Choose **Everything**, **Column names**, **Categorical columns values**, or **Numeric columns values**. **Whole value only** and **Match case** are off initially. Use Previous and Next to move through matches; pressing Enter again advances when the search settings are unchanged. Search moves the view to a matching cell. It does not filter the table.

3.5 Save and reopen a project

Choose **Save as...** for the first save and select a project path. HypoNum adds the `.hnum` extension when needed. Thereafter **Save** writes to that path. **Import project...** restores a saved project. When current work is unsaved, the confirmation offers to save all tabs, continue without saving, or cancel.

Important

New-format projects embed captured original inputs by default, along with recorded edits and retained results. Original paths and checksums identify the source; reopening does not require those paths to remain available. Database data are retained query snapshots. Keep the project itself when transferring work, and retain any separately required trusted plug-ins and compatible software environment for an explicit reproduction run.

The project also retains operations, exclusions, formulas and saved analysis or report results. On opening, HypoNum validates the saved package and its retained inputs. It restores saved results without rerunning analyses or executing arbitrary embedded Python. Refreshing a database remains a separate action.

Close project closes the entire workspace. Closing one dataset tab is a different action, but saving from its confirmation saves the complete project, not just that tab.

3.6 Export a workbook

File > Export Excel workbook... writes the active tab or all open tabs. Decide explicitly whether to include user-excluded rows. Numeric, Boolean, date and time cells retain their types; the export freezes headers and adds filters. Text that begins like an Excel formula is kept as literal text.

An exported workbook is a data delivery, not a replacement for the project. Keep the project when the recipient needs exclusions, recorded operations or retained analysis results.

3.7 Import statistical data files

File > Import data... also reads SPSS `.sav`, `.zsav` and `.por`; Stata `.dta`; SAS `.sas7bdat` and transport `.xpt` files. A SAS `.sas7bcat` format catalogue can accompany a SAS dataset. These imports do not require the originating application.

Inspect the **Variables** and **Data preview** tabs before importing. They show variable types, labels, missing-value definitions, measurement levels and the first observations. A bounded preview of a wide file is not the whole table. The summary identifies the file format/version and whether the encoding is declared by the file or assumed from your choice. Check accented text carefully: decoding successfully does not prove that an assumed encoding is correct. Invalid byte sequences are refused, with a location, rather than silently replaced.

3.7.1 Preserve missingness and its meaning

System-missing values become empty cells. SPSS user-missing values normally retain their stored codes and become the column's declared missing values; analyses treat them as missing. For example, a retained value 99 can continue to mean "not asked" rather than an observed score of 99.

Stata extended missing values `.a`–`.z` and SAS special missing values `.A`–`.Z` and `._` become empty measurements plus a linked text column, such as `income_missing_code`. That column keeps each row's original missing code and its available labels. Ordinary system-missing rows have no special code. The reasons therefore remain distinguishable without introducing artificial numeric measurements.

Convert user-missing values and missing-value codes to missing is off by default. Turning it on discards those distinctions; the import notes identify and count the affected values. SAS transport imports do not expose special missing codes, and SPSS portable imports do not expose user-missing declarations. Read the preview's limitation notes before accepting such a file.

3.7.2 Check types, labels and dates

Variable labels, value labels, measurement levels and display settings become column properties in one recorded edit. Declared levels are distinguished from inferred ones. Each column gets its own label copy: editing one does not edit other columns that shared a source label set. The

import record preserves the original definitions. A column copy is limited to 5,000 value labels and 256 characters per label, with counted shortening or omission notices.

Stata integer storage becomes integer columns. SPSS and SAS numeric storage normally becomes decimal-number columns. The SPSS option to store F-format whole-number columns as integers is off by default and requires every stored value to be integral. A display with zero decimals alone does not prove that the underlying values have no fractional parts.

Recognized date/time formats are converted using their package's origin: SPSS counts seconds from 14 October 1582; Stata and SAS use 1 January 1960 with units determined by the format. For example, 21,915 SAS or Stata date days mean 1 January 2020. Durations can exceed one day. Stata's leap-second-counting `\%tC` remains a numeric millisecond column with an explicit note; it is not silently interpreted as ordinary `\%tc`.

Original display formats and label sets are retained as provenance. They cease to apply after an incompatible type, format or label edit, while the historical record remains. Saving and reopening keep both the current properties and this record without needing the original source file.

3.8 Import open interchange formats

The same **File > Import data...** command reads the following formats. OpenDocument `.ods` workbooks use the worksheet-selection workflow described above; their formula cells supply stored values, without recalculation.

Format	Supported tabular content
Parquet and Arrow IPC	`.parquet`, `.pq`, `.arrow`, `.feather` (version 2), `.ipc`: one rectangular table.
R	`.rds`, `.RData`, `.rda`: a data frame; choose the object when the file contains several.
MATLAB	`.mat`: a two-dimensional numeric, logical or homogeneous-column cell array, a structure of equal-length columns, or a supported structure array.
HDF5	`.h5`, `.hdf5`: a one-dimensional compound dataset or a group of equal-length one-dimensional datasets.

The object list includes unavailable objects with their reasons. Choose an object and inspect **Columns**, **Data preview** and the conversion notes. For a MATLAB array, declare whether rows or columns are observations; rows are the default. Use generated names or a compatible names variable. For an R data frame with nonautomatic row names, the default adds a `row_names` column; choose its name deliberately if identifiers matter. Saved projects reopen their retained data; replay uses the recorded object selection and options.

R factors normally become integer level codes with value labels, including unused levels; ordered factors are ordinal. If all levels cannot be represented within the label constraints, the factor becomes text with a note. R ordinary `NA` and `NaN` become missing; infinities remain non-finite values. R date-times become UTC instants and the original time-zone declaration is noted. Parquet/Arrow decimals become floating-point values with a precision warning, and categorical columns become plain text. These conversions are reasons to review the preview, not to assume that all source semantics survive.

Nested, list, complex, sparse or nonrectangular content is not automatically flattened into invented observations. Unsupported objects are refused with conversion guidance. HDF5 external/soft links and external data are not followed. Encrypted Parquet, Feather version 1 and Arrow streams are not supported. The row ceiling is 1,000,000; R whole-file decoding and some MATLAB reads also have a 512 MiB budget. A format's own stricter limit still applies.

3.8.1 Tagged missing values from R

Haven-labelled R data frames may distinguish `NA(a)`, `NA(b)` and other tagged reasons. The measurement becomes missing while a linked nominal text column retains the exact printed tag. For example:

Source score	Imported score	`score_missing_code`
1	1	empty
`NA(a)`	missing	`NA(a)`
3	3	empty
ordinary `NA`	missing	empty
`NA(b)`	missing	`NA(b)`

One printable ASCII character is supported, with case and spaces significant. Nonprintable tags, incompatible tagged-label column types and conflicting labels for one tag are refused. Labels are compared in full; the import record keeps full labels even when the displayed copy is shortened to 256 characters. Labels for unused tags are retained and identified in the notes.

Sorting, filtering, saving and replay keep a row's reason with its measurement. Deleting the linked column loses the row-level reasons. SPSS or SAS export can retain R tags as that separate text column beside missing measurements. Stata export refuses R tags, including tags present only in labels: R tags are not Stata extended-missing codes. An intentional deletion of the code column permits system-missing export but loses those distinctions.

3.9 Exchange files without promising lossless round trips

File > Export data... writes the active dataset as SPSS system, compressed-system or portable data; Stata versions 114, 117 or 118; or SAS transport versions 5 or 8. It does not write SAS `.sas7bdat` files. The default Stata version is 118 and the default SAS transport version is 5.

Before choosing a destination, inspect **Variables** for renamed columns and written types, and **Notes** for refusals and conversions. Excluded rows are omitted unless **Include rows excluded from analyses** is selected. Export reads stored values, including retained user-missing codes; it does not export an analysis view in which those codes have already become nulls. The option to write user-missing values as system-missing when the format cannot declare them is off by default.

Target	Principal information to check
SPSS system files	Variable/value labels, measurement levels and supported user-missing declarations; package-specific name and label lengths still apply.

Target	Principal information to check
Stata	Value labels require eligible whole-number codes. Linked Stata missing codes fold back into `.a`–`.z`. R tags and SAS `._` are refused; SAS letter codes require the disclosed mapping.
SAS transport	No value labels or user-missing declarations. Version 5 has eight-character names, 40-character variable labels and 200-byte text limits.
SPSS portable	Eight-character names, ASCII text and limited string width; value labels, user-missing declarations and measurement levels are not written.

Invalid or long names are mapped and listed; uniqueness is enforced without regard to case. Label shortening is disclosed. Infinities and integers beyond the exact double-precision range are refused. Text missingness becomes empty text, and NaN becomes system-missing, with notices. Stata date-times retain whole milliseconds; finer precision is lost. Zoned date-times are exported as UTC instants. Formats lacking an encoding declaration need care when another reader assumes a different encoding.

A completed export records its filename, format, size, checksum, column mapping and notices in the project. Undo removes that record, not the already written external file. Reproducing the project checks the recorded dataset state and returns the export record; it does not overwrite or recreate that external file. A cancelled or failed export does not publish a partial file under the chosen name.

For a simple illustration, a measurement with stored values 1, 99 and 3 and declared user-missing value 99 can keep that declaration in SPSS system data. For a target unable to declare it, export initially refuses. Explicitly choosing system-missing conversion writes a missing measurement at that row and records the loss of the original code. Neither choice should be mistaken for an ordinary observed value of 99.

File > Export Parquet file... is a separate table export. It keeps supported column values/types but does not carry the project's column-property definitions or row identities. A `.hnum` project remains the format for preserving the full desktop workflow and its results.

CHAPTER 4

Editing values and columns

Use editing to correct a known value or construct a documented variable. Keep the distinction between a correction, an analytical exclusion and a transformation: they answer different questions about the data.

4.1 Undo, redo and row selection

Undo reverses the latest undoable project change; **Redo** reapplies a change that was undone. Each command is disabled when its history is empty. They do not reverse edits made outside HypoNum to an external file.

The **Edit** selection commands act on the active table:

Invert selection
Select the unselected rows and clear the selected rows.

Exclude selected rows
Keep the observations in the table but omit them from plots and analyses that respect user exclusions.

Include selected rows
Restore previously excluded observations.

Delete selected rows
Remove those observations from the prepared table.

Delete selected columns
Remove the selected column or columns. A protected dependency can prevent deletion.

Selection alone does not change inclusion. Before reporting an analysis, check its exclusion count and the scientific reason for each exclusion.

4.2 Create and name a column

Right-click a column header and choose **New column...**. The new column appears immediately before that column, with missing values. Its name must be non-empty and unique. Choose **Continuous numeric** for measured quantities, **Fixed numeric** for discrete numeric levels, or **Categorical** for labels and groups. Continuous numeric is the default.

Rename column... changes one name and updates stored formula references. For a batch change, choose **Edit > Rename columns...**, enter a regular expression in **Pattern** and its **Replace with** text, then inspect the old-to-new preview. **Ignore case** is optional. A large preview may show only part of the list, although the match count is complete.

Names are changed simultaneously. An invalid pattern, empty or duplicate resulting name, or a pattern with no matches prevents the operation. The internal row identifier is not renamed. Formula references change with the names; a plot using a renamed axis is cleared so that it cannot continue to display an obsolete column reference.

4.3 Calculate a column with a formula

Choose **Formula...** from the target column's context menu. The **Numeric** and **Text** tabs provide compatible-column insertion, syntax checking and a preview of eight current and calculated values. **Apply formula** replaces the target values throughout the table and retains the formula for recalculation. The formula language and its functions are described in appendix A.

Check units before applying a calculation. A conversion from millimetres to centimetres divides by 10; subtracting a target from a measurement produces a deviation in the original units. Neither operation changes the observational unit represented by a row.

Formula columns cannot be edited cell by cell. **Remove formula** retains the latest calculated values and makes those cells editable. This changes the column from a maintained calculation into ordinary stored data.

4.4 Recode categories

Recode values... is available for categorical and fixed-numeric columns without formulas. Assign a target value to each distinct existing value. Several values assigned to the same target are combined.

Worked example

Choose **File > Import data...** and select `sample-data/manual/plant-labels-three.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Right-click the **plant** column header and choose **Recode values...**. In the **New value** fields, enter **North** for each of the existing values North, NORTH and N; finish each edit by pressing Tab. Enable **Recode into a new column**, enter `plant_clean` as its name, and click **Apply recoding**. Inspect the table: the original **plant** column retains all three spellings and all three **plant_clean** cells contain North. A source uses `North`, `NORTH` and `N` for the same plant. After checking their meaning, map all three to `North`. Choose **Recode into a new column** to retain the original labels alongside the cleaned labels. Compare group counts before using the new column in an analysis.

The new-column option creates a text column immediately after the original. Its first suggested name is, for example, `Plant_Type_1`. If that name already exists, the suggestions are `Plant_Type_1_1`, `Plant_Type_1_2`, and so on. Edit the name before applying. Formulas and plots using the original column continue to use that original data.

4.5 Column order and row order

Change order... moves a column to the first or last position, or before or after another named column. The first position is the default. This changes column order only.

Sort changes row order using the selected column. Text categories offer **A to Z** and **Z to A**; other orderable columns offer **Ascending** and **Descending**. Sorting a wholly missing column leaves its row order unchanged. In a time-series or process-monitoring study, confirm that the resulting order still represents time before running an order-dependent analysis.

4.6 Direct edits and filling downward

A directly entered value is parsed using the column's type. An incompatible entry produces **Value needs attention**; it does not silently change the type.

Right-click a loaded non-empty cell to choose **Fill column below with '<value>'**. Dependent formulas recalculate. The command is unavailable for formula columns, empty cells and the last row.

Important

Fill downward immediately overwrites existing and missing values from the clicked row through the final row, including rows beyond the visible page, without confirmation. Check the starting row and extent first. For selective missing-data treatment, use chapter 6.

CHAPTER 5

Column meaning, case weights and analysis groups

The stored values, their labels and the population used for an analysis are different things. A code 99 can be stored, displayed as “No answer” and treated as missing by an analysis. Review all three before interpreting a summary or exporting the data.

5.1 Describe a column

Select a column and choose **Data > Column properties...**. Set its variable label, typed value labels, measurement level, displayed decimals and any user-missing codes. A declared level can be nominal, ordinal or scale. With no declaration, numeric storage is inferred as scale and other storage as nominal; the inference does not establish scientific measurement properties. Numeric category codes are not necessarily distances.

A column accepts at most 5,000 value labels, with labels up to 256 characters. User-missing definitions allow three individual codes, or a numeric closed range plus one code. They preserve stored values. A valid value entered later is not made missing unless it matches the current declaration.

Property edits are recorded and undoable. Renaming preserves properties; deleting the column removes them. A type change retains only compatible properties. When an operation creates a new table, unchanged columns keep their source properties as they were when the operation ran. This includes labels and user-missing declarations, which can affect the analysis population. Newly computed or replaced columns start without properties.

Splitting retains each column’s properties. In a merge, carried columns keep their own source properties, including a renamed right-hand column; a combined key uses the left table’s properties and reports a conflict when they differ. Concatenation uses the first input table containing each column and reports conflicting definitions. If the combined column changes type, incompatible value labels and missing codes are removed with a note.

When stacking, identifier columns keep their properties. The stacked value column keeps properties only when all stacked columns agree; otherwise it starts without them and the result reports the difference. When pivoting, each new value column inherits the original value column’s properties. Group summaries keep grouping-column properties; calculated summaries start without them. Check these notes and the resulting columns before analysing.

Older saved operations retain the original behaviour: their derived tables do not acquire properties merely because the project is reopened or replayed. To use the current rules, create and review a new operation.

Edit > Show value labels changes the grid display for the current table. Calculations, editing, search and sorting still use stored values. Header badges **L** and **M** identify value labels and user-missing codes. Analysis results count user-missing values separately from ordinary empty cells.

Frequency and crosstab procedures can explicitly count user-missing codes as valid categories; that choice is retained with the result.

5.2 Declare frequency weights

Data > Weight cases... declares an integer frequency-weight column. A row of weight w represents w identical cases. Fractional weights are refused; zero weights contribute nothing; negative, missing or non-finite weights are omitted and counted. A banner shows the active declaration.

Supported procedures compute the same result as explicit case replication. For values 2 and 5 with weights 3 and 1, the represented cases are $(2, 2, 2, 5)$, giving mean 2.75 and sample variance 2.25. A one-sample t calculation has three degrees of freedom, not one. This is appropriate only if the weights really count independent cases. Survey sampling weights do not have that meaning; use the Survey methods for design-based inference.

Frequencies, crosstabs, the hypothesis-test window, correlation and other explicitly supported table procedures honour this declaration. The result identifies whether it ran weighted, the weight column, retained rows and represented case total. The total before a method's own missing-value omissions need not equal its final analysis n.

An unsupported procedure is refused unless the declaration explicitly allows it to run unweighted; the resulting output then states that choice. The expanded case total cannot exceed 1,000,000. This protects expansion size and is not a promise that every rank or resampling procedure is fast at that limit. Save the project to preserve the declaration and result receipts.

5.3 Run the same settings by group

Data > Split file... declares one or two categorical columns for separate analyses with identical settings, up to 50 groups. It does not physically create separate dataset tabs. Group summaries precede stacked results. A group that cannot support the chosen procedure retains its error message while other groups continue; an unsupported procedure is refused.

Groups use typed values. Missing split values form their own group and do not merge with a literal text value such as "(missing)". With two grouping columns, category pairs remain distinct even when their printed labels contain separators. Each group's result preserves its own population and settings for saving and replay. Check both the split-file and weight banners when a result's population is unexpected.

5.4 Define a multiple-response question

Data > Define multiple response sets... names the columns of a check-all-that-apply question. A multiple-dichotomy set counts a chosen value, such as 1 or Yes, in each column. A multiple-category set pools values across columns sharing a label scheme. Renaming updates a set's column references; deleting a required column removes the affected set.

Use **Statistics > Descriptives & tables > Multiple response frequencies...** or **Statistics > Descriptives & tables > Multiple response crosstab...** after defining the set. The case base consists of rows with at least one counted response. A row answering No to every dichotomy is outside that base. Rows with no counted response and rows missing all responses are counted separately.

Worked example

Choose **File > Import data...** and select `sample-data/manual/response-slots-two.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding,

decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Data > Define multiple response sets...**. Enter `choices` in **Name**, choose **Multiple category** in **Type**, and tick only **first** and **second** under **Columns**. Click **Save set**, then **Close**. Choose **Statistics > Descriptives & tables > Multiple response frequencies...**, set **Response set** to **choices**, and click **Run analysis**. Inspect the response counts and both percentage columns. The two response slots contain (A, A) in the first case and (A, B) in the second. There are four responses: three A and one B. Response percentages are therefore 75% and 25%. A is selected by two of two valid cases (100%) and B by one of two (50%). Case percentages total 150% because categories overlap; neither category exceeds 100% of cases.

Crosstabs distinguish deduplicated case pairs from occurrence pairs. On a response base, repetitions multiply the paired response count; on a case base, a case counts once for each category pair. The denominators match the selected base. Frequency weights and split file are supported. No ordinary chi-square test is supplied because a case can contribute to several cells and the cell counts are dependent.

CHAPTER 6

Preparing and combining tables

The **Data** commands help you review questionable records and organize tables for analysis. Cleaning changes the active table; split, concatenate, reshape and merge commands create new tabs and preserve their input tables.

6.1 Choose the condition before the action

Open **Data > Remove duplicates...**, **Data > Missing values...** or **Data > Outlier values...**. These dialogs follow the same sequence: choose the **Columns**, specify the **Definition** for an outlier check, choose a **Selection method**, then choose an **Action**.

For duplicates and missing values, the column choices are **All columns**, **Numeric columns only** and **Chosen columns**. Outlier checks offer **All numeric columns** or **Chosen numeric columns**. Use the column selector to add the particular columns you want to check.

The live summary reports how many included rows match and how many excluded rows were left out. Column lists in messages show at most five names, followed by "and N more" when additional columns are affected. Read the action explanation, then use **Select matching rows** to highlight the matches without changing data or exclusions. The primary button applies the chosen action to the matching rows; an existing grid selection does not define the population for the check.

Exclude rows
Keep the rows in the table but omit them from analyses, as with **Edit > Select > Exclude selected rows**.

Remove rows
Delete the matching rows from the table.

Fill blank cells
Available for missing values. Fill only the eligible blank cells in the chosen columns of matching rows.

Replace cells with blanks
Available for outliers. Clear only the outlying cells in the chosen columns of matching rows.

These changes can be undone and are recorded for project reproduction. The dialogs initially offer **Remove rows**; inspect the condition and choose the action deliberately before applying it.

Important

Previously excluded rows do not participate in these cleaning commands. They cannot match a condition, serve as the retained row of a duplicate group, supply a fill value or contribute to an outlier limit. Their cells are left unchanged. Include a row again first if it should participate.

6.2 Decide what a duplicate means

Two equal values are not necessarily duplicate observations. Repeated measurements from the same participant may be essential to the study, whereas two copies of the same laboratory record may be an error. Choose columns that identify an observation at the intended level.

Whole row repeats another row
Compare the chosen columns together as a single key. A row matches if that combination repeats another row retained under the **Row to keep** rule. This is the opening selection method. Despite its name, it compares only the chosen columns; choose **All columns** to compare complete records. Repeated matching missing entries can form part of a duplicate key.

Any chosen column has a repeated value
Check each chosen column separately. A row matches if, in at least one column, its value repeats a value retained in another row. Blank cells never count as repeated values under this method.

Row to keep chooses the first row, the last row, or the row with the lowest or highest value in an order column. The default is first. Lowest and highest require a numeric or date/time order column; ties retain the earliest source row. With the per-column method, this decision is made separately in each column. A row retained for one column can still match as a repeat in another column.

All column checks use the same original included rows. Removing a match in one column does not cause the other columns to be checked again, so changing the order of the chosen columns does not change the result.

Worked example

Choose **File > Import data...** and select `sample-data/manual/duplicates-three.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Data > Remove duplicates...**. Set **Columns** to **Chosen columns** and add only **A** and **B** to the chosen list. The imported records, in order, are:

Record	A	B
1	1	x
2	1	y
3	2	y

Set **Selection method** to **Any chosen column has a repeated value**, **Row to keep** to **Keep first remaining row**, and **Action** to **Remove rows**. Click **Select matching rows** to inspect records 2 and 3, then click **Remove rows**. Column A flags record 2; column B flags record 3. Both records match. **Remove rows** therefore leaves only record 1, even though record 2 was the first occurrence of y in column B. To compare the other rule, import the CSV again, reopen **Data > Remove duplicates...**, choose only A and B, and set **Selection method** to **Whole row repeats another row**. The match count is zero because all three pairs differ.

Worked example

Choose **File > Import data...** and select `sample-data/manual/replacement-two.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Specimen S1 has results recorded at hour 9 and hour 11. For this exercise, suppose the hour-11 row is a documented correction of the first row. Choose **Data > Remove duplicates...**, **Chosen columns**, and add only **specimen**. Set **Selection method** to **Whole row repeats another row**, **Row to keep** to **Keep row with highest value in column**, and **Order column** to **hour**. Set **Action** to **Remove rows** and click **Remove rows**. The remaining row is S1, hour 11, value 2.8. If both are valid measurements at different occasions, do not deduplicate on specimen alone. If the second row is a corrected replacement and recording time establishes that fact, choose the specimen column as the key and keep the row with the highest recording time. The software cannot decide which interpretation is correct.

Use the per-column method cautiously with a low-cardinality column, such as shift or machine: many legitimate observations share that value. Inspect the overall match count and the highlighted records before excluding or removing them.

6.3 Handle missing values deliberately

In **Missing values...**, a blank means a missing cell (null or a numeric NaN). A text string or a code such as 999 does not become missing merely because it represents an unknown value in your study.

The default selection method is **All chosen columns are blank**; a row must be blank in every chosen column. **Any chosen column is blank** matches a row with at least one blank among those columns. Choose whether to exclude the matching rows, remove them, or fill their blank cells.

6.3.1 Which cells are filled

Fill blank cells changes only blank cells of the chosen columns in matching included rows. Observed cells and unchosen columns are preserved. With the **All** selection method, a partially observed row does not match, so even its blank cells remain unchanged. With **Any**, such a row matches, but only its blank cells are candidates for filling.

The fill value comes from observed values in included rows, including rows that do not themselves match the blank-selection condition. Excluded rows never supply values.

Worked example

Choose **File > Import data...** and select `sample-data/manual/missing-exclusion-four.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Click the row header of record 3, whose **value** is 100, and choose **Edit > Select > Exclude selected rows**. Open **Data > Missing values...**. Choose **Chosen columns** and add only **value**; set **Selection method** to **Any chosen column is blank**, **Action** to **Fill blank cells**, and the fill method to **Using the mean**. Click **Fill blank cells**. The included values are $2, 4$ and a blank; record 3 remains excluded. The blank receives $(2 + 4)/2 = 3$. Including the excluded value would have given $106/3 \approx 35.33$, but it does not participate. A blank in an excluded row would remain blank.

6.3.2 Fill methods and column types

Fill method	Applicable columns and calculation
Median (default)	Numeric, date, date-time, time and duration columns; use the middle observed value, or the midpoint of the two middle values.
Mean	The same column types as median; use their arithmetic average.
Most frequent value	Any column type; use its most frequent observed value. If frequencies tie, use the value encountered first.
4 nearest neighbors	Numeric columns; average values supplied by up to four similar rows, as described below.

A method never substitutes another method for an unsupported column type. For example, median and mean leave text cells blank; choose **Using the most frequent value** explicitly if that is the intended rule. The dialog names columns that the selected method cannot fill. A column with no observed included value remains blank.

Filling preserves each column's data type. Whole-number results are rounded to the nearest integer, with halfway values rounded away from zero. Dates are rounded to the nearest day; a midpoint between two days becomes the later date. The dialog explains this rounding. Other numeric and temporal results are stored at the precision supported by their column type.

Worked example

Choose **File > Import data...** and select `sample-data/manual/missing-rounding-four.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. For each calculation below, open **Data > Missing values...**, choose **Chosen columns**, and add only the named column. Set **Selection method** to **Any chosen column is blank** and **Action** to **Fill blank cells**. Choose the specified fill method, then click **Fill blank cells** and inspect the formerly blank cells:

- **mean_integer**: **Using the mean**; replacement 3.
- **median_integer**: **Using the median**; both replacements 3.
- **median_negative**: **Using the median**; both replacements -3.
- **mean_fractional**: **Using the mean**; replacement $16/3$.
- **mean_whole**: **Using the mean**; replacement 5.

Clear the previously chosen column when moving to the next calculation. For the whole-number column containing $2, 3, 5$ and a blank, the mean is $10/3$, so the replacement is 3. The median of 2 and 3 is 2.5, which also becomes 3; a midpoint of -2.5 becomes -3. These are rounded replacements, not changes to the observed cells.
For a column that allows fractional values, $2, 4, 10$, blank gives a mean fill of $16/3 \approx 5.3333$ or a median fill of 4. The same mean in a whole-number column would be stored as 5.

6.3.3 Choose what makes neighboring rows similar

Using the 4 nearest neighbors estimates each blank numeric cell from other rows. The **Compare rows using** line initially shows **All numeric columns**. Use **Choose columns...** to select numeric comparison variables that have a meaningful relationship to the quantity being filled. The target column itself does not contribute to similarity.

Comparison variables are centered on their medians and scaled by their interquartile ranges, using included rows. A constant or effectively zero scale is replaced by one. Distances use the squared standardized differences in the comparison variables observed in both rows, with an adjustment for the number of variables shared. Thus a variable measured in large units does not dominate solely because of its units, although the choice of variables still matters.

An eligible donor is an included row with an observed target value and at least one comparison value in common with the row being filled. Up to four nearest eligible donors contribute, with equal weight. If fewer than four can be compared, use those available. Replacements are returned to the target column's original units and then rounded if its type requires it. Values filled during this operation do not become new donor observations for other cells in the same operation.

If no donor can be compared, the cell stays blank and the message reports the unfilled cells. There is no fallback to the column mean. In particular, choosing only the target column supplies no comparison variables.

Worked example

Choose **File > Import data...** and select `sample-data/manual/neighbors-seven.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Six observed rows contain the following measurements:

Record	Height	Age	Weight
1	180	20	1.0
2	178	60	2.0
3	160	40	3.0
4	182	50	4.0
5	150	41	5.0
6	100	39	6.0

A seventh row has Height 179, Age 41 and a blank Weight. The file stores Weight with fractional values. Open **Data > Missing values...**, choose **Chosen columns** and add only **Weight**. Set **Selection method** to **Any chosen column is blank**, **Action** to **Fill blank cells**, and fill method to **Using the 4 nearest neighbors**. Beside **Compare rows using**, click **Choose columns...** and add only **Height** as a comparison column. Click **Fill blank cells**; the seventh Weight becomes 2.5. For the Age comparison, import the original CSV again and repeat these steps, choosing only **Age** as the comparison column. The seventh Weight then becomes 4.5.
Using Height alone selects records 1, 2, 4 and 3, giving $(1 + 2 + 4 + 3)/4 = 2.5$. Using Age alone selects records 5, 3, 6 and 4, giving $(5 + 3 + 6 + 4)/4 = 4.5$. A whole-number Weight column would instead store 3 and 5, respectively.

These are preparation operations, not a general solution to uncertainty from missing data. Filling a cell does not create another independent measurement. Constant replacements can reduce

apparent variation and distort correlations. Record which cells were filled and why the method is reasonable. For prediction, estimate preprocessing from training data when the workflow supports it; using evaluation data to construct replacements can make predictive accuracy look too good. Multiple imputation is a separate analysis with different assumptions and uncertainty calculations.

Note

When repeating work from an earlier version, distinguish a newly applied fill from reproduction of an already recorded fill. Current fills use only included rows, leave unsupported text columns blank, preserve whole-number and date types, and leave cells without comparable neighbors unfilled. Earlier fills could use excluded observations, substitute the most frequent text value, promote a column's type, or substitute a column mean when no neighbor could be compared. Reproducing an older recorded fill retains its earlier behavior; applying the dialog again uses the current rules.

6.4 Inspect potential outliers

In **Outlier values...**, choose the numeric columns, a definition and a positive threshold factor k. The default is **Median / MAD** with $k = 3.5$. One definition and factor apply to all chosen columns, but each column's limits use its own included finite values. Run separate checks if different columns need different rules. Null, NaN and infinity neither contribute to a limit nor receive an outlier flag.

Rule	Default k	Flag when
Median / MAD	3.5	$\lvert 0.6745(x - \tilde{x})/\mathrm{MAD}\rvert > k$, with MAD = median $\lvert x - \tilde{x}\rvert$. Zero MAD gives no flags.
Interquartile range	1.5	$x < Q_1 - k(Q_3 - Q_1)$ or $x > Q_3 + k(Q_3 - Q_1)$.
Mean / standard deviation	3	$\lvert x - \bar{x}\rvert > ks$, with sample standard deviation s.
Absolute value from zero	5	$\lvert x\rvert > k\lvert T\rvert$, where T is the mean, median (default) or sample standard deviation.

A value exactly on a boundary is retained. Quartiles use linear interpolation between ordered values: percentile p has one-based position $1 + (n-1)p$. The standard deviation uses denominator $n - 1$. Zero MAD gives no flags and is named in the dialog; this often occurs in rounded or discrete data and does not establish the absence of unusual observations. A column with no finite values, or too few values for the selected statistic, cannot supply a classification.

Any chosen column has an outlier, the default selection method, matches a row if at least one chosen column flags it. **All chosen columns have an outlier** requires every chosen column to flag the row. A blank or non-finite cell therefore prevents that row from matching **All**. A column that cannot be classified also supplies no flags.

Choose **Exclude rows**, **Remove rows** or **Replace cells with blanks**. The last action preserves the row and its ordinary values, clearing only the outlying cells in chosen columns of matching rows. Clearing creates missing data; it does not establish what the correct replacement should be.

Worked example

Choose **File > Import data...** and select `sample-data/manual/outliers-eight.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Open **Data > Outlier values...**, choose **Chosen numeric columns** and add **Height** and **Weight**. Set the definition to **Interquartile range**, factor to 1.5, **Selection method** to **Any chosen column has an outlier**, and **Action** to **Replace cells with blanks**. Click **Select matching rows** to inspect the first two rows, then click **Replace cells with blanks**. The original eight records have Height values $250, 250, 175, 176, 174, 175, 177, 173$ and Weight values $80, 300, 78, 79, 81, 80, 82, 78$, in that order. Choose both columns and the interquartile-range rule with $k = 1.5$.

For Height, $Q_1 = 174.75$, $Q_3 = 195.25$, so the fences are 144 and 226. For Weight, the quartiles are 78.75 and 81.25, giving fences of 75 and 85. The first two heights and the second weight are flagged.

With **Any chosen column has an outlier**, replacing cells with blanks clears those three cells; record 1 keeps its Weight of 80. Import the original CSV again and repeat the same settings except choose **All chosen columns have an outlier**; click **Replace cells with blanks**. Only record 2 matches and only its two cells are cleared. Record 1 remains unchanged.

Checking many columns with the **Any** rule can flag ordinary rows by chance. These rules identify values for investigation, not a multiple-test procedure with a stated overall false-alarm rate.

An unusual value may reveal an instrument failure, a mixture of populations, an omitted explanatory variable, or a real rare event. Keep the original data and compare conclusions with and without a justified exclusion when that sensitivity matters.

6.5 Split a table

Split tab... previews the proposed output tabs and row counts. **Create tabs** creates them without changing the source.

Method	Result and settings
Selected and remaining rows	Two tabs; available when some but not all rows are selected. This is the opening choice when a selection exists.
Categorical values	One tab per distinct value and a separate tab for missing values, up to eight output tabs. Default without a selection.
Numeric split value	Values below a finite threshold in one tab, values at or above it in another, and, only when missing rows exist, missing values in a third.
Random sample and remainder	Choose a percentage strictly between 0 and 100 (default 50), or a positive row count smaller than the table. The non-negative integer seed defaults to 1729. Each output keeps original row order.

A random row split is suitable only when rows can be split independently for the intended purpose. It is not a substitute for keeping a participant's repeated observations together or reserving later times for a forecast evaluation.

6.6 Concatenate tables

Use **Concatenate tabs...** to append observations with compatible meanings. Select at least two tabs. **Select all** and **Clear** change the input list.

Common columns only, the default, retains names present in every input; column order follows the first selected tab. **All columns** retains names in first-appearance order and inserts missing values where an input lacks a column. Rows follow selected-tab order. Inspect the projected row and column counts, then choose **Create tab**.

Before combining monthly files, check units and definitions as well as names. Two columns both called `temperature` cannot safely be combined if one contains degrees Celsius and the other degrees Fahrenheit.

6.7 Choose a merge relationship

Merge tabs... combines different columns by matching identifiers. The opening mode, **Legacy first-match**, is a one-key inner match: only identifiers present on both sides appear, and a repeated identifier uses its first occurrence. Choose the left and right tabs and their identifier columns. Missing keys do not match, incompatible types fail, and colliding non-key names receive suffixes. Output follows the retained left-side order.

The preview reports matches, unmatched rows, output columns and duplicate warnings. **Auto-identify** proposes a key pair using name similarity, uniqueness, overlap and missingness. Its confidence value measures the suggestion, not whether the keys denote the same entity. Confirm that meaning before choosing **Create tab**. To change the keeper for a duplicate key, prepare the input with **Remove duplicates...** first.

Use **General join** when a left or full join, multiple keys or an explicit one-to-many relationship is required. The next sections show why that choice changes which observations survive.

6.8 Stack and pivot measurements

A wide table places repeated measurements in separate columns. A long table places them in one value column and records the occasion or measurement name in another column. Choose the arrangement that the intended analysis requires; reshaping changes layout, not the meaning of an observation.

6.8.1 Worked example: two visits

Choose **File > Import data...** and select `sample-data/manual/visits-three.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. The imported wide table is shown below. A dash denotes a missing value.

Person	Before	After
A	10	12
B	8	9
C	11	—

Choose **Data > Stack measurements**, select Person as the identifier and Before followed by After as the measurements. Keep `measurement` and `value` as the generated column names. Click **Preview**,

check the projected six rows, then click **Create new tab**. Inspect the new **Stacked dataset** tab. The six output rows are:

Person	measurement	value
A	Before	10
B	Before	8
C	Before	11
A	After	12
B	After	9
C	After	—

Stacking retains a row for a missing measurement. Output follows the selected measurement order and, within each measurement, the source row order. Active user exclusions are omitted and counted.

6.8.2 Stack settings

Select one or more identifier columns and at least two measurement columns using the available and selected lists. They start empty. Measurements must have compatible types: numeric columns need a lossless common numeric representation, while text columns need a compatible text family. Mixed text and numeric measurements are refused.

The output tab defaults to **Stacked dataset**. Change its name and generated column names if needed. Column roles must not overlap, and output names must be unique and must not use reserved internal names.

6.8.3 Pivot back to wide form

To reproduce the two-visit example from this subsection alone:

Choose **File > Import data...** and select `sample-data/manual/visits-three.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Open **Data > Stack measurements**, add **Person** as identifier and **Before**, then **After**, as measurements. Keep generated names **measurement** and **value**. Click **Preview**, then **Create new tab**. With **Stacked dataset** active, open **Data > Pivot measurements**, add **Person** as identifier, choose **measurement** as the measurement-name column and **value** as the value column, and enter `Before,After` as the measurement names. Click **Preview**, then **Create new tab**. The three output rows are A: 10, 12; B: 8, 9; C: 11, blank, in columns Before and After.

Choose **Data > Pivot measurements**. Select identifiers, the measurement-name column and the value column. Enter the measurement names as a comma-separated list in the desired output order, for example `Before, After`. Surrounding spaces and empty list entries are removed; this entry control cannot represent a measurement name containing a comma. The output tab defaults to **Pivoted dataset**.

Every identifier/measurement combination must be unique. If A has two Before rows, decide what they represent before pivoting: replicate measurements may need a replicate identifier, while duplicate records may need correction. HypoNum refuses the duplicate cell instead of silently averaging or selecting its first value.

Missing identifier values form a group. Missing measurement names, or names outside the declared list, are refused. Absent combinations become missing cells. Identifier groups follow their first appearance; columns follow the declared measurement order.

6.8.4 Preview, create and retain

Preview reports projected output rows, cells and width, with at most 100 preview rows and 16 displayed columns. The default ceilings are one million source rows, two million output rows, eight million output cells and 128 user columns. A changed setting invalidates the previous preview.

Choose **Create new tab** to create the checked result, or **Add to results** to retain its summary. **Cancel work** stops pending work. The recorded operation retains its input relationships and exclusions for save, reopen, undo and redo. Inspect the new table before using it in a repeated-measures analysis; layout alone does not establish the correct subject and occasion roles.

6.9 General joins in Merge tabs

Choose **General join** inside **Merge tabs...** when the relationship between tables matters. Table A initially refers to the active dataset and Table B to another open dataset. Confirm both before selecting keys. Defaults are an inner join, a one-to-one relationship and suffix `_B` for colliding right-side names.

6.9.1 Worked example: customers and orders

Choose **File > Import data...** and select `sample-data/manual/customers-three.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Keep that tab open. Choose **File > Import data...** again, select `sample-data/manual/orders-four.csv`, use comma, UTF-8, decimal period, header row 1, first data row 2, blank missing codes and empty type overrides, then click **Import**. Choose **Data > Merge tabs...** and set mode to **General join**. Explicitly choose the imported customers tab as **Table A** and the imported orders tab as **Table B**. Table A contains customers C1, C2 and C3, each once. Table B contains the following orders:

Customer	Order	Amount
C1	O1	10
C1	O2	15
C2	O3	8
C4	O4	20

Select Customer on both sides and declare **1:m**: one customer may have many orders. Set the join type to **Inner**, check the three-row preview, then click **Create tab**. To try **Left** and **Full**, reopen **Data > Merge tabs...** for each run and select the original customers and orders tabs again, mode **General join**, key **Customer** on both sides, relationship **1:m**, and the requested join type. Check the preview, then click **Create tab**.

- An inner join returns O1, O2 and O3: three matched rows.
- A left join adds C3 with missing order fields: four rows.
- A full join also adds C4/O4, with missing customer-table fields: five rows.

The legacy first-match mode would retain only the first matching order for C1. Choosing a general join therefore changes the analysis population in a substantive way.

6.9.2 Keys and cardinality

Select one key on each side. **Add key pair** adds another ordered pair; **Remove** removes it. Every pair must be complete, and a column cannot occur twice in the same side's key list. A composite key matches only if every component matches.

Declare **1:1** for unique keys on both sides, **1:m** for uniqueness on Table A, or **m:1** for uniqueness on Table B. The complete eligible inputs are checked, including keys that have no partner on the other side. A duplicate on a side declared unique prevents the join. Guided many-to-many expansion is not offered.

Text keys match exactly, including spaces and case. Null and floating-point NaN keys never match. Compatible numeric and date/time keys may be converted losslessly; incompatible types, precision loss and incompatible time zones are refused. Active exclusions are omitted and recorded.

6.9.3 Read the preview

An inner join keeps matching pairs. A left join also keeps unmatched A rows; a full join keeps unmatched rows from both sides. Check source counts, matched pairs, unmatched rows, duplicate-key counts and projected output size. A high match rate does not establish that the key represents the right entity.

Table A columns come first. Identically named paired keys are combined; other name collisions receive the right suffix. Suffixes retain letters, digits and underscores, replacing other characters with underscores and adding a leading underscore when needed. An empty normalized suffix becomes `_B`; further collisions receive additional disambiguation.

Output follows A row order and then matching B row order. Unmatched B rows come last in a full join. Defaults allow one million rows per input and twenty million combined input cells; output is limited to one million rows, twenty million cells and 256 columns. Counts describe the complete eligible inputs, while retained preview data are limited to 200 rows.

Settings changes refresh the preview. **Create tab** creates the result, leaving source tabs unchanged; closing the dialog cancels pending work. The created table retains the selected settings and source-row relationships for save, reopen, undo and redo.

CHAPTER 7

Create derived columns with a preview

The **Data > Transform** commands create new columns while preserving their sources. Inspect the eight-row preview, name the output and apply. The outputs form one recorded, undoable edit. They are computed once and do not update when source cells later change; apply the transformation again to refresh them. Column formulas have their own recalculation behaviour. All rows participate, including rows excluded from analyses.

7.1 Recode values explicitly

Recode into new column... reads ordered rules. The first matching rule determines the new numeric or text value; no matching rule gives missing. Rules can match a stored value, inclusive range, lower or upper tail, system-missing, user-missing, or all remaining values. The last choice must be last and includes missing values.

A stored user-missing code also matches an ordinary value/range rule. Place a user-missing rule first when it must stay missing. For source $(1, 2, 99, 4, \text{empty}, 2, 3, 5)$ with 99 declared missing, use: user-missing to blank; system-missing to blank; 1–2 to 1 labelled Low; all remaining values to 2 labelled High. The new column is $(1, 1, \text{missing}, 2, \text{missing}, 1, 2, 2)$. This is distinct from the column menu's in-place recoding command.

Integer source matching preserves the distinction between adjacent large 64-bit integers; enter integer boundaries as exact integer values. For example, 9007199254740992 and 9007199254740993 can match different rules. Numeric output uses double precision, however: it cannot preserve every large integer exactly. Choose text output when those output digits are identifiers that must remain exact. Matching precision and output representation are separate choices.

Automatic recode... instead assigns whole-number category codes and keeps the original values or source labels as value labels. Choose ascending, descending or frequency order; tied frequencies use ascending value order. The codes identify categories, not quantitative distances.

7.2 Ranks and rank-derived scores

Rank cases... ranks numeric columns, optionally within declared groups. The default gives rank 1 to the smallest value and average ranks to ties. Missing measurements and missing group values receive missing scores.

Choose **File > Import data...** and select `sample-data/manual/ranks-seven.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Data > Transform > Rank cases...**. Select **value** under **Columns to rank**, leave groups empty, choose **Smallest value** for **Rank 1 goes to**, and **Mean rank** for **Tied values get**. Select only **Rank** under **Scores to create**. Inspect the preview, then click **Create columns** and read the generated rank column in source row order. For $(20, 10, 30, 20, \text{empty}, 20, 40)$, the ranks are $(3, 1, 5, 3, \text{missing}, 3, 6)$.

Lowest, highest, sequential and dense tie rules answer different questions. Sequential ties follow table order; dense ranks number distinct values consecutively. Dense ranks support only the rank itself. For positional rank r among n valid values, fractional rank is r/n and percent rank is $100r/n$. These are declared plotting conventions, not the percentage strictly below the observation.

Normal scores use $\Phi^{-1}(p)$ with Blom's $p = (r - 3/8)/(n + 1/4)$ by default. Tukey, Rankit and van der Waerden positions are alternatives. Savage scores at position j are $\sum_{i=1}^{j} 1/(n - i + 1) - 1$; average ties average the positional scores. Percentile groups use $\lfloor kr/(n + 1) \rfloor + 1$ for the selected k. Ten untied observations and four groups give group sizes 2, 3, 3 and 2; equal-sized or fully occupied groups are not guaranteed.

7.3 Standardize with a stated denominator

Standardize... creates $z = (x - \bar{x})/s$ for each selected numeric column, optionally within groups. Choose sample SD (divisor $n-1$, the default) or population SD (divisor n). Missing observations are omitted from these summaries and receive missing scores. Zero-spread groups, and one-observation groups with sample SD, receive missing scores.

Choose **File > Import data...** and select `sample-data/manual/standardize-eight.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Data > Transform > Standardize...**. Select **value** under **Columns to standardize**, leave groups empty, and choose **Population SD (divides by n)**. Inspect the preview and click **Create columns**. The generated scores use the same row order. For $(2, 4, 4, 4, 5, 5, 7, 9)$ the mean is 5, population SD is 2 and population z-scores are $(-1.5, -0.5, -0.5, -0.5, 0, 0, 1, 2)$. Sample-standardized scores instead divide the deviations by $\sqrt{32/7}$. Standardization changes units; it does not make a distribution normal.

7.4 Count matching cells

Count values... counts, within each row, selected columns satisfying any stated value/range/missingness rule. Choose columns of compatible numeric or text type. A cell matching several rules counts once. The output is an integer between zero and the number of columns, never missing. Recode and Count read stored values and supply explicit missing-value rules; the rank, standardization and binning calculations treat declared missing codes as missing.

7.5 Choose and retain bin boundaries

Visual binning... creates ordinal bin numbers and labels. Choose equal widths, equal percentiles, mean and selected sample-SD offsets, or entered cut points. The default is four bins. Equal-percentile cuts use linearly interpolated quantiles; duplicate cuts merge and the reduced bin count is reported. Equal-width and equal-percentile methods refuse a constant column.

By default an observation exactly on a cut goes in the lower bin. With cuts 20 and 40, values $(10, 20, 30, 40, 50)$ give $(1, 1, 2, 2, 3)$. Excluding upper endpoints instead gives $(1, 2, 2, 3, 3)$. Missing values stay missing. At most 99 cuts are supported. The chosen boundaries, endpoint rule and labels are retained; replay does not silently learn new cuts from later data. Binning loses within-bin information, so keep the original column.

CHAPTER 8

From a question to an analysis

The methods in this book answer different questions. A descriptive summary characterizes the observed data; an interval expresses uncertainty about a specified quantity; a fitted model describes a relationship under stated assumptions. Choosing a sophisticated method cannot repair a mismatch between the question, the sample and the design.

The desktop and this book group related methods into Explore, Statistics, Survey, Time Series, Quality, Design and Predict. Follow the menu path printed with each procedure. Cross-listed commands have one primary explanation and references from their other locations.

8.1 Write the question in measurable terms

Specify the response, the observational unit, the population of interest and the comparison or prediction you need. Asking whether a process improved is incomplete until improvement has a direction and a meaningful size. Asking whether the mean filling mass increased by at least 0.5 g identifies a quantity and units, but still needs a sampling design and a time period.

Distinguish these common aims:

Describe.
Summarize the data you have, including variation, shape, missingness and unusual observations.

Compare or estimate.
Estimate a population quantity or contrast, with uncertainty appropriate to the sampling or experimental design.

Explain an association.
Describe how the response varies with predictors, accounting for the chosen covariates. An observational association is not automatically a causal effect.

Predict.
Assess how well a procedure predicts outcomes it has not used for fitting or selection.

Monitor.
Compare an ordered process with a specified stable reference. A control limit is not a product specification.

8.2 Choose the analysis population

Most desktop tools begin with the active table after user exclusions. They then apply their own requirements to the selected columns. Always read the included, omitted and analyzed counts. Changing a predictor can change the fitted population if that predictor has missing values.

Use independent-group procedures for different independent units. Use paired procedures for measurements genuinely matched on a unit or design. Use repeated-measures, mixed, panel or survey methods when their dependence or sampling structure is part of the question. Treating correlated readings as independent commonly overstates the information.

For a weighted analysis, establish what the weights mean. Survey weights, frequency counts and inverse-variance weights answer different questions and cannot be exchanged just because all are positive numbers.

8.3 Inspect the data before fitting

Check units, impossible values, missing-value codes and the number of observations in each group. Plot the response and relevant predictors. For time series, inspect order, spacing and changes over time. For experiments, compare the collected runs with the intended design.

Do not use a normality test as an automatic switch that decides the entire analysis. The relevant assumption may concern model errors, paired differences or a sampling approximation. A small sample may reveal little about the tails, while a large sample may detect a departure with little practical consequence. Combine the design, the plot and the model's sensitivity to departures.

8.4 Understand an estimate and its interval

An estimate names a quantity: a mean, a difference, a ratio, a regression coefficient or a survival probability. Interpret its sign, reference group and units before reading its p-value.

A 95% confidence procedure is designed to cover the fixed target in 95% of repetitions under its assumptions. The realized interval does not assign a 95% probability to that fixed target. It also does not contain 95% of individual observations.

A prediction interval concerns a future observation or other declared future outcome and usually includes more variation than a confidence interval for a mean. A tolerance interval seeks to cover a declared fraction of a population with a declared confidence. These intervals serve different purposes; always identify which one is displayed.

Worked example

Suppose a mean difference is estimated as 0.4 g with a 95% confidence interval from −0.1 to 0.9 g. The data remain compatible with both a small decrease and an increase under the model. They do not establish zero difference. If an increase of 0.5 g is required to justify a change, the interval also shows why the practical decision remains uncertain. These illustrative values explain interpretation; they are not the output of a supplied dataset.

8.5 Hypotheses, direction and multiplicity

Choose the null quantity, significance level and alternative before examining the result. A two-sided alternative allows a departure in either direction. **Less than** and **Greater than** are directional claims; the order of the groups determines their meaning.

A p-value measures how unusual the observed discrepancy would be under the specified null model, using the selected statistic and reference distribution. It is neither the probability that the null hypothesis is true nor the size or importance of the effect. Report the estimate and interval alongside it.

A nonsignificant difference does not establish equivalence. Equivalence requires scientifically justified margins and an appropriate equivalence procedure. With several comparisons, define the family of claims and use the procedure's simultaneous or multiplicity-adjusted results. Selecting only the most favorable comparison after looking at all of them changes the evidential meaning of an ordinary single-comparison p-value.

8.6 Missing data and prediction

Complete-case analysis uses only rows observed on its required variables. Its interpretation depends on why the values are missing. It does not automatically recover the original population.

Correlation uses pairwise-complete finite rows, so matrix entries can describe different subsets. The exploratory **Fit Y by X** workflow offers deletion and single-imputation choices; its missing-value treatment occurs before its training/validation partition. Its displayed split therefore should not be treated as an assessment of a fully independent data-preparation procedure.

For predictive evaluation, preparation learned from data belongs inside the training process. A test set used to select variables, transformations or models has become development data. Use the dedicated validated prediction workflow for its stated procedure and read its limitations before making a performance claim.

8.7 Read the complete result

Examine the method and population disclosure, estimates and intervals, diagnostics, warnings, and any rows marked unavailable. Nonconvergence, separation, rank deficiency or an unsupported design can make a requested quantity unavailable. A displayed number is not a reason to ignore its status.

Save useful results with **Add to results** when offered. In the report, state the question, data period and units; included observations and exclusions; method and essential settings; estimates and uncertainty; diagnostic concerns; and the conclusion supported by those results. A saved result remains evidence for its original data and settings when the source later changes.

Part II

Explore

CHAPTER 9

Exploring distributions and relationships

Explore before choosing a statistical model. Look for scale, shape, missing values, groups and unusual observations, then ask whether the pattern is consistent with how the data were collected. A visual pattern can suggest a question; it does not establish a cause.

Explore uses included rows and links selectable observations to the grid. Read the disclosure beneath a plot: a statistic can use all eligible rows while the picture shows a bounded sample.

9.1 Choose a plot for the question

Open **Explore > Plot...** and assign **Plot Y**. Double-clicking a numeric column, dragging it onto the canvas, or focusing it and pressing Space or Enter assigns Y and opens the default box plot. Repeating this action replaces Y and returns to a one-variable view.

Question	Useful view
Where is a variable centred and how variable is it?	Box plot, with raw points where practical.
Is its distribution skewed, rounded or multimodal?	Histogram, violin or dotplot; compare more than one binning or view.
Do groups differ?	Numeric Y with categorical X; add **Then by** for a second categorical split.
Are two measurements related?	Numeric X and Y scatterplot.
How does an ordered trajectory change?	Line plot, only when X order has substantive meaning.
How frequent are categories?	Categorical Y bar chart, Pareto chart, or counted bars.

Numeric X and categorical split roles are mutually exclusive. A categorical Y produces a univariate count bar chart, using all retained rows. Categories are alphabetical, with missing values last; more than 36 categories is refused. X, split, colour and facet roles are unavailable for that view. For split numeric plots, more than 36 observed group combinations is also refused. **Clear axes** removes the setup; **Restore plot** retrieves the last exact setup until another plot is made.

9.1.1 Read a distribution

Box plots use linearly interpolated (type-7) quartiles. Whiskers extend to the most extreme observations within 1.5 times the interquartile range below the first or above the third quartile. These quartiles are not Tukey hinges; for values 1, 2, 3, 4, the quartiles are 1.75 and 3.25. A point beyond a whisker is an unusual observation under that display rule, not automatically an error. A violin adds an estimated density, whose shape depends on smoothing. Histograms count observations in intervals: a peak that disappears after a small change of bins deserves caution.

Box and violin plots can overlay bounded raw points. Histogram bars always show exact counts. A dotplot stacks observations in numeric bins; when a stack is too large to draw fully, its exact count remains labelled.

9.1.2 Worked exploration: compare two groups

Choose **File > Import data...** and select `sample-data/manual/paired-measurements.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Explore > Plot...**. In the column list, select **x** and click **Plot Y**; select **g** and click **Plot X**. Keep the box-plot view and enable **Overlay points**. The plot updates from these role selections; inspect the two groups and their overlapping observations. The small dataset used in the statistical examples contains eight observations in group A and eight in B. Their numeric x values are:

$$A : 3.1,\ 4.2,\ 5.4,\ 5.8,\ 6.3,\ 7.2,\ 7.7,\ 8.5,$$
$$B : 5.2,\ 5.7,\ 6.1,\ 7.4,\ 8.2,\ 8.8,\ 9.1,\ 10.3.$$

Here **g** supplies categorical X and **x** supplies numeric Y. The means are 6.025 and 7.600, but the ranges overlap substantially. A mean difference alone does not tell you whether the populations differ precisely enough for the study's purpose. A confidence interval, the design and the practical size of the difference complete that assessment.

9.2 Scatterplots and descriptive overlays

With numeric X and Y, a scatterplot shows paired observations. A curved pattern can coexist with a small linear correlation. A single unusual point can dominate an apparently strong linear relationship. Use linked selection to inspect such points in the original rows.

The optional ordinary least-squares fit includes a 95% confidence band for the *mean response*. It is not a prediction interval for individual new observations. Fitting is unavailable while colour or facet roles are assigned.

LOESS adds a local linear smoother using a deterministic bounded sample of finite pairs, span 0.75 and no robustness iterations. Its fit sample size is disclosed. The curve is descriptive and has no confidence band.

Data ellipse uses the mean and sample covariance of all finite pairs to draw a 95% bivariate-normal data contour. This describes a model-based region for observations, not uncertainty in the mean. At least three valid pairs and nonsingular covariance are required.

Other numeric-X/Y views serve specific questions:

Line
Connects the bounded displayed points in increasing X order. It may skip undrawn observations, so do not use it to audit every event.

Stacked area
Sums nonnegative Y at each X by the selected category. Band thickness represents contribution to the total.

Bubble
Adds a numeric Size column, with circle area proportional to its value; the size range is disclosed.

Binned scatter
Counts every valid pair in a square grid cell. Selecting a cell selects its contributing source rows.

Marginal
Combines a bounded scatter display with exact X and Y histograms.

9.3 Colour, facets and linked selection

For numeric X and Y, assign a nonnumeric column to **Color by**, **Facet by**, or both using different columns. A colour role distinguishes categories within a panel; a facet role makes a panel for each category. Numeric codes remain numeric and cannot take these roles. Each column can hold only one role. **Clear** removes a role; choosing categorical X switches to a split plot and removes both.

Categories are ordered by text, with missing values last as **(missing)**. Empty text is a category. A literal text value `(missing)` remains distinct from a missing value in this workflow. Turn off **Show missing categories** to omit missing colour or facet values; the disclosure counts the omission. Each role and the observed combination of roles are limited to 36 categories or groups.

Panels share X and Y scales. Axis limits, reference lines and grids apply to every panel; points outside a limit are not drawn. Large arrangements scroll so that panels remain readable. Category order uses every included row, helping keep a category's appearance consistent across axis choices. The first twenty categories use distinct palette colours; later categories reuse colours with square markers. The legend may shorten long labels; **Show group summary** retains their full names.

All groups share a 5,000-point display budget, allocated so that small groups can remain visible. Selection is deterministic for unchanged data. The group summary separates source rows, valid X/Y pairs and drawn points. These counts describe complete groups, not only the plotted sample.

Click or drag a rectangle to select plotted rows. Grid selection highlights the same observations across panels, with dark outlines around selected points. Colour swatches under the group summary change the current display; these changes last until reopening the plot window and are not saved in the project.

Add to results retains one chart per facet, labelled with the facet value and without category colours, plus exact group counts. Saved plot settings retain role columns and the missing-category choice. Renaming, recoding or deleting a role column clears the plot. A restored scatterplot requires its X column still to be numeric.

9.4 Compare several variables

Multi-Plot makes one card per selected numeric column. Choose Box, Violin or Histogram for all cards. Histogram bin counts range from 1 to 64, default 24. **Overlay points** affects box and violin cards only.

Each card reports *N*, mean, median, standard deviation, and the displayed mean-plus-or-minus-four-standard-deviation values. Those are descriptive limits, not automatically control limits, confidence intervals or specification limits. Statistics use all valid included values; raw-point overlays and violin density use the disclosed bounded sample. **Add to results** retains the selected card type.

Parallel coordinates place Y and X on the first two axes and accept one to eight further numeric axes through **Add axis**. Only rows finite on all selected axes are eligible; at most 1,000 source-linked rows are drawn. Each axis uses its own complete-row range, with constant axes centred. A line identifies a row; selecting it selects that row in the grid. Rescaling axes changes visual slopes, so compare relative patterns rather than treating a steep segment as a statistical coefficient.

LOESS, data ellipse and parallel-coordinates setups can be saved and replayed as plots. Their report snapshots are not yet available.

9.5 Variation between groups and occasions

Variation charts offers Individual values, Mean interval plot and Multi-vari chart. Individual values show observations and group means. Mean interval plots use 95% t confidence intervals for means. Overlapping individual-value distributions and overlapping mean intervals answer different questions.

In a Multi-vari chart, **Within panel** names the inner group and **Panels / time** the outer group. Dividers and labels separate the levels. Use a grouping that reflects the study, such as parts within production days. Incomplete and excluded rows are counted. **Add to results** saves the chart snapshot.

9.6 Prioritize categories with a Pareto chart

Choose a categorical cause, defect or reason. **Count rows** treats each row as one event. An optional finite, nonnegative frequency column supports already summarized data. Bars descend by frequency, with cumulative percentage on the right axis and an 80% guide.

For defect counts 12, 5 and 3, the total is 20 and cumulative percentages are 60%, 85% and 100%. The first two categories exceed the 80% guide. This is a prioritization convention, not a significance test or evidence that those categories cause 85% of cost. A frequency chart and a cost-weighted chart answer different questions.

For long lists, set a maximum and combine the remainder into **Other**. The exact table reports frequencies, cumulative values, combined categories and omitted rows.

9.7 Investigate more than one mode

Explore > Multimodality... searches numeric columns for evidence of multiple distributional modes. Inspect its row counts, held-out counts, mode summaries, plot data and warnings. Several peaks can arise from different populations, rounding or sampling variation. They do not by themselves prove different physical causes.

Select a mode row and choose **Explain multimodality**, or open the same command from the Explore menu. Each finite response value is assigned to the density basin defined by the plotted kernel-density valleys. Modes are numbered from low to high; one basin leaves no split to explain.

A shallow tree searches other columns for rules distinguishing those basins. Categorical splitters are limited to eight levels by default; numeric columns with no more than that many distinct values are treated as categorical. Continuous numeric thresholds are off by default. Time splitters can use an automatically selected time unit and, for sufficiently long spans, hour, weekday or month.

Defaults reserve 25% of rows by a seeded permutation. A split must reduce training entropy by at least 0.02 of the node's entropy, produce at least two branches meeting a 2% minimum leaf fraction, and also reduce held-out entropy. Growth stops at depth three, a pure node or the absence of a qualifying split. Equal gains are resolved by column name.

Read each leaf rule as a conjunction of its conditions. The result reports its predicted majority mode, row count, training purity and held-out purity. A separate table ranks each candidate's root-level gain. The held-out check helps reject unstable splits, but it does not make a descriptive

search causal or guarantee performance on new populations. Save the findings with **Add to results**.

9.8 Save, reuse and export a plot

Copy plot and **Export PNG** preserve the complete high-resolution rendering, including labels and disclosures. Pixel dimensions depend on the plot surface size and the space needed for its title, legend and footnote; they are not fixed at 1600 by 1000. A scrolling facet export includes every panel and the legend. Plot choices are restored with a saved project when its source remains available.

Save as quick plot records a named setup for reuse on tables with the same column names. It retains axes, categorical splits and filters, plot type, bins, overlays, fit and axis settings, including supported per-category colours. These user-level presets are distinct from temporary colour/facet swatch changes. Up to twelve quick plots are retained on the machine, most recent first; saving an existing name replaces it. **Explore > Quick plot** applies the chosen preset. Missing columns are reported instead of silently substituting others. **Remove quick plots** removes presets without changing data or projects.

CHAPTER 10

Counts, percentages and cumulative distributions

Choose **Explore > Bars, pie, heatmap and ECDF...** to summarize groups or compare distributions using retained rows after exclusions. Aggregated bars and cells describe groups, not individual source rows.

10.1 Set the roles and the summary

A bar chart uses one group column, or two for grouped or stacked series. A pie uses one mutually exclusive category column and counts rows. A heatmap uses different row and column category columns. An empirical cumulative distribution function (ECDF) uses a numeric value and optionally a group column.

For categorical bars, choose count, mean, median or sum. Heatmaps support count, mean and median, but not sum. Counts need no value column; sums add values, rather than count observations. Mean, median and sum use valid values; missing values remain in source counts but not in the summary.

Order categories alphabetically or by first appearance. Bars and pies can also use descending frequency. Missing groups can be displayed last as **(missing)** or omitted with their count disclosed. In this workflow a literal category already named `(missing)` conflicts with that label; rename it or change the missing-group treatment before proceeding.

10.2 State the percentage denominator

Counted bars can show within-group, within-series or overall percentages. Choose one denominator for the entire chart.

Worked example

Choose **File > Import data...** and select `sample-data/manual/plant-outcomes-fifty.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Explore > Bars, pie, heatmap and ECDF...**. Set **Graph** to **Categorical bars**, **Category column** to **Plant**, **Bars: series (optional)** to **Outcome**, and **Aggregation** to **Counts**. Choose **Within group** for **Percentages**, **Alphabetical** for **Ordering**, and **Show as (missing)** for **Missing groups**. Keep **Stack the series** selected and the default row and cell ceilings. Click **Draw graph**. The 50 source rows have these counts:

Plant	Pass	Fail	Total
A	18	2	20
B	24	6	30
Total	42	8	50

For the Plant A/Fail cell, within-plant percentage is $2/20 = 10\%$; within-outcome percentage is $2/8 = 25\%$; overall percentage is $2/50 = 4\%$. These are three valid answers to different questions. The first estimates the plant's observed failure proportion; the second describes its contribution to all observed failures. To reproduce the other two percentages, return to the setup window, change only **Percentages** to **Within series**, and click **Draw graph**; then repeat with **Overall**. Read the Plant A/Fail row in the accompanying result table: 10%, 25% and 4%, respectively.

Within-group percentages sum to 100% within each group; within-series percentages within each series; overall percentages across the whole chart. Percentage bars require both a category and a series column. Percentages are available for counts, not mixed with value summaries. The accompanying table gives values, valid N, source and contributing rows, and the percentage denominator where applicable.

10.3 Use a pie for a simple part-to-whole summary

A pie shows counts and percentages for one category column. Include missing values as a slice or omit them explicitly. Above twelve categories, the eleven largest remain separate and the rest form **Other**, with exact combined count and percentage. A literal category named `Other` stays distinct.

Prefer a count or percentage table when close values need precise comparison. The table and CSV export list the displayed slices, including the combined slice. **Add to results** saves both chart and table.

10.4 Read a two-way heatmap

A heatmap places two categorical variables on its axes and uses a sequential scale to represent the selected summary. It includes every combination of the observed row and column levels. An empty counted cell is zero; an empty mean or median cell has no estimate and is drawn neutrally. Do not interpret an absent estimate as a measured zero.

The full table reports each cell's result, source and contributing counts. Compare denominators: a large sum may reflect more observations rather than a larger typical response.

10.5 Read an empirical cumulative distribution

For n finite observations, the ECDF at x is

$$F_n(x) = \frac{\#\{i : x_i \le x\}}{n}.$$

It answers, for example, what fraction of observed delivery times was at most five days. It is a description of the sample, not automatically a confidence bound for a population percentile.

10.5.1 Worked example: tied observations in an ECDF

Choose **File > Import data...** and select `sample-data/manual/ecdf-four.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Explore > Bars, pie, heatmap and ECDF...**. Set **Graph** to **Empirical CDF**, **ECDF: value** to **value**, and **ECDF: group (optional)** to **(None)**. Keep **Ordering** at **Alphabetical**, **Missing groups** at **Show as (missing)**, and the default row and cell ceilings. Select **Cumulative proportion (0-1)** and click **Draw graph**. For these $1, 1, 2, 4$ observations, the curve jumps to 0.50 at 1, to 0.75 at 2 and to 1.00 at 4. Both observations tied at 1 contribute to the same jump. Choose cumulative proportions or counts; an optional group column draws up to 36 series.

Up to 5,000 distinct values are displayed. Above that, a deterministic reduction chooses displayed values, but each displayed cumulative value remains exact. The reduction is disclosed and point selection is disabled. Without reduction, a step can select a source row only when exactly one row has that value. Tied steps do not pretend to identify one row. The series table reports full-data N, distinct count, minimum, quartiles, median and maximum.

10.6 Limits and retained results

Bar and heatmap axes permit up to 36 categories. Pie categories beyond the twelve-slice limit are combined as described above. Default calculation ceilings are one million source rows, twenty million selected input cells and two million output cells. Exceeding a ceiling stops the request; the calculation is not silently truncated.

The window displays at most 200 table rows; **Export full CSV** provides the complete table. **Add to results** retains a bounded table, settings, source context and separate excluded/missing counts. Reopening restores the saved snapshot without rerunning.

10.7 Matrix, probability and spatial displays

The exploration-graph commands complement the general Plot window. Select variables according to the question and inspect the included-row count before interpreting the geometry.

A matrix plot uses 2–8 numeric columns and complete rows. When drawing is limited to 5,000 observations, the same deterministic row sample is shared across panels; axis ranges and diagonal histograms still use the declared analysis population. While the result is current, clicking a point selects its row in the data grid; selecting rows in the grid rings their points across the matrix panels. The plot reports selected rows that it does not draw, including rows outside the sample, rows with missing values and excluded rows. Selection does not change the plotted sample or the analysis population and is not a recorded data operation. Pairwise visual association need not persist after adjustment or imply causation.

Stem-and-leaf displays truncate at the chosen leaf unit, keeping negative and positive stems distinct. Thus with unit 1, 9.99 appears as stem 0, leaf 9, not stem 1, leaf 0. The automatic unit and split stems are display choices. Omitted leaves or lines at display limits are counted. Read the key before reconstructing a value from its stem and leaf.

A probability plot compares ordered observations with fitted distribution quantiles using Benard positions $(i - 0.3)/(n + 0.4)$. Complete-data positive families require values strictly above their declared origin for the log plotting scale. The maximum-likelihood line and pointwise percentile intervals are not least-squares plotting lines, simultaneous confidence bands or prediction limits

for future observations. Ties retain their ranks. Each group can have its own fit; a pooled fit is a separate scientific assumption.

Contour and surface plots interpolate measured Z over X, Y using linear triangles, after rescaling the two coordinates for triangulation. A plane is reproduced exactly: at the centre $(0.5, 0.5)$ of observations from $z = 2x + 3y$, the interpolated value is 2.5. Outside the observed convex hull grid values remain blank. Conflicting responses at the same location require an explicit mean-aggregation choice. This display is piecewise planar, not a smooth fitted response model. Three-dimensional scatter is an orthographic display whose overlap depends on viewing direction.

After running a stored-model profiler, **Predict > Overlaid contour of profiled models...** evaluates its retained mean responses over two numeric factors while holding other inputs fixed. The common acceptable region reflects positive desirability for every response. All models must use both varied factors. The region contains no prediction uncertainty; it is not a guaranteed joint acceptance region for future units.

10.8 Editing a plot without changing the analysis

Plot options controls title and subtitle, axis limits, tick steps, reference/specification/target lines, annotations, legend and footnote where the plot supports them. Positions and limits are in data units. The supported options vary with aggregation and faceting; controls do not manufacture underlying measurements or change the numerical analysis.

Fixed limits disclose how many drawn marks lie outside the range. For points at doses 1–5, an X maximum of 3.5 hides the last two and reports two of five plotted marks outside. In a sampled plot that count refers to drawn marks; the population/sampling disclosure retains its separate meaning. Do not use clipping to imply that omitted observations never existed.

Saving a quick plot, adding it to results and exporting preserves the supported presentation settings. Changing X or moving to/from a histogram clears inapplicable axes and annotations; changing Y starts a new presentation. Edits to a retained result chart affect that saved chart, not its source analysis or every other copy. Inspect the exported figure at its intended size for readable labels, legend and clipping disclosures.

Part III

Statistics

CHAPTER 11

Comparing groups and testing claims

A test compares the observed data with a specified null model. Choose the comparison from the study design before examining which choice gives a small p-value. Report an estimate and uncertainty in meaningful units, then use the test as one part of the argument.

11.1 A small dataset for following the calculations

The examples use the synthetic file `sample-data/manual/paired-measurements.csv`. Each row contains two paired numeric measurements x, y, a group g, a second factor h and a binary outcome. The values are deliberately small enough to inspect. They illustrate methods; they are not observations from a clinical or industrial study.

Row	x	y	g	h	outcome
1	3.1	2.4	A	L	yes
2	4.2	3.8	A	R	no
3	5.4	4.9	A	L	yes
4	5.8	5.1	A	R	yes
5	6.3	5.8	A	L	no
6	7.2	6.8	A	R	yes
7	7.7	7.1	A	L	no
8	8.5	8.0	A	R	yes
9	5.2	4.7	B	L	yes
10	5.7	5.1	B	R	no
11	6.1	5.8	B	L	no
12	7.4	6.6	B	R	no
13	8.2	7.5	B	L	yes
14	8.8	8.1	B	R	no
15	9.1	8.8	B	L	no
16	10.3	9.5	B	R	yes

Import with the first row as column names. Keep x, y numeric and g, h, outcome categorical. The printed Row numbers are reading aids, not an additional analysis variable. There are no missing or excluded observations in these examples.

11.2 Select the design before the test

Open **Statistics > Compare means & ANOVA > Hypothesis tests...**. The opening tab is **Hypothesis test**, with one-sample t, target 0, two-sided alternative and $\alpha = 0.05$. Tukey comparisons are initially enabled where applicable.

Situation	Setup
One numeric population vs a target	One-sample *t*: response, null mean, alternative and alpha.
Two independent groups	Two-sample *t*: numeric response and exactly two groups. Welch unequal-variance is default; pooled variance is optional.
Matched observations	Paired *t*: two different numeric columns, with one pair per row.
One binary proportion	One-proportion: outcome, success value, null proportion, alternative and alpha.
Two binary proportions	Two-proportion: outcome, two-level group, success value and Automatic, *z*, or Fisher exact method.
Several independent means	One-way ANOVA: response and group. Classic ANOVA is default; Welch is optional.
Two factors	Two-way ANOVA: response, row and column factors, including their interaction; optional Tukey comparisons for each factor.
One categorical distribution	Chi-square goodness-of-fit: outcome and expected proportions in displayed level order. Blank means equal; entered proportions must be positive and sum to one.
Two categorical variables	Chi-square association: distinct factors.
Group variances	Equal-variance test: Brown–Forsythe default, Bartlett, or a two-group *F* test.
Rank-based independent comparison	Mann–Whitney for two groups, Kruskal–Wallis for several, or Mood's median test.
Rank or sign comparison of pairs	Wilcoxon signed-rank or sign test.

An independent-groups test assumes that observations are independent across and within groups. A paired test assumes independent pairs, while allowing dependence within a pair. Repeated rows from the same person, machine or site need a design that represents that dependence.

11.3 One mean: calculate the effect and its uncertainty

Choose **File > Import data...** and select `sample-data/manual/paired-measurements.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Compare means & ANOVA > Hypothesis tests...**. On **Hypothesis test**, choose **Means, variances and proportions** as **Test family** and **One-sample t-test** as **Test**. Set **Response** to **x**, **Null mean** to 6.2, **Alternative** to **Two-sided** and **Alpha** to 0.05. Click **Run test**; read the estimate, test statistic and interval. The 16 values have mean 6.8125 and sample standard deviation 1.9259197. The estimated difference from the target is 0.6125. The standard error is $s/\sqrt{n} = 0.4814799$, so

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = 1.2721195, \qquad \nu = n - 1 = 15.$$

The two-sided p-value is 0.2226990. The 95% interval for the difference is

$$0.6125 \pm t_{0.975,15}(0.4814799) = [-0.4137502,\ 1.6387502].$$

Adding 6.2 gives the corresponding interval for the population mean, [5.7862498, 7.8387502].

The interval includes zero difference and allows differences in either direction. This is insufficient evidence of a mean difference at 5%; it is not proof that the mean equals 6.2. A claim of practical equivalence requires an explicit acceptable margin and an equivalence procedure.

The exact small-sample t model assumes independent normal errors. Inspect shape and unusual observations, especially for small samples. A large sample can make mean inference more stable without repairing dependence, biased sampling or a badly chosen population.

11.4 Independent groups and paired observations

11.4.1 A difference between independent means

Choose **File > Import data...** and select `sample-data/manual/paired-measurements.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Compare means & ANOVA > Hypothesis tests...**. On **Hypothesis test**, set **Test family** to **Means, variances and proportions** and **Test** to **Two-sample t-test**. Set **Response** to **x**, **Group** to **g**, **Method** to **Welch (unequal variance)**, **Alternative** to **Two-sided**, and **Alpha** to 0.05. Click **Run test**. Read the group summaries and the A-minus-B contrast. For x by g, both groups have $n = 8$:

$$\bar{x}_A = 6.025, \quad s_A = 1.8013884, \qquad \bar{x}_B = 7.600, \quad s_B = 1.8126539.$$

Welch's statistic uses the standard error $\sqrt{s_A^2/n_A + s_B^2/n_B}$ and the Satterthwaite degrees of freedom,

$$\nu = \frac{(s_A^2/n_A + s_B^2/n_B)^2}{(s_A^2/n_A)^2/(n_A - 1) + (s_B^2/n_B)^2/(n_B - 1)}.$$

For the A-minus-B direction the difference is -1.575, $t = -1.7431919$, $\nu = 13.99946$, and two-sided $p = 0.1032105$. The observed group B mean is larger, but these data give limited precision about the population difference. Always read the displayed contrast direction before interpreting its sign.

Pooled t replaces the two variance estimates by one weighted estimate and uses $n_A + n_B - 2$ degrees of freedom. Choose it when a common variance is scientifically defensible, not solely because a preliminary variance test was nonsignificant.

11.4.2 The same rows treated as pairs

Choose **File > Import data...** and select `sample-data/manual/paired-measurements.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Compare means & ANOVA > Hypothesis tests...**. On **Hypothesis test**, choose **Means, variances and proportions** and **Paired t-test**. Set **Response** to **x**, **Paired response** to **y**, **Alternative** to **Two-sided**, and **Alpha** to 0.05. Click **Run test** and inspect the mean paired difference and its test. The analysis uses $d_i = x_i - y_i$, not two independent samples:

$$\bar{d} = 0.5625, \quad s_d = 0.1627882, \qquad t = \frac{\bar{d}}{s_d/\sqrt{16}} = 13.8216401.$$

The two-sided p-value is approximately 6.126×10^{-10}. The strong paired association makes the difference much less variable than either column alone. Pairing is justified by how observations were collected, not by choosing the method that gives the smallest p-value.

The paired t assumption concerns the distribution of differences. Normal-looking individual columns do not establish normal differences, and vice versa.

11.5 ANOVA and multiple comparisons

One-way classic ANOVA compares between-group mean variation with pooled within-group variation. With k groups and N observations,

$$F = \frac{\sum_g n_g(\bar{y}_g - \bar{y})^2/(k-1)}{\sum_g \sum_i (y_{gi} - \bar{y}_g)^2/(N-k)}.$$

To reproduce the numerical example in this section,

Choose **File > Import data...** and select `sample-data/manual/paired-measurements.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Compare means & ANOVA > Hypothesis tests...**. On **Hypothesis test**, select **Means, variances and proportions** and **One-way ANOVA**. Set **Response** to **x**, **Group** to **g**, **Method** to **Classic ANOVA**, and **Alpha** to 0.05. Click **Run test** and read the omnibus test table. For x by g, $F = 3.0387181$, with degrees of freedom 1 and 14, and $p = 0.1032096$. With two groups this equals the squared pooled two-sample t statistic. Its slight difference from Welch's result reflects their variance conventions.

A significant omnibus test says that not all group means are equal; it does not identify which comparisons matter. Tukey comparisons control a family of pairwise mean comparisons under the classic common-variance model. Select the family before interpreting its adjusted intervals. Welch ANOVA addresses unequal variances; do not interpret an ordinary pooled Tukey calculation as automatically providing its post hoc analogue.

In two-way ANOVA, first examine interaction: a treatment effect that changes with a second factor cannot generally be summarized by one unqualified main-effect statement. Report the design and relevant simple comparisons.

11.6 Proportions, tables and variances

Declare which outcome counts as success before running a proportion test. A proportion is a count divided by a stated number of eligible trials. For example, 8 successes among 16 trials gives 0.5; it does not by itself give a sufficiently precise population estimate. The model assumes the trials supply the intended independent binomial information.

For two proportions, the selected method can be Automatic, z, or Fisher exact. Read the result's method line. A small-sample exact contingency-table test and a large-sample approximation need not produce the same p-value.

Chi-square association compares observed counts with the counts expected under independence, $E_{ij} = n_{i+}n_{+j}/N$. Goodness-of-fit compares one categorical distribution with externally specified proportions. Small expected counts weaken ordinary chi-square approximations. A statistically detectable association can still be practically small.

11.6.1 Worked example: two binary outcomes

This example compares the fraction of yes outcomes in two independent groups. Each row represents one observed unit; there are no frequency weights. Start with no excluded rows and follow these steps:

1. Choose **File > Import data...**, select `sample-data/manual/binary-outcomes-twenty.csv` from the supplied sample-data folder, and import its comma-separated, header-bearing data. Make this table active. Check that it has 20 rows and columns **group** and **outcome**: group A has eight yes and two no; group B has one yes and nine no.
2. Choose **Statistics > Compare means & ANOVA > Hypothesis tests...**. On the **Hypothesis test** tab, set **Test family** to **Means, variances and proportions** and **Test** to **Two-proportion test**. This selection displays the categorical **Outcome** control needed for these data.
3. Set **Outcome** to **outcome**, **Group** to **group**, **Success value** to **yes**, **Method** to **Fisher exact**, **Alternative** to **Two-sided**, and **Alpha** to **0.05**.
4. Click **Run test**. In the results, check the two group counts, the A-minus-B direction, the Fisher method label and the values below.

The observed difference A minus B is $0.8 - 0.1 = 0.7$. The reported difference interval, approximately $[0.265827, 0.865180]$, uses the stated Newcombe score convention; choosing Fisher's test does not turn it into an exact conditional interval for the difference. The two-sided Fisher p-value is 0.00547749, with sample odds ratio $8 \times 9/(2 \times 1) = 36$.

For a second calculation on the same active 20-row table, close the result window and choose **Statistics > Compare means & ANOVA > Hypothesis tests...** again. On the **Hypothesis test** tab, select **Means, variances and proportions** and **Chi-square association**. Set **Row factor** to **group**, **Column factor** to **outcome**, and **Alpha** to **0.05**; click **Run test**. This gives Pearson $\chi^2 = 9.89899$ on one degree of freedom, $p = 0.00165370$, without a continuity correction. Two expected cells are 4.5, so the small-expected-count warning applies. The different p-values reflect different reference calculations, not different observations. Report the selected test and the effect estimate together.

Brown–Forsythe examines absolute deviations from group medians and is the default variance comparison. Bartlett and the two-group variance-ratio F test rely more strongly on normality. A variance test evaluates variation, not a difference in means. The two-group F test requires positive observed variance in both groups. A constant group is refused whichever group is listed first.

11.7 Rank and sign procedures

Rank methods use different information and hypotheses from mean tests. Mann–Whitney compares two independent distributions through ordering. Kruskal–Wallis extends a pooled-rank comparison to several groups. A location or median interpretation requires additional shape assumptions.

Choose **File > Import data...** and select `sample-data/manual/paired-measurements.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Compare means & ANOVA > Hypothesis tests...**. On **Hypothesis test**, select **Nonparametric: independent samples** as **Test family**, then **Kruskal–Wallis test** as **Test**. Set **Response** to **x**, **Group** to **g**, and **Alpha** to 0.05. Click **Run test**. The Kruskal–Wallis omnibus result gives

$H = 2.1617647$, with asymptotic $p = 0.1414821$. This is not an alternative way of calculating the same mean test; ranking changes the information used.

11.7.1 Wilcoxon signed-rank

HypoNum forms first-minus-second differences, removes and counts zero differences, and assigns average ranks to the remaining absolute differences. An all-zero set is refused. It reports W^+, the sum of positive-difference ranks, and matched rank-biserial effect $(W^+ - W^-)/(W^+ + W^-)$.

For at most 50 nonzero differences, inference is exact conditional on the observed absolute differences under independent fair signs, including tied ranks. The one-sided tails are inclusive; the two-sided value is twice the smaller inclusive tail, capped at one. Above 50, the normal approximation uses

$$E(W^+) = \frac{m(m+1)}{4}, \qquad \mathrm{Var}(W^+) = \frac{m(m+1)(2m+1)}{24} - \frac{\sum_g (t_g^3 - t_g)}{48},$$

with a half-unit continuity correction and tie-group sizes t_g. The result states which inference was used. Symmetry of the difference distribution is required for the usual zero-location interpretation.

The sign test uses only the directions of nonzero differences. It avoids the signed-rank symmetry interpretation but discards magnitude information. Choose the method because its estimand and assumptions fit the question.

11.7.2 Mood's median test

Values above the pooled median form one row of a count table; values at or below it form the other. Ties count below. HypoNum reports Pearson chi-square on $k - 1$ degrees of freedom without a continuity correction. An expected count below 5 triggers a warning; data all on one side are refused. With ties or unequal shapes, this is not an exact test of equal population medians.

11.8 Read and report the result

Start with the estimate and its direction, units and confidence interval. Then report the test statistic, degrees of freedom where applicable, p-value, method and population counts. Review group summaries, assumption checks and comparison tables before accepting a conclusion. A small p-value does not measure practical importance or the probability that the null hypothesis is true.

11.9 Plan a future sample

The dialog's second tab plans one-sample, two-sample and paired t tests, one- and two-proportion tests, and one-way ANOVA. Choose Required sample size or Achieved power. Defaults are standardized effect 0.5, alpha 0.05, power 0.80, sample size 30, three ANOVA groups, equal group sizes and two-sided alternatives.

Use Cohen's d for mean tests, h for proportions and f for ANOVA. For paired d, standardize by the standard deviation of differences. For two-group procedures, entered sample size refers to group 1; the group-2/group-1 ratio determines the other size. Achieved power requires an integer sample size of at least two. The ANOVA group-count range is 2–10,000.

Set the smallest effect that matters and a plausible planning variance. An effect estimated from a small pilot can be very uncertain; compare several defensible values. Power computed from the effect observed in the completed study adds little to its estimate and confidence interval. The Design part develops prospective planning in more detail.

11.10 Establishing a sufficiently small difference

A conventional difference test asks whether a contrast is zero. Failure to reject can result from an imprecise study. Equivalence asks whether the plausible difference is small enough for the declared purpose.

Choose **Statistics > Compare means & ANOVA > Equivalence tests...**. The initial mode is **Independent two-sample**. Select a numeric response, grouping column and two explicit groups. The contrast is Group A minus Group B. Enter justified difference bounds $L < U$ before examining the result.

For independent groups,

$$SE = \sqrt{s_A^2/n_A + s_B^2/n_B},$$

with Welch–Satterthwaite degrees of freedom. The two one-sided tests use

$$t_L = (\widehat{\Delta} - L)/SE, \qquad t_U = (\widehat{\Delta} - U)/SE.$$

Equivalence requires both one-sided p-values strictly below alpha and the matching $100(1-2\alpha)\%$ interval strictly inside (L, U). At alpha 0.05 the interval is 90%. Contact with a bound does not establish equivalence.

For example, a 90% interval $[-0.3, 0.4]$ lies within margins $[-0.5, 0.5]$ and meets the interval condition. An interval $[-0.8, 0.2]$ includes zero but fails that condition. These illustrate the decision, not fitted datasets.

Each group needs at least two finite observations. Missing values, exclusions and observations in other groups are counted separately. Bounds are differences in response units, not values inferred from the observed significance test. One-sample, paired and ratio modes are in chapter 32.

CHAPTER 12

Regression for a numeric response

Regression describes how a numeric response changes with selected predictors. It can estimate an adjusted association or predict a response; those purposes need different checks. An observational regression coefficient is not automatically a causal effect.

12.1 Set up the response and predictors

Open **Statistics > Correlation & regression > Regression...**. Assign at least one continuous numeric response to Y and numeric or categorical predictors to X. Each response is fitted separately using the shared predictors. A column cannot take both roles. Double-click, drag, or use Ctrl/Shift selection in the searchable list to assign columns; repeating a Y assignment replaces the response set.

Nonnumeric predictors use indicator columns, with the first alphabetically sorted level as reference. Missing categories are omitted. Numeric codes remain numeric here: codes 1, 2 and 3 imply a numeric relationship unless you choose a workflow that explicitly declares them as factor levels. **Linear effects** supports that declaration and whole-factor tests.

Setting	Meaning and initial choice
Missing values	Drop rows, by default, over every selected X and Y. Average of 4 neighbors uses up to two finite neighbours on either side within a column. KNN imputation uses five neighbours jointly in the selected numeric matrix. Unfillable rows are reported.
Algorithm	Ordinary Least Squares (OLS), by default. Ridge and Lasso standardize predictors and select a penalty; classical coefficient t tests and model F tests are unavailable for them.
Model type	Linear by default; Polynomial, Exponential, Logarithmic and Power are alternatives.
Polynomial degree	2 by default; range 2–6. Powers are added for each numeric predictor, not for categorical indicators.
Interactions	Off initially; pairwise products of base predictor terms, available with at least two factors.
Validation holdout	0% initially; range 0–50%. Positive values reserve rows for the dialog's validation calculation.
Cross-validation folds	5 initially; choices 2–5, for Ridge or Lasso when holdout is zero.
Confidence level	95% initially; range 50–99%, for OLS intervals.
Standard errors	Classical by default; HC3 or Cluster-robust (CR1) are OLS alternatives. CR1 additionally requires a cluster column that is neither Y nor X.

12.2 Worked example: a fitted line

Choose **File > Import data...** and select `sample-data/manual/paired-measurements.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Correlation & regression > Regression...**. Assign **y** to Y and only **x** to X. Choose OLS, Linear, Drop rows, no holdout, classical standard errors and 95% confidence. Click **Fit model**. Check that 16 observations were fitted, then read the parameter estimates and the fitted line. The fitted equation is

$$\hat{y} = -0.4599528 + 0.9849472x.$$

A one-unit increase in x corresponds to an estimated 0.98495-unit increase in mean y, within the scope of this fitted relationship. The intercept refers to $x = 0$, outside the observed range 3.1–10.3; its numerical availability does not make that extrapolation scientifically meaningful.

Term	Estimate	Standard error	95% interval
Intercept	-0.459953	0.157007	[-0.796699, -0.123207]
x	0.984947	0.022229	[0.937270, 1.032624]

The residual sum of squares is 0.3848933, $R^2 = 0.9929195$ and the displayed RMSE is $\sqrt{\text{SSE}/16} = 0.1550994$. The residual variance used for classical coefficient inference instead divides by the 14 residual degrees of freedom, giving 0.02749238. These two denominators answer different questions.

At $x = 6$, the fitted mean is 5.4497304. Its confidence interval would describe uncertainty about a mean response at that predictor value. An individual new observation has additional residual variation and needs a prediction interval. The displayed fitted-line band is a mean-response band.

High R^2 says the line describes most variation in this constructed sample. It does not establish agreement between the two measurements, good future performance or absence of bias.

12.3 Model scale and assumptions

Ordinary least squares minimizes the sum of squared residuals. With design matrix X, it estimates the coefficients of $y = X\beta + \varepsilon$. Classical small-sample inference assumes an appropriate mean function, independent normally distributed errors with constant variance, and an identifiable design.

Polynomial terms describe curvature without requiring a nonlinear coefficient fit. Exponential and power models fit $\log(Y)$ and require positive responses. Logarithmic and power models require positive X and reject categorical predictors. Linear and polynomial models permit categorical predictors; their indicators are not raised to powers.

For exponential and power fits, reported metrics use the log-response scale. Exponentiating a fitted log mean gives a geometric-scale prediction; without an appropriate adjustment it is not generally the arithmetic mean on the original scale. Interval limits are calculated on the log scale and exponentiated.

12.4 Read diagnostics before drawing conclusions

The primary view is a fitted relationship for one factor, or Actual by Predicted for several. Under **Diagnostics**, inspect **Residuals by predicted**, **Normal Q-Q of residuals**, and optional **VIF in parameter estimates**. Penalized models offer **Coefficient trace**. Use **Export PNG** for the displayed plot and **Add to results** to retain the result.

Curvature in residuals suggests a misspecified mean. A changing vertical spread suggests unequal variance. Time or group dependence must also be assessed from collection order and design; a residual-versus-fitted plot cannot establish independence. A Q–Q tail departure can signal unusual errors or a few influential observations.

The fit uses its complete eligible population. Scatter and residual displays are limited to 2,000 points and Q–Q displays to 1,000; a smooth fitted curve and its band are evaluated on bounded grids. The graphic is not a row-by-row audit of a very large dataset.

12.4.1 Collinearity and VIF

For an encoded predictor x_j,

$$\mathrm{VIF}_j = \frac{\mathrm{SST}_j}{\mathrm{SSE}_j} = \frac{1}{1 - R_j^2},$$

where the auxiliary regression predicts that column using an intercept and the other predictor columns. Large VIF indicates that the available design provides little separate information about its coefficient.

A constant or exactly reproduced column has undefined or infinite VIF. HypoNum also reports unavailable when the auxiliary residual is unresolved at working precision, rather than inventing a large finite value. Finite large VIF values are retained. Units and origins alone should not determine this diagnosis. For a multi-level factor, these are the VIFs of its encoded columns, not a single generalized factor VIF.

12.5 Choose a covariance model

On the same fitted rows, changing the standard-error mode preserves OLS coefficients, fitted values, residuals and sums of squares. It changes the estimated covariance of the coefficients and inference based on it. CR1 can additionally omit rows without a cluster identifier.

Let X have N fitted rows and K columns including the intercept, $B = (X^\mathsf{T}X)^{-1}$, residuals e_i, and leverage $h_i = x_i B x_i^\mathsf{T}$. Then:

Classical
$V = s^2 B$, where $s^2 = e^\mathsf{T}e/(N-K)$. Coefficient inference uses t_{N-K}.

HC3

$$V = B\left[\sum_i x_i^\mathsf{T} x_i \frac{e_i^2}{(1-h_i)^2}\right] B.$$

This addresses heteroskedastic covariance under the model's independence conditions. HypoNum uses t_{N-K} reference inference. A row with $1 - h_i \le 10^{-10}$ makes HC3 unavailable and stops the fit.

Cluster-robust CR1
For cluster score $u_g = \sum_{i \in g} x_i^\top e_i$,

$$V = \frac{G}{G-1}\frac{N-1}{N-K} B \left(\sum_g u_g u_g^\top \right) B.$$

Coefficient inference uses t_{G-1}. At least two clusters are needed; fewer than 30 triggers a warning about unreliable small-cluster inference.

HC3 does not correct correlation among rows. CR1 allows within-cluster dependence under an independent-cluster framework; choose clusters from the sampling or dependence structure, not by searching for favourable standard errors. Neither method repairs a wrong mean function or selection bias.

Cluster values are compared exactly as text. Missing values, NaN and empty text have no cluster identity and are omitted. This omission happens *after* the missing-value strategy, so such rows can have contributed to an earlier imputation. The omission count includes rows also missing X or Y. VIF uses the resulting fitted population.

Both robust modes require full rank. Aliased terms stop those fits. Classical OLS can retain a minimum-norm solution for a singular design; individual aliased coefficients then lack a unique interpretation. Robust modes are unavailable for Ridge and Lasso.

12.5.1 Joint tests and confidence bands

For q slopes and their covariance block V_{SS}, the robust joint test uses

$$W = \hat{\beta}_S^\top V_{SS}^{-1} \hat{\beta}_S, \qquad F = W/q.$$

Its reference degrees of freedom are $(q, N-K)$ for HC3 and $(q, G-1)$ for CR1. A singular covariance block makes the test unavailable; CR1 also requires $q \leq G-1$. Read the stated reason when no test is shown. The ordinary ANOVA table retains its classical F test and its constant-variance, independent-row interpretation.

For design row x_0, the mean-response interval is

$$\hat{y}_0 \pm t_{1-\alpha/2,\nu} \sqrt{x_0 V x_0^\top},$$

with $\nu = N-K$ for classical/HC3 and $G-1$ for CR1. The band identifies its covariance convention. It is pointwise mean uncertainty, not a simultaneous band or a robust individual prediction interval.

12.6 Understand this dialog's validation

Missing values are handled before any split in this workflow, including selected numeric response values when an imputation option is chosen. Imputed responses are not additional observations. Consider the missingness mechanism before treating the resulting inference as ordinary observed-data inference.

With positive holdout, Ridge and Lasso tune their penalty on the same holdout later used to report validation metrics. Those metrics therefore include model-selection optimism. With zero holdout, penalty tuning uses the selected shuffled seeded K-fold procedure and displays a refit.

For a separate test set and training-fitted preprocessing, use **Validated regression**, described in section 16.1. In this Regression dialog, a random row holdout can split one cluster between

training and validation. CR1 standard errors use training rows and training clusters, but the validation is not cluster-independent. Do not confuse a dependence-aware covariance with a dependence-aware prediction split.

12.7 Search for a predictor subset

Open **Model selection...** for one numeric response and numeric candidate predictors. The defaults are Stepwise, direction Both and AIC. Forward and Backward directions and BIC are alternatives. Best subsets searches all subsets, up to 15 candidates; stepwise allows up to 20.

One common complete-case population is fixed across candidates. Constant or collinear candidates, insufficient rows, a non-varying response and failed fits are refused. The selected model is refitted, with fit metrics, coefficient estimates, ordinary tests and 95% intervals. Stepwise reports the path; best subsets reports the best model at each size.

Lower AIC or BIC indicates a preferable relative fit under the selected criterion. It does not validate the chosen model on new data. Ordinary coefficient p-values after selection do not account for the search that selected those predictors. Retain the candidate set and selection procedure, and evaluate predictive claims with independent data or a suitable nested validation design.

12.8 Whole-factor linear effects

Choose **Statistics > Compare means & ANOVA > Linear effects (ANOVA / ANCOVA)...** for an additive model of one continuous response with categorical factors and numeric covariates. Declare factors explicitly, including numeric codes that denote levels rather than quantitative distances.

The fit uses one complete-case population after active exclusions. Factor levels are sorted deterministically, the first becomes the reference, and the result identifies the encoded comparisons. The design must have full rank, positive residual degrees of freedom and positive residual variance. Aliased or constant designs are refused.

12.8.1 A factor is more than one coefficient

A factor with a levels generally contributes $a - 1$ coefficients. Its whole-term Type II partial F test asks whether those coefficients are jointly needed after the other additive terms. The numerator degrees of freedom are $a - 1$; the denominator uses the full model's residual degrees of freedom.

Choose **File > Import data...** and select `sample-data/manual/paired-measurements.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Compare means & ANOVA > Linear effects (ANOVA / ANCOVA)...**. Set **Response** to **x**; select only **g** as a categorical factor and leave covariates empty. Click **Run linear effects** and inspect the factor test and adjusted means for the 16 complete rows. The adjusted means are 6.025 for A and 7.600 for B; the factor test is the same one-way classic ANOVA shown earlier. If another covariate is added, the adjusted means answer a different question because they condition on that covariate.

12.8.2 Interpret adjusted means and comparisons

Numeric covariates are held at their retained complete-case means. Predictions are averaged equally over the levels of every other factor. These are reference-grid means, not necessarily raw means or predictions averaged over the observed frequency distribution.

The display shows individual 95% t intervals. Pairwise differences use the same OLS covariance and receive Holm-adjusted p-values within each factor's family. Their displayed 95% intervals remain individual and non-simultaneous. A Holm-adjusted p-value does not turn an unadjusted interval into a simultaneous interval.

Select a factor to view adjusted predictions and intervals in response units. Read the conditioning convention and exact table beside the plot. The additive model assumes that each factor effect applies consistently across the other terms; use a suitable interaction model when that is not credible.

12.9 General linear models

Use **General linear model...** under **Statistics > Compare means & ANOVA** when the mean model needs interactions, quadratic covariates, one-level nesting or precision weights. The response is continuous. This is a Gaussian linear-model workflow; the word *general* does not make it a generalized linear model for binary or count responses.

12.9.1 Build a model from declared terms

Confirm the response shown on opening. Add numeric covariates and categorical factors through the available and selected lists. A numeric column assigned as a factor is treated as discrete levels. An intercept and main effects are included automatically.

Additional terms start empty. Choose **Interaction**, **Quadratic** or **Nested**, select the relevant columns and press **Add term**. For nesting, choose the parent first and child second. Remove individual terms or use **Clear terms**. HypoNum accepts these declared terms through the controls; this workflow does not evaluate a free-form formula.

An interaction allows one effect to change with another predictor. For example,

$$E(Y) = \beta_0 + \beta_A A + \beta_x x + \beta_{Ax} Ax$$

has slope β_x at $A = 0$ and $\beta_x + \beta_{Ax}$ at $A = 1$. A single common-slope interpretation would omit that feature of the model.

Nested child contrasts are formed separately inside each parent. A machine uniquely belonging to a plant can be nested within plant; a treatment appearing in every plant is crossed with plant. Nested children do not also receive crossed main effects. Nesting chains and crossed interactions involving a nested child are refused.

12.9.2 Choose the hypothesis convention

Categorical terms use sum-to-zero contrasts. The initial test is Type III. Its all-other-term partial Wald F tests condition each term on the remaining model, including interactions. Type II uses covariance-orthogonal contrasts that preserve marginality: a main effect is not adjusted for interactions containing it. Quadratic numeric terms remain adjustment terms in Type II tests.

The conventions test different hypotheses in some designs. State the chosen convention with the model terms; do not switch types merely to obtain a preferred p-value. When interaction is material, interpret the relevant conditional comparisons rather than an isolated averaged main effect.

12.9.3 Precision weights

The initial weight is **(none)**, giving ordinary least squares. A selected numeric weight must be finite and strictly positive for every eligible weight row; an invalid value stops the fit. Weights

represent inverse *relative error variances*. They are not survey sampling weights or counts of repeated observations.

For example, weights 1 and 4 assert that the second observation has one quarter the error variance of the first, under the model. The weighting assumption needs evidence from the measurement process. Observed response size alone does not justify assigning it a larger weight.

12.9.4 Adjusted predictions and output

Numeric covariates are held at complete-case means; other factors receive equal level weights in the reference grid. Means and contrasts retain response units, and residual variance has squared response units. Individual intervals are fixed at 95%; within-factor comparison p-values use Holm adjustment. The intervals are not simultaneous.

Fit x on factor g alone in the small paired-measurements file, with no weights. The adjusted means equal the raw group means because there are no additional terms to adjust for. Adding covariates or an unequally weighted reference grid changes what those predictions mean.

Fit model opens model summaries, term tests, adjusted means, contrasts and row-use counts. The factor-selectable margins plot shows the same reference-grid estimates and individual intervals as the table. **Cancel** stops pending work; **Add to results** saves settings and the result snapshot. Editing the source clears the live result.

There is no imputation. Incomplete response, covariate or factor rows are omitted and counted separately from user exclusions. Full rank and positive residual degrees of freedom are required. Limits are one million source rows, twelve factors, 32 covariates, 32 levels per factor, 128 terms and twenty million design cells. Singular or unsupported models fail explicitly.

12.10 Fitting a nonlinear curve

A nonlinear mean function can express a scientific relationship more directly than a high-degree polynomial. In an exponential decay, for example, the rate has an interpretation as a rate of change and the limiting value has the response's units. The statistical model is

$$y_i = f(x_i, \theta) + \varepsilon_i, \qquad E(\varepsilon_i) = 0, \quad \mathrm{Var}(\varepsilon_i) = \sigma^2.$$

The observations are independent and the error variance is constant on the *original response scale*. Logging the response would fit a different error model.

12.10.1 Choose a curve and plausible starting values

Open **Statistics > Correlation & regression > Nonlinear regression...**. Select distinct numeric X and response columns. Excluded rows and rows with missing or non-finite selected values do not enter the fit. Choose a curve, enter starts and any bounds, then choose **Fit model**.

Curve	Meaning and initial parameters
Exponential decay	$f(x) = c + ae^{-kx}$, starts $a = 1, k = 1, c = 0$. a, c have response units; k has reciprocal-X units.
Asymptotic growth	$f(x) = c + a(1 - e^{-kx})$, the same starts. The value at zero is c, and the limiting value is $c + a$.

Curve	Meaning and initial parameters
Four-parameter logistic	$f(x) = \text{lo} + (\text{hi} - \text{lo})/[1 + (x/\text{mid})^{\text{slope}}]$. Starts are $0, 1, 1, 1$ in that order. X must be positive. The small-X limit is hi and the large-X limit lo; their ordering is free.
Custom expression	Enter an expression and one to eight parameter names. The initial expression is `c+a*exp(-k*x)`.

Changing the curve resets starts and bounds. A finite start must be strictly inside its bounds and must give finite fitted values for the selected X. Blank bounds mean unbounded. Library rate, midpoint and logistic slope parameters initially have lower bound 10^{-12}; choose scientifically appropriate bounds in the actual measurement units. The logistic midpoint has X units and its slope is dimensionless.

The confidence level starts at 95% and accepts whole percentages 51–99. The solver evaluation budget starts at 1,000 and can be set from 1 to 10,000. Increasing a budget cannot resolve a structurally unidentified model. Try several scientifically plausible starts when competing local solutions are possible; the workflow does not automatically search many starts.

Custom expressions allow numbers, `x`, declared parameter names, parentheses, arithmetic `+-*/**`, and the functions `exp`, `log` and `sqrt`. Names must be distinct ASCII identifiers; leading underscores and the variable/function names are reserved. Every declared parameter and X must occur. There are limits of 2,048 characters, 128 syntax nodes and 16 nesting levels. This is a restricted formula language. Invalid domains or overflow produce an error with guidance to revise the data, starts or bounds. Custom constraints remain your choice.

12.10.2 Worked example: decay toward a background level

Choose **File > Import data...** and select `sample-data/manual/nonlinear-decay.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Correlation & regression > Nonlinear regression...**. The file contains 30 constructed observations with X evenly spaced from 0 to 5. Use the file's full precision. Representative values, rounded here for inspection, are

$$(x, y) = (0, 5.0300),\ (1.03448, 2.71843),\ (2.06897, 1.79415),\ (3.10345, 1.30425),\ (5, 1.05047).$$

Select X `x`, response `y`, and Exponential decay. Enter starts $a = 4$, $k = 0.8$, $c = 1$, bounds $0 < a < 10$, $0.01 < k < 3$, $0 < c < 5$, confidence 95%, and budget 1,000. Click **Fit model**. Inspect the convergence status, coefficient estimates and residual sum of squares. The fitted mean is

$$\hat{y} = 1.0118683 + 4.0126115\, e^{-0.8125882x}.$$

The original-scale residual sum of squares is 0.01184819 and the displayed RMSE is $\sqrt{\text{SSE}/30} = 0.01987309$. The fitted level tends toward about 1.012 response units. The fitted amplitude is about 4.013, so the fitted value at zero is about 5.024. The rate is about 0.813 per X unit. Its half-life, $\log(2)/\hat{k}$, is about 0.853 X units.

Inspect residuals against X for changing spread or remaining curvature. A small residual sum of squares in this constructed example does not establish that the same curve is appropriate for a new scientific process. Extending far beyond X=5 assumes the fitted asymptote remains meaningful.

12.10.3 Understand the uncertainty

The fit minimizes SSE $= \sum_i [y_i - f(x_i, \theta)]^2$. At least $n > p$ eligible observations and varying X are required, where p is the number of parameters. A rank-deficient sensitivity matrix means the data do not identify every parameter locally; fitting fails.

With a usable, full-rank Jacobian J, no active parameter bound and residual degrees of freedom $n - p$, the local covariance is

$$s^2 = \frac{\text{SSE}}{n-p}, \qquad V = s^2 (J^\mathsf{T} J)^{-1}.$$

The parameter interval is

$$\hat{\theta}_j \ \pm \ t_{1-\alpha/2, n-p} \sqrt{V_{jj}}.$$

For the mean response, with gradient $g(x) = \partial f(x, \theta)/\partial \theta$ at the fitted parameters, the interval is

$$\hat{f}(x) \ \pm \ t_{1-\alpha/2, n-p} \sqrt{g(x)^\mathsf{T} V g(x)}.$$

These are local linear approximations, not intervals for a new individual observation or simultaneous bands over all X. Strong curvature, a weakly identified parameter or a solution on a bound can invalidate the approximation. At an active bound or a Jacobian condition number above 10^{10}, HypoNum withholds covariance and intervals and explains why. An exact fit gives zero-width local intervals; that numerical result does not establish zero uncertainty about the real process.

12.10.4 Read, save and reuse the result

The result includes starts, bounds, estimates, approximate intervals, convergence information, Jacobian rank and condition, SSE, RMSE and original-scale residuals. A successful local fit can depend on the chosen starts and bounds. Exhausted budgets and non-convergence produce an error.

The row table displays up to 200 source rows, preserving missing/excluded positions; plots display up to 400 eligible observations. Every eligible row contributes to fitting. The fitted line connects mean values at the displayed X values; available live shading represents mean-response uncertainty.

Choose **Add to results** to retain the parameter table, row preview, fitted-mean plot and residual plot. The saved fitted-mean plot does not retain the live confidence shading. Reopening a project restores the saved result. Choose **Restore fit** explicitly to load its retained parameters. **Predict new X** applies them to the same-named numeric X column of the active table, without requiring Y or refitting. Values outside the fitted X range receive an extrapolation warning.

A source or exclusion change makes the previous calculation stale for recording. Closing the calculation window cancels ongoing work. The source limit is one million rows and twenty million selected cells. This workflow does not provide weighted or robust fitting, automatic multistart selection, profile-likelihood or bootstrap intervals, or creation of prediction columns.

12.10.5 Further reading

The NIST/SEMATECH Engineering Statistics Handbook discusses nonlinear least-squares regression, starting values and model assumptions. Use the separate validation report for retained numerical reference evidence.

12.11 Comparing two measurement methods

When both measurements contain error, ordinary regression of Y on X treats one side as error-free and can distort the estimated relationship. Deming regression models a straight line with error in both coordinates. It requires a known ratio of their error variances, independent measurement pairs, and constant error variances over the measurement range.

Open **Statistics > Correlation & regression > Deming measurement-method comparison...**. Select two numeric columns measured on the same items: one row is one pair. Choose the ratio, interval method and confidence level, then run the fit.

12.11.1 Declare the error ratio in the correct direction

HypoNum uses

$$\lambda = \frac{\mathrm{Var}(\text{Y measurement error})}{\mathrm{Var}(\text{X measurement error})}.$$

The initial value is 1, which assumes equal error variances in the entered units. Obtain a defensible ratio from replicate measurements or relevant measurement-system evidence. Sample variance across the items includes real item differences and does not generally estimate measurement-error variance.

If X error SD is 2 units and Y error SD is 1 unit, enter $\lambda = 1^2/2^2 = 0.25$. Some references use the reciprocal ratio; in that convention the same study has ratio 4. Changing units also changes the variance ratio. Checking **Comparable units** enables the identity line $y = x$; it does not convert either column's units.

Setting	Initial choice and meaning
X and Y measurements	First and second numeric columns; verify the pairing.
Variance ratio	1; a positive finite Y-error/X-error variance ratio. The entry field has minimum 10^{-12}.
Comparable units	Off. Turn on only when the identity line has a meaningful measurement interpretation.
Coefficient intervals	Jackknife t (leave-one-out), initially. Paired percentile bootstrap and None are alternatives.
Confidence	95%; whole percentages 51–99.
Bootstrap resamples	1,000 when that method is selected; range 200–5,000.
Seed	0 for the bootstrap; range 0–2,147,483,647. Retain it to reproduce the resampling.

12.11.2 Worked example: ten paired measurements

Choose **File > Import data...** and select `sample-data/manual/deming-measurements.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Correlation & regression > Deming measurement-method comparison...**. Set X to **method_a**, Y to **method_b**, and the Y-error/X-error variance ratio to **0.25**. Under **Coefficient intervals**, choose **Jackknife t (leave-one-out)** and **Confidence (%)** 95. Click **Compare methods**; inspect the coefficient table's estimates, standard errors and limits. These published example values are reproduced in the NIST Dataplot Deming-fit documentation, which credits NCSS.

X	Y	X	Y
7.0	7.9	8.2	7.3
8.3	8.2	10.2	10.2
10.5	9.6	10.3	10.6
9.0	9.0	7.1	6.3
5.1	6.5	5.9	5.2

Select X `method_a`, Y `method_b`, ratio 0.25, Jackknife t intervals and confidence 95%. The fitted line is

$$\hat{y} = -0.0897449 + 1.0011942x.$$

Coefficient	Estimate	Jackknife SE	95% interval
Intercept	-0.089745	1.721987	[-4.060655, 3.881165]
Slope	1.001194	0.187177	[0.569563, 1.432825]

Both intervals use 8 degrees of freedom. The estimated line is close to identity, but the intervals are wide. Including 0 for the intercept and 1 for the slope does *not* demonstrate equivalence. A method-agreement decision needs acceptable differences specified in measurement units, a suitable design and an analysis targeted to that decision.

12.11.3 How the line and residuals are defined

For the eligible complete pairs, let S_{xx}, S_{yy}, S_{xy} denote centred sums of products. Then

$$D = S_{yy} - \lambda S_{xx}, \qquad b = \frac{D + \sqrt{D^2 + 4\lambda S_{xy}^2}}{2S_{xy}}, \qquad a = \bar{y} - b\bar{x}.$$

At least three complete pairs and variation in both measurements are required. An undefined or effectively vertical line is refused. No ordinary least-squares fit is substituted.

The fitted line at an observed X is $a + bx$. The estimated latent points, which account for error on both axes, are

$$\hat{x}_i = \frac{\lambda x_i + b(y_i - a)}{\lambda + b^2}, \qquad \hat{y}_i = a + b\hat{x}_i.$$

The displayed residuals $x_i - \hat{x}_i$ and $y_i - \hat{y}_i$ keep their respective measurement units. Swapping the two measurements requires replacing λ by $1/\lambda$.

12.11.4 Interpret the interval method

For a coefficient θ, write $\hat{\theta}_{(-i)}$ for its estimate after deleting pair i, and $\bar{\theta}_{(\cdot)}$ for the mean of those estimates. The jackknife method uses

$$SE_J^2 = \frac{n-1}{n} \sum_i (\hat{\theta}_{(-i)} - \bar{\theta}_{(\cdot)})^2, \qquad \hat{\theta} \pm t_{1-\alpha/2,n-2} SE_J.$$

The centre is the full-sample estimate. At least four pairs are needed for these intervals. If any deletion makes the fit undefined, both coefficient intervals are withheld with a reason. The approximation treats λ as fixed and does not include uncertainty from estimating that ratio.

The paired bootstrap samples whole pairs with replacement. It reports percentile intervals only when at least $\max(100, \lceil 0.9B \rceil)$ of B requested fits are valid. Invalid draws and their count are disclosed; an interval based on the remaining draws is conditional on their validity. The requested total of pair selections is limited to five million.

Both methods provide approximate, marginal coefficient intervals. In the retained study with 40 pairs, equal error variances, latent X uniform on $[0, 10]$ and 1,000 simulated datasets, the 95% jackknife intervals covered slope and intercept in 94.0% and 93.8% of datasets. The percentile slope interval with 500 resamples covered in 92.7%. The percentile result therefore carries a measured-coverage caution. These results describe that design; they are not guarantees for other sample sizes, error distributions or variance ratios.

12.11.5 Keep a reviewable result

The result shows the pair plot, fitted line, optional identity line, coefficients, uncertainty method, row counts, projected values and residuals. The table previews up to 200 source rows and the plot up to 400 pairs; all eligible pairs contribute. Missing and excluded pairs remain identifiable.

Choose **Add to results** to retain the tables, plot, ratio, unit declaration, confidence settings and resampling settings, when used. A saved historical result retains its original interval method. New analyses default to the jackknife. Changing the source or exclusions requires a new calculation before recording. The source limit is one million rows and twenty million selected cells. Weights, automatic unit conversion and nonlinear measurement relationships are outside this workflow.

12.11.6 Sources

The NIST Dataplot Deming-fit reference supplies the numerical example and jackknife calculation. See also NCSS, Deming Regression, and Linnet (1993), *Clinical Chemistry* 39(3), 424–432. Check each source's variance-ratio direction before reproducing a calculation.

CHAPTER 13

Models for categories and counts

A binary outcome, an ordered rating and an event count require different response models. The link function connects predictors to a probability or mean; it also determines how to interpret a coefficient.

Open these workflows under **Statistics > Generalized & mixed models**. Start with the response definition and the observational unit. Repeated outcomes from one subject do not become independent simply because the response is binary or a count; the dependent-data chapter describes alternatives.

13.1 Binary logistic regression

Choose **Binary logistic...**, an outcome with exactly two complete values, the **Event to model**, and at least one predictor. The model is

$$\log \frac{p_i}{1-p_i} = \beta_0 + x_i^{\mathsf{T}}\beta, \qquad p_i = \frac{\exp(\eta_i)}{1+\exp(\eta_i)},$$

where p_i is the probability of the declared event. Reversing the event changes the interpretation and signs of the effects.

Numeric predictors enter continuously. Categorical predictors use indicators, with the first sorted level as reference. The workflow allows up to 20 predictors and 20 levels per categorical predictor. Rows missing any required field are omitted together. Constant predictors, collinear encoded columns and designs without more complete rows than coefficients are refused.

13.1.1 Odds ratios and probabilities

A coefficient β_j gives odds ratio $\exp(\beta_j)$ for a one-unit increase, holding other predictors fixed. An odds ratio of 2 is not a doubling of probability. As an illustrative calculation, a starting probability of 0.20 has odds $0.20/0.80 = 0.25$; doubling those odds gives probability $0.5/1.5 = 1/3$. At a starting probability of 0.80, doubling odds instead gives $8/9$.

Results include coefficients, standard errors, z tests, p-values, odds ratios and fixed 95% odds-ratio intervals. Overall summaries include the likelihood-ratio test, partial comparison with an intercept-only model through McFadden R^2, log likelihood, AIC and BIC. McFadden R^2 is not the fraction of response variance explained in ordinary regression.

The classification threshold defaults to 0.5 and must lie strictly between zero and one. It changes predicted class labels, confusion counts, accuracy, sensitivity, specificity and predictive values; it does not change the fitted coefficient meaning. ROC and AUC summarize discrimination over thresholds. These fit-sample summaries are not independent evidence of future accuracy or probability calibration.

A separation warning indicates unstable coefficients or odds ratios even when classification looks excellent. Sparse events and too many predictors can produce substantial uncertainty. Read fit warnings before interpreting a large odds ratio.

13.2 Ordinal logistic regression

Choose **Ordinal logistic...** for an ordered outcome with at least three and at most twenty levels. Confirm the low-to-high **Level order**; enter a comma-separated order when text sorting does not express the intended ranking. Numeric labels are naturally ordered.

The proportional-odds model uses common slopes across cut points:

$$\log \frac{P(Y_i \le j)}{P(Y_i > j)} = \theta_j - x_i^\top \beta.$$

Thus $\exp(\beta_j)$ is the multiplier of the odds of being above any given cut point for a one-unit predictor increase. One coefficient applies to all thresholds. This is a substantive assumption about the response, not a consequence of having ordered labels.

Select one or more predictors, up to twenty, with at most twenty levels per categorical predictor. Incomplete rows are omitted and counted; constant or collinear predictors, insufficient rows and failed fits are refused.

13.2.1 Read thresholds and effects separately

The result identifies the ordered levels, coefficients, standard errors, z tests, p-values, cumulative odds ratios and fixed 95% intervals. A separate table gives fitted thresholds. Thresholds locate category boundaries on the latent logit scale; they are not predictor effects. Model summaries include row count, McFadden R^2, likelihood-ratio p-value, AIC and BIC.

For an illustrative coefficient $\log(1.5)$, the higher-category odds multiply by 1.5 at every cut point. If the effect is positive at one threshold and negative at another in the underlying population, this proportional-odds structure cannot describe both correctly. The reported odds ratio does not certify that the assumption holds.

Complete or quasi-complete separation means the joint cumulative-link likelihood lacks a finite maximum. A verified case stops the fit and identifies involved levels. HypoNum checks the joint model with common slopes and ordered thresholds, rather than treating separate binary threshold fits as the same model. Near separation can leave a finite but poorly determined fit and is reported as a warning.

13.3 Nominal logistic regression

Choose **Nominal logistic...** for an unordered outcome with three to twelve categories. Select the response, predictors and reference category; the default reference is the first naturally sorted category. There may be up to twenty predictors and twenty levels per categorical predictor.

For non-reference category j,

$$\log \frac{P(Y_i = j)}{P(Y_i = r)} = \beta_{0j} + x_i^\top \beta_j.$$

There is a coefficient set for each category against reference r. Changing the reference changes those comparisons, not the fitted category probabilities.

The coefficient table reports estimates, standard errors, z tests, p-values, relative-risk ratios and fixed 95% intervals. An exponentiated coefficient multiplies the probability ratio $P(Y = j)/P(Y = r)$; it is not itself a probability or a multiplier for $P(Y = j)$ alone. Row and category counts, McFadden R^2, likelihood-ratio p-value, AIC and BIC complete the summary.

Rows incomplete in required fields are omitted. The response cannot also be a predictor. Constant/collinear encoded columns, insufficient rows and failed fits are refused. Verified complete

or quasi-complete separation stops the fit; no penalized estimate is substituted. Large coefficients alone are not treated as a proof of separation. If numerical verification cannot resolve the separation check, a warning accompanies the fit's own convergence diagnostics. Sparse categories remain a concern even without certified separation.

13.4 Poisson regression and exposure

Choose **Poisson (count)...** for a numeric event count. The log-link model is

$$Y_i \sim \text{Poisson}(\mu_i), \qquad \log \mu_i = \beta_0 + x_i^{\mathsf{T}}\beta + \log E_i,$$

where optional exposure $E_i > 0$ is entered as an offset with coefficient fixed at one. Without exposure, the response is a count per observation; with exposure it describes a rate per exposure unit.

Worked example

Suppose two machines have expected event rate 0.2 per operating hour. At 10 and 50 hours their expected counts are 2 and 10. The greater count does not imply a less reliable machine; it reflects more exposure. A coefficient of $\log(0.8)$ would multiply the rate by 0.8 at fixed predictors and exposure, a 20% reduction in expected rate. These values illustrate the model, not a fitted dataset.

Choose at least one predictor, up to twenty, with at most twenty levels per categorical predictor. Required missing or non-finite fields are omitted and counted. Counts must be finite and nonnegative; noninteger values are accepted with a warning because ordinary Poisson outcomes are counts. Nonpositive exposure, constant or collinear predictors, insufficient rows, a nonvarying response or a failed fit prevents a result.

The Poisson variance equals its conditional mean. The result reports Pearson dispersion and flags overdispersion or underdispersion, alongside deviance, log-likelihood-based summaries and coefficient inference. A dispersion warning may reflect omitted predictors, dependence or an inappropriate count model. It does not automatically select a replacement.

Rate ratios and fixed 95% intervals exponentiate coefficients and their limits. Report the exposure unit: a rate per hour and a rate per year use different intercepts even if relative effects agree.

13.5 Negative binomial regression

Choose **Negative binomial (count)...** when the NB2 variance structure is appropriate:

$$\log \mu_i = \beta_0 + x_i^{\mathsf{T}}\beta + \log E_i, \qquad \operatorname{Var}(Y_i) = \mu_i + \alpha\mu_i^2, \quad \alpha > 0.$$

Optional exposure has the same rate interpretation. The extra parameter allows conditional variance above the mean; it does not explain dependence or a separate structural-zero process.

Coefficients and dispersion are estimated jointly by maximum likelihood. HypoNum obtains coefficient standard errors from the coefficient block of the inverse *joint observed information*, including estimated dispersion. Software that conditions on fitted dispersion and uses expected information can give different finite-sample standard errors despite similar estimates.

Wald intervals use the selected confidence level. Exponentiating a coefficient or its limits gives a rate ratio or rate-ratio limits. These are asymptotic model-based intervals, not exact or robust ones.

13.5.1 The Poisson boundary

As α approaches zero, NB2 approaches Poisson. Ordinary NB2 Wald covariance need not exist at that boundary. A tiny fitted dispersion and similar likelihoods alone do not establish a boundary: HypoNum also checks the one-sided dispersion score against numerical error.

A verified Poisson boundary stops NB2 inference; an unresolved boundary also withholds Wald inference with that explanation. A singular information matrix prevents the fit. Choosing Poisson is an analyst's subsequent modeling decision, not a silent substitution by the software.

CHAPTER 14

Repeated observations and dependent data

Repeated measurements carry information about change within an entity, but they usually do not supply as many independent pieces of information as the same number of unrelated observations. Identify the participant, machine, site or other unit that generates dependence before choosing a model.

Panel fixed effects estimate within-entity slopes while absorbing separate entity intercepts. Gaussian mixed models describe variation in random intercepts or slopes. Generalized mixed models do so for binary or count responses. GEE instead targets population-average effects with a working correlation and sandwich covariance. Repeated-measures covariance models directly describe residual correlation across occasions.

These are different statistical assumptions, not interchangeable corrections to an ordinary regression. Decide whether the scientific target is a within-entity association, a conditional group-specific effect, or an average effect across the population. Record the grouping and occasion definitions, exclusions and any incomplete observation patterns.

14.1 One-way panel fixed-effects regression

Use panel fixed effects when repeated observations on the same entities permit a within-entity comparison. The model

$$y_{gt} = \alpha_g + x_{gt}^{\mathsf{T}}\beta + \varepsilon_{gt}$$

assigns a separate intercept to every entity and common slopes to the numeric predictors. Time-invariant differences between entities are absorbed by those intercepts. A slope describes association with changes *within* an entity, not an overall comparison between entities.

The model does not establish causation or remove time-varying confounding. The Time ID identifies occasions and validates keys; it does not automatically add time fixed effects or a time trend.

14.1.1 Prepare and run a panel

Open **Statistics > Generalized & mixed models > Panel fixed-effects regression...**. Select one numeric Response, Entity ID, Time ID and one to 32 numeric additive Predictors. Entity and time may be numeric or categorical identifiers. All roles must be distinct and each valid entity/time pair unique.

Opening setup resets predictors, leaves entity and time unselected, shows the first available numeric response, and restores 95% confidence. Confidence choices are 90, 95 and 99%. The effective desktop maximum defaults to one million source rows; the row control can lower it. The calculation also limits selected input to twenty million cells and complete-case entities to 10,000 before singleton removal.

Choose **Run panel fixed effects**. Cancel or close setup to stop pending work. A cancelled or superseded calculation cannot replace a newer result.

14.1.2 Understand which rows contribute

Exclusions are removed first. Remaining model rows need finite response and predictors and nonmissing entity/time IDs. Duplicate valid entity/time keys stop the calculation *before* numeric complete-case omission: a duplicate record does not become an acceptable key merely because its response is missing.

An entity with only one complete row is a singleton and is then removed and disclosed. Unbalanced panels are allowed. At least two retained entities, positive residual degrees of freedom and a full-rank within design are required. Empty or all-singleton data, invalid roles, unresolved rank or non-finite/zero clustered standard errors prevent inference.

14.1.3 Within transformation and inference

The fit subtracts each entity's mean from its response and predictors:

$$\tilde{y}_{gt} = y_{gt} - \bar{y}_g, \qquad \tilde{x}_{gt} = x_{gt} - \bar{x}_g.$$

Predictors with no within-entity variation are absorbed and named in the result. Least squares on the remaining demeaned columns estimates the slopes. Entity intercepts are reconstructed as $\hat{\alpha}_g = \bar{y}_g - \bar{x}_g^{\mathsf{T}}\hat{\beta}$.

Within R^2 compares the fitted within residual sum of squares with $\sum_{gt}(y_{gt}-\bar{y}_g)^2$. It is unavailable when that denominator is zero and should not be read as the overall variation explained by the entity intercepts.

Uncertainty uses entity-clustered CR1. With N retained rows, G entities and k remaining slopes, the finite-sample correction is

$$c = \frac{G}{G-1}\frac{N-1}{N-G-k}.$$

This counts all G absorbed intercepts as well as the k slopes. Some panel software uses another degrees-of-freedom convention, so matching slopes need not imply matching standard errors. Coefficient tests and confidence limits use t_{G-1}. Fewer than 30 entities triggers a small-cluster warning.

14.1.4 Worked example: changes within four entities

Choose **File > Import data...** and select `sample-data/manual/panel-measurements.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Generalized & mixed models > Panel fixed-effects regression...**. Set **Response** to **y**, **Entity ID** to **entity**, **Time ID** to **period**, and **Confidence** to 95%. Add only **x** to the numeric predictors and click **Run panel fixed effects**. Inspect the coefficient and entity-intercept tables in the results. The complete data are shown by entity; each cell is x, y.

Period	A	B	C	D
1	0, 5.3	1, 8.1	2, 12.2	0, 10.9
2	1, 6.3	3, 11.6	1, 10.8	2, 14.4
3	2, 8.1	2, 10.2	4, 14.7	1, 12.3
4	3, 9.3	4, 13.1	3, 13.3	3, 15.4

Choose y as response, entity and period as the identifiers, x as the predictor, and 95% confidence. All sixteen rows remain. Demeaning gives total predictor sum of squares 20 and response/predictor cross-product 29.3, so the slope is $29.3/20 = 1.465$.

The retained entity intercepts are 5.0525, 7.0875, 9.0875 and 11.0525 for A through D. Residual sum of squares is 0.8555. Here $N - G - k = 11$, correction $c = 20/11$, and the entity-clustered standard error is 0.09599953. Reference inference has only three degrees of freedom. This example demonstrates the calculation; four entities provide weak grounds for reliable cluster-robust inference, as the displayed warning states.

14.1.5 Inspect and retain the result

Read the slopes, covariance convention, absorbed predictors, singleton counts and entity/time coverage together. Fitted values and residuals retain source-row alignment; excluded, missing and singleton rows have missing output with a reason.

Coverage and row-fit tables display at most 200 entries, with a disclosure; counts refer to the full calculation. Results use six significant digits and support Escape to close. **Add to results** saves the settings, coefficients, all entity intercepts and coverage, warnings and bounded row-fit table. A reopened report is a saved snapshot, marked stale when the source changes, not an automatic refit.

14.2 A Gaussian model with a random intercept

Repeated observations within a group often share a common deviation. A random-intercept model represents that dependence as

$$y_{gi} = x_{gi}^{\mathsf{T}}\beta + b_g + \varepsilon_{gi}, \qquad b_g \sim N(0, \sigma_b^2), \quad \varepsilon_{gi} \sim N(0, \sigma_e^2).$$

Groups are independent; observations are independent conditional on their group effect under this model. The random effects are assumed independent of the specified predictors. This differs from absorbing a separate fixed intercept and estimating only within-entity slopes.

Choose **Statistics > Generalized & mixed models > Mixed models (random intercept)...**. Confirm the numeric response and group column. With no fixed effects selected, the fit has a fixed intercept and a random group intercept. Add numeric or categorical fixed effects using the transfer lists. A numeric column assigned as categorical denotes discrete levels.

Fixed effects are additive. This workflow offers no interactions, random slopes, observation weights or additional grouping levels. The group identifier cannot also be an ordinary fixed factor. Categorical effects compare each level with the first sorted text level; the reference is disclosed.

14.2.1 Choose ML or REML

Restricted maximum likelihood (REML) is the opening default. Maximum likelihood (ML) is an explicit alternative. Both constrain $\sigma_b^2 \geq 0$ and compare an interior variance solution with the zero boundary. A small positive optimum is retained rather than rounded to zero.

REML accounts for estimating fixed effects in its variance criterion. ML is appropriate when comparing likelihoods of models with different fixed-effect designs under otherwise compatible likelihood assumptions; REML likelihoods based on different fixed effects do not provide that ordinary comparison.

The numerical profile can have multiple peaks for unbalanced data. HypoNum searches and refines candidate peaks, compares them with the zero boundary and repeats with a denser search. An unresolved maximum, an increasing upper search edge or non-finite conditioning produces an error. This numerical search is disclosed with results and is not a proof that every possible narrow peak has been found.

Zero within-group residual variation can make the likelihood unbounded when unfitted within-group contrasts remain. That is different from a legitimate zero *between-group* variance. HypoNum refuses unbounded or numerically unresolved cases instead of returning a tiny arbitrary residual variance.

14.2.2 Read coefficients, variances and ICC

Fixed-effect standard errors use the fitted covariance as a plug-in asymptotic estimate. Intervals are fixed 95% normal intervals; they do not apply a Kenward–Roger or Satterthwaite small-sample correction.

The residual variance is $\hat{\sigma}_e^2$, random-intercept variance is $\hat{\sigma}_b^2$, and

$$\widehat{\mathrm{ICC}} = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_b^2 + \hat{\sigma}_e^2}.$$

ICC is the model's within-group correlation. For illustrative variances 4 and 1, ICC is $4/5 = 0.8$. The two variance estimates have squared response units; ICC has no units. Variance-component significance tests are not reported.

When the constrained optimum has $\hat{\sigma}_b^2 = 0$, the result states that boundary. It is not evidence that between-group variability is known to be exactly zero in the population.

14.2.3 Population and conditional predictions

A population fitted value is $x_{gi}^{\mathsf{T}}\hat{\beta}$. A conditional fitted value also uses the estimated group adjustment,

$$\hat{b}_g = \frac{\hat{\sigma}_b^2}{\hat{\sigma}_e^2 + n_g\hat{\sigma}_b^2} \sum_{i \in g} (y_{gi} - x_{gi}^{\mathsf{T}}\hat{\beta}).$$

The adjustment shrinks the observed group residual mean toward zero. With five observations, variances 4 and 1, and average residual 2, it is $(20/21)2 = 1.904762$. This is illustrative arithmetic. A small or noisy group receives stronger shrinkage.

At zero group variance, population and conditional fitted values coincide. For a new group, the old groups' fitted adjustments are not observed information about that new group. Do not report a conditional fit to existing groups as if it were independently measured new-group accuracy.

14.2.4 Worked comparison with panel fixed effects

Choose **File > Import data...** and select `sample-data/manual/panel-measurements.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported

order. Choose **Statistics > Generalized & mixed models > Mixed models (random intercept)...**. Set **Response** to **y**, **Group** to **entity**, and **Estimator** to **REML**. Add only **x** to the numeric fixed effects; leave categorical fixed effects empty. Click **Fit model**. The sixteen complete rows belong to four entities. Check the fixed coefficients and the population and conditional predictions in the results. The fixed intercept is 8.0699592 and slope 1.4650204. Estimated group variance is 6.6489278, residual variance 0.07776861 and ICC 0.9884388.

For entity A's first row, $x = 0, y = 5.3$, the population fit is 8.0699592. Adding its estimated group adjustment −3.0086921 gives conditional fit 5.0612671 and conditional residual 0.2387329. The slope is close to the panel within estimate for this constructed dataset, but the models make different assumptions and use different covariance calculations. Only four groups are present; treat the asymptotic intervals and variance estimates with the caution that a small number of groups requires.

14.2.5 Population checks and saved-result controls

Remove explicit exclusions first, then omit missing or non-finite required values without imputation. Every analyzed group must retain at least two observations, with at least two such groups overall. The fixed design must have full rank and more rows than its rank.

Limits are one million source rows, twenty selected fixed-effect columns, twenty retained levels per categorical effect and 100 encoded fixed columns. Singular, ill-conditioned or failed fits stop with a reason.

Fit model shows coefficient, variance, group-count and row-diagnostic tables. Diagnostics distinguish population and conditional predictions and preserve source-row identity. **Cancel** stops pending work; **Add to results** saves settings, counts, warnings and the result. Source edits clear the live result. Older saved results can identify an earlier optimizer and should be read as the calculation actually retained.

14.3 Nested, crossed and random-slope mixed models

Use this Gaussian mixed-model workflow when a single random intercept does not describe the design. It always includes a fixed intercept and additive numeric/categorical fixed effects. Choose one of two random structures: a correlated random intercept and slope within one group, or two to three independent random-intercept components.

14.3.1 Random slopes

A random-slope model allows groups to differ in both level and response to a numeric predictor:

$$y_{gi} = \beta_0 + \beta_1 x_{gi} + \cdots + b_{0g} + b_{1g} x_{gi} + \varepsilon_{gi}, \qquad \begin{pmatrix} b_{0g} \\ b_{1g} \end{pmatrix} \sim N(0, G).$$

The covariance in G describes association between group intercepts and slopes. The slope predictor must also be selected as a numeric fixed effect. Each of at least two groups needs at least three complete rows and within-group predictor variation.

Centre the predictor at a scientifically useful value when appropriate. The random intercept then describes variability at that value. Changing the origin changes intercept/slope covariance, while an equivalent reparameterized model preserves its marginal fit and predictions.

14.3.2 Crossed and nested components

In variance-components mode, declare two or three independent components. A crossed component uses the global levels of a column. A nested component uses the parent/child pair, so child label 1 in plant A is a different random unit from child label 1 in plant B.

For example, operators who measure every part are crossed with parts. Machines that belong to one plant are nested in plant. The program does not infer these relationships from repeated labels. Choosing the wrong structure changes which observations share a random effect and can seriously change uncertainty.

Add a component row and confirm its name, level column and relation. Names default to `component1`, `component2` and `component3`; the initial relation is Crossed and the first available level column is selected. A nested row also needs a parent column. Names must be unique and level columns distinct. Remove unwanted rows before fitting.

14.3.3 Setup and population

The dialog opens in random-slope mode with REML. Response, group and slope selectors return to their initial choices; confirm them. Fixed-effect lists and component lists start empty. Reopening clears the previous component rows and pending state.

Numeric roles require numeric finite values. Categorical roles use their genuine text labels, with sorted complete-case levels and the first as reference. Null, NaN and infinite numeric codes count as missing. Explicit exclusions are removed before complete-case encoding. No missing values are imputed.

The selected source is limited to one million rows and twenty million cells. Random-slope fitting permits at most 10,000 rows. Variance-components fitting permits at most 1,000 source rows and 128 total indicator columns. Fixed designs permit 128 encoded columns, at most 32 selected source columns and 64 levels in a categorical source. These limits are fixed in the desktop workflow.

14.3.4 Estimate and inspect the covariance

REML is the default; ML is available explicitly. The fit compares interior and explicit boundary covariance candidates, using dispersed starting values. A small positive variance is retained. A tied objective is not deliberately resolved in favour of a simpler model. The reported fit has the lowest attained deviance among the accepted attempts.

A *boundary* status requires both an exact boundary estimate and verified constrained-optimum conditions. A *near boundary* status can indicate small relative variance, nearly singular correlation, or an unresolved derivative sign; it is a diagnostic, not a variance-component test and it does not change the estimates.

If either random variance is zero, intercept/slope correlation is undefined and remains unavailable. It is not zero correlation. Different covariance matrices achieving the same likelihood can leave a model unidentified; that condition is refused. An increasing search edge, unresolved residual variance or failed convergence does not become an accepted boundary estimate. The numerical maximization is disclosed and is not a global-optimality certificate.

14.3.5 Uncertainty, units and predictions

Fixed-effect intervals are fixed 95% plug-in normal intervals. They do not use Kenward–Roger or Satterthwaite small-sample adjustments. AIC is reported for ML only. The workflow provides no

variance-component p-values, inverse-Hessian variance intervals or boundary likelihood-ratio tests.

Intercept and residual variances have squared response units. Slope variance has squared response units per squared predictor unit; intercept/slope covariance has squared response units per predictor unit. Always retain the predictor's origin and units with a covariance estimate.

Population predictions use fixed effects. Conditional predictions also use estimated group/component adjustments from the fitted Gaussian covariance, without artificial covariance jitter. Both fitted values and residuals retain source-row alignment.

Results report design rank and condition, population counts, group or component levels, covariance estimates, residual variance, log likelihood, applicable AIC and numerical status. **Add to results** retains those settings and warnings with bounded prediction summaries. An old saved result can identify an earlier optimizer; read it as the retained calculation rather than assuming it used the present method.

14.4 Binomial and Poisson mixed models

Choose **Statistics > Generalized & mixed models > Binomial and Poisson mixed models...** for a binary or count response with one random intercept per independent group. The conditional models are

$$\operatorname{logit}(p_{gi}) = x_{gi}^{\mathsf{T}}\beta + b_g, \qquad \text{or} \qquad \log \mu_{gi} = x_{gi}^{\mathsf{T}}\beta + b_g + \log E_{gi}, \quad b_g \sim N(0, \tau^2).$$

Fixed effects are additive and include an intercept. Group-specific effects and population-average effects differ because the inverse link is nonlinear.

14.4.1 Declare the response correctly

The default is Binomial with a logit link and one row per trial. Declare the event value explicitly; after mapping and exclusions, both event and non-event outcomes must be present. Alternatively, choose integer successes and a trials column, with positive integer trials and successes between zero and trials.

Poisson uses a log link and nonnegative integer counts. Optional exposure must be finite and strictly positive and enters as a log offset. Without exposure, each row has exposure one. The Group role is a label, including when it contains numeric IDs; it cannot also be a fixed factor. Numeric and categorical predictors start unselected; categorical levels use the first sorted reference. Confidence defaults to 95%.

Rows missing response, trials, exposure, group or predictors are omitted and counted by reason after exclusions. There is no imputation. Invalid counts or exposure are errors, not a request to repair values automatically. All-event/all-non-event binomial data and all-zero Poisson data cannot support a finite ordinary intercept estimate.

14.4.2 Check the design and numerical status

At least three groups and at least one group with repeated observations are required. Fewer than 30 groups triggers an advisory warning. A rank-deficient or badly conditioned fixed design prevents the fit. This workflow has no observation weights or additional offsets.

For binomial data, verified complete or quasi-complete separation in the fixed design is sufficient to refuse the unpenalized fit. Ignoring random intercepts in that particular certificate does not invalidate the refusal. If the check exceeds its 500,000-inequality limit or its candidate cannot be verified, the result is unavailable rather than described as non-separated.

The fit maximizes the marginal likelihood, integrating over the Gaussian random intercept. It uses adaptive Gauss–Hermite quadrature, starting with 15 nodes and comparing refits at 31 and, if needed, 61 nodes. The higher-node fit is retained when successive calculations agree in likelihood and parameters. The result records that accuracy check. A failed check can leave provisional estimates but provides no coefficient inference.

Several starting standard deviations are used. Agreement between two quadrature node counts alone is not proof of a unique optimum. Convergence, curvature and additional quadrature checks must also pass before Wald inference is reported. Read their status and warnings.

14.4.3 A zero variance is a separate statistical case

The ordinary generalized linear model at $\tau = 0$ is fitted with the same likelihood. The boundary is compared with a resolved interior search and with the one-sided variance score at zero. A nonpositive local score alone cannot exclude a better positive-variance optimum.

A verified zero-variance result is reported explicitly. Merely reaching the positive search's lower standard-deviation bound does not prove zero variance. Failed interior convergence, unresolved integration or an upper search bound produces an unresolved decision, with inference withheld. No reduced model is substituted silently, and ordinary boundary Wald inference is not supplied.

14.4.4 Interpret fixed effects and uncertainty

At a regular interior optimum, coefficient standard errors come from the inverse joint observed-information matrix, including uncertainty in the random-intercept parameter. Tests use normal z reference inference and selected-confidence Wald intervals. They are asymptotic, not small-sample corrections.

For binomial models, exponentiated coefficients are conditional odds ratios; for Poisson models they are conditional rate ratios. Exponentiate interval endpoints, not standard errors. The random-intercept standard deviation is an estimate without a test or interval in this workflow.

Inference is unavailable when covariance is non-positive-definite, curvature is unstable, integration accuracy is unverified, convergence fails or the variance decision is boundary/unresolved. An unavailable interval is not an interval of width zero.

14.4.5 Conditional and population-average predictions

A conditional prediction uses the fitted group's posterior-mode intercept. A population-average prediction integrates over the random-intercept distribution. For Poisson,

$$E(Y \mid x, E) = E \exp(x^{\mathsf{T}}\beta + \tau^2/2).$$

Thus a fixed-part mean of 2 and random-intercept variance 0.5 imply population mean $2e^{0.25} = 2.56805$, not 2. This is an illustrative calculation of the model's averaging rule.

For binomial responses, the average is $E_U[\operatorname{expit}(x^{\mathsf{T}}\beta + \tau U)]$, evaluated numerically. Setting the random intercept to zero does not generally give that average. A new group has no fitted group intercept, so only its population-average prediction is available. This workflow supplies no prediction intervals.

14.4.6 Retain and report the analysis

Add to results saves the summary, fixed effects, quadrature-accuracy table and the first 200 group intercepts, with complete settings for rerunning the study on its data. Report response coding, exposure units, group counts, numerical status and coefficient interpretation together.

Random slopes, multiple grouping factors, negative-binomial or ordinal mixed responses and automatic term selection are outside this workflow. Functional numerical checks are distinct from the selected independent comparisons described in the separate validation report; a successful fit is not by itself an independent validation claim.

14.5 Population-average GEE for clustered responses

Choose **Statistics > Generalized & mixed models > Population-average GEE (clustered)...** to model the average binary or count response while allowing observations within a cluster to be correlated. Clusters must be independent. Unlike a generalized mixed model, GEE estimates no cluster-specific random effects. Its coefficients describe a marginal mean.

14.5.1 Choose roles and working correlation

Select a response, cluster identifier and predictors. Binary/binomial with a logit link is the default; Poisson with a log link is the alternative. Binary input is 0/1, or integer successes with positive integer trials. Grouped binomial rows are not expanded into individual trials. Poisson counts must be nonnegative integers; optional strictly positive finite exposure enters as a log offset.

Numeric predictors enter continuously; text predictors use indicators against the first sorted category. Cluster IDs are labels even when numeric. At least three clusters are required. Confidence defaults to 95%, with range 50–99.99%.

Working correlation defaults to Exchangeable, giving the same working correlation to any pair of observations in one cluster. Independence is an explicit alternative; it does not remove the cluster-based sandwich covariance. No structure is selected automatically. If every cluster has only one observation, exchangeable correlation is unestimable and Independence must be chosen explicitly.

14.5.2 What GEE estimates

For cluster i, the mean is $\mu_i = m_i\pi_i$, with logit $\pi_i = X_i\beta$, or $\mu_i = E_i \exp(X_i\beta)$ for Poisson exposure. With diagonal working variances A_i, the covariance is

$$V_i = A_i^{1/2} R_i A_i^{1/2}, \qquad R_i = I \quad \text{or} \quad R_i = (1-\rho)I + \rho \mathbf{1}\mathbf{1}^{\mathsf{T}}.$$

Binomial diagonal variance is $m_i\pi_i(1-\pi_i)$; Poisson variance is μ_i, with working scale one. If $D_i = \partial\mu_i/\partial\beta^{\mathsf{T}}$, the estimating equation is

$$\sum_i D_i^{\mathsf{T}} V_i^{-1}(y_i - \mu_i) = 0.$$

The working covariance helps weight observations; the marginal mean model must still be scientifically appropriate.

Exchangeable ρ uses Pearson residuals $e_{ij} = (y_{ij} - \mu_{ij})/\sqrt{A_{ij}}$:

$$\widehat{\phi} = \frac{\sum e_{ij}^2}{N}, \qquad \hat{\rho} = \frac{\sum_i \sum_{j<k} e_{ij} e_{ik}}{\left[\sum_i n_i(n_i-1)/2\right]\widehat{\phi}}.$$

No design degrees of freedom are subtracted from these denominators. This is a declared moment convention, not an exact unbiased estimator; other software can differ slightly. Scale normalizes the correlation estimate while the working covariance retains scale one.

The correlation is constrained to the positive-definite range allowed by the largest cluster, and an active constraint is disclosed. An unresolved residual scale or correlation makes the fit unavailable.

14.5.3 Use the reported covariance convention

HypoNum uses the Mancl–DeRouen bias-reduced sandwich. Write $r_i = y_i - \mu_i$, $B = \sum_i D_i^\mathsf{T} V_i^{-1} D_i$, $H_i = D_i B^{-1} D_i^\mathsf{T} V_i^{-1}$ and $r_i^* = (I - H_i)^{-1} r_i$. Then

$$\widehat{\mathrm{Var}}(\hat{\beta}) = B^{-1} \left[\sum_i D_i^\mathsf{T} V_i^{-1} r_i^* r_i^{*\mathsf{T}} V_i^{-1} D_i \right] B^{-1}.$$

It is not the ordinary uncorrected sandwich. Singular information, singular leverage correction, non-finite covariance or materially indefinite covariance prevents covariance-based inference; the matrix is not repaired and another covariance is not substituted.

Coefficient tests and intervals use t_{G-p}, where G is the cluster count and p the fixed-design rank. This is an approximate finite-cluster convention. With $G \leq p$, estimates can be shown but inference is unavailable. Fewer than 30 clusters triggers an advisory, not a universal validity threshold. No multi-parameter tests are supplied.

For example, twelve clusters and a rank-three design leave nine reference degrees of freedom, even if those clusters contain hundreds of rows. More rows within the same clusters do not create more independent clusters. A positive, finite coefficient variance retains its inference however small; an exactly zero coefficient variance in an otherwise valid covariance leaves that coefficient's inference unavailable.

Exponentiate estimates and interval endpoints for marginal odds ratios or rate ratios. Do not exponentiate a standard error. Conditional odds ratios from a mixed model need not equal GEE marginal odds ratios even with the same named predictors.

14.5.4 Convergence, missingness and prediction

The iteration checks both parameter changes and the remaining estimating equations. A stalled or nonconverged finite iterate is labelled provisional, has no inference and cannot be used for prediction. The working correlation is not switched to obtain convergence.

Required missing values are omitted without imputation and counted. Ordinary complete-case GEE does not generally correct informative missingness; its mean estimating equations require an appropriate missingness condition given the included covariates. Do not treat dependence-aware standard errors as a missing-data adjustment.

Population-average predictions use the retained encoding and exposure. There are no cluster adjustments or prediction intervals. AR(1), unstructured correlation, survey weights, ordinal responses, model-based covariance, normal-reference inference and automatic correlation selection are outside this workflow.

14.6 Repeated measures with a residual covariance

This workflow models Gaussian responses observed at occasions within independent subjects:

$$y_i = X_i \beta + e_i, \qquad e_i \sim N(0, V_i).$$

The fixed effects describe the mean; V_i describes remaining within-subject variation. An occasion variable can serve both as a covariance-schedule identifier and, if selected, a numeric fixed effect for the mean. Those roles do different jobs.

Open **Statistics > Generalized & mixed models > Repeated measures (residual covariance)...**. Choose a numeric Response, Subject and Occasion, then fixed numeric or categorical effects. A column is categorical only when assigned that role; its first sorted level is the reference. An intercept is always included.

14.6.1 Check the observation schedule

Each subject may have at most one row per occasion. Duplicate subject/occasion keys are refused with their identities. At least three subjects and a full-rank fixed design are required. Excluded and incomplete rows are omitted and counted, without imputation.

Absent occasions remain absent. Each subject uses the submatrix of the occasion covariance corresponding to its actual observations. Likelihood interpretation assumes missingness is ignorable under the specified model, commonly expressed as missing at random given that model. An observed-data fit alone cannot establish this assumption. Source row order does not change the fit; the analysis orders by subject and occasion.

14.6.2 Choose a covariance structure

Choice	Occasion covariance
Independent, equal variance	Default: $\Sigma = \sigma^2 I$.
AR(1) on declared schedule	$\Sigma_{jk} = \sigma^2 \rho^{\lvert k_j - k_k \rvert}$, using integer schedule positions. Gaps retain their distance; ρ may be negative.
Independent, occasion-specific variances	$\Sigma = \mathrm{diag}(\sigma_j^2)$.
Unstructured	A positive-definite covariance with each occasion pair estimated separately; each pair must be co-observed in at least one subject.

Optional random intercept is off initially. It adds $\tau^2 \mathbf{1}\mathbf{1}^\mathsf{T}$ to the subject covariance and may accompany the first three choices. It cannot accompany Unstructured, because that decomposition is not identified.

14.6.3 Worked covariance calculation

Suppose schedule spacing is one, observed occasions are 0, 1 and 3, $\sigma^2 = 4$, $\rho = 0.5$, and there is no random intercept. The implied covariance is

$$\Sigma = \begin{pmatrix} 4 & 2 & 0.5 \\ 2 & 4 & 1 \\ 0.5 & 1 & 4 \end{pmatrix}.$$

The gap from 1 to 3 gives correlation $0.5^2 = 0.25$, not 0.5. A subject observed only at 0 and 3 uses the first and third rows/columns of this matrix. This is an illustrative covariance calculation, not a fit to a dataset.

AR(1) requires occasions on an equally spaced declared schedule: $k_j = (t_j - t_{\min})/s$ must resolve to an integer. Irregular times require an explicit mapping before using this structure. Numerically ambiguous schedule positions are refused rather than rounded into a different design.

14.6.4 Identification and boundaries

If all observed within-subject lags are even, ρ and $-\rho$ produce the same covariance and its sign is unidentified. The AR(1) fit is refused. Lags 2, 3 and 5 can identify the sign despite having no lag-one pair; failure of one zero-correlation starting value alone does not prove structural nonidentification. The fit tries alternative starts and reports when identification cannot be established numerically.

No subject with two occasions, an unobserved unstructured covariance pair, or a response fitted exactly by the fixed effects prevents the relevant fit. The covariance structure is never simplified silently.

14.6.5 Estimation

REML is the default; ML is explicit. For covariance parameters, fixed effects are estimated by generalized least squares:

$$C = \left(\sum_i X_i^\mathsf{T} V_i^{-1} X_i \right)^{-1}, \qquad \hat{\beta} = C \sum_i X_i^\mathsf{T} V_i^{-1} y_i.$$

The likelihood uses observed subject vectors, with their own covariance submatrices. With a random intercept, the zero-variance fit is compared with the interior; an equally good boundary fit is reported as zero, not hidden as a tiny positive variance. Results retain convergence, iterations and maximum gradient.

14.6.6 Satterthwaite inference

For a coefficient or contrast vector c, its plug-in variance is $v = c^\mathsf{T} C c$. If A approximates the covariance of the natural covariance parameters and $g = \partial v / \partial \theta$, the Satterthwaite reference degrees of freedom are

$$\nu = \frac{2v^2}{g^\mathsf{T} A g}.$$

Tests and confidence limits use t_ν. The approximation allows different coefficients to have different degrees of freedom. Under independent equal variance with REML it reduces to $N - p$. In a balanced random-intercept design, between-subject and within-subject coefficients draw information from different strata.

This is not an exact small-sample method or the Kenward–Roger adjustment used in some designed-experiment workflows. Fewer than ten subjects triggers a warning; exceeding ten does not guarantee accuracy.

The analysis withholds inference at a covariance boundary, after nonconvergence, with singular parameter information, or when ν is not finite and positive. Point estimates remain, explicitly provisional when required. This is the workflow's boundary policy, not a theorem that fixed-effect inference can never exist at a boundary. A simpler model may be requested explicitly; it is not substituted automatically. No joint multi-degree-of-freedom tests are offered.

14.6.7 Read covariance and residual diagnostics

Results separate fixed-effect inference from covariance estimates, approximate covariance-parameter standard errors and boundary flags. They show the model's occasion covariance/correlation, residuals by occasion with empirical and model variances, and within-subject whitened residuals against fitted marginal means.

Mean confidence intervals use the same contrast rule. Future-observation intervals and imputation of missing occasions are unavailable. AIC and BIC refer to the selected likelihood criterion. Under ML, the parameter count includes fixed and covariance parameters and BIC uses N. Under REML, it counts covariance parameters and uses $N - p$; compare only models with the identical fixed design. Every declared covariance parameter is counted, including a boundary estimate. Read the displayed parameter count and BIC sample size when comparing software.

Add to results retains the summary, fixed effects, covariance parameters, occasion covariance, contrasts and residual plot with settings. Reopening restores the saved result; rerunning is explicit. There is no automatic choice of covariance structure.

CHAPTER 15

Relationships and multivariate structure

Multivariate methods examine several measurements together. Start by deciding what should be summarized: association between variables, a small set of composite measurements, shared latent variation, or similarity between observations. These choices lead to different models.

Aim	Method and interpretation
Describe two-variable association	Correlation measures linear or monotonic association; inspect the scatterplot as well.
Compress total variation	Principal components form uncorrelated linear combinations of the selected measurements.
Describe shared variation	Factor analysis models common factors plus variable-specific variation.
Find similarity groups	Hierarchical clustering gives nested groups; *k*-means gives a chosen number of groups.
Map a categorical table	Correspondence analysis summarizes departures from independence.
Compare several responses	MANOVA tests a multivariate response effect under its model assumptions.

A mathematically clear pattern is not automatically a scientifically meaningful construct or a verified class. The measurement definitions and study design remain essential.

15.1 Correlation

Select at least two numeric columns in **Correlation...**. Add columns by double-clicking or selecting them and choosing **Add**. An optional **Primary** star restricts calculations to that variable against the others; click the star again to return to a full symmetric matrix.

Pearson correlation is

$$r = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x})^2 \sum_i (y_i - \bar{y})^2}}.$$

It measures linear association and is unchanged by positive changes of measurement units. Spearman correlation applies the same idea to average ranks, capturing monotonic association. Sample covariance instead divides the cross-product sum by $n - 1$; its units are the product of the two measurement units.

Each pair uses rows finite in both variables. Different matrix cells can therefore have different populations when missingness differs. A constant variable or fewer than two paired observations gives an undefined coefficient. Pairwise deletion can also produce a matrix that is unsuitable as a covariance model for a later analysis.

15.1.1 Worked example: association is not agreement

Choose **File > Import data...** and select `sample-data/manual/paired-measurements.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Correlation & regression > Correlation...**. Double-click **x**, then **y**, in **Available columns**; select no other columns. Set **Metric** to **Pearson (r)**, leave scatter plots enabled, and click **Run correlation**. Inspect the x/y matrix cell. To examine the difference using the same table, choose **Statistics > Compare means & ANOVA > Hypothesis tests...**. On **Hypothesis test**, set **Test family** to **Means, variances and proportions**, **Test** to **Paired t-test**, **Response** to **x** and **Paired response** to **y** (the test uses a zero null difference), **Alternative** to **Two-sided**, and **Alpha** to 0.05. Click **Run test** and inspect the estimated first-minus-second difference. The file has 16 x, y pairs. Its Pearson correlation is 0.9964535, yet $x - y$ averages 0.5625. The two series track one another closely while differing systematically. If the columns represented two measurement methods, high correlation alone would not establish interchangeability. Examine differences and the measurement-error model as well.

Scatter panels are on by default, limited to ten selected columns and 400 deterministic display points per panel. The coefficients use the complete pair populations. The heatmap uses those same coefficients: Pearson and Spearman scales run from -1 to 1, while covariance uses the largest absolute value to define a symmetric scale. Undefined cells remain neutral. Neither the colour nor the coefficient proves causation.

15.2 Principal component analysis

Choose **Principal component analysis...** and two to 64 numeric measurement columns. PCA uses rows finite in every selected column after exclusions; it does not fill missing values. The fit centres each column.

With the default correlation scaling, it also divides each column by its complete-case sample standard deviation:

$$z_{ij} = (x_{ij} - \bar{x}_j)/s_j, \qquad S = Z^{\mathsf{T}}Z/(n-1).$$

With covariance scaling, Z denotes centred values in their original units. Correlation scaling gives differently measured variables comparable variance; covariance scaling gives more influence to variables with larger numerical variation. That is an analytical choice, not a cosmetic one. A constant column cannot be standardized.

For eigenvector v_k and eigenvalue λ_k,

$$Sv_k = \lambda_k v_k, \quad v_k^{\mathsf{T}} v_k = 1, \qquad t_k = Zv_k, \quad \text{variance fraction}_k = \frac{\lambda_k}{\sum_j \lambda_j}.$$

HypoNum reports unit-eigenvector loadings. Some references multiply these by $\sqrt{\lambda_k}$ and use the same word *loading*; check that convention before comparing tables.

Choose a component count or retain all available components, at most $\min(n-1, p)$. The results give eigenvalues, individual and cumulative variance, loadings, means/scales and the first two score axes. The calculation uses a full singular-value decomposition on all complete rows; only the score display is limited to 5,000 rows. Limits are one million source rows and twenty million selected numeric cells.

A component sign is chosen so that its largest absolute loading is positive. Changing all signs on an axis would describe the same component. Nearly equal eigenvalues permit substantially

different axes within the same subspace; avoid attaching a unique interpretation to an unstable individual axis.

15.2.1 Worked example: three measurements

Choose **File > Import data...** and select `sample-data/manual/multivariate-measurements.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Multivariate > Principal component analysis...**. Double-click **x**, **y** and **z** in **Available columns** to add them to the selected list. Set **Scaling** to **Correlation (standardized)** and **Components** to 2. Click **Run principal components**; inspect the eigenvalues, loadings and row-aligned scores. The twelve complete input rows are:

x	y	z	x	y	z
1.0	2.0	3.1	5.0	7.2	4.8
1.2	2.1	2.8	4.8	6.9	5.1
0.8	1.8	3.0	9.1	3.0	8.9
1.1	2.2	3.3	8.8	3.2	9.1
4.9	7.1	5.2	9.2	2.9	8.8
5.2	6.8	5.0	9.0	3.1	9.2

Read down the left group and then the right group. The file also contains other variables; only these three enter this calculation. Means are $(5.008333, 4.025000, 5.691667)$ and sample standard deviations $(3.414796, 2.245045, 2.590001)$. Eigenvalues are 1.997459, 0.997908, 0.004633, accounting for 66.58197%, 33.26361%, 0.15442% of standardized variation. The first two unit-eigenvector loading columns are

$$\begin{pmatrix} 0.706726 & -0.005286 \\ 0.136588 & 0.982179 \\ 0.694178 & -0.187874 \end{pmatrix}.$$

The first row's scores are approximately $(-1.647389, -0.691712)$. The first axis mainly combines x and z; the second mainly represents y. Two components retain about 99.8456% of the standardized variation, but this small constructed sample establishes no population model.

15.2.2 Use scores without changing their definition

Create score columns projects the retained fit population using the stored means, scales and basis. Excluded or incomplete rows receive missing scores, and original row alignment is retained through sorting. The action supports undo, redo and replay without refitting. **Add to results** retains the analysis tables and plots. After changing the source, run a new analysis deliberately.

Choose the number of components using the purpose, scree, cumulative variation and interpretability together. A high percentage of explained variation does not prove that the discarded dimensions are irrelevant to a later outcome: PCA does not use that outcome.

15.3 Exploratory factor analysis

Factor analysis asks whether several standardized measurements share a smaller number of common sources of variation. For an orthogonal factor model,

$$R \approx LL^{\mathsf{T}} + \Psi,$$

where L contains loadings and diagonal Ψ contains unique variances. Communality is the sum of squared loadings for a measurement; on this standardized scale it accounts for the variance assigned to the common factors. This differs from PCA's decomposition of total variation.

Choose three to 32 numeric columns and an explicit factor count from 1 to $\min(p - 1, 16)$. HypoNum fits maximum likelihood to the complete-case sample correlation matrix, with more complete rows than columns. Constant columns and singular or non-positive-definite correlation matrices are rejected. No covariance-scale option or missing-value imputation is offered. The input limits are one million rows and twenty million selected cells.

The default displays the unrotated fitted frame. Optional **Varimax** applies an orthogonal rotation: it can make loading patterns easier to describe while preserving communalities, uniquenesses and reproduced common covariance. Both frames are retained. Factors are ordered by the sum of squared loadings, with the largest absolute loading positive; ties follow column order.

15.3.1 Read the fit status before the loadings

A converged admissible fit, a nonconverged fit and a verified Heywood boundary are different outcomes. A zero or negative unique variance cannot support ordinary interior-model interpretation. Boundary and nonconverged fits provide no scores. A very small *positive* uniqueness can still be admissible and receives a near-boundary warning. A suspected boundary that cannot be verified remains unresolved rather than being declared successful.

A model with zero degrees of freedom need not have an admissible solution: the exact correlation solution may require an impossible negative unique variance. Reducing the factor count, reconsidering nearly redundant measurements, or obtaining better data may be more meaningful than seeking an optimizer setting that suppresses a warning.

Results include loadings, communalities, uniquenesses, contributed common variance, standardization and a correlation-eigenvalue scree for context. Those correlation eigenvalues are not a list of independently fitted factors. Tied common-variance contributions can make factor orientation ambiguous. The workflow does not fit a confirmatory model or provide intervals for loadings or scores.

15.3.2 Understand a factor score

For standardized row vector z, the regression (Thurstone) score is

$$\hat{f} = zW, \qquad W = (LL^{\mathsf{T}} + \Psi)^{-1}L.$$

It is a model-based summary, not an observed latent truth. With varimax, the weights correspond to the displayed rotated loadings. **Create score columns** uses retained weights without refitting; omitted rows stay missing. Two or more cleanly converged factors permit a first-two-factor scatter display of at most 5,000 source-linked rows. The fit uses all complete rows. Retain results and warnings together.

For example, hypothetical one-factor loadings $0.8, 0.7, 0.6$ imply communalities $0.64, 0.49, 0.36$, uniquenesses $0.36, 0.51, 0.64$ and reproduced off-diagonal correlations $0.56, 0.48, 0.42$. These are model-arithmetic examples, not a fit to the PCA dataset above. They show why a loading of 0.8 means 64% common variance in this one-factor orthogonal model, rather than 80%.

15.4 Hierarchical clustering

Hierarchical clustering builds nested groups. Choose two to 64 numeric columns and decide whether to cluster **Rows** or **Variables**. Both modes use one complete finite population after exclusions.

Row mode uses Euclidean distances. **Standardize columns**, on by default, divides by complete-case sample standard deviations so that measurement units do not dictate distance. Ward linkage is the default; average and complete linkage are also offered. Ward merges minimize increases in within-cluster squared variation. Reported merge height is the square root of twice that increase.

Variable mode uses $d = 1 - r$, with the distance clipped to $[0, 2]$ for rounding. A correlation of 1 gives distance zero and a correlation of -1 gives distance two: reversing a variable's direction matters. Average linkage is the default; complete linkage is available. Ward is unavailable for this distance convention, which does not supply the required Euclidean geometry. Standardization does not apply because correlation distance is scale-invariant.

Average linkage uses the average pair distance across the two clusters; complete linkage uses the largest. These can yield different groups. Select the rule because its notion of similarity fits the question.

15.4.1 Read and cut a dendrogram

Leaves identify observations or variables; joins occur at their linkage heights. **Cut k** optionally requests k groups. HypoNum does not estimate an optimal k. Cluster numbers follow first-leaf order, and the display marks the cut with a dashed line. Clicking a row leaf selects its source row.

The merge schedule, cluster sizes and omission counts accompany the tree. Large tables may be shortened for display, while every merge is drawn. For deterministic ties, complete rows are ordered by stable identity before distance calculation. The default is at most 2,000 clustered rows; the hard ceiling is 5,000 rows or 12.5 million pairwise cells, about 100 MiB of distances. Oversized requests fail before allocation and are not sampled. Variable mode clusters at most 64 columns.

Create label column is available for a row-mode cut. It reuses that cut, assigning missing labels outside the fit population. Save the linkage, scaling, cut and warnings with the result; cluster membership alone omits the decisions that produced it.

15.5 K-means clustering

K-means instead seeks k centroids to reduce

$$\sum_{g=1}^{k} \sum_{i \in C_g} \|z_i - \bar{z}_g\|^2.$$

The criterion favours compact Euclidean groups. It can be unsuitable for elongated or markedly unequal-density groups, and it does not establish that a population contains k natural classes.

Select two to 64 numeric columns and an explicit **Clusters (k)**. All complete finite rows after exclusions enter the fit. Standardization is on by default; a constant column is rejected in that mode. Without standardization, constant columns contribute no separation. A requested k exceeding the number of complete or distinct complete rows is refused, and an empty-cluster solution is not silently relabelled as a smaller k.

The fit uses Lloyd iterations with k-means++ initialization. **Seed** defaults to 0 and **Restarts** to 10, with at most 100 restarts. Retain both settings when comparing runs. A local solution can depend on initialization; reproducibility does not prove global optimality.

15.5.1 Worked example: a transparent partition

Choose **File > Import data...** and select `sample-data/manual/cluster-points.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Multivariate > K-means clustering...**. Double-click **x** and **y** in **Available columns** to select both. Set **Clusters (k)** to 2, **Seed** to 0 and **Restarts** to 10; clear **Standardize columns**. Click **Run k-means**. Inspect the cluster centroids, sizes and inertia. The four points are

$$(0,0), \quad (0,1), \quad (10,10), \quad (10,11).$$

Select both columns, $k = 2$, seed 0, 10 restarts and turn standardization off. The retained centroids are $(0, 0.5)$ and $(10, 10.5)$, with two observations in each cluster. Each point is 0.5 from its centroid, so inertia is $4(0.5^2) = 1$. The first two rows receive label 1 and the last two label 2. This constructed example explains the criterion; it does not demonstrate that a real population contains two groups.

Read cluster sizes, original-unit centroids, inertia and iteration count. Centroids are ordered lexicographically to number clusters, with row identity breaking remaining ties. The silhouette summary, when defined, uses a disclosed deterministic sample of at most 2,000 complete rows; the fit itself is unsampled. For observation i, the silhouette compares its mean within-cluster distance a_i with its smallest mean distance b_i to another cluster:

$$s_i = (b_i - a_i) / \max(a_i, b_i).$$

A large value suggests separation under this distance, not an independently validated class.

The first-two-column scatter displays at most 5,000 rows and supports linked selection. **Create label column** uses retained membership without refitting; omitted rows stay missing. Label creation supports undo, redo and replay. Source changes require a new analysis.

15.6 Simple correspondence analysis

Correspondence analysis (CA) turns a two-way count table into a low-dimensional map of row and column profiles. It is useful for seeing which categories depart from the average pattern. It is a descriptive decomposition; this workflow reports no chi-square p-value and performs no multiple correspondence analysis.

Choose **Statistics > Multivariate > Correspondence analysis...**. Use **Two categorical columns**, the default, for record data, or **Numeric count table** for a table of counts.

15.6.1 Prepare counts and categories

In categorical mode, choose different row and column category columns. An optional frequency column supplies exact nonnegative integer counts; otherwise each eligible record counts once.

In count-table mode, choose a row-label column and at least two count columns. Row labels must be unique. Enter count-column names as a comma-separated list; its order determines displayed column order. Roles start unselected and frequency defaults to none. Missing categories or row labels are omitted and counted, while exclusions are counted first.

Counts must be finite nonnegative integers, with total at most $2^{53} - 1$. Input limits are one million source rows, twenty million selected cells, 200 categories per axis and 40,000 dense cells. At least two positive-mass categories on each axis and a positive total are needed. Zero-margin categories remain in the tables but have unavailable coordinates, contributions and squared cosines.

15.6.2 Worked example: a two-by-two table

Choose **File > Import data...** and select `sample-data/manual/category-profiles.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Multivariate > Correspondence analysis...**. Set **Input mode** to **Numeric count table**, **Row-label column** to **Row**, **Selected count columns** to `Yes, No`, and **Dimensions** to **2**. Click **Run correspondence analysis**. Check the reported dimension count, inertia and row/column coordinates against the calculation below. The complete count table is:

Row	Yes	No
A	30	10
B	10	30

Choose Numeric count table, Row as the label, Yes and No as the count columns, and request two dimensions. Each row total is 40 and each column total is 40, so independence would give 20 in every cell. The Pearson statistic is

$$\chi^2 = 4(10^2/20) = 20, \qquad \text{total inertia} = 20/80 = 0.25.$$

Only one positive-inertia dimension exists. HypoNum reports that the two-dimension request has been reduced to one. Its singular value is 0.5, and the first-axis coordinates are +0.5 for A and Yes, and −0.5 for B and No. The second axis is labelled unavailable.

A contains 75% Yes whereas B contains 25% Yes. The map summarizes that profile contrast. Coincidence between A and Yes on this particular one-dimensional map does not create a general distance measure between a row category and a column category.

15.6.3 Calculation and map interpretation

Let N be the count matrix, n its total, $P = N/n$, and r, c the row and column masses. The analysis decomposes

$$D_r^{-1/2}(P - rc^{\mathsf{T}})D_c^{-1/2} = U\,\mathrm{diag}(s)V^{\mathsf{T}}.$$

Principal coordinates are $F = D_r^{-1/2}U\,\mathrm{diag}(s)$ for rows and $G = D_c^{-1/2}V\,\mathrm{diag}(s)$ for columns. Axis inertia is s_k^2; total inertia is χ^2/n. Row contribution to axis k is $r_i F_{ik}^2/s_k^2$, with the analogous column expression. Squared cosine compares displayed coordinate variation with the category's full chi-square distance.

Read contributions to identify which categories define an axis and squared cosines to judge how well the displayed axes represent a category. A category close to the origin can have little distinctive profile, even if its frequency is large.

Within the row set, and separately within the column set, retained-axis distances approximate chi-square profile distances. Distances across the two sets do not directly measure association. Exact independence gives zero inertia and no informative axes, a valid descriptive result.

15.6.4 Display and retain

Dimensions default to two, with a range of one to ten, truncated to the positive-inertia rank with a warning. Signs are deterministic; tied axes can still rotate within their common subspace. Categories follow first source appearance, or selected count-column order. The map has at most 400 category marks and the complete category tables remain available.

The calculation can be cancelled. Closing setup or results, Escape or Cancel stops active work; an obsolete calculation cannot replace a newer request. **Add to results** retains the inertia table, map, category tables, settings, population counts and source context.

15.7 Multivariate analysis of variance

MANOVA asks whether a fixed-effect design changes several continuous responses jointly. It can be useful when responses describe complementary aspects of the same observational unit. It does not remove the need to understand which response patterns matter.

Choose **Statistics > Multivariate > Multivariate analysis of variance...**, with two to eight numeric responses, categorical factors and numeric covariates. Every selected response and predictor uses the same complete-case population after exclusions. A row missing any required value is omitted from the joint analysis, not used for only part of the response vector.

15.7.1 Specify the design

The default is Type III with sum coding. Type II uses the same marginality convention as the general linear model in section 12.9. Selected predictors receive main effects. Add interactions between distinct selected predictors, quadratic terms for a covariate, or nested terms with parent factor first and child second. Identical duplicate terms are ignored; Remove and Clear terms edit the list.

MANOVA here is unweighted ordinary multivariate least squares. Precision and survey weights are unavailable. General-linear limits apply: one million rows, twenty million selected/design cells, twelve factors, 32 covariates, 32 levels per factor and 128 terms. Setup resets roles and additional terms on opening. Intervals are fixed at 95%.

15.7.2 Read the joint tests

Let E be the residual sum-of-squares-and-cross-products matrix and H the corresponding hypothesis matrix for a term. If λ_j are the nonzero eigenvalues of $E^{-1}H$, the two reported statistics are

$$\text{Pillai trace} = \sum_j \frac{\lambda_j}{1+\lambda_j}, \qquad \text{Wilks lambda} = \prod_j \frac{1}{1+\lambda_j}.$$

Larger Pillai and smaller Wilks values indicate stronger departures from the tested joint null, but their scales are not p-values. Read each accompanying F approximation, numerator and denominator degrees of freedom, and p-value.

For illustration, eigenvalues 0.5 and 0.2 give Pillai = 0.5 and Wilks = $1/1.8 = 0.555556$. Significance cannot be inferred from those values alone: the reference approximation also uses model and sample dimensions. These illustrative eigenvalues are not a fit to the small dataset elsewhere in this manual.

Rank-deficient designs, insufficient residual degrees of freedom, singular/non-finite residual covariance or invalid approximation degrees of freedom prevent the analysis. No covariance jitter or alternative test is silently substituted. Residual covariance conditioning is checked

after conversion to a correlation matrix, with smallest eigenvalue greater than 10^{-10}, so merely changing response units does not evade the check. An exactly null term has Pillai 0, Wilks 1 and $p = 1$ at the appropriate positive approximation degrees of freedom.

15.7.3 Interpret follow-ups

Choose **File > Import data...** and select `sample-data/manual/paired-measurements.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Multivariate > Multivariate analysis of variance...**. Add **x** and **y** as responses and **g** as the only categorical factor; leave covariates empty. Use Type III sums of squares and click **Run MANOVA**. All 16 rows are complete. Inspect the multivariate term table for **g** and the response follow-up table. Pillai is 0.1855259 and Wilks is 0.8144741; both give $F(2, 13) = 1.4806095$, $p = 0.2634526$. The response-specific ordinary p-values are 0.1032096 for x and 0.1112060 for y, both becoming 0.2064192 in the two-test Holm family. The joint test and the separate response tests ask different questions; neither should be selected after looking for the smallest value.

The residual covariance is E divided by residual degrees of freedom. Its diagonal entries have squared response units; off-diagonal entries have the product of the two response units.

Response-specific ordinary F follow-ups use the same joint population. Their p-values form one Holm-adjusted family over every response-by-term test. They are secondary to the multivariate assessment. Adjusted means and their displayed 95% intervals are individual, unadjusted quantities, not simultaneous confidence regions. Choose a response in the margins plot to view its estimates and exact interval endpoints.

Reordering responses changes their output order, not the joint tests. Retain complete-case counts, model terms, test convention and warnings. Saved results reopen without refitting and become stale when their source changes.

15.7.4 Adjusted means and declared comparison families

For nested factors, the parent adjusted mean averages equally over the legal child levels within that parent. A child belonging to another parent is not inserted as an artificial combination. Other nuisance factors receive equal reference-grid weights and numeric covariates use their recorded reference values. Changing the follow-up choice does not change this mean or its standard error. Earlier saved analyses retain their recorded grid convention.

Choose all-pairs or comparisons to a named control to obtain max-t adjusted comparisons. Each response and factor defines a separate family, using the fitted model's covariance and residual degrees of freedom; no second model is fitted. These follow-ups are available without first requiring a significant joint MANOVA test. The ordinary displayed mean intervals remain individual 95% intervals. Comparison intervals have their stated family coverage.

Without across-family allocation, simultaneous protection applies within each declared family only. Selecting allocation divides 0.05 by the number of declared response/factor families before calibrating each family. Inspect the family definition, control level, numerical integration precision and any refusal. A numerical probability error bound measures the integration calculation, not sampling uncertainty or model validity.

15.8 Further reading

NIST/SEMATECH, sections 6.5.5 and 6.5.5.1, explains PCA eigenvalues and scaling. The R statistical references describe maximum-likelihood factor analysis and score conventions, and linkage definitions and Ward conventions. Reference-software defaults may differ from HypoNum's stated choices.

CHAPTER 16

Building and evaluating predictive models

A predictive model should be judged on observations that did not help choose or fit it. Training error answers how closely a model follows known data; test error estimates how it may perform on comparable new observations. Neither is a confidence interval for a regression coefficient.

16.1 A training, tuning and test workflow

Open **Statistics > Correlation & regression > Validated regression...**. Select one numeric response, numeric and/or categorical predictors, OLS, Ridge or Lasso, a missing-predictor policy, a test fraction and a seed. The test fraction starts at 25% and ranges from 5% to 50%. Keep the seed in the study record.

The random split is intended for observations that can be treated as independent. Repeated measurements from the same person should not be randomly dispersed between training and test sets to claim performance on new people. Time-ordered forecasting needs evaluation in time order. The split offered here does not implement those alternative designs.

16.1.1 Keep preprocessing inside training

Rows require a finite observed response and nonmissing categorical predictors. **Drop rows** also requires finite numeric predictors. **Training-mean** and **Training-KNN(5)** fill missing numeric predictors using a transformation learned on training rows. The response is never imputed. Learned transformations are then applied to test rows.

Categorical reference levels come from the first alphabetical training level. A category present only in validation or test rows stops the fit with an explanation; it is not silently assigned a known effect. Ridge and Lasso also learn predictor scaling from training data.

For penalized models, each training fold refits the entire preprocessing pipeline. The candidate penalty ladder contains 40 geometric points spanning three decades and is derived from the standardized training data. Mean fold RMSE chooses the penalty; ties favour the larger penalty. The selected pipeline is refitted to all training rows and applied once to the held-out test rows.

Important

After inspecting a test result, repeatedly changing predictors or settings to improve that result turns the test set into part of model selection. Use a genuinely new evaluation sample to assess the final choice. The software's split cannot prevent choices made outside the calculation.

16.1.2 Worked interpretation of prediction errors

Suppose four untouched test responses are $10, 12, 14, 16$, and a retained model predicts $11, 11, 13, 17$. This small arithmetic example explains the reported measures; it is not a fitted

HypoNum study. Residuals $y - \hat{y}$ are $-1, 1, 1, -1$, so

$$\text{RMSE} = \sqrt{\frac{1}{4}\sum_i (y_i - \hat{y}_i)^2} = 1, \qquad \text{MAE} = \frac{1}{4}\sum_i |y_i - \hat{y}_i| = 1.$$

Both have response units. The test mean is 13 and its centred sum of squares is 20, giving

$$R^2_{\text{test}} = 1 - \frac{4}{20} = 0.8.$$

A test R^2 can be negative: this means its squared prediction error exceeds that of predicting the test sample's mean. That comparison is descriptive; the actual deployable constant baseline must use a mean learned from training observations.

RMSE penalizes a few large errors more strongly than MAE. A single split can have substantial sampling variation, especially in a small dataset. A reported decimal value does not supply its own uncertainty interval.

16.1.3 Read the separate result tables

Results distinguish training metrics, the tuning criterion and full penalty ladder, and untouched-test RMSE, MAE and R^2. The coefficient table is expressed in original units; retained settings include training fill values, category mappings and scaling. The actual-versus-predicted view displays a bounded set of test observations.

The name **Validated regression** denotes this predictive split-and-tune workflow. It does not itself establish independent numerical qualification or scientific validity. Classical coefficient standard errors, p-values and confidence intervals are not supplied. For inferential OLS, use the regression chapter and its design assumptions.

16.2 Regression trees

A regression tree divides predictor space into regions and predicts a constant response within each leaf. It can express thresholds and interactions without choosing a polynomial equation, but its stepwise prediction can change abruptly and its fitted splits can be unstable in small samples.

Open **Statistics > Multivariate > Regression tree...**. Select one numeric response and one to 64 distinct numeric predictors. Rows missing any selected value, containing non-finite values, or excluded from the active population are omitted; no imputation is performed.

Setting	Initial value	Allowed values
Test fraction	25%	5–50%
Training cross-validation folds	5	2–10
Maximum depth	8	1–8
Minimum leaf size	1	1–100,000
Seed	0	0–2,147,483,647

The depth search runs from 1 through the selected maximum. Leaf-size candidates start at the selected minimum and double, with at most eight candidates and none above 100,000. The

default candidate sizes are 1, 2, 4, 8, 16, 32, 64 and 128. Training-only cross-validation selects the smallest mean RMSE; exact ties prefer shallower depth, then larger leaves. Enough complete training rows must remain for the requested folds.

16.2.1 An example of what a leaf predicts

Consider an illustrative tree that splits at temperature 50. Its lower-temperature training responses are 8, 10 and 12, so that leaf predicts their mean, 10. Its higher-temperature responses are 18, 20 and 22, so the other leaf predicts 20. A new row is sent through the same split and receives the corresponding constant. This arithmetic explains the prediction rule; the fitted workflow chooses its actual splits from training data.

A deeper tree could separate more of these observations and reduce training error. That alone is not a reason to prefer it. Cross-validation and an untouched test sample are intended to reveal the cost of such extra detail.

16.2.2 Evaluate and retain the tree

The result shows the selected settings, full tuning grid, split nodes, leaf predictions, feature importances, row-use counts and warnings. It reports training and test RMSE and MAE, with a training-mean predictor evaluated on the same rows as the corresponding comparison. Compare test performance with that simple baseline before interpreting a complex tree as useful.

Feature importance is a measure of the fitted tree's use of a predictor, not a causal effect or a coefficient in response units. Correlated predictors may divide or substitute for one another's importance. A constant training response can produce a one-leaf tree and is disclosed.

Add to results retains the result and settings. **Create score column** separately requires an explicit output name and supports undo, redo and saved-project replay. Rows outside the eligible population or without usable predictors receive missing scores. A source change clears the live fitted result.

The fit accepts up to one million source rows and twenty million selected predictor cells. Predictors use single-precision numerical values during tree fitting; values outside the finite range are refused. As with other random-split prediction workflows, dependent groups and ordered future forecasting require a different evaluation design.

16.3 Classification trees

A classification tree predicts a category and a probability vector. The largest probability determines the predicted class. A high accuracy can coexist with poor probability estimates, particularly when one class is common, so HypoNum reports several complementary measures.

Choose **Statistics > Multivariate > Classification tree...**. Select one response and numeric and/or categorical predictors. The response can have two to twenty classes. Displayed class labels are treated as text and sorted lexicographically; that order is retained in probability vectors and result tables. For a binary response, explicitly choose the positive class. Multiclass analysis does not silently select a class for binary ROC measures.

16.3.1 Settings and eligible rows

Opening setup clears variable selections and restores defaults. The initial test fraction is 25%, five cross-validation folds, maximum depth 8, minimum leaf size 1 and seed 0. The allowed ranges match the regression tree. Class weights start at **None**; **Balanced** estimates weights from the rows

used to fit each fold and the final training fit. Custom weights and probability calibration are unavailable.

Missing responses and incomplete predictors are counted separately. Numeric predictors must be finite and representable in single precision. Categorical predictors are labels, not numbers with an implied distance. The workflow does not impute missing values.

Limits are one million rows, 64 source predictors, 20 classes, 32 learned levels per categorical predictor, 128 encoded features, twenty million source predictor cells and 511 tree nodes. Each class must have enough training rows for the requested folds; the software does not reduce the fold count silently.

16.3.2 Fit, select and test

The reproducible train/test split and cross-validation folds are stratified by class. Both portions must contain every class. A tree uses Gini impurity to choose splits. Candidate depth and leaf-size settings are selected by smallest mean validation log loss; ties favour shallower trees, then larger leaves.

Categorical encodings are learned independently within each training fold. The final encoding and class weights are learned on all training rows, then used for the final fit and untouched-test evaluation. A previously unseen categorical level is encoded with all its indicator values zero. Results warn about such values and count affected categorical cells; one row may contribute more than one cell.

16.3.3 Read a confusion matrix

Rows of the confusion matrix are actual classes and columns are predicted classes. For example, this illustrative binary test result contains 40 rows:

	Predicted negative	Predicted positive
Actual negative	18	2
Actual positive	5	15

Accuracy is $(18 + 15)/40 = 0.825$. Positive-class precision is $15/(15 + 2) = 0.8824$, whereas recall is $15/(15 + 5) = 0.75$. Precision asks how often a positive prediction is right; recall asks how many actual positives are found. An undefined denominator is shown as unavailable, not an invented zero.

16.3.4 Evaluate probabilities as well as labels

For the probability assigned to the actual class y_i, log loss is

$$-\frac{1}{n}\sum_i \log p_{i,y_i}.$$

HypoNum clips probabilities below double-precision machine epsilon for this calculation. Confidently assigning a tiny probability to the correct class has a large cost. Selection by log loss therefore differs from selection by accuracy.

The multiclass Brier score is

$$\frac{1}{n}\sum_i \sum_k \left(p_{ik} - \mathbf{1}[y_i = k]\right)^2.$$

For a binary row whose actual class is positive and predicted positive probability is 0.8, its contribution is $(0.8-1)^2+(0.2-0)^2=0.08$. This sums both class terms; a convention using only the positive-class term would instead give 0.04. State the convention when comparing software.

Binary ROC and precision–recall curves use the declared positive class and retain at most 1,000 points each. Results report ROC AUC and average precision. Multiclass results do not show binary curves. Probability ties for the predicted label follow the retained class order.

16.3.5 Save results and create class scores

The result contains population counts, selected settings, the candidate table, training and test metrics, confusion and per-class tables, warnings and the fitted tree. **Add to results** retains the snapshot.

Create scores applies the retained model to the selected target table. Enter an output name; the displayed `predicted_class` suggestion is only a placeholder. The target needs the retained predictor columns. Missing or unusable predictors give missing labels; unseen categories use the declared all-zero encoding. A one-row target is permitted.

Scoring uses the fitted tree without refitting. It preserves the target's row alignment and supports undo, redo and project replay. Existing columns cannot be overwritten with conflicting scores. Check that the target represents the population and measurement procedures for which the model was developed; a successful scoring calculation does not establish that transfer is scientifically appropriate.

CHAPTER 17

Survival curves and event-time regression

An event-time study records how long each subject is observed and whether the event occurs. A subject observed for 8 months without the event supplies information through month 8; treating that duration as an event or deleting the subject would answer a different question.

17.1 Kaplan–Meier survival curves

Open **Statistics > Survival > Survival curves (Kaplan-Meier)...**. Choose positive numeric durations and a status column, then identify the displayed labels for **Event value** and **Censored value**. Conventional 1/0 and True/False codes can be preselected, but verify them. A different or misspelled status is refused rather than assumed censored.

An optional group column produces up to ten separate curves. Numeric group levels are ordered numerically and other levels alphabetically. Confidence defaults to 95%. Rows missing time, status or an assigned group are omitted and counted; zero, negative or infinite durations are invalid.

The usual interpretation assumes independent subjects and non-informative right censoring: conditional on the study design or relevant grouping, censoring should not preferentially remove subjects with different unobserved event risks. A subject leaving because of worsening health, for example, may violate that assumption.

17.1.1 Worked example: events and censoring at the same time

Choose **File > Import data...** and select `sample-data/manual/survival-six.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Survival > Survival curves (Kaplan-Meier)...**. Set **Time** to **time**, **Status** to **status**, **Event value** to 1 and **Censored value** to 0; leave **Group (optional)** unset and retain 95% confidence. Click **Run analysis**. Inspect the survival curve and the time-by-time risk, event and censor counts. The six observations are

$$(5,1),\ (8,1),\ (8,0),\ (12,1),\ (15,0),\ (20,1).$$

At time t_j, n_j subjects remain at risk immediately before the time and d_j have events. The survival estimate is

$$\hat{S}(t) = \prod_{t_j \le t} \left(1 - \frac{d_j}{n_j}\right).$$

Time	At risk	Events	Censored	Survival
5	6	1	0	0.833333
8	5	1	1	0.666667
12	3	1	0	0.444444
15	2	0	1	0.444444
20	1	1	0	0

Both rows at time 8 are at risk for that time's event. The censor at 15 causes no downward step. The estimated median is 12, the first time survival reaches or falls below 0.5. Its lower 95% confidence limit is 5; its upper limit is unavailable under this short follow-up and endpoint convention. Six observations are useful for understanding the arithmetic, not for a precise event-time estimate.

17.1.2 Pointwise uncertainty and the median

Greenwood's variance uses the event risk sets. Away from the endpoints,

$$\widehat{\mathrm{Var}}\{\hat{S}(t)\} = \hat{S}(t)^2 \sum_{t_j \le t} \frac{d_j}{n_j(n_j - d_j)}.$$

Confidence limits use a log-log transformation to stay on a probability scale. They are pointwise intervals, not a simultaneous band. Before the first event and when the estimate reaches zero, unavailable standard errors and limits are shown as blank. A group without events has a descriptive curve at 1 with inference unavailable.

If the curve never reaches 0.5, the median is *not reached*. It is not extrapolated beyond follow-up. Median limits are found by inverting the lower and upper pointwise curves at 0.5. A limit can be not reached, or not estimable at a zero-survival endpoint. If survival is exactly 0.5 over an interval, HypoNum uses its first time; some software uses an interval midpoint.

17.1.3 Comparing two curves

With exactly two analyzable groups, the log-rank test compares observed events with expected events under equal event hazards over follow-up, using the shared risk sets. Its large-sample statistic has a chi-square reference with one degree of freedom. The result includes observed and expected events, the statistic and p-value. It does not identify the time at which curves differ or estimate a constant difference in survival probability.

A log-rank test is unavailable for one group, more than two groups, or when no event occurs while both groups remain at risk. Crossing curves and time-varying effects deserve direct examination; a nonsignificant overall test does not establish equivalence.

17.1.4 Read and retain the output

The display shows step curves, censor marks, a reference at 0.5, optional interval shading, numbers at risk, group summaries and a life table. Long curves and tables are reduced for display, with that reduction stated; all eligible observations enter estimation and testing. The source limit is one million rows and the displayed life table has at most 200 rows.

Add to results retains summaries, a curve for each group, the available log-rank table and the bounded life table. A source change requires a new calculation. This workflow supports right censoring; delayed entry, competing risks and interval censoring need different methods.

17.2 Cox proportional-hazards regression

A Cox model relates event risk to fixed numeric or categorical predictors:

$$h(t \mid x) = h_0(t) \exp(x^\mathsf{T}\beta).$$

The baseline hazard $h_0(t)$ is unspecified. For two observations differing by one unit of predictor x_j, with other predictors held fixed, the hazard ratio is e^{β_j}.

Open **Statistics > Survival > Cox proportional hazards...**. Choose positive duration and status columns, explicitly assign the event/censor labels, and select one or more predictors. Categorical comparisons use the retained reference level shown with results.

17.2.1 Interpret a hazard ratio

An illustrative coefficient of $\log(1.5) = 0.405465$ corresponds to hazard ratio 1.5: at a given time, among subjects still at risk and under the model assumptions, the instantaneous event rate is multiplied by 1.5. It is not a statement that survival time falls by 50%, or that the probability of the event by a particular date rises by 50%. A one-unit predictor change must be stated in its measurement units.

HypoNum reports coefficients, standard errors, hazard ratios and fixed 95% asymptotic intervals, along with the partial log likelihood and row counts. Coefficient intervals have the form

$$\hat{\beta}_j \pm 1.96\, SE(\hat{\beta}_j);$$

hazard-ratio limits are their exponentials. If exponentiation overflows, the finite coefficient remains visible and the corresponding hazard-ratio field is unavailable with a warning.

17.2.2 Check the model's assumptions

The model assumes independent subjects, appropriate censoring and proportional hazards: the hazard ratio for a fixed contrast stays constant over time. Efron's method handles tied event times. Schoenfeld residuals are displayed for a bounded set of event rows. A systematic residual trend over time can suggest a proportional-hazards problem. A quiet plot, especially with few events, does not prove the assumption.

Missing duration, status or predictor values and excluded observations are counted. Nonpositive durations, unknown status labels, no-event data, singular designs and failed convergence prevent a fitted result. The number and distribution of events, not merely the total number of rows, determine how much information is available for covariate effects. The supported model uses fixed covariates; do not interpret its output as an analysis with time-varying predictors or competing events.

CHAPTER 18

Multiple imputation and missing-data uncertainty

18.1 Multiple imputation with pooled regression

Choose **Statistics > Missing data > Multiple imputation (pooled regression)...** when a defensible imputation model can use observed information to address missing values in an OLS or binary logistic analysis. Multiple imputation creates several plausible completed datasets, fits the same analysis to each and combines estimates and uncertainty. It does not turn a guessed value into a measured observation.

18.1.1 State the missingness assumption

The analysis assumes values are missing at random given the included variables: conditional on those variables, missingness does not further depend on the unobserved value. Observed data alone cannot verify that assumption. Include the outcome, its predictors and appropriate auxiliary variables that predict values or missingness.

If a participant is more likely to omit an especially high unobserved response even after conditioning on the available variables, this ordinary imputation model may be inadequate. Sensitivity to alternative missingness assumptions is a substantive study decision.

18.1.2 Choose variables and analysis

Select two to thirty imputation variables and declare each Numeric, Binary, Nominal or Ordinal. Only selected variables enter the imputation models; all missing cells in them are imputed. Ordinal is treated as nominal in this version, with that limitation disclosed: category order does not inform the imputation model.

Select a numeric outcome for OLS or a binary outcome for logistic, and one to 25 other imputation variables as predictors. For logistic analysis, explicitly select an observed event category. It has no automatic default and is retained for every completed analysis. Categorical analysis predictors use indicators against the first sorted category.

Defaults are twenty imputations, forty sweeps per chain, seed 20260923, and 95% confidence. Imputations range from two to 100; sweeps must be even, from forty to 200. The first half are warm-up and the completed dataset is taken after the final sweep. Confidence ranges from 50–99.99%. Retain the seed and settings for reproducibility within the same numerical runtime.

Excluded rows enter neither the models nor completed datasets. The source limit is one million rows. Empty or non-finite numeric cells and empty categorical cells are missing. Observed numeric values must be exactly representable in the working numeric format; unsupported large integers are refused instead of being silently changed. Observed values are preserved.

18.1.3 What chained equations do

Each completed dataset comes from its own seeded chain. Missing cells are first filled by draws from observed values of the same column. A sweep revisits incomplete variables from fewest to most missing cells, breaking ties by name. For each target, the other selected variables at their current values act as predictors. Moving a column in the table does not change that visit order, and complete variables remain unchanged.

Numeric variables use predictive mean matching. A regression is fitted to observed target values, and coefficient/variance uncertainty is drawn. A missing row's predicted value is compared with observed-row predictions; one of the five nearest observed donors is chosen at random and its actual value copied. Distance ties are randomized across all tied donors, including those at the fifth-nearest boundary.

This keeps numeric imputations among observed values, but does not make the imputation model automatically suitable. It needs enough observed targets for a full-rank regression with positive residual degrees of freedom. When its residual variance is zero, coefficient drawing degenerates to the fitted value; that special convention does not supply proper uncertainty for a deterministic relationship. Passive derived-variable constraints are unavailable.

Binary and nominal targets use regularized logistic models with normal-prior slopes and intercepts, then draw from a normal approximation to their posterior and sample categories from predicted probabilities. This regularization belongs to the *imputation* model; the final logistic analysis remains unpenalized. An incomplete categorical target needs at least two observed categories; unobserved categories are not invented.

18.1.4 Fit every completed dataset

The same analysis design is fitted to every completed dataset. An OLS fit must have an identifiable design and appropriate residual degrees of freedom. An unpenalized logistic fit must pass separation, convergence and information checks. A failed or inconclusive separation check prevents the logistic analysis; it is not interpreted as evidence that separation is absent.

If any completed dataset's analysis fails, the study stops and names that dataset. Successful completions are not pooled after silently dropping the others. No penalized or reduced analysis model is substituted.

18.1.5 How pooling includes missing-data uncertainty

For coefficient vectors $\widehat{Q}_j$ and covariances U_j, Rubin's rules are

$$\bar{Q} = \frac{1}{m}\sum_j \widehat{Q}_j, \qquad \bar{U} = \frac{1}{m}\sum_j U_j, \qquad B = \frac{1}{m-1}\sum_j (\widehat{Q}_j - \bar{Q})(\widehat{Q}_j - \bar{Q})^{\mathsf{T}},$$

$$T = \bar{U} + (1 + 1/m)B.$$

Within-imputation variance $\bar{U}$ represents ordinary analysis uncertainty. Between-imputation variance B represents variation between plausible completions. The pooled standard error of coefficient a is $\sqrt{T_{aa}}$, not a standard error from a selected completed table.

As an illustrative scalar calculation, three estimates $1.0, 1.1, 1.2$ with variance 0.04 each give $\bar{Q} = 1.1$, $\bar{U} = 0.04$, $B = 0.01$, $T = 0.0533333$. The pooled standard error is 0.230940. Taking one estimate with standard error 0.2 would omit the between-imputation contribution.

For each coefficient, with scalar diagonal entries,

$$\lambda = \frac{(1 + 1/m)B}{T}, \qquad \nu_{\text{old}} = \frac{m-1}{\lambda^2}, \qquad \nu_{\text{obs}} = \frac{\nu_{\text{com}} + 1}{\nu_{\text{com}} + 3}\nu_{\text{com}}(1 - \lambda), \qquad \nu = \left(\nu_{\text{old}}^{-1} + \nu_{\text{obs}}^{-1}\right)^{-1}.$$

This is the Barnard–Rubin small-sample convention. OLS uses complete-data $\nu_{\text{com}} = n - p$; logistic uses $\nu_{\text{com}} = \infty$ and thus $\nu = \nu_{\text{old}}$. If $B = 0$, the old degrees of freedom are infinite, but OLS still uses the finite observed-data expression. Only when both are infinite is normal reference inference used.

Intervals and two-sided tests use t_ν. For logistic coefficients, exponentiate the pooled log-odds estimate and interval endpoints. Do not average completed-dataset odds ratios as if that were the same pooling rule. Nonpositive total variance or invalid degrees of freedom makes inference unavailable.

18.1.6 Check simulation precision and chain behaviour

The relative increase in variance is $r = (1+1/m)B/\bar{U}$. The reported fraction of missing information is $\{r + 2/(\nu + 3)\}/(1 + r)$, or $r/(1 + r)$ for infinite ν; it is unavailable when $\bar{U} = 0$.

Monte Carlo standard error of the pooled coefficient is $\sqrt{B/m}$. Its ratio to pooled standard error is reported when $T > 0$, with an advisory above 10%. In the three-estimate illustration the ratio is 25%, suggesting more imputations are needed for stable numerical reporting. Increasing the number of imputations addresses this simulation precision, not a wrong missingness model.

Trace plots record every chain's imputed numeric mean/standard deviation or category proportions after every sweep, drawing at most ten chains. Stable traces after warm-up are consistent with, but do not prove, convergence. They do not verify missing at random. Review observed and imputed distributions and the scientific plausibility of the selected variables as well as numerical diagnostics.

When no included selected value is missing, HypoNum fits the analysis once and returns its ordinary inference rather than manufacturing repeated completions.

18.1.7 Retain results and inspect a completed table

Add to results saves pooled coefficients, variance components, variable summaries, settings and completed-data identities. Source changes make the result stale and disable this action until rerun.

Create completed dataset tab regenerates the chosen completion in a new tab with every source column. Selected observed cells remain unchanged; their missing cells receive that completion's imputations. Unselected columns remain source data. Types are retained when they can represent all completed values exactly; otherwise a numeric column widens to floating point or a category to text, with a disclosure.

The recorded operation includes the chosen completion and excludes the same rows. Replay checks that the regenerated completion agrees with the retained identity; changed data cannot silently create a different table. Undo removes the new tab. Use a single completion for inspection; use the pooled analysis for inferential conclusions.

18.2 Missing value analysis before imputation

Open **Statistics > Missing data > Missing value analysis...** to examine which values are missing before choosing an imputation model. Choose numeric measurements as quantitative variables and category codes as categorical variables. The command produces diagnostics; it does not fill cells or create an imputed data set. Declared user-missing codes are treated according to the active column definitions.

The univariate table gives present/missing counts and percentages. For quantitative variables it also gives mean, standard deviation and counts outside the $Q_1 - 1.5\,IQR$ and $Q_3 + 1.5\,IQR$

fences, using interpolated quartiles. These fences flag observations for investigation; they do not prove errors. The pattern table counts each distinct combination of present and missing selected variables, including patterns hidden by a marginal percentage alone.

By default, a variable missing in at least 5% of cases defines a split between its present and missing cases. Each other quantitative variable is compared across that split using Welch's test, with a difference defined as present minus missing. Holm adjustment covers the whole computed family by default. Reducing the missingness threshold creates more exploratory comparisons; it does not remove multiplicity. Categorical cross-tabs show missingness within each category, including the category variable's own missing values.

Worked pattern example. Choose **File > Import data...** and select `sample-data/manual/missing-patterns-six.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Missing data > Missing value analysis...**. Add **x** and **y** to **Quantitative variables**; leave categorical variables empty. Set the missing-percentage threshold to 5, keep **t-test p-values** at **Holm-adjusted across the family**, and choose **Tolerance 1e-6, at most 100 iterations** for **EM convergence**. Click **Run analysis** and inspect the univariate counts and pattern table. The exact linear relationship among complete pairs makes normal-model covariance inference unsuitable here; the exercise concerns missingness counts, not evidence for a missing-data mechanism. For six rows, use these (x, y) pairs:

$$(1,2),\ (2,4),\ (3,\text{missing}),\ (4,8),\quad (\text{missing},10),\ (\text{missing},\text{missing}).$$

Both columns are missing in 2 of 6 cases (33.33%). Nevertheless there are four patterns: three complete, one missing only x, one missing only y, and one missing both. Complete-case analysis retains three rows. The marginal percentages alone do not reveal that reduction.

18.2.1 Normal-model estimates and Little's test

Expectation–maximization estimates the quantitative means, covariances and correlations under a multivariate normal model using observed components. A row missing every quantitative value contributes no likelihood information. The default tolerance is 10^{-6} with at most 100 iterations. The alternate SPSS-style setting uses 10^{-4} and 25 iterations. Convergence requires both scaled parameter changes and the observed log-likelihood change to meet the tolerance; reaching the cap is explicitly reported as nonconvergence. Variables never observed together can leave covariances unidentified.

For missingness pattern p with n_p rows, observed-variable set O_p, and observed mean vector $\bar{y}_p$, Little's statistic is

$$d^2 = \sum_p n_p(\bar{y}_p - \hat{\mu}_{O_p})'\hat{\Sigma}^{-1}_{O_pO_p}(\bar{y}_p - \hat{\mu}_{O_p}), \qquad df = \sum_p |O_p| - k,$$

where k is the number of quantitative variables. Its chi-square reference is an approximation under the normal model. The test is withheld for nonconverged EM, unidentified or singular required covariance, or nonpositive degrees of freedom. A fully observed table provides no informative MCAR test.

Rejecting missing completely at random does not distinguish missing at random from missing not at random. Failure to reject does not establish MCAR. Use knowledge of collection, eligibility and reasons for nonresponse alongside these diagnostics. The EM estimates are neither completed

values nor draws from a multiple-imputation distribution. Continue to the imputation workflow only after deciding what missingness assumptions the analysis can defend.

CHAPTER 19

Fitting probability distributions

A distribution fit describes the spread and shape of a measurement. It can support probability calculations or suggest a useful model, but a close fit to one sample does not establish independent observations, process stability or a correct scientific mechanism.

19.1 Choose measurements and families

Choose **Statistics > Descriptives & tables > Distribution fit...** and select continuous numeric columns. Numeric codes, categories, temporal columns and internal identity columns are not eligible measurements here. Each column needs at least eight finite observations. Select at least one candidate family.

Normal, Student t, Laplace, Logistic, Lognormal, Gamma, Weibull and Exponential begin selected. A candidate can still be unavailable when its model has no finite maximum-likelihood estimate or its support does not contain the observations. The result states why.

19.2 Decide the origin before fitting

Lognormal, Gamma, Weibull and Exponential fit $X = \theta + Y$ with a *fixed* origin θ. **Origin (fixed), default 0** accepts a finite value, including a negative one. Blank means zero. It never asks the fit to estimate a location. The result names the origin, for example *Weibull (origin fixed at 5)*.

Choose the origin from physical meaning or independent knowledge. Selecting a sample minimum or trying origins until the fit looks best introduces an unaccounted selection step. All positive-support families and columns in one run share the origin; run columns separately if their justified origins differ.

Lognormal, Gamma and Weibull require every $y_i = x_i - \theta$ to be strictly positive and finite, with at least two distinct values. Exponential permits zero distance from the origin but needs a positive mean. Observations are not removed or shifted to make a family eligible. A shifted distribution does not model censoring, truncation or a detection limit merely by having a nonzero origin.

19.3 Worked example: a normal likelihood fit

Choose **File > Import data...** and select `sample-data/manual/distribution-normal-eight.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Descriptives & tables > Distribution fit...**. Add **value** to the selected columns. Keep **Normal** checked and clear all seven other family checkboxes. Click **Fit distributions** and inspect the normal fit's parameters, information criteria and goodness-of-fit result. The complete values are

$$8,\ 9,\ 9,\ 10,\ 10,\ 11,\ 11,\ 12.$$

Select `value` and Normal only. The maximum-likelihood estimates are

$$\hat{\mu} = 10, \qquad \hat{\sigma} = \sqrt{\frac{12}{8}} = 1.2247449.$$

This scale divides the centred sum of squares by n. A descriptive sample SD divides by $n - 1$, so the two quantities differ.

The log likelihood is −12.9733687. With two fitted parameters, AIC is 29.9467374 and BIC is 30.1056205. The Anderson–Darling statistic is 0.2540888. With the default 9,999 bootstrap samples and seed 0, the goodness-of-fit p-value is 0.7707; the 95% interval for Monte Carlo tail-probability error is $[0.762309, 0.778887]$.

The result finds no clear mismatch in this small sample. It does not prove normality, especially in the tails. AIC and BIC have no stand-alone passing threshold; selecting Normal only does not compare it with another family.

19.4 Parameter estimates and model comparison

For positive-support models write $y_i = x_i - \theta$. The lognormal estimates are

$$\hat{\mu} = \overline{\log y}, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_i (\log y_i - \hat{\mu})^2.$$

Gamma shape a solves

$$\log a - \psi(a) = \log \bar{y} - \overline{\log y}, \qquad \widehat{\text{scale}} = \bar{y}/a.$$

Weibull shape k solves

$$\frac{1}{k} + \overline{\log y} - \frac{\sum_i y_i^k \log y_i}{\sum_i y_i^k} = 0, \qquad \widehat{\text{scale}} = (\overline{y^k})^{1/k}.$$

Exponential scale is $\bar{y}$, its mean above the origin. The likelihood is for the original measurements; translation has Jacobian 1.

For log likelihood ℓ and q estimated parameters,

$$\text{AIC} = 2q - 2\ell, \qquad \text{BIC} = q \log n - 2\ell.$$

The fixed origin is not estimated: $q = 2$ for Lognormal, Gamma and Weibull, and $q = 1$ for Exponential. Candidates are ranked by BIC, then AIC for an exact BIC tie. Smaller values favour one eligible candidate over another fitted to the same observations. Comparisons across different populations or response scales need additional justification.

19.5 Why some familiar candidates are unavailable

Free-location Gamma and Weibull likelihoods can diverge when shape is below 1 and location approaches the smallest observation. A free-location Lognormal likelihood can also grow without bound. There is then no finite global maximum-likelihood solution. A local optimizer returning numbers would not resolve that problem. Free-location Exponential has a finite boundary estimate, but this workflow offers the declared-origin form.

Student t with free location, scale and degrees of freedom has an unbounded likelihood too. For distinct observations, place location on one value, choose $0 < \nu < 1/(n-1)$, and let scale s approach zero. The likelihood behaves as $s^{(n-1)\nu-1}$ and diverges. Student t is consequently listed as not compared; HypoNum does not silently fix its degrees of freedom or rank a local solution as a global fit.

19.6 Goodness of fit and simulation uncertainty

For the leading single-mode fit, the unmodified Anderson–Darling statistic at fitted CDF F is

$$A^2 = -n - \frac{1}{n}\sum_i (2i-1)\left[\log F(x_{(i)}) + \log\{1 - F(x_{(n+1-i)})\}\right].$$

The calibration simulates 9,999 samples from the fit with seed 0 and refits each using the same estimator and fixed origin. If K simulated statistics are at least the observed statistic, $p = (K+1)/10,000$. Ties count and the reported value cannot be zero. A failed simulated fit is not dropped or redrawn: the p-value becomes unavailable, with the failure count disclosed.

The 95% exact binomial interval describes simulation error, not uncertainty about the fitted parameters. Interpretation bands use 0.01 and 0.05. A categorical assessment is given only when the whole simulation interval lies in one band; otherwise the result says simulation uncertainty crosses a threshold.

This is an approximate parametric-bootstrap check. Choosing the best of several candidates does not make its p-value selection-adjusted. An Exponential observation exactly at the origin has $F = 0$, making this statistic undefined. The likelihood fit remains available but the check is withheld.

19.7 Multimodality and graphical review

Before ordinary fitting, a majority decision from kernel-density, Gaussian-mixture and bimodality-coefficient detectors can identify multiple modes. Results then include an overall Gaussian mixture and descriptive fits within hard-assigned modes. These are exploratory summaries, not a maximum-likelihood mixture of arbitrary selected families.

The main table shows column, N, mode count, leading fit, BIC, AIC, p-value and interpretation. Double-click or choose **View details** for density and Q–Q diagnostics, parameters and warnings. A running batch can be cancelled. Inspect the region relevant to the decision: a plausible centre can coexist with poor extreme-tail probabilities.

CHAPTER 20

Extended regression and count-table models

An analysis is specified by its response scale, mean model, error model and population. A familiar label such as regression does not make methods interchangeable. The following windows expose these choices explicitly.

20.1 Enter predictors in planned blocks

Statistics > Correlation & regression > Linear regression (blocks and diagnostics)... fits cumulative blocks on one common complete-case population. A predictor belongs to one block. Earlier blocks remain in the model when a later block uses forward, backward or stepwise selection. Enter-all is the default; selection uses Gaussian AIC or BIC, including the error-variance parameter. Reported tests after selection do not account for the search.

For successive models, the change statistic is $F = [\Delta R^2/(p_k - p_{k-1})]/[(1 - R_k^2)/(n - p_k)]$, with p_k counting the intercept and all coefficients. It is meaningful because the models use the same rows. The output also distinguishes standardized, studentized and deleted residuals, leverage, Cook's distance and DFBETAs. The default case screen uses $|e_i/s| > 3$, not a studentized residual.

Collinearity diagnostics include the intercept in an uncentred design with columns scaled to unit length. Large condition indices and shared variance proportions identify potentially unstable coefficient directions. A rank-deficient design is not repaired by silently omitting a predictor. Durbin–Watson describes adjacent residuals in table order and has no reported p-value; that order must represent time for a temporal interpretation.

20.2 Know which scale a curve fits

Statistics > Correlation & regression > Curve estimation... offers named linearized curves. Linear, quadratic and cubic models fit the original response; power, exponential and growth forms fit a logarithmic response. The logistic growth curve requires a declared upper bound. Without a predictor, the default sequence retains original row positions, including gaps caused by exclusions; renumbering retained rows is an explicit alternative.

For a power curve $Y = b_0 t^{b_1}$, least squares fits $\log Y = \log b_0 + b_1 \log t$, requiring positive Y and t. Values $t = (1, 2, 4)$ and $Y = (3, 12, 48)$ lie on $Y = 3t^2$. That exact relationship illustrates the transformation, while a noisy log-scale fit targets a back-transformed median under symmetric transformed errors. It is not an original-scale least-squares mean fit.

The reported R^2, tests and intervals use the named fitting scale and cannot rank models fitted to different response transformations as if they shared a loss. All candidates use the same retained population; a curve with incompatible support becomes unavailable instead of dropping additional observations. For original-scale nonlinear least squares, choose the separate nonlinear-regression procedure.

20.3 Estimate a variance power deliberately

Statistics > Correlation & regression > Weight estimation... uses $\mathrm{Var}(e_i) = \sigma^2 z_i^p$ for a strictly positive variance predictor z. The fitting weights are z_i^{-p}. The default power grid is -2 to 2 in steps of 0.5. Its Gaussian profile likelihood includes both weighted residual error and the weight determinant:

$$\ell(p) = -\frac{n}{2}\{\log(2\pi) + \log(WSSE(p)/n) + 1\} + \frac{1}{2}\sum_i \log w_i.$$

Choosing the smallest WSSE alone would answer a different optimization problem. An optimum at a grid endpoint is reported as a constrained-grid result. Coefficient uncertainty treats the selected power as fixed and does not include the uncertainty of that selection. These variance weights are different from the Data menu's integer frequency weights. A constant response is refused. Weight estimation also refuses a fit whose positive residual variance cannot be resolved at working precision, including a perfect fit. Such a fit cannot supply a usable finite Gaussian profile-variance estimate.

20.4 Two-stage least squares requires valid instruments

Statistics > Correlation & regression > Two-stage least squares... separates endogenous regressors, included exogenous regressors and excluded instruments. Included exogenous terms also serve as instruments. There must be enough excluded instruments and the retained instrument/regressor relation must have sufficient rank. Those conditions do not establish instrument validity.

With structural regressors X and instrument projection P_Z, $\hat{\beta} = (X'P_ZX)^{-1}X'P_Zy$. Residuals for uncertainty are $u = y - X\hat{\beta}$, not the residuals of regressing on the fitted regressors. Choose classical or HC1 covariance and report that choice. First-stage strength statistics assess relevance, not exclusion restrictions. With several endogenous regressors, separate first-stage tests do not guarantee joint strength. The Sargan test remains a homoskedastic diagnostic; selecting robust coefficient covariance does not turn it into Hansen's J. Structural-residual R^2 may be negative.

20.5 Generalized linear models

Statistics > Generalized & mixed models > Generalized linear models... fits $g(\mu_i) = o_i + x_i'\beta$, with optional offset o_i on the link scale, factors, numeric covariates and two-way interactions. The selected family determines the response support and variance function.

Family	Variance function	Default link
Normal	1	Identity
Binomial	$\mu(1-\mu)/m$	Logit
Poisson	μ	Log
Negative binomial	$\mu + \mu^2/\theta$	Log
Gamma	μ^2	Inverse
Inverse Gaussian	μ^3	Inverse squared
Tweedie, $1 < p < 2$	μ^p	Power $1 - p$

Binomial data are 0/1 or events with declared trials. Counts must be nonnegative; gamma and inverse-Gaussian responses must be positive. Tweedie allows zeros and positive observations

with a supplied variance power. Choose only an offered family/link combination; constraints on the inverse link's branch remain part of the model.

Dispersion is fixed at one for binomial and Poisson; negative binomial has its separate estimated or fixed θ. Other families use the stated moment/deviance or maximum-likelihood dispersion choice. The covariance and reference distribution are reported separately. HC0 is an asymptotic sandwich choice, not a small-sample exact correction. Moment-dispersion t/F inference is generally approximate outside the ordinary normal identity-link model.

The separate **Statistics > Generalized & mixed models > Negative binomial (count)...** workflow reports joint observed-information uncertainty for coefficients and estimated NB2 dispersion α. The general GLM window uses conditional-θ expected-information coefficient uncertainty, where $\alpha = 1/\theta$. These are distinct covariance conventions even when fitted means agree. Choose the intended convention and report it. Similarly, the separate binary and count windows use first-level reference coding, while the general GLM window uses sum coding; coefficient meanings must be compared on a common basis.

Tweedie is fitted by quasi-likelihood here, without a normalized density; log-likelihood, AIC, BIC and likelihood-ratio tests are consequently absent. For likelihood families, a robust covariance choice does not make likelihood-ratio tests robust. Nonconvergence, support failures, unidentified designs and infinite-estimate boundaries are reported rather than replaced with another model.

Marginal means average the *linear predictor* over an equal-weight factor grid, at covariate means and the mean offset, then invert the link. They need not equal the arithmetic average of response-scale predictions. For log-scale predictors $\log 2$ and $\log 8$, that procedure gives 4; averaging their response means gives 5. State which estimand you report. Type-III factor tests use the declared coding at covariate zero and can differ from comparisons at covariate means when interactions are present.

Add fitted means and residuals retains fitted means, their conditional mean intervals and residuals. An inverse-link interval reaching a pole can be unbounded and is labelled accordingly. A future-response prediction interval is offered only for the normal identity-link case. Confidence in a mean is not uncertainty of a new individual response.

20.6 Loglinear models for count tables

The **Statistics > Generalized & mixed models > Loglinear** commands model expected counts $\log \mu_c = o_c + x_c'\lambda$ in tables of two to six factors. Supply individual cases or a count column. A generating term such as `A*B` includes its lower-order terms, preserving hierarchy. Choose Poisson, multinomial or product-multinomial sampling according to which totals were fixed by design.

Structural zeros identify impossible cells and remove them from the sample space; observed sampling zeros remain observations. A pseudocount is an explicit, labelled analysis choice. Degrees of freedom use allowable cells minus design rank. Sampling-zero patterns can put maximum likelihood on a boundary: extended fitted counts may then remain available while ordinary finite coefficients are withheld.

For an independence table with row totals 20 and 20 and column totals 18 and 22, expected counts are $(9, 11)$ in each row. Compare those expectations with observed counts using residuals and the reported Pearson and likelihood-ratio statistics; sparse expected cells limit their chi-square approximations.

The logit form fixes the explanatory-factor interaction and expresses dependent-factor category logits against a reference. Hierarchical model selection removes eligible terms by likelihood-ratio comparisons while retaining sampling-design terms. Its final tests remain exploratory after the search. A saturated model's perfect fit is not evidence of predictive performance on new tables.

20.6.1 Reference categories, profiles and exact-zero functionals

In **Statistics > Generalized & mixed models > Loglinear > General...**, **Reference levels (factor=level)** sets explicit factor references. Changing coding changes coefficient labels, not fitted counts or equivalent estimable contrasts. With **Coefficient intervals: Wald and profile likelihood**, a profile interval retains values v whose constrained deviance is at most $\chi^2_{1,1-\alpha}$ above the fitted minimum. The same likelihood is refitted under the constraint. These are regular asymptotic intervals, not exact small-sample limits. A failed bracket or refit is unresolved, not an infinite limit. Boundary inference remains withheld. Parameters fixed by multinomial or product-multinomial sampling are not profiled.

A requested functional names a coefficient and its weight. Separate weighted terms within one named request with semicolons; separate complete named requests with `|`. For example, the request name precedes the colon in `cancel:1A[A=a1];-1A[A=a1]`. Its coefficient vector must lie in the support design's row space. An exactly zero or cancelling vector is an algebraic identity: value and standard error are zero, with no stochastic z test, p-value or profile interval. This fact also holds at a boundary; it does not restore regular inference for other functionals.

Worked example

Choose **File > Import data...** and select `sample-data/manual/loglinear-six.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Generalized & mixed models > Loglinear > General...**. Select factors **A** and **B**, count column **n**, independence terms `A;B`, Poisson sampling and last-level references. Leave structural-zero and pseudocount settings empty. In the functional request enter `zero:0A[A=a1]|cancel:1A[A=a1];-1A[A=a1]`; click **Run analysis**. Both functionals are estimable and identically zero, with estimate, standard error and exact lower/upper values all zero. Its test and profile limits are absent for each request. The six observed counts are $(25, 15)$, $(10, 20)$ and $(30, 12)$; the zero weight gives an identity for any valid fit, not evidence of no effect between distinct categories.

CHAPTER 21

CHAID, nearest neighbours and finite mixtures

These methods describe or predict patterns across independent rows. Define the target, available predictors and intended future population before fitting. A random holdout assesses exchangeable future rows; it does not protect against subject duplication, a changing process or future information already present in a predictor.

21.1 CHAID trees

Choose **Statistics > Multivariate > CHAID tree...**. The target can be categorical or quantitative. Declare ordered predictors deliberately. Numeric predictors are divided into equal-frequency bins, 10 by default, which are treated as ordered. Missing predictor values form a floating category that may merge with any category.

At each node, categories with similar target distributions are merged using Pearson chi-square tests for a categorical target or one-way F tests for a quantitative target. Nominal categories may merge freely; ordered categories merge only with neighbours, apart from the floating missing category. Ordinary CHAID stops merging at the chosen merge threshold. Exhaustive CHAID continues down to two groups and retains the most significant grouping encountered along that merge sequence; it does not search all partitions. The default **Category search: Merge-only** retains that behavior. The opt-in **Kass merge/re-split** mode adds binary re-splitting of compounds containing at least three original categories, following the step described by Kass (1980, p. 121). It supports categorical targets and refuses the separate exhaustive merge-history option.

Within-compound re-split alpha defaults to 0.01 and must be strictly smaller than the merge threshold. Pairwise merging uses raw p greater than the merge threshold; a compound is re-split only when its best admissible binary partition has raw p below the re-split threshold. Successful re-splits return to merging. Nominal partitions count each complementary pair once. Ordered partitions use contiguous cuts; the floating missing category may join either side or stand alone. Ties follow the deterministic category and partition order.

Maximum category tests per predictor/node defaults to 100000. Exhausting this budget or revisiting a partition refuses the analysis rather than declaring convergence. This is a bounded stepwise search, not a global optimum. A published worked-table control does not establish independent validation of every tree or its post-selection inference.

The overall split test includes a Bonferroni multiplier for the candidate category reduction. For c ordered categories reduced to r groups this multiplier is $\binom{c-1}{r-1}$; nominal and floating-category reductions have different combinatorial counts. The smallest adjusted p-value selects the split if the split threshold and size rules permit it. These local adjustments are not a familywise guarantee for the entire discovered tree.

The initial merge and split thresholds are both 0.05; parent/child minima are 100/50 rows and maximum depth is 3. The program does not silently shrink these minima for a small study. An

unsplit tree can be the correct outcome. A seeded random split reserves 30% for assessment by default. Read held-out misclassification or mean squared error separately from training error. The tables show node conditions, counts, predictions, merge tests and reasons for stopping. Added predictions and node numbers include training rows, whose predictions remain in-sample.

21.2 Nearest neighbours

Choose **Statistics > Multivariate > Nearest neighbours...** and a numeric feature set. Features are standardized using training means and standard deviations. Euclidean and city-block distances impose different neighbourhoods; choose a distance for the scientific meaning of the features, not solely its best observed score. A constant feature cannot contribute a standardized distance.

For regression, the prediction is the mean of the k nearest training targets. For classification it is their majority class. A tied vote follows the nearest neighbour among the tied classes; equal-distance ordering follows training-row order. For example, with an already standardized one-dimensional training set $x = (0, 1, 3)$, targets $(0, 2, 6)$ and $k = 2$, a query at $x = 1.2$ predicts $(2 + 0)/2 = 1$. This is a neighbourhood average, not a fitted straight line.

The default split reserves 20% for a test set, then 25% of the remainder for validation, leaving 60% for training. The grid $k = 1, 3, 5, 7, 9, 11, 15$ is selected by validation error, ties favouring smaller k. The untouched test set estimates the selected model's error. The selected search mode and work budget bound the distance calculation. The extended exact-search option supports larger training sets within its displayed limits while retaining exact neighbour and boundary-tie rules; it is not an approximate search. Feature-importance comparisons remove one feature at a time while keeping selected k fixed and compare validation errors; they are not causal effects. Added scores for training rows may use the row itself among its neighbours.

21.3 Model-based clustering

Choose **Statistics > Multivariate > Model-based clustering...**. Select either all scale variables or all categorical variables. A Gaussian mixture uses diagonal covariance by default, with a full-covariance option. A latent class model uses conditional independence of categorical variables within each class. The separate TwoStep-style mode can combine scale and nominal variables; it does not reinterpret that combination as a Gaussian mixture.

For K components with probabilities π_g and component densities or probabilities f_g, the likelihood is

$$L = \prod_i \sum_{g=1}^{K} \pi_g f_g(y_i), \qquad \pi_g > 0, \quad \sum_g \pi_g = 1.$$

The default search fits $K = 1, \ldots, 6$ from 10 seeded starts per K and chooses the smallest $BIC = -2 \log L + m \log n$, where m counts free parameters. EM may find different local maxima. The selected fit is the best successful converged start under the recorded policy, not a proven global maximum. Inspect failed and unconverged starts before treating a choice of K as substantive evidence for distinct populations.

Variance floors prevent numerical collapse; components with insufficient effective membership, invalid covariance or a variance at its floor fail. The result lists candidate likelihoods, parameter counts, BIC, component weights and profiles. For a two-component observation with weighted likelihoods $\pi_1 f_1(y) = 0.12$ and $\pi_2 f_2(y) = 0.08$, posterior membership is $0.12/(0.12 + 0.08) = 0.60$ in component 1. Calling its hard assignment "1" does not remove the 40% conditional allocation uncertainty. Saved memberships remain blank for missing required values or unseen categories.

Cluster names and order are arbitrary; interpret profiles and stability rather than attaching meaning to numeric labels.

The latent-class observed-item option evaluates each row's likelihood from its observed items. Wholly missing rows are excluded from fitting and BIC sample size; they do not create observed evidence for class membership. This option needs an appropriate ignorable missingness interpretation and conditional independence of items within class. One reference dataset with partial missingness cannot establish those assumptions for another study.

TwoStep-style fitting uses seeded-order cluster-feature preclustering, minimum merge-distance agglomeration, a BIC-change stage and distance-change refinement. It records standardization, seed, tree/leaf bounds, merge threshold and rebuild policy; no outlier handling is supplied. The criterion is explicitly identified and is not ordinary mixture log likelihood. Frozen scoring returns merge distance, not posterior probability. If a required ratio is undefined or its extra agglomerative state is absent, automatic selection is refused without a different selector being substituted. For example, two available CF leaves are not evidence that only one cluster exists. A genuine one-leaf fit and a defined first-stage one-cluster choice are identified separately. Labels remain arbitrary.

Large-sample silhouette summaries may use a seeded sample of focal rows against the full fitted clusters. Read the sample size, finite-population adjustment and reported sampling error. Sampling focal rows does not mean clusters were refitted to that subset.

CHAPTER 22

Distances, canonical relationships and category quantification

22.1 Multidimensional scaling

Choose **Statistics > Multivariate > Multidimensional scaling...** to represent object dissimilarities by distances in one to six dimensions. Input can be a square symmetric nonnegative matrix with zero diagonal and no missing entries, or scale variables from which distances are computed. Rows and columns of a supplied matrix must refer to objects in the same order. Off-diagonal zeros are legitimate; they are not missing values.

For computed distances, complete rows are used. Standardization to unit standard deviation is initially selected; retaining original units instead lets their scales influence the map. Euclidean, city-block and correlation distances answer different similarity questions. The correlation option compares profiles using $\sqrt{2(1-r)}$. A supplied distance matrix is never standardized as though its columns were measurements.

Classical scaling eigendecomposes $B = -JD^{(2)}J/2$, where $J = I - \mathbf{1}\mathbf{1}'/n$. Coordinates use eigenvectors multiplied by square roots of positive eigenvalues. Negative eigenvalues reveal departures from Euclidean distances; the reported goodness-of-fit ratios use both the absolute and the positive eigenvalue totals. For distances among points $0, 1, 3$ on a line, the one-dimensional centered coordinates are $(-4/3, -1/3, 5/3)$, up to reflection, and reproduce distances $1, 2, 3$ exactly.

Ordinal scaling preserves dissimilarity order and fits monotone disparities while minimizing stress. Secondary ties, the default, assign equal disparities to equal dissimilarities; primary ties allow them to differ. Seeded additional starts can improve a local solution. Read convergence and the actual definition of reported stress: raw stress, normalized stress and Kruskal stress-1 have different denominators. The Shepard diagram helps reveal poorly represented pairs. A map's reflection, rotation or translation does not change its distances.

22.2 Canonical correlation

Choose **Statistics > Multivariate > Canonical correlation...** and two sets of quantitative variables. The method finds linear combinations $U = a'X$ and $V = b'Y$ with maximal correlation, then successive pairs uncorrelated with preceding pairs. Complete independent rows are required. Within-set linear dependence or a perfect canonical correlation prevents the stated inferential calculation and is refused.

Read standardized coefficients together with structure correlations (each variable with its own canonical variate) and cross-structure correlations. Large canonical correlation alone does not mean that most original-variable variance is explained. Redundancy for a set is its mean squared structure correlation multiplied by the squared canonical correlation. For example, a mean squared structure correlation of 0.4 and canonical correlation 0.8 give redundancy $0.4(0.8)^2 = 0.256$.

Bartlett's sequential test of the remaining canonical correlations uses

$$-\left(N-1-\frac{p+q+1}{2}\right)\sum_{i=k}^{\min(p,q)}\log(1-r_i^2)\mathrel{\dot\sim}\chi^2_{(p-k+1)(q-k+1)}.$$

It assumes independent multivariate-normal observations and is approximate. The sign of a canonical pair is arbitrary; small coefficient differences can also reflect unstable or highly correlated predictors.

22.3 Categorical principal components

Choose **Statistics > Multivariate > Categorical principal components...**. CATPCA assigns numeric category values to improve a low-dimensional representation. Nominal variables permit unrestricted category values; ordinal variables impose a monotone order; numeric variables retain a linear transformation. Verify the ordering of ordinal text or numeric categories before fitting. The implementation uses one quantification per variable shared across dimensions, complete rows and equal variable weights.

Alternating least squares updates category quantifications and object scores. It may find a local optimum. Read the achieved loss and convergence, component eigenvalues, loadings and category values. Numeric-only inputs reduce to correlation-based PCA. The CATPCA alpha printed for a component is a function of its eigenvalue; it is not evidence that all items measure one construct. Added object scores remain blank for unsupported categories or incomplete rows.

22.4 Multiple correspondence analysis

Choose **Statistics > Multivariate > Multiple correspondence analysis...** for several categorical variables. The analysis operates on their complete indicator matrix: each row has one active category per variable. With Q variables and J active categories, raw inertias sum to $(J-Q)/Q$. Use category contributions and representation quality alongside coordinates; nearby rare categories can have a visually striking but weakly supported relationship. Supplementary variables are projected after fitting and do not define the active axes.

Missing values can be excluded or treated as a distinct category by the declared option. A missing category describes a response pattern; it is not an imputation. Empty active categories are removed. The current analysis accepts 2–20 variables, at most 200 active categories and at most 10 dimensions.

The Benzécri adjustment retains only $\lambda_s > 1/Q$ and replaces raw inertia by

$$a_s=\left[\frac{Q}{Q-1}\left(\lambda_s-\frac{1}{Q}\right)\right]^2.$$

Greenacre's percentage convention uses a different adjusted total. These are reporting conventions for interpreting inertia, not changes to the principal coordinates. Although a Burt cross-product can be used in the calculation, the reported raw inertias are those of indicator MCA, not the squared inertias of a separate correspondence analysis of the Burt table.

CHAPTER 23

Grouped survival and changing covariates

23.1 Actuarial life tables

Choose **Statistics > Survival > Life tables...** when survival is to be summarized over declared intervals. Select time, status and explicit event and censor values. Other status values are errors. Define increasing boundaries, or a width and last boundary that is an integer number of widths. Intervals are closed on the left and open on the right: an event at 2 belongs to $[2, 4)$ rather than $[0, 2)$. A final open interval follows the last boundary. Missing required fields omit the row and are counted.

For n_i subjects entering, w_i withdrawals and d_i events in an interval, the effective risk count is $n'_i = n_i - w_i/2$, conditional event probability $q_i = d_i/n'_i$, and conditional survival $p_i = 1 - q_i$. Survival through the finite interval is $S_i = \prod_{j \le i} p_j$, with Greenwood standard error

$$SE(S_i) = S_i \sqrt{\sum_{j \le i} \frac{q_j}{n'_j p_j}}.$$

The half-withdrawal adjustment assumes approximately uniform withdrawal within the interval and noninformative censoring. For width b_i, midpoint hazard is $2q_i/[b_i(1 + p_i)]$ and density is $S_{i-1}q_i/b_i$. Neither a plotted connecting line nor this midpoint approximation supplies an exact within-interval survival curve.

For example, 16 entrants, one withdrawal and three events in $[0, 2)$ give $n' = 15.5$, $q = 3/15.5 = 0.193548$, survival 0.806452, midpoint hazard 0.107143 per time unit and density 0.096774 per time unit. The last open interval has no endpoint or width: survival is shown at its start only, and hazard/density are unavailable. Zero-event and all-fail boundaries can make standard-error formulas undefined; unavailable does not mean zero uncertainty. Optional strata produce separate tables and a grouped Gehan (Lee–Desu) comparison. Each time is represented by its interval start; at a tied grouped time, an event precedes a withdrawal. Pairwise scores sum to zero across all subjects. With group score sums U_j, group sizes n_j and total size N, the statistic is

$$(N - 1)\frac{\sum_j U_j^2/n_j}{\sum_i S_i^2},$$

compared with chi-square on groups minus one degrees of freedom. If every score is zero, the test is unavailable. This is distinct from exact-time log-rank inference, and no proprietary-package output equivalence is claimed.

The actuarial median linearly interpolates the first finite interval whose end survival reaches one half, using that interval's own width. An exact one-half endpoint gives the earliest crossing. If the crossing occurs only in the open last interval, no point is interpolated; the result states that the median exceeds the last boundary. If survival remains above one half, it is not reached. No median standard error or confidence interval is supplied.

23.2 Cox regression with time-dependent covariates

Choose **Statistics > Survival > Cox regression with time-dependent covariates...**. The model is $h_i(t) = h_{0s}(t)\exp\{z_i(t)'\beta\}$, with optional stratum-specific baseline hazards and common coefficients. Independent subjects may contribute several episodes, each $(\text{start}, \text{stop}]$. An event at t uses covariates in force immediately before t.

A treatment change at day 50 is represented by an old-treatment episode $(0, 50]$ and a new-treatment episode starting at 50. A measurement taken at an episode's endpoint belongs on the next episode. Using future or endpoint information for the preceding period can create time-related bias. Episodes may not overlap within a subject. Gaps are allowed and disclosed; a subject is not at risk in a gap. At most one event is supported, on the last episode. Missing information on any episode excludes that subject as a whole.

The alternative one-row-per-subject layout begins at origin 0. Declared time functions can add $xg(t)$ using $g(t) = t$, $\log(t - t_0)$, or $\mathbf{1}\{t > \tau\}$. The origin or change time is supplied, not estimated by a search. A log-time function requires the relevant times to exceed its origin. The log hazard ratio becomes $\beta_x + \gamma_x g(t)$. The application expands episodes where required while preserving subject exposure; the displayed preflight count must fit the selected row bound.

Efron or Breslow handling of tied events is explicit. The fit maximizes partial likelihood. Non-convergence or a detected monotone-likelihood direction is refused; a finite displayed coefficient alone does not establish that a finite likelihood maximum exists. The separation check is sufficient, not exhaustive. Model-based covariance uses the information matrix; robust covariance aggregates score contributions over whole subjects. Treating episodes as independent observations would give the wrong clustering unit. Wald intervals for hazard ratios use the chosen covariance and a log scale.

For models without a declared time-varying effect, the proportional-hazards score diagnostic can use the Kaplan–Meier, rank or identity time transform. The rank transform depends on episode representation. For a model already containing the declared time interaction, that diagnostic is withheld; interpret the time-interaction coefficient under its chosen form instead.

A singular or numerically unstable information matrix can also prevent the proportional-hazards diagnostic from being estimated. In that case its p-value is withheld with a reason and the fitted Cox model is retained. Events occurring at only one distinct time provide no variation in the centred time transform; this alone does not invalidate the fit.

Predicted survival requires a complete covariate path over contiguous segments from time zero, including the stratum where needed. It combines the fitted baseline increments with the path's risk at each event time:

$$\hat{S}(t) = \exp\left\{-\sum_{t_k \le t} \Delta\hat{\Lambda}_0(t_k)\exp[z(t_k)'\hat{\beta}]\right\}.$$

A hypothetical path is not an estimated causal intervention. Recurrent events, overlapping follow-up and unspecified future covariate values are outside this command's supported interpretation.

CHAPTER 24

Grouped summaries and cross-tabulation

A grouped summary makes the population visible before a model is fitted. It can reveal unequal group sizes, missing measurements, sparse categories or a pattern hidden by an overall average. Choose **Statistics > Descriptives & tables > Grouped summaries and cross-tabs...**.

24.1 Summarize a numeric response

The initial task is **Grouped numeric summaries**, with the first numeric column selected and no groups or frequency column. Choose the response and up to three grouping columns. Without a group, the result describes all eligible rows. With groups, each observed combination has a row, ordered by its text labels.

Each row gives source count, valid N, missing N, mean, sample SD, SE, a confidence interval for the mean, median, quartiles, minimum, maximum and range. Confidence starts at 95%; 90% and 99% are available. For an ordinary group of n observations,

$$s^2 = \frac{\sum_i (x_i - \bar{x})^2}{n-1}, \qquad SE(\bar{x}) = \frac{s}{\sqrt{n}}, \qquad \bar{x} \pm t_{1-\alpha/2,n-1} SE(\bar{x}).$$

The interval concerns the population mean. It is not a range containing 95% of individual measurements. Exact small-sample calibration assumes independent normal observations; otherwise a suitable large-sample approximation is needed.

Choose **File > Import data...** and select `sample-data/manual/summary-five.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Descriptives & tables > Grouped summaries and cross-tabs...**. Set **Table type** to **Grouped numeric summaries**, **Numeric value** to **value**, **Frequency** to **No frequency**, and **Confidence** to 95. Leave group columns empty and click **Run tabulation**. Inspect the single summary row. The five values $8, 9, 10, 11, 12$ have mean 10, sample variance 2.5, sample SD 1.581139 and SE 0.707107. The central observations are fairly close to 10, but the mean's interval and the range 8–12 answer different questions.

24.2 Understand frequency counts

An optional frequency column represents repeated identical observations. It must contain finite nonnegative integers.

Choose **File > Import data...** and select `sample-data/manual/frequency-three.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Descriptives & tables > Grouped summaries and cross-tabs...**. Set **Table type** to **Grouped numeric summaries**, **Numeric value** to **value**, **Frequency** to **frequency**, and **Confidence** to 95.

Leave group columns empty and click **Run tabulation**. The values 8, 10 and 12 have frequencies 1, 3 and 1, the summary describes the expanded sequence $8, 10, 10, 10, 12$, with valid $N = 5$. These are replication counts, not survey weights or inverse variances.

Zero-frequency rows contribute to source-row accounting but not to moments, quantiles, extrema or range. Each frequency and total replicated N must be at most $2^{53} - 1$. Median and quartiles use linear, type-7 quantiles of the replicated observations without expanding the data.

A group with only missing responses stays visible with valid N=0 and unavailable statistics. Rows without a group value are omitted from group rows and disclosed in overall missing accounting. Exclusions are separate from missingness.

24.3 Build a categorical table

Choose **Categorical cross-tabulation**. Assign a row category and, optionally, a column category. Margins start on; row, column and overall percentages start off and can be selected independently.

Choose **File > Import data...** and select `sample-data/manual/binary-outcomes-twenty.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Descriptives & tables > Grouped summaries and cross-tabs...**. Set **Table type** to **Categorical cross-tabulation**, **Cross-tab row** to **group**, **Cross-tab column** to **outcome**, and **Frequency** to **No frequency**. Enable **Row percentages**, **Column percentages**, **Overall percentages**, and **Include margins**. Click **Run tabulation** and inspect the A/yes cell; the display may list no before yes alphabetically. The counts, arranged here as Yes then No, are

	Yes	No	Total
A	8	2	10
B	1	9	10
Total	9	11	20

the A/Yes cell is 80% of row A, $8/9 = 88.89\%$ of the Yes column, and 40% of all observations. Always state the denominator when communicating a percentage.

The table includes every observed row and column level. A missing combination is an observed zero, not proof of a structurally impossible event. A zero denominator makes a percentage unavailable. Frequency counts, when chosen, supply both cell totals and denominators. The table is descriptive; an association test is a separate analysis.

24.4 Display, export and save

The default limits are one million source rows, twenty million selected input cells and two million computed output cells. Numeric summary rows emit 15 cells and categorical rows emit 7, including displayed margins. Exceeding a limit stops the calculation; it does not silently reduce the analyzed population.

The window displays at most 200 result rows and states any reduction. **Export full CSV** writes the complete computed table to the entered path. **Add to results** retains the bounded 200-row view, settings, counts and its reduction notice. Reopening a project restores that snapshot; a new calculation is needed to recover a full live CSV export.

CHAPTER 25

Descriptive reports and categorical association

Use the **Statistics > Descriptives & tables** commands to review variables and distributions before fitting a model. Labels, missing-value definitions and the population in force when a report runs are retained with its result. Displaying a descriptive report does not establish the assumptions of a later inferential analysis.

25.1 Choose a report for the question

Codebook... describes each variable's storage type, measurement level, labels, display settings and missing-value specification. Categorical variables include counts and labels, including labelled categories with no cases; scale variables include numerical summaries.

Case summaries... lists selected variables for a bounded number of cases, 100 by default, with optional groups. Its numerical summaries use all eligible cases, not just the displayed rows. Each original row is listed once, with its frequency weight when one is declared. With frequency weights, the mean uses valid values with positive weights. Their weighted sum of squared deviations is divided by the weight sum W minus one; the square root gives the sample standard deviation, which is unavailable when $W \leq 1$. The number of listed rows, the positive-weight valid row count and W are different quantities; zero-weight rows can appear in the listing but do not contribute to the numerical summaries.

Frequencies... separates percent of all rows from valid percent, and lists user-missing categories separately from system-missing counts. Optional statistics, percentiles and plots use the stated population. Its percentile convention interpolates at $(n + 1)p$, with boundary handling, so it need not equal the type-7 percentile used by some other commands. Do not compare rounded percentiles without matching their definitions.

Explore... combines descriptives, individual mean intervals, a 5% trimmed mean, robust location estimates, extreme cases, normality checks, boxplots and Q–Q displays, optionally by a factor. The trimmed mean removes 5% from each end with fractional weight at the boundary. Boxplot whiskers extend to observed values within 1.5 IQR; points beyond 3 IQR are marked as extremes. These labels flag observations for investigation, not deletion.

The location estimates use a fixed 1.4826 MAD scale and Huber, Tukey, Hampel or Andrews weights. A nonconstant sample with zero MAD cannot supply these estimates. The fitted-normal K–S correction is an approximation; a small normality-test p-value does not by itself identify an appropriate replacement model. Review sample size, the shape and the intended inferential procedure.

25.2 Build a custom table with explicit denominators

Custom tables... permits up to three nested categorical variables on each axis, an optional layer, counts and percentages, and selected numerical summaries. Choose row, column, layer or

table denominators deliberately. Column percentages are taken within the outer row group; a displayed percentage without that context is ambiguous.

Optional column comparisons use two-proportion tests for proportions or Welch tests for numerical means. Holm adjustment is the default over all comparisons in the layer; Bonferroni and unadjusted comparisons are choices. A letter names a column with a significantly smaller value. Shared or absent letters never establish equality or equivalence. Inspect the number of comparisons and expected-count diagnostics alongside any chi-square test.

25.3 Crosstab statistics depend on scale and direction

Crosstab statistics... provides Pearson chi-square and nominal, ordinal and directional association measures. State which variable is predicted by which for asymmetric measures. For text categories, explicitly declare their order before requesting ordinal measures. Alphabetical ordering of High, Low and Medium is a display rule, not a measurement scale.

Nominal measures include phi, Cramér's V, the contingency coefficient, Goodman–Kruskal lambda/tau and uncertainty coefficients. Ordinal measures include gamma, Somers' d and Kendall's τ_b and τ_c. The linear-by-linear statistic uses integer category scores in the declared order. Eta uses between-category over total variation in the direction whose response categories are numeric.

The standard error for a confidence interval describes the alternative distribution. When an ordinal test uses a separate null standard error, the result labels it separately. They are not interchangeable. Lambda is nondifferentiable at tied maxima: its point estimate can remain available while its delta-method standard error and interval are withheld.

For a 2×2 table with cells a, b, c, d, the odds ratio is $ad/(bc)$. For rows $(12, 8)$ and $(6, 14)$ it is 3.5. This compares odds, not risks: the first-column risk ratio is $(12/20)/(6/20) = 2$. The row/column orientation determines the reported direction. Zero cells may make log-scale intervals unavailable. The optional 0.5 correction adds 0.5 to all cells only when explicitly selected.

25.4 Kendall and partial correlation

Statistics > Correlation & regression > Kendall's tau-b... accounts for ties in both variables. Select pairwise or listwise deletion. Exact inference requires no ties; the tie-corrected normal approximation and seeded Monte Carlo permutation procedure have their own stated scope. The latter uses $(1 + \text{exceedances})/(1 + \text{draws})$.

Statistics > Correlation & regression > Partial correlation... correlates the residuals after regressing each response on an intercept and the declared controls, using complete cases. If the controls' numerical rank is k, the usual test has $n - k - 2$ degrees of freedom. Redundant controls do not each consume another degree. A response completely explained by the controls has no residual variation and cannot have a partial correlation. The test's linear, normal-error assumptions and the observational nature of adjustment still matter; it is not a causal effect.

CHAPTER 26

Distribution diagnostics

A distributional assumption should be checked in the context of the analysis that needs it. For a regression model, normality may concern errors conditional on predictors, not the marginal response column. A small p-value identifies evidence against an assumption; a large one does not prove the assumption true.

Open **Statistics > Descriptives & tables > Distribution diagnostics...**. Select a numeric column, the diagnostic and alpha. The initial diagnostic is Shapiro–Wilk and alpha is 0.05; the alpha field accepts whole percentages 1–49. Inspect the plot together with the statistic and the scientific context.

26.1 Normality

Shapiro–Wilk uses 3–5,000 finite observations. Larger samples are refused because the supported calibration does not provide the required p-value accuracy beyond that range. Constant data are unsuitable for either normality diagnostic.

Anderson–Darling compares the fitted normal cumulative distribution with the ordered observations, giving substantial attention to tails. For fitted CDF F,

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}(2i-1)\left[\log F(x_{(i)}) + \log\{1 - F(x_{(n+1-i)})\}\right].$$

HypoNum displays the raw statistic. Its fitted-normal p-value calibration uses $A^2(1 + 0.75/n + 2.25/n^2)$ and a piecewise tail approximation. Mean and variance are estimated from the same observations, so a known-parameter reference distribution would be inappropriate.

26.1.1 Worked example: identify the calibration

Choose **File > Import data...** and select `sample-data/manual/stephens-weights-eleven.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Descriptives & tables > Distribution diagnostics...**. Set **Numeric column** to **value**, **Diagnostic** to **Anderson-Darling normality**, and **Alpha** to 1% (enter 1 in the percentage control). Click **Run diagnostic**. Inspect the raw statistic, p-value and decision, rather than comparing the raw statistic to a cutoff for a differently corrected statistic. Different published Anderson–Darling corrections can give different borderline decisions. For Stephens's eleven-weight Illustration 2, this calibration gives $p = 0.010454$, which does not reject at 0.01. The 1974 paper's modified statistic, 1.095, exceeds its corresponding 1% cutoff, 1.092. Reproduce a reference by matching the statistic, parameter estimators and calibration, not only the test name.

A normal Q–Q plot should be approximately straight under a normal model. Tail curvature can matter even when a small sample gives a nonsignificant test. In a very large sample, a tiny unimportant departure can be detected. Choose an analysis that addresses the consequence of the departure; the diagnostic does not automatically transform the data.

26.2 One suspected outlier: Grubbs

The two-sided Grubbs test evaluates one most-extreme standardized deviation,

$$G = \frac{\max_i |x_i - \bar{x}|}{s},$$

under independent normal observations. It requires at least three nonconstant complete values. The adjustment uses $\alpha/(2n)$ for the two-sided search over candidates. Results show the statistic, critical value, adjusted p-value, candidate value and any ties for the largest deviation.

Choose **File > Import data...** and select `sample-data/manual/grubbs-five.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Descriptives & tables > Distribution diagnostics...**. Set **Numeric column** to **value**, **Diagnostic** to **Grubbs single-outlier**, and **Alpha** to 5%. Click **Run diagnostic** and inspect the candidate value and statistic. For $10, 10, 10, 10, 15$, the mean is 11, sample SD is $\sqrt{5}$, and the candidate value 15 has $G = 4/\sqrt{5} = 1.788854$. This arithmetic identifies the statistic; the test's reference calculation is still required for its decision. A flagged row stays in the source. Investigate recording error, measurement conditions and the analysis design before excluding it. Repeatedly deleting flagged observations and rerunning a single-outlier test changes its error properties.

26.3 A Poisson count model

For Poisson goodness of fit, select nonnegative integer counts. The estimated rate is the sample mean. The procedure pools adjacent integer bins until expected counts reach the chosen threshold, initially 5, and includes an infinite final tail. It then compares observed and expected counts using

$$X^2 = \sum_j \frac{(O_j - E_j)^2}{E_j}, \qquad df = \text{number of pooled bins} - 2.$$

One degree of freedom is used by the total and one by the estimated rate.

This model assumes a common Poisson rate for comparable observations. Different exposures can create variation that a single-rate count model does not represent. A zero fitted rate, insufficient pooled degrees of freedom or counts above the supported maximum 100,000 prevent this calculation. The result explains the reason rather than substituting a test.

26.4 Data use and retained output

Null and non-finite numeric values are counted as missing; exclusions are counted separately. Every retained observation enters the test. Source limits are one million rows and twenty million selected cells. Plots contain at most 400 points and displayed diagnostic tables at most 200 rows, with reductions stated. The saved test metadata retains all pooled bins even if the separate table card is shorter.

Add to results saves the selected method, settings, assumptions, statistic, decision, warnings and bounded plot/table snapshot. Reopening displays that saved result. To assess changed observations, run the diagnostic again and explicitly add the new result.

CHAPTER 27

Probability calculations and random data

Choose **Design > Simulation > Probability distributions...**. Its **Probability calculator** and **Random data** tabs work without a dataset. Calculations use specified parameters; they do not estimate them from an open table.

27.1 Specify the distribution

The initial distribution is Normal with mean 0 and SD 1. Changing family supplies example parameters; replace them with values appropriate to the calculation. Every parameter must be finite.

Family	Parameters	Support
Normal	Mean μ, SD $\sigma > 0$	All real values
Student t	Degrees of freedom $\nu > 0$, location 0 and scale 1	All real values
Chi-square	Degrees of freedom $k > 0$	$x \geq 0$
F	Positive numerator and denominator degrees of freedom	$x \geq 0$
Binomial	Integer trials $1 \leq n \leq 10^9$, probability $0 \leq p \leq 1$	$0, \ldots, n$
Poisson	Mean $0 < \lambda \leq 10^9$	Nonnegative integers
Exponential	Scale $\theta > 0$, equal to its mean; rate is $1/\theta$	$x \geq 0$
Weibull	Shape $k > 0$, scale $\lambda > 0$, two-parameter form	$x \geq 0$

An undefined moment is labelled undefined: Student t, for example, has no mean when $\nu \leq 1$ and no variance when $\nu \leq 2$. The SD is not always the named scale parameter.

27.2 Density, probability and quantiles

Enter comma-separated X values or quantile probabilities, up to 200 of each, and optionally interval bounds. The initial probability is 0.95. Choose **Calculate**.

For continuous distributions, density $f(x)$ is not a probability at that point. It can exceed 1. A discrete mass $P(X = x)$ is a probability and is zero at a noninteger value for the supported count families.

The cumulative column is $P(X \leq x)$; the upper-tail column is $P(X > x)$. Base-10 log probabilities help display extremely small tails even when the ordinary probability rounds or underflows to zero. For counts, $P(X \leq x) = P(X \leq \lfloor x \rfloor)$.

A lower-tail quantile for $0 < p < 1$ is

$$Q(p) = \inf\{x : P(X \leq x) \geq p\}.$$

For a discrete family it is the smallest integer reaching that probability; the achieved probability can exceed p. The upper-tail quantile similarly finds the smallest value whose strict upper-tail probability is no greater than p. Read the achieved probability column to see a discrete step.

27.3 Worked probability calculations

Choose **Design > Simulation > Probability distributions...**; no table is required. Open **Probability calculator**, select **Normal**, enter mean 100 and SD 2, and enter 104 in **Values of x**. Enter 0.95 in **Probabilities for quantiles**, enable **Calculate an interval probability**, set lower bound to 96 and upper bound to 104, and select both endpoint inclusion boxes. Click **Calculate** and inspect the value, interval and quantile tables. The upper-tail probability is about 0.0227501. The interval from 96 to 104 has probability about 0.954500. The 0.95 lower-tail quantile is about 103.2897. These statements concern the specified population distribution, not a confidence interval for a mean estimated from observations.

For the discrete comparison, open **Design > Simulation > Probability distributions...**, choose **Probability calculator** and **Binomial**, enter trials 10 and success probability 0.5. Clear **Values of x** and **Probabilities for quantiles**; enable **Calculate an interval probability**, enter lower bound 2 and upper bound 5. Clear **Include the lower endpoint**, select **Include the upper endpoint**, and click **Calculate**. Then reopen the same setup with those parameters, select the lower endpoint, clear the upper endpoint, and click **Calculate** again. The two interval results are

$$P(2 < X \leq 5) = \frac{582}{1024} = 0.5683594, \qquad P(2 \leq X < 5) = \frac{375}{1024} = 0.3662109.$$

Endpoint choices matter for discrete distributions. Both endpoint-inclusion boxes start selected. For continuous families, including or excluding a single endpoint does not change an interval's probability. A blank bound represents an unbounded side; lower cannot exceed upper. The complement is reported separately.

27.4 Read the plot and retain the calculation

The result shows support, mean, SD, value and quantile tables, and density/mass or cumulative plots. A selected interval is hatched. Plots use 201 points initially and no more than 512. A sparse discrete display or clipped tall density is disclosed, along with the probability represented by the plotted range. Use the numerical table for exact requested values.

Edit X or a probability in the result table and press Enter to recalculate with the same distribution. Clearing a cell removes that row; **Add value** and **Add probability** extend the request. **Add to results** retains the calculation and entered parameters. **Export PNG** saves the plot. The saved calculation is explicitly based on parameters, not a dataset.

27.5 Generate reproducible random values

On **Random data**, enter a table name, row count, comma-separated column names and seed. Defaults are **Random data**, 100 rows, **Value** and a freshly chosen seed; **New seed** changes it.

Up to 20 columns and one million cells are allowed. Names must be distinct ignoring case, no longer than 64 characters, and must not begin with the reserved `__tabelz` prefix. At most 16 dataset tabs can be open.

Generate creates a preview with counts, sample mean, sample SD, minimum and maximum beside theoretical quantities. Changing a setting withdraws the old preview. **Create dataset tab**

adds those previewed values as a new table. Normal and other continuous values use floating point; Binomial and Poisson values are integer counts.

Each column has its own seeded random stream. Adding rows or columns preserves existing generated values under the recorded algorithm. Keep the distribution, parameters, seed and algorithm together. The saved project retains the table and the generation operation. Undo closes the generated tab; redo restores it. Replay compares regenerated values with their retained fingerprint and reports a difference instead of silently accepting changed numbers.

A random sample need not match its theoretical mean or SD exactly. Preview statistics neither qualify the generator nor establish that real measurements follow the same law.

27.6 Simulating inputs through a model

Choose **Design > Simulation > Simulate model...**. Supply a retained linear, binary logistic or Poisson model, or a formula over named inputs. Declare every input once, as a fixed value or a distribution. The calculation propagates those inputs through the model. Fixed fitted parameters are the default. Optional coefficient and fitted-input uncertainty are described below; their interpretation differs from a confidence interval for risk.

A declaration such as `speed~normal(10,2)unitsm/s` assigns mean 10 and SD 2. `temperature=21.5unitsC` fixes an input. `life~weibullfit[Lifehours]` estimates a positive-support Weibull distribution once from the eligible complete column. Normal, exponential, Weibull, Poisson and binomial input fits are supported; other calculator families require supplied parameters. A fitted binomial also needs its number of trials. A categorical model input is fixed at a fitted level. Bracket names containing spaces.

The default target is the conditional mean: fitted linear response, logistic event probability or Poisson expected count. **Outcome draws** additionally simulates normal residuals for OLS, Bernoulli outcomes for logistic models, or Poisson counts. WLS and arbitrary formulas do not supply that outcome distribution. Add an explicit error input to a formula if that is the intended model. Row-order functions such as lag and cumulative sum are inapplicable to independently evaluated draws and are refused.

27.6.1 Dependence, limits and interpretation

Inputs are independent unless continuous-input Spearman targets are supplied, for example `speed,load=0.4`. A Gaussian copula transforms a target s to latent normal correlation $r = 2\sin(\pi s/6)$, then transforms normal probabilities through each input's inverse distribution function. The complete latent matrix must be positive definite; individually valid pairwise correlations can be jointly impossible. The program does not silently repair that matrix. Fixed and discrete inputs cannot participate in these correlation targets. Inspect achieved as well as requested rank correlations; finite simulation draws will not match them exactly.

The default is 10,000 draws with seed 20260928. Draws range from 100 to 1,000,000 subject to the displayed work budget. Retain the input declaration order and seed for reproduction. Undefined or nonfinite formula outcomes cause refusal with a count, rather than silent removal of inconvenient draws. Check the reported share of extrapolated model inputs and the plausibility of the assumed distribution tails.

The result gives target mean, SD, Monte Carlo standard error of the mean, interpolated percentiles and threshold exceedance proportions. Its histogram describes the simulated target under the declared model. Tail percentiles with fewer than ten supporting draws are flagged; those with

fewer than one are withheld. The tornado chart ranks signed input–target Spearman associations, not variance contributions or causal effects.

For independent draws with fixed fitted parameters, and k outcomes strictly greater than a threshold among n draws, $\hat{p} = k/n$. The reported exact binomial interval describes Monte Carlo uncertainty conditional on the simulation model. No exceedances in 10,000 draws give a 95% interval from zero to approximately 0.00036882, not proof of zero risk. This interval does not account for uncertain model coefficients, uncertain input parameters or an incorrect model.

27.6.2 Latin hypercube sampling and approximate rank pairing

The optional replicated Latin hypercube method stratifies each conditional input's innovation probabilities. Its error estimates come from independent replicate summaries; the dependent rows within one replicate do not justify the independent-binomial interval above. Retain replicate count, size and seed.

The explicitly named Iman–Conover normal-score pairing preserves the sampled continuous marginal values and reorders them. It whitens the empirical Pearson correlation of normal scores, then uses the requested rank matrix directly. A direct score target r has limiting achieved Spearman correlation $6\arcsin(r/2)/\pi$, generally different from r. Increasing the number of draws does not remove this systematic discrepancy. Requested and achieved correlations are reported separately. Tied continuous samples and matrices that are not positive definite are refused without jitter or matrix repair. Pairwise correlations and margins do not identify joint tail dependence. The separately sine-converted rank-pairing variant is not offered.

27.6.3 Optional fitted-parameter mixtures

Coefficient uncertainty uses joint normal draws with the full retained covariance. Input-parameter uncertainty uses expected-information normal approximations on log or logit coordinates where appropriate, retaining within-fit covariance. Nonregular boundary fits are refused, not clipped. Residual variance, dispersion and encoding remain fixed. Different fits are treated as independent even if they used overlapping observations.

A new parameter vector is drawn independently for each simulation row. The resulting distribution is a pointwise mixture of fitted-response means, or of future outcomes if that separate option is selected. It is not a confidence distribution for an integrated mean, exceedance probability or trajectory sharing a fitted parameter. Those targets would require an outer parameter draw shared across the relevant inner integrations; that workflow is not supplied. Reported Monte Carlo errors measure numerical precision for the implemented mixture, not parameter-confidence limits for risk. Uncertain fitted margins combined with rank-correlation targets remain unavailable. Under Latin hypercube sampling, only conditional innovations are stratified, not the integrated uncertain mixture margin.

Worked distribution example. Choose **Design > Simulation > Simulate model...**; this formula example needs no input table or stored fit. Set **Model** to **Formula over named inputs**, **Formula** to `3+2*x`, and **Target name** to **Y**. Under **Inputs**, enable **Edit as text** and enter `x~normal(10,2)`. Leave Spearman targets empty, enter **arbitrary units** in **Target units**, set **Draws** to 10000, **Random seed** to 20260928, **Exceedance thresholds** to 27, and **Confidence level** to 0.95. Click **Run analysis** and inspect the mean, standard deviation and exceedance estimate. The theoretical values below are the comparison targets, rather than exact finite-draw outputs. With the stated seed, this edition produces mean 22.984318, sample SD 4.021208, and 1,592 exceedances: estimated probability 0.1592, with Monte Carlo 95% interval approximately $[0.152079, 0.166520]$. The exact model

output has mean 23 and SD 4. Its probability of exceeding 27 is $1 - \Phi(1) = 0.158655$. With 10,000 draws, the mean's Monte Carlo SE is about $4/100 = 0.04$ and the exceedance SE about 0.00365. Seeded results fluctuate around these model values. Increasing draws reduces simulation noise; it does not improve the adequacy of the assumed normal input or the fixed linear relationship.

CHAPTER 28

Known-sigma, variance and event-rate inference

Choose **Statistics > Compare means & ANOVA > Z, variance and Poisson-rate tests...**. These methods answer different questions and require different information. A known measurement SD supports a mean Z test; a sample SD does not become known merely because it can be calculated.

28.1 Choose raw rows or an entered summary

The initial method is **Known-sigma one-sample Z**, input is **Raw table rows**, alternative is two-sided and confidence is 95%. The confidence field accepts 51–99%; the corresponding alpha is one minus confidence. Less and greater alternatives are also available.

Validated summary accepts the necessary numerical summary directly and works without an open table. Its result concerns those entered values. Active filters or exclusions from an unrelated table do not apply. Record where the summary came from and verify that its counts and units match the study.

The null mean starts at 0 and the null rate ratio at 1. Known sigma, null variance, null rate, group labels and summary quantities must be supplied where required. Raw analyses allow one million source rows and twenty million selected cells. Missing values and explicit exclusions are reported separately.

28.2 A mean with known sigma

With sample mean $\bar{x}$, sample size n, known SD $\sigma > 0$ and null mean μ_0, the statistic and two-sided $1 - \alpha$ confidence interval are

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}, \qquad \bar{x} \pm z_{1-\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

The reference is exact for independent normal observations with known sigma; otherwise a justified normal approximation to the mean is needed.

To run the following entered-summary example, choose **Statistics > Compare means & ANOVA > Z, variance and Poisson-rate tests...**. Set **Method** to **Known-sigma one-sample Z**, **Input** to **Validated summary**, **Alternative** to **two-sided**, and **Confidence (%)** to 95. Enter mean 10.2, sample size 25, known sigma 0.5 and null mean 10, then click **Run test**. No table is required. For this summary with $\bar{x} = 10.2$, $n = 25$, $\sigma = 0.5$ and $\mu_0 = 10$, $SE = 0.1$, $z = 2$, the two-sided p-value is about 0.04550 and the 95% interval is $[10.0040, 10.3960]$. Use a one-sample t analysis when the population SD must instead be estimated from the sample.

28.3 A single population variance

For independent normal observations, test a positive null variance σ_0^2 using

$$X^2 = \frac{(n-1)s^2}{\sigma_0^2}, \qquad df = n - 1.$$

Raw observations require at least two finite values. Summary mode requires a nonnegative sample variance and sample size at least two. A two-sided interval for the population variance is

$$\left[\frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}\right].$$

The interval is generally asymmetric. Selecting **Standard deviation** takes the square root of the estimate and both endpoints; it does not change the hypothesis test. One-sided intervals have an unbounded endpoint where appropriate.

To run the variance example, choose **Statistics > Compare means & ANOVA > Z, variance and Poisson-rate tests...**. Set **Method** to **One-variance chi-square**, **Input** to **Validated summary**, alternative to **two-sided**, confidence to 95%, and **Display** to **Variance**. Enter variance 4, sample size 10 and null variance 2; click **Run test**. No table is required. With $n = 10$, $s^2 = 4$, $\sigma_0^2 = 2$, the statistic is 18 on 9 degrees of freedom. The variance is in squared measurement units; its square root is in the response's original units. This variance test is sensitive to nonnormality.

28.4 One Poisson event rate

Select nonnegative integer event counts and positive exposures. For total count C and total exposure E,

$$\hat{\lambda} = C/E.$$

Under independent Poisson counts with a common rate per exposure unit, Garwood limits provide exact count-based uncertainty. For two-sided confidence $1 - \alpha$,

$$\lambda_L = \begin{cases} 0, & C = 0, \\ \chi^2_{\alpha/2,2C}/(2E), & C > 0, \end{cases} \qquad \lambda_U = \chi^2_{1-\alpha/2,2(C+1)}/(2E).$$

Zero observed events therefore do not establish a zero underlying rate.

To reproduce the zero-event example without a table, choose **Statistics > Compare means & ANOVA > Z, variance and Poisson-rate tests...**. Set **Method** to **One Poisson rate**, **Input** to **Validated summary**, alternative to **two-sided**, and confidence to 95%. Enter count 0, exposure 100, null rate 0.01, and exposure units **hours**. Click **Run test** and inspect the rate interval. Zero events in 100 hours gives a 95% two-sided upper limit $-\log(0.025)/100 = 0.036889$ events per hour. A 95% one-sided upper limit instead uses $-\log(0.05)/100 = 0.029957$. To obtain that one-sided upper limit, reopen the same menu, repeat the entered summary and settings, choose alternative **less**, and click **Run test**. State which interval is being reported.

The one-rate test requires a positive null rate. Its two-sided p-value doubles the smaller inclusive Poisson tail, capped at 1; one-sided alternatives use the appropriate tail. This test convention and the central interval construction are distinct.

28.5 Compare two Poisson rates

Assign two explicit groups, or enter two count/exposure summaries. The estimated ratio is $R = (C_1/E_1)/(C_2/E_2)$. For null ratio R_0, condition on the total count:

$$C_1 \mid (C_1 + C_2) \sim \text{Binomial}\left(C_1 + C_2,\ \frac{R_0 E_1}{R_0 E_1 + E_2}\right).$$

The two-sided exact test sums binomial outcomes with probability no greater than the observed outcome's probability. Clopper–Pearson binomial limits transform to ratio limits through $pE_2/\{(1-p)E_1\}$. These central limits need not invert that probability-ordered two-sided test.

If group 2 has zero events, the observed ratio is unbounded; two zero counts provide no estimable comparison. For an executable two-rate example, open **Statistics > Compare means & ANOVA > Z, variance and Poisson-rate tests...**. Choose **Two Poisson rates**, **Validated summary**, **two-sided**, confidence 95%, null ratio 1 and exposure units **hours**. Enter Count 1 = 12, Exposure 1 = 300, Count 2 = 5, Exposure 2 = 250, then click **Run test**. No table is required. The component estimates are 0.04 and 0.02 per hour, giving estimated ratio 2. That point estimate alone does not establish a difference in population rates.

28.6 Units, limits and saved results

Exposure labels such as hours or days are descriptive; changing a label does not convert the values. Converting all hours to days multiplies rates and their limits by 24. Convert a one-rate null value too. A common conversion leaves a two-rate ratio unchanged.

Each count and the aggregate across both groups must not exceed $2^{53} - 1$; invalid integers, nonpositive exposure and non-finite totals are refused. Results retain the method, alternative, null, units, settings, uncertainty convention and population accounting. **Add to results** saves a snapshot; a summary snapshot retains the entered numbers and does not invent an underlying table population.

CHAPTER 29

Blocked ranks and ordered runs

Friedman's test compares treatments within complete independent blocks. A runs test examines ordering in a binary sequence. Their designs and null hypotheses differ. Choose **Statistics > Compare means & ANOVA > Friedman and runs tests...**.

29.1 Friedman: one observation per treatment and block

Assign a numeric response, treatment and block. The initial mode is Friedman with asymptotic inference. At least two complete blocks and three treatments are required. Each block must have exactly one finite response for every treatment. Missing cells, duplicates and incomplete blocks stop the analysis; the workflow does not silently remove incomplete blocks.

Rank the responses within each block, averaging tied ranks. For n blocks, k treatments and treatment rank sums R_j,

$$Q = \frac{12}{nk(k+1)} \sum_j R_j^2 - 3n(k+1).$$

With ties, divide this by $1 - \sum(t^3 - t)/\{n(k^3 - k)\}$, where t is a within-block tie-group size. An all-tied unestimable design is refused. The reference is an asymptotic chi-square distribution with $k - 1$ degrees of freedom; small samples receive a caution.

29.1.1 A transparent rank calculation

Choose **File > Import data...** and select `sample-data/manual/friedman-nine.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Compare means & ANOVA > Friedman and runs tests...**. Set **Test** to **Friedman complete-block**, **Response** to **response**, **Treatment** to **treatment**, and **Block** to **block**. Keep the source-row limit at least 9 and the block limit at least 3. Click **Run test**; inspect the treatment rank sums and the asymptotic chi-square result. Consider three blocks with treatment A/B/C values

	A	B	C
1	1	2	3
2	2	1	3
3	1	3	2

These already equal their within-block ranks. Rank sums are 4, 6 and 8, giving $Q = 8/3 = 2.666667$, $df = 2$ and asymptotic $p = 0.26360$. The example explains arithmetic; three blocks provide a weak basis for the large-sample approximation.

Blocks should be independent and within-block treatment observations comparable. Ranking does not turn nonrandom treatment allocation into a causal study.

29.2 Runs: preserve the observation order

Choose a categorical sequence with exactly two observed categories, or leave the selector at **(Numeric response)** and choose a numeric response. Numeric signs are defined relative to the observed median by default, or a specified finite reference. Values exactly at the reference are omitted and counted.

A run is a maximal uninterrupted stretch of one category. To run the six-observation example:

Choose **File > Import data...** and select `sample-data/manual/runs-six.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Compare means & ANOVA > Friedman and runs tests...**. Set **Test** to **Runs sequence**, **Categorical sequence (optional)** to **sequence**, **Order** to **(Source row order)**, **Inference** to **Exact**, and **Alternative** to **Two-sided**. Keep the exact-sequence limit at 200 and source-row limit at least 6. Click **Run test** and inspect the observed and expected run counts. The sequence A, A, B, B, A, B has four runs: AA, BB, A and B. For category counts n_1, n_2, $n = n_1 + n_2$,

$$E(R) = 1 + \frac{2n_1n_2}{n}, \qquad \mathrm{Var}(R) = \frac{2n_1n_2(2n_1n_2 - n)}{n^2(n-1)}.$$

With three observations in each category, the expected number is 4 and variance is 1.2. Too few runs can suggest clustering; too many can suggest alternation.

Source row order is the default. An optional order column uses its native order with source-order ties. Never sort the response to prepare the test.

Missing or non-finite retained values stop the calculation: dropping one would join its neighbours and change the sequence. An intentional omission must be an explicit row exclusion; the result then concerns that retained sequence. At least one observation of each category must remain.

29.3 Exact and approximate inference

Exact inference enumerates runs conditional on category counts, up to 200 observations. The two-sided convention sums run-count outcomes whose probabilities are no greater than the observed outcome's probability. Fewer-runs and more-runs alternatives use their respective tails.

Asymptotic inference standardizes by the mean and variance above, without continuity correction. It can be inaccurate for short or highly unbalanced sequences. The null treats category ordering as exchangeable conditional on counts; the declared observation order must have a substantive meaning.

29.4 Record population and limits

The source-row field allows up to one million rows and selected input cells are bounded at twenty million. The block limit is at most 10,000; the exact-runs limit ranges from 3 to 200.

Results retain counts, method convention, order, reference, tie handling, statistics and warnings. Rank-summary tables show at most 200 rows, with any reduction stated; calculations use the complete eligible design. **Add to results** retains the displayed snapshot. A source change requires a new run to record updated evidence.

CHAPTER 30

Permutation tests

A permutation test compares an observed statistic with the values obtained by reassigning labels, or by reversing paired differences, under a specified null symmetry. The reference distribution comes from those assignments. It does not require a normal model for the response, but it does require the appropriate exchangeability or symmetry.

Choose **Statistics > Compare means & ANOVA > Permutation tests...**. The initial design is independent two-sample, alternative two-sided, inference Monte Carlo, 2,000 reassignments and seed 0.

30.1 Independent and paired designs

Independent mode needs a numeric response, group column and two exact observed group labels. Its statistic is

$$T = \bar{x}_A - \bar{x}_B.$$

Labels are reassigned over the pooled responses with group sizes fixed. Under the null, the observations must be exchangeable across those labels. Equal means alone, with arbitrary unequal distributions or variances, does not supply that condition.

Paired mode needs two distinct numeric columns aligned within rows. Its statistic is the average first-minus-second difference, $T = \bar{d}$. Reassignment reverses signs of whole differences; pairs are never broken or rematched. The usual sign-flip null requires symmetric differences, or a design that justifies the corresponding randomization.

Missing or non-finite values omit a whole pair in paired mode. Other group labels stay outside the independent comparison and are counted. At least two eligible observations per group, or two pairs, are required. Clustered, stratified and weighted randomization designs are not offered.

30.2 Worked example: enumerate every assignment

Choose **File > Import data...** and select `sample-data/manual/permutation-four.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Compare means & ANOVA > Permutation tests...**. Set **Design** to **Independent two-sample**, **Response** to **value**, **Group column** to **group**, **Group A label** to **A**, and **Group B label** to **B**. Choose **Two-sided** for **Alternative**, **Exact** for **Inference**, and seed 0. Click **Run test**. Expect statistic -3, six assignments and $p = 0.333333$. The following enumeration explains that result.

For groups $A = (1, 2)$ and $B = (4, 5)$, $T_{\text{obs}} = 1.5 - 4.5 = -3$. There are $\binom{4}{2} = 6$ possible choices of the two observations labelled A. Their mean differences are

$$-3,\ -1,\ 0,\ 0,\ 1,\ 3.$$

Two are at least as extreme in absolute value as the observed difference. Thus the exact two-sided p-value is $2/6 = 1/3$. The minimum possible two-sided value here is large because the randomization set is so small.

Exact independent inference enumerates $\binom{n}{n_A}$ assignments; paired inference enumerates 2^n sign patterns. The limit is 20,000 assignments. Above it, choose Monte Carlo explicitly. Tails include ties, with relative numerical tolerance 10^{-12} around the observed magnitude. Two-sided tails use $|T| \geq |T_{\text{obs}}|$; less and greater alternatives use their respective directions.

30.3 Monte Carlo inference

Monte Carlo mode samples assignments with the recorded seed. If r of m sampled statistics are at least as extreme,

$$\hat{p} = \frac{r+1}{m+1}.$$

The plus-one convention prevents a reported value of zero. With the initial 2,000 draws, the smallest possible value is 1/2001, about 0.000500. Changing the seed can change the simulated result; increasing draws reduces Monte Carlo variability.

The draw count ranges from 2 to 10,000 and the desktop seed from 0 to 2,147,483,647. The histogram uses 40 bins initially, with range 1–64. It displays a null reassignment distribution, not a bootstrap sampling distribution. A bootstrap resamples observations with replacement for a different purpose.

30.4 Read and save

Results show the statistic, alternative, exact or Monte Carlo convention, assignment count, seed, p-value, group summaries and the histogram with its observed-statistic marker. There are no parametric degrees of freedom. Read the eligible, missing, excluded and outside-group counts together.

Limits are one million source rows and twenty million selected cells. Cancellation prevents publication of a partial simulated result. **Add to results** records the settings, population and result snapshot; source changes do not alter an already saved result.

CHAPTER 31

Comparisons with a control and between unequal groups

Multiple comparisons form a family of claims. Choosing that family determines the adjustment and the meaning of simultaneous confidence intervals. Use Dunnett when the planned comparisons are treatments against one control. Use Games–Howell for all pairs when equal variances and equal sample sizes are not assumed.

Open **Statistics > Compare means & ANOVA > Dunnett and Games-Howell comparisons...**. Assign numeric response and group, then select the comparison family. Dunnett and alpha 0.05 are the initial choices. Alpha is entered as a whole percentage from 1 to 49. Dunnett additionally needs an explicit control label.

31.1 Dunnett: several treatments against one control

The independent-normal model uses a common variance. Each contrast is a treatment mean minus the declared control mean, with direction named in the table. Dunnett's single-step adjustment accounts for the dependence induced by reusing the same control group.

If the groups are Control, Low dose and High dose, the requested family has two comparisons: Low minus Control and High minus Control. It does not also test High minus Low. Changing the control or restricting groups changes the family. Declare these choices before examining outcomes.

Results include differences, pooled standard errors, adjusted p-values and simultaneous confidence intervals. The numerical integration seed starts at 12345 and ranges from 0 to 2,147,483,647. Retain it for reproducing the numerical calculation. Integration or optimization warnings remain part of the result; the software does not substitute Tukey's method.

31.2 Games–Howell: all pairs with unequal variances

For groups i, j, define

$$SE_{ij} = \sqrt{s_i^2/n_i + s_j^2/n_j}, \qquad \nu_{ij} = \frac{(s_i^2/n_i + s_j^2/n_j)^2}{(s_i^2/n_i)^2/(n_i - 1) + (s_j^2/n_j)^2/(n_j - 1)}.$$

Games–Howell combines these pair-specific quantities with the studentized-range distribution for the all-pairs family. Its intervals and p-values are adjusted for that family.

For illustrative group summaries $\bar{y}_A = 10, s_A = 2, n_A = 10$ and $\bar{y}_B = 12, s_B = 3, n_B = 15$, the A-minus-B difference is -2 and $SE_{AB} = \sqrt{0.4 + 0.6} = 1$. The degrees of freedom are about 22.9927. The corresponding studentized-range statistic is $\sqrt{2}\,|-2|/1 = 2.8284$. The family size is still needed to obtain the adjusted inference; a simple unadjusted t interval is not its substitute.

At least two observations per eligible group are needed. Two zero-variance groups do not supply an estimable Welch comparison and are refused. One zero-variance group can be compared

when the other supplies a positive finite variance contribution. Approximate inference still depends on suitable independent within-group observations.

31.3 Specify eligible groups and inspect the results

Leave **Eligible groups** empty to use every observed nonempty level, or enter exact labels separated by semicolons. This entry format cannot represent a label containing a semicolon. A requested group with no complete observations is an error. Missing response/group values and explicit exclusions are counted.

The limit is one million source rows, 32 eligible groups and twenty million projected cells. Results and saved contrast tables display at most 200 contrasts, with any bound stated. The plot displays estimates and contrast labels, without confidence whiskers; read interval endpoints in the table.

Add to results retains the selected family, control where relevant, groups, alpha, confidence, seed, assumptions and bounded output. A saved project restores the snapshot; another calculation is explicit. Report the contrast direction, magnitude, interval and family adjustment together rather than presenting adjusted p-values alone.

CHAPTER 32

One-sample and paired equivalence

The two-one-sided-test principle also applies to a difference from a target or to within-row paired differences. Choose **Statistics > Compare means & ANOVA > Equivalence tests...** and change **Test type** to **One-sample difference**, **Paired difference** or **Paired log-ratio**.

32.1 Choose the contrast and margins

Setup initially uses independent samples, alpha 0.05 and bounds $-1, 1$. Replace these generic bounds with margins justified for the study. Alpha must be strictly between 0 and 0.5.

One-sample mode analyzes $d_i = x_i - \mu_0$; the finite target initially equals zero. Paired difference uses first minus second measurement within the same source row. Paired log-ratio uses

$$d_i = \log x_{i,1} - \log x_{i,2},$$

requiring positive measurements and positive ratio bounds. Columns are never independently sorted or re-paired.

Missing or non-finite values omit the whole pair. A finite zero or negative value is an error in log-ratio mode. At least two contrasts with positive sample variation are required; zero standard error is refused. Limits are one million rows and twenty million selected cells.

32.2 Worked example: paired differences

Choose **File > Import data...** and select `sample-data/manual/paired-measurements.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Compare means & ANOVA > Equivalence tests...**. Select **Paired difference**, first measurement `x`, second `y`, alpha 0.05 and bounds 0.4 and 0.8. Click **Run analysis**. Inspect the contrast estimate, its 90% interval, and both one-sided p-values; all sixteen pairs should be analyzed. These margins are teaching assumptions, not a recommended measurement-agreement standard.

The sixteen differences give

$$\bar{d} = 0.5625, \qquad SE = 0.04069705, \qquad df = 15.$$

The 90% interval is $[0.4911560, 0.6338440]$. The lower-bound test has $p = 0.00058807$, and the upper-bound test has $p = 0.00001639$. Both pass the one-sided threshold and the interval is strictly within $(0.4, 0.8)$, so the declared equivalence condition is established.

This does not establish interchangeability around a zero difference. The acceptable region explicitly allowed a positive offset. Margins must express the scientific question.

32.3 The matching interval

For a one-sample or paired contrast,

$$SE = s_d/\sqrt{n}, \qquad t_L = (\bar{d} - L)/SE, \quad t_U = (\bar{d} - U)/SE, \quad df = n - 1.$$

The lower-bound test uses the upper t tail and the upper-bound test uses the lower tail. The interval is

$$\bar{d} \pm t_{1-\alpha,n-1}SE.$$

Its confidence is $1 - 2\alpha$, not $1 - \alpha$. The calculation assumes independent observations or independent paired contrasts and an approximately normal contrast population.

Log-ratio bounds become $\log L, \log U$. Exponentiating the estimated mean log difference gives the geometric mean ratio; exponentiating its interval gives ratio limits. Ratios 0.8 and 1.25 have geometric mean 1, although their arithmetic mean is 1.025. State which quantity is estimated. This paired calculation does not fit a crossover-study model.

32.4 Report the declared decision

Results include mode, scale, estimate, standard error, degrees of freedom, both p-values, interval, bounds, decision and row counts. Retain these together. Reporting only *equivalent* conceals the margins and precision of the evidence.

Add to results stores the current settings and result. Source or exclusion changes require a new calculation before recording. Reopening restores the snapshot without refitting. Cancellation or closing setup during work prevents recording a partial result.

CHAPTER 33

Repeated-measures ANOVA and comparison families

Statistics > Compare means & ANOVA > Repeated-measures ANOVA... uses one row per subject and one numeric column per within-subject condition. Define one or two within factors and map every level combination to its measurement column. The first factor varies slowest in the cell grid. The limits are 20 levels per within factor and 60 cells overall. Optional between-subject factors and covariates describe independent subjects.

33.1 Define the population and the reference values

A subject missing any required measurement, factor or covariate is omitted. For an unqualified complete-case interpretation, missingness must not select subjects according to their responses. The long-form residual-covariance model can address different incomplete-data assumptions; it is a separate choice and is not substituted automatically.

Covariates default to their means for within-effect tests. Testing at zero is an explicit alternative whose meaning depends on the covariate's units and origin. Follow-up simple effects and interaction contrasts use the parent model's recorded reference: zero for uncentered covariates, otherwise their means. The result states the actual reference values and averaging weights. Between-cell averaging is equal by default; observed-size averaging targets the observed group composition instead. These are different scientific estimands, not alternative numerical approximations to one target.

The fitted model is $Y = XB + E$, with subjects in rows and conditions in columns. Between factors use sum-to-zero coding and their factorial interactions; the covariates share slopes across between groups. The residual degrees of freedom are $\nu = N - \operatorname{rank}(X)$. An unidentified design, including required empty between cells, is refused rather than silently dropping terms.

33.2 Read both within-subject inferential approaches

For an effect with q within contrasts and between-hypothesis rank r, the univariate statistic compares hypothesis and residual sums of squares:

$$F = \frac{\operatorname{tr}(H)/(rq)}{\operatorname{tr}(E_W)/(\nu q)}.$$

Sphericity requires equal variance in orthogonal within-contrast directions. For eigenvalues λ_j of the residual contrast covariance, Greenhouse–Geisser epsilon is $\hat{\epsilon} = (\sum_j \lambda_j)^2/(q \sum_j \lambda_j^2)$. Corrected tests multiply both degrees of freedom by their epsilon and leave F unchanged. The result also shows the Huynh–Feldt/Lecoutre and lower-bound corrections. Mauchly's test uses its stated first-order approximation; failing to reject it in a small sample is weak evidence. With only one contrast ($q = 1$), sphericity is automatic and epsilon is one.

Pillai, Wilks, Hotelling–Lawley and Roy tests provide the multivariate approach. Singular residual covariance can make them unavailable while the univariate calculations remain available.

Roy's multi-dimensional F approximation is an upper bound, with a corresponding lower-bound p-value. Do not choose whichever approach gives the smallest p-value.

Worked example

Choose **File > Import data...** and select `sample-data/manual/repeated-five.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Compare means & ANOVA > Repeated-measures ANOVA...**. Enter **Time** as **First within factor** and `Before,After` in **Its levels, in order**. Leave quantitative level values and the second within factor empty. In **Column for each cell**, map Before to **before** and After to **after**. Leave **Between-subject factors** and **Covariates** empty, set **Alpha** to 0.05, and leave simple effects and interaction contrasts off. Keep marginal-mean pairwise adjustment at **Bonferroni**. Click **Run analysis**; inspect the Time row in **Within-subject tests (univariate)**. For these five subjects, before values $(10, 11, 12, 13, 14)$ and after values $(11, 13, 13, 16, 16)$ give differences $(1, 2, 1, 3, 2)$. The mean increase is 1.8 and its standard error is $\sqrt{0.7/5} = 0.37417$. With one two-level within factor and no between factors, the within test is the paired-t test squared: $F = 23.14286$ on $(1, 4)$ degrees of freedom. There is no sphericity correction to estimate for two levels. The repeated-measures design, not the count of ten recorded cells, determines the independent sample size.

33.3 Contrasts, simple effects and interaction contrasts

Marginal-mean comparisons use the fitted joint covariance, never an independent-groups standard error for paired conditions. The available adjustments here are Bonferroni, Sidak and Holm, or no adjustment. Mean intervals are individual intervals; Holm supplies adjusted p-values without automatically producing simultaneous intervals.

Polynomial contrasts require declared numeric level values and their spacing. Otherwise choose an ordered named family such as Helmert, repeated or simple contrasts. A label such as Visit 1, Visit 2, Visit 3 does not establish equally spaced measurement times.

Optional simple effects test one factor at each level of another using the parent model and its residual covariance. Optional interaction contrasts test crossed contrast vectors, including within-by-within and between-by-within combinations. They are not separate fits of each slice. The result identifies the family and adjustment. Single-level conditioning is supported; more elaborate conditioning on multi-factor cells is not. Automatic uses polynomial contrasts for ordered factors with numeric level values and a recorded Helmert basis otherwise, including unordered between factors. Explicit Polynomial without level scores is unavailable; another family is not substituted. An unavailable requested contrast retains its row, reason and place in the declared adjustment family.

33.4 Independent-group post hoc procedures

Statistics > Compare means & ANOVA > One-way ANOVA with post hoc and contrasts... offers independent-group procedures and planned contrast families. Choose a method whose variance and comparison assumptions match the design.

Choice	Appropriate interpretation
LSD	Individual pooled-error comparisons; no general familywise-error protection.
Bonferroni, Sidak, Holm	Declared comparison family; Sidak needs its dependence assumptions. Holm is a step-down adjustment of p-values.
Tukey/Tukey–Kramer	All pairwise comparisons with a pooled error term.
Dunnett	Multiple groups compared with one stated control.
Scheffé	Protection over the applicable contrast space, not only the displayed pairs.
Games–Howell, Tamhane T2, Dunnett T3/C	Independent groups without equal variances; Dunnett C supplies intervals without adjusted p-values.
SNK, Duncan, REGW	Range/subset procedures. SNK and Duncan lack strong familywise-error control. Subsets do not establish equivalence.

Custom contrast coefficients must be finite, nonzero as a vector and sum to zero. Algebraic orthogonality of coefficient vectors differs from uncorrelated contrast estimates when group sizes or variances differ. The output reports the relevant distinction.

For supported maximum-t calculations, numerical integration reports estimated precision and the finite computation budget. A decision too close to its numerical uncertainty is labelled unresolved. This uncertainty describes the probability calculation, separately from sampling uncertainty. The wider library's comparison options do not imply that every one is available in the repeated-measures or adjusted-means windows.

33.5 Analysis of means

Choose **Statistics > Compare means & ANOVA > Analysis of means (ANOM)...** to compare every group with the overall centre using simultaneous decision limits. A group outside its limit identifies a departure from that centre, not a pairwise difference from every other group. Alpha is 0.05 by default; 0.01 and 0.10 are available.

For normal one-factor data with independent equal-variance errors, k groups of sizes $n_i \geq 2$ share pooled SD s_p and weighted grand mean $\bar{\bar{y}}$. The limits are

$$\bar{\bar{y}} \pm h s_p \sqrt{\frac{N - n_i}{N n_i}}.$$

The joint critical value depends on the actual group sizes and residual degrees of freedom. Reusing a balanced-design constant for unequal groups does not preserve the simultaneous interpretation. Numerical root tolerance is not an error bound for the probability calculation beneath it.

With group observations $m - 1, m - 0.5, m, m + 0.5, m + 1$ and means $m = (10, 10, 10, 12)$, the centre is 10.5 and $s_p = \sqrt{0.625}$. At 0.05, $h \simeq 2.74337$ gives common limits 9.6600 and 11.3400; only the fourth mean is outside. These are implemented critical-value calculations, not a claim of agreement with an unavailable printed table.

Two-factor ANOM requires a complete balanced replicated design. It shows factor A, factor B and interaction charts; alpha applies separately to each chart, not jointly across all three. Interaction

critical values can use seeded numerical integration with an estimated numerical standard error. Values close enough to a boundary to be affected by that numerical uncertainty are marked. Layouts exceeding the supported interaction rank are refused.

Binomial ANOM pools successes and trials and compares group proportions with normal-approximation limits. Poisson ANOM pools counts and exposure and compares rates in the declared exposure unit. Each group needs adequate expected counts: at least five expected successes and failures for binomial data or five expected events for Poisson. These guards are minimum conditions, not guarantees of nominal small-sample error. Zero pooled variation or rate, inadequate groups and unsupported unbalanced two-factor designs are refused with the reason.

CHAPTER 34

Distribution, ordered and paired categorical tests

The **Statistics > Compare means & ANOVA > Hypothesis tests...** window groups tests into means/-variances/proportions and one-sample, independent-sample and related-sample nonparametric families. Select the family before the test; the visible fields follow that choice. Included complete observations and any declared integer frequency weights define the analysis population. A weight represents replicated independent observations, not survey-design information.

Every result identifies how its p-value was obtained: exact enumeration or formula, a named asymptotic approximation, Monte Carlo, or withheld with a reason. The seeded simulation default is 9,999 draws with seed 0. A plus-one Monte Carlo value is $(1+b)/(1+B)$ for b simulated statistics at least as extreme as observed among B draws. It cannot be smaller than $1/(B+1)$. Simulation error and sampling uncertainty are different quantities.

34.1 Distributional hypotheses

The one-sample binomial test can define success by a category or by a numeric cut point: values at or below the cut point are successes. Specify the null probability and alternative. The exact two-sided test sums null probabilities no larger than that of the observed count; it is not always twice one tail. For three successes in three trials against probability 0.5, that two-sided p-value is 0.25.

The one-sample Kolmogorov–Smirnov statistic compares the empirical and null distribution functions: $D = \sup_x |F_n(x) - F_0(x)|$. A fully specified continuous null permits the classical Kolmogorov distribution. If normal, uniform or exponential parameters are estimated, their uncertainty changes the null distribution: the program fits, simulates and refits each bootstrap sample. It does not use the classical known-parameter table for an estimated null. A Poisson null is discrete and is simulated even when its mean is supplied. Ties against a continuous null are disclosed.

Two-sample KS compares $\sup_x |F_1(x) - F_2(x)|$. Untied observations use the continuous-null exact route when feasible, otherwise the named approximation. Tied observations use a conditional reassignment distribution, exact when the lattice calculation is feasible and Monte Carlo otherwise. The result can detect differences beyond location; a rejection is not specifically a test of medians.

Moses' test examines whether an experimental group extends beyond both ends of a declared control group. Inspect the control choice and trimmed and untrimmed rank spans. The conventional span distribution requires untied pooled observations. With ties within either group or across groups, the displayed spans use average ranks and the p-value is withheld; it is not reported as an approximate test. The two-sample Wald–Wolfowitz test counts group runs in pooled value order. With ties shared across groups, ordering can change the run count. The program reports the possible range and withholds its p-value.

34.2 Ordered groups and rank comparisons

Jonckheere–Terpstra uses a declared group order. Its statistic

$$J = \sum_{a<b} \sum_{x\in a, y\in b} \left(\mathbf{1}\{x < y\} + \tfrac{1}{2}\mathbf{1}\{x = y\} \right)$$

counts evidence for increasing outcomes along that order. The default uses the tie-corrected asymptotic variance; the permutation option enumerates when feasible and otherwise samples with the recorded seed. Reversing the declared order reverses a directional interpretation. Do not choose the order after examining outcomes and interpret the resulting test as planned.

Kruskal–Wallis can be followed by Dunn comparisons using pooled mean ranks, the pooled tie correction and Holm adjustment over all pairs by default. This is different from independently ranking each pair. A median test counts values above versus at or below the pooled median. Yates correction is available only for a two-group table.

For related measurements, select one column per condition and one row per independent block. Incomplete blocks are omitted. Friedman ranks within blocks, applies its tie correction and reports Kendall's concordance W. Conover complete-block comparisons use the within-block rank variation and Holm adjustment; Nemenyi is a different simultaneous comparison option. If every block ranks treatments identically, a comparison denominator can vanish and is withheld. Kendall's W measures concordance, not agreement on the original measurement scale.

34.3 Paired and stratified categories

For McNemar, let b, c be the two discordant cells in a paired binary table. The default exact test conditions on $b + c$ and tests a binomial probability of one half. With $(b, c) = (1, 7)$, the two-sided exact value is $2(\binom{8}{0} + \binom{8}{1})/2^8 = 0.0703125$. Concordant pairs affect the paired risk difference $(b - c)/n$ but not this conditional test. The optional chi-square approximation and continuity correction are separately labelled. The approximate risk-difference interval is not an inversion of the exact conditional test.

Stuart–Maxwell and Bhapkar test marginal homogeneity for paired tables with more than two categories. Unconnected category groups reduce the effective degrees of freedom; a table without discordance provides no evidence against homogeneity. Cochran's Q extends the binary related-sample comparison to several columns. Pairwise follow-ups use exact McNemar tests with adjustment over the declared family.

Cochran–Mantel–Haenszel conditions on strata. For 2×2 strata it reports the common odds ratio, its interval and the chosen continuity correction. Zero or infinite estimates do not receive an invented finite interval or an automatic half-cell correction. Inspect the heterogeneity diagnostic before describing one odds ratio as common. General rectangular tables use a covariance-rank test and do not acquire a binary odds ratio.

Cochran–Armitage trend requires ordered groups and explicit scores. The scores' direction defines the alternative. Exact enumeration is used when feasible; larger exact requests use the labelled Monte Carlo route. Its unconditional and fixed-margin variance choices differ. For three groups of two subjects with event counts $(0, 1, 2)$ and scores $(1, 2, 3)$, the conditional two-sided exact value is 0.20. The ordering, sample allocation and conditioning are part of that result.

34.4 A variance confidence set can be disconnected

The Bonett–Banga–Fox variance-ratio procedure allows nonnormal populations with finite relevant moments and adequate positive-variance samples. It is an approximation, not an exact

normal-theory F interval. Inverting its test can produce several disjoint accepted ranges. Read every reported component of the confidence set; filling the gap between components changes the set. An unresolved numerical boundary is reported as unresolved rather than silently joined or discarded. As with other variance methods, heavy tails and a few influential observations warrant sensitivity analysis.

CHAPTER 35

Forests, penalized classification and conditional quantiles

These procedures answer different questions. A regression forest predicts a numeric response flexibly; penalized logistic regression predicts an event probability; quantile regression describes a chosen part of the conditional response distribution. Select the target before selecting the algorithm. The random training/test splits described here require independent rows. Repeated subjects, batches and time order require a different validation design.

35.1 A regression random forest

Choose **Statistics > Multivariate > Regression random forest...**. Select one numeric response, numeric and/or categorical predictors, declare independent observations and set the seed (default 0). Missing responses are omitted and counted. Missing numeric predictors use fitting-row medians; categorical missingness is a separate level. Infinite predictors are refused. There must be at least 25 usable responses: at least 20 training and five test observations. The current forest limit is 20,000 response-bearing rows.

The forest averages 500 regression trees, each grown on a bootstrap sample. At a split, a random subset of encoded predictors is considered. This reduces dependence between trees, while averaging reduces the variability of an individual tree. It does not provide a causal interpretation or an automatic uncertainty interval.

35.1.1 Keep tuning separate from evaluation

A seeded split first reserves 20% of the usable rows, rounded up. Five-fold cross-validation on the remaining rows selects among 12 combinations: minimum leaf sizes 1, 3 and 5; candidate-feature fractions 0.5 and 1; and depth 10 or unlimited. Preprocessing is fitted within each training fold. Categories absent from a fitted sample cannot be scored as an invented reference category.

The selection score is pooled out-of-fold root mean squared error,

$$\mathrm{RMSE}_{\mathrm{CV}} = \sqrt{\frac{\sum_i (y_i - \hat{y}_i^{(-k(i))})^2}{n_{\mathrm{train}}}}.$$

It is not the unweighted mean of fold RMSEs. If two folds have squared-error totals 1 and 18 over one and two observations, the pooled score is $\sqrt{19/3} = 2.5166$, whereas averaging their RMSEs gives 2. Near-ties prefer larger leaves, then shallower depth, then the smaller feature fraction. A candidate with a failed fold cannot win.

The selected settings are fitted on all training rows and evaluated once on the reserved test rows. Compare test RMSE, MAE and R^2 with the training-mean baseline. Revising settings after seeing test performance uses up that test set's role as an untouched evaluation.

35.1.2 Interpret importance and score columns

Permutation importance reports the change in pooled validation RMSE when a predictor is shuffled within each validation fold. A categorical predictor is shuffled as one variable. The mean and sample standard deviation across ten permutations describe the permutation calculation; the SD is not a population confidence interval. Correlated predictors can share or conceal importance, and negative importance is retained.

Add evaluated-fit scores applies the training-only forest. Its training predictions are in-sample; only its reserved test predictions were held out. **Add deployment-refit scores** applies a separate fit on all usable response rows with the selected settings. That refit has no held-out evaluation. Excluded rows and rows missing the response receive empty cells for these actions. Saved scores use the retained forest, without refitting. The spread between trees is not presented as a prediction interval. Classification forests, sample weights and grouped or temporal forest validation are not offered.

35.2 Penalized logistic regression

Choose **Statistics > Generalized & mixed models > Penalized logistic (elastic net)...**. Identify the binary outcome, its event value and the predictors. Confirm independent observations. The default seed is 20260923 and the default classification threshold is 0.5. A different threshold needs a substantive decision rule; changing it does not change the probabilities.

With $y_i = 1$ for the selected event and $\eta_i = b_0 + x_i^{\mathsf{T}}\beta$, the objective is

$$\frac{1}{n}\sum_i \{\log(1 + e^{\eta_i}) - y_i\eta_i\} + \lambda\left\{\frac{1-\rho}{2}\|\beta\|_2^2 + \rho\|\beta\|_1\right\}.$$

The intercept is unpenalized. Ridge uses $\rho = 0$, lasso $\rho = 1$ and elastic net values between them. The fixed grid combines $\lambda = 10^{-4}, 10^{-3}, 10^{-2}, 0.1, 1, 10, 100$ with $\rho = 0, 0.25, 0.5, 0.75, 1$.

A stratified test sample is reserved first. Five stratified training folds select the lowest pooled log loss, with near-ties preferring the larger penalty, then larger ρ. Each class needs at least five training rows and one test row. Missing outcomes are omitted; missing numeric predictors use training medians. Numeric centring and scaling and categorical reference levels are learned on the fitting rows only. A new category or previously unseen missing level prevents evaluation for that fit.

The results compare log loss, Brier score and ROC AUC with the training prevalence baseline. A reliability plot groups test probabilities only when there is adequate support; tied probabilities stay together. A small or flat plot does not establish calibration. Coefficients are penalized estimates: ordinary logistic standard errors and p-values do not accompany them.

Worked example

For $x = (1, -1, 2)$, $y = (1, 0, 0)$, $b_0 = 0.5$, $\beta = 0.3$, $\lambda = 0.1$ and $\rho = 0.25$, the linear predictors are $(0.8, 0.2, 1.1)$. The penalty is $0.1(0.75 \cdot 0.09/2 + 0.25 \cdot 0.3) = 0.010875$. Add this to the *average* of the three logistic losses, not their sum. This calculation explains the objective; it is not a fitted-model example.

Add to the open table writes the retained model's event probabilities using the predictors, without requiring an outcome or refitting. Unscorable rows remain empty and are counted. Training-row probabilities remain in-sample. Saving, undoing and replaying the column preserve the selected fit.

35.3 Quantile regression

Choose **Statistics > Correlation & regression > Quantile regression...**. Select the response, optional predictors and one to five distinct quantiles between 0.01 and 0.99, separated by semicolons. The default is the median, 0.5. The interval level defaults to 0.95 and the seed to 0. An optional 25% holdout evaluates predictions on rows withheld before fitting.

For quantile τ, the coefficients minimize check loss:

$$\sum_i \rho_\tau(y_i - x_i^\top \beta), \qquad \rho_\tau(u) = u\{\tau - \mathbf{1}(u < 0)\}.$$

Positive residuals receive weight τ and negative residuals weight $1 - \tau$. At $\tau = 0.5$ this is half the absolute-error loss. It targets a conditional quantile rather than the conditional mean fitted by least squares.

Complete finite rows enter the fit; exclusions are counted. The design must have full rank and more rows than terms. Up to 10,000 complete rows are supported. The computation rescales the fitting data for numerical stability and checks its returned solution; a failed fit is not replaced with OLS.

Worked example

Choose **File > Import data...** and select `sample-data/manual/quantile-four.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Correlation & regression > Quantile regression...**. Set **Response** to **y**, leave numeric and categorical predictors empty, enter 0.5 in **Quantiles**, set **Interval level** to 0.95, **Observations** to **Independent rows**, and **Seed** to 0. Leave **Hold out 25% of rows to evaluate the fitted quantiles** off. Click **Run analysis** and inspect the intercept and non-uniqueness results. Any intercept from 2 to 3 minimizes check loss at 2. HypoNum reports the lower empirical quantile, 2, and the minimizing interval $[2, 3]$. This interval describes non-uniqueness of the optimum, not sampling uncertainty. A bootstrap confidence interval is withheld here because only two observations support each tail.

Coefficient confidence intervals use 999 seeded IID pairs-bootstrap samples and linearly interpolated percentile endpoints. A failed or rank-deficient resample withholds the corresponding intervals; failed draws are not replaced. Intervals are also withheld when $\min(n\tau, n(1-\tau)) < 10$. That operational threshold does not establish coverage above ten. The reported near-optimal coefficient ranges are a separate sensitivity calculation at the stated numerical tolerance, not confidence intervals or a proof of uniqueness.

Separate quantile fits may cross. Crossings are counted at the evaluated rows and plotted points, and fitted values are kept unchanged. Absence of crossings at those points does not establish their absence elsewhere. Held-out results report mean check loss and the fraction at or below each predicted quantile. **Add conditional-quantile scores** adds one retained-fit column per quantile for rows with usable predictors and known categories; other rows stay empty. These are conditional-quantile estimates, not confidence bounds for the coefficients.

CHAPTER 36

Bootstrap uncertainty for an estimate

Choose **Statistics > Compare means & ANOVA > Bootstrap an estimate...** to estimate sampling variability by resampling the observational units and recomputing a statistic. The design determines what may be resampled independently. Resampling individual visits from a repeated-subject study would generally understate dependence between visits.

36.1 Select the statistic and the resampling unit

Available statistics include means, medians, trimmed means, standard deviations, quantiles, correlations, regression coefficients, two-group mean differences, medians of paired differences and selected reliability/agreement coefficients. Select the required response, grouping, predictor or rater columns. Paired calculations keep both measurements together. Regression uses whole cases, retaining their response and predictors and the original category coding.

Choose rows, rows within strata, whole clusters, or clusters within strata. Every stratum needs at least two resampling units, and each cluster must belong to one stratum. A cluster sample retains the number of clusters but may change the number of rows. These schemes assume independent resampling units within strata. They do not apply survey weights or finite-population corrections. For a complex survey, use the Survey methods appropriate to that design.

The default is 5,000 replicates, with at most 10,000. Record the seed, interval level, statistic and resampling scheme. More replicates reduce simulation error; they do not repair a sample or resampling design that cannot represent the target population.

36.2 Separate sampling uncertainty from simulation error

For observed estimate $\hat{\theta}$ and B bootstrap estimates θ_b^*, the reported bias is $\overline{\theta}^* - \hat{\theta}$ and the bootstrap standard error is the sample SD of the B estimates. The percentile interval takes the $\alpha/2$ and $1 - \alpha/2$ replicate quantiles, with linear interpolation.

For example, the sample $(1, 2, 3)$ has mean 2. Its empirical resampling distribution has variance $2/3$, so the exact SD of the mean of three independent draws from that distribution is $\sqrt{2}/3 = 0.4714$. A finite bootstrap run approximates this SD and will vary with the seed. This example explains resampling arithmetic; it does not establish interval coverage for samples of three observations.

The displayed Monte Carlo standard errors, approximately $SE/\sqrt{B}$ for bias and $SE/\sqrt{2(B-1)}$ for the standard error, describe variation between replicate sets. They are not an additional confidence interval for the population parameter.

36.3 Use BCa only where supported

The optional bias-corrected and accelerated (BCa) interval adjusts the percentile levels using the fraction of replicates below the observed estimate and a delete-one-unit jackknife. With bias

correction z_0 and acceleration a, an endpoint uses probability

$$\Phi\left[z_0 + \frac{z_0 + z_\gamma}{1 - a(z_0 + z_\gamma)}\right], \qquad \gamma \in \{\alpha/2, 1 - \alpha/2\}.$$

BCa is not offered for medians, quantiles, medians of paired differences, or clusters within strata. It needs at least 200 replicates and warns below 1,000. When its correction cannot be calculated, it is unavailable; a percentile interval is not silently substituted.

36.4 Read failed-draw and endpoint warnings

Some resamples lose a factor level, have a constant variable, lack a group or produce separated logistic data. Such failures are counted with their reasons. The observed estimate remains available, but bootstrap summaries for that statistic are withheld rather than based on the successful draws alone. An endpoint with fewer than one expected replicate beyond it is withheld; fewer than ten indicates poor resolution. Report these limitations alongside the estimate instead of reporting only an apparently precise interval.

CHAPTER 37

Combining evidence across studies

Statistics > Meta-analysis... combines independent study estimates. Use one row per study and ensure that the effects have the same scientific meaning, units and direction. A pooled association cannot correct biased study designs. Several comparisons sharing participants or a control group are correlated and need a model that accounts for that correlation.

37.1 Prepare comparable study effects

Choose generic effect and standard error; two-group sizes, means and SDs; or two-group event counts and totals. Continuous outcomes support mean differences, Cohen's d and Hedges' g. Binary outcomes support log odds ratios, log risk ratios and risk differences. Ratios are analysed on the log scale; exponentiation converts estimates and interval endpoints back to ratios. A log ratio of zero corresponds to a ratio of one.

For log ratios, studies with no events in either group or all events in both groups are excluded and listed. Other zero-cell studies use the declared 0.5 correction in all four cells and are marked. Risk differences use the original counts; a study with zero estimated variance cannot receive an inverse-variance weight and is excluded. Inspect this accounting before interpreting a pooled estimate.

37.2 Choose the pooling and uncertainty model

The common-effect model weights y_i by $w_i = 1/v_i$ and estimates $\hat{\mu} = \sum_i w_i y_i / \sum_i w_i$. It assumes the studies estimate one common underlying effect. The random-effects model instead uses $w_i = 1/(v_i + \hat{\tau}^2)$, allowing true effects to vary between studies. REML is the default estimator of τ^2; DerSimonian–Laird, Paule–Mandel and Sidik–Jonkman are alternatives.

The default random-effects inference is modified Hartung–Knapp: multiply $1/\sum_i w_i$ by $q = \max\{1, \sum_i w_i (y_i - \hat{\mu})^2/(k-1)\}$ and use a t_{k-1} reference distribution. Ordinary Hartung–Knapp removes the floor at one; normal inference is another explicit choice. A zero heterogeneity estimate is a boundary result, not proof of identical effects.

Worked example

Choose **File > Import data...** and select `sample-data/manual/meta-three-studies.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Meta-analysis...**. Set **Data layout** to **Effect and standard error per study**, **Effect** to **effect**, **Standard error** to **se**, and **Model** to **Common (fixed) effect**. Set **Confidence level** to 0.95; common-effect inference uses the normal reference automatically. Leave subgroup and moderator roles unset. Click **Run analysis**. Inspect the pooled effect, standard error, interval and heterogeneity table. For independent generic effects $(0.2, 0.4, 0.6)$ with standard errors $(0.1, 0.2, 0.1)$, common-effect weights are $(100, 25, 100)$. The pooled effect is

0.4, its standard error is $1/\sqrt{225} = 0.06667$, and its normal 95% interval is approximately $[0.2693, 0.5307]$. Cochran's Q is 8 on two degrees of freedom. The precision of the pooled mean does not erase this study disagreement; review the scientific reasons for heterogeneity and a random-effects analysis.

37.3 Read heterogeneity and prediction separately

The result includes Cochran's Q, τ^2 and τ with Q-profile intervals, and I^2 and H^2 derived from heterogeneity relative to the typical within-study variance. I^2 is not the percentage of studies that are biased or the probability that an effect is real.

For at least three studies, a random-effects prediction interval approximates the true effect in a new comparable study:

$$\hat{\mu} \pm t_{k-2,1-\alpha/2}\sqrt{\hat{\tau}^2 + \widehat{\mathrm{Var}}(\hat{\mu})}.$$

It describes neither an individual outcome nor the new study's observed estimate. Its coverage is approximate, especially with few studies or uncertain heterogeneity. The forest plot shows study intervals, weights, pooled estimates and the prediction interval together.

37.4 Moderators, sensitivity and small-study checks

Subgroups are pooled separately with their own heterogeneity estimates. Meta-regression relates effects to study-level numeric or categorical moderators. A study-level association need not describe the same association among individuals. The moderator test, residual heterogeneity and stated inference method should accompany any fitted slope.

In **Predict at**, separate prediction points with semicolons and join moderator assignments within a point with `&`; for example, `ablat=20;ablat=30&alloc=random`. Supply every moderator and use analysed factor levels. The result separates uncertainty in the average effect at those settings from prediction of a new study's true effect.

Cumulative and leave-one-out analyses show sensitivity to study order and individual studies. Egger, Begg and trim-and-fill calculations investigate funnel asymmetry; they do not diagnose publication bias uniquely. They have little power with fewer than ten studies, and apparent asymmetry can reflect clinical heterogeneity, design differences or chance. Report them as exploratory checks alongside the study selection process.

CHAPTER 38

Bayesian estimation and model comparison

The **Statistics > Bayesian statistics** commands combine a likelihood with an explicitly selected prior. They report posterior summaries and, where defined, Bayes factors for named models. A credible interval concerns the posterior under that model and prior. A Bayes factor compares two models' predictive support; it is neither a p-value nor a posterior model probability by itself.

38.1 Declare the scientific parameter and its prior

The available procedures cover one-sample, paired and independent-sample t tests; Pearson correlation; a binomial proportion; a Poisson rate; a normal mean with known SD; one-way fixed-factor ANOVA; and linear regression model comparison. Prepare the same observational design needed by the corresponding likelihood. Bayesian inference does not make paired observations independent or identify an exposure that was not recorded.

Procedure	Prior and principal choice
t tests	Cauchy prior on standardized effect, default scale $1/\sqrt{2}$; declare the null mean or paired difference and the comparison direction.
Pearson correlation	Stretched beta prior on correlation in $(-1, 1)$; default width 1 is uniform.
Binomial proportion	Beta(a, b), default Beta$(1, 1)$; name the event and point-null proportion.
Poisson rate	Gamma shape and rate, both positive, per unit of exposure. The Jeffreys option supplies a posterior only.
Normal mean, known SD	Normal prior on the mean; supply the known observation SD, prior mean and prior SD.
One-way ANOVA	Shared scale for fixed level effects, default 0.5; at most ten levels.
Linear regression	Zellner–Siow or hyper-g prior; declare candidate models and their prior probabilities, including the intercept-only model.

For independent-group effects, verify which group is subtracted from which. Changing the sign convention changes directional interpretation even when a symmetric prior leaves a two-sided Bayes factor unchanged.

38.2 Work through a beta–binomial update

Choose **File > Import data...** and select `sample-data/manual/bayes-binomial-ten.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Statistics > Bayesian statistics > Binomial proportion...**. Set **Variable** to **y**, **Event value** to **1**, **Null proportion** to 0.5, **Beta prior a** and **Beta prior b** to 1, and the credible level to 0.95. Click **Run analysis**. Inspect the posterior mean and SD and the Bayes factor against the point null. Six independent trials succeed and four fail. Under a uniform $\mathrm{Beta}(1,1)$ prior, the binomial likelihood updates the proportion to $\mathrm{Beta}(7,5)$. Its posterior mean is $7/12 = 0.58333$, rather than the sample proportion 0.6. Its posterior SD is

$$\sqrt{\frac{7 \cdot 5}{12^2 \cdot 13}} = 0.13674.$$

The imported ten rows provide the binary response; the success category is 1, the two prior shapes are 1, and the point null is 0.5. The Bayes factor for the beta-prior alternative against that point null is

$$\mathrm{BF}_{10} = \frac{B(7,5)/B(1,1)}{0.5^{10}} = 0.44329.$$

The data therefore have greater predictive support under this point-null model than under this particular alternative prior. This statement depends on the prior's spread; it does not assert that the null is true.

38.3 Interpret posterior summaries and Bayes factors

The results show the posterior mean, SD, median and equal-tailed credible interval. A highest-density interval is given when one interval can represent the supported posterior shape; it is withheld for a multimodal shape that needs disjoint regions. A boundary mode may produce an interval starting at that boundary. Inspect the prior/posterior plot together with the summaries.

The result names both models in BF_{10}, shows its reciprocal and logarithm, and gives numerical-integration error where relevant. Posterior odds equal prior odds multiplied by the Bayes factor. For example, prior odds 1:1 and $\mathrm{BF}_{10} = 3$ imply posterior odds 3:1, or alternative-model probability 0.75. Different prior model odds produce different posterior probabilities from the same Bayes factor.

An improper prior's arbitrary normalization prevents a point-null Bayes factor from being defined. This is why the Poisson Jeffreys option reports posterior results without inventing such a comparison.

38.4 Keep model selection uncertainty visible

Regression compares up to 64 declared models using at most ten predictors, with posterior model probabilities under the stated model-prior weights. Coefficient summaries for the most probable model are conditional on that model; they do not automatically average over model uncertainty. Two-factor Bayesian ANOVA, random effects and models needing several effect-scale parameters are not offered here.

The calculations use deterministic integration rather than Monte Carlo chains. Repeating the settings therefore does not supply a convergence diagnostic from independent chains. Investigate integration warnings and repeat a scientifically meaningful prior-sensitivity analysis before interpreting a precise-looking numerical Bayes factor.

CHAPTER 39

Guided statistical studies

Guided studies connect a question, collection design, analysis population, method, diagnostics and interpretation. They use the same methods described in this manual; a completed stage does not independently validate a method or prove its assumptions. Review and accept the proposed method and settings before fitting. Later changes need a recorded reason, and earlier outcomes remain in the study history.

39.1 Comparing groups

Statistics > Guided study > Guided comparison of groups... starts with difference, estimation or equivalence, outcome type and collection design. Independent continuous groups route to Welch procedures without a preliminary equal-variance test; paired measurements route to within-unit differences. Pairing uses unit identity and condition, not row position. Duplicate unit/condition observations are refused. Incomplete pairs are identified and require confirmation of the retained population.

Prespecify directional group order, a planned comparison family, and any equivalence margins in response units. Reversing group order reverses a difference and its interval; an odds ratio is reciprocated. A rank-test rejection is not automatically evidence of a median difference. Planned Welch pairs retain their original multiplicity family even when one pair cannot be estimated. The all-pairs Games–Howell route and planned pairs need no significant omnibus gate.

At 0.05, one TOST equivalence claim uses a 90% interval. An interval strictly inside prespecified margins supports that claim. A nonsignificant difference test does not establish equivalence, and a descriptive comparison with a meaningful-difference threshold is not a substitute for TOST. The guide refuses unsupported repeated/clustered designs, paired binary outcomes and proportion-equivalence questions instead of substituting a convenient test.

39.2 Association and prediction

Statistics > Guided study > Guided association or prediction... separates adjusted association from performance on new units. Causal intervention questions are outside its supported claims. Declare independence, clustering or repeated observations before selecting a route; unknown dependence is not silently interpreted as independence.

Independent quantitative estimation uses OLS with HC3 uncertainty; binary and count outcomes use their stated generalized models. A high Poisson dispersion raises an attention flag, not an automatic negative-binomial substitution. Repeated quantitative outcomes use the declared random-intercept model. Clustered binary/count questions distinguish marginal GEE from conditional mixed-model effects. More complex residual covariance or survey design requires the corresponding separate workflow. Name a focal term for the interpretation, or read the entire coefficient table without letting its first row choose the scientific question.

For prediction, all predictors must be available at the intended prediction time. The guide reserves a seeded 20% test set before development and uses five training folds. Quantitative

candidates are OLS, ridge and lasso, compared by the mean of fold RMSEs. For fold RMSEs $(0, 0, 0, 0, 10)$ that score is 2; it is not the pooled root mean square 4.472. Binary prediction uses logistic regression and fold-mean log loss, Brier score and AUC. Failed folds make a candidate ineligible rather than changing its estimator.

Lock the procedure before final evaluation. Reusing reserved outcomes for threshold choice, recalibration or another evaluation makes them development-used, a status retained after reopening. Deployment fitting on all rows is a separate step. A descriptive calibration table does not itself establish calibrated probabilities.

39.3 Missing-data analysis

Statistics > Guided study > Guided missing-data analysis... covers purpose, missingness audit, design assumptions, imputation specification, chain diagnostics, pooled analysis, sensitivity and export. Descriptive work can end at the audit; prediction goes to a workflow that learns preprocessing within training data. Inferential pooling supports the declared additive OLS or binary logistic model with independent rows.

Review each variable's role and imputation kind, including auxiliaries. Numeric category codes do not become measurements by default. An ordinal declaration currently uses nominal imputation and discloses loss of ordering. Identifiers, derived/passive variables, unsupported dependent designs and structural not-applicable blanks in a used variable block the imputation route. An all-missing column supplies no observed basis for its conditional model. Acknowledging MAR records an assumption, not a test result.

Defaults are 20 imputations, 40 sweeps and seed 20260923. Inspect chain traces, donor reuse and pooled-coefficient Monte Carlo error. More imputations address Monte Carlo precision; more sweeps address chain behaviour. A failed chain prevents pooling over just the successful subset. Completed values are plausible draws, not recovered observations.

Complete-case sensitivity uses the same outcome, predictors and contrasts. Its different population limits the comparison; agreement does not prove MAR, and the guide offers no formal MNAR adjustment. Exported imputations carry source row identities, imputation numbers and original-missingness flags. Analyse each completed data set and pool appropriately; stacking them as one larger independent sample understates uncertainty.

39.4 Time-series studies

Time Series > Guided time-series study… audits order, duplicate/missing times, spacing and missing responses before fitting. It does not quietly sort, interpolate or infer a season. Confirm a season or explicitly choose none. The default horizon is 12 audited periods, with a reserved final block of that length and expanding rolling-origin windows. Fewer than five origins are labelled limited evidence; overlapping origins are dependent.

Training-history exploration and candidate comparison use development rows only. Naive benchmarks stay in the comparison. A candidate must succeed with acceptable fit status at every scheduled origin. Seasonal MASE divides each origin's absolute errors by that origin's mean absolute lag-m training difference; without a season it uses lag 1. If any required scale is zero, MASE is unavailable for the common comparison and MAE is used instead. For $(1, 10, 2, 11, 3, 12)$ with season 2 the scale is 1, so an absolute error of 2 gives MASE 2; using lag 1 would change the metric.

The final reserved block is evaluated after locking the method. A later deployment fit uses current history and issues a separate frozen forecast with cutoff, targets, units and prediction intervals.

Monitoring joins actuals by time; revisions do not erase issued forecasts. Log-transformed forecasts back-transform to original-scale medians. Predictor-conditional forecasts require the separate dynamic-regression workflow with future predictor paths. Rolling coverage is descriptive covered/usable/scheduled counts, not an independent qualification of forecast interval coverage. Earlier saved comparison studies retain their labelled legacy evaluation convention.

39.5 Reliability and experimental studies

The guided reliability path records observation type, censoring assumptions, population, units and objective before choosing nonparametric survival, parametric life, covariate, accelerated-life or demonstration-planning routes. Use its stated diagnostics: an exploratory residual-correlation screen is not the formal proportional-hazards score test available in another workflow. Extrapolated use conditions and assumed known shapes remain explicit.

Guided experiment planning declares the required model before ranking designs. Budget, estimability, blocks, hard-to-change factors and physical constraints must all be feasible; an unsupported requirement is not dropped. Screening, interaction estimation, response-surface optimization, mixtures and robust noise performance require different designs. In guided analysis, adopt the active table's recorded design and reconcile planned run identities, actual settings and responses. A duplicated identity, wrong block or changed whole-plot structure needs explicit resolution. Permitted deviations remain visible and the fit uses recorded settings. Distinguish prospective randomized experiments from observational run sheets; a fitted response surface alone does not establish an intervention effect.

39.5.1 Simulation precision of forecast endpoints

When the guided forecast reports a Monte Carlo precision band for a simulated quantile, it uses binomial order-statistic ranks for independent draws at that single horizon. It does not provide a simultaneous band over forecast times. A required rank of zero or $N + 1$ lies outside the simulated sample: the precision band, ranks and width are then unavailable, rather than replaced by the sample minimum or maximum. Increasing the supported simulation count (100–2000) can improve precision. The warning for fewer than 20 expected draws in a tail is a diagnostic heuristic, not a guarantee of accuracy.

39.5.2 Supported experiment-planning criteria

Choose a supported design family and inspect its declared estimation model before comparing its criterion value. D-optimality targets the determinant of information for that model; I-optimality targets average prediction variance over its stated candidate region. In **Design > Guided study > Guided experiment planning...**, enter the **Blocks** and **Runs for an optimal-design alternative**, then choose **Optimal criterion**. D is the objective default for screening/effects; I is the default for prediction/optimization. Hard-to-change and mixture requirements use their separate design families.

For fixed blocks, capacities are balanced in advance; remainder runs go to the earlier block IDs. The search chooses treatment settings within those fixed slots jointly, allows repeated candidate settings, and randomizes the selected runs within blocks afterward. It does not assign blocks after optimizing an unblocked design. Feasible candidates retain the planned terms, region constraints and forbidden-corner exclusions.

Let Z contain the block indicators and F the treatment columns without an intercept. The combined model $X = [Z, F]$ must have full column rank. D maximizes $\log\det\{F'(I - P_Z)F\}$, where $P_Z = Z(Z'Z)^{-1}Z'$. I averages $\ell(x)'(X'X)^{-1}\ell(x)$ uniformly over the feasible candidate grid,

with $\ell(x) = (1/B, \ldots, 1/B, f(x))'$. This is prediction of the equal-weight mean over the B named fixed blocks, per unit common error variance. It is neither the average of separate blockwise variances nor prediction in an unseen random block.

Inspect the achieved criterion, capacities, model rank and stopping reason. The deterministic search has at most eight coordinate-exchange sweeps and one million candidate evaluations. A search limit returns the checked full-rank design achieved so far and labels the limit; it does not certify convergence or a global optimum. Insufficient runs or a candidate region that cannot estimate the declared model is refused. Earlier saved planning calls retain their original algorithm, including the old blocked-optimal refusal; reopening them does not silently generate a new design.

Part IV

Survey

CHAPTER 40

Survey summaries and scale consistency

Survey analysis starts with the sampling design and the meaning of each response. Weighted summaries describe the population represented by a declared design. Item consistency addresses a different question: whether the responses to a set of items vary together as a summative scale. The two workflows are presented together here, with their distinct assumptions and handling of incomplete responses.

40.1 Survey means and proportions

Survey weights represent a sampling design. Unequal weights change the estimate, and clustering and stratification change its uncertainty. Treating survey weights as replicated observations or as precision weights would give a different analysis.

Choose **Survey > Survey means and proportions...**. Assign four distinct roles: numeric response, positive survey weight, stratum and primary sampling unit (PSU). Numeric identifiers can serve as stratum or PSU labels. Roles are explicit; the software does not infer a design from column names.

The initial estimate is Mean with 95% confidence. Binary proportion is the alternative, requiring observed responses exactly 0 or 1. Confidence accepts whole percentages 51–99. Opening setup resets these choices.

40.1.1 Worked example: eight observations in four PSUs

Choose **File > Import data...** and select `sample-data/manual/survey-design-eight.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Survey > Survey means and proportions...**. Set **Response (numeric)** to **response**, **Survey weight (numeric)** to **weight**, **Stratum** to **stratum**, and **PSU** to **psu**. Select **Mean**, retain 95% confidence, and click **Run survey descriptives**. Check that eight observations, two strata and four stratum-specific PSUs are analyzed, then compare the mean, standard error and interval printed below.

Stratum	PSU	Response	Weight
A	1	2.0	1.00
A	1	3.0	1.50
A	2	4.5	2.00
A	2	5.0	0.75
B	1	5.5	1.25
B	1	7.0	2.50
B	2	8.0	1.00
B	2	9.5	1.75

The weight total is 11.75 and weighted response total 68.25, so

$$\hat{\mu} = 68.25/11.75 = 5.8085106.$$

PSU 1 in stratum A differs from PSU 1 in stratum B: there are four design PSUs, not two. The Taylor standard error is 0.6577195. With $df = 4-2 = 2$, the 95% interval is $[2.9785722, 8.6384490]$. The wide interval reflects the small number of independent design units despite eight measured observations. These constructed data demonstrate accounting, not an adequate survey design.

40.1.2 The supported variance estimator

The implemented design is one-stage sampling with replacement. For observed responses and weights w_j,

$$W = \sum_j w_j, \qquad \hat{\mu} = \frac{\sum_j w_j y_j}{W}.$$

For PSU i in stratum h, define the linearized total

$$u_{hi} = \sum_{j \in hi} \frac{w_j(y_j - \hat{\mu})}{W}.$$

If m_h original valid design PSUs occur in stratum h,

$$\widehat{V}(\hat{\mu}) = \sum_h \frac{m_h}{m_h - 1} \sum_i (u_{hi} - \bar{u}_h)^2, \qquad df = \sum_h m_h - H.$$

The interval is $\hat{\mu} \pm t_{1-\alpha/2,df}\sqrt{\widehat{V}}$. Multiplying every weight by one common positive finite constant leaves the estimate and SE unchanged.

Binary proportion applies the same formulas to a 0/1 response. Its ordinary-scale t interval can extend outside [0,1]. HypoNum displays that approximation without clipping and gives a warning. Clipping would conceal the interval's limitations.

40.1.3 Missing responses and valid design information

Weights, strata and PSUs must be valid on every source row, including rows without a response. Weights must be finite and strictly positive; design labels must be nonmissing and nonempty. Each stratum needs at least two original PSUs and an observed response. Singleton strata are refused.

Missing or non-finite responses are omitted from the response-complete estimand and disclosed. Their original PSUs remain in the design count. A PSU with no observed response contributes a zero linearized total. This convention does not correct bias caused by informative nonresponse; the suitability of weights and response adjustment remains a study question.

Active row exclusions and domain filters are refused because this workflow does not implement design-aware subpopulation inference. Deleting observations before import also changes the supplied design; it is not a substitute for a domain analysis. Finite population correction, multi-stage designs, replicate weights, survey regression and weighted population-total estimation are unavailable.

40.1.4 Read and save

The result lists source and analyzed counts, missing responses, observed weight total, strata, original PSUs, design degrees of freedom, estimate, SE, interval and warnings. Limits are one million source rows, twenty million selected cells and 10,000 design PSUs.

Add to results saves the estimate with these design details and settings. Reopening restores the snapshot; rerunning explicitly uses the current source. Closing setup or selecting Cancel stops active work.

40.2 Item consistency and summative scores

Cronbach's alpha summarizes how a declared set of items varies together. It does not establish a single underlying construct, a valid measurement scale or agreement with an external standard. Content, design and dimensionality require separate consideration.

Choose **Survey > Item consistency (Cronbach's alpha)...**. Select two to 64 numeric item columns. Every calculation uses the same complete-case rows across those items. Missing and excluded counts are reported; nothing is imputed.

40.2.1 Align the direction of items

Declare oppositely directed items **Reverse-coded** and supply shared scale bounds Low and High. The analyzed value becomes

$$x_{\text{adjusted}} = \text{Low} + \text{High} - x.$$

On a 1–4 scale, 1 becomes 4, 2 becomes 3, and vice versa. All observed values of a reversed item must lie within the declared bounds. Reverse coding should follow the item's wording and measurement design, not a search for whichever sign maximizes alpha.

40.2.2 Worked example

1. Choose **File > Import data...** and select `sample-data/manual/item-consistency-four.csv`. In the preview, use delimiter comma, encoding **utf-8**, decimal point, header row 1 and first data row 2; click **Import**. Make the imported table active and keep all four rows included. Its three columns are **x**, **y** and **z**, with the values printed below.
2. Choose **Survey > Item consistency (Cronbach's alpha)...**. Double-click **x**, **y** and **z** in **Available columns** to add them to **Items**. Under **Reverse-coded items**, tick only **z**. Set **Shared scale bounds** to **1** and **4**.
3. Click **Run item analysis**. Check that four complete respondents and three items were analyzed; compare the alpha and item statistics below.
4. To reproduce the last column of the table, enter **mean_score** in **Score column name**, choose **Mean of items**, and click **Create score column** in the results window. Inspect the new column in the active table. Reverse coding is used in this score; the original **z** values remain as imported.

Row	x	y	z	Adjusted z	Mean score
1	1	2	4	1	1.333333
2	2	2	3	2	2.000000
3	3	4	2	3	3.333333
4	4	4	1	4	4.000000

Raw alpha is 0.975 and standardized alpha is 0.9753844. Corrected item-total correlations are 0.9761871 for x and adjusted z, and 0.8944272 for y. Deleting y leaves two identical adjusted items and alpha 1. That numerical result is a useful warning: a perfect coefficient can arise from redundant questions. It does not imply an ideal instrument. Four constructed respondents provide arithmetic transparency, not a reliable empirical validation of a scale.

40.2.3 Raw and standardized alpha

For k adjusted items and total score $T = \sum_j X_j$, raw alpha is

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_j s_j^2}{s_T^2}\right),$$

using sample variances on the common complete cases. Standardized alpha uses mean pairwise item correlation $\bar{r}$:

$$\alpha_{\text{std}} = \frac{k\bar{r}}{1 + (k-1)\bar{r}}.$$

Raw alpha depends on item scales; standardized alpha uses correlations. Choose interpretation in relation to how the intended score is formed.

Negative coefficients are reported and explained, not clipped. They can indicate items working against one another or a coding problem. A zero standardized denominator makes that coefficient unavailable. If total-score variance is zero, raw alpha is undefined and the analysis is refused.

A constant item remains in the declared item set. Raw alpha remains calculable if the total varies; one varying and one constant item yield raw alpha 0. Correlations involving a constant item, the mean inter-item correlation and standardized alpha of the declared set are unavailable. Other pairwise correlations remain visible.

40.2.4 Item diagnostics and score creation

Each item has an adjusted mean and SD, correlation with the sum of the *other* items, and alpha after deletion. The corrected correlation avoids correlating an item with a total that already contains itself. It is unavailable when that item or the sum of the others is constant. Alpha after deletion is blank for a two-item scale.

Review the correlation table and item content before deleting an item. Repeatedly selecting items for high alpha in one sample can give an optimistic picture of a new sample's score consistency.

Create score column adds an explicitly named mean score, initially, or a sum score. In the worked example the sums are 4, 6, 10 and 12. Only analyzed complete rows receive scores; omitted rows remain missing. The original table's row alignment is preserved through sorting. The score operation supports undo, redo and saved-project replay.

Add to results retains bounded summary and correlation tables. Source changes require a new run. Limits are one million source rows and twenty million selected numeric cells.

CHAPTER 41

Inference from a complex sample

Sampling weights identify the represented population. Strata and sampling units determine the design uncertainty. Two analyses with the same weighted mean can legitimately have different standard errors because their sampling designs differ. Integer case replication is not a substitute for this design.

41.1 Save the table's sampling design

Survey > Define design... records a positive finite sampling-weight column, optional first-stage strata and primary sampling units (PSUs), and optional finite-population correction (FPC). Without a PSU column, each row is a sampling unit. An FPC is either a population count of units or a sampling fraction, with the interpretation declared explicitly. It must be consistent within the relevant stratum.

A second sampling stage can supply its own strata, units and FPC within PSUs. When unit codes restart in each stratum, select that option so that the stratum and unit jointly identify a PSU. Otherwise reused codes are refused as nonnested units. A single observed PSU does not prove certainty: use the documented certainty, stratum-combination or centring convention only when it describes the design, and report it.

Alternatively, supply replicate-weight columns and declare their method, whether they are complete weights or multipliers, variance scale, replicate-specific factors and centring convention. Fay replication also needs its Fay factor. These conventions are not inferred from column names. Replicate columns can be supplied or created by the explicitly supported one-stage designs described below. Their sampling interpretation is required in either case; a set of weights alone does not identify that design.

The design is an undoable table declaration. Analyses keep a copy of the design they used, so a later declaration change does not rewrite retained results. Deleting a required design column removes the current declaration.

41.2 Estimate a domain without discarding its design context

Choose the Survey means/totals/proportions, ratios, table or regression command after saving the design. Use the domain selector for a subpopulation instead of first deleting other rows. Nondomain sampling units still contribute to the variance structure. The design's degrees of freedom and the domain's response count describe different quantities.

For weights $(2, 1)$ and values $(10, 20)$, the weighted mean is $(2 \cdot 10 + 1 \cdot 20)/3 = 13.3333$. The represented total is 40. Those two rows alone do not specify a survey standard error: strata, PSUs, sampling fractions or replicate conventions are still needed.

Taylor linearization aggregates influence contributions within PSUs and estimates their between-PSU variation within strata, applying declared FPCs. A second-stage contribution is added where the first-stage fraction permits it; without a first-stage FPC, the with-replacement ultimate-cluster approximation already includes later-stage variability.

Replicate variance instead recomputes the estimate for every supplied replicate and uses

$$\widehat{V} = c \sum_r c_r (\hat{\theta}_r - \tilde{\theta})(\hat{\theta}_r - \tilde{\theta})',$$

where centring is at the full estimate or replicate mean as declared. The scale c is part of the method; an incorrect default can change uncertainty without changing the point estimate.

41.3 Tables and survey models

Survey crosstabs include design-adjusted Rao–Scott inference. Their reference degrees of freedom depend on generalized design effects; the ordinary Pearson chi-square is shown only as an unadjusted comparison.

Survey linear regression uses weighted least squares with design-based covariance. Logistic regression uses a weighted pseudo-likelihood and design covariance. By default, linear/logistic coefficient tests use a t reference with $D - \operatorname{rank}(X) + 1$ degrees of freedom, where D is the design degrees of freedom and X is the fitted model matrix. Term tests use F with the same denominator degrees of freedom. Results distinguish the design and test degrees of freedom; choosing asymptotic normal inference explicitly instead uses normal/chi-square references. Ordinal proportional-odds and survey Cox models use asymptotic normal/chi-square inference. A positive ordinal coefficient shifts probability toward higher ordered response levels; a Cox coefficient exponentiates to a hazard ratio, not a survival probability.

Complete or quasi-complete separation is refused rather than replaced by a penalized model. A large sampling-weight total does not create a large number of independently sampled PSUs. Read design degrees of freedom, missingness, domain size and convergence before interpreting significance.

41.4 Supplementary proportion intervals and design effects

In **Survey > Means, totals and proportions...**, the primary symmetric t interval remains visible. **Supplementary proportion interval** adds either a delta-logit interval or a Korn–Graubard beta/effective-size interval. For an interior estimated proportion p with design variance V, delta-logit transforms $\operatorname{logit}(p) \pm t_{D,1-\alpha/2}\sqrt{V}/[p(1-p)]$ back to the probability scale. The beta method uses

$$n_e = \frac{p(1-p)}{V}\left(\frac{t_{n_s-1,\alpha/2}}{t_{D,\alpha/2}}\right)^2$$

and beta quantiles with parameters $(n_e p, n_e(1-p)+1)$ for the lower limit and $(n_e p+1, n_e(1-p))$ for the upper limit. Here D remains the saved design's degrees of freedom and n_s counts domain rows before item missingness. The actual observed response count is reported separately. Neither interval has exact binomial coverage for an arbitrary survey. At $p = 0$ or 1, nonpositive variance, or insufficient degrees of freedom, the supplementary interval is unavailable with a reason.

Design-effect comparator compares the design variance of a mean or proportion with a declared simple-random-sampling (SRS) variance. With replacement, the comparator is the normalized weighted dispersion divided by the actual observed $n - 1$. DEFF is the variance ratio and DEFT is its square root. A common rescaling of weights does not change this comparator. Without replacement, the additional factor is $1 - f$; this option requires one-stage independent sampled rows, an explicit row FPC, inverse-probability population-scale weights and a complete sample without a domain or item nonresponse restriction. Totals and counts have no supported SRS comparator here. An unavailable comparator is not replaced by another one.

Worked example

Choose **File > Import data...** and select `sample-data/manual/survey-proportion-twelve.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Survey > Define design...**. Set the weight column to **w**, first-stage strata to **h** and first-stage units to **psu**; leave the second stage, FPC and supplied replicate columns empty. Save the design. Choose **Survey > Means, totals and proportions...**, leave numeric columns empty and select **yes** under categorical columns. Leave the domain empty, choose **SRS with replacement (normalized weights allowed)** and **Delta-logit t interval**, set confidence to 0.95 and run the analysis. For category 1, $p = 0.6428571$, $V = 0.0060001041$, $D = 2$ and DEFF $= 0.2874717$. The supplementary limits are $(0.2965347, 0.8848750)$. Run again with **Korn-Graubard beta/effective-n interval**: the limits are $(0.2978434, 0.9032352)$. These are two uncertainty conventions for the same estimated proportion, not two different datasets.

41.5 Count, nominal and interacting survey models

Survey > Count regression... fits a log mean to nonnegative integer counts. Poisson is the default. A strictly positive optional exposure enters as its logarithm with coefficient fixed at one; exponentiated slopes then multiply a rate per exposure unit. Without exposure they multiply a mean count. NB2 has working variance $\mu + \alpha\mu^2$. Externally fixed alpha gives conditional inference; jointly estimated alpha uses the complete coefficient/dispersion score covariance and re-estimates dispersion in every replicate. These choices are not interchangeable. Coefficient t/F inference uses the explicit approximate convention $D - \text{rank}(X) + 1$; estimating dispersion does not automatically subtract another degree of freedom. Normal/chi-square inference is an explicit alternative. An invalid reference df, nonconvergence, separation or dispersion boundary is refused.

Worked example

Choose **File > Import data...** and select `sample-data/manual/survey-counts-twenty-four.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Survey > Define design...**: weight **w**, first-stage strata **h**, first-stage units **psu**, no FPC, second stage or replicate columns. Save. Choose **Survey > Count regression...**: response **y**, family **Poisson**, covariate **x**, no factors, interactions, exposure or domain, confidence 0.95 and the default finite-df reference. Run the analysis. The intercept is approximately 0.29633547 and the slope 1.26594185. Choose **NB2 (jointly estimated dispersion)** and run again: the intercept is approximately 0.352367, slope 1.166084 and alpha 0.794435. These fitted coefficients describe the declared survey model; agreement with this example does not guarantee small-PSU interval coverage.

Survey > Multinomial regression... fits 3–12 observed response categories jointly. Enter the reference category explicitly. For each other category j, $\log[P(Y = j)/P(Y = \text{ref})] = x'\beta_j$. The design covariance includes cross-category blocks; term tests span all associated category coefficients. Coefficient intervals are asymptotic normal and joint tests asymptotic chi-square on the independent estimable restrictions. Displayed design df do not change that reference into a finite-sample test. Exponentiated slopes multiply relative category probabilities, not an absolute probability.

The linear, logistic, ordinal, count, nominal and Cox windows offer **Interactions**. Select all component factors/covariates first; enter `group:x` for a pair. A triple such as `a:b:x` requires its three lower-order pairs as well. Factors use their fitted reference coding; interactions multiply the corresponding model columns. Names containing colons cannot be expressed with this GUI shorthand.

Survey Cox additionally permits delayed-entry episodes and stratified baseline hazards. Keep complete subjects as the sampling units; episodes do not become independent PSUs. Whole-subject missingness and subject-level replicate refits preserve that distinction. Repeated-episode data refuse generated/calibrated row-weight provenance that cannot establish subject consistency. A survival path or recurrent-event model is not implied by these extensions.

41.6 Generate replicate weights from a declared sampling plan

Survey > Generate replicate weights... requires a complete saved Taylor design and explicit confirmation of first-stage SRS PSUs, not PPS. Exclusions, split subsets, multistage generation and noncertainty singleton strata are refused. The original full weight remains the full-sample weight. JK1 deletes each PSU of an unstratified sample; JKn deletes within strata, inflating the other PSUs by $n_h/(n_h - 1)$. Their recorded variance factors include the declared first-stage FPC. BRR and Fay require two noncertainty PSUs per stratum and no nonzero sampling fraction in those strata. Fay's default factor is 0.5. Balanced signs determine the full replicate weights. Rao–Wu here is a with-replacement one-stage SRS-PSU bootstrap: each stratum draws $n_h - 1$ PSUs with replacement. Defaults are 100 replicates and seed 970149; finite-population and more general PPS/multistage versions are absent.

Inspect method, counts, variance scale, replicate-specific scales, centring and degrees of freedom before **Add weights and design**. This creates the columns and declaration as one undoable edit, without overwriting columns. Changing required source values invalidates the recorded plan.

Worked example

Choose **File > Import data...** and select `sample-data/manual/survey-replication.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Survey > Define design...**: weight **w**, first-stage strata **h**, **PSUs p**, no FPC, second stage or supplied replicates. Save. Choose **Survey > Generate replicate weights...**; select **Stratified jackknife (JKn)**, tick **I confirm first-stage SRS PSUs (not PPS)**, keep prefix **Replicate** and full-sample centring, then run. Inspect six replicate columns and design df 4; click **Add weights and design**. Choose **Survey > Means, totals and proportions...**, select numeric column **x**, leave categorical columns/domain and supplementary options empty, set confidence 0.95 and run. The mean is 5.75 and its variance approximately 0.375529832. Starting again from the original imported table and declaring **N** as first-stage population-count FPC, save the design, then repeat **Generate replicate weights...** with the same JKn settings and apply **Add weights and design** before rerunning the mean. This regenerated JKn variance is approximately 0.281647374; the mean is unchanged. Simply estimating from the Taylor design after changing the FPC is a different calculation and gives variance approximately 0.282343107.

41.7 Calibrate to fixed known population totals

Survey > Calibrate weights... starts from the complete original design. Enter **Known totals**, for example `group[A]=18` and `group[B]=22` on separate lines, only when these are known population counts. This notation illustrates input syntax, not data for a worked fit. Every observed category level needs a positive target; numeric totals use `x=175`, and population size uses `(Intercept)=40`. Targets estimated in another sample require extra uncertainty not supplied by this fixed-total procedure.

Poststratification uses one complete cell column and factors $g_h = N_h / \sum_{i \in h} w_i$. Raking balances complete category margins. Linear calibration balances auxiliary totals while minimizing quadratic weight change; optional positive lower/upper bounds constrain the factors, not absolute weights. Defaults are tolerance 10^{-9}, 100 iterations and prefix **Calibrated**. Infeasible targets, missing auxiliaries or unresolved convergence refuse the plan; post-fit trimming is not substituted.

The recorded calibrated design retains the original design and auxiliary projection needed for Taylor variance. Merely replacing the weight column does not reproduce that variance. Every supplied replicate is recalibrated to the same totals and bounds; one failed replicate refuses the entire plan. No failed replicate is dropped or survivor scale renormalized. Original design df remain an explicit first-order convention after calibration; they are not increased from recalibrated matrix rank or automatically reduced by the number of auxiliary columns. Review balance and then choose **Add weights and calibrated design**. Domains are estimated afterward, with the full design context retained.

CHAPTER 42

Scale reliability and rater agreement

Reliability concerns reproducibility under a specified measurement model. Agreement additionally asks whether measurements are interchangeable on their original scale. Neither establishes validity or a single underlying construct. Use one row per independently sampled target and one column per item or rater. Select **Survey > Scale reliability (split-half, Guttman, ICC)...** or **Survey > Rater agreement...**. The separate **Quality > Measurement systems > Attribute Agreement (MSA)...** workflow is described in section 55.1.

42.1 Item consistency and declared halves

Select the numeric items, identify any reverse-coded items, and supply the actual lower and upper response bounds. Reversal uses $x^* = L + U - x$ once. Cases with an incomplete selected item set are omitted from scale reliability. Check their number before interpreting a coefficient.

For k items with covariance matrix S and total-score variance $s_T^2 = \mathbf{1}'S\mathbf{1}$, Cronbach's coefficient is

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\operatorname{tr}(S)}{s_T^2}\right).$$

Standardized alpha instead uses the mean inter-item correlation $\bar{r}$: $k\bar{r}/[1+(k-1)\bar{r}]$. Standardizing gives each item equal variance; it changes the meaning of a total made from items with different units. Guttman's coefficients give other covariance-based summaries. In particular, the reported split-half coefficient uses the declared split; the program does not search all possible splits and call the largest one a fixed estimate.

Choose first/second half, odd/even items, or an explicit first half. The output identifies the actual members, the two half alphas, correlation of half totals, Spearman–Brown coefficients and Guttman's split-half value. For equally long halves with correlation r, Spearman–Brown is $2r/(1+r)$. Unequal-length correction has a different formula and assumptions; compare it with the named equal-length estimate rather than interchanging them. Negative or undefined estimates should prompt inspection of coding, covariance and the measurement model, not automatic truncation to zero.

The parallel-item model requires equal variances and covariances; the strictly parallel model also requires equal means. Their likelihood-ratio tests use the multivariate normal model. Item-mean, nonadditivity, Friedman and binary Cochran comparisons answer different questions. A significant difference between item means does not by itself invalidate every use of a sum score.

42.2 Intraclass correlations

Specify whether the raters are a random sample or the particular fixed raters of interest, whether agreement or consistency is required, and whether future use involves one rating or an average of k ratings. The result names all three choices. A two-way random-rater interpretation extends to other raters from the same population; a mixed model conditions on the chosen raters.

With target, rater and residual mean squares MS_R, MS_C, MS_E, and n targets, the two-way single-rating coefficients are

$$ICC(C,1) = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E}, \qquad ICC(A,1) = \frac{MS_R - MS_E}{MS_R + (k-1)MS_E + k(MS_C - MS_E)/n}.$$

The corresponding average-rating coefficients are

$$ICC(C,k) = \frac{MS_R - MS_E}{MS_R}, \qquad ICC(A,k) = \frac{MS_R - MS_E}{MS_R + (MS_C - MS_E)/n}.$$

For the one-way model, replace residual variation by within-target variation: $ICC(1,1) = (MS_R - MS_W)/(MS_R + (k-1)MS_W)$ and $ICC(1,k) = (MS_R - MS_W)/MS_R$. The normal random-effects assumptions underlie the intervals. One-way and consistency intervals use their F pivots; agreement intervals use an approximate degrees-of-freedom calculation and are labelled accordingly. Undefined zero-over-zero cases remain undefined.

Worked example. Choose **File > Import data...** and select `sample-data/manual/icc-three-targets.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Survey > Scale reliability (split-half, Guttman, ICC)...**. Select **a** and **b** as **Items**; leave reverse-coded items empty. Set **Intraclass correlation model** to **Two-way random (raters sampled)**, **Intraclass correlation type** to **Absolute agreement**, and **Confidence level** to 0.95. Click **Run analysis** and read the single- and average-measure ICC estimates. Reopen the same menu, select the same items and model, change the type to **Consistency**, and run again. The file contains three targets rated twice, with rows $(1,2), (2,3), (3,4)$. The second rater is always one unit higher. Then $MS_R = 2$, $MS_C = 1.5$ and $MS_E = 0$. Consistency is 1 for either unit, while agreement is $ICC(A,1) = 2/3$ and $ICC(A,2) = 0.8$. Perfect ordering can coexist with a systematic difference that matters when substituting one rater for another. This boundary example illustrates the estimates; it is not a useful sample for assessing interval performance.

42.3 Categorical agreement

Cohen's kappa uses units rated by both of two raters:

$$\kappa = \frac{p_o - p_e}{1 - p_e}, \qquad p_o = \sum_{ij} w_{ij} p_{ij}, \quad p_e = \sum_{ij} w_{ij} p_{i\cdot} p_{\cdot j}.$$

Unweighted agreement uses 1 on the diagonal and 0 elsewhere. Linear or quadratic weights require a declared category order; numeric-looking codes do not by themselves establish equally spaced severity. Choose the interval method explicitly: the large-sample variance and a bootstrap over whole units are different uncertainty calculations.

Fleiss' kappa uses complete sets of ratings with a common number of raters. Krippendorff's alpha allows incomplete units with at least two ratings and uses nominal, ordinal or interval disagreement as selected. An interval metric requires meaningful numeric distances; an ordinal metric requires meaningful order. Bootstrap resampling keeps all ratings of each sampled unit together. Resampling individual ratings would discard within-unit dependence and answer a different question.

For a two-rater table with cells $(40, 10; 5, 45)$, agreement is 0.85. The marginal chance agreement is 0.50, so unweighted kappa is 0.70. Report both agreement and kappa: prevalence and the

marginal distributions affect kappa. If chance agreement is 1, kappa is undefined. None of these coefficients identifies which rater is correct without an external standard.

The model distinctions follow Shrout and Fleiss (1979), McGraw and Wong (1996, including their correction), and Krippendorff's coincidence-based definition of agreement. See the references appendix.

Part V

Time Series

CHAPTER 43

Understanding an ordered series

Time series record measurements in an order whose spacing has meaning. A forecast uses what was available at a cutoff to predict later values. Start by establishing that order and the measurement units, before choosing a model from **Time Series**.

43.1 Declare time, calendar and missingness

Choose a numeric, date or datetime time column, or explicitly use current row order. An explicit axis must be sorted, unique and regularly spaced. The analysis does not silently sort, aggregate or impute it. Missing responses are rejected initially. If omission is selected, the retained time spacing is checked again.

A missing Tuesday measurement is different from a weekend on which the process was never observed. Use a weekday calendar for Monday–Friday data. Compare & Validate and advanced studies additionally support named exchange calendars and time zones, including holidays and special sessions.

Frequency conversion is explicit in Compare & Validate: native, daily, weekly, monthly, quarterly or annual, with first, last, mean or sum aggregation; last is the initial choice. A monthly sum of sales and a monthly last stock price have different meanings. Choose aggregation from the measurement definition.

Changing response clears units, transformations and future scenarios. Changing time clears calendar and vintage settings; changing a predictor clears its future scenario. Switching tables resets the study context. Confirm the complete setup after such a change.

43.2 A small series used throughout these chapters

Choose **File > Import data...** and select `sample-data/manual/time-series-ten.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. The constructed values, in period order, are

Period	1	2	3	4	5	6	7	8	9	10
Value	20	22	21	24	23	25	27	26	29	30

There is an upward movement with short reversals. Ten values are enough to explain calculations but too few for a convincing seasonal study or a rich model.

43.3 Autocorrelation and partial autocorrelation

Choose **File > Import data...** and select `sample-data/manual/time-series-ten.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click

Import, make the new table active, and keep all its rows included and in their imported order. Choose **Time Series > ACF & PACF...**. Set **Series** to **value**, **Time or order** to **(use current row order)**, and **Maximum ACF / PACF lag** to 4. Keep **Missing series values** at **Reject and explain**. Click **Run** and inspect the ACF table. Autocorrelation at lag k relates the series to itself k periods earlier. For the displayed sample ACF,

$$r_k = \frac{\sum_{t=k+1}^{n}(y_t - \bar{y})(y_{t-k} - \bar{y})}{\sum_{t=1}^{n}(y_t - \bar{y})^2}.$$

The common denominator keeps the full-series normalization. It differs from recalculating a separate Pearson correlation over each shortened pair of windows.

For the ten values, $\bar{y} = 24.7$, and lags 0–4 give

$$1, \quad 0.5795205, \quad 0.3858142, \quad 0.1641359, \quad -0.0795205.$$

The pointwise 95% white-noise reference width is about $1.96/\sqrt{10} = 0.61980$. It is not a simultaneous band over the examined lags. A trending series can show serial correlation because its mean changes; it should not be treated automatically as stationary noise.

Partial autocorrelation at lag k describes the remaining linear relationship after accounting for the intermediate lags. ACF and PACF patterns can inform model candidates, but a noisy sample pattern does not uniquely identify an ARIMA order. The default maximum lag is 24, subject to the available sample.

43.4 Stationarity and differencing

Choose **File > Import data...** and select `sample-data/manual/time-series-ten.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Time Series > Stationarity & Differencing...**. Set **Series** to **value**, **Time or order** to **(use current row order)**, **Maximum ordinary differences to test** to 1, and **Stationarity specification** to **Level (constant)**. Leave **Apply one seasonal difference** off and **Missing series values** at **Reject and explain**. Click **Run**. The calculation below explains the first-difference series assessed by this run; ten observations provide very limited information for stationarity tests. The augmented Dickey–Fuller (ADF) test has a unit-root null. KPSS has a level- or trend-stationarity null, depending on the selected deterministic specification. Their different nulls are useful; they are not two interchangeable votes for the same hypothesis.

HypoNum labels a differencing candidate when ADF does not reject a unit root and KPSS rejects the chosen stationarity null at 5%. That is evidence relevant to differencing, not proof of random-walk increments or a guarantee that forecasts improve.

Ordinary first differences are $\Delta y_t = y_t - y_{t-1}$. For the imported values

$$20, 22, 21, 24, 23, 25, 27, 26, 29, 30,$$

they are

$$2, \ -1, \ 3, \ -1, \ 2, \ 2, \ -1, \ 3, \ 1.$$

Their mean is $10/9$, the change from first to last value divided by nine intervals. A seasonal difference is $y_t - y_{t-m}$, where m is the declared season length. Seasonal and ordinary differences express different patterns and may be used together.

The setup examines up to two ordinary differences initially, with an optional seasonal difference. Results show the original and each supported stage and recommend the smallest meeting the

combined criterion. Review all stages. Excess differencing can introduce avoidable noise. **Create new column** explicitly creates a live calculated difference column, with missing leading values. It recalculates with source edits. No column is created when the original already meets the criterion or no supported transformation is available.

43.5 Seasonal decomposition

Seasonal Decomposition & Adjustment separates an observed series into trend, seasonal components and remainder using robust STL/MSTL. Enter the seasonal periods, separated by commas, and choose robust estimation as appropriate. Multiple periods must represent plausible, adequately observed cycles.

A decomposition is descriptive. The remainder need not be independent noise, and a smooth trend is not itself a forecast. Endpoint estimates can change when another cycle becomes available; the result discloses that revision. For validation, a decomposition used as preprocessing must be re-estimated using only each training window.

CHAPTER 44

Fitting and interpreting forecasts

A forecast combines a model, a cutoff, a horizon and information available at that cutoff. Its interval describes future observations under the model assumptions; it is not a confidence interval for a coefficient.

44.1 Begin with meaningful benchmarks

Open **Forecast** under **Time Series**. Choose the response, time order, model, horizon, holdout and interval level. Initial horizon and holdout are both 12; confidence is 95%. For the ten-value teaching series, select current row order, horizon 3, holdout 0 and a specified model.

Model	Forecasting idea
Naive	Repeat the last observed value: $\hat{y}_{T+h} = y_T$.
Seasonal naive	Repeat the most recent observed value for that season; the season length must be declared.
Mean	Predict the training mean at every horizon.
Drift	Continue the average first-to-last change: $\hat{y}_{T+h} = y_T + h(y_T - y_1)/(T-1)$.
Simple exponential smoothing	Update a level from recent observations with declining weights on older information; forecasts have no explicit trend.
Holt	Estimate level and trend; damped Holt reduces the projected trend as horizon grows.
Additive Holt–Winters	Combine level, trend and an additive seasonal pattern.
ARIMA / seasonal ARIMA	Model dependence after the specified ordinary and seasonal differences.

A benchmark is useful even when it is scientifically simple. A more complicated fit should earn its complexity in an appropriate evaluation, not merely describe training data more closely.

44.2 Worked example: three different assumptions

Choose **File > Import data...** and select `sample-data/manual/time-series-ten.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Time Series > Forecast...**. Set **Series** to **value**, **Time or order** to **(use current row order)**, future periods to 3, **Assessment holdout** to 0, **Forecast interval** to **95%**, and **Missing series values** to **Reject and explain**. Choose **Naive (last value)** in **Forecasting model** and click **Forecast**. Inspect the forecasts for periods 11–13. Reopen **Time Series > Forecast...** and repeat these settings with

Historical mean, then with **Drift**, clicking **Forecast** for each model. Each fit uses all ten observations. For the ten-value series, Naive predicts 30 at periods 11–13. Mean predicts 24.7 at all three. Drift estimates change $10/9 = 1.111111$ per period and predicts 31.111111, 32.222222 and 33.333333.

Model	Period 11	Period 12	Period 13
Naive	30.0000	30.0000	30.0000
Mean	24.7000	24.7000	24.7000
Drift	31.1111	32.2222	33.3333

The drift fit follows the rising pattern; the mean model assumes a stable common mean. This explains their difference without declaring a winner from ten already observed values. With confidence 95%, the period-11 intervals are $[26.1905, 33.8095]$ for Naive, $[16.7875, 32.6125]$ for Mean, and $[27.6165, 34.6057]$ for Drift. At period 13 the drift interval widens to $[26.7028, 39.9638]$.

44.3 What the prediction intervals assume

For the Mean model, independent $N(\mu, \sigma^2)$ observations give

$$\bar{y} \pm t_{1-\alpha/2,n-1}\, s\sqrt{1 + 1/n},$$

with sample variance divisor $n - 1$. The extra 1 represents variation in a new observation; $1/n$ represents uncertainty in the estimated mean. The interval has the same width at every horizon. Exact normal-model coverage does not carry over to autocorrelated observations or a model selected after inspecting the data.

Drift fitted values are one-step predictions using the preceding value and estimated drift. Its forecast-error scale grows with

$$\sqrt{h\left(1 + \frac{h}{T-1}\right)}.$$

Both accumulated innovation variation and drift estimation matter. Other benchmark intervals use normal approximations. All displayed intervals are pointwise rather than simultaneous statements about the entire future path.

44.4 Fit ARMA or ARIMA explicitly

Choose **Fit ARMA / ARIMA** to set p, d, q. Here p counts autoregressive lags, d ordinary differences and q moving-average innovation lags. With $d = 0$, the model is ARMA. Seasonal ARIMA additionally specifies seasonal orders P, D, Q and period m.

For an AR(1) around a stable mean,

$$y_t - \mu = \phi(y_{t-1} - \mu) + \varepsilon_t.$$

An illustrative $\phi = 0.6$ makes a deviation's expected size after h periods equal to 0.6^h times its current size. If innovation variance is 4, the stationary marginal variance is $4/(1 - 0.6^2) = 6.25$. This is model arithmetic, not an estimate for the teaching series.

The deterministic term can be none, constant or drift where supported. A constant is available with total differencing $d + D = 0$; drift with $d + D = 1$. Automatic settings retain a constant for an undifferenced series. Changing deterministic terms changes the model.

Fits use Gaussian maximum likelihood in state-space form with exact diffuse initialization for integrated states. The reported likelihood for a complete series corresponds to the differenced-series likelihood convention, retaining parameter-dependent contributions. AIC counts fitted ARMA, deterministic and innovation-variance parameters; BIC uses the effective observation count after integration. Historical saved fits retain their original estimates and method record.

44.5 Automatic selection and diagnostics

Automatic selection initially considers the listed benchmarks, smoothing methods, ARIMA and seasonal ARIMA. A temporal holdout ranks successful candidates by MASE where available, then MAE and RMSE, with AIC/BIC shown where defined. The holdout chooses the model; it is therefore not an untouched final evaluation sample.

Nonconverged candidates remain visible with warnings. In this edition the automatic workflow can select one when no converged candidate is available. Treat that output as unresolved numerical evidence, not an accepted final forecast.

Results separate Summary, Charts, Parameters, Diagnostics, Validation and Warnings. Inspect observed/fitted values, holdout and future predictions, intervals, residual ACF/PACF and residual tests. Ljung–Box assesses remaining serial dependence; Jarque–Bera concerns normality; ARCH-LM concerns conditional variance patterns. A nonsignificant diagnostic does not prove a model adequate.

Ljung–Box omits state-space burn-in values and adjusts degrees of freedom for fitted ordinary and seasonal AR/MA terms. The lag table initially emphasizes the twelve strongest nonzero lags outside its reference band; **Show all lags** and **Export all lags** recover the complete table.

CHAPTER 45

Evaluating and preserving a forecast

A model that follows the observed history closely may forecast poorly. Evaluation asks what it would have predicted at earlier cutoffs, using only information available at those cutoffs. Keep model development, a final assessment and the forecast for future use distinct.

45.1 Plan the evaluation before comparing models

Choose **Time Series > Compare & Validate…**. Set the initial training length, forecast horizon and interval between forecast origins. The initial settings are 30 training observations, a three-period validation horizon and an origin step of one. An expanding window retains all preceding history; a sliding window retains a fixed recent training length.

Reserve the final observations before choosing transformations or models. The initial final-evaluation length is six. A period used to choose a model is development data, even if its controls call it a holdout. Changing a model after seeing the final period uses that period for development too.

At each origin the procedure fits the requested model to its training history and forecasts the following horizon. Overlapping horizons are useful, but their errors are dependent: the number of forecast cells is not the number of independent replications. Inspect performance separately by horizon. A model useful one month ahead need not be useful twelve months ahead.

45.2 Read accuracy in the response's units

For errors $e_i = y_i - \hat{y}_i$ over N forecast cells,

$$\mathrm{MAE} = \frac{1}{N}\sum_i |e_i|, \qquad \mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_i e_i^2}.$$

MAE describes average absolute error. RMSE weights larger errors more heavily. Both retain the response unit; compare them on the same response definition, origins and horizons.

The mean absolute scaled error (MASE) divides each absolute error by a training-history benchmark scale before averaging. The calculation in this edition uses the lag-one scale

$$s_o = \frac{1}{T_o - 1}\sum_{t=2}^{T_o} |y_t - y_{t-1}|$$

within origin o's training window of length T_o. This is a nonseasonal scaling convention, including when a seasonal model is fitted. A seasonal-naive forecast can separately appear as a benchmark in the results.

Important

A constant training history gives a zero MASE scale. In this edition, cells without a positive scale are omitted from the scaled-error average. Its MASE can therefore describe fewer

cells than its MAE. Review the origins and compare MAE on a common set of observations when scaled errors are unavailable. Do not treat a partial-origin MASE as a score for the complete planned evaluation.

Coverage is the fraction of forecast cells whose observed value falls inside the reported prediction interval. In this edition the denominator includes every evaluated forecast cell. Review unavailable intervals and the forecast table before interpreting it as coverage of usable intervals. Coverage on one short, overlapping evaluation is descriptive; it does not establish that a nominal 95% procedure attains 95% long-run coverage.

45.3 Follow a small rolling evaluation

Choose **File > Import data...** and select `sample-data/manual/time-series-ten.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Time Series > Compare & Validate...**, set **Series** to **value**, **Forecasting model** to **Naive (last value)**, and **Time or order** to **(use current row order)**. Set the future horizon to three, the interval to 95%, an expanding window, initial length five, validation horizon two, step one and a reserved final block of two observations (**Untouched final evaluation** $= 2$). Use **Target transformation: None**, annualization 0, and **Missing series values: Reject and explain**. Click **Run** and inspect the rolling-origin predictions, pooled metrics and final evaluation. The first eight observations form development history.

Training ends at row	Horizon	Forecast	Actual	Error
5	1	23	25	2
5	2	23	27	4
6	1	25	27	2
6	2	25	26	1

The pooled MAE is $9/4 = 2.25$ and RMSE is $\sqrt{25/4} = 2.50$. The two training scales are $7/4 = 1.75$ and $9/5 = 1.80$, giving

$$\text{MASE} = \tfrac{1}{4}(2/1.75 + 4/1.75 + 2/1.80 + 1/1.80) = 1.27381.$$

Each origin contributes two forecast cells. The repeated actual value 27 is forecast from two different information sets.

After the procedure is fixed, the naive model fitted through row eight forecasts 26 for rows nine and ten. Their observed values are 29 and 30, so final-period MAE is 3.50. A later fit through all ten rows forecasts 30 for each of the three future periods. That later fit is distinct from the fit assessed on the reserved block; its use of rows nine and ten does not improve the earlier assessment retrospectively.

45.4 Fit transformations inside each training history

Compare & Validate supports log transformation, differencing, simple returns, log returns, percentage growth, annualization, detrending and within-window seasonal adjustment. Each

transformation is fitted using the corresponding training history. A trend or seasonal pattern estimated once from the complete series would leak later information into earlier forecasts.

Forecasts return to original units. Differences and returns require the last observed level and, in general, a joint sequence of future increments. Adding or multiplying separately chosen marginal interval endpoints does not give a valid interval for the resulting level.

For a log-normal forecast, a log-scale mean of $\log 100$ and log-scale standard deviation 0.2 imply an original-scale median of 100 and mean $100\exp(0.2^2/2) = 102.02$. The distinction answers two different questions. Select the requested original-scale centre deliberately.

Some transformations allow an exact marginal median; more complicated compositions require simulated joint forecast paths. The median at each horizon is a marginal summary, not one jointly most likely trajectory. The mean averages reconstructed levels. The calculation checks numerical precision of the reconstruction and warns when its target precision is not established. A reproducible seed does not by itself establish numerical accuracy.

Annualized simple returns and growth rates may require a fractional power when returning to the original scale. Gaussian predictive uncertainty can place values below the supported lower boundary. Such a reconstruction is refused when it is undefined. Log-return annualization is linear and avoids that particular domain problem. Even an unannualized simple-return model can imply negative future levels; examine whether the model suits the quantity.

45.5 Use information that was available at the origin

Revised economic data need more than a measurement date. Publication-date and optional real-time start/end columns describe when a value became available. A vintage-aware evaluation reconstructs the history available at its cutoff. A publication delay may require a forecast across a gap before the scored target period begins.

The reported evaluation uses the latest supplied vintage for the actuals being assessed. Preserve the data vintage with the study. Do not describe a forecast based on later revised predictors as one that could have been issued in real time.

Future predictor scenarios have their own availability dates. An archived scenario available by an origin can support an honest historical conditional forecast. Realized future predictor values answer a conditional hindsight question; label that distinction.

45.6 Review warnings, then save the complete study

Failed origins remain visible with their reasons. Inspect convergence and diagnostic warnings at the origin level, not only on the final fit. An output table from an unresolved fit does not establish a usable numerical forecast.

Save as Validated Study carries a direct fit into the evaluation workflow. The command name does not certify predictive accuracy. **Saved Studies** reopens the response, model, transformations, units, calendar, recorded data identity and results. Reproduction checks that the supplied data match the recorded identity.

Use **Save Study** to retain the specification, **Add to Report** to include selected results, and **Export** for the underlying tables. Validation exports include rolling forecasts, final-evaluation forecasts and future forecasts as separate CSV outputs, along with diagnostics and reproducibility information.

Before comparing saved studies, check target definition, units, horizon, training-window rules and evaluation population. The application flags incompatible settings; a common chart alone does not make two scores comparable.

CHAPTER 46

Predictors, multiple series and changing volatility

46.1 Conditional forecasts with predictors

Choose **Time Series > Dynamic Regression…** when the forecast question includes other measured series. Specify response, time axis, predictors and the future predictor path. Available features include levels, lags, rolling means and rolling standard deviations, with additional distributed lags and window lengths.

A simple model might be

$$y_t = \beta_0 + \beta_1 x_t + n_t, \qquad n_t \text{ follows an ARIMA model.}$$

The predictor explains a conditional mean; the error model addresses remaining serial dependence. Adding predictors does not remove the need to inspect residuals or justify their forecast-time availability.

Suppose a fitted conditional mean is $10 + 2x$. A future predictor value of 3 gives conditional response mean 16; a scenario value of 5 gives 20. These are two different scenarios, not competing estimates of one known future outcome. The arithmetic illustrates interpretation; it is not a fitted example.

With a supplied future path alone, intervals condition on that path. If future predictor standard errors are supplied, enable uncertainty draws to propagate the supported predictor uncertainty; the initial draw count is 500. That propagation still depends on the uncertainty model and does not account for every possible scenario or misspecification. Record known versus estimated predictor values, units, assumptions and the seed when draws are used.

Build lagged and rolling features using information available at each forecast origin. A future rolling predictor can require multiple future input values. Supply the complete required path; do not substitute later observed values without changing the interpretation to a hindsight conditional evaluation.

46.2 Joint dynamics: VAR and VECM

Choose **Time Series > Multivariate Models — VAR / VECM…** for several series observed on the same regular time axis.

A vector autoregression of order p has the form

$$\mathbf{y}_t = \mathbf{c} + A_1 \mathbf{y}_{t-1} + \cdots + A_p \mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t.$$

Every component can depend on lagged values of every component. The number of coefficients grows quickly with the number of series and lags. A short history may not support the desired model.

If integrated series have stable long-run combinations, a vector error-correction model can express both that relation and short-run changes:

$$\Delta \mathbf{y}_t = \boldsymbol{\alpha}\boldsymbol{\beta}^{\top}\mathbf{y}_{t-1} + \sum_{j=1}^{p-1} \Gamma_j \Delta \mathbf{y}_{t-j} + \text{deterministic terms} + \boldsymbol{\varepsilon}_t.$$

The cointegration rank is the number of independent long-run relations. Select it explicitly or review Johansen-based selection under the chosen deterministic terms. A different constant or trend convention can change the rank assessment.

The lag-order control denotes the corresponding VAR order in levels. Thus VAR order one gives zero lagged differences in the VECM. Review the reported convention instead of adding a difference lag by habit.

Inspect stability, residual whiteness and normality diagnostics before interpreting forecasts. Ordered impulse responses depend on the ordering of the variables; the initial band calculation uses 500 draws. An ordering should reflect a defensible identification assumption, not a search for the most attractive response. Granger results concern incremental predictability under the model; they do not establish causation.

46.3 Volatility: GARCH

Choose **Time Series > Volatility Models — GARCH…** for a series whose conditional variation changes over time, commonly returns rather than raw price levels.

Write $y_t = \mu + \varepsilon_t$ with $\varepsilon_t = \sigma_t z_t$. The conditional variance follows

$$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2.$$

Here p counts lagged squared innovations and q counts lagged conditional variances. The initial orders are $(1, 1)$. Choose Gaussian or standardized Student-t innovations and a constant or zero mean. The Student-t choice permits heavier tails; it does not make every outlier compatible with the fitted process.

For a GARCH(1,1) illustration with $\omega = 0.1$, $\alpha = 0.1$, $\beta = 0.8$, previous squared innovation 4 and previous conditional variance 1, the next conditional variance is $0.1 + 0.1(4) + 0.8(1) = 1.3$. The corresponding conditional standard deviation is $\sqrt{1.3} = 1.1402$ in the response's units. This is a model calculation, not a forecast of the response mean.

Numerical fitting may scale the data internally; reported volatility is returned to the declared return unit. Returns expressed as fractions and percentages therefore differ by a factor of 100 in standard deviation and 10,000 in variance. Review that unit before comparing outputs.

The fit initializes variance with a weighted backcast of early squared innovations, held fixed during estimation. Each evaluation window uses its own training data for that initialization. Robust covariance describes coefficient uncertainty under the fitted specification; it does not certify the distribution of every future tail event. Inspect convergence, standardized residuals and remaining ARCH effects.

46.4 A unit root with one possible break

Choose **Time Series > Unit Root with a Break — Zivot–Andrews…** to assess a unit-root null against a stationary alternative with one allowed break. Choose a break in the constant, trend, or both.

The calculation uses Baum's approximation to the Zivot–Andrews procedure. It chooses the augmentation lag once by AIC in a no-break ADF regression with constant and trend, then uses that lag for all candidate breaks. The maximum candidate lag is

$$\min\left(\left\lceil 12(n/100)^{1/4}\right\rceil, \lfloor n/2\rfloor - 3\right).$$

A common comparison sample is used in that lag selection. Candidate break positions cover the central 70% of the observations (trim fraction 0.15). The statistic is the smallest lagged-level t statistic over these positions. Critical values and p-values interpolate a simulated reference table.

Important

Read the reported break row with the selected break terms. For a break in the constant, or in constant and trend, the reported row is the last row before the break terms become active. For a trend-only break, the trend-break term is already active at the reported row. This index convention affects how the date is described.

The original procedure selects the lag separately at each candidate break; that is not the approximation used here. A selected break is not a general change-point analysis, evidence of several breaks, or a causal explanation of an event. Review the series, specification and substantive timing together.

46.5 Sources and further study

Hyndman and Athanasopoulos, *Forecasting: Principles and Practice*, 3rd edition, develops benchmark forecasts, residual diagnosis, forecast evaluation, ARIMA and dynamic regression with worked examples. Box, Jenkins, Reinsel and Ljung, *Time Series Analysis: Forecasting and Control*, 5th edition, develops the underlying ARIMA modelling framework. Lütkepohl, *New Introduction to Multiple Time Series Analysis* (2005), treats VAR, VECM and impulse-response interpretation.

Bollerslev, *Generalized autoregressive conditional heteroskedasticity* (1986), *Journal of Econometrics* 31, pp. 307–327, introduces GARCH. Zivot and Andrews, *Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis* (1992), *Journal of Business & Economic Statistics* 10, pp. 251–270, develops the break test. The lag-selection approximation and row convention described above are part of this manual's calculation definition; do not infer that every implementation of the named test uses the same convention.

46.6 Deterministic trend analysis

Time Series > Trend analysis... fits a trend to positions $t = 1, \ldots, n$ in table order or a chosen ordering column. Excluded or missing responses retain their positions, so removing a value does not compress elapsed index time.

Linear and quadratic models use original-scale least squares. Log-linear exponential growth fits $\log y = a + bt$ and displays $e^a(e^b)^t$; with symmetric log errors this is a median trend without bias adjustment. The original-scale exponential instead minimizes squared response residuals for $y = ce^{dt}$, a different mean model under additive errors. Log-linear fitting refuses zero/negative responses.

The logistic S-curve $y = K/(1 + Ae^{-rt})$ estimates positive K, A; r permits growth or decay. Early exponential-looking observations may not determine an asymptote. An unresolved fit is refused. For a fitted curve, automatic extrapolation is withheld when asymptote relative SE exceeds 0.25 or covariance is unavailable. This screen is an operational caution, not a test or a guarantee of successful extrapolation below the threshold.

MAE, MSE, RMSE and MAPE describe in-sample original-scale errors. Zero observed responses make MAPE unavailable. For observed $(2, 4, 5)$ and fitted $(1, 4, 7)$, MAE is 1, MSE is $5/3$ and RMSE is 1.291; these are not held-out forecast errors. Extrapolation extends the deterministic curve without prediction intervals. Use Forecast for a stated stochastic prediction model. Adding a detrended column writes row-aligned residuals $y - \hat{y}$, leaving unsupported rows blank, with the recorded calculation retained for replay.

CHAPTER 47

Cross-correlation

Cross-correlation asks whether changes in one ordered series are associated with earlier or later values of another. It is useful for exploring delays, but a peak does not establish a causal effect. Shared trends, seasonal patterns and serial dependence can all produce a strong association.

Choose **Time Series > Cross-correlation…**. The table must expose at least two numeric columns.

47.1 Choose the series and their time axis

Select the first series X and second series Y deliberately: their order determines the sign of the lead–lag statement. The setup initially selects the first two numeric columns, current row order, a strict calendar and maximum lag 30.

With a time column, keys must be present, unique, sorted and regular. The weekday calendar additionally excludes weekend observations. The procedure does not interpolate time keys. Without a time column, one lag means one table-row position.

Each series may be left unchanged or first-differenced; both initially use no differencing. Prewhitening is initially off. If enabled, choose an AR order from 1 to 20; the initial order is one. The maximum lag is bounded by series length and a ceiling of 250.

47.2 Interpret the sign and overlap

At signed lag k, HypoNum pairs x_t with y_{t+k} and calculates the Pearson correlation over the valid overlapping pairs. A positive peak at $k = d$ is labelled X leads Y by d: the first series is associated with the second series d periods later. A negative peak at $k = -d$ means Y leads X by d. Lag zero has no lead–lag separation.

Each lag uses its own paired means and standard deviations. Edge observations leave the overlap as the lag grows. Lags with fewer than three valid pairs, or with a constant variable within that overlap, have no reported correlation. An omitted correlation is not zero. Inspect overlap counts alongside the coefficients.

47.3 Reproduce a one-period delay

Choose **File > Import data...** and select `sample-data/manual/cross-correlation-eight.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Time Series > Cross-correlation...**. This constructed dataset makes the alignment visible:

Row	X	Y
1	2	0
2	−1	2
3	3	−1
4	0	3
5	4	0
6	−2	4
7	1	−2
8	5	1

Choose columns **x** and **y**, current row order, maximum lag two, no differencing and no prewhitening. Click **Run cross-correlation** and inspect the lag table's coefficient and overlap count at lags 0 and +1. At lag +1 the seven paired values are identical:

$$(x_1, \dots, x_7) = (y_2, \dots, y_8) = (2, -1, 3, 0, 4, -2, 1).$$

The reported correlation is 1.000 with seven pairs, and the peak label says that x leads y by one. At lag zero the correlation is −0.588665; alignment, rather than a new dataset, changes the question being answered.

The example demonstrates the sign convention. Its deliberately copied values do not establish a predictive relationship in an observed process.

47.4 Treat reference bands as a limited guide

The displayed nominal band at a lag with m overlapping pairs is $\pm 1/\sqrt{m}$. At the example's lag +1, it is ± 0.377964. **It is not a 95% confidence band.** Its white-noise interpretation is inappropriate for strongly autocorrelated series, and inspecting many lags adds a multiple-comparison problem. A line outside the band is a prompt to investigate.

When prewhitening is enabled, each differenced series receives its own OLS AR(p) fit on a common consecutive block. Correlations then use the two residual series. This convention differs from a procedure that fits a filter to X and applies that same filter to both series. Choose the convention for the scientific question, and retain the transformation settings with the result.

47.5 Keep gaps in their original positions

Excluded rows and rows with missing or non-finite values remain unavailable positions. The series does not close up around a gap. Thus a lag of two always means two original positions, even when an intermediate observation is unavailable.

First differencing needs two adjacent available values. Its first position and differences adjacent to a missing value are unavailable. For prewhitening, unavailable end positions shorten the common block; an interior gap refuses the autoregressive fit because it would break the consecutive-time assumption. Calendar validation still checks every row's time key.

The result distinguishes analyzed, missing, excluded and transformation-unavailable rows. Check that those counts explain the source population and inspect the lag-specific overlap separately. If no lag has enough information, the result reports insufficient data.

The workflow accepts at most one million source rows and twenty million selected cells. It refuses an unsupported size instead of silently truncating a series.

Part VI

Quality

CHAPTER 48

Control charts

A control chart compares an ordered process with a reference pattern of variation. It helps answer a practical question: *does the next result look consistent with the process we intended to monitor?* A signal marks a reason to investigate. The chart does not identify the cause, and its limits do not define acceptable product.

This chapter takes a study from chart selection and baseline definition through calculation, interpretation and a recorded response. It covers I–MR, subgroup mean and dispersion charts, attribute charts, and CUSUM/EWMA. The worked examples use complete synthetic datasets so that the arithmetic and the desktop settings can be checked together.

48.1 Choose the question and the chart

A line filling packages might change its average fill, become more variable, or produce more packages below a declared requirement. These are different outcomes. A chart of measured fill monitors a continuous variable; a chart of nonconforming packages monitors a classification that depends on the requirement and inspection rule. Keep the definition and measurement units constant through the study.

Observation at each point	Starting chart
One continuous measurement	Individuals and moving range (I–MR).
A short-time sample of measurements	X-bar/R for smaller subgroups; X-bar/S for larger subgroups.
Nonconforming units out of inspected units	p for fractions; np for counts at constant sample size.
Defects, with more than one possible per unit	c at constant exposure; u when exposure varies.
Attribute rates with additional subgroup variation	Laney p′ or u′, after investigating the origin of that variation.
Opportunities or elapsed time through the next rare event	g for whole opportunities; t for continuous elapsed time.
Continuous measurements, with small sustained shifts of interest	CUSUM or EWMA, with a stated target and process sigma.

For a variables chart, examine dispersion as well as location. A quiet mean chart can accompany a change in spread. A nonconforming *unit* is counted once on a p or np chart, even if it has several defects. A c or u chart counts those defects individually.

48.1.1 Control limits and specifications answer different questions

Control limits describe variation under the chart's reference model. Specification limits express an external requirement. A stable process can consistently make unacceptable product, and an unstable process can temporarily remain within specification. HypoNum may overlay the

applied lower and upper specifications (LSL and USL), but these lines neither replace control limits nor generate control-rule violations.

48.1.2 What makes a useful baseline?

In a retrospective, *Phase I* study, examine a historical period to understand variation, investigate signals and decide which operating conditions the reference should represent. In *Phase II* monitoring, compare subsequent observations with that established reference. Do not continually include a developing shift in the data used to estimate its own limits: the reference can then move with the change.

A baseline must represent the process, measurement system and sampling practice intended for monitoring. The ability to compute limits from a small sample does not make that sample adequate. The short examples in this chapter teach calculations; they do not qualify an industrial baseline or establish a false-alarm rate. See Montgomery's treatment of Phase I and Phase II studies (reference [1]).

Variables-chart limits rely on a suitable within-process variation model, conventionally independent normal observations. Time dependence, a skewed distribution, changing operating conditions or measurement rounding can alter their performance. Investigate these features before interpreting a pattern as a change in the process mean.

48.2 Prepare the table and set up the analysis

The current desktop path is **Quality > Statistical process control > Assess stability / control charts...**. Select **Variables (measurements)**, **Attributes (counts)**, or **Time-weighted (CUSUM / EWMA)** in **Chart family**.

1. Put the table in process order. **Time or order (optional)** supplies labels; selecting a timestamp does *not* sort the rows. Use unique, meaningful labels when you will define stages or investigate specific observations.
2. Select **Process variable (Y)** for measurements, or the count and denominator fields for attributes. Check units, coding and exclusions.
3. Define the sampling structure. For variables, choose **Individual measurements (I-MR)**, **Constant subgroup size**, or **By grouping column**.
4. To estimate limits from a reference period, enable **Calculate control limits from a historical subset** under **Historical baseline (Phase 1)**. Choose the column, comparison and value.
5. Choose the run rules or the time-weighted parameters, then select **Run analysis**. Read the baseline and omitted-data disclosures before interpreting the plot.

Without a historical subset, all usable included observations estimate the reference. With a subset, the other usable observations remain plotted and are compared with that reference. Selecting a baseline is therefore different from excluding rows from the analysis.

Text baseline fields allow equality or inequality. Ordered fields also allow $<$, $\leq$, $\geq$, $>$ and an inclusive **Between** range. Enter dates as `YYYY-MM-DD`. A subset with no usable baseline observations is refused; individual chart families have further requirements for estimating variation.

48.2.1 Subgroups should reflect how the process was sampled

A rational subgroup collects measurements obtained under sufficiently similar conditions that its within-group variation represents the short-time noise of interest. For example, five packages

taken close together each hour can reveal changes between hours. Combining one package from each of five shifts into a single subgroup answers a different question and can hide the change being sought.

The constant-size mode defaults to five observations and allows sizes 2–25. Original consecutive row positions define the blocks before missing measurements are removed. A block that is incomplete is dropped and counted; a later observation is not pulled into that block. Sizes 2–10 select X-bar/R; sizes 11–25 select X-bar/S.

With a grouping column, equal labels form a subgroup, in order of first appearance. Labels reused far apart in the table are still combined. Usable groups need at least two measurements and cannot exceed 25. The integer part of the median usable group size selects R (2–10) or S (11–25); a median of 10.5 therefore selects R. Different usable sizes produce different control limits. A subgroup contributes to the baseline only when all its retained measurements belong to the historical subset.

Important

Confirm the analyzed order and the omitted observations. An I–MR chart removes non-finite measurements before pairing retained values; a missing reading can therefore make the values on either side adjacent. Constant-size subgroups preserve their original block boundaries instead. These are different data-handling rules.

48.3 Individuals and moving range

Use I–MR when each sampling occasion supplies one continuous measurement. The Individuals chart shows location; the Moving Range chart shows the absolute change between successive usable measurements. It estimates short-time variation from pairs, not from the overall standard deviation of a potentially shifting series.

Let $\bar{x}_B$ be the mean of the baseline measurements and $\overline{MR}_B$ the baseline mean moving range. For a new desktop analysis, a baseline moving range uses adjacent retained measurements only when *both* belong to the baseline. A finite non-baseline measurement between two baseline measurements breaks that pairing. If no eligible pair remains, the limits cannot be estimated. The displayed moving ranges still cover the complete retained sequence.

With $d_2 = 2/\sqrt{\pi}$ and $d_3 = \sqrt{2 - 4/\pi}$, the calculations are

$$\hat{\sigma} = \frac{\overline{MR}_B}{d_2}, \qquad LCL_X,\ UCL_X = \bar{x}_B \mp 3\hat{\sigma}, \tag{48.1}$$

$$CL_{MR} = \overline{MR}_B, \qquad LCL_{MR} = 0, \tag{48.2}$$

$$UCL_{MR} = \left(1 + \frac{3d_3}{d_2}\right)\overline{MR}_B. \tag{48.3}$$

The factors are approximately 2.65868 for the Individuals half-width as a multiple of $\overline{MR}_B$, and 3.26653 for the MR upper limit. HypoNum uses unrounded factors; tables printed as 2.66 and 3.267 can give slightly different last digits.

48.3.1 Worked example: a change in shaft diameter

Choose **File > Import data...** and select `sample-data/manual/spc-individuals.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported

order. Choose **Quality > Statistical process control > Assess stability / control charts...**. A technician records one diameter per sampling period, in millimetres. Periods 1–20 supply the teaching baseline; periods 21–26 are subsequent observations. All 26 rows are shown below; the two column groups continue the same sequence.

Period	Diameter (mm)	Baseline	Period	Diameter (mm)	Baseline
1	10.00	1	14	9.99	1
2	10.03	1	15	10.02	1
3	9.98	1	16	9.96	1
4	10.01	1	17	10.01	1
5	9.99	1	18	10.00	1
6	10.04	1	19	10.04	1
7	10.00	1	20	9.97	1
8	9.97	1	21	10.02	0
9	10.02	1	22	10.04	0
10	10.01	1	23	10.09	0
11	9.98	1	24	10.10	0
12	10.00	1	25	10.11	0
13	10.03	1	26	10.12	0

Settings. Choose **Variables (measurements)** and **Individual measurements (I-MR)**. Set the process variable to `diameter_mm`, the order label to `period`, and the historical subset by enabling **Calculate control limits from a historical subset**, choosing **baseline**, equality, and value 1. Clear Rules 2–4 and leave Rules 5–8 off, so only Rule 1 is active. Keep all rows included and apply no specification overlays. Select **Run analysis**.

Calculation. The 20 baseline measurements sum to 200.05 mm. The 19 eligible moving ranges sum to 0.69 mm. Consequently,

$$\bar{x}_B = 10.0025, \qquad \overline{MR}_B = \frac{0.69}{19} = 0.0363157895, \qquad \hat{\sigma} = 0.0321840305 \text{ mm}.$$

The following values are rounded for display. Decisions use the unrounded quantities.

Chart (mm)	Lower limit	Center	Upper limit
Individuals	9.905948	10.002500	10.099052
Moving Range	0.000000	0.036316	0.118627

Read the result. Periods 24, 25 and 26 exceed the Individuals upper control limit. Period 23, at 10.09 mm, is close to the limit but remains below it. No moving range exceeds its upper limit. The sequence is therefore consistent with a developing change in location that does not require an unusually large single step. A quiet MR chart does not cancel the Individuals signals.

The pattern alone cannot establish a mechanical cause. Check the measurement method, adjustments, material and operating conditions around the onset of the change. Evaluate product conformance using the applicable specifications separately.

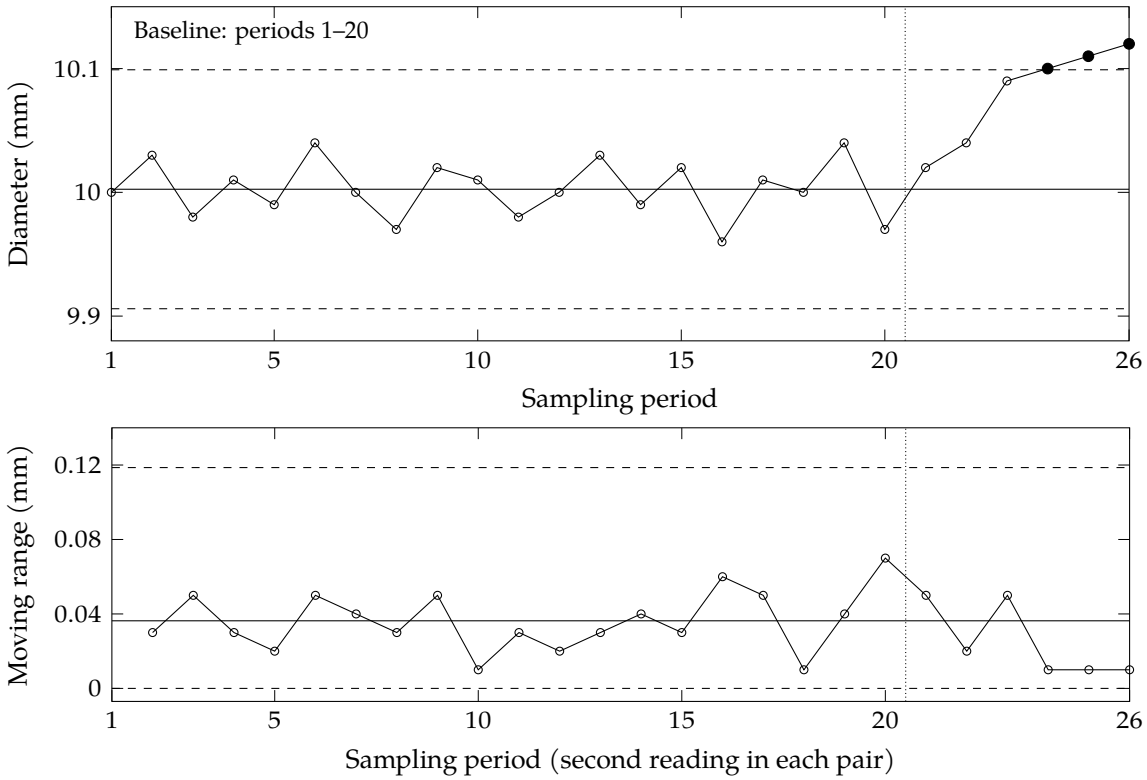


Figure 48.1. The same observations viewed for location and successive variation. Solid horizontal lines are centers; dashed lines are control limits. The vertical dotted line ends the baseline. Filled points exceed a limit. Explanatory illustration drawn from the example calculations; the application's plot styling differs.

An example of a defensible report. Using periods 1–20 as the reference, the estimated center was 10.0025 mm and the Individuals limits were 9.905948–10.099052 mm. With Rule 1 alone, periods 24–26 signaled above the upper limit. The moving-range chart had no limit breach. The later observations require investigation before the reference is revised.

Note

The dialog normally enables Rules 2–4 as well. This example deliberately disables them to isolate a point-beyond-limits decision. With Rule 3 enabled, the strictly increasing sequence ending at period 25 also signals. Record the chosen rules whenever results are compared.

48.4 Subgroup means and dispersion

When each sampling occasion produces several measurements, a subgroup chart separates variation *within* that occasion from differences *between* occasions. The X-bar chart plots subgroup means. Its companion plots subgroup ranges (R) or sample standard deviations (S). The subgroup definition is part of the statistical design, not merely a way of reducing the number of plotted points.

48.4.1 How the limits are calculated

For m baseline subgroups with sizes n_j and means $\bar{x}_j$, the center of the mean chart is the observation-weighted grand mean

$$\bar{x}_B = \frac{\sum_{j=1}^{m} n_j \bar{x}_j}{\sum_{j=1}^{m} n_j}.$$

For an R chart, R_j is the largest minus the smallest measurement in group j. Let $d_2(n)$ and $d_3(n)$ be the mean and standard deviation of the range of n independent standard normal observations. HypoNum estimates one common within-process standard deviation:

$$\hat{\sigma}_R = \frac{1}{m} \sum_{j=1}^{m} \frac{R_j}{d_2(n_j)}.$$

At a plotted subgroup of size n_i, the mean limits and range lines are

$$LCL_{\bar{X},i},\ UCL_{\bar{X},i} = \bar{x}_B \mp \frac{3\hat{\sigma}_R}{\sqrt{n_i}}, \tag{48.4}$$

$$CL_{R,i} = d_2(n_i)\hat{\sigma}_R, \tag{48.5}$$

$$LCL_{R,i} = \max\{0, d_2(n_i) - 3d_3(n_i)\}\hat{\sigma}_R, \tag{48.6}$$

$$UCL_{R,i} = \{d_2(n_i) + 3d_3(n_i)\}\hat{\sigma}_R. \tag{48.7}$$

At constant size these reduce to the familiar $\bar{x}_B \pm A_2\bar{R}$ and $D_3\bar{R}, D_4\bar{R}$ expressions. For varying sizes, the range center changes as well as its limits. Compare each point with its own lines.

For an S chart, use the within-group sample standard deviation

$$s_j = \sqrt{\frac{\sum_{k=1}^{n_j} (x_{jk} - \bar{x}_j)^2}{n_j - 1}}, \qquad c_4(n) = \sqrt{\frac{2}{n-1}} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)}.$$

Here Γ is the gamma function and c_4 corrects the normal-theory bias of s. Replace $\hat{\sigma}_R$ by $\hat{\sigma}_S = m^{-1} \sum_j s_j / c_4(n_j)$ in the mean-chart limits. The S-chart center is $c_4(n_i)\hat{\sigma}_S$; its limits are

$$\max\{0, c_4(n_i) - 3\sqrt{1 - c_4(n_i)^2}\}\hat{\sigma}_S, \qquad \{c_4(n_i) + 3\sqrt{1 - c_4(n_i)^2}\}\hat{\sigma}_S.$$

For constant n, these are also written $B_3\bar{S}, B_4\bar{S}$, with mean-chart limits $\bar{x}_B \pm A_3\bar{S}$.

The normal-range constants are evaluated at full precision for supported sizes 2–25. In particular, $D_4(5) = 2.1144991451$, whereas the NIST printed table lists 2.115 (reference [2]). This table value is not the ordinary three-decimal rounding of the full-precision result. Use the stated precision and convention when reproducing limits.

48.4.2 Worked example: five packages per sampling occasion

Choose **File > Import data...** and select `sample-data/manual/spc-subgroups.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Quality > Statistical process control > Assess stability / control charts...**. The file `spc-subgroups.csv` contains seven groups of five fill measurements, in grams. The first six groups form the teaching baseline. The table below gives every measurement in file order and the resulting mean and range. In the CSV, each measurement occupies its own row.

Group	Measurements (g)					Mean (g)	Range (g)
1	49.8	50.1	50.0	49.9	50.2	50.0	0.4
2	50.1	50.3	49.9	50.0	50.2	50.1	0.4
3	49.9	49.8	50.2	50.1	50.0	50.0	0.4
4	50.0	49.7	50.1	49.9	50.3	50.0	0.6
5	50.2	50.0	49.9	50.1	49.8	50.0	0.4
6	49.9	50.2	50.0	50.1	49.8	50.0	0.4
7	50.7	50.5	50.8	50.6	50.4	50.6	0.4

Settings. Choose **Variables (measurements)**, process variable `measurement`, order label `subgroup`, and **Constant subgroup size** of 5. Enable **Calculate control limits from a historical subset**, choose **baseline**, equality, and value 1. Clear Rules 2–4 and leave Rules 5–8 off, so only Rule 1 is active. Apply no specification overlays. Do not sort the measurements within or across the original groups. Select **Run analysis**.

Calculation and output. The six baseline means average to 50.0166667 g. The six ranges sum to 2.6 g, so $\bar{R} = 0.4333333$ g. The resulting limits are

Chart (g)	Lower limit	Center	Upper limit
X-bar	49.766712	50.016667	50.266622
R	0.000000	0.433333	0.916283

Group 7 has a mean of 50.6 g and signals above the X-bar upper limit. Its range is 0.4 g, consistent with the reference dispersion. The comparison locates the change between subgroups, rather than suggesting that the five measurements in group 7 have become more spread out.

What changes if group sizes vary? With a grouping column, the mean of a larger group has a smaller standard error under the common-sigma model. Its mean-chart limits therefore become narrower. Raw ranges tend to increase with group size, so the R-chart center also rises. A

staircase in these lines is an expected consequence of changing information per subgroup; it is not itself evidence of a process change.

Important

Do not pool different products, settings or measurement systems merely to reach a subgroup size. If those differences matter scientifically, identify them in the design. Repeated values caused by rounding can also make within-group variation appear smaller than it is.

48.5 Counts, proportions and rates

An attribute chart starts with an operational definition of an event: what counts as a nonconforming unit, a defect or a rare event? Keep inspection criteria and opportunities consistent. A change in inspection intensity can change the observed rate without a change in the underlying process.

Choose **Attributes (counts)**, then the chart type and **Count column**. For p and np, this is the number of nonconforming units; for c and u, it is the number of defects. For g and t, the field instead selects opportunities or elapsed time between events. Each row represents one plotted subgroup or interval.

When a denominator is required, select a sample-size or inspection-area column, or enter a **Constant sample size** with **(none / use constant)** selected. For p and np, inspected-unit counts must be positive whole numbers. Exposure for a u chart can be continuous, such as metres of material inspected.

48.5.1 Binomial and Poisson limits

Let B denote the baseline rows. For d_i nonconforming units out of n_i inspected units, the pooled baseline proportion is

$$\bar{p} = \frac{\sum_{i \in B} d_i}{\sum_{i \in B} n_i}.$$

A p chart plots $p_i = d_i/n_i$ with center $\bar{p}$ and limits

$$\max\left\{0, \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}}\right\}, \qquad \min\left\{1, \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n_i}}\right\}.$$

An np chart plots d_i at a *constant* sample size n. Its center is $n\bar{p}$ and its limits are $n\bar{p} \pm 3\sqrt{n\bar{p}(1-\bar{p})}$, with the lower limit clipped at zero. HypoNum does not clip this chart's upper limit at n. Use p when the number inspected varies.

For defects with constant opportunity, a c chart plots the count c_i. Its center is the baseline mean count $\bar{c}$, with limits $\max(0, \bar{c} - 3\sqrt{\bar{c}})$ and $\bar{c} + 3\sqrt{\bar{c}}$. For varying exposure e_i, a u chart instead plots $u_i = c_i/e_i$, uses $\bar{u} = \sum_{i \in B} c_i / \sum_{i \in B} e_i$, and has limits

$$\max\{0, \bar{u} - 3\sqrt{\bar{u}/e_i}\}, \qquad \bar{u} + 3\sqrt{\bar{u}/e_i}.$$

These constructions use binomial or Poisson variation, respectively. For very small expected counts, their discreteness and asymmetry matter: three-sigma limits do not guarantee normal-distribution tail probabilities.

Rows with null or non-finite counts, negative counts, or unusable denominators are omitted and counted. A denominator that is zero or negative is unusable. Remaining counts must be whole

numbers, except for elapsed time in a t chart. A nonconforming count greater than the number inspected is an error. The analysis needs at least two usable rows and a nonempty baseline. These numerical checks cannot establish that all rows represent comparable opportunities.

48.5.2 Worked example: unequal numbers inspected

Choose **File > Import data...** and select `sample-data/manual/spc-proportions.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Quality > Statistical process control > Assess stability / control charts...**. In `spc-proportions.csv`, periods 1–6 form the teaching baseline. Periods 7–8 are monitoring observations. A unit is either conforming or nonconforming under the same inspection rule throughout.

Period	Nonconforming	Inspected	Baseline	Proportion	Upper limit
1	4	100	1	0.0400	0.098788
2	6	150	1	0.0400	0.088000
3	3	80	1	0.0375	0.105727
4	9	200	1	0.0450	0.081569
5	5	120	1	0.0417	0.093666
6	3	100	1	0.0300	0.098788
7	14	100	0	0.1400	0.098788
8	8	200	0	0.0400	0.081569

Settings. Set **Chart family** to **Attributes (counts)**. Choose the p chart, `nonconforming` as the count, `inspected` as the sample-size column, and `period` for the order labels. Enable **Calculate control limits from a historical subset**, choose **baseline**, equality, and value 1. Retain all rows, clear Rules 2–4 and leave Rules 5–8 off so only Rule 1 is active. Select **Run analysis**.

Calculation and interpretation. The baseline contains 30 nonconforming units out of 750 inspected, so $\bar{p} = 0.04$. With 100 inspected units, the upper limit is

$$0.04 + 3\sqrt{\frac{0.04(0.96)}{100}} = 0.098787754.$$

With 200 inspected units, it falls to 0.081569219. Every lower limit in this example is zero after clipping. Period 7, at 0.14, exceeds its upper limit; period 8, at 0.04, does not.

The center is a pooled proportion, not an unweighted average of the six percentages. Larger inspected samples contribute more information. Likewise, a single displayed summary limit cannot replace the sample-size-specific limit at the point being assessed.

48.5.3 Laney p-prime and u-prime

A Laney chart adjusts the sampling-model spread using observed subgroup-to-subgroup variation. Before choosing it, examine whether mixing populations, dependence or a genuine process change explains the apparent excess variation. A dispersion adjustment does not repair an unsuitable baseline.

For p′, standardize each baseline proportion as

$$z_i = \frac{p_i - \bar{p}}{\sqrt{\bar{p}(1-\bar{p})/n_i}}, \qquad \hat{\sigma}_z = \frac{\overline{|z_i - z_{i-1}|}}{2/\sqrt{\pi}}.$$

For u′, replace the numerator by $u_i - \bar{u}$ and the denominator by $\sqrt{\bar{u}/e_i}$. Multiply the ordinary p or u standard error by $\hat{\sigma}_z$ before forming and clipping the limits. For example, $\hat{\sigma}_z = 1.5$ multiplies an unclipped half-width by 1.5; a positive value of 0.8 multiplies it by 0.8. HypoNum retains positive estimates below one.

The moving-range average uses consecutive source rows that are both usable baseline subgroups. An omitted or excluded row breaks adjacency for this Laney calculation. If standardization is undefined, no eligible pair remains, or the estimated dispersion is zero or non-finite, HypoNum reports the chart as unavailable with a reason. It does not replace it with an ordinary p or u chart. Laney's paper gives the standardized construction and its interpretation (reference [4]).

48.5.4 Rare events: opportunities and elapsed time

The g chart counts whole opportunities *through* the next event: an event at the very next opportunity is recorded as 1. Do not supply the number of event-free opportunities, whose support starts at zero, without converting that convention first.

With baseline mean $\bar{g}$, HypoNum fits the positive-support geometric model using $\hat{p} = 1/\bar{g}$. Its center is $\bar{g}$, with control limits given by the 0.00135 and 0.99865 geometric quantiles. For example, $\bar{g} = 10$ gives $\hat{p} = 0.1$, center 10, lower limit 1 and upper limit 63. An interval of 63 is on the limit; 64 is beyond it. The discrete tail probabilities need not equal the nominal probabilities.

The t chart uses positive elapsed times, in one consistent unit. It fits an exponential model with scale equal to the baseline mean $\bar{t}$, uses that mean as its center, and sets limits at

$$LCL = -\bar{t}\log(1 - 0.00135), \qquad UCL = -\bar{t}\log(1 - 0.99865).$$

For a mean interval of 10 hours, the limits are approximately 0.013509 and 66.076507 hours. The lower limit is close to zero because the exponential distribution is asymmetric.

Both charts require strictly positive intervals. The probability limits, mean centers and geometric support are part of the method. They are not interchangeable with a median center, zero-based geometric count or mean-plus/minus-three-standard-deviations design. Longer intervals can be desirable when events are failures, yet can also indicate incomplete reporting; investigate the operational meaning of either tail.

48.6 Small sustained changes: CUSUM and EWMA

A Shewhart chart compares each plotted statistic with its limits. CUSUM and EWMA also use the preceding observations, which can help reveal a small persistent departure before an individual measurement looks extreme. Both methods require a reference target T and a process standard deviation σ on the measurement scale.

Select **Time-weighted (CUSUM / EWMA)**. Enter a target and sigma when they come from an established reference. If left blank, the target is the baseline mean and sigma is estimated as $\overline{MR}/(2/\sqrt{\pi})$ from successive selected baseline measurements. This automatic estimate pairs the baseline readings in their retained order, including across intervening non-baseline readings. It differs from the new I–MR desktop baseline-pair rule. Use a contiguous, suitable reference period when estimating these parameters automatically. A supplied sigma must be positive; a baseline with no variation cannot supply an automatic sigma.

48.6.1 CUSUM: accumulate departures beyond an allowance

For the tabular CUSUM, the allowance $K = k\sigma$ absorbs small departures from the target, and the decision interval is $H = h\sigma$. The default multipliers are $k = 0.5$ and $h = 5$. Starting with $C_0^+ = C_0^- = 0$, calculate

$$C_i^+ = \max\{0, C_{i-1}^+ + (x_i - T) - K\}, \tag{48.8}$$
$$C_i^- = \max\{0, C_{i-1}^- + (T - x_i) - K\}. \tag{48.9}$$

The upper accumulator responds to sustained high measurements; the lower accumulator responds to sustained low measurements. Both are plotted as nonnegative quantities. A signal occurs when an accumulator is *greater than* H, not when it merely equals H. The allowance must be nonnegative and the decision multiplier positive.

These settings define a detection design, not an estimate of the process's current mean. A change in k or h changes its operating properties. Choose a design for the shift size and false-signal behavior relevant to the application, rather than tuning it to make an already observed signal disappear.

48.6.2 EWMA: give recent observations more weight

An exponentially weighted moving average begins at $z_0 = T$ and follows

$$z_i = \lambda x_i + (1 - \lambda) z_{i-1}, \qquad 0 < \lambda \leq 1.$$

The default weight is $\lambda = 0.2$. A smaller weight retains more history. With limit multiplier L (default 3), the finite-time limits used by HypoNum are

$$T \pm L\sigma\sqrt{\frac{\lambda}{2 - \lambda}\left[1 - (1 - \lambda)^{2i}\right]}.$$

The limits widen from their initial values toward $T \pm L\sigma\sqrt{\lambda/(2 - \lambda)}$. An early plotted point must be assessed against its finite-time limits, not these asymptotes. Here i counts usable observations from the start of the calculation; calendar time does not change the weight.

The Nelson sequence rules do not apply to CUSUM or EWMA. Their reported signals are their own decision- or control-limit breaches. Published average run lengths depend on initialization, limits and parameter knowledge. For example, Lucas and Saccucci give $L = 2.962$ at $\lambda = 0.20$ for a zero-state in-control average run length of 500 with their asymptotic-limit design (reference [5], Table 3). That calibration is not a performance claim for HypoNum's default $L = 3$, finite-time limits or an estimated baseline.

48.6.3 Worked example: compare two ways of accumulating a shift

Choose **File > Import data...** and select `sample-data/manual/spc-time-weighted.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Quality > Statistical process control > Assess stability / control charts...**. The file `spc-time-weighted.csv` contains the following 12 fill measurements in grams. For this teaching example, suppose the reference target is 100 g and the reference sigma is 2 g. These are supplied values, not estimates from the 12 observations.

Settings. Select **Time-weighted (CUSUM / EWMA)**, process variable `measurement` and order label `period`. Leave the historical subset off and enter **Target** 100 and **Sigma** 2. Choose **CUSUM** under **Method**, enter **Slack k** 0.5 and **Decision interval h** 5, then click **Run analysis**. Reopen **Quality > Statistical process control > Assess stability / control charts...** on the same table. Select **Time-weighted (CUSUM / EWMA)**, process variable **measurement**, order label **period**, no historical subset, target 100 and sigma 2. Choose **EWMA**, enter **Weight** 0.2 and **Limit width L** 3, then click **Run analysis**. Keep all rows included for both runs and compare their first signal below.

Period	x_i (g)	C_i^+ (g)	C_i^- (g)	z_i (g)	EWMA LCL	EWMA UCL
1	100	0	0	100.0000	98.8000	101.2000
2	101	0	0	100.2000	98.4633	101.5367
3	99	0	0	99.9600	98.2820	101.7180
4	100	0	0	99.9680	98.1755	101.8245
5	102	1	0	100.3744	98.1104	101.8896
6	103	3	0	100.8995	98.0699	101.9301
7	102	4	0	101.1196	98.0445	101.9555
8	104	7	0	101.6957	98.0283	101.9717
9	103	9	0	101.9566	98.0181	101.9819
10	102	10	0	101.9652	98.0116	101.9884
11	103	12	0	102.1722	98.0074	101.9926
12	104	15	0	102.5378	98.0047	101.9953

Follow one calculation. The CUSUM allowance is $K = 0.5(2) = 1$ g and its decision interval is $H = 5(2) = 10$ g. At period 6, $C_6^+ = \max(0, 1 + 103 - 100 - 1) = 3$ g. For EWMA, $z_5 = 0.2(102) + 0.8(99.968) = 100.3744$ g. The table carries the remaining recurrences without resetting after a signal.

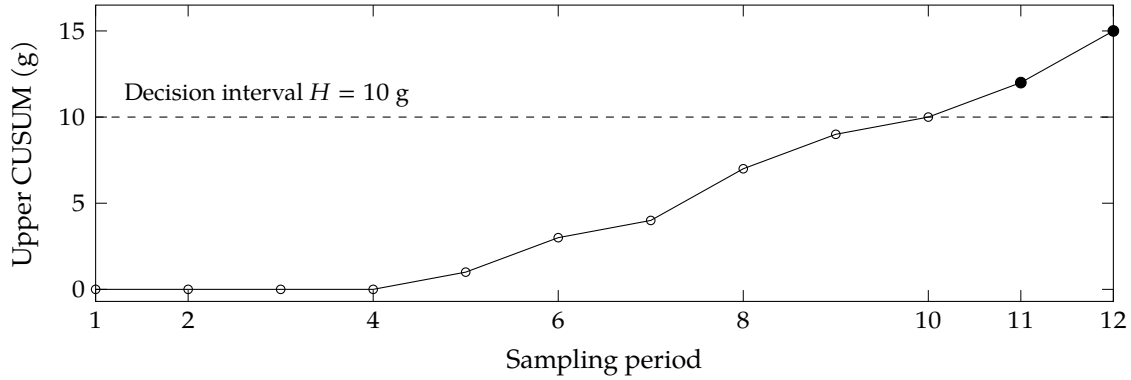


Figure 48.2. The CUSUM accumulates departures from the target after allowing 1 g per observation. Period 10 meets the boundary; the filled points at periods 11 and 12 exceed it. Explanatory illustration drawn from the example calculations; the application's plot styling differs.

Read the result. At period 10, $C_{10}^+ = 10$ g is exactly on the decision boundary, so it does not signal. Periods 11 and 12 exceed it. The lower CUSUM stays at zero throughout. The EWMA also first crosses its upper limit at period 11; its period-10 value is still below the applicable 101.9884 g upper limit.

Neither calculation proves that the underlying change began at the first flagged period. Accumulation takes time. Inspect the observations leading up to that point and the relevant process

records. The two methods first signal together in this constructed example; they need not do so for another sequence.

48.7 Recognize and interpret sequence signals

A process can change without producing a single extreme observation. Nelson's supplementary rules look for patterns in a sequence (reference [3], Figure 1). In the variables and attribute workflows, Rule 1 is always active; the dialog initially enables Rules 2–4 and leaves Rules 5–8 off. Its nine-point same-side rule is the Nelson convention.

Rule	Default	Condition
1	Always	A point strictly above its upper control limit or below its lower control limit.
2	On	Nine consecutive points strictly on the same side of the center. A point on the center interrupts the run.
3	On	Six consecutive points strictly increasing or strictly decreasing. Equal successive values interrupt the trend.
4	On	Fourteen consecutive points alternating upward and downward. A tie interrupts the alternation.
5	Off	At least two of three consecutive points more than two sigma from the center on the same side.
6	Off	At least four of five consecutive points more than one sigma from the center on the same side.
7	Off	Fifteen consecutive points strictly within one sigma of the center.
8	Off	Eight consecutive points strictly outside the one-sigma zone, with at least one on each side of the center in that same window.

A sequence rule flags the point completing its window, then any later point that continues to satisfy it. It does not retrospectively label every earlier point in the window as a new signal. One observation can complete several patterns and carry several rule numbers. The companion R, S or MR chart is checked for limit breaches; the optional sequence rules are scanned on the primary chart.

48.7.1 Use the scale of the plotted statistic

The sigma used for zones is the standard deviation of the statistic being plotted. On an X-bar chart it is $\hat{\sigma}/\sqrt{n_i}$; on an ordinary p chart it is $\sqrt{\bar{p}(1-\bar{p})/n_i}$. It is not necessarily the standard deviation of individual measurements.

Clipping a limit at zero or one does not redefine this sigma. In particular, dividing the distance from a clipped limit to the center by three gives the wrong zone width. The same principle applies when approved limits are reused with different sample sizes. When sigma is zero, Rules 5–8 have no usable zones.

Strict comparisons matter at the boundaries. A point exactly one sigma from the center is neither strictly inside nor strictly outside that zone for Rules 7 and 8. For Rule 8, eight values all 1.5 sigma above the center do *not* qualify: both sides are required. Eight values alternating between -1.5 and $+1.5$ sigma do qualify. A point within the one-sigma zone breaks that eight-point pattern.

48.7.2 A signal initiates an investigation

More enabled rules create more opportunities to signal, including signals caused by ordinary variation. A chart using several rules does not inherit the false-alarm probability of a single independent three-sigma comparison. Autocorrelation, estimated parameters and discrete or asymmetric counts change the interpretation further. Nelson specifically discusses restrictions when extending his tests to attribute charts.

A same-side run can be consistent with a location change; a long alternation can suggest systematic cycling or adjustment. These patterns are prompts for checking process knowledge, not unique diagnoses. Record the suspected cause and the evidence supporting it. Do not adjust a process solely because an automated rule supplied a label.

Earlier saved analyses can retain the previous Rule 8 convention, which also signaled an eight-point outside-zone run entirely on one side. Reopening that result preserves its recorded calculation. Run a new analysis to apply the current both-sides definition, and state the convention when comparing historical results.

48.8 From a chart to a documented decision

Read the baseline description and data-handling warnings first. Confirm which observations estimated the center and variation, how many subgroups were retained, and whether limits vary with size. Then examine both location and dispersion before reviewing individual violations. An empty violation table means that no enabled rule signaled under the selected design; it does not prove independence, conformance or adequate process capability.

The violation table identifies the chart, point, label, plotted value and rule. The plotted value is the chart statistic: it can be a subgroup mean, proportion or accumulator rather than an original measurement. Use the source observations and their labels to investigate it.

48.8.1 Save the result and its context

Export violations writes the violation table as CSV. **Export plot** writes the displayed plot as a PNG image. **Info** opens the method explanation. **Add to results** retains the chart snapshot with the project. A saved snapshot records the analysis that was run; a fresh analysis is needed to assess changed data or settings.

A useful process-monitoring report records the measurement and its units, subgroup design, order and exclusions, baseline period, chart family, rule selection, center and limits, and any supplied target or sigma. Include the investigation and resulting disposition, not only an image of flagged points. For varying sample sizes, retain the denominators so that point-specific limits can be reproduced.

48.8.2 Approve a reference for subsequent monitoring

Where available, **Approve limits...** saves the current reference as a control-plan revision. Enter the revision, approving person and rationale or approval reference; associate a process stage when appropriate. Approval records an operational decision. It does not turn an unsuitable baseline into a statistically adequate one.

Use approved limits applies the recorded reference. For p, u, X-bar, R and S charts, the reference retains the underlying center, variation and size model. Applying it to current subgroup sizes therefore reconstructs the appropriate limits at each point; it does not freeze one displayed staircase value for every future sample size.

An approved revision remains fixed until a new revision is approved explicitly. Earlier results can also preserve an older baseline moving-range convention that paired the baseline readings in order across non-baseline observations. Keep that convention with the result when reproducing its limits.

48.8.3 Separate a justified process stage

Use **Define stage...** to identify a stage by its first point label and, optionally, its last point label. Record the process change or other reason for the boundary. **Recalculate stages** estimates each stage independently and restarts the run-rule windows at its first point.

Separate estimation answers a different question from monitoring against a fixed reference. A post-change stage can describe a newly established operating state; it should not be introduced solely to remove evidence of an unexplained change. The available stage and approval controls depend on the chart result.

48.8.4 Record the response to a signal

Annotate signal... links the investigation to the relevant observations. Record the cause or working explanation, containment, disposition, any exclusion reason, and the reaction-plan reference. Prior approval and disposition revisions remain in the project history.

In the diameter example, a reasonable first record identifies periods 24–26 and the high-side limit breach, notes the unchanged MR result, and lists the measurement and operating records to inspect. A cause is recorded as established only when supported by that investigation. If observations are excluded from a later baseline, retain the reason and compare the revised study with the original.

For a worked route through measurement evidence, Phase I monitoring, capability and approved limits, see section 61.3. The following SPC sections address short-run monitoring, moving averages and multivariate charts.

48.9 Short-run standardized Z–MR monitoring

Several products can share a monitoring chart when each has a defensible process centre and standard deviation. Standardization expresses every reading in its product's own standard-deviation units:

$$z_i = \frac{x_i - \mu_{p_i}}{\sigma_{p_i}}.$$

The study requires the product parameters in advance. It does not estimate or pool them from the readings being monitored.

Select the short-run monitoring command in **Quality > Statistical process control**. Assign distinct Measurement and Product columns, optionally a Stage, and supply one parameter record for every active product. The initial selectors are the first two columns; confirm them.

48.9.1 Supply meaningful product parameters

The parameter editor accepts a structured list with product name, centre, positive sigma and an optional unit label.

For the worked example, enter:

```
[{"product":"A","center":10,"sigma":1},
 {"product":"B","center":100,"sigma":5}]
```

This is the content of the desktop parameter field. Product names must be unique, centres and sigmas finite, and every monitored product must be covered. A unit label documents the raw scale; it performs no conversion. The form can begin empty, but running requires one to 200 records.

Centres should describe the stable process being monitored. A specification midpoint or tolerance divided by six is not automatically the process mean or standard deviation. Changing those parameters changes the monitoring question.

48.9.2 Follow a short mixed-product run

Choose **File > Import data...** and select `sample-data/manual/short-run-four.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Quality > Statistical process control > Short-run standardized Z-MR...**. Set **Measurement column** to **y**, **Product column** to **product**, and leave the optional stage unset. Paste this complete list into **Known product parameters (JSON)**:

```
[{"product":"A","center":10,"sigma":1},
 {"product":"B","center":100,"sigma":5}]
```

Leave **Reset MR at product change** off and click the primary **Run Z–MR** button. Compare the standardized observations with the table below and check that the last observation signals.

Row	Product	Reading	Centre	Sigma	Z
1	A	10	10	1	0
2	B	105	100	5	1
3	A	11	10	1	1
4	B	120	100	5	4

The standardized moving ranges are 1, 0 and 3. The last Z value is above three, so it signals. None of the moving ranges exceeds its limit.

With known independent-normal standardized variation,

$$d_2 = 2/\sqrt{\pi} = 1.128379\ldots, \qquad d_3 = \sqrt{2 - 4/\pi} = 0.852502\ldots.$$

The Z centre is zero with limits -3 and $+3$. The MR centre is d_2, its lower limit zero, and its upper limit

$$d_2 + 3d_3 = 3.685886566\ldots.$$

Only strict limit breaches signal. A value exactly on a limit does not breach it. These are three-standard-deviation range limits, not an exact tail-probability or average-run-length guarantee. They use full-precision moments rather than combining separately rounded printed constants.

48.9.3 Decide whether a product change resets the range

Initially a product transition contributes a moving range because both readings use standardized units. Turning on **Reset MR at product change** starts a new MR sequence at every change; the choice is saved and disclosed. For the four-row example, every row changes product, so that option would leave no adjacent within-product ranges.

Original source row order defines the sequence. Sorting only a display does not redefine it. Missing and excluded rows break MR adjacency. A changed stage also breaks adjacency; a missing stage marks the row missing, and a stage label returning later begins a new segment. This chart does not close the sequence around a missing position.

48.9.4 Inspect original readings with standardized signals

The detailed result links each statistic to the original measurement, product, supplied centre and sigma, preceding row, stage and reason. Unknown products, invalid parameters and non-finite active measurements fail explicitly; no unit sigma or zero centre is substituted. At least two eligible measurements are required.

The table initially displays 200 source rows; each chart displays at most 400 eligible points, and the signal table shows its first 200 signals. Total counts cover the complete calculation. The source limit is one million rows and twenty million selected cells.

Add to results requires the same source and exclusions as the calculation. A new run supersedes an older result; Escape closes the window and cancels an active run. The study changes no measurements and carries no specification or compliance decision.

The range moments follow the normal span-two convention described in the NIST/SEMATECH *e-Handbook of Statistical Methods*, section 6.3.2.2. Applying supplied product scales is a declared standardization model whose adequacy must be assessed for the process.

48.10 Finite-window moving-average charts

A finite moving average smooths an ordered sequence over the most recent w retained observations. Unlike EWMA, an observation loses all influence when it leaves that finite window.

Choose **Quality > Statistical process control > Finite-window moving-average control chart...**. The desktop initially uses width three, a moving-range baseline standard deviation and expanding startup. Width one is supported and gives the individuals-chart statistic.

48.10.1 Establish the baseline before monitoring

Use a Phase-I marker to separate reference and monitoring observations. Only reference rows estimate the frozen mean and standard deviation. Alternatively supply both a historical mean and a positive historical sigma; neither can be entered alone. If neither is supplied, all finite rows estimate the baseline and remain eligible for signals. The result labels this single-sequence mode.

Marker values must be boolean or the trimmed, case-insensitive labels 0, 1, false or true. Other values are refused as ambiguous. The baseline moving-range estimator is $\overline{MR}/(2/\sqrt{\pi})$ at full precision. A constant or otherwise unusable estimated baseline is refused.

48.10.2 Follow startup and a complete window

With k values currently available in a window,

$$M_t = \frac{1}{k}\sum_{j=0}^{k-1} x_{t-j}, \qquad LCL_t = \mu_0 - \frac{3\sigma_0}{\sqrt{k}}, \qquad UCL_t = \mu_0 + \frac{3\sigma_0}{\sqrt{k}}.$$

Expanding startup uses $k = 1, 2, \ldots, w$ before the first full window. The decreasing limit width reflects the number of observations actually averaged.

Choose **File > Import data...** and select `sample-data/manual/moving-average-four.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Quality > Statistical process control > Finite-window moving-average control chart...**. Set **Measurement column** to **y**, **Window width** to 3, **Historical mean** to 10 and **Historical sigma** to 2. Leave **Stage column** and **Phase-I marker column** unset. Click **Run chart**. Read the plotted averages in imported row order and compare their limits with this table:

Row	Reading	k	Average	Lower	Upper
1	10	1	10	4	16
2	12	2	11	5.757359	14.242641
3	14	3	12	6.535898	13.464102
4	16	3	14	6.535898	13.464102

The fourth moving average exceeds its upper limit. Its window contains 12, 14 and 16; the first value has left the average. The baseline remains 10 and two throughout.

Successive averages overlap and are correlated. Only limit breaches are signalled; ordinary independent-point run rules and their run-length interpretations cannot simply be added.

48.10.3 Recognize gaps, stages and saved conventions

Missing, non-finite and excluded rows are omitted before this chart's sequence is built. The retained observations become adjacent. An optional Stage column resets the window at each change. A later return to an earlier label still begins a new segment.

This adjacency rule differs from charts that retain a missing position as a break. Confirm that averaging across omitted positions matches the intended use of the measurements. An exclusion request needs reliable row identity; it is refused when that identity is unavailable.

Saved calculations can also use a full-window convention: partial means remain visible during the first $w - 1$ positions, but limits and breach decisions are withheld until a complete window exists. The desktop setup described here uses expanding startup. Montgomery's Example 9.3 uses expanding startup too; its separate early Shewhart option is not the same as withholding early decisions.

The result records the baseline, startup rule, stages, population and source rows. Chart/export observations are limited to 500 rows, and report tables to 200 rows and sixteen columns; truncation is disclosed. Read those display limits alongside the retained counts and settings. Saved results keep their original baseline even when later data change.

Montgomery, *Introduction to Statistical Quality Control*, 6th edition, Example 9.3, pp. 428–429, provides a published moving-average example and startup discussion.

48.11 References and further reading

The references below support the statistical constructions and their interpretation. The CSV studies in this chapter are constructed teaching examples; their numbers are not attributed to these papers. For retained comparisons against published numerical examples, consult the Quality chapter of the companion statistical validation report.

[1] Montgomery, D. C. (2009). *Introduction to Statistical Quality Control*, sixth edition. Wiley. Chapters 5–7 develop the use of control charts and the variables and attribute families; Chapter 9 treats CUSUM and EWMA. Useful for baseline design, rational subgrouping and the distinction between monitoring and capability.

[2] NIST/SEMATECH. *e-Handbook of Statistical Methods*, Section 6.3.2.1, Shewhart X-bar and R and S control charts. Gives normal-theory constructions and a printed factor table. Distinguish the table's displayed constants from a full-precision calculation when comparing final digits.

[3] Nelson, L. S. (1984). The Shewhart control chart—tests for special causes. *Journal of Quality Technology*, 16(4), 237–239. doi:10.1080/00224065.1984.11978921. Figure 1 defines the eight patterns; Figure 2 and the accompanying text discuss interpretation and applicability.

[4] Laney, D. B. (2002). Improved control charts for attributes. *Quality Engineering*, 14(4), 531–537. Develops standardized attribute charts and the moving-range dispersion adjustment used in p′ and u′ monitoring.

[5] Lucas, J. M., and Saccucci, M. S. (1990). Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics*, 32(1), 1–12. Stable article record. Discusses EWMA design and operating properties. Table 3 supplies the specific average-run-length calibration discussed in this chapter.

[6] NIST/SEMATECH. *e-Handbook of Statistical Methods*, Sections 6.3.2.3, CUSUM control charts, and 6.3.2.4, EWMA control charts. Further explanation of the two monitoring constructions.

CHAPTER 49

Between/within and zone monitoring

Rational subgroups distinguish short-term dispersion within a sampling occasion from changes between occasions. Keep subgroup identity and time order intact. A control limit concerns process behaviour; a specification limit concerns the required product and serves a different purpose.

49.1 Between/within I–MR–R/S charts

Choose **Quality > Statistical process control > Between/within I-MR-R/S chart...**. Select a numeric measurement, a subgroup ID or fixed consecutive size, optional stage and either an R or S dispersion chart. Subgroups contain 2–25 finite measurements; at most 5,000 subgroups are charted. Fixed-size groups are assigned before exclusions and missing-value removal, so later measurements do not move into earlier sampling periods. Incomplete groups are dropped and counted. Each stage needs at least two retained groups.

Three displays answer different questions. The I chart monitors subgroup means $\bar{x}_i$ around their *unweighted* mean $c = \sum_i \bar{x}_i/k$, with limits $c \pm 3\overline{MR}/d_2(2)$. The MR chart monitors $MR_i = |\bar{x}_i - \bar{x}_{i-1}|$ around $\overline{MR}$, with upper limit $D_4(2)\overline{MR}$. The R or S chart monitors within-group dispersion using the mean of $R_i/d_2(n_i)$ or $s_i/c_4(n_i)$. Its limits change with n_i. Rule 1 marks points beyond their three-sigma limits; other run rules are not applied by this command.

The size-weighted mean of all individual measurements is reported separately. For unequal sizes it need not equal the I-chart centre. Applying one MR scale to means with different sampling precision is approximate, and the result warns about unequal pointwise false-alarm behaviour. Moving ranges join consecutive retained groups and can span dropped sampling periods. Stage boundaries restart the calculations.

49.1.1 Keep three meanings of spread separate

The I-chart scale estimates variation of a *subgroup mean*. The R/S scale estimates within-subgroup dispersion. The between, within and total component SDs come from a constrained random-intercept REML fit $y_{ij} = \mu + u_i + e_{ij}$, with $\sigma^2_{\text{total}} = \sigma^2_u + \sigma^2_e$. Do not add the I-chart variance to a within variance: it already contains within variation divided by subgroup size.

For balanced groups with a positive between component, $\hat{\sigma}^2_u = (MS_B - MS_W)/n$ and $\hat{\sigma}^2_e = MS_W$. At a zero between-component boundary the constrained REML within estimate can change as well. Zero chart scales leave the affected limits unavailable; they are not turned into a numerical pass. A previously retained result preserves the estimator identified in its own method notes.

Worked example

Choose **File > Import data...** and select `sample-data/manual/between-within-twenty-four.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Quality > Statistical process control > Between/within**

I-MR-R/S chart.... Set **Measurement** to **value**, **Subgroups** to **Rows sharing a subgroup ID column**, and **Subgroup ID** to **subgroup**. Leave **Stage** and approved control-plan limits unset. Choose **R chart (ranges)** for **Within-subgroup chart** and click **Run analysis**. Inspect the I, MR and R limits and the separate variance-component estimates. The six groups contain four measurements each, $m-1, m, m+1, m$, with means $m = (10, 12, 9, 11, 13, 10)$. Every range is 2, the centre is 10.8333 and $\overline{MR} = 2.4$. The I-chart limits are approximately 4.4525 and 17.2142, and the R-chart upper limit is 4.5641. No group exceeds them. Here $MS_B = 26/3$ and $MS_W = 2/3$, giving between SD $\sqrt{2}$ and total individual SD $\sqrt{8/3}$. These are distinct from the I-chart scale $2.4/d_2(2) = 2.1269$.

49.2 Accumulate zone scores

Quality > Statistical process control > Zone chart... scores the distance and side of each subgroup mean. For subgroup-mean monitoring, $z_i = (\bar{x}_i - c)/(\hat{\sigma}_w/\sqrt{n_i})$; the centre is the size-weighted mean of the baseline observations. The alternative I chart of subgroup means uses its recorded MR scale. This is not a raw-individual zone chart.

The four absolute-z zones have boundaries 1, 2 and 3. A point on a boundary belongs to the inner zone. Default weights are 0, 2, 4 and 8, with signal score 8. Scores accumulate while points remain on the same side. A strict side change starts a new score with that point's weight. Exactly at the centre, zero is added and the previous side and score remain. With default accumulation, subsequent same-side points continue to signal once the threshold is reached. **Reset after signal** restarts the next point from zero. Each stage starts a new accumulation.

For eight successive points at $z = 1.5$, the scores are 2, 4, 6, 8, 10, 12, 14 and 16 under the defaults; the fourth starts signalling. A point beyond three sigma on the opposite side immediately starts at 8 and also signals. The table reports each point's zone, weight, cumulative score and reset event. Read it together with the measurement plot.

49.3 Monitor against an approved baseline

Run a baseline chart and choose **Approve limits...** in its results. Supply the revision and approver, with a stage and rationale where needed. The control plan keeps prior approvals. A chart already using approved limits cannot be approved again as a newly estimated baseline.

In a monitoring setup, select **Use approved limits from the control plan**. The selected entry supplies its baseline centre and scales; monitoring observations do not re-estimate them. The retained result records the exact entry used, even if a later approval changes the control plan. Between/within monitoring does not estimate new variance components.

Entries are restricted to their dataset, characteristic and chart. Dispersion estimator and zone scoring choices must match. Between/within monitoring sizes must stay within the baseline size range; an I-of-means zone chart requires its baseline's single size. Subgroup-mean zone scales use each accepted size's square root. Monitoring covers one stage per run and starts its MR sequence within the monitoring data, without a bridge from the last baseline subgroup.

For a baseline with means $(10, 12, 9, 11)$, the I centre is 10.5 and $\overline{MR} = 7/3$. Its limits remain approximately 4.2964 and 16.7036 when new means $(20, 20, 21)$ arrive: all three signal. Re-estimating limits from those new means would answer a different question and conceal the comparison with the approved baseline.

CHAPTER 50

Multivariate subgroup monitoring

Correlated measurements can change together even when separate univariate charts are difficult to interpret. This workflow pairs a location statistic for each subgroup mean with a spread statistic for its covariance.

Choose **Quality > Statistical process control > Hotelling T-squared subgroups and generalized variance...**. It monitors future, fixed-size rational subgroups against a frozen reference baseline. The individual-observation Hotelling chart uses a different limit and is a separate choice.

50.1 Declare complete reference and monitoring groups

Select two to 64 numeric variables, a Subgroup column, fixed subgroup size n and a Reference indicator with its reference value. The subgroup column must differ from the variables and indicator. The proposed size is the most common label count, with the smaller count breaking a tie; confirm it against the physical grouping.

Size must be two to 100. The initial location false-alarm level is $\alpha = 0.0027$; a selected value must lie strictly between zero and 0.5. The source ceiling is one million rows.

Rows sharing a subgroup label form one group. A reference row matches the declared indicator value as text; all members of a group must share the same reference or monitoring role. Mixed roles stop the calculation.

Excluded rows are removed first. Missing or blank group labels are counted as unassigned. More than n remaining rows in a group is an error. A shorter group, or a missing/non-finite monitored value within it, drops that whole group as incomplete. No smaller group replaces it. Review excluded, unassigned, incomplete, reference and monitoring counts before using the chart. Monitoring groups follow their first appearance in the table.

50.2 Locate a change in the multivariate mean

Let m complete reference subgroups each contain n observations of p variables. The baseline mean is the average of their mean vectors and $\bar{S}$ is the average of their sample covariance matrices, each using divisor $n - 1$.

The covariance has $m(n-1)$ degrees of freedom. That number must be at least p and the matrix must be nonsingular. A singular matrix is refused rather than replaced by an arbitrary generalized inverse.

For a monitoring subgroup,

$$T_j^2 = n(\bar{\mathbf{x}}_j - \bar{\bar{\mathbf{x}}})^\top \bar{S}^{-1} (\bar{\mathbf{x}}_j - \bar{\bar{\mathbf{x}}}).$$

The Phase-II upper limit for a future subgroup mean is

$$UCL_{T^2} = \frac{p(m+1)(n-1)}{mn - m - p + 1} F_{1-\alpha}(p, mn - m - p + 1).$$

A strict exceedance signals; univariate zone/run rules are not added. This interpretation assumes the model of independent normal rational subgroups and an applicable reference process.

50.3 Locate a change in joint spread

Generalized variance is the determinant $|S_j|$ of a subgroup covariance matrix. For an underlying covariance Σ,

$$E|S_j| = b_1|\Sigma|, \qquad \operatorname{Var}|S_j| = b_2|\Sigma|^2,$$

where

$$b_1 = \frac{\prod_{i=1}^{p}(n-i)}{(n-1)^p},$$

$$b_2 = \frac{\prod_{i=1}^{p}(n-i)\left\{\prod_{j=1}^{p}(n-j+2) - \prod_{j=1}^{p}(n-j)\right\}}{(n-1)^{2p}}.$$

Estimate $|\Sigma|$ by the mean reference determinant divided by b_1. The centre is therefore the mean reference determinant; the upper and lower limits are

$$\widehat{|\Sigma|}(b_1 + 3\sqrt{b_2}), \qquad \max\{0, \widehat{|\Sigma|}(b_1 - 3\sqrt{b_2})\}.$$

A value above the upper limit, or below a positive lower limit, signals. This determinant distribution is skewed. Its three-SD limits do not inherit the location chart's α.

If $n \le p$, each subgroup covariance is singular, although the pooled reference covariance can support the location chart. The spread chart is then unavailable with an explanation. It is also unavailable when every reference determinant is zero.

50.4 Separate a centre shift from a spread shift

Choose **File > Import data...** and select `sample-data/manual/subgroup-hotelling-sixteen.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Quality > Statistical process control > Hotelling T-squared subgroups and generalized variance...**. Add **x** and **y** to **Monitored numeric variables**. It contains two reference groups and two monitoring groups, each with four rows and variables **x** and **y**. Select Subgroup **subgroup**, size four, Reference indicator **role**, reference value **reference**, and explicitly change α to 0.05. Click **Run analysis**. Inspect the monitoring statistics and limits separately for T-squared and generalized variance.

Each reference group contains

$$(\sqrt{1.5}, 0),\ (-\sqrt{1.5}, 0),\ (0, \sqrt{1.5}),\ (0, -\sqrt{1.5}),$$

giving mean zero and sample covariance I. Monitoring group M1 adds $(1, 0)$ to those four points. M2 doubles the original four coordinates, retaining mean zero.

Group	T^2	$\lvert S\rvert$	Result
M1	4	1	No limit breach
M2	0	16	Spread above its upper limit

Here $m = 2$, $n = 4$, $p = 2$. The location upper limit is 20.830086. For spread, $b_1 = 2/3$, $b_2 = 28/27$, the centre is one, the lower limit zero and upper limit 5.582576. M2's unchanged mean hides its spread shift from the location statistic, but the determinant chart detects it.

These two reference groups make the arithmetic transparent. They do not provide adequate evidence of a stable operating baseline or demonstrate the long-run detection performance of a chart.

50.5 Review and retain the baseline

The result shows counts, baseline mean/covariance, limits, both charts, warnings and signals linked to subgroup member rows. The window initially lists fifty monitoring statistics; report tables retain up to 200 statistics and 200 signals. At most 2,000 monitoring groups are stored for display, while breach counts cover all groups.

Add to results requires the source and exclusions still to match the calculation. A rerun on the same population reproduces the baseline and limits; reopening preserves the saved result. A new run or source change supersedes an older result.

This workflow does not provide retrospective Phase-I limits, variable subgroup sizes, principal-component selection or a decomposition of a signal into responsible variables. Investigate the underlying measurements after a signal.

Montgomery, *Introduction to Statistical Quality Control*, section 11.3, gives the Phase-II location construction. Alt and Grimshaw, *Multivariate Quality Control* (2013), in the *Encyclopedia of Operations Research and Management Science*, pp. 1014–1017, describes the multivariate and generalized-variance methods, especially pp. 1016–1017.

CHAPTER 51

Multivariate EWMA

A multivariate exponentially weighted moving average accumulates information about small, sustained shifts in several correlated variables. It monitors individual observation vectors against a frozen reference mean and covariance.

Choose **Quality > Statistical process control > Multivariate EWMA (MEWMA)...**. Use the subgroup Hotelling workflow for rational subgroups; MEWMA here has one observation vector at each monitoring step.

51.1 Set a weight and a documented decision limit

Select two to 64 numeric variables and a reference indicator/value. The initial smoothing weight is $\lambda = 0.2$, with $0 < \lambda \leq 1$. One weight applies to every variable. Smaller weights retain more of the earlier history.

Enter a positive decision limit h; it has no default. The initial **Uncalibrated trial limit** choice makes no false-alarm or average-run-length claim and carries no calibration source note. **User-calibrated limit** requires a source note of at most 500 characters explaining the published or simulated calibration, its weight, dimension and target in-control run length.

HypoNum records that explanation as user supplied; it does not verify the calibration. A limit suitable for another weight, dimension or baseline-estimation scheme need not be suitable here. The source limit is one million rows.

51.2 Freeze the reference population

A row is a reference row when its indicator equals the declared reference value as text. Excluded rows are removed first; incomplete or non-finite measurement vectors are omitted and counted. The complete reference rows estimate mean $\bar{\mathbf{x}}$ and sample covariance S with divisor $m - 1$.

With p variables, more than p reference rows and a nonsingular covariance are required. A singular baseline is refused. Monitoring values never revise this baseline.

Complete monitoring rows follow table order. Reference rows interspersed among them are skipped by the recursion, with a warning. Missing monitoring rows do not advance the step. The result partitions source rows into excluded, incomplete, reference and monitoring populations.

51.3 Follow the recursion and its changing covariance

Starting at the reference mean,

$$\mathbf{Z}_0 = \bar{\mathbf{x}}, \qquad \mathbf{Z}_i = \lambda \mathbf{x}_i + (1 - \lambda)\mathbf{Z}_{i-1}.$$

The finite-time covariance factor is

$$c_i = \frac{\lambda\{1 - (1 - \lambda)^{2i}\}}{2 - \lambda}, \qquad \Sigma_{Z_i} = c_i S.$$

The monitoring statistic is

$$T_i^2 = (\mathbf{Z}_i - \bar{\mathbf{x}})^\mathsf{T} \Sigma_{Z_i}^{-1} (\mathbf{Z}_i - \bar{\mathbf{x}}).$$

The factor begins at λ^2 and approaches $\lambda/(2-\lambda)$. Using the steady-state covariance from the first point would understate the early statistic.

With $\lambda = 1$, the recursion has no memory and its statistic equals the individual Hotelling statistic for the same row and baseline. This identity concerns the statistic; it does not supply a calibration for an arbitrarily chosen h.

51.4 Calculate the first two monitoring points

Choose **File > Import data...** and select `sample-data/manual/mewma-six.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Quality > Statistical process control > Multivariate EWMA (MEWMA)...**. Its four reference vectors are

$$(\sqrt{1.5}, 0),\ (-\sqrt{1.5}, 0),\ (0, \sqrt{1.5}),\ (0, -\sqrt{1.5}),$$

with mean zero and sample covariance I. The two monitoring vectors are both $(1, 0)$. Select **x** and **y**, Reference indicator **role**, reference value **reference**, weight 0.2, trial limit 9 and no stage. Set **Limit provenance** to **Uncalibrated trial limit**, then click **Run analysis**. Inspect the two monitoring statistics and the decision limit; this chosen trial limit is not an average-run-length calibration.

At the first point, $\mathbf{Z}_1 = (0.2, 0)$ and $c_1 = 0.04$, so $T_1^2 = 0.2^2/0.04 = 1$. At the second,

$$\mathbf{Z}_2 = (0.36, 0), \qquad c_2 = 0.0656, \qquad T_2^2 = 0.36^2/0.0656 = 1.975610.$$

Neither exceeds nine. The accumulating statistic illustrates memory even though the new observation repeats the previous one.

The four-row baseline and trial limit are deliberately chosen for arithmetic. They do not establish stability, an in-control false-alarm probability or an average run length.

51.5 Use stages for an intentional restart

An optional Stage column restarts the recursion at the baseline mean whenever the label changes between consecutive complete monitoring rows. The new segment's first step is one. A label returning after another stage begins another segment and produces a warning. Blank monitoring-stage labels are incomplete rows.

All stages use the same frozen baseline. A signal itself does not reset the chart. The statistic continues accumulating until an explicit stage change, so a sustained shift can produce repeated signals.

A signal is a strict $T_i^2 > h$. Successive statistics are dependent; the decision limit is not a per-point false-alarm quantile. Report the declared calibration basis with the signal count.

51.6 Retain the signal and source context

The window shows baseline and limit information, the chart, row-linked signals, the first fifty monitoring statistics, stage/step information and warnings. Report tables display up to 200 statistics and 200 signals; at most 2,000 monitoring points are stored for display. The total signal count covers every analyzed monitoring point.

Add to results requires unchanged source and exclusions. A new run or table change supersedes older results. Reopening preserves baseline, limit provenance, settings and displays; recalculation on the same data reproduces the statistics.

Lowry, Woodall, Champ and Rigdon, *A Multivariate Exponentially Weighted Moving Average Control Chart* (1992), *Technometrics* 34(1), pp. 46–53, gives the covariance construction in equations (2.3)–(2.5) and a bivariate example in Table 4. That calculation check does not calibrate limits for different operational conditions.

CHAPTER 52

Planning a measurement-system study

A measurement system can add random variation, systematic bias or inconsistent classifications to a process study. Choose an investigation that matches both the decision and the way measurements are collected.

Question	Suitable starting point
Can every appraiser measure every part repeatedly?	Crossed Gage R&R; inspect a gage run chart first.
Does testing destroy the specimen, with different parts assigned to each appraiser?	Nested/destructive Gage R&R, when the declared replication represents comparable material.
Does a site or fixture add a distinct source?	Expanded multi-factor Gage R&R with an explicit crossed or nested structure.
Are readings biased against known values?	Linearity and bias, or a Type-1 study of one standard.
Does the gage drift over time?	Gage stability in acquisition order.
Do appraisers repeat the same categories?	Attribute Agreement.
At what known magnitude does a binary gage switch?	Attribute gage response curve.

Do not relabel a crossed design as nested to make a calculation run. The model must follow the physical experiment. The guided measurement-system study in section 61.2 helps collect the design declaration and choose the matching analysis.

52.1 Choose parts, appraisers and trials deliberately

Use parts that represent the variation relevant to the decision. A very broad part range can make measurement error look small as a percentage of total variation, while a narrow range can do the opposite. Neither changes the gage's absolute measurement error.

Choose appraisers, fixtures and operating conditions that represent routine use. Repeated readings should include the actions whose repeatability matters: removing and replacing a part, resetting a gage or preparing a sample when those actions belong to the method. Avoid letting an appraiser copy a previous reading.

For destructive measurement, repeat specimens within a declared part must be comparable enough for the within-part model. A statistical nesting label cannot make materially different specimens replicates. For all designs, consider resolution, reference uncertainty, within-part heterogeneity and stability alongside the numerical output.

52.2 Create a randomized collection worksheet

Choose **Design > Create MSA study worksheet...**. No existing table is required. The command creates design rows and blank measurement cells; it does not generate observed readings.

Control	Initial choice and allowed setup
Design	Crossed; alternatives are Nested and Attribute agreement.
Parts	10; from 1 to 10,000, per appraiser for Nested. Optional distinct comma-separated labels override the count for other layouts.
Operators	3; from 1 to 1,000. Optional distinct labels override the count.
Trials per cell	2; from 1 to 100.
Randomization seed	2026; integer from zero through $2^{63} - 1$.
Blind part labels	Off; optional seeded labels with a recorded mapping.
Standard values	Blank; one per part, for Attribute agreement only.
New table name	MSA study worksheet; 1–128 characters.

Labels can contain at most 64 characters. Nested worksheets require at least two appraisers and two parts per appraiser, and reject an explicit part-label list. Crossed and attribute worksheets assign every appraiser to every part. A nested worksheet assigns distinct parts to each appraiser.

Run order is randomized within appraiser blocks using the recorded seed. The same settings and seed reproduce that order. The table contains Part, Operator, Trial, TrialID and a blank Measurement column, or Value for attribute ratings; declared standard values and blinded labels are also retained. The total design is limited to one million rows.

Choose **Create dataset tab** to add the sheet as a separate editable table. It does not replace the active table. Its creation can be undone and redone. When an analysis is opened on that worksheet, matching measurement, part, operator, trial and standard roles are suggested. Confirm them before analyzing the completed readings.

52.3 Inspect individual readings first

Choose **Quality > Measurement systems > Gage run chart...**. Assign Measurement, Part and Operator, with an optional Trial column and a numeric reference value.

Panels follow natural part order, sharing one vertical scale and consistent appraiser markers. Short segments mark part means; the horizontal reference is the supplied value or the overall included-row mean.

Within each part/appraiser cell, horizontal position follows Trial order, not elapsed time or trial-number gaps. Without Trial, the result discloses its use of current within-cell row order.

Inspect appraiser shifts, inconsistent repeats, isolated readings and part-specific disagreement. Unbalanced cells are disclosed; rows with missing measurement, part, operator or declared trial are omitted and counted, as are exclusions.

Studies above 5,000 included readings are refused. Each page shows up to twelve part panels. **Select these rows** selects their source rows; **Add to results** retains the setup, counts and means. These exploratory summaries provide neither a variance decomposition nor gage qualification.

CHAPTER 53

Repeatability and reproducibility

Repeatability concerns repeated measurements under the same conditions; reproducibility adds appraiser or other declared condition differences. Part-to-part variation belongs to the specimens.

For a balanced crossed study, a useful model is

$$y_{ijk} = \mu + P_i + A_j + (PA)_{ij} + \varepsilon_{ijk},$$

with independent random part, appraiser, interaction and repeatability terms. The interaction allows appraisers to disagree differently on different parts.

Choose **Quality > Measurement systems > Assess measurement / Gage R&R...**. Select the governing standard before interpreting method names, percentages or guidance.

53.1 Set up a crossed study

The initial **AIAG MSA, 4th ed.** standard starts with ANOVA and interaction threshold 0.25. **HypoNum extended** starts at 0.05 and offers the uncertainty and unbalanced-study options below. These profiles are not interchangeable.

Control	What to specify
Measurement	Numeric response; initially the first numeric column.
Part	Specimen identifier; initially the first column.
Appraiser (or fixture)	Measurement condition; initially the second column. Its displayed role name follows the standard.
Reference value	Optional numeric reference, constant within each part, for error and X–Y displays.
Tolerance	Optional positive width $USL - LSL$; needed for tolerance-based guidance and sorting.
Total variation	Initially From the study. Other bases use a surrogate process spread, Pp target or specification tolerance.
Intended use	Initially analyzing/controlling the process. Under a profile using this choice, sorting reads guidance against tolerance instead.
Process standard deviation	Required for the Range method; optional historic process SD for the pooled-SD study.
Interaction threshold, spread	Start from the standard's values. Only permitted per-study changes retain its standard identity.
Confidence, unbalanced design	Available when the selected standard permits them; HypoNum extended offers these choices.
PV corrected for EV	Initially off; optional Average and Range correction.

Every appraiser must measure every part with the required replication. Review missing or uneven cells before proceeding. A setup outside the standard's permitted choices is refused with an instruction to derive a separate user standard.

53.2 Keep variance, spread and percentages distinct

For component variance v_j, standard deviation is $\sqrt{v_j}$ in measurement units, and study spread is $m\sqrt{v_j}$ for the declared multiplier m. Percent contribution uses variance:

$$100\, v_j / v_{\text{total}}.$$

Percent study variation uses standard deviations:

$$100\, \sqrt{v_j / v_{\text{total}}}.$$

The latter percentages do not add to 100. A component contributing 9% of variance contributes 30% of total standard deviation.

Percent tolerance compares the gage spread with the requirement width. The AIAG profile fixes this comparison to 6σ even when a historical 5.15 multiplier is chosen for the separate spread columns. HypoNum extended instead uses its declared spread multiplier. Read the reported denominator and basis.

53.3 Work through a complete crossed study

Choose **File > Import data...** and select `sample-data/manual/crossed-gage-ninety.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Quality > Measurement systems > Assess measurement / Gage R&R...**. It contains ten parts, three appraisers and three repeated readings in each of thirty cells: ninety measurements in all.

Select Standard **HypoNum extended**, Method ANOVA, Measurement **y**, Part **part**, Operator **operator**, Tolerance 12 and total variation From the study. Use spread multiplier 6, interaction threshold 0.05, confidence 95% and leave unbalanced data disabled. Click **Run study** and inspect the variance-components table.

Source	Variance	% contribution	% study variation
Repeatability	0.256058	7.99994	28.2842
Reproducibility	0.077552	2.42294	15.5658
Total Gage R&R	0.333610	10.42288	32.2845
Part-to-part	2.867139	89.57712	94.6452
Total	3.200749	100.00000	100.0000

The interaction p-value is 0.904825, so interaction is pooled under this declared threshold. The total gage standard deviation is $\sqrt{0.333610} = 0.577590$, its six-SD spread is 3.465539 and its percentage of tolerance is $100(3.465539/12) = 28.8795\%$. The number of distinct categories is four.

The gage's 10.42% variance contribution and 32.28% study variation describe the same decomposition; reporting them as the same percentage would be an error. Its tolerance percentage also differs from its total-variation percentage because it uses a different denominator.

Under HypoNum extended's total-variation guidance, 32.28% is in the above-30% band. The numerical example illustrates calculation and interpretation; choosing sampled parts and a measurement decision requires the actual process context.

53.4 HypoNum extended estimation and uncertainty

Balanced crossed studies use ANOVA. The full table includes Part, Operator, Operator-by-Part, Repeatability and Total. When the interaction p-value is at least the declared threshold, its sum of squares and degrees of freedom join the error term. Otherwise interaction contributes to reproducibility. Main-effect F statistics use the interaction, or pooled error after pooling. Negative method-of-moments components are reported at zero.

For an unequal-replication study, **Allow unbalanced** uses REML and requires at least two trials in each observed part/appraiser cell. This option changes the fitting method; it does not fill missing cells.

The interval-scope column distinguishes component intervals and aggregate bounds. In the unbalanced workflow, repeatability's point estimate is the REML residual variance, while its chi-square bounds use pooled within-cell squared deviations with $N - c$ degrees of freedom, where c is the number of observed cells. Other component intervals use the actual-design REML profile, with an actual-design parametric bootstrap at a numerical boundary. Reproducibility, total Gage R&R and total variation sum component bounds; these are explicitly aggregate bounds, not directly profiled intervals for the total.

For positive gage variation the profile uses

$$ndc = \max\left(1, \left\lfloor \sqrt{2}\, SD_{\text{part}} / SD_{\text{GRR}} \right\rfloor\right).$$

If gage variation is zero and part variation positive, ndc is unbounded by the estimate; it is not proof of perfect measurement. When both vanish, ndc is undefined. Percentages with zero denominators are also undefined, while a zero numerator over a positive denominator remains zero.

The profile's guidance uses below 10%, 10–30% and above 30% total gage study variation, and applies the same bands to an upper-bound percentage when available. These are conventional labels rather than a substitute for measurement decision risk. Older saved reports keep their recorded results and their pre-standards HypoNum extended label; their historic treatment of degenerate zero-denominator cases can differ.

53.5 Read and retain the whole result

Run study starts the calculation; **Cancel run** stops it. Read the headline together with standard identity, decision basis, ndc, assumptions, interaction decision, charts, interval scope, warnings and row accounting.

Add to results retains the effective configuration and standard with the tables and charts. Rerunning a report uses those recorded settings, including when its standard file is absent on another machine. The next section describes the profiles and their methods in detail.

53.6 Choose and retain a measurement standard

A standard profile specifies conventions such as the spread multiplier, interaction threshold, guidance bands and available methods. HypoNum includes **AIAG MSA, 4th ed.** and **HypoNum**

extended. The former covers the documented replicable-variable methods; the latter preserves the extended crossed-study conventions. ISO 22514-7 is not currently an available profile.

A profile name identifies these calculation conventions. It does not certify complete conformity with an external standard. A user-derived profile has its own name and records its differences; changing a convention does not retain an unqualified external label.

53.6.1 A saved result keeps its governing settings

The saved result retains standard name and version, lineage, per-study choices, effective settings and an identity for those settings. A label might describe a company profile derived from AIAG with different guidance bands and a different spread multiplier.

A report rerun uses the recorded configuration. If its standard file is absent on the new machine, it uses the recorded settings and explicitly labels that fact. It does not silently substitute the currently installed default. This is particularly important when comparing studies collected under different company procedures.

53.6.2 Compare the two built-in profiles

The following settings explain why the same measurements can produce different labels, rows or uncertainty displays. A change to display precision changes presentation; a change to interaction pooling can change the estimated components.

Convention	AIAG MSA, 4th ed.	HypoNum extended
Initial spread multiplier	6; historical 5.15 allowed	6
Percent tolerance	Fixed six-SD basis	Declared spread basis
Initial guidance bands	10% and 30%	10% and 30%
Decision basis	Intended use	Total variation
Guidance display	Decision phrases and caution	Short conventional levels
ndc multiplier	1.41	$\sqrt{2}$
ndc rounding	Truncate, minimum one	Truncate, minimum one
ndc guidance	At least five	No separate threshold
Initial interaction threshold	0.25	0.05
Interaction display	Separate interaction row and explanation	Within reproducibility
Main-effect F statistics	Not reported	Reported
Confidence limits	Not reported	Component/aggregate methods described above
Initial confidence level	95% recorded	95%; per-study choice
Zero repeatability mean square	Withhold the ANOVA model	Limiting-ratio treatment
Unbalanced design	Refused	Refused unless REML is explicitly enabled
Range constants	K1/K2/K3: four decimals; Range d_2^*: five	Same stored precision
Range/average chart factors	Report-form conventions	Report-form conventions

Convention	AIAG MSA, 4th ed.	HypoNum extended
Constants beyond tabulated sizes	Calculated by the stated formulas	Same
Display precision and layout	Report-form conventions	Five significant digits; HypoNum table
Small-design warning	Fewer than ten parts; part/appraiser product at most fifteen	No matching profile warning
Assumption statement	Shown	No profile-specific statement
Appraiser role label	Appraiser (or fixture)	Operator
Out-of-control ranges	Flag and calculate	Flag and calculate
Negative part-variance estimate	Boundary zero	Boundary zero
Average/Range PV correction	Off initially	Off initially
Methods offered	ANOVA, Average/Range, Range, worksheet, pooled SD	ANOVA
Graphical analysis, Gage R, Cp impact	Offered	Not offered by this profile
Gage R readings and estimator	Ten; sample SD	Recorded convention: ten; sample SD
Worksheet constants	Section E tables	Same recorded convention
Worksheet category explanations	Shown	Not shown
Pooled-SD chart correction	Pooled SD and $c_4(\nu + 1)$	Same recorded convention
Historical descriptive h index	Shown	Not shown

AIAG permits per-study interaction thresholds from zero to one, spread multipliers 6 or 5.15, and the optional part-variation correction. HypoNum extended also permits a positive spread multiplier up to 100, confidence levels 0.51–0.999 and unbalanced REML. Other changes require a derived user standard.

A standard cannot alter mathematical identities, input validity, the units of variance and SD, or the distinction between a descriptive index and calibrated inference. In particular, a pooled SD with ν degrees of freedom uses $c_4(\nu + 1)$; choosing a label does not change that sample-size convention.

53.6.3 Create a company profile

Open **Edit > Settings... > Gage R&R standards**. The list shows built-ins and the newest version of each user standard. The selected profile's settings and explanations appear beside it; differences from its base show the original value.

Use **Save as...** to derive a new profile. Edit within the allowed setting ranges and review the differences before saving. **Edit...** followed by **Save new version** creates another version of a user profile while retaining earlier versions. Built-ins are read-only.

A profile name must be nonempty, at most sixty characters and unique. It cannot reuse a built-in name or contain AIAG or ISO as a separate letter sequence. For example, Visor is allowed but ISO_Plant is not. This prevents a modified procedure acquiring an external-standard label.

Import... checks the file before installing it; **Export...** writes a user standard for transfer. Built-ins ship with the application. User profiles live in the user's standards library, outside projects; results still retain their own effective settings. Damaged files, invalid settings or inconsistent identity checks produce an explicit error.

Consider a profile derived from AIAG with spread multiplier four and guidance boundaries 10/20. Its spread columns become two-thirds of the six-SD values, while total-variation percentages, contribution, ndc and the fixed six-SD tolerance comparison stay unchanged. The same percentage is then read against the newly declared guidance. This makes the distinction between a numerical estimate and a decision convention visible.

53.7 Methods in the AIAG profile

The principal source is AIAG's *Measurement Systems Analysis Reference Manual*, 4th edition: Chapter II Section D for guidance; Chapter III Section B for replicable-variable studies; Chapter IV Sections E and H for the two optional studies; and Appendices A–E for spread, capability impact, constants, repeatability screening and the part-variation correction. The following definitions identify HypoNum's particular conventions where printed examples or alternative software choices differ.

The profile does not cover every procedure in that reference. Bias, linearity, stability, attribute and non-replicable studies have the separate workflows described elsewhere in this manual.

53.7.1 Spread and range constants

All component calculations use standard deviations in the measurement unit. A displayed spread is a separately identified multiple of SD. The initial multiplier is six; historical 5.15 is labelled as such. Percent tolerance remains $100(6\sigma)/(USL - LSL)$ in this profile.

For m independent normal readings, $d_2(m)$ is the mean range and $d_3(m)$ its standard deviation after dividing by σ. For an average of g independent ranges,

$$d_2^*(m, g) = \sqrt{d_2(m)^2 + d_3(m)^2/g}.$$

The squared estimator $(\bar{R}/d_2^*)^2$ is unbiased for variance under those assumptions. The square root is not an unbiased SD estimator, and the independent-range result does not transfer to overlapping moving ranges.

Average and Range uses

$$K_1 = 1/d_2(r), \qquad K_2 = 1/d_2^*(k, 1), \qquad K_3 = 1/d_2^*(n, 1),$$

for r trials, k appraisers and n parts. Each is rounded to four decimal places under this profile. The Range method uses $d_2^*(2, g)$ rounded to five places. For sizes outside the published tables, the same moment formulas are evaluated numerically.

Range-chart D_4 is 3.27 for two trials and 2.58 for three, following the report form. With four or more trials it uses $1 + 3d_3/d_2$ rounded to three decimals. $D_3 = \max(0, 1 - 3d_3/d_2)$ is zero below seven trials. The average-chart factor is $A_2 = 3/(d_2\sqrt{r})$ to three decimals. These profile constants need not equal the separately documented full-precision constants of a general SPC calculation.

The SD bias factor is

$$c_4(N) = \sqrt{\frac{2}{N-1}}\,\frac{\Gamma(N/2)}{\Gamma((N-1)/2)}.$$

It is indexed by sample size, not degrees of freedom. The corresponding B_5 and B_6 are $c_4 \mp 3\sqrt{1-c_4^2}$, with the lower factor floored at zero.

53.7.2 Select the total-variation basis

The options below all define an SD, despite the name total variation. In these formulas GRR and PV also denote standard deviations.

Basis	Total-variation SD
From the study	$TV = \sqrt{GRR^2 + PV^2}$, with PV estimated from the parts.
Surrogate process	A supplied process six-SD spread divided by six; the process should be relevant and stable.
Pp target	$TV = (USL - LSL)/(6Pp)$. Ppk also depends on centering and cannot determine this spread.
Specification tolerance	$TV = (USL - LSL)/6$, the documented sorting basis when process Pp is below one.

For an external TV, part variation is $\sqrt{TV^2 - GRR^2}$. If GRR exceeds the supplied TV, the nonnegative decomposition is incompatible. PV is shown at zero and ndc at one for positive GRR, but the actual gage percentage and supplied TV remain visible and overall guidance is withheld. A boundary display does not repair an incompatible process input.

Percent EV, AV, GRR and PV equal 100 times component SD divided by TV. Variance contributions use TV squared. Percentages of SD do not sum to 100.

53.7.3 Read ndc and guidance in context

For positive GRR,

$$ndc = \max\{1, \operatorname{trunc}(1.41\,PV/GRR)\}.$$

The constant is exactly 1.41 and the value is truncated. The stated guidance is at least five categories. When GRR vanishes with positive PV, the estimate has no finite category bound; when both vanish it is undefined. Neither condition establishes a perfect measuring instrument.

For process analysis or control, guidance uses percent total variation. For sorting, it uses percent tolerance. Below 10% is the generally acceptable band; 10–30%, inclusive, calls for an application-specific decision with the customer; above 30% is the unfavourable band. Read these as guidance, not an automatic acceptance certificate. Process charts should first support stability, and final acceptance depends on the measurement's purpose and environment.

The balanced methods require every appraiser to measure every part the same number of times. Uneven or absent cells are named rather than filled. Fewer than ten parts or at most fifteen part/appraiser combinations trigger warnings. Replicable parts, negligible within-part variation and stable conditions remain assumptions to justify from the experiment.

53.7.4 ANOVA

For n parts, k appraisers and r trials, the mean squares for Part, Appraiser, Interaction and Error have degrees of freedom $n-1$, $k-1$, $(n-1)(k-1)$ and $nk(r-1)$. The interaction F statistic is MS_{AP}/MS_e. Keep interaction when its upper-tail $p < \alpha$ and pool otherwise; the initial threshold is 0.25. Pooling adds interaction and residual sums of squares and degrees of freedom, giving $nkr - n - k + 1$ pooled error degrees of freedom.

Variance component	Interaction retained	Interaction pooled
Repeatability EV^2	MS_e	MS_{pool}
Interaction INT^2	$(MS_{AP} - MS_e)/r$	0
Appraiser AV^2	$(MS_A - MS_{AP})/(nr)$	$(MS_A - MS_{\text{pool}})/(nr)$
Part PV^2	$(MS_P - MS_{AP})/(kr)$	$(MS_P - MS_{\text{pool}})/(kr)$

Negative component estimates are reported at zero with their boundary disclosed. Total gage SD is $\sqrt{EV^2 + AV^2 + INT^2}$. The result explains the interaction decision, threshold, F and p, and whether using 0.05 would change the decision. Pooling is a modelling decision, not proof of absent interaction.

Zero repeatability mean square makes the ordinary interaction test undefined. Under this profile the model and components are withheld. Only the interaction F is reported; main-effect F statistics are not part of the profile's result. Confidence limits are not supplied under this profile because the construction of the source's printed limits is not established. Their absence does not mean the estimates are exact.

Read the component table together with the interaction plot of cell means and residuals about those cell means. The row structure separates EV, AV, INT, GRR, PV, TV and ndc.

53.7.5 Average and Range

Calculate each cell's trial range and average them to obtain $\bar{\bar{R}}$. Let $\bar{X}_{\text{DIFF}}$ be the range of appraiser averages and R_p the range of part averages. Then

$$EV = \bar{\bar{R}}K_1, \qquad AV = \sqrt{(\bar{X}_{\text{DIFF}}K_2)^2 - EV^2/(nr)},$$

$$GRR = \sqrt{EV^2 + AV^2}, \qquad PV = R_pK_3.$$

A negative AV radicand gives a reported zero boundary with a warning.

The optional Appendix E part correction, initially off, replaces PV by

$$\sqrt{(R_pK_3)^2 - EV^2/(kr)}.$$

A negative corrected variance is shown as a truncated boundary estimate. The correction changes PV, TV and ndc and is named in the result.

The range limit is $D_4\bar{\bar{R}}$. Ranges above it are flagged while the calculation remains available. Investigate the readings and repeat the collection where justified; the workflow does not automatically discard observations. This method does not estimate interaction. Prefer the ANOVA analysis when interaction is a relevant concern.

53.7.6 Graphical analysis

With ANOVA and Average and Range, average and range charts can be stacked or separated by appraiser. Average limits use $\bar{\bar{X}} \pm A_2\bar{\bar{R}}$; range limits use $D_3\bar{\bar{R}}$ and $D_4\bar{\bar{R}}$, without extra trend rules.

Unlike a production control chart, an MSA average chart is expected to show substantial between-part separation when the chosen parts represent process variation. Many part averages outside the repeatability-based limits can therefore be desirable. Ranges, by contrast, investigate consistency of repeat measurement.

Additional displays include part run charts, part/appraiser scatter and whisker charts, errors and normalized histograms, average-versus-size plots, and appraiser comparison plots with the equality line. Errors use known references when supplied, otherwise part averages.

53.7.7 Range approximation

The Range method needs exactly two appraisers and one reading from each per part. It is a quick overall-gage approximation; it cannot separate repeatability and reproducibility. For g parts,

$$GRR = \bar{R}/d_2^*(2, g).$$

The percentage uses a supplied process standard deviation. A differently replicated design should use a supported Average and Range or ANOVA study.

53.7.8 Gage R screening

The Appendix D screen uses one part, one operator, ten readings and a supplied tolerance. Repeatability is the sample SD with divisor nine; the percentage is $100(6s)/(USL - LSL)$. It is preliminary and does not provide final gage acceptance.

Its I–MR chart retains acquisition order. The individuals half-width is $2.66\overline{MR}$ and the MR upper limit is $3.267\overline{MR}$ under this profile. The plotted scale is therefore moving-range based while the reported repeatability SD is the sample SD. HypoNum does not use an unspecified d_2^* adjustment for overlapping moving ranges. Ten points without a signal cannot establish stability.

53.7.9 Control-chart worksheet

The Chapter IV Section E study uses its own printed constants for sizes two through ten and calculated values beyond those sizes. The range chart has centre $\bar{R}$ and upper limit $D_4\bar{R}$; its printed two-reading factor is 3.268. The average limits are $\bar{\bar{X}} \pm A_2\bar{R}$.

The component calculations are

$$\hat{\sigma}_e = \bar{R}/d_2, \quad \hat{\sigma}_A = R_A/d_2^*, \quad \hat{\sigma}_m = \sqrt{\hat{\sigma}_e^2 + \hat{\sigma}_A^2}, \quad \hat{\sigma}_p = R_p/d_2^*.$$

Here the appraiser component is not corrected for repeatability. Signal-to-noise is $\hat{\sigma}_p/\hat{\sigma}_m$. The result shows both the unrounded 1.41 times this ratio and ndc truncated with the profile's minimum of one.

The decision guidance remains ndc at least five. The worksheet additionally explains the historical category scale: fewer than two is inadequate for process control; two supports only high/low distinctions; three adds a middle category; four or more gives better separation. Those descriptions do not replace the declared decision threshold.

53.7.10 Pooled standard-deviation study

The Chapter IV Section H study, following ASTM E691, needs at least two appraisers and three readings per cell. For each part g, let $s_{\bar{x}g}$ be the SD of appraiser means and let s_{Eg} be the root mean of their within-cell variances. With r readings per cell,

$$s_{Ag} = \sqrt{\max(0, s_{\bar{x}g}^2 - s_{Eg}^2/r)}, \qquad GRR_g = \sqrt{s_{Eg}^2 + s_{Ag}^2}.$$

The correction divides by the number of readings, not appraisers. Pooled estimates are root mean squares across parts. The first running row contains one part; actual across-part pooling begins with the second.

For m appraisers and g parts, the chart estimate is $\hat{\sigma} = s_p/c_4(\nu + 1)$ with $\nu = mg(r - 1)$. Average limits are $\bar{\bar{X}} \pm 3\hat{\sigma}/\sqrt{r}$; the S-chart centre is $c_4(r)\hat{\sigma}$, with $B_5(r)\hat{\sigma}$ and $B_6(r)\hat{\sigma}$ limits. These assume independent normal within-cell errors with common variance. A supplied historic process SD supplies the percentage denominator.

The h and k consistency displays include their dot plots, median k and the historical h summary $\bar{h}\sqrt{g}$. That summary is descriptive: it has no normal-test p-value, significance limit or calibrated acceptance interpretation. No z statistic for k is supplied. Zero denominators leave the affected statistics unavailable.

If appraisers use different sites or equipment, the study cannot separate effects that are confounded by that design. Pooled repeatability averages their within-appraiser variances; it does not isolate an equipment-only effect.

53.7.11 Measurement error's effect on Cp

The optional Appendix B panel uses observed **Cp**, not Cpk, because Cpk also contains a centering term. Let $\rho = \sigma_{\text{meas}}/\sigma_{\text{obs}}$. Under independent, zero-mean additive measurement error on the same population and timescale,

$$Cp_{\text{actual}} = \frac{Cp_{\text{observed}}}{\sqrt{1 - \rho^2}}.$$

For observed Cp 1 and $\rho = 0.30$, this model-implied correction is $1/\sqrt{0.91} = 1.048285$. It is a point estimate under the error model, not proof of capability.

On a tolerance basis, $q_{\text{tol}} = 6\sigma_{\text{meas}}/T$ and $\rho = Cp_{\text{observed}} q_{\text{tol}}$. If $\rho = 1$, no finite corrected index exists; if $\rho > 1$, the inputs are incompatible. The calculation does not clamp the radicand to create a result.

When observed Cp, observed SD and tolerance are all entered, they must agree with $Cp = T/(6\sigma_{\text{obs}})$ within 0.1%, allowing for input rounding. Otherwise correct an input or supply only the SD or tolerance basis. Within that allowance, the calculation uses one consistent set and states the adjustment, so the two bases agree. An entered GRR proportion is converted to its measurement-SD equivalent on the fixed six-SD basis.

The panel's family of curves shows observed against corrected Cp for GRR proportions 10–90%. A study marker is used only when its denominator is the observed process SD. Changing the separate study-spread multiplier does not change this fixed six-SD correction.

53.7.12 Comparing results with printed examples

Use the declared precision and method when checking a published example. HypoNum calculates with full precision except where the profile explicitly defines rounded constants or report-form factors. Intermediate rounded values can give different last digits.

The profile uses truncation for ndc, corrected component identities for the pooled-SD study and the tolerance-based Cp panel, and computed range moments where printed constants conflict with those moments. For example, the computed $d_2^*(3,3)$ is $\sqrt{2/3 + (6+\sqrt{3})/\pi} = 1.7685743\ldots$. Rounding both component moments first can give the printed 1.76858.

These conventions identify the implemented calculation. Reproducing selected worked examples is narrower evidence than independent qualification of a measurement system or full conformity with the reference manual.

CHAPTER 54

Other measurement designs

54.1 An additional site or fixture factor

Choose **Quality > Measurement systems > Expanded gage R&R (multi-factor)...** when a third factor must be represented explicitly.

Select one of four structures: Classic crossed (no site), Fully crossed, Site nested in part, or Operator nested in site. The structure describes the physical collection design; it is never inferred from the table.

The classic crossed option uses the corresponding crossed Gage R&R calculation. The other structures use REML variance components. Optional Operator-by-Part, Operator-by-Site and Part-by-Site interactions must be identifiable in the selected structure. They require at least two trials per cell and cannot duplicate a nesting term or repeatability. An unsupported declaration is refused with the reason.

Repeatability is residual variance; reproducibility combines operator-related components. Crossed site and part-by-site variation count as **Site (measurement system)** and enter total Gage R&R. Site-within-part variation instead belongs to part-to-part variation. That distinction can change the gage conclusion, so confirm what a site label means before fitting.

The component table reports variance contribution, standard deviation, study spread (initially six SDs), study-variation percentage, optional tolerance percentage and ndc. The design table reports levels, observed cells, balance and trial counts. On the REML path only repeatability has the pooled chi-square interval; other component intervals are marked unsupported. Worksheet metadata are not automatically linked in this workflow.

54.2 Nested and destructive studies

Choose **Extended gage studies (MSA)...**; Nested/destructive Gage R&R is the opening study. Assign Measurement, Part and Operator, with an optional positive tolerance.

Each part belongs to one operator. Every operator must have the same number of at least two parts, and each part the same number of at least two trials. When an actual measurement destroys a specimen, those trials must represent physically comparable material within the declared part. Do not create fictitious repeats by assigning a common label to unrelated specimens.

The model separates Operator, Part within Operator and Repeatability. Its results include repeatability, operator reproducibility, total Gage R&R, within-operator part variation and total variation. This workflow uses a six-SD study spread and the 10/30% study-variation guidance; tolerance adds a separate percentage.

Nested part and operator effects cannot provide the same information as measuring identical parts across all operators. Interpret reproducibility within the design's comparability assumptions.

54.3 Linearity and bias

In Extended gage studies choose **Gage linearity and bias**. Select a numeric reference-value column and optionally enter tolerance. At least six complete readings over at least three distinct known reference values are required.

For each reading, bias is measured value minus reference:

$$b_i = y_i - r_i.$$

The analysis fits bias as a linear function of the reference value. It reports average bias with a 95% t interval and a test against zero, and tests whether bias changes across the reference range. The displayed linearity is the fitted slope times the observed reference range.

For example, a bias model $b = -0.2 + 0.01r$ changes by $0.01(30 - 10) = 0.2$ measurement units across references 10 to 30. A small average bias can hide such a range-dependent pattern. This arithmetic illustrates the interpretation of the fitted slope.

The detailed table gives count, mean bias, standard deviation, interval and p-value at each reference; the plot adds the fitted line. With tolerance, absolute linearity is also expressed as a percentage of tolerance. A nonsignificant slope does not establish constant bias when the range or replication is inadequate. Reference values are treated as known; uncertainty in a reference requires a separate assessment.

54.4 Type-1 study of one standard

Choose **Type-1 gage study**, enter the Known reference and a positive Tolerance, and select at least ten finite repeat readings. The sample standard deviation must be nonzero.

This study fixes the spread multiplier at six and allocates 20% of tolerance to the gage. With $T = USL - LSL$, sample SD s and bias $B = \bar{y} - r$,

$$C_g = \frac{0.20T}{6s}, \qquad C_{gk} = \frac{0.10T - |B|}{3s}.$$

C_g describes repeatability relative to the allowance; C_{gk} reduces it for absolute bias. The displayed assessment is Acceptable only when both are at least 1.33.

54.4.1 Worked example: bias at one reference

Choose **File > Import data...** and select `sample-data/manual/type1-gage-twenty-five.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Quality > Measurement systems > Extended gage studies (MSA)...**. Set **Study** to **Type-1 gage study**, **Measurement** to **y**, **Known reference** to 10 and **Tolerance** to 0.6. Click **Analyze** and inspect the bias, standard deviation, C_g and C_{gk} results. The twenty-five readings have bias 0.0192352 and sample SD 0.0153854, giving $C_g = 1.299931$ and $C_{gk} = 0.883190$. Both miss the stated criterion; accounting for bias makes the second index substantially smaller. The bias test has $p \approx 1.84 \times 10^{-6}$.

The result also includes the mean, reference, bias interval, standard deviation and study spread. This one-standard study does not establish linearity across the measurement range or reproducibility across appraisers.

54.5 Check stability in acquisition order

Choose **Gage stability** in Extended gage studies. A known reference and a Run order/time column are optional; at least ten measurements are required.

With an order column, complete readings are sorted by it. Without one, current row order is used. The order field actively sorts the stability sequence, unlike an SPC field that merely labels an already ordered sequence.

The centre is the mean and the within estimate is $\overline{MR}/(2/\sqrt{\pi})$. The individuals limits are mean $\pm 3\hat{\sigma}$. Signals identify readings beyond the limits and six successively rising or falling readings. A regression on sequence position estimates drift per observation; a p-value below 0.05 adds a drift warning. It is not automatically a drift rate per hour when observation spacing varies.

A supplied reference also gives mean bias. Inspect the plot and recorded order labels together with the signal table. A short sequence without a signal cannot establish long-term stability.

54.6 A preliminary repeatability screen

The separate **Gage R repeatability screening...** command uses one part and one operator, the profile's required number of readings (ten under AIAG), and a required tolerance. It is explicitly preliminary. Its sample-SD estimate and I–MR diagnostic are described in section 53.7.8; it does not provide final gage acceptance.

CHAPTER 55

Agreement and binary gage response

55.1 Repeatability of categorical decisions

Choose **Quality > Measurement systems > Attribute Agreement (MSA)...**. Arrange the table in long form: one row per item's rating by an appraiser in a trial.

Select distinct Rating, Part/item and Appraiser columns. Trial and Known standard are optional. An explicit Trial field is preferable; otherwise order is inferred within each item/appraiser pair. The known standard must be consistent within an item.

After incomplete records are omitted, the design must be fully crossed and balanced: each appraiser rates every item once per trial. A single trial supports between-appraiser comparison, but supplies no genuine repeat observation for within-appraiser repeatability.

The initial scale is Nominal—exact agreement. Ordinal alternatives use linear or quadratic weighting for Cohen kappa. Numeric-looking categories are ordered numerically and other categories alphabetically; confirm the order shown in the result. An alphabetical order need not match the intended severity order.

55.2 Read exact agreement and kappa together

Exact agreement answers how often decisions match. For two rating streams, Cohen's kappa adjusts for agreement expected from their category margins:

$$\kappa = \frac{p_o - p_e}{1 - p_e}.$$

The chance model uses those margins; it is not a universal correction for every sampling scheme or decision cost.

Consider the following two-rater table for twenty items:

	B: fail	B: pass	Total
A: fail	8	1	9
A: pass	2	9	11
Total	10	10	20

Observed agreement is $(8 + 9)/20 = 0.85$. Expected agreement from the margins is $(9/20)(10/20) + (11/20)(10/20) = 0.50$, so $\kappa = 0.70$. The three disagreements remain important even if the summary label sounds favourable. Neither rater is a known standard in this example.

HypoNum reports within-appraiser exact agreement across trials and its Wilson 95% interval, with Cohen kappa where two streams can be compared. Between-appraiser agreement compares the same item and trial: two appraisers use Cohen kappa; more than two use Fleiss kappa.

Fleiss kappa stays unweighted even when an ordinal scale is selected. A supplied standard adds agreement, Cohen kappa and a confusion table against that reference.

Kappa can be undefined when the margins have no variation, including a study with 100% exact agreement. Read the counts and category frequencies rather than replacing the missing value by one. The displayed bands are below 0 poor; 0–0.2 slight; 0.2–0.4 fair; 0.4–0.6 moderate; 0.6–0.8 substantial; and at least 0.8 almost perfect. They describe a conventional scale, not the cost of a wrong release or rejection decision.

Use blinded, randomized trials and include relevant boundary cases. **Add to results** retains the design, settings, omitted rows, warnings and interpretation.

55.2.1 Try an ordinal agreement study

Choose **File > Import data...** and select `sample-data/manual/ordinal-agreement-thirty.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Quality > Measurement systems > Attribute Agreement (MSA)...**. Choose Rating **rating**, Item / part **item**, Appraiser **appraiser**, no Trial or Known standard, and Ordinal—linear weighting. Click **Analyze agreement** and inspect the between-appraiser agreement and weighted kappa. There are fifteen items, two appraisers and one rating from each.

Item	P	Q	Item	P	Q
S01	3	1	S09	3	3
S02	1	2	S10	3	3
S03	3	3	S11	1	2
S04	3	3	S12	3	3
S05	1	1	S13	3	2
S06	3	2	S14	3	2
S07	2	2	S15	1	2
S08	2	2			

Eight of the fifteen item pairs agree exactly, giving 53.3333%. The linear-weighted Cohen kappa is 0.381443. Weighting gives partial credit to adjacent-category disagreements, so this is not the unweighted kappa of the same table. With only one trial, within-appraiser repeatability kappa is unavailable and the result warns accordingly.

55.3 A binary gage's switching curve

Choose **Attribute gage response curve (MSA)...** from the same menu group when known magnitudes, rather than reference categories, are available. Each row supplies a numeric Reference value, whole Pass count and positive whole Trials count. Repeated reference values are pooled.

Enter the known threshold at which the gage should switch. The initial event is Pass; Fail models its complement. The initial transition probabilities are 5% and 95%, and confidence is 95% (supported levels are at least 50% and below 100%).

The grouped binomial probit model is

$$P(\text{event} \mid x) = \Phi(\beta_0 + \beta_1 x).$$

Its response point at probability p is

$$x_p = \frac{\Phi^{-1}(p) - \beta_0}{\beta_1}.$$

The 50% point locates the middle of the switch. Bias is that point minus the specified threshold. Transition spread is the distance between the declared lower and upper response points; the probit scale is $1/|\beta_1|$.

For a model with $\beta_0 = -5$, $\beta_1 = 1$ and threshold 5, the midpoint is 5, bias is zero, and the 5/95% points are 3.355146 and 6.644854. Their spread is 3.289707 measurement units. These are calculations for a specified curve, not fitted readings. A zero midpoint bias does not imply a sharp switch: the spread answers a different question.

Reversing the event reverses the curve's direction. Its midpoint and spread remain unchanged, while the p response point becomes the original $1 - p$ point.

55.4 Why a threshold interval may be unbounded

Response points involve a ratio of fitted coefficients. Their uncertainty is therefore evaluated with Fieller's method, rather than a symmetric error bar that assumes the slope is well separated from zero. Bias uses the corresponding shifted interval. The spread and probit-scale intervals invert the slope's Wald interval.

When the slope cannot be distinguished from zero at the selected level, the response-point result says **No bounded Fieller interval**. The full confidence set may be disjoint or the entire real line; the display does not distinguish those shapes. The spread has no finite upper limit in this case. Do not interpret unavailable finite endpoints as a precise threshold.

These intervals use asymptotic normal uncertainty for the fitted binomial-model coefficients. They do not provide an exact finite-sample coverage guarantee.

55.5 Fit a transition curve from counted trials

Choose **File > Import data...** and select `sample-data/manual/gage-response-eleven.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Quality > Measurement systems > Attribute gage response curve (MSA)...**. The complete constructed study has twenty trials at each reference:

Reference	Passes	Trials	Reference	Passes	Trials
9.80	0	20	10.04	13	20
9.84	0	20	10.08	18	20
9.88	0	20	10.12	19	20
9.92	4	20	10.16	20	20
9.96	8	20	10.20	20	20
10.00	8	20			

Choose Reference **reference**, Pass count **passes**, Trials **trials**, modeled event Pass, known threshold 10, transition probabilities 5/95% and confidence 95%. Click **Run analysis** and inspect the midpoint, bias and transition spread.

The fitted midpoint is 9.999984 with Fieller interval [9.982337, 10.017632]. Bias is approximately −0.000016. The transition points are 9.886076 and 10.113892, giving spread 0.227816. Its interval is [0.185613, 0.294860].

The near-zero estimated bias does not make the transition spread negligible. Compare both quantities with the decision's requirements. The Pearson goodness-of-fit result is 5.212782 on nine degrees of freedom, with $p = 0.815379$. Eight of the eleven levels have an expected pass or fail count below five, so the result explicitly warns that this p-value is approximate. It is not proof of fit.

55.6 Check support and model adequacy

At least three distinct reference levels and both outcomes are required. All-pass, all-fail or separated outcomes do not have a finite fitted probit maximum under the supported model. The workflow refuses the fit and recommends more measurements in the transition region. It does not report an artificial steep curve.

Counts outside their allowed domain produce a row-specific error. Missing values are omitted and counted; excluded rows do not enter the fit. The goodness-of-fit table pools by reference level. Its Pearson chi-square has $k - 2$ degrees of freedom for k levels. A p-value below 0.05 warns about mismatch and reports dispersion; the coefficient intervals are not inflated for that extra variation.

Also review warnings for fewer than two mixed-outcome reference levels, small expected counts and response points outside the tested range. The plot shows observed proportions, the fitted curve with a pointwise band, the threshold and midpoint, together with trial counts. Coefficients, response points, residuals and fit diagnostics accompany it.

Add to results retains those tables, the curve and its setup. An agreement study and a magnitude-based response curve answer different questions; neither alone certifies compliance with a named measurement-system standard.

CHAPTER 56

Specifications and capability

Capability compares process output with a requirement. A specification limit comes from engineering or a customer; a control limit describes expected process behaviour. Neither can substitute for the other.

Establish how the characteristic is measured, which specification applies, and whether the process is stable before interpreting an index as evidence about future production. The guided process-quality study in section 61.3 links these decisions.

56.1 Define the requirement

Choose **Quality > Define characteristic / apply spec...**. Import a CSV, TSV or TAB file containing variable names and their limits. Preview the source, assign roles and inspect the proposed changes before applying them.

Control	Meaning and initial choice
Variable	Source field naming each target measurement column; a likely name field is suggested.
LSL, USL	Lower and upper specification fields. At least one side is required; a blank value means that side is deliberately absent.
Target, units, revision	Optional information identifying the intended centre, measurement unit and requirement revision.
Effective from/to	Optional ISO dates defining the applicable period.
Product/customer	Optional information identifying the applicable product or customer.
Transpose sheet	Initially off; use when the source arranges variables across columns and its first column should become headings.
Ignore whitespace	Initially on; matching ignores surrounding and internal whitespace without renaming the measurement column.
Case insensitive	Initially on; matching ignores letter case.

Variable, LSL and USL roles must use distinct source fields. The preview displays at most ten rows and ten columns; check the role selectors even when the automatic suggestions look plausible.

A populated limit must be a finite number. Malformed limits leave the complete specification row unresolved: an invalid upper limit must not silently become a one-sided requirement. Invalid targets, target/limit conflicts and invalid or reversed dates also need correction. Other valid rows can still be matched.

Review the list of mappings to add, change or remove. Applying replaces the mapping as one operation; cancelling retains the previous mapping. The requirement is stored with the table and can appear as an SPC overlay. Applying it changes no measurement values.

56.2 Choose the population and variation estimate

Open **Capability...** from the same menu group. Continuous capability is selected when applied specifications exist; valid specifications start a Normal calculation on opening the window. Choose the model and confirm the included observations.

For ordered individuals, overall variation is the ordinary sample standard deviation s_o. Within variation uses consecutive moving ranges:

$$\overline{MR} = \frac{1}{N-1}\sum_{i=2}^{N}|x_i - x_{i-1}|, \qquad s_w = \frac{\overline{MR}}{d_2}, \qquad d_2 = \frac{2}{\sqrt{\pi}} = 1.128379\ldots.$$

The calculation uses the full-precision constant, not the rounded display value. Moving ranges depend on order; use the actual production sequence.

With rational subgroups, the selected estimator instead combines subgroup ranges, standard deviations or pooled within-subgroup variation. For unequal sizes, ranges are first divided by their own $d_2(n_i)$, or standard deviations by $c_4(n_i)$. Averaging raw ranges of different subgroup sizes would mix different scales.

Overall variation includes changes between production occasions. A large gap between within and overall indices therefore deserves investigation of time order, grouping and process changes.

56.3 Understand the normal indices

With included mean $\bar{x}$ and within standard deviation s_w,

$$C_{PL} = \frac{\bar{x} - LSL}{3s_w}, \qquad C_{PU} = \frac{USL - \bar{x}}{3s_w}, \qquad C_{PK} = \min(C_{PL}, C_{PU}).$$

For two limits,

$$C_P = \frac{USL - LSL}{6s_w}.$$

The corresponding overall indices P_{PL}, P_{PU}, P_{PK} and P_P replace s_w with s_o.

C_P describes the tolerance width relative to within variation; C_{PK} also accounts for the mean's position between the limits. A negative side index means that the estimated mean is beyond that limit. It remains negative in the output.

A one-sided requirement has only its corresponding side index. C_P and P_P require two limits and are unavailable with one. Zero variation and other undefined quantities remain unavailable; a favourable number is not substituted. Equal limits imported from a specification sheet can still define exact-target pass/fail counts. Their tolerance width is zero, so capability indices and their intervals are withheld with an explanation. With no usable observations, conformance is not assessed.

56.4 Work through thirty measurements

Choose **File > Import data...** and select `sample-data/manual/normal-capability-thirty.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Quality > Define characteristic / apply spec...** and select `sample-data/manual/normal-capability-spec.csv`. Set **Variable** to **Variable**, **LSL** to **LSL**, and **USL** to **USL**; leave optional roles unset and **Transpose sheet** off. Click **Review changes**, check that column **value**

receives LSL 8.7 and USL 11.3, then click **Confirm changes**. Close the specification dialog. Choose **Quality > Capability > Assess capability...**, open the **Continuous** tab, set **Model** to **Normal**, **Sampling** to **Individuals in order**, **Within sigma** to **Moving range**, and **Production order** to **(source row order)**.

This numerical example uses 95% confidence and a lower-bound criterion of 1.33: enter 0.95 in **Confidence** and 1.33 in **Minimum Ppk lower bound**. To reproduce the isolated calculation, clear **Require stable SPC** and **Require acceptable MSA**, then click **Run study**. Inspect the row for **value**. That setup supplies no stability or measurement-system evidence; it is an illustration of the indices, not a process-qualification study.

Quantity	Result
Included observations	30
Mean	9.988813
Within standard deviation	0.203834
Overall standard deviation	0.346648
C_P	2.125914
C_{PK}	2.107619
P_P	1.250067
P_{PK}	1.239309
Observed values outside specifications	0

The lower side is slightly more limiting because the mean is below the midpoint 10. For example,

$$C_{PL} = \frac{9.988813 - 8.7}{3(0.203834)} \approx 2.10762.$$

The within indices look much better than the overall indices because the within spread is smaller. That contrast is a reason to inspect the ordered process, not permission to report only the larger index. Zero observed defects among thirty values does not imply a zero population defect probability.

56.5 Keep estimation uncertainty visible

Normal C_{PK} and P_{PK} intervals use a large-sample approximation:

$$\hat{C} \pm z_{1-\alpha/2}\sqrt{\frac{\hat{C}^2}{2\nu} + \frac{1}{9N}}.$$

Here N is the sample size used for the mean and ν describes the standard-deviation estimator. For P_{PK}, $\nu = N - 1$. The lower endpoint of a two-sided 95% interval uses $z_{.975} = 1.959964$; it is not a one-sided 95% lower bound.

For within variation, the calculation accounts for the estimator used:

- With $m = N - 1$ span-two moving ranges, define $A = \pi/2 - 1$ and $B = \sqrt{3}/2 + \pi/12 - 1$. Then $v = \{mA + 2(m-1)B\}/m^2$ and $\nu = 1/(2v)$. Adjacent moving ranges share an observation.
- For the average of g subgroup estimates $R_i/d_2(n_i)$, $v = g^{-2}\sum_i\{d_3(n_i)/d_2(n_i)\}^2$.
- For the average of $s_i/c_4(n_i)$, $v = g^{-2}\sum_i\{1/c_4(n_i)^2 - 1\}$. These two methods also use $\nu = 1/(2v)$.
- The uncorrected pooled standard deviation uses $\nu = \sum_i(n_i - 1)$.

Here v is the estimator's variance divided by the underlying σ^2, under independent normal observations. The resulting degrees of freedom are not rounded. Even when v is calculated exactly, the index interval remains approximate. Coverage can differ from nominal, particularly near exact centering where the minimum of the two side indices is not smooth. Unavailable within uncertainty does not fall back to the overall standard deviation. The result records sample size, estimator and degrees of freedom.

Bissell's 1990 example uses fifteen subgroups of five with limits 5.49 and 5.79. Its overall sample-standard-deviation index is called C_{pk} in that paper; it corresponds to this manual's P_{PK}. Its range, subgroup-SD and pooled estimates correspond to within C_{PK}. The paper prints index ± 2 standard errors; a configured 95% normal interval here uses 1.959964 instead. For its pooled row, the printed standard error is 0.086, while recalculation from the stated formula and $\nu = 60$ gives about 0.0868. Preserve those conventions when comparing rounded published results.

56.6 Separate conformance from a capability claim

Observed nonconformance counts measurements beyond the supplied limits. Predicted nonconformance integrates a fitted distribution beyond those limits and multiplies by one million to obtain PPM. These quantities answer different questions. Z_{bench} expresses the modelled nonconformance through a standard normal tail; no 1.5-sigma shift is applied.

The result separates conformance, stability, MSA and capability status. For a Normal-model qualification claim, the lower endpoint of the two-sided P_{PK} interval must meet the declared criterion and the required stability and MSA evidence must pass. The reported C_{PK} interval is not the qualification gate. Use the customer or process requirement for the threshold.

Require stable SPC and **Require acceptable MSA** initially remain selected. Keep them selected when the intended claim depends on that evidence. Disabling an evidence requirement does not demonstrate that its assumption holds.

Inspect the characteristic's detailed sigma estimates, probability plot, fitted parameters, bounds, sample guidance and warnings. A rejected normality check, a failed distribution fit, too little variation or too few observations leaves affected conclusions warned or unavailable.

56.7 Counts: defectives and defects

The Attribute tab uses event counts and exposure rather than individual continuous measurements.

Choice	Input and interpretation
Binomial defectives (initial)	Nonnegative whole defective-unit counts and positive inspected counts; defectives cannot exceed inspected units.
Poisson defects	Nonnegative whole defect counts and positive exposure. **Use exposure column** is initially on; switching it off assigns unit exposure to every row.
Confidence level	Fixed at 95% in this dialog.

For Binomial data,

$$\hat{p} = \frac{\text{total defective units}}{\text{total inspected units}}.$$

The result gives a Wilson 95% interval, PPM defective and yield $100(1-\hat{p})$, assuming independent inspected units. Five defective units among 200 inspected give $\hat{p} = 0.025$, 25,000 PPM and observed yield 97.5%. A defective unit counts once even if it has several defects.

For Poisson data,

$$\hat{u} = \frac{\text{total defects}}{\text{total exposure}}.$$

The exact Poisson rate interval, defects per million exposure units and modelled zero-defect yield $100e^{-\hat{u}}$ assume a constant Poisson rate. Eight defects in 200 units of exposure give $\hat{u} = 0.04$ and zero-defect yield 96.079% for one exposure unit. This modelled yield is different from counting defective units.

Z.bench converts yield fraction Y to a standard-normal score, $\Phi^{-1}(Y)$, when $0 < Y < 1$. For binomial data $Y = 1 - \hat{p}$; for Poisson data $Y = e^{-\hat{u}}$. No 1.5-sigma shift is added. At yield zero or one the score is unavailable rather than a finite number. Missing/non-finite counts or denominators omit the corresponding pair; invalid complete counts or nonpositive denominators are errors.

56.7.1 Worked example: incomplete count and exposure pairs

Choose **File > Import data...** and select `sample-data/manual/attribute-counts-four.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Quality > Capability > Assess capability...**, open the **Attribute** tab, set **Count model** to **Binomial**, **Events** to **defectives**, and **Inspected** to **inspected**. Click **Run study**. The attribute calculation uses 95% confidence: three complete rows give $3/150 = 0.02$, yield 98% and $Z_{\text{bench}} = 2.05375$, with Wilson 95% bounds 0.006825 and 0.057147. In the same window, set **Count model** to **Poisson**, **Events** to **defects**, enable **Use exposure column**, and select **exposure** for the exposure column. Click **Run study** again: a different set of three complete rows gives $9/30 = 0.3$, zero-defect yield 74.0818%, $Z_{\text{bench}} = 0.64587$, and rate bounds 0.137179 and 0.569493. Each run omits one incomplete pair, without turning a missing count into zero.

Use **Run study** to refresh the selected calculation and **Add to results** to retain the completed study, settings and warnings.

56.8 Capability for nonnormal measurements

Choose a distribution from the measurement mechanism and the observed shape, then examine its fit. A model chosen solely for a favourable capability index gives a misleading assessment. The **Model** choice in Capability is explicit and saved with the study; the calculation does not silently select another family when fitting fails.

Model	Support and interpretation
Normal	Whole real line; within indices use the selected within estimator, overall indices use the sample standard deviation.
Lognormal, Weibull, Gamma	Strictly positive measurements; zero-location distributions fitted in original units. Overall percentile indices are reported; within indices are unavailable.
Logistic	A symmetric distribution on the whole real line with fitted location and scale; indices use original-unit percentiles.

Model	Support and interpretation
Box–Cox	Positive observations transformed with an estimated power. Normal capability uses transformed values and supported transformed limits. The power and any unsupported nonpositive limit are disclosed.
Distribution-free	Observed nonconformance is available without a fitted family. Extreme percentile indices require at least 741 usable rows; a smaller sample does not receive extrapolated tail indices.
Johnson	Choose SU, SB or shifted SL explicitly. Each has a different support and transformation.

56.8.1 Use percentile indices on the original scale

For a fitted distribution with quantiles q_p,

$$P_p = \frac{USL - LSL}{q_{.99865} - q_{.00135}}, \qquad P_{pl} = \frac{q_{.5} - LSL}{q_{.5} - q_{.00135}}, \qquad P_{pu} = \frac{USL - q_{.5}}{q_{.99865} - q_{.5}},$$

and P_{pk} is the minimum of the available side indices. The two-sided width index requires both limits. A skewed distribution has different distances from its median to its two tail quantiles; replacing both by $3s$ would impose a different model.

For example, if a fitted distribution has lower tail quantile 2, median 5 and upper tail quantile 11, with limits 1 and 12,

$$P_p = 11/9 = 1.2222, \quad P_{pl} = 4/3 = 1.3333, \quad P_{pu} = 7/6 = 1.1667, \quad P_{pk} = 1.1667.$$

This is an arithmetic illustration of the percentile definition. Its centre and tails must come from the same fitted distribution.

Predicted PPM combines the fitted lower and upper tail probabilities. Observed nonconformance remains a separate count. A specification outside a bounded distribution's support still has a well-defined original-unit tail probability: no mass below a lower support boundary, and all mass below an upper support boundary. The calculation preserves those tail directions.

56.8.2 Interpret the Johnson families

Johnson SU has unbounded support. For parameters $a, b > 0$, location ξ and scale $\lambda > 0$,

$$z = a + b \operatorname{asinh}\{(x - \xi)/\lambda\}.$$

Johnson SB has support $(\xi, \xi + \lambda)$ and

$$u = (x - \xi)/\lambda, \qquad z = a + b \log\{u/(1 - u)\}.$$

The shifted SL family has support (ξ, ∞) and uses a shifted-lognormal shape $s > 0$:

$$z = \frac{\log\{(x - \xi)/\lambda\}}{s}.$$

This corresponds to Johnson's $\gamma^* + \eta \log(x - \xi)$ with $\eta = 1/s$ and $\gamma^* = -\log(\lambda)/s$.

For all three families the transformation can be expressed as $z = \Phi^{-1}\{F_J(x)\}$ and inverted with $x = F_J^{-1}\{\Phi(z)\}$. The fitted parameters and support accompany the result. Values at or outside

an SB boundary map to the corresponding infinite tail; they are not ordinary finite transformed measurements.

As a direct transformation example, SU with $a = 0$, $b = 1$, $\xi = 0$, $\lambda = 1$ maps $x = 0$ to $z = 0$ and $x = \sinh(1) = 1.175201$ to $z = 1$. SB with the same parameters maps $x = 0.5$ to $z = 0$ and $x = e/(1 + e) = 0.731059$ to $z = 1$. These are supplied-parameter calculations, not fitted models.

To create a transformed column, open the operation catalogue with **Alt+O**, choose **Johnson transformation**, and enter the family and parameters explicitly. The form initially suggests source **x**, target **Transformed**, SU, $a = b = 1$, location zero and scale one. Choose the actual source and a new target name; this action applies the supplied parameters without fitting them. To change parameters, undo the operation and create a new column.

The operation preserves source-row correspondence and exclusions. Creating a transformed column does not run capability, and a capability study does not alter the source measurements.

56.8.3 Understand uncertainty and fitted-model conditions

Fitting rejects insufficient or constant data, invalid support and non-finite or inadmissible parameters. Positive-support families cannot use nonpositive observations. The selected model's failure remains visible instead of triggering a substitute family.

For fitted-model P_{pk} bounds, at least twenty usable observations allow 399 seeded bootstrap refits. Too few valid refits make the bound unavailable. The initial confidence level is 95% and the minimum lower-bound criterion is 1.33.

The displayed goodness-of-fit checks are descriptive. Johnson's Shapiro–Wilk check uses scores obtained after estimating the transformation on the same rows. The other fitted families also use fitted-data diagnostics. These are not calibrated nominal qualification tests, and their p-values should not be read as if the distribution had been fixed independently of the data.

A passing fitted-model bound with the required stability and MSA evidence is labelled conditional on the fitted model. It does not establish unconditional predictive capability. Review support, probability plots, original-unit tails and process knowledge together.

Add to results records limits, model, parameters, uncertainty, diagnostics and the population used. The source revision and exclusion state must still match the calculation. Reopening retains the saved fit; an explicit replay refits the selected family to the supplied data.

56.8.4 Sources

Slifker and Shapiro, *The Johnson System: Selection and Parameter Estimation* (1980), *Technometrics* 22(2), pp. 239–246, gives supplied-parameter SU and SB examples on p. 243. Agreement with those coordinate transformations does not by itself establish the accuracy of maximum-likelihood fitting or family choice. The family selection in this workflow remains explicit.

56.9 Within and between batch capability

A batch can be internally consistent while batch centres vary substantially. This study separates those two sources of variation instead of averaging the batches' capability indices.

Choose **Quality > Capability > Within/between batch capability...**. Select the numeric response, batch identifier and at least one finite specification limit. The dialog uses 95% confidence. Its recorded minimum criterion initially equals 1.33, but this workflow does not make a qualification or pass/fail decision.

A batch identifier denotes a grouping in a random-effects model; it is not automatically a rational subgroup or a time window. Each retained batch needs at least two complete finite observations. Smaller batches are dropped and listed. At least two retained batches and five retained observations are required. Missing, non-finite and excluded rows are accounted for separately from dropped-small-batch rows.

56.9.1 The variance-components model

For observation j in batch i,

$$y_{ij} = \mu + u_i + \varepsilon_{ij}, \qquad u_i \sim N(0, b), \qquad \varepsilon_{ij} \sim N(0, e).$$

Batches and within-batch errors are assumed independent. The fitted nonnegative REML components define

$$\sigma_{\text{within}} = \sqrt{\hat{e}}, \qquad \sigma_{\text{total}} = \sqrt{\hat{e} + \hat{b}}.$$

The reported indices use the analyzed grand mean and the indicated spread. Inspect the labels: the overall calculation uses total variance, not the average of individual batch indices.

At an estimated zero between-batch variance, the residual variance is recomputed as $\text{SST}/(N-1)$ on the retained observations. Within and total point estimates then coincide, and the boundary is disclosed. This differs from truncating an ANOVA component while leaving its residual mean square unchanged. A small positive fitted component is retained.

56.9.2 Three batches with a common within spread

Choose **File > Import data...** and select `sample-data/manual/batch-capability-nine.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Quality > Capability > Within/between batch capability...**. Set **Response column** to **y**, **Batch column** to **batch**, **Lower specification limit** to **6**, and **Upper specification limit** to **18**. Click **Run capability**. Inspect both the within-batch and total-spread results; they describe different sources of variation. Its complete constructed dataset is:

Batch	Measurements		
A	9	10	11
B	11	12	13
C	13	14	15

Choose response **y**, grouping **batch**, LSL 6 and USL 18. The mean is 12. The fitted within variance is 1 and between-batch variance is $11/3$, so

$$\sigma_{\text{within}} = 1, \qquad \sigma_{\text{total}} = \sqrt{14/3} = 2.160247.$$

The centred within C_p and C_{pk} are both 2.0000; the total-spread indices are both 0.925820. Between-batch shifts substantially change the conclusion.

Only three batches give weak information about a population of batches. The simultaneous 95% within-C_p interval is approximately $[0.7371, 3.3933]$ and the total-C_p interval is $[0.0908, 3.3933]$. The corresponding C_{pk} intervals are much wider, about $[-13.7387, 3.3933]$ for both in this example. Their negative lower endpoint is retained. The favourable within point estimate does not remove the uncertainty about the process mean or between-batch variation.

56.9.3 How the simultaneous bounds are constructed

These are conservative Gaussian simultaneous bounds for within and total C_p and C_{pk}. Conditional on the retained batch sizes and under the stated independent Gaussian model, all four intervals cover together with probability at least $1 - \alpha$, including at $b = 0$. They need not be centred on the constrained-REML estimates and are not equal-tail intervals.

Let G be the number of batches, $N = \sum_i n_i$, $d = N - G$, W the within-batch sum of squares, m_i the batch means, and $Q = \sum_i (m_i - \bar{m})^2$ their unweighted sum of squared deviations. The reported centre is $\hat{\mu} = \sum_i n_i m_i / N$. Writing $q(k, p)$ for a chi-square quantile,

$$e_L = \frac{W}{q(d, 1 - \alpha/6)}, \qquad e_U = \frac{W}{q(d, \alpha/6)},$$

$$b_L = \max\left(0, \frac{Q}{q(G-1, 1-\alpha/6)} - \frac{e_U}{n_{\min}}\right),$$
$$b_U = \max\left(0, \frac{Q}{q(G-1, \alpha/6)} - \frac{e_L}{n_{\max}}\right).$$

The mean interval is

$$\hat{\mu} \pm z_{1-\alpha/6} \sqrt{e_U/N + b_U \sum_i n_i^2 / N^2}.$$

Within indices take their extremes over $[e_L, e_U]$ and this mean interval; total indices use $[e_L + b_L, e_U + b_U]$. C_p has no mean term. One-sided specifications do not define it. C_{pk} retains its sign when taking these extrema.

The construction uses the chi-square law of W/e, bounds on the distribution of Q from the smallest and largest retained batch sizes, and normality of $\hat{\mu}$. Allocating the six tail errors with a union bound establishes the joint level. A lower bound of zero for b does not declare its estimate zero.

No within-batch variation, or an unrepresentable endpoint, makes the affected bounds unavailable with a warning. Previously saved results retain their recorded interval convention; check that convention before comparing them with a new result. If an older setup description calls this a large-sample normal approximation, use the uncertainty method identified in the result: the current calculation is the simultaneous construction above.

56.9.4 Inspect the batch table before reporting

Review each batch's count, mean, within standard deviation and mean effect relative to the analyzed grand mean. The variance view is a diagnostic display; it does not replace the fitted variance-components calculation.

Add to results retains the selected response and batch, limits, population, method, warnings, indices and batch displays. Save/reopen preserves those results; replay uses the recorded method and supplied data context. Missing stable row identity prevents applying an ambiguous exclusion request.

56.10 Review capability with stability and model fit

Quality > Capability > Capability sixpack... assembles the matching I/MR or subgroup mean/dispersion charts, recent subgroup measurements, a capability histogram, the chosen model's probability plot and a capability spread display. Choose specifications, optional target, model and subgrouping before running it. The component plots use the underlying analyses' retained

estimates; assembling them does not create a new capability estimator or establish process stability.

The normal overall spread is mean plus/minus three overall SDs; the normal within spread uses within sigma. For a nonnormal model, the overall spread is its fitted 0.135th to 99.865th percentiles. It is labelled as percentiles, not as mean plus/minus three SDs. A Box–Cox inverse outside its mathematical domain is unavailable rather than clipped.

With both specification limits and a target, descriptive overall Cpm is

$$C_{pm} = \frac{USL - LSL}{6\sqrt{s^2 + (\bar{x} - T)^2}},$$

where s is the overall sample SD. Limits 4 and 16, mean 10 and SD 2 give Cpm 1 at target 10 and $1/\sqrt{1.25} = 0.8944$ at target 11. No Cpm interval is supplied by borrowing the interval for Cp or Cpk. Normal within PPM is a model projection using within sigma, distinct from observed or expected overall nonconformance.

Compare capability of populations... compares groups or populations before and after a declared split point using the same specifications, model and subgrouping. It places estimates and intervals side by side; it does not test equality of capability between populations.

Capability of several characteristics... applies a single-characteristic analysis separately to each selected column and its own limits/target. Every column needs at least one limit. Review each row's population and warning. Nothing is pooled across characteristics, and passing several individual indices does not establish the joint probability that a product meets all specifications.

56.11 Sources and interpretation

Bissell, *How Reliable is Your Capability Index?* (1990), *Applied Statistics* 39(3), pp. 331–340, develops the normal-index uncertainty approximation. Montgomery, *Introduction to Statistical Quality Control*, 8th edition, develops process stability, capability and rational subgrouping. A numerical index should be read within that process context rather than as a standalone quality certificate.

CHAPTER 57

Reliability engineering

Life-distribution studies relate failure times to the population and conditions being studied. The observation scheme matters: an exact failure time, a surviving unit and a failure known only to lie in an interval contribute different information. Start with the section that matches how the observations were collected, then consider an acceleration model or a prospective demonstration plan.

For background on censored likelihoods and extrapolation to use conditions, see NIST/SEMATECH, *e-Handbook of Statistical Methods*, Section 8.4.1.2 and Section 8.4.3.

For Kaplan–Meier survival curves and Cox regression, see the Statistics part. The methods below focus on engineering life distributions and reliability planning.

57.1 Right-censored life distributions

A test may end before every unit fails. A unit still operating at 500 hours has a lifetime *greater than* 500 hours; recording it as a failure at 500 would systematically shorten the apparent life. Right-censored analysis retains the information that the unit survived that long.

Choose **Quality > Reliability engineering > Censored life distribution families...**. The window compares exponential, Weibull, lognormal and loglogistic distributions on the same retained observations. Decide first whether the units represent one population under comparable operating conditions. Different failure mechanisms or stress conditions may require separate populations or an explicit covariate model.

57.1.1 Prepare the lifetime and status columns

Select a lifetime column with finite, strictly positive durations and a status column. Enter the status value meaning an observed failure and the value meaning right censoring; the defaults are 1 and 0. The values must differ, and the desktop compares the entered values as text. Set confidence to the desired level; its default is 95%, with supported values strictly between 50% and 100%.

At least three observations and two failures are required. Missing or non-finite lifetime entries and statuses outside the two selected values are omitted from the common fit. Explicit row exclusions remain excluded. Check the displayed counts of retained observations, failures, censored observations and omissions before reading the model comparison.

Important

Censoring must be non-informative for the model being fitted. Removing a unit because it appears about to fail conveys different information from ending every unit's test at a scheduled date. A status code alone does not make those observation mechanisms interchangeable.

57.1.2 What the likelihood uses

For recorded time t_i and failure indicator δ_i, the contribution is

$$L_i(\theta) = f(t_i;\theta)^{\delta_i} S(t_i;\theta)^{1-\delta_i}, \qquad S(t) = \Pr(T > t).$$

A failure contributes a density; a censored observation contributes a survival probability. The product of these terms is maximized within each family.

The following parameterizations distinguish scales from means and medians. All scale and shape parameters are positive.

Family	Survival function	Interpretation
Exponential	$S(t) = \exp(-t/\lambda)$	Constant hazard $1/\lambda$; mean life λ.
Weibull	$S(t) = \exp\{-(t/\eta)^{\beta}\}$	Scale η and shape β; hazard decreases, is constant or increases as $\beta < 1$, $\beta = 1$ or $\beta > 1$.
Lognormal	$S(t) = \overline{\Phi}((\log t - \mu)/\sigma)$	μ and σ describe log lifetime; median life is $\exp(\mu)$.
Loglogistic	$S(t) = \{1 + (t/\alpha)^{\beta}\}^{-1}$	Scale α is median life; β controls shape.

Here $\overline{\Phi} = 1 - \Phi$. The density is $f(t) = -S'(t)$ and the hazard is $h(t) = f(t)/S(t)$, the instantaneous failure rate conditional on survival.

For a fraction failed p, write $B_p = F^{-1}(p)$. The four percentile formulas are

$$B_p^{\mathrm{Exp}} = \lambda[-\log(1-p)], \qquad B_p^{\mathrm{Wei}} = \eta[-\log(1-p)]^{1/\beta},$$
$$B_p^{\mathrm{LN}} = \exp\{\mu + \sigma\Phi^{-1}(p)\}, \quad B_p^{\mathrm{LL}} = \alpha\{p/(1-p)\}^{1/\beta}.$$

Thus B_{10} uses $p = 0.1$ and median life uses $p = 0.5$. A B_{10} estimate is a population percentile under the fitted model, not a guaranteed minimum lifetime for a particular unit.

57.1.3 Worked example: six units with two censored observations

Choose **File > Import data...** and select `sample-data/manual/survival-six.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Quality > Reliability engineering > Censored life distribution families...**. The complete teaching dataset is:

Time	Status	Observation
5	1	Failure
8	1	Failure
8	0	Right censored
12	1	Failure
15	0	Right censored
20	1	Failure

Select **time** as lifetime, **status** as event, failure value 1, right-censored value 0 and 95% confidence. Click **Fit families**. Read each family's convergence status and compare its parameter and life-percentile estimates with the calculation below.

For the exponential family the calculation can be followed by hand. Total time on test is $5 + 8 + 8 + 12 + 15 + 20 = 68$, and four units failed. Therefore

$$\hat{\lambda} = \frac{68}{4} = 17, \qquad \hat{B}_{10} = 17[-\log(0.9)] \simeq 1.7911, \qquad \widehat{M} = 17\log 2 \simeq 11.7835.$$

The censored units contribute their full observed times to 68 without increasing the failure count. Treating them as failures would instead divide by six and fit a different observation model.

The exponential 95% scale interval is approximately 6.3804 to 45.2949. Its AIC is 32.6657. On these observations the Weibull fit has the smallest AIC, 31.4963, with fitted shape 2.33092. The difference is modest, and the sample is too small for the ranking alone to settle model choice.

There are only four failures. The result warns about uncertain tail-life estimation, and the small B_{10} estimate is largely a consequence of the exponential assumption. Compare the other families and survival diagnostics; this small example teaches censoring arithmetic and is not a demonstration that an exponential model is adequate for engineering use.

57.1.4 Uncertainty and comparison

The comparison gives each family's convergence state, parameter estimates, log likelihood, AIC, BIC, B_{10}, median and available intervals. For k fitted parameters and n retained observations,

$$\mathrm{AIC} = 2k - 2\ell, \qquad \mathrm{BIC} = k\log n - 2\ell.$$

Smaller AIC ranks a candidate favourably *among these families on these same rows*. It neither tests absolute model adequacy nor establishes a physical failure mechanism.

The exponential log-scale standard error is $1/\sqrt{r}$, where r is the number of failures. Its scale interval is

$$\hat{\lambda}\exp\left\{\pm z_{1-\alpha/2}/\sqrt{r}\right\},$$

and the corresponding percentile intervals follow by scaling. Other families use observed likelihood curvature and log-scale delta-method intervals for life percentiles. These are approximate pointwise Wald intervals; they do not include uncertainty from selecting a family by AIC.

A numerical search must reach a resolved finite fit with usable curvature. Some observations, such as coincident exact failure times with vanishing dispersion, may not support a finite estimate. An unresolved candidate retains its reason and is excluded from best-by-AIC selection. An interval that cannot be represented remains explicitly unavailable. Very small fitted dispersion also prompts a warning to check time resolution and the assumed family; the warning itself does not replace the estimate with a capped value.

57.1.5 Use the diagnostic curves and save the interpretation

Compare the fitted survival curves with the Kaplan–Meier steps. At tied failure times the survival update uses the risk set at that time, rather than treating tied failures as unrelated ordered observations. Inspect the observed time range and censoring pattern before extending a curve into the tail. The diagnostic table also reports fitted survival and hazard where defined. Its displayed points are bounded, so the plot is a diagnostic summary rather than a replacement for the input table.

Use **Add to results** to retain the comparison, censoring disclosure, settings, warnings and source population. Saving and reopening displays the saved fit without recalculating it. Refit explicitly after changing data, status meanings or exclusions, and describe why the model is appropriate alongside any reported life estimate.

57.2 Mixed exact and interval-censored life data

Periodic inspection often reveals only that a failure occurred between two visits. Other units may have failed before their first inspection or may still operate at the last visit. These records belong in the likelihood with their actual information; replacing every interval by its midpoint invents exact failure times.

Choose **Quality > Reliability engineering > Mixed censored life...**. Select distinct lower-bound and upper-bound columns and a confidence level (default 95%, strictly between 50% and 100%). HypoNum compares the same four parametric families as Section 57.1.

57.2.1 Encode what is known about each lifetime

Lower bound	Upper bound	Meaning
Positive finite time	The same time	Exact failure.
Positive finite time	Missing or positive infinity	Right censored: the unit survived through the lower bound.
Zero or missing	Positive finite time	Left censored: failure occurred by the upper bound.
Positive finite time	A larger finite time	Interval censored: failure lies between the two inspections.

A narrow interval with different endpoints remains interval censored. For these continuous lifetime families, including or excluding an individual finite endpoint does not change an interval probability. Equality of the two endpoints, however, denotes an exact failure with a density contribution.

Both bound cells missing causes that source row to be omitted. Negative lower bounds, non-positive upper bounds, reversed bounds and invalid infinities are errors. A lower bound of zero with no finite upper endpoint contains no useful positive lifetime information and is not a valid observation. Retained rows must include at least three observations and at least two exact, left-censored or interval-censored records. Right censoring alone does not satisfy the latter requirement. Check all four censoring counts and the omissions before fitting.

57.2.2 Follow one contribution of each type

For lower bound a_i, upper bound b_i, density f, distribution F and survival S, the likelihood contribution is

$$L_i(\theta) = \begin{cases} f(a_i;\theta), & a_i = b_i > 0, \\ S(a_i;\theta), & a_i > 0,\ b_i = +\infty, \\ F(b_i;\theta), & a_i = 0,\ 0 < b_i < +\infty, \\ F(b_i;\theta) - F(a_i;\theta), & 0 < a_i < b_i < +\infty. \end{cases}$$

The fitted parameters maximize $\sum_i \log L_i(\theta)$. Stable probability differences preserve narrow intervals without reclassifying them as exact observations.

Worked example

Consider the following four records, with all times in hours:

Lower	Upper	Interpretation
5	5	Exact failure at 5
6	10	Failed between inspections at 6 and 10
blank	8	Already failed at first inspection at 8
15	blank	Still operating at 15

To see the observation model, temporarily consider an exponential distribution with scale 10 hours. This is a *candidate parameter value*, not the fitted answer. Its four contributions are

$$\begin{aligned} f(5) &= 0.1e^{-0.5} \simeq 0.0606531, \\ F(10) - F(6) &= e^{-0.6} - e^{-1} \simeq 0.1809322, \\ F(8) &= 1 - e^{-0.8} \simeq 0.5506710, \\ S(15) &= e^{-1.5} \simeq 0.2231302. \end{aligned}$$

Only the exact record contributes a density, with units of inverse hours. The other three contributions are probabilities. If you convert hours to minutes, the exact density changes its units while the interval probabilities describe the same events.

Choose **File > Import data...** and select `sample-data/manual/mixed-censoring-four.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Quality > Reliability engineering > Mixed censored life...**. Select **lower** as the lower-bound column and **upper** as the upper-bound column, and retain 95% confidence. Click **Fit life models**. Inspect the censoring counts and family comparison. The output counts one record of each type. The fitted exponential scale is approximately 10.4559 hours, with interval 3.33405 to 32.7905 and $B_{10} = 1.10164$ hours. All four families converge in this example; exponential has the smallest AIC, 15.2117. The warning about fewer than five failure-indicative records applies: this is an encoding example with very limited information.

57.2.3 Read the fitted comparison

The family parameterizations and B_{10}/median formulas are those in Section 57.1. Weibull and loglogistic use positive shape and scale; lognormal uses log-time location and positive dispersion. The comparison reports convergence, log likelihood, AIC, BIC, parameter estimates and available life-percentile intervals.

An unresolved search is not evidence that a maximum cannot exist. Some datasets do lack a finite identifiable fit, however. For example, pure interval records with a common overlap may allow dispersion to shrink without identifying an ordinary interior maximum. HypoNum preserves candidate-specific failure reasons and does not rank a failed fit as best. An advisory warning for unusually small dispersion asks you to check inspection resolution and family choice.

Intervals use observed likelihood curvature and approximate normal inference on fitted or transformed parameters. B_{10} and median intervals use the delta method on log life. They are conditional on the candidate family, without family-selection coverage. Fewer than five failure-indicative observations prompts an additional uncertainty warning.

The results contain the fit comparison and censoring summary. This workflow does not supply a nonparametric interval-censored survival estimator or a survival/hazard plot. The separate Kaplan–Meier workflow is appropriate for its supported exact/right-censored observation scheme; it does not turn these interval records into nonparametric exact failure times.

57.2.4 Retain the inspection information

Use **Add to results** to retain both the fit comparison and the censoring roles/counts. Keep the inspection protocol with the report: the validity of each endpoint and the observation mechanism are substantive assumptions. An informative inspection schedule or incorrect bound can matter more than the formal width of a model-based interval.

Saved reports retain settings, source population, exclusions and warnings. Reopening shows the saved comparison. A changed source requires an explicit new fit; the saved numbers do not silently update.

57.3 Accelerated life at declared stress conditions

An accelerated test exposes units to stresses intended to produce failures sooner than normal use. The statistical task is to relate the tested lifetimes to use conditions. A smooth fitted curve is insufficient: the same failure mechanism and the declared stress relationship must remain credible over the extrapolation.

Choose **Quality > Reliability engineering > Physical accelerated life...**. The workflow fits Weibull and lognormal accelerated failure-time models to one common set of exact/right-censored observations.

57.3.1 Set the observation and stress roles

Select lifetime, event and stress columns. Lifetimes must be positive and finite; stress must be finite and vary across retained observations. The failure and right-censored values default to 1 and 0 and must differ. At least five retained observations, including at least three failures, are needed. Missing lifetime, event or stress information is omitted from the common fit, and explicit exclusions remain excluded.

Setting	Default	Meaning
Relationship	Arrhenius	Choose the stress transformation before fitting.
Stress units	Celsius	Arrhenius accepts Celsius or Kelvin. The other relationships use the entered numeric stress units.
Use stress	25	Condition at which to report life and survival.
Reference high stress	120	Comparison condition for the acceleration factor.
Confidence	95%	Level for approximate intervals; strictly between 50% and 100%.

The default use and reference numbers are temperature defaults, not recommended settings for your experiment. Enter conditions appropriate to the selected relationship and the actual test.

57.3.2 Distinguish the three relationships

Write x for the transformed stress:

$$x = \begin{cases} 1/T, & \text{Arrhenius, with absolute temperature } T, \\ \log S, & \text{inverse-power stress } S, \\ S, & \text{linear stress } S. \end{cases}$$

Celsius input is converted by $T = S + 273.15$; Kelvin is already absolute. Observed, use and reference Arrhenius temperatures must exceed zero Kelvin. Inverse-power stresses must be strictly positive. Linear stress can be negative but must be finite. A constant transformed stress cannot identify a slope.

For Arrhenius, a fitted slope on $1/T$ has a different numerical meaning from a slope on Celsius. For inverse power, the power exponent is the slope on $\log S$. Choosing a transformation declares the relationship; HypoNum does not select a physical law from the best-looking life estimate.

57.3.3 Understand the common-dispersion model

Both families use log-location

$$m_i = \beta_0 + \beta_1 x_i.$$

For Weibull, scale is $\eta_i = \exp(m_i)$ and shape k is common across stresses:

$$S(t \mid x_i) = \exp[-\{t/\exp(m_i)\}^k].$$

For lognormal, $\log T$ has mean m_i and common standard deviation σ:

$$S(t \mid x_i) = \overline{\Phi}\{(\log t - m_i)/\sigma\}.$$

Both use the density contribution for a failure and survival contribution for a right-censored observation. A common shape or dispersion means that stress changes the time scale while preserving the fitted distribution's shape. That assumption can be consequential when extrapolating.

At transformed use stress x_u, let $m_u = \beta_0 + \beta_1 x_u$. Then

$$B_{10}^{\mathrm{Wei}} = \exp(m_u)[-\log(0.9)]^{1/k}, \quad M^{\mathrm{Wei}} = \exp(m_u)[-\log(0.5)]^{1/k},$$
$$B_{10}^{\mathrm{LN}} = \exp\{m_u + \sigma\Phi^{-1}(0.1)\}, \quad M^{\mathrm{LN}} = \exp(m_u).$$

The reported acceleration factor compares use stress with the selected reference stress x_h:

$$AF_{u:h} = \exp\{\beta_1(x_u - x_h)\}.$$

It is a fitted time-scale ratio. With common dispersion within a family, the same ratio applies to corresponding life percentiles. It is not the ratio of the observed failure-time medians.

Worked example

Suppose an inverse-power relationship has fitted slope $\beta_1 = -2$, use stress 2 and reference stress 4 in the same stress units. The time-scale ratio is

$$AF_{u:h} = \exp\{-2(\log 2 - \log 4)\} = 4.$$

Every corresponding percentile of that fitted family is four times as large at stress 2 as at stress 4. This calculation illustrates the interpretation of a fitted slope; it supplies no evidence that the inverse-power relationship holds outside the tested stresses.

57.3.4 Read fit status, intervals and extrapolation together

Choose **File > Import data...** and select `sample-data/manual/accelerated-life-thirty-two.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Quality > Reliability engineering > Physical accelerated life...**. These are synthetic teaching observations, not evidence for a physical acceleration law. Select **time**, **status** and **stress**; use failure 1, censoring 0, **Linear** relationship, use stress 50, reference high stress 80 and 95% confidence. Click **Fit accelerated life** and inspect the fitted-family and use/reference-stress prediction tables. The additional **x** column is not used. There are 32 observations, with 26 failures and six censored units.

The Weibull fit has shape 3.86956, slope 0.00715852, use-stress $B_{10} = 9.15819$ and median 14.9019 time units. Its median interval is 13.3331 to 16.6553. The lognormal median is 14.2586, with interval 12.9483 to 15.7015. The Weibull acceleration factor is 0.806739, below one: the fitted positive slope makes the reference stress 80 have a longer time scale than use stress 50 in these synthetic data. The name of the analysis does not force a physically accelerating effect.

The family table shows regression coefficients, shape or dispersion, likelihood, fit status, and use-stress life and acceleration estimates. A resolved fit requires finite likelihood curvature and a satisfactory numerical optimum. Exact stress-time relations with vanishing dispersion can prevent a finite ordinary estimate. Failed candidates retain their reason instead of being silently replaced by the other family. Unusually small dispersion prompts an advisory check of measurement resolution and model assumptions.

Use-stress B_{10}, median and acceleration-factor intervals are approximate log-scale Wald intervals based on the fitted covariance and delta method. A bound remains unavailable if the needed covariance or propagated variance is unusable, or the value cannot be represented. These intervals condition on the family and stress transformation. They do not include uncertainty about a wrong acceleration mechanism.

The prediction table includes use and reference stresses with point estimates; the use-stress intervals are in the fitted-family table. An estimate outside the observed stress range is marked as extrapolation. The stress diagnostic table reports observations, failures, censored records and the median of observed *failure times* at each stress. That last summary excludes censored times and is not a censored-population median estimate.

57.3.5 Interpret survival bands correctly

Use-stress survival curves keep the families identified. The accompanying bounds are pointwise intervals at the chosen confidence level. For Weibull, delta-method uncertainty is calculated for

$$g(t) = k\{\log t - m_u\},$$

then transformed through $\exp\{-\exp(g)\}$ with endpoint order reversed. For lognormal, uncertainty in $z(t) = (\log t - m_u)/\sigma$ is transformed through $\overline{\Phi}(z)$, again reversing the order. Bounds therefore remain between zero and one.

A pointwise 95% band does not have 95% simultaneous coverage of the whole curve. It also excludes family-selection uncertainty. Unavailable bounds are identified in the survival table; the plot may omit the band while retaining the fitted curve. A family that cannot supply finite display points is disclosed rather than drawn using a substitute.

Use **Add to results** to retain the fit, stress diagnostics, curves, settings, source population and warnings. Save the project with the test protocol and explain the physical basis for the stress relationship. A reopened report displays its recorded fit; changing data or stresses requires a new calculation.

57.4 Reliability demonstration and precision plans

A demonstration plan asks how much testing is needed to support a declared reliability requirement under a specified pass/fail rule. A precision plan asks how accurately a life percentile will be estimated. These are different planning questions, even when both use a lifetime distribution.

Choose **Design > Power & sample size > Reliability demonstration and estimation plans...**. The inputs are assumptions; no active data table is required and no lifetime distribution is fitted from observations.

57.4.1 Choose the planning question

Mode	Question answered
Binomial demonstration	How many units must complete the mission time with no more than the allowed failures?
Duration conversion	Under an exponential or declared-shape Weibull model, how can test duration and sample size be exchanged?
Precision simulation	How many units make a life-percentile estimate sufficiently close to its true value with the requested probability?

Demonstration defaults are confidence 0.95, target reliability 0.90 at mission time 1000, and zero allowed failures. Distribution defaults to exponential; a Weibull plan requires a declared positive shape (default 2). The workflow does not estimate that shape. The default duration-conversion stopping scheme is **Time censored**, with **Sample size** as the search direction. Alternatively, search for duration at a fixed sample size.

Confidence and target reliability lie strictly between zero and one; mission/test times and shape are positive and finite. Allowed failures, sample sizes and search limits are bounded integers. An unmet criterion inside the declared limits is reported as infeasible; the planner does not weaken the requirement.

57.4.2 Demonstrating mission-time reliability

Let R_0 be target survival probability at the mission time, C the confidence and f the largest failure count allowed for acceptance. Under the boundary reliability R_0,

$$X \sim \text{Binomial}(n, 1 - R_0).$$

The plan selects the smallest integer n for which

$$\Pr(X \leq f) \leq 1 - C.$$

Thus a process at the boundary has at most $1 - C$ probability of passing this rule. Confidence is not the probability that the reliability of the particular population exceeds R_0 after a test.

For zero allowed failures,

$$n = \left\lceil \frac{\log(1 - C)}{\log R_0} \right\rceil.$$

Worked example

Choose **Design > Power & sample size > Reliability demonstration and estimation plans...**. No table is required. Set **Planning mode** to **Binomial demonstration**, confidence to 0.95, target reliability to 0.90, mission time to 1000 hours and allowed failures to 0 and **Maximum n** to 100000. Click **Run planning**. The required plan tests 29 units for 1000 hours each. It passes only if all 29 survive.

$$0.9^{28} \simeq 0.052335 > 0.05, \qquad 0.9^{29} \simeq 0.047101 \leq 0.05.$$

The reported achieved confidence is $1 - 0.9^{29} = 0.952899$. Twenty-eight units attain only 0.947665, below the requested 0.95.

The independence and common-population assumptions are essential. Testing 29 units from one unusually favourable batch may not support a claim about production containing substantial between-batch variation. If one failure is allowed, more units are required: with $R_0 = 0.95$ and $C = 0.90$, the corresponding binomial plan needs 77 units.

57.4.3 Exchange duration and sample size only under a model

For mission time t_0, a declared Weibull shape β gives

$$R(t) = R_0^{(t/t_0)^\beta}.$$

The exponential case has $\beta = 1$. A time-censored plan substitutes this survival probability into the binomial calculation, solving either for sample size at a fixed duration or for duration at a fixed sample size. The result displays the chosen plan and a nearby shorter or smaller plan so that the criterion can be checked.

To reproduce a duration calculation without a table, choose **Design > Power & sample size > Reliability demonstration and estimation plans...**. Set **Planning mode** to **Duration conversion**, confidence to 0.95, target reliability to 0.90, mission time to 1000, and allowed failures to 0. Choose **Exponential**, **Time censored**, and search **Test duration**; enter sample size 10 and leave optional duration bounds blank. Click **Run planning**. Ten independent units under this exponential model require

$$t = \frac{1000 \log(0.05)}{10 \log(0.9)} \simeq 2843.3 \text{ hours per unit.}$$

This equivalence depends on the exponential lifetime model. It does not authorize accelerating environmental stress without a separate stress model.

57.4.4 The exponential total-time-on-test rule

The **Failure terminated** option is restricted to the exponential model. Let $\theta_0 = -t_0 / \log R_0$ be the boundary mean life. The total accumulated time-on-test threshold is

$$T_{\text{req}} = \frac{\theta_0}{2} \chi^2_{C,\,2(f+1)}.$$

Stop and reject at the $(f+1)$-th failure; accept if the accumulated time reaches T_{req} with at most f failures. Under the specified exponential exposure model, the Poisson–chi-square identity sets the boundary probability of acceptance to $1 - C$.

To run a total-time example, open **Design > Power & sample size > Reliability demonstration and estimation plans...**. Choose **Duration conversion**, **Exponential**, and **Failure terminated (exponential)**. Enter mission time 1000, target reliability 0.90, confidence 0.95, allowed failures 1 and sample size 10; click **Run planning**. With $t_0 = 1000$, $R_0 = 0.90$, $C = 0.95$ and $f = 1$, $\theta_0 \simeq 9491.2$ hours and $T_{\text{req}} \simeq 45025$ unit-hours. The result also divides this total by the declared number of units. Treat that duration as a planning conversion: actual calendar stopping time depends on failures, the rejection rule and how exposure is accumulated. The division alone is not a general expected calendar stopping-time formula.

57.4.5 Plan percentile-estimation precision by simulation

Precision simulation defaults to percentile 0.10 (B_{10}), relative margin 0.10, required coverage 0.90, censor factor 5, seed 42 and 1000 replicates. Supported replicate counts are 100–5000.

For a true percentile q, success in one replicate means that its estimate $\hat{q}$ lies within the declared relative margin:

$$|\hat{q} - q|/q \leq r.$$

The requested coverage is the probability of this estimation event. It is separate from the confidence of a demonstration test and from Monte Carlo uncertainty in estimating that probability.

Simulation uses scale one because relative precision is scale invariant under the declared exponential or fixed-shape Weibull family. Lifetimes are right censored at the censor factor times the true percentile. Each replicate estimates the percentile with shape held fixed. A replicate with zero failures has no finite estimate and counts as unsuccessful.

The search scans candidate sample sizes upward and returns the first whose simulated success proportion reaches the target. Common random numbers link comparisons across sizes. This is the first passing size for the recorded simulation, not an exact population-power minimum. For B replicates and success proportion $\hat{p}$, the reported Monte Carlo standard error is

$$\sqrt{\hat{p}(1-\hat{p})/B}.$$

A threshold close to $\hat{p}$ can yield a different selected size under another seed; the result warns when Monte Carlo uncertainty crosses the target. Examine that uncertainty before committing to a test budget.

57.4.6 Retain assumptions with the plan

Use **Add to results** to preserve the selected mode, model, stopping scheme, assumptions, search limits, achieved criterion and any simulation seed. Record the reliability population, mission definition and failure definition in the study protocol. The calculated plan is statistical planning evidence under these assumptions, not a regulatory approval.

57.5 Broader life families and nonparametric support

The life-distribution commands distinguish exact, right-censored, left-censored and interval-censored observations. A fitted likelihood uses $f(t)$, $S(L)$, $F(U)$ and $F(U) - F(L)$ respectively. Supply interval endpoints and frequencies according to the actual observation scheme; replacing interval failures by midpoints changes the likelihood.

The extended families include Weibull, lognormal, loglogistic and exponential on positive support, and normal, logistic and smallest-extreme-value families on the real line. A fixed declared threshold shifts the positive-support families; it is not estimated and its uncertainty is not included. A right-censoring bound below that threshold contributes no information. A failure wholly outside the model support is refused. Real-support families are not silently truncated at zero.

For exponential failures $(10, 11, 12)$ at a declared threshold of 10, the shifted mean and fitted scale are 1. An exact exponential failure at its threshold is allowed; the Weibull, lognormal and loglogistic workflows require exact failures strictly above it. Check this distinction before using a common threshold across candidate families.

Quality > Reliability engineering > Nonparametric life estimate (Kaplan-Meier or Turnbull)... uses Kaplan–Meier for exact/right-censored records and explicitly identifies Turnbull when left or interval censoring requires it. Failure precedes censoring at tied KM times. KM log-minus-log intervals are unavailable at survival 0 or 1. Turnbull estimates masses on maximal intersection intervals, checks likelihood optimality and does not assign an observed failure time within each interval. Open-ended support can leave a tail unidentified. No Greenwood band is offered for Turnbull.

Quality > Reliability engineering > Distribution ID plot... fits the selected candidates to the same sample. Probability-plot positions derived from Turnbull masses use interval upper ends as a plotting convention only. Inspect the underlying intervals. The complete-data Anderson–Darling statistic and the censored restricted-domain discrepancy are different quantities. For the latter, read the time domain and fitted probability mass it covers; a small discrepancy over a small identified domain does not establish good fit over the unobserved tail. An empty comparison domain is unavailable, and no calibration p-value is supplied for that discrepancy. The program lists failed families rather than silently selecting among survivors.

57.6 Regression with life data

Quality > Reliability engineering > Regression with life data... adds numeric and categorical predictors with common dispersion. Positive-support models use $\log(T - \gamma) = x'\beta + \sigma W$; real-support models use $T = x'\beta + \sigma W$. Weibull shape is $1/\sigma$. For physical acceleration, explicit presets transform temperature to inverse kelvin or positive stress to its logarithm. Verify input units before using a preset.

The coefficient table, dispersion, likelihood and requested conditional percentiles describe a resolved maximum-likelihood fit under noninformative censoring. Percentile confidence intervals use the full fitted covariance, including coefficient–dispersion terms. They are not intervals for an individual future lifetime. Unseen factor levels are refused and numeric settings outside observed ranges are marked extrapolation. Rank deficiency, flat likelihood and unresolved optima stop fitting rather than producing a different undisclosed model.

CHAPTER 58

Recurring failures, degradation and warranty

58.1 Repairable systems

Choose **Quality > Reliability engineering > Repairable systems (recurrent events)...**. The active table holds events with unit identity and process age. A second table holds one observation window per unit, including units with no failures. Use a shared age origin, with $0 \le a_i < b_i$ and every event in $(a_i, b_i]$. Duplicate or ambiguous unit identities and events outside their windows are refused. Repairs are minimal: age does not restart. Renewal, repair-cost, frailty and covariate models are outside this workflow.

With $Y_i(t) = \mathbf{1}\{a_i < t \le b_i\}$, the mean cumulative function is

$$\hat{M}(t) = \sum_{u \le t} \frac{dN(u)}{Y(u)}, \qquad Y(u) = \sum_i Y_i(u).$$

Its robust variance aggregates contributions over independent units, including zero-event units, and uses the units contributing to each connected risk segment. Pointwise log-scale intervals require at least three such units and nonzero estimable variation. This minimum is not a coverage guarantee. A period with nobody at risk breaks the curve; estimation restarts as an increment in the next supported segment. A line never bridges unobserved exposure as though failures had been observed there.

Up to 20,000 events the influence-sum variance is evaluated in rational arithmetic before final floating conversion. Above that limit, a compensated calculation uses a numerical refusal screen. The screen is heuristic, not a proof of accuracy for every unscreened result. Unresolved variance and its interval are withheld while the MCF remains; a small positive value is not silently declared zero. The risk chart reports counts at event ages, not the entire risk-set step function. The transformed-event display is descriptive and does not provide a goodness-of-fit test for censored transformed gaps.

An optional homogeneous Poisson process estimates rate as total events over total exposure, with exact Poisson rate limits. The power-law process $\Lambda(t) = e^a t^b$ has decreasing intensity for $b < 1$ and increasing intensity for $b > 1$. This recurrent-event intensity is not a first-failure survival hazard. Power-law shape uncertainty uses profile likelihood; an unresolved numerical search limit is not a finite statistical bound. Expected-count intervals describe the expected number, not a prediction interval for the actual future count.

For three units each observed from 0 to 10 hours, with events at age 4 in unit 2 and ages 3 and 7 in unit 3, the MCF steps are $1/3, 2/3, 1$. Unit 1 has no events but remains in every risk set. The constant-rate estimate is $3/30 = 0.1$ per hour; its central 95% Poisson interval is approximately 0.0206 to 0.2922 per hour. Deleting unit 1 would inflate the rate to 0.15.

58.2 Degradation-based lifetime

Choose **Quality > Reliability engineering > Degradation-based lifetime...** when a measured trajectory approaches a known failure threshold. Declare repeated or destructive measurement, time, response, direction, known baseline y_0 and threshold h. The required dropout declaration asserts that missing readings are not caused by degradation; the software cannot verify it.

For direction $s = 1$ or -1 and $D = s(h - y_0) > 0$, the model is

$$y_{ij} = y_0 + se^{\eta_i + u_i} t_{ij} + e_{ij}, \qquad u_i \sim N(0, \tau^2), \quad e_{ij} \sim N(0, \sigma_e^2).$$

Failure is threshold crossing of the noise-free trajectory, so $\log T \sim N(\log D - \eta, \tau^2)$. It is not a stochastic first-passage model. Repeated units need at least two distinct positive observed ages. Destructive data require a calibrated positive measurement SD, at least three distinct positive ages and replication at one positive age. An unobserved whole unit is reported and removed; partially inadequate units are not silently switched to another study type.

Optional temperature acceleration uses inverse kelvin and a declared use temperature; at least two informative temperatures are needed. Extrapolation beyond their range is flagged. Negative/noisy readings and units trending opposite the declared direction remain in the likelihood.

The marginal fit checks agreement of successively more accurate integration and optimization calculations. Unresolved integration, search boundaries or information matrices withhold acceptance or intervals. A genuine $\tau = 0$ fit gives a point lifetime; numerical difficulty is not evidence for zero between-unit variability. Interior intervals for lifetime quantiles or reliability are confidence intervals for population quantities, not prediction intervals for future readings.

Residual plots against time and fitted values show conditional residuals about each unit's empirical-Bayes trajectory. They are not marginal errors for a new unit. The displayed first units are a deterministic subset in analysis order, not a random sample. Rate variation and measurement SD remain separate quantities.

For $D = 10$, rate log-mean $\eta = \log 2$ and $\tau = 0.5$, median life is 5, the 10th percentile is $5e^{0.5z_{0.1}} = 2.634$, the 90th is 9.490, and $P(T > 5) = 0.5$. At $\tau = 0$, lifetime is exactly 5 under the model and the strict survival probability $P(T > t)$ is zero at and after 5.

58.3 Warranty triangles and predictions

Choose **Quality > Reliability engineering > Warranty data pre-processing...**. Rows are shipment cohorts in shipment order. Select shipped counts and return columns in calendar-period or age-period order, period width and units. Units are assumed shipped at the period start; reporting delay is not modelled. The final calendar period is excluded as still in progress by default. Treating it as complete is an explicit alternative that can overstate exposure. Blank observed cells are unknown, not zero returns, and are refused.

Returns become interval-censored failures; unreturned units become right-censored survivors at the last included date. For a width w, a return in calendar period j from cohort i has interval $((j-i)w, (j-i+1)w]$. The published censoring table retains counts without expanding units into rows. Inspect the reconciliation of failures, survivors, excluded returns and excluded shipments.

Quality > Reliability engineering > Warranty prediction... fits Weibull, lognormal, exponential or loglogistic life to those censored data. Enter a future horizon, optional warranty duration, future shipments and unit cost. For n_j survivors of age a_j, expected failures through horizon h before warranty limit L are

$$n_j \left[1 - \frac{S(\min(a_j + h, L))}{S(a_j)} \right],$$

zero once the warranty has expired. If the denominator is numerically zero, the conditional expectation is undefined, and an all-cohort total is withheld; a labelled subtotal must not be read as the whole fleet.

For example, under an exponential life with mean 100 months, 100 survivors of any current age with at least 10 months of warranty remaining have $100(1 - e^{-10/100}) = 9.5163$ expected failures in the next 10 months. The fitted-parameter confidence interval for expected count incorporates shared parameter covariance across cohorts. It is not a prediction interval for realized returns. Multiplying by a fixed unit cost changes the units but does not add uncertainty in repair prices or claim-reporting delays.

CHAPTER 59

Attribute acceptance and tolerance intervals

Choose **Quality > Manufacturing decision tools...**. This window groups three compact procedures in tabs: **Acceptance sampling**, **Tolerance interval** and **Censored reliability**. They address different questions. The first plans a lot decision, the second estimates population coverage, and the third fits a basic Weibull lifetime model.

59.1 Plan acceptance from defective-unit counts

An attribute sampling plan chooses n units to inspect and an acceptance number c. Accept the lot if the sample contains at most c defective units; reject if it contains more. Count each defective unit once, regardless of how many defects it contains.

Enter an acceptable quality level (AQL), worse lot quality (LTPD), producer risk α and consumer risk β, all as fractions. The default values are

$$\text{AQL} = 0.01, \quad \text{LTPD} = 0.08, \quad \alpha = 0.05, \quad \beta = 0.10.$$

Leave lot size blank for a binomial plan, or enter the finite lot size for sampling without replacement.

The operating characteristic is $P_a(p)$, the chance of accepting when the defective fraction is p. The plan seeks

$$P_a(\text{AQL}) \geq 1 - \alpha, \qquad P_a(\text{LTPD}) \leq \beta.$$

For the binomial model,

$$P_a(p) = \sum_{j=0}^{c} \binom{n}{j} p^j (1-p)^{n-j}.$$

These are probabilities of the sampling rule under stated lot quality. AQL does not mean that every accepted lot has no more than that fraction defective.

Worked example

Choose **Quality > Manufacturing decision tools...** and open the **Acceptance sampling** tab; no table is required. Enter AQL 0.01, LTPD 0.08, producer risk 0.05 and consumer risk 0.10 as fractions, and leave lot size blank. Click **Find plan**. The plan inspects $n = 65$ units and accepts when at most $c = 2$ are defective. Its achieved probabilities are

$$P_a(0.01) = 0.972407, \qquad P_a(0.08) = 0.099099.$$

Thus the producer risk at AQL is about 2.76% and the consumer risk at LTPD about 9.91%, both within the requested limits.

Enter lot size 1000 with the same quality/risk inputs and click **Find plan** again. The finite-lot plan is $n = 64, c = 2$, with acceptance probabilities 0.978289 and 0.097362. This changes the

probability model to hypergeometric sampling without replacement; it is not merely a label attached to the binomial plan.

For lot size N with D defectives,

$$P_a(D) = \sum_{j=0}^{c} \frac{\binom{D}{j}\binom{N-D}{n-j}}{\binom{N}{n}},$$

with impossible count combinations contributing zero. The declared quality fractions are converted to whole defective counts by rounding their products with lot size. Inspect the achievable quality points for a small lot.

The sampling assumption is a representative random sample with an appropriate independent-unit or finite-lot model. A biased collection procedure cannot be repaired by the operating-characteristic calculation. For background and a published single-sampling example, see NIST/SEMATECH, Section 6.2.3.2.

59.2 Estimate an interval covering a population fraction

On **Tolerance interval**, select a numeric measurement column, **Method** (Normal or Distribution-free), **Population content** and **Confidence**. These last two are fractions, initially 0.99 and 0.95. **Sides** offers Two-sided, Upper bound or Lower bound for normal data; distribution-free mode uses two sides.

A tolerance interval covers a population fraction, rather than merely estimating its mean. Content 0.95 and confidence 0.95 mean a procedure designed to cover at least 95% of the population with 95% confidence. Neither level is the chance that the next single observation lies in the particular realized interval.

For a normal two-sided interval, the stated Howe HE2 approximation is

$$\bar{x} \pm ks, \qquad k = z_{(1+P)/2}\sqrt{\frac{(n-1)(1+1/n)}{\chi^2_{1-C,n-1}}}.$$

Here P is content and C confidence. This is Howe's simple equation 2.13 approximation, not the paper's recommended improved expression and not a claim of universal conservatism. The one-sided factor is the noncentral-t quantity

$$k = \frac{t_{C,n-1,\sqrt{n}z_P}}{\sqrt{n}},$$

giving upper bound $\bar{x} + ks$ or lower bound $\bar{x} - ks$. Normality and independent sampling are substantive assumptions.

Worked example

Choose **File > Import data...** and select `sample-data/manual/tolerance-thirteen.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. The file contains the integers 1 through 13 in column **x**. Choose **Quality > Manufacturing decision tools...** and open the **Tolerance interval** tab. Set **Measurement** to **x**,

Method to **Normal**, **Sides** to **Two-sided**, **Population content** to 0.95 and **Confidence** to 0.95. Click **Calculate** and inspect the interval endpoints and factor.
The sample mean is 7 and sample SD is approximately 3.89444. The HE2 factor is 3.08209, giving

$$[-5.00301,\ 19.00301].$$

The lower endpoint is negative because a fitted normal population extends below zero; HypoNum does not clip the interval to the smallest observation. This synthetic sequence illustrates the formula, not evidence of normality.

A distribution-free interval instead uses ordered sample values. For an iid continuous population and sample range at size n, its coverage confidence for content P is

$$\Pr\{\text{Binomial}(n, P) \leq n - 2\}.$$

More inward order statistics can be selected only when the required confidence remains supported.

Choose **File > Import data...** and select `sample-data/manual/tolerance-forty-six.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. The file contains integers 1 through 46 in **x**. Choose **Quality > Manufacturing decision tools...** and the **Tolerance interval** tab. Set **Measurement** to **x**, **Method** to **Distribution-free**, **Population content** to 0.90 and **Confidence** to 0.95. Click **Calculate**. The interval is $[1, 46]$, with achieved confidence 0.951996. The confidence is discrete in the chosen order statistics; it need not equal the requested level exactly. A small dataset may be insufficient even when its full observed range is used. Do not replace the requested content/confidence with an unlabelled easier target.

To plan the sample size before collecting these measurements, use Chapter 64.

59.3 The compact censored-Weibull panel

On **Censored reliability**, select **Time / cycles column**, **Status column** and **Value meaning failure** (default 1), then **Fit Weibull**. The panel uses fixed 95% confidence. For non-Boolean statuses, comparison ignores surrounding whitespace and case; all other complete statuses are treated as censored. Check spelling and coding before fitting: this panel supplies only the failure-value control. For Boolean status, blank, true or 1 as the failure entry selects True; another entry selects False. Null time/status pairs are omitted.

The two-parameter Weibull survival function is

$$S(t) = \exp\{-(t/\eta)^k\},$$

with positive scale η and shape k. Exact failures contribute density and right-censored observations contribute survival. At least three positive finite lifetimes and two observed failures are required.

Choose **File > Import data...** and select `sample-data/manual/survival-six.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Quality > Manufacturing decision tools...** and the **Censored reliability** tab. Set **Time / cycles** to **time**, **Status column** to **status**, and **Value meaning failure** to 1. Click **Fit Weibull**. Check the four failures and

two censored records, then inspect the fitted parameters and life-percentile table. The fitted shape is approximately 2.33092, scale 15.1424, $B_{10} = 5.76647$ and median 12.9392. The panel's approximate 95% B_{10} interval is $[2.44521, 13.59889]$. Its warning about fewer than five failures applies.

This compact panel uses a bounded Weibull fit: shape 0.05–100 and scale 10^{-12}–10^{15} in the entered time units. Its intervals use an approximate optimizer-derived log-parameter covariance. The separate **Censored life distribution families...** workflow (Section 57.1) compares four families and has different numerical-fit and interval conventions. Do not expect its intervals to be identical to this panel's, and investigate a fit close to a numerical bound.

59.4 Save the selected study

Enter a **Study name** and choose **Save to project** to retain the completed study with its parameters and result. **Export study...** writes the study as JSON. Keep the sampling design or inspection protocol with those results: the saved numerical settings alone cannot establish how observations were obtained.

CHAPTER 60

Acceptance sampling by variables

A variables plan uses measured values, such as diameter or strength, to decide whether to accept a lot. Its operating characteristic describes the probability of acceptance for different fractions nonconforming. This probability is a property of the plan and its assumptions; it does not tell you that a particular accepted lot contains no nonconforming units.

Choose **Quality > Acceptance sampling by variables...**. This is a planning calculation from entered assumptions. You can run it without an open data table. The supported plan has one sample and one specification limit, with independent normally distributed measurements.

60.1 State the two quality points

The *acceptable quality level* (AQL) is the fraction nonconforming at which the producer wants a high probability of acceptance. The *lot tolerance percent defective* (LTPD) is a worse quality level at which the consumer wants a low probability of acceptance. The plan meets

$$P_a(\text{AQL}) \geq 1 - \alpha, \qquad P_a(\text{LTPD}) \leq \beta,$$

where α is producer risk and β is consumer risk. AQL must be strictly smaller than LTPD. Enter fractions in the units indicated by the form: the default quality points are 1% and 8%, not 1 and 8 as probabilities.

A *known* standard deviation is a justified process value supplied before inspecting the sample. Selecting this option because a sample standard deviation is available would use the wrong operating characteristic. Choose **Unknown sigma** when the inspection sample will supply s.

Setting	Default	Interpretation
Specification side and limit	Upper; 100	Acceptability is below an upper limit or above a lower limit.
Sigma status and value	Known; 1	Known sigma must be positive. Unknown sigma uses the future sample standard deviation; the sigma entry is cleared and disabled.
AQL and LTPD	1%; 8%	Two process fractions nonconforming.
Producer and consumer risk	5%; 10%	Upper bounds on rejecting AQL quality and accepting LTPD quality.
Maximum sample size	5000	Largest sample considered in the search.

The specification limit must be finite. Risks lie strictly between zero and one. Confirm the normal-process assumption in the setup. This confirmation records an assumption; the planner has no observations with which to test it.

60.2 How the acceptance rule is constructed

Write U for an upper specification limit, σ for known process standard deviation and $\bar{x}$ for the inspection-sample mean. Accept when

$$\bar{x} + k\sigma \leq U.$$

For a lower limit L, accept when $\bar{x} - k\sigma \geq L$. A larger k requires the sample mean to lie farther inside the specification and makes acceptance less likely.

Let p be the true fraction beyond the selected limit and $z_p = \Phi^{-1}(1-p)$. Under the normal model, the known-sigma operating characteristic is

$$P_a(p) = \Phi\{\sqrt{n}(z_p - k)\}.$$

For an upper limit, for example, $\mu = U - z_p\sigma$ and $\bar{X} \sim N(\mu, \sigma^2/n)$ give this expression directly.

With unknown sigma, accept when

$$\bar{x} + ks \leq U \quad \text{or} \quad \bar{x} - ks \geq L,$$

according to the selected side. In this case

$$P_a(p) = \Pr\left\{T_{n-1,\sqrt{n}z_p} \geq \sqrt{n}\,k\right\},$$

where $T_{\nu,\delta}$ denotes a noncentral t variable with ν degrees of freedom and noncentrality δ. Estimating sigma adds uncertainty; substituting s into the known-sigma probability formula would omit it.

The planner considers integer sample sizes from 2 through the chosen maximum. For each n, the consumer-risk condition gives a lower bound on k and the producer-risk condition gives an upper bound. It selects the first sample size for which these bounds overlap. The known-sigma bounds follow from normal quantiles; the unknown-sigma bounds are solved numerically.

The displayed constant is normally rounded to four decimals and checked again against both risk requirements. When that precision cannot represent a feasible value, the result retains a more precise value and explains why. Use the reported constant; rounding it again can change the achieved risks. If the sample-size cap is insufficient, the planner reports that outcome without changing the requested risks.

60.3 Worked example: an upper dimensional limit

Worked example

Choose **Quality > Acceptance sampling by variables...**; no table is required. Set **Specification side** to **Upper limit**, **Specification limit** to 100, **Sigma status** to **Known sigma**, and **Known sigma** to 1. A measured characteristic has upper specification limit 100 and a justified known process standard deviation of 1, in the same measurement units. Use the default AQL of 1%, LTPD of 8%, producer risk of 5%, consumer risk of 10% and maximum sample size of 5000. Enter 1, 8, 5 and 10 in the respective percentage fields (not 0.01, 0.08, 0.05 and 0.10), then click **Find plan** and inspect the sample size and acceptance constant.
The plan is

$$n = 11, \qquad k = 1.8109.$$

Measure 11 independently sampled units and accept when

$$\bar{x} + 1.8109(1) \leq 100, \qquad \text{equivalently} \qquad \bar{x} \leq 98.1891.$$

The achieved acceptance probabilities are 0.956325 at AQL and 0.089154 at LTPD. Thus the actual rejection probability at AQL is about 4.37%, and the acceptance probability at LTPD is about 8.92%.
Now change only **Sigma status** to **Unknown sigma** and click **Find plan** again. The plan becomes $n = 28$, $k = 1.8212$, with acceptance probabilities 0.951477 and 0.098049. Its rule is $\bar{x} + 1.8212s \leq 100$. There is no single fixed cutoff for $\bar{x}$ because the observed s is part of the decision.

The larger unknown-sigma sample reflects the uncertainty in estimating dispersion. It is not a different specification or a relaxation of the quality requirements.

60.4 Read and retain the operating characteristic

The result gives the sample size, acceptance constant, inspection rule and achieved probabilities at both quality points. Read the operating-characteristic curve between and beyond those points. AQL and LTPD are design points on a continuous curve, not a guarantee that all intermediate-quality lots receive the decision you would prefer.

The probability table and curve identify the normal or noncentral-t calculation used. The default display includes 101 quality points, including AQL and LTPD; the supported requested range is 2 to 10001 points. Use **Add to results** to retain the plan, settings, assumptions and curve. Saving and reopening the project preserves that plan. Because its inputs are entered assumptions, changes to an unrelated data table do not alter it.

Important

The calculations assume a normal process and representative independent sampling. They include no finite-lot correction. This workflow supplies one-sided single-sample plans; it does not implement two-sided inspection or a MIL-STD-414 or ANSI/ASQ Z1.9 standards-table procedure. When a contract requires a named standard, these calculated plans do not establish conformity with that standard.

CHAPTER 61

Guided quality studies

A guided study records a chain of decisions: what is being measured, how the observations were collected, which method is appropriate, what the diagnostics show and what conclusion the evidence supports. It uses the same numerical methods as the individual analysis windows. Its additional value is the explicit connection between those decisions.

The stages appear on the left, with the selected stage on the right. Use **Back** and **Next stage** to move through them. Each stage names its current state, including **Needs input**, **Ready to run**, **Running**, **Complete**, **Complete — needs attention**, **Unavailable** or **Out of date**. Completing a stage means that its calculation or declaration is recorded; it does not automatically mean the evidence is satisfactory.

61.1 Review methods and preserve the reasoning

Before a fitting stage runs, inspect the proposed method and effective settings and choose **Accept method and settings**. If an earlier answer changes those settings, review them again. Changing the method or settings after a result exists requires a reason, so the final interpretation can be understood alongside the analysis history.

A blocking diagnostic stops the dependent calculation. An attention diagnostic permits it while retaining the warning. An unknown check remains unknown. For example, an absent measurement-system assessment does not become satisfactory because a process-capability calculation completes.

Changes to roles, declarations, exclusions or methods mark dependent stages **Out of date**. Unaffected stages keep their results; display-only labels do not force a new calculation. A changed source table makes all computed stages out of date. Previous results remain in history, but out-of-date numbers are not used for current conclusions. A cancelled or superseded run cannot replace the current study.

Suggested next steps follow the evidence. A stage link has a **Go** button; a linked study has an **Open** button, which opens or creates it with the applicable shared answers. If reserved evaluation data are reused to change an analysis, their untouched-evaluation status is lost; fresh data are needed to restore that status.

61.2 A measurement-system study

Choose **Quality > Guided study > Guided measurement-system study...**. Begin with the intended measurement decision. Process discrimination and sorting against a tolerance place different demands on the gage.

Record the characteristic, unit, instrument and method, then declare whether results are numeric measurements or categorical judgments. Numeric measurements require a further objective: repeatability/reproducibility, Type-1 bias and repeatability on one reference, linearity across the range, or stability over time. Having a reference value does not override the selected objective. Changing instrument, result type or objective makes later stages out of date; characteristic and

unit edits label the measurement study, while their effect on a linked process study is checked separately.

61.2.1 Let the physical design determine the analysis

Declared collection design	Supported analysis
Categorical judgments, with nominal or ordered categories	Attribute agreement; category codes are not treated as continuous measurements.
One reference part, repeatedly measured by one operator, with Type-1 objective	Type-1 bias and repeatability.
Known references across the operating range	Gage linearity and bias.
A reference measured over time	Gage stability.
The same parts remeasured by every operator	Crossed Gage R&R.
Different parts assigned to operators	Nested Gage R&R, with the part-assignment assumptions disclosed.
Destructive measurements on homogeneous batches or matched specimens	Nested Gage R&R, conditional on the asserted comparability.
Destructive measurements with no comparable units across operators	No supported design; revise the collection plan.

Declaring a measurement destructive is insufficient to identify a design: what is comparable across operators matters. For nested studies, operator differences can be confounded with the parts assigned to them. Exchangeable, representative, replicated parts or justified matched specimens are needed for the intended interpretation.

A Type-1 study with several parts or operators does not establish the single-part/single-operator result. Analyze those studies separately or choose a Gage R&R objective. To address more than one measurement question, start a separate **New study** for each. Do not combine metrics with different denominators into one pass/fail value.

For new collection, use **Design > Create MSA study worksheet...**. Randomized order helps protect the study from time-related effects, but choosing parts representative of the process remains a substantive task.

61.2.2 Declare the standard and audit the collected layout

For crossed Gage R&R, select **AIAG MSA 4th edition** or **HypoNum extended** and the total-variation basis: study parts, a known process spread, a target P_p with tolerance, or tolerance divided by six. The tolerance is needed for sorting, Type-1 analysis and any basis that uses it. ISO 22514-7 is unavailable. The guided crossed analysis uses ANOVA under the selected profile; other Gage R&R methods remain in the individual window (Section 53.6).

Select the measurement, part, operator and optional trial roles explicitly. The layout audit compares observations with optional planned numbers of parts, operators and trials. Inspect unbalanced cells, duplicate part/operator/trial cells, omitted rows and any trial order inferred from row order. Missing readings are not filled and unmatched observations are not deleted to make a balanced design.

The analysis retains the actual readings used, variance components, denominators and profile-specific interpretation. A withheld analysis makes the stage **Unavailable** with its reason. A larger estimated repeatability component suggests an investigation priority; it is not a significance test against reproducibility and does not identify a physical cause.

For Type-1, linearity and stability, the saved 0.05-level bias, slope and drift tests are screens. Non-significance does not establish absence of an important problem; significance does not establish practical importance. Compare estimates and intervals with the intended measurement requirement. Several unadjusted screens do not have a common simultaneous 5% error rate.

61.2.3 Know what can support a linked process decision

Only Gage R&R supplies an automatic measurement-system decision for a linked process-quality study. Crossed studies use the chosen profile and %GRR basis; nested studies use their own assessment.

For process discrimination, an acceptable %GRR band is insufficient when ndc is below the applicable guideline or unavailable. The AIAG and nested guideline is five categories; inadequate discrimination makes the decision conditional. For sorting against tolerance, low ndc does not replace the tolerance-based decision, although it warns about the sampled process range.

A Type-1 result covers one reference and operator, not reproducibility. Linearity and stability are partial evidence. Attribute agreement does not automatically establish accuracy or discrimination. These results remain not assessed for the linked capability qualification.

A completed Gage R&R or Type-1 study with named characteristic and unit can offer navigation to a linked process study. For Type-1 this link conveys navigation only. To use Gage R&R evidence, confirm the same characteristic, unit, method, instrument and range, and that the gage was assessed for process discrimination. A sorting-only decision is insufficient. The confirmed mapping is retained. Later changes to the linked characteristic, unit, instrument, intended use or result make the process measurement stage and its dependants out of date.

61.3 From measurement to stability and capability

Choose **Quality > Guided study > Guided measurement → stability → capability study...**. The sequence separates requirements, evidence and estimates.

First declare what the characteristic represents: a measurement, defective units out of inspected units, defects over exposure, or a count whose meaning is not yet established. Measurements need at least one specification limit; with two limits, LSL must be below USL. A specification expresses the requirement and never supplies a statistical control limit.

Then enter measurement-system evidence as a linked study, a declared result with its source, or **Unknown**. Unknown evidence permits descriptive calculations while preventing a capability claim.

61.3.1 Choose a chart from sampling design

Declared observations	Default chart
Individual measurements	I–MR.
Consecutive subgroups of size 2–10	$\bar{X}$–R.
Consecutive subgroups of size 11–25	$\bar{X}$–S.

Declared observations	Default chart
Subgroups named by a column	$\bar{X}$–R or S from modal subgroup size.
Defective units with inspected totals	p.
Defects with exposure	u.

np, c and Laney p′ or u′ are explicit advanced choices. An unknown count meaning must be resolved before choosing its model. Integer-valued measurements prompt a declaration check; integer storage alone does not make a variable a defect count. Subgroups larger than 25 are unsupported here. Time-weighted and multivariate charts use their separate windows.

61.3.2 Establish the Phase I population

The baseline can be all rows or the first N rows. Rule 1, beyond three sigma, is always enabled and is the guided study's default. Nelson rules 2–4 or 2–8 are explicit saved choices. These defaults differ from those of the standalone SPC window.

A signal makes the stability stage **Complete — needs attention**. No signal means no signal under the selected rules, rather than proof of stability. For the guided assessment, the default baseline adequacy screen requires at least 25 rational subgroups or 100 individual readings, unless documented historical evidence supports a shorter baseline. These are operational screening thresholds, not universal statistical guarantees. Count charts also need meaningful totals/exposure and representative time coverage.

Excluding a signal requires a recorded assignable cause. The point remains visible, while the exclusion changes the limit-estimation and capability population. For an MR signal, both contributing readings are identified; the cause determines whether the earlier, later or both readings should leave the baseline. The study never attributes the cause automatically. I–MR baseline ranges use adjacent readings that both belong to the baseline, without bridging an excluded reading.

61.3.3 Separate calculated capability from qualification

Review the chosen capability model before fitting: normal, lognormal, Weibull, gamma, logistic, Box–Cox, Johnson or distribution free. Keep within-subgroup C_p, C_{pk} distinct from overall P_p, P_{pk}.

For defective units, the model reports a binomial proportion; for defects over exposure, a Poisson rate. Qualification uses a one-sided upper bound at the declared confidence, while the two-sided interval remains a descriptive result. The binomial bound uses the Wilson convention with $z_{1-\alpha}$. For k Poisson defects over exposure E, the bound is

$$\frac{\chi^2_{1-\alpha,\,2(k+1)}}{2E}; \qquad k = 0: \quad \frac{-\log \alpha}{E}.$$

The upper endpoint of a central 95% interval is a 97.5% one-sided bound, not the 95% one-sided bound.

Enter the governing criterion and its source: a minimum lower confidence bound for P_{pk}, or a maximum upper bound for nonconforming rate per million. The guided study supplies no default criterion. Calculated indices do not depend on this decision threshold.

Not assessed.
A required criterion, measurement assessment, stability assessment or usable confidence bound is missing.

Below the criterion.
The lower P_{pk} bound is below the minimum, or the upper rate bound exceeds the maximum.

Not qualified; descriptive only.
The numerical bound meets the criterion but the measurement or stability evidence is unacceptable. Conditional measurement suitability retains that status and requires its conditions to be resolved or justified for the intended use.

Conditional on the fitted model.
A fitted non-normal model meets the criterion with acceptable supporting evidence.

Qualified.
The required numerical and supporting conditions are met.

A failed normality diagnostic can prompt comparison with a scientifically justified non-normal model. Selecting the distribution solely for its largest capability index would change the question.

Worked example

Choose **File > Import data...** and select `sample-data/manual/normal-capability-thirty.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Quality > Guided study > Guided measurement → stability → capability study...**. Use **New study** if this window already contains a different study.

1. In the question stage, choose **Characteristic: value**, **It records: Measurements (continuous)**, lower specification 8.7 and upper specification 11.3. Leave target and units blank. Click **Run this stage**, then **Next stage**.
2. Choose **Measurement evidence: Unknown**, click **Run this stage**, then **Next stage**. The attention status is expected.
3. Choose **Collected: One measurement at a time** and point labels **Row order**. Click **Run this stage**, then **Next stage**.
4. Leave **Phase I baseline** blank to use all 30 rows and leave **Documented baseline evidence** blank. Choose **Rule 1 only**, click **Run this stage**, and inspect the short-baseline diagnostic. Click **Next stage** without excluding any readings.
5. Choose **Capability model: Normal**, enter minimum Ppk lower bound 1.33, criterion source **Teaching example**, and confidence 0.95. Click **Accept model**, then **Run this stage**. Inspect the indices and the qualification explanation.

The file contains 30 individual measurements. Its indices can be calculated, but 30 readings alone do not meet this guided study's default 100-reading baseline screen. Leaving measurement suitability unknown also prevents qualification. Adding a favourable point estimate does not supply either missing piece of evidence. Use this example to see the missing-evidence messages; use an adequate baseline, measurement assessment and stated criterion for the process decision.

61.3.4 Approve limits and record the next action

The improvement stage records whether a change is needed, made or deferred, alongside the $P_p - P_{pk}$ centring gap and overall/within spread ratio. After a process change, the earlier indices describe the earlier process; collect appropriate new observations and rerun.

Approval of Phase II limits is refused while Phase I signals remain unexplained. Record approver, revision, rationale and optional process stage. Approved centre and limits are frozen. Rerunning with the same revision, even after recomputing a changed baseline, reuses the approved limits. A new revision approves the current baseline limits. Each approval retains its limit basis, rules and history.

The conclusion names the estimate, uncertainty, method, population, exclusions, major concern and next action. Actions may include resolving measurement evidence, investigating a signal, collecting data, comparing a justified model or following the monitoring reaction plan.

Add study to report retains every stage's state and key results with the conclusion and full study record. Saving the project preserves studies and the current stage. A stage saved while running reopens out of date, so an unfinished calculation cannot masquerade as a completed result.

Part VII

Design

CHAPTER 62

Designing experiments

An experiment changes selected factors deliberately so that their effects can be separated from background variation. Good analysis begins with the collection plan: the experimental unit, factor ranges, randomization, replication, blocking and the effects that must be estimable.

The **Design** menu groups **Design of Experiments (DOE)**, **Power & Sample Size** and **Simulation**. Planning tools use entered assumptions and can run without a data table. Design creation opens a new run sheet; analysis uses the responses collected on that sheet.

62.1 Choose a design for the question

Choose **Design > Design of experiments (DOE) > Design of experiments (DOE)...**. The **Create design** and **Analyze design** tabs separate planning from fitting.

Design family	Typical purpose and principal consideration
Full factorial	Study all combinations of declared factor levels, including interactions supported by the run count.
Regular fractional factorial	Reduce runs through a fraction of a two-level factorial. Read the defining relation and alias groups before interpreting an effect.
General factorial	Study factors with declared numeric or categorical levels; categorical contrasts differ from a linear numeric slope.
Plackett–Burman or definitive screening	Screen several factors with a limited run budget. The supported model and aliasing determine which conclusions can be separated.
Central composite or Box–Behnken	Fit a quadratic response surface within a planned quantitative region.
Simplex lattice or simplex centroid	Study mixture proportions constrained to sum to one, with mixture-model terms.
Two-level split plot	Separate hard-to-change whole-plot factors from subplot factors; the two error strata affect inference.
Constrained D- or I-optimal	Select runs from a bounded candidate region for a specified model, optimizing parameter information or prediction precision respectively.

A fractional design saves runs by confounding some effects. A screening result therefore does not establish that every associated interaction is absent. A response-surface design addresses curvature that a simple two-level main-effects model cannot identify. Mixture proportions require their own parameterization because increasing one component must decrease others.

The factor controls change with family: resolution/foldover, axial placement, mixture order, hard-to-change factors and constraints appear where applicable. The validated run count updates as factors, replicates or centre points change.

62.2 Create a reproducible run sheet

For numeric factors, enter physically feasible **Low** and **High** settings. They map to coded values -1 and $+1$:

$$x = \frac{\text{setting} - (\text{Low} + \text{High})/2}{(\text{High} - \text{Low})/2}.$$

General-factor levels are entered with semicolons; mixture proportions must sum to one. An optimal-design constraint can use factor names, arithmetic, comparisons and logical `and`, `or`, `not`; for example, `Temperature+Pressure<=250`. Such a constraint is meaningful only if its units and physical interpretation make sense.

Randomization is enabled by default and retains its seed. Leave it enabled unless the execution requires a justified restricted order. Randomization does not repair systematic failure to follow the planned settings. A replicate is another execution of the experimental condition; repeated readings of one experimental unit do not automatically provide independent experimental replication.

Creating the design opens a new editable dataset without overwriting the active table. Enter responses in the run sheet and save the project to preserve its design metadata. A measurement-system collection sheet is available separately from **Design > Create MSA study worksheet...** (Chapter 52).

62.3 Analyze responses using the design's model

For a generated sheet, choose the response. The restored family metadata identifies its factors and model: main/two-factor terms for factorials, main effects for screening, quadratic terms for definitive screening, a full quadratic response surface, categorical general-factor terms, quadratic Scheffé mixture terms, or a whole-plot random-intercept split-plot model. Fractional results retain defining relations and aliases; aliased effects are not independent estimates.

For an imported or prepared sheet, use existing-table mode. Select 2–6 numeric factors and one of:

- main effects;
- main effects and all two-factor interactions;
- a full quadratic response surface.

Enter finite Low/High bounds containing the observations. These also bound prediction displays and optimization. An optional numeric run-order column supports order diagnostics; a numeric or categorical block supplies fixed block effects. If the model is rank deficient, simplify it or collect additional informative runs. A successful calculation cannot recover effects that the design does not identify.

62.3.1 Worked example: why an interaction changes the recommendation

Choose **File > Import data...** and select `sample-data/manual/factorial-eight.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported

order. Choose **Design > Design of experiments (DOE) > Design of experiments (DOE)...** and open the **Analyze design** tab. This imported CSV opens in existing-table mode because it has no saved generated-design definition. These synthetic responses form a 2^2 design with two executions at each setting:

A	B	y
10	1	50.0303
20	1	53.0651
10	3	41.1580
20	3	53.9118
10	1	48.6747
20	1	54.1154
10	3	39.9531
20	3	51.4940

Set **Response** to **y**, **Response type** to **Continuous**, and **Model** to **Main + interactions**. In **Available factors**, select **A** and click **Add**; select **B** and click **Add**. Set A's **Low**/**High** to 10/20 and B's to 1/3. Set **Run order** to **(Row position)**, **Block** to **(None)**, confidence to 95%, and leave **Analyze variability (mean + SD)** off. Click **Analyze design** and inspect the coded coefficient and fit-summary tables.

The fitted coded model is

$$\hat{y} = 49.0503 + 4.096275x_A - 2.421075x_B + 1.9774x_Ax_B.$$

The main effects are twice their coded coefficients: A = 8.19255, B = −4.84215; the interaction effect is 3.9548. Residual degrees of freedom are 4, residual standard deviation is 1.13128, and $R^2 = 0.976467$.

The four cell means explain the interaction:

	B=1	B=3
A=10	49.35250	40.55555
A=20	53.59025	52.70290

Increasing B reduces the mean by 8.79695 at A=10 but by only 0.88735 at A=20. Reporting only B's average effect would hide this difference. Within the specified rectangle, the largest fitted value is 53.59025 at A=20, B=1. Confirm that setting with additional data before treating it as an established process optimum.

The fitted model has a separate mean for every cell, so replication estimates pure error but leaves no separate lack-of-fit degrees of freedom for this four-cell model. A high R^2 is not a substitute for additional settings when curvature is the scientific concern.

62.3.2 Use the diagnostics to challenge the model

Results include coefficients/effects and intervals, term tests, Pareto plots, rank and balance, detectable-effect or power summaries, fitted response displays, actual-versus-predicted values and residual diagnostics. Check residuals against run order for drift and against fitted values for changing spread. Read normal-probability plots in light of sample size.

Pure-error/lack-of-fit and centre-curvature results require the relevant replication and design structure. They are unavailable when the design does not separate those components. Prediction surfaces remain inside the declared region, but being inside a rectangular range does not itself establish adequate joint support or physical feasibility.

62.4 Respect the whole-plot error in split-plot experiments

With hard-to-change factors, several subplot observations share a whole plot. A simple model is

$$y_{ij} = x_{ij}^{\mathsf{T}}\beta + b_i + \varepsilon_{ij}, \quad \mathrm{Var}(b_i) = \tau^2, \quad \mathrm{Var}(\varepsilon_{ij}) = \sigma^2.$$

More subplots improve within-plot information; they do not create more independent whole plots.

The fit estimates whole-plot and residual variances by restricted maximum likelihood, allowing the whole-plot variance to reach zero. **Split-plot inference** defaults to **Kenward-Roger**: covariance is adjusted for estimated-variance uncertainty, with coefficient-specific denominator degrees of freedom and t inference. The alternative **Asymptotic Wald** uses normal-reference inference. Under Kenward–Roger, Pareto bars use equivalent normal scores so one threshold identifies the terms with $p < \alpha$.

At an estimated zero whole-plot variance, coefficients remain available but their standard errors, tests and intervals are withheld with the boundary explanation. A design with no degrees of freedom in a required stratum, or an exact fit to all within-plot observations, cannot support the ordinary variance calculation.

Two lack-of-fit questions have distinct scope:

Subplot lack of fit.
A least-squares F comparison with a full treatment model includes fixed whole-plot intercepts. It requires nested models and within-plot replication, not balance. It can detect only within-plot lack of fit and can remain available even if the mixed fit is refused.

Omnibus lack of fit.
The Goos–Gilmour approach fits separate treatment means with the whole-plot random intercept and any replicate blocks, estimates variance under that full model, then tests the proposed model's restrictions with a Kenward–Roger F. It spans both strata. It is unavailable when that full model's whole-plot variance is at zero.

62.5 Binary responses and variability

62.5.1 Events rather than continuous responses

Choose **Response type: Binary (0/1 or events)**. A 0/1 column represents one trial per run. For grouped counts, select an integer trials column with $0 \leq \text{events} \leq \text{trials}$. The default link is logit; probit is an explicit alternative.

For logit,

$$\log\{p/(1-p)\} = x^{\mathsf{T}}\beta.$$

Coefficients and their Wald intervals are on the link scale; response displays, predicted settings and goal calculations use probabilities. Grouped counts and equivalent event-level observations give the same coefficient estimates for the same design. Separation and failed convergence are disclosed as failures. Binary split-plot responses are not supported in this workflow.

62.5.2 Model the spread as well as the mean

For continuous responses, select **Analyze variability (mean + SD)**. The mean fit remains separate. Exact factor combinations define settings; their replicated observations supply counts, means and standard deviations. Singleton or zero-spread settings cannot support the dispersion fit and are identified as excluded.

The variability model assumes independent, normally distributed replicates within each setting. With $\nu = n - 1$ degrees of freedom, it fits $\log(s^2) + \log(\nu/2) - \psi(\nu/2)$, where ψ is the digamma function, using degrees-of-freedom weights. Coefficient covariance uses the log-variance sampling variance $\psi_1(\nu/2)$, where ψ_1 is the trigamma function. Coefficient tests and intervals use an approximate normal reference, which can be poor with few replicates. Coefficients are on the log-variance scale; their standard-deviation multipliers describe relative spread. Exponentiating fitted log variance produces SD predictions, not unbiased estimates of SD. Changing measurement units rescales those predictions, not the relative effects. Earlier saved analysis and optimization recipes retain their recorded numerical conventions; newly created analyses use the corrected log-variance moments.

The variability card reports usable settings, effects, predicted-SD profiles and settings with the highest/lowest predicted spread. Insufficient replication or rank deficiency withholds that fit. A standard-deviation goal in optimization must be positive and uses the **SD** flag; it cannot also be a binary-probability goal.

62.6 Optimize several responses without hiding an unacceptable one

Optimize responses requires at least two completed numeric responses. For each, declare maximize, minimize or target, meaningful endpoints and a positive importance weight. Mark probability goals as **Binary (0–1)** and dispersion goals as **SD** where appropriate.

Each individual desirability d_i lies between zero and one, with linear segments corresponding to exponent-one Derringer–Suich desirabilities. The combined value is

$$D = \left(\prod_i d_i^{w_i} \right)^{1/\sum_i w_i}, \qquad w_i > 0.$$

A zero individual desirability makes $D = 0$ exactly. For equal weights, desirabilities 0.5 and 0.8 yield $\sqrt{0.4} \simeq 0.6325$. If one becomes zero, the combined value is zero regardless of the other.

When every evaluated setting has zero desirability, no candidate is recommended; a displayed setting supports inspection. Weights express preferences, not confidence or evidence of meeting a target.

The individual functions are described by NIST/SEMATECH's discussion of multiple-response desirability. The positive-weight geometric combination follows G. C. Derringer, *A Balancing Act: Optimizing a Product's Properties*, *Quality Progress* 27(6), 1994, pp. 51–58. Saved results retain their original numerical convention.

62.7 Extend an executed design

The **Augment design** tab creates a new dataset, preserving the original sheet, responses and order. Add replicate sets, centre points, a complementary fractional-factorial foldover, or optimal runs conditional on the executed design.

Replicating all settings adds pure-error information. Centre points support curvature questions; requesting zero is an explicit no-op. A foldover can separate main effects from two-factor interactions for an appropriate regular fraction, but is refused if the complementary fraction is already

present or the observed runs no longer match the recorded fraction. Optimal augmentation uses an explicit bounded candidate set and optional constraint and reports model rank, *D*-efficiency and candidate count before and after.

Only appended runs are randomized, using the retained seed. **RunOrder** and **StdOrder** continue; a new block label can distinguish the added period. The combined sheet obeys the run limit. Augmentation rank/alias reports support two-level numeric factors.

62.8 Related planning workflows

For interval precision, see chapter 64; for detection and equivalence power, chapter 65; for reliability demonstration, section 57.4. Probability calculations and seeded simulations are in chapter 27. Simulations illustrate assumptions; they do not establish them for real data.

CHAPTER 63

Robust settings and Taguchi designs

63.1 Robust parameter design

Choose **Design > Design of experiments (DOE) > Robust parameter design...** to choose controllable settings that perform well while specified noise factors vary. The workflow creates a crossed experiment, analyses collected responses and plans fresh confirmation runs. Declare 2–6 two-level numeric factors in total, including at least one control and one noise factor. Noise probabilities describe use conditions, are normalized within each factor and assume independence across noise factors. A dynamic design adds a signal with 3–10 levels and declared weights.

The default two independent replicates are whole experimental runs, not repeated readings of one run. Generated rows retain planned identities and seeded random order. Repeated readings are averaged within a run; unequal reading counts are disclosed because their mean variances can differ. Record reasons for unobserved responses. Those reasons document the loss, but do not establish that the missingness is ignorable.

The static OLS model includes all main effects and two-factor interactions of the factors coded at $-1, +1$. A dynamic model adds that basis multiplied by the signal in its original units. Full rank and positive residual degrees of freedom are required. No term is silently removed to make a fit possible. For fitted response $f(c, z)$ at control setting c and noise setting z,

$$\bar{f}(c) = \sum_z w_z f(c, z), \qquad V(c) = \sum_z w_z [f(c, z) - \bar{f}(c)]^2 + \hat{\sigma}^2.$$

Closest-to-target minimizes $(\bar{f} - T)^2 + V$. Low-response preference minimizes $\bar{f} + k\sqrt{V}$; high-response preference maximizes $\bar{f} - k\sqrt{V}$. The weight k expresses a preference, not a confidence multiplier. A dynamic design minimizes weighted squared departure from the declared target line plus residual variance.

For a simple comparison, setting A has mean 10, noise variance 4 and residual variance 1; B has mean 11, noise variance 0 and residual variance 1. At target 10 their expected squared losses are 5 and 2, respectively. B is preferred under that loss despite its larger mean bias. A different target or noise distribution can change the decision.

The ranking recommends among tested control settings. Predictions for unobserved settings are identified separately. The 999 seeded parametric bootstrap refits give pointwise intervals conditional on the fixed Gaussian model; they do not account for optimism from selecting the best setting. Failed refits or exact residual fits withhold intervals. Independent-run pure error permits a lack-of-fit comparison where its degrees of freedom and variance are adequate.

Select a ranked setting and create a confirmation plan. Enter new independent runs under the declared noise conditions. Evaluation compares their errors and observed objective with the original predictions; it does not refit the design data and relabel them confirmation. Hard-to-change factors, blocks, mixtures and constrained randomization need another experimental design.

63.2 Taguchi orthogonal arrays

Design > Design of experiments (DOE) > Taguchi design... supplies two-level, multilevel and mixed arrays, including L4, L8, L9, L12, both L16 forms, L18, L25, L27, L32 and a separately identified mixed-level L36. Assign named control factors to compatible columns. Pairwise balance makes main effects orthogonal; it does not remove interaction aliases. Read the estimable degrees of freedom and alias disclosures before assigning factors. Unused columns are not independent experimental error.

Choose repeated responses or an outer noise array and optional dynamic signal levels. The resulting table records the correspondence between each response column and its repeat, noise and signal settings. New L27 designs use the corrected standard column order; earlier recorded designs retain their recorded order on replay. The supplied L36 is not claimed to have the same column identities as a printed Taguchi L36 table.

Design > Design of experiments (DOE) > Analyze Taguchi design... offers explicitly named signal-to-noise summaries:

Smaller is better	$-10\log_{10}(n^{-1}\sum y_i^2)$,
Larger is better	$-10\log_{10}(n^{-1}\sum y_i^{-2})$,
Nominal, variance only	$-10\log_{10}s^2$,
Nominal, mean/SD scaling	$10\log_{10}(\bar{y}^2/s^2)$.

Dynamic analysis uses fitted signal slope squared over residual mean square, with a declared intercept or forced origin. Choose the formula for the physical objective and scale. Zero responses can invalidate larger-is-better; zero variance or a zero logarithm argument is unavailable. Negative dB and zero dB are valid finite results, not errors.

For responses $(1, 2, 3)$, smaller-is-better is $-10\log_{10}(14/3) = -6.6901$ dB; sample SD is 1 and mean is 2, so nominal mean/SD scaling is $10\log_{10}4 = 6.0206$ dB. These summarize different objectives on the same data.

The analysis rechecks balance after edits or exclusions. Confounded level averages remain descriptive, while orthogonal ANOVA and additive prediction are refused for a nonorthogonal layout. A saturated inner array has no independent error degrees of freedom for testing its signal-to-noise ratios; more readings within an inner run do not create those degrees of freedom. Pooling selected small effects makes subsequent tests exploratory. Confirm the chosen settings with independent runs, and consider the expected-loss robust workflow when its model and objective match the decision.

63.3 Power for the planned or actual design

Open **Design > Power & sample size > Two-level factorial and Plackett-Burman designs...** for standalone planning. Define the high-minus-low difference Δ of one effect and error SD σ. Under ± 1 coding its coefficient is $\Delta/2$. With model matrix X, effect selector c and $\nu = N - \operatorname{rank}(X)$, the noncentral t shift is

$$\eta = \frac{\Delta/2}{\sigma\sqrt{c'(X'X)^{-}c}}.$$

Blocks, aliases, retained interactions and curvature consume degrees of freedom or change the variance. Power for an aliased effect assumes its other aliased contributions absent or negligible. A saturated or confounded choice is refused. Alpha applies to the selected effect, not every model term.

In an unblocked 2^3 full factorial repeated twice, the full corner model has 16 runs, 8 residual degrees of freedom and $\eta = 2$ when $\Delta = \sigma$. Adding centre points can increase degrees of freedom; in blocked designs it can also alter effect variance through block allocation. Searches for replicates or centre points report the candidates actually examined and achieved power. A calculator-only design outside the run-sheet catalogue is labelled accordingly.

After creating a supported design, **Power for this design...** on the DOE analysis tab uses the actual included runs, recorded coding, blocks and centre flags. Edits and exclusions can change its power. The comparison with the regenerated catalogue design is reported, rather than replacing actual rows by an ideal layout. This linked command computes power only; planning a different size uses the standalone command.

CHAPTER 64

Planning for estimation precision

A precision plan starts with the uncertainty you can tolerate. For example, you may want a 95% interval for a mean with half-width at most 0.5 mm. That question differs from having 80% power to reject a null hypothesis.

Choose **Design > Power & sample size > Sample size for precision and tolerance...**. The calculation uses entered assumptions and needs no open table. The assumed spread, mean or proportion is a design input, not an estimate from data that have yet to be collected.

64.1 Match the quantity to the intended interval

Planning quantity	Target and required assumptions
Mean precision	Absolute half-width in measurement units, or half-width relative to an assumed nonzero mean; enter an assumed positive SD.
Proportion precision	Absolute half-width as a proportion, or half-width relative to an assumed positive proportion.
Normal tolerance interval	A tolerance factor or raw-unit width/extent, with normality, assumed SD, population content and confidence declared.
Distribution-free tolerance interval	Sample size supporting the requested population content and confidence; no raw-unit width is predicted.

The form initially selects mean precision, absolute criterion, target 1, confidence 0.95 and assumed SD 1. Assumed mean defaults to 1; assumed proportion to 0.5. Normal tolerance defaults are content 0.90 and two sides. The sample-size search defaults to 2–100000; supported limits are integers from 2 through 1000000.

Choose a target that has meaning in the selected units. A proportion half-width of 0.05 means five percentage points, not 0.05 percentage point. Relative precision is a fraction of the entered mean magnitude or proportion and is undefined at zero. Confidence and tolerance content lie strictly between zero and one; SD and numerical width/factor targets are positive.

64.2 Mean precision and a complete planning example

For assumed SD s_0, confidence C and sample size n, the Student-t planning half-width is

$$h_n = t_{(1+C)/2,n-1}\frac{s_0}{\sqrt{n}}.$$

An absolute target requires $h_n \le h_0$. A relative target requires $h_n/|\mu_0| \le r_0$, using an entered nonzero assumed mean μ_0.

Worked example

Choose **Design > Power & sample size > Sample size for precision and tolerance...**. No table is required. Suppose the measurement SD is expected to be 2 mm and you want a 95% mean interval with half-width no more than 0.5 mm.
Select **Mean precision**, **absolute**, target 0.5, confidence 0.95 and assumed SD 2. Set **Minimum n** to 2 and **Maximum n** to 100000. Click **Run planning** and inspect the chosen and previous sample sizes. The result is $n = 64$:

$$h_{64} = 0.499585 \text{ mm}, \qquad h_{63} = 0.503693 \text{ mm}.$$

The previous integer misses the target and 64 meets it.

This uses the assumed SD as if it were the value in the future interval formula. The actual sample SD is random, so the plan does not guarantee that every realized interval will have this width. Explore plausible larger SDs when the planning estimate is uncertain.

64.3 Proportion precision uses the Wilson convention

Let p_0 be the entered proportion, $z = z_{(1+C)/2}$ and $D = 1 + z^2/n$. The Wilson planning half-width is

$$h_n = \frac{z}{D}\sqrt{\frac{p_0(1-p_0)}{n} + \frac{z^2}{4n^2}}.$$

Absolute planning compares h_n with the target; relative planning compares h_n/p_0. The assumed proportion can range from zero to one, but a relative target requires it to be nonzero.

The half-width describes the Wilson interval's span, not equal distances on either side of the raw observed proportion. An assumed $p_0 = 0.5$ gives the largest half-width at a fixed n under this formula. For rare events, examine the discrete possible counts and an appropriate exact-binomial design as well; a continuous planning proportion does not ensure that the anticipated count is an integer.

64.4 Tolerance content is different from confidence

A confidence interval for the mean addresses a population parameter. A tolerance interval aims to cover a specified fraction P of the population, with confidence C that at least that content is covered. A 90%-content, 95%-confidence tolerance interval has two distinct percentages, not one combined level.

For a two-sided normal tolerance interval, this planner uses Howe's simple HE2 approximation, equation 2.13:

$$k_n = z_{(1+P)/2}\sqrt{\frac{(n-1)(1+1/n)}{\chi^2_{1-C,n-1}}}.$$

It is a named approximation and is not described as a universally conservative or exact normal tolerance factor.

For a one-sided normal bound, the factor is

$$k_n = \frac{t_{C,n-1,\sqrt{n}z_P}}{\sqrt{n}},$$

with a noncentral-t quantile. Choose the lower or upper side explicitly.

A factor target compares k_n with the requested factor. For assumed SD s_0, a two-sided width target compares $2k_n s_0$ with the requested full width. A one-sided target compares the extent $k_n s_0$ with its target. Consequently, a width of 2 does not have the same meaning for one and two sides. The selected convention is retained with the plan.

64.5 Distribution-free planning specifies content, not width

For independent observations from a continuous population, the full-sample range has the planner's maximum attainable confidence

$$\Pr\{\text{Binomial}(n, P) \leq n - 2\}$$

of containing at least fraction P. The search finds the first n meeting the requested confidence and records its central order-statistic convention.

Without a population scale or distribution, this calculation cannot predict the interval's width in millimetres, hours or another response unit. Width and factor criteria are therefore unavailable in distribution-free mode. Ties, rounding and a discrete population require attention to the continuous-population interpretation.

64.6 Read the candidate check and preserve assumptions

Results retain the requested target, achieved precision/confidence, content where applicable, selected n, assumptions and warnings. When $n - 1$ is inside the search range, the table shows that it fails the chosen criterion. If the declared minimum binds, the result makes no claim about smaller excluded sample sizes. A target not reached at the maximum produces an explicit infeasible outcome.

Use **Add to results** to save the plan as entered planning parameters. A reopened report retains those parameters and numbers without recalculation. It has no fitted data population. To change an assumption, run a new plan and compare its sample size with the earlier one.

A calculation can be cancelled. Closing the setup cancels an active run; Escape closes the setup or result without adding a report. A calculation that exceeds the three-minute limit stops visibly without publishing a partial plan.

CHAPTER 65

Prospective power and sample size

Power is the probability that a specified test rejects its null hypothesis under a specified alternative. Planning therefore needs a scientifically meaningful effect and a variability model. A calculation based on an effect estimated after seeing the same test result is not a prospective design.

Choose **Design > Power & sample size > Extended power and power curves...**. The workflow takes entered parameters and does not use the active table. Choose the test family, design, alternative and quantity to solve for before entering numerical assumptions.

65.1 Choose a family and count the units correctly

Family	Entered effect and design convention
Raw mean / t	Mean difference and SD; one-sample, paired or independent groups. Independent groups use a common SD and pooled degrees of freedom.
Raw proportion	Alternative and null proportions; one-sample or independent groups with an arcsine/normal-reference convention.
Raw ANOVA	Group means, group sizes and common SD; group sizes determine the effect weighting and the solve reports total N.
Known-sigma Z	Mean difference and known sigma; one-sample, paired or independent groups with a normal reference.
Normal variance	True-to-null variance ratio; one-sample normal data with a chi-square reference.
Poisson rate	Alternative and null rates and exposure per observation; one-sample independent Poisson counts.
Mean equivalence (TOST)	True difference, lower/upper equivalence margins and SD; one-sample, paired or independent common-SD groups.

For one-sample and paired designs, n is respectively the number of observations or pairs. A paired mean calculation uses the SD of the within-pair *differences*, not the marginal SD of either measurement. For independent groups,

$$n_1 = n, \qquad n_2 = \lceil n \times \text{allocation ratio} \rceil.$$

The second group must have at least two observations. Independent t and TOST planning use $n_1 + n_2 - 2$ degrees of freedom. Unequal-variance or Welch planning is not substituted for this assumption.

Paired-proportion planning is unavailable because no paired binary covariance model is declared. Normal variance and Poisson planning support one sample only.

65.2 Understand the initial settings

The initial setup is **Raw mean / t**, **Required sample size**, **One sample** and a two-sided alternative. Alpha is 0.05 and target power 0.80. The sample-size entry is 30 and allocation ratio is 1. The visible search bounds are 2–1000; effect bounds are 0.001–10. Power curves default to 41 points and can have at most 201.

Family-specific initial assumptions are:

- mean difference 1 and SD/sigma 1;
- alternative/null proportions 0.6/0.5;
- ANOVA means 1 and 3, group sizes 20 and 20;
- variance ratio 2;
- Poisson rates 0.2 and 0.1 per exposure unit, with exposure 10;
- TOST true difference 0.1 and margins −0.5 and 0.5.

These are editable starting values, not recommendations for a particular study. Alpha must lie strictly between zero and 0.5, target power between zero and one, and SD/sigma be positive. The displayed confidence control defaults to 95% and supports 51–99%.

65.3 Follow a prospective mean calculation

A raw mean effect becomes

$$d = \frac{\mu_1 - \mu_0}{s}.$$

For a one-sample two-sided t test, the alternative statistic has noncentrality $\delta = \sqrt{n}\, d$ and $n - 1$ degrees of freedom. Power adds its two probabilities beyond the central-null critical values.

Worked example

Choose **Design > Power & sample size > Extended power and power curves...**. No table is required. Plan to detect a mean difference of 1 unit when the observation SD is 2 units. Select **Raw mean / t**, **One sample**, two-sided, **Confidence (%)** 95 (alpha 0.05) and target power 0.80. Enter difference 1 and SD 2. The standardized effect is $d = 0.5$.
Set **Solve for** to **Required sample size**, search bounds to 2 and 1000, and click **Run**. The result is $n = 34$ and achieved power 0.807778. Then set **Solve for** to **Power at n**, enter 30 in the sample-size field, and click **Run** again; power is 0.753965. Thus 30 observations do not meet this particular 80% target.

These probabilities depend on the entered difference and SD. Inspect a power curve over plausible sample sizes and repeat the calculation with less favourable variability assumptions. Power alone does not establish that the target effect is practically important or that the test assumptions will hold.

For known-sigma Z planning, the normal reference replaces the t reference. Select it only when the known-sigma assumption is justified. For a directional alternative, preserve the sign of the effect: a negative change is not evidence for a greater-than alternative.

65.4 Proportions, ANOVA, variance and event rates

Raw-proportion planning uses the signed arcsine effect

$$h = 2\{\arcsin\sqrt{p} - \arcsin\sqrt{p_0}\},$$

with proportions strictly inside $(0, 1)$. This is the stated normal-reference convention, not an exact binomial power calculation.

Raw ANOVA uses the entered group sizes as weights:

$$\bar{\mu} = \frac{\sum_j n_j \mu_j}{N}, \qquad f = \frac{\sqrt{\sum_j n_j (\mu_j - \bar{\mu})^2 / N}}{s}.$$

The solve reports total observations and requires N greater than the number of groups. Keep the weighting and common-SD assumptions with the plan rather than describing f as a universal property of the means alone.

Normal-variance planning uses a true-to-null variance ratio and central chi-square null critical values. Enter a ratio, not a raw null variance. The test requires the normal-population variance model.

Poisson planning combines n observations with exposure E_0 each, so total exposure is nE_0. Null rejection cutoffs are integers with tail probability no greater than alpha, or alpha/2 for each two-sided tail. The achieved null error and alternative power can therefore be conservative and can jump as n changes. A smooth continuous approximation would hide those jumps.

65.5 Equivalence asks for an interval inside the margins

TOST declares equivalence only when both one-sided tests reject. Enter the assumed true difference separately from the lower and upper margins. A true difference of 0.1 and margins $-0.5, 0.5$ means that 0.1 is the planning alternative, while the margins define acceptable equivalence; they do not all describe the same effect.

At alpha 0.05 for each one-sided test, the corresponding two-sided mean interval has 90% confidence. It must lie strictly within both margins. The setup's 95% confidence control sets the 0.05 one-sided test level; it does not change that TOST interval to a 95% interval.

Power integrates the joint acceptance event over the estimated-variance distribution, using the chi-square reference appropriate to the selected design. Multiplying the two one-sided rejection probabilities would ignore their dependence. TOST uses the two-sided selection because its rule already comprises two one-sided tests.

For an independent design, the SD is common and degrees of freedom are pooled. For paired observations it is the difference SD. A favourable planned equivalence probability remains conditional on those assumptions.

65.6 Choose a solve or curve with supported meaning

Required sample size returns an integer and recalculates its power, checking the previous integer within the feasible search range. **Power at n** evaluates a declared design. **Effect at n** is available for mean, proportion and Z families. It searches positive effects for greater/two-sided alternatives and negative effects for less alternatives.

ANOVA effect solving is unavailable because the entered means and group weighting define a separate raw-unit conversion. Variance, Poisson and TOST also have no generic effect solve or effect-axis curve: their ratio, rate and margin assumptions have distinct meanings.

Sample-size searches are bounded. Discrete Poisson and joint TOST searches scan ascending sizes rather than assuming a smooth monotone curve; they have a 20000-evaluation budget. An exhausted budget or a target outside the declared limits is an explicit failure. Do not interpret it as a calculated sample size or an invitation to lower the target without justification.

Curves retain their numerical coordinates with the settings. The display table shows at most 200 rows while the saved plot can contain all 201 supported points. These calculations are deterministic; they are not a Monte Carlo power study.

65.7 Save the design assumptions

Use **Add to results** to retain the requested and achieved values, family, alternative, allocation, effect assumptions, search limits and warnings. The source is entered planning parameters, with no observed data population. Reopening preserves that plan; changing assumptions requires an explicit new calculation.

Runs can be cancelled and stop visibly at the three-minute calculation limit. An invalid or unsupported request is reported rather than silently mapped to a different statistical test.

65.8 Correlation, regression and repeated-measures planning

The extended-power window also plans correlations, regression R^2, a balanced one-way contrast, within-subject effects, exact McNemar tests and exact one-proportion tests. Choose the family, task, direction and effect definition before entering a sample size. Its sizes mean pairs, subjects, trials, or observations per group according to the selected family.

For one Pearson correlation, Fisher's approximation treats $\operatorname{atanh}(r)$ as normal with standard deviation $1/\sqrt{n-3}$. The standardized alternative shift is

$$\delta = [\operatorname{atanh}(\rho) - \operatorname{atanh}(\rho_0)]\sqrt{n-3}.$$

Power is the shifted normal probability beyond the chosen critical values. For two independent correlations, divide their Fisher-transform difference by $\sqrt{1/(n_1-3)+1/(n_2-3)}$. The second size is rounded upward from the allocation ratio; inspect the achieved ratio. These are large-sample Pearson calculations. The separately selected exact, partial and Spearman modes below use different assumptions. Dependent-correlation comparisons are not substituted for independent samples. To reproduce the numerical case without a table, choose **Design > Power & sample size > Extended power and power curves...**. Set **Family** to **Correlation (Fisher z)**, **Solve for** to **Required sample size**, **Design** to **One sample**, and **Alternative** to **two-sided**. Enter correlation 0.3, null correlation 0, target power 0.8, confidence 95%, and search bounds 4 and 1000. Click **Run**. At two-sided 0.05, $\rho = 0.3$ against 0 requires 85 pairs for power at least 0.8 under this approximation (approximately 0.80035).

For regression, define the full and reduced population explained variances:

$$f^2 = \frac{R^2_{\text{full}} - R^2_{\text{reduced}}}{1 - R^2_{\text{full}}}, \quad \lambda = Nf^2, \qquad P\{F_{u,N-p,\lambda} > F_{1-\alpha;u,N-p}\}.$$

Here p is the full design rank including intercept, and u is the rank difference being tested. Count indicator and interaction dimensions rather than factor names. "From zero" means an intercept-only reduced model; "increase" includes at least one reduced-model predictor dimension. The noncentral F calculation conditions on the stated fixed-design effect.

A balanced contrast uses coefficients c_i summing to zero, group means μ_i, common within-group SD σ, and n observations per group. Its noncentral t shift is $\sum_i c_i\mu_i/[\sigma\sqrt{\sum_i c_i^2/n}]$ with $k(n-1)$ degrees of freedom. Scaling all coefficients changes the contrast's units but not its power. This is one planned contrast, not simultaneous power for a post-hoc family.

For a within-subject effect, the univariate planning approximation uses population Cohen's f, m measurements, correlation ρ, correction ϵ and k balanced groups:

$$\lambda = \frac{Nmf^2\epsilon}{1-\rho}, \qquad \nu_1 = (m-1)\epsilon, \quad \nu_2 = (N-k)(m-1)\epsilon.$$

For the within-by-between interaction, multiply numerator df by $k-1$. The declared total must divide equally among the groups. Sample-size searches respect that allocation. Do not enter an observed partial eta-squared as though it were the specified population f. Examine several plausible correlations and sphericity corrections when these are uncertain.

65.9 Exact binary planning and search claims

Paired-proportion planning requires both discordance probabilities p_{10} and p_{01}, nonnegative with sum at most 1. It averages conditional exact McNemar rejection over $D \sim \text{Binomial}(N, p_{10}+p_{01})$ and, given $D = d$, $B \sim \text{Binomial}(d, p_{10}/(p_{10}+p_{01}))$. No discordance means no rejection. The reported attained size conditions on the supplied total discordance probability. A one-sided plan carries a notice that the current hypothesis window performs the matching McNemar analysis only two-sided.

The one-proportion exact family enumerates counts whose exact test p-value is strictly less than α, then sums their alternative probabilities. It reports the rejection region and attained size. To reproduce the first example without a table, choose **Design > Power & sample size > Extended power and power curves...**. Set **Family** to **One proportion (exact binomial)**, **Solve for** to **Power at n**, **Design** to **One sample**, **Alternative** to **greater**, and **Test method** to **Exact binomial**. Enter proportion 0.8, null proportion 0.65, sample size 20 and confidence 95%; click **Run**. For null 0.65, alternative 0.8, 20 trials and a greater alternative at 0.05, the region is $x \geq 17$: power is 0.41145 and attained size 0.04438. Optional score-normal and continuity-corrected approximations are explicitly different procedures.

Exact discrete power need not increase at every successive integer sample size because rejection regions change. Read the result's search statement:

- A justified monotone search can state that larger sizes maintain the target.
- A bounded ascending search identifies the first qualifying size and, when the whole selected range is assessed, the earliest size maintaining the target through that upper bound only.
- A longer search that stops at its first success says stability was not assessed. Failure within a range does not prove impossibility beyond it.

For a separate sample-size example, reopen **Design > Power & sample size > Extended power and power curves...**. Choose **One proportion (exact binomial)**, **Required sample size**, **One sample**, **two-sided** and **Exact binomial**. Enter proportion 0.6, null proportion 0.5, target power 0.8, confidence 95%, and search bounds 2 and 1000. Click **Run** and inspect both reported size criteria. For exact two-sided power at alternative 0.6 against 0.5, target 0.8 and search 2–1000, the first success is 199; the target is maintained from 210 through 1000. These are two distinct planning answers.

Effect solves report both admissible sides of a two-sided null. Where a branch is not known monotone, the result is the first crossing detected on the searched grid; earlier crossings or tangencies between points are not excluded. The recorded range, direction and definition are part of the answer.

65.10 Additional correlation and allocation choices

The correlation method is an explicit choice. Exact Pearson planning numerically integrates the sampling density under bivariate normality; "exact" describes that sampling model, not distribution-free validity. The null density sets one-sided or equal-tail two-sided critical correlations, and the alternative density determines power. The supported effective sample size is 4–10000 and correlations must have absolute value at most 0.995. An unresolved integral is refused rather than replaced by Fisher's approximation.

Partial Pearson planning subtracts the number q of prespecified independent control variables from the original complete-case sample size N. The intercept is not a control. It requires $N > q+3$, full rank and the declared joint-normal model. Report original N and q, not just their difference.

Spearman planning offers Bonett–Wright and Fieller–Hartley–Pearson normal approximations on the Fisher-transform scale, with variance respectively $(1 + \rho_s^2/2)/(N-3)$ and $1.06/(N-3)$. Null variance sets the rejection boundary and alternative variance determines power. These continuous-data approximations do not guarantee finite-sample rank-test power with heavy ties. Caruso–Cliff is unavailable pending the required original-source control.

For one planned contrast, explicit positive integer allocation ratios change the sample-size unit to total N. The ratios are reduced by their greatest common divisor, and only feasible total-size multiples are considered. Variance is $\sigma^2 \sum_i c_i^2/n_i$, with residual degrees of freedom $N-k$. Inspect the actual group sizes and the previous admissible size's power.

For a repeated-measures between-group effect under equal-variance compound symmetry, population Cohen's f uses the single-measure marginal SD and $\lambda = Nmf^2/[1 + (m-1)\rho]$, with degrees of freedom $k-1, N-k$ and $\epsilon = 1$. A prespecified partial-eta-squared sum-of-squares convention instead uses $\lambda = \mathrm{df}_{\mathrm{error}}\eta_p^2/(1-\eta_p^2)$ at the declared residual degrees of freedom. These are different effect definitions; do not apply a correlation gain twice or interpret observed post-hoc power as a prospective design calculation.

65.11 A distinct paired-proportion approximation

The explicitly named Connor approximation uses total pairs N, $\psi = p_{10}+p_{01}$ and $\delta = p_{10}-p_{01}$. Its normal null variance is ψ/N, and alternative variance is $(\psi-\delta^2)/N$, without continuity correction. Zero variance is refused. This does not condition on an observed number of discordant pairs or round $N\psi$ to such a count. Unconditional exact McNemar power remains the initial method. The approximation's original worked-source comparison remains outside the assessed validation scope.

Part VIII

Predict

CHAPTER 66

Fit once and score another table

Scoring applies an already fitted model to new predictor values. It answers what that model predicts for the new rows; it does not estimate new coefficients or establish predictive accuracy on those rows.

Choose **Predict > Stored regression scoring and influence...**. This workflow retains the model's coefficients, predictor definitions, categorical reference levels, uncertainty information and fitting population so that later scores use the same model.

66.1 Fit a model with explicit response meaning

Select the response and at least one predictor. The default is ordinary least squares (OLS), with no predictors initially selected. Supported choices are:

OLS.
An additive linear mean model for a numeric response.

WLS.
A linear mean model with a positive numeric weights column. Weights describe relative precision, not replicated frequency counts.

Binary logistic.
A two-level response, with the event value entered explicitly as it appears in the response. Predictions are probabilities of that event.

Poisson.
Nonnegative integer counts. An optional positive exposure column supplies a log offset; predictions are expected counts at that exposure.

Numeric predictors keep their original units. Categorical predictors use sorted observed levels, with the first as reference and one indicator for every other level. The response cannot also be a predictor. Exposure is not a predictor and is never inferred from an arbitrary column. For Poisson, non-intercept coefficients are log-rate effects and their exponentials are rate ratios.

The desktop reports 95% intervals. OLS lets you choose mean-response confidence or individual prediction intervals; mean response is the default. WLS provides model-based mean uncertainty. Logistic and Poisson intervals are on the probability and expected-count scales, respectively.

The fitting population excludes user-excluded rows and omits missing responses/predictors or missing, non-finite or nonpositive weights/exposures. Review the omission counts. A non-numeric weight or exposure column, rank deficiency, a constant numeric predictor, an absent requested event, failed GLM convergence or logistic separation stops the fit. OLS/WLS need positive residual variance for their reported uncertainty.

The supported setup has at most 32 predictors, 32 levels per categorical predictor and 256 encoded features. Interactions and automatic transformations are not supplied by this stored-scoring model; use a method that explicitly represents the intended relationship.

66.2 A complete training example

Choose **File > Import data...** and select `sample-data/manual/scoring-train.csv`. Use delimiter comma, encoding **utf-8**, decimal point, header row 1 and first data row 2; click **Import**. Make the imported table active, keep all five rows included, and check its values:

x	y
0	1
1	2
2	2
3	4
4	5

Choose **Predict > Stored regression scoring and influence...**. In its setup, set **Model** to **ols**, **Response** to **y**, select **x** as the only predictor, and set **Interval** to **mean**. The confidence level is fixed at 95%. Click **Fit and save model**. Inspect the fitted coefficients and the score row for $x = 2$ in the results. The fitted equation is

$$\hat{y} = 0.8 + 1.0x.$$

Residual sum of squares is 0.8, with three residual degrees of freedom, so $s^2 = 0.8/3 = 0.266667$.

At $x = 2$, the predicted mean is 2.8. Its 95% mean-response interval is approximately $[2.06505, 3.53495]$. Selecting the individual interval instead gives $[0.99974, 4.60026]$. The wider interval allows for the scatter of a new observation as well as uncertainty in the estimated mean.

This small synthetic dataset makes the calculations visible. It does not supply a reliable assessment of prediction performance for a real process.

66.3 Why the two OLS intervals differ

For design row x_0 including the intercept and coefficient covariance V, the mean-response standard error is

$$\mathrm{SE}_{\mathrm{mean}} = \sqrt{x_0^{\mathsf{T}} V x_0}.$$

The individual prediction standard error adds residual variation:

$$\mathrm{SE}_{\mathrm{individual}} = \sqrt{x_0^{\mathsf{T}} V x_0 + s^2}.$$

Both use the Student-t critical value with residual degrees of freedom $n - p$, where p includes the intercept. An individual interval refers to a future observation under the same residual model.

WLS mean uncertainty depends on the declared weights; an individual WLS interval is unavailable here. Logistic and Poisson mean limits use normal Wald intervals on the link scale and then transform them to response units, including the Poisson exposure. These are not individual-outcome prediction intervals.

66.4 Inspect influence without automatically deleting observations

OLS diagnostics identify fitted rows through their original row identities. Let e_i be a residual, h_i leverage, $d = n - p$ and $s^2 = \mathrm{SSE}/d$. The internal standardized residual and Cook's distance are

$$r_i = \frac{e_i}{s\sqrt{1 - h_i}}, \qquad D_i = \frac{e_i^2 h_i}{p\, s^2 (1 - h_i)^2}.$$

They require positive residual variance and $h_i < 1$. Leverage measures the design position; a large response residual is a different property.

The externally studentized residual uses the variance after deleting that row:

$$s^2_{(i)} = \frac{\mathrm{SSE} - e_i^2/(1-h_i)}{d-1}, \qquad t_i = \frac{e_i}{s_{(i)}\sqrt{1-h_i}}.$$

It needs $d > 1$ and positive deleted variance. DFFITS is $t_i\sqrt{h_i/(1-h_i)}$; DFBETAS describe standardized changes in individual coefficients.

A diagnostic can be unavailable even when another is meaningful. Zero deleted variance, for example, does not erase valid full-fit diagnostics. Numerically unresolved denominators are identified rather than replaced with a small arbitrary value. A row is described as determining a coefficient alone only when deleting it reduces design rank.

With only one residual degree of freedom, defined internal standardized residuals all have magnitude one and Cook's distance depends only on the design. The result discloses this limitation; these values are not calibrated outlier tests. Earlier saved results may label the internal standardized quantity as studentized residual.

Investigate influential observations through collection records, model assumptions and sensitivity analysis. Influence alone does not justify deletion. WLS, logistic and Poisson results do not receive these OLS influence interpretations.

66.5 Score a new table using the saved fit

The following steps include the training prerequisite so that this section can be followed on its own.

1. Choose **File > Import data...** and select `sample-data/manual/scoring-train.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order.
2. Choose **Predict > Stored regression scoring and influence...**. Set **Model** to **ols**, **Response** to **y**, select only predictor **x**, and set **Interval** to **mean**. Click **Fit and save model**. The training predictions are 0.8, 1.8, 2.8, 3.8 and 4.8. In the result window click **Add report**, then **Close**.
3. Choose **File > Import data...** and select `sample-data/manual/scoring-target.csv`. Use comma, UTF-8, decimal period, header row 1, first data row 2, blank missing codes and empty type overrides. Click **Import**. Keep this target tab active; its **x** column contains 2, 5 and a blank, and it has no response column.
4. Choose **Predict > Stored regression scoring and influence...** again. In the saved-model list, select the OLS report just created and click **Restore saved model**. Click **Score current table** and inspect **Predictions and response intervals** in the result window.

Saving the project after **Add report** also retains the model for a later session. On reopening that project, activate the target table, open the same menu, select that saved report, click **Restore saved model**, then **Score current table**. The target needs the fitted predictors and retained Poisson exposure column, if applicable. It does not need the original response.

The saved OLS fit gives these target results:

x	Prediction	95% mean-response interval
2	2.8	[2.06505, 3.53495]
5	5.8	[4.07638, 7.52362]
blank	unavailable	unavailable

All target rows retain their identities, including the incomplete row. The $x = 5$ prediction is outside the training range 0–4 and depends on extending the linear relationship. A finite prediction and interval do not establish that this extrapolation is adequate.

Missing or non-finite target predictors, missing/nonpositive exposure, and excluded target rows have unavailable scores and disclosed counts. A missing required column, incompatible type or unknown categorical level in an otherwise scoreable row is an error. The model never creates new category coefficients from the target data.

Choose a new column name (default **model_prediction**) and **Add score column** to attach the row-aligned predictions to the table. The operation retains the frozen fit and supports undo, redo and save/reopen replay. Fitting and target populations remain distinct.

66.6 Retain the model as well as its displayed rows

The results display the first 200 score/diagnostic rows and at most 400 influence marks. Saved reports retain the displayed table, model and provenance needed to reproduce scoring; the live influence plot is a session view. A bounded display does not mean that only those rows were fitted.

Source and score populations support up to 1000000 rows, with a 20000000 selected-cell input budget. Large models or influence outputs can exceed output/report limits even within those counts; an explicit error asks you to reduce the calculation. No partial model is published.

A run can be cancelled and has a three-minute calculation limit. A late response from a changed source or superseded request cannot replace a newer result. A saved report remains the recorded fit until you explicitly fit another model.

CHAPTER 67

Profiles and multiresponse desirability

A prediction profile asks how a retained model's mean prediction changes when one predictor varies and the others are held fixed. Desirability adds explicit response goals and searches for a setting that balances them. Both calculations depend on the fitted models and the operating region.

Choose **Predict > Stored-model profiler and desirability...**. The menu requires a current or saved compatible OLS, WLS, binary logistic or Poisson fit from the stored-regression workflow. The profiler uses its saved coefficients and encoding, without refitting or modifying a table. Compatible fits can also be recovered from a saved profiler report.

67.1 Select compatible models and an operating region

Choose one to eight models with distinct response names. A shared predictor must have the same numeric/categorical role in every applicable model. There can be at most 16 predictor roles overall. Changing model selection resets the entered settings, so establish that selection first.

Numeric bounds initially use the intersection of available original fit ranges, with the midpoint held fixed. Enter finite ordered bounds and a held value inside them. For older saved fits without ranges, the initial 0–1 range and held value 0.5 are editable assumptions, not recovered training information. An empty shared range requires compatible models or a reconsidered analysis.

For categorical predictors, choose permitted levels from the common retained levels and a permitted held level. At least one must remain, and the product of the level counts cannot exceed 128 combinations. The fit's original category references remain in effect.

A Poisson model with exposure requires a positive exposure setting. Its default is 1 and maximum is 10^9. This exposure is held fixed for that model's profiles; changing it changes expected counts, not the fitted rate relationship.

67.2 Give the response goals their own units

Each response has a goal, lower/upper thresholds and positive relative weight. Defaults are maximize, thresholds 0 and 1, and weight 1. Replace these generic thresholds with meaningful response values. For a target goal, a blank target uses the midpoint; an explicit target must lie strictly inside the two bounds.

OLS/WLS goals use response units, logistic goals use the specified event probability, and Poisson goals use expected counts at the entered exposure. A coefficient's log-odds or log-rate scale is not the goal scale.

Maximize desirability is zero at/below the lower threshold, one at/above the upper threshold and linear between. Minimize reverses the direction. Target desirability is zero at/outside the two bounds and one at the target, with linear segments to either side. The joint value is

$$D = \left(\prod_i d_i^{w_i} \right)^{1/\sum_i w_i}.$$

A zero response desirability makes the joint value zero. The weights state priorities and do not express precision, probability or confidence.

67.3 Follow the saved five-row model

Choose **File > Import data...** and select `sample-data/manual/scoring-train.csv` in the supplied sample-data folder. In the preview, use a comma delimiter, **utf-8** encoding, decimal point, **Header row** 1 and **First data row** 2. Leave **Missing codes** blank and **Type overrides** as `{}`. Click **Import**, make the new table active, and keep all its rows included and in their imported order. Choose **Predict > Stored regression scoring and influence...**. Set **Model** to **ols**, **Response** to **y**, select only predictor **x**, and set **Interval** to **mean**. Click **Fit and save model**. The fitted relationship is $\hat{y} = 0.8 + x$, on $x = 0, \ldots, 4$. In the results click **Add report**, then **Close**. Choose **Predict > Stored-model profiler and desirability...**.

Select that model, numeric x range 0–4 and held x=2. Set a maximize goal for y with lower threshold 1, upper threshold 5 and weight 1. Retain confidence 0.95, seed 42, 5000 numeric search samples and 1024 local evaluations. Click **Run**; inspect the held-setting profile and the best-setting summary.

At the held value $x = 2$, the predicted mean is 2.8 and desirability is

$$d = \frac{2.8 - 1}{5 - 1} = 0.45.$$

The best setting found is x=4, with predicted mean 4.8 and desirability 0.95. Its pointwise 95% mean-response interval is approximately $[3.52702, 6.07298]$. The goal threshold 5 is a preference boundary; it is neither a control limit nor an interval endpoint.

The model is linear in this example, so the upper x boundary maximizes its fitted mean. With several responses or nonlinear inverse links, the best joint setting can lie elsewhere. A high desirability value is still a model-based recommendation requiring practical confirmation.

67.4 Read a profile as a held-setting comparison

Numeric profiles evaluate 101 equally spaced values. Categorical profiles evaluate each permitted level. All other predictors remain at their entered held values, so the profile is not a path through a series of separately optimized settings.

For OLS/WLS, the predicted mean is $x^{\mathsf{T}}\hat{\beta}$. Logistic uses the inverse-logit and Poisson the exponential inverse link, with the entered exposure multiplier where present. The profile table gives exact categorical names even when the plot uses ordered positions.

OLS/WLS mean intervals use the retained coefficient covariance and residual t degrees of freedom. They remain mean intervals even when the original OLS scoring setup requested individual prediction intervals. Logistic/Poisson limits use approximate normal Wald inference on the link scale, transformed to response units. Unavailable or overflowing bounds are disclosed.

All these intervals are pointwise and conditional on the retained fit. They are not simultaneous bands, causal effects or intervals adjusted for choosing the best setting after a search.

67.5 Understand the search and its limits

The search enumerates permitted categorical combinations. Each combination receives seeded uniform numeric candidates, the held setting and the all-lower/all-upper corners, with optional bounded Nelder–Mead refinement. The defaults are seed 42, 5000 numeric candidates and 1024

local evaluations. Numeric candidates can range from 128 to 10000 and local evaluations from 0 to 4096; zero disables local refinement. Budgets are divided across categorical combinations.

A finite improvement can be retained when local refinement stops without convergence, with a warning. The result is the *best setting found*, not a guarantee of the global numeric optimum. Evaluation counts include search/refinement work, not the separate profile points. Ties retain the first best setting.

If every evaluated setting has zero joint desirability, the result says there is no acceptable candidate. A displayed setting then permits inspection of the failed goals and is not a recommendation.

Ranges extending beyond retained fitting ranges are warned. With older models lacking such evidence, extrapolation cannot be certified. Even a setting inside every marginal range can lie in a sparsely observed joint combination. Statistical support, physical feasibility and causal interpretation require separate judgment.

67.6 Save the scenarios with their retained models

The result displays settings, predictions and intervals, individual and joint desirabilities, categorical combinations and evaluation counts. Choose a predictor/response profile to inspect its mean and desirability plots and values; its table can show all 101 numeric points.

Add to results adds the summary, mean-response and profile cards together. They retain the frozen models, operating settings, goals, seed, numerical results and warnings. Their source is a retained-model scenario, not the current table's population. Save/reopen preserves those scenarios without recalculation; replay uses the retained fits. HTML, PDF and LaTeX exports include the labeled tables and method disclosures.

Cancel calculation, closing the setup, a new run or a dataset/project reset cancels the active request. Calculation failures leave the data and saved models unchanged. Displayed tables identify truncation and oversized report cards fail visibly rather than dropping model information.

CHAPTER 68

Scoring discriminant and PLS models

A retained discriminant-analysis or partial-least-squares model can apply its fitted relationship to another table. Use the scoring controls for the retained fit in its analysis workflow. The target needs the required predictor columns but does not need the original response.

The scoring calculation reuses the fitted preprocessing as well as the coefficients. For example, centring by the target table's mean would change the model's meaning; HypoNum uses the stored training values. Scoring does not retrain the classifier, select a different PLS component count or estimate new category boundaries from the target.

68.1 Missing predictors preserve the target population

For a model fitted with drop-row handling, an incomplete target row receives an unavailable prediction. This includes missing, NaN or infinite required predictors. Other complete rows receive the same score they would receive if evaluated alone. A table with every row incomplete therefore produces an aligned set of unavailable outcomes, not a different fitted model.

For a model fitted with mean imputation, missing predictors use the retained training means. They are not replaced with means from the scoring table. A required predictor column absent from the table is an error.

Input row identities and their original order are preserved. Review the incomplete-row disclosure before joining predictions to other information or comparing scored and unscored groups.

68.2 Interpret what the saved fit predicts

Discriminant scoring returns a predicted class and available posterior probabilities under the retained classifier. Such a probability depends on its training assumptions, class structure and priors. It does not establish performance in a population whose class frequencies or predictor relationships have changed.

PLS scoring predicts the response with the retained preprocessing, loadings and selected components. Choosing those components on training data is a separate step from evaluating prediction on fresh observations. Keep the model's fitting and evaluation populations identified when reporting a new-table prediction.

For the fitting assumptions and examples, see the multivariate Statistics chapters. Treat new scores as outputs of the saved model; use an appropriate independent evaluation when predictive accuracy is the question.

Part IX

Results and support

CHAPTER 69

Build and share an analysis report

A useful report connects the statistical result to its question, population, settings and interpretation. Retaining only a picture of a chart loses much of that context.

Choose **Results > Open results workspace**. Most analysis windows offer **Add to results**; Time Series uses **Add to Report**, and some retained-model workflows use **Add report**. These actions preserve the numerical result, settings, warnings, tables and plot information. They do not take a screenshot or rerun the calculation.

69.1 Assemble a readable report

Create or rename a report and select it in the list on the left. The report scrolls vertically on the right. Add result cards from the relevant analysis windows, reorder them into an explanatory sequence, and add notes giving the question and interpretation. Remove obsolete cards or reports deliberately. These edits are project changes, so save the project to retain them.

A practical order for a process study is:

1. the measurement question, units, population and collection period;
2. measurement-system evidence and limitations;
3. the baseline control chart and any explained exclusions;
4. capability estimates with their model and uncertainty;
5. the decision criterion, interpretation and next action.

This order lets a reader assess the evidence supporting a conclusion before encountering a qualification label.

For a predictive model, include how training and evaluation data were separated, the actual evaluation results and the population to which a new prediction applies. A fitted coefficient table alone does not establish prediction performance.

69.2 Distinguish a saved snapshot from current data

Cards show their title, data context and a short disclosure, together with **Saved snapshot** or **Source changed**. A source-change warning does not alter the saved result.

For example, after retaining a capability analysis, change a specification or exclude an observation. The existing card remains evidence of the earlier calculation. Run the new analysis explicitly and add its result, then explain which result supports the current decision. A saved snapshot is not a live spreadsheet formula.

Inspect warnings and row counts as part of the result. A displayed table may contain only a bounded subset; respect its truncation disclosure when describing the underlying calculation. The model and complete retained numerical content depend on the particular workflow.

69.3 Export for reading or further typesetting

Export a report to HTML, PDF or editable LaTeX. The export preserves report order, headings, notes, tables, plot aspect ratios and source-change disclosures. Open the exported file and check page breaks, table labels and plot readability before distributing it.

HTML supports convenient reading in a browser. PDF gives a fixed page layout. LaTeX supports further typesetting, with editable PGFPlots chart source; the software requirements for compiling that export are in the separate *API and programming reference*.

An export communicates the retained result. To reproduce the recorded calculation and its prerequisites, use the separate project/target reproduction workflow described in Chapter 70. Keep the saved project when reproducibility matters, rather than relying on a report image alone.

CHAPTER 70

Plug-ins, help and recovery

70.1 Run a trusted local analysis plug-in

The bundled **Plug-ins > Sample: Column summary** demonstrates the workflow. Open a table, choose a numeric column and histogram options, then select **Run analysis**. Results use the familiar bounded scalar, table, histogram or scatter displays.

For an additional local plug-in, install its folder as a direct child of the platform's plug-in directory:

Windows.
`%LOCALAPPDATA%\HypoNum\plugins`

macOS.
`~/Library/ApplicationSupport/HypoNum/plugins`

Linux.
`$XDG_DATA_HOME/tabelz/plugins`, falling back to `~/.local/share/tabelz/plugins`.

Choose **Plug-ins > Reload plug-ins** after installation or a manifest change. Reloading reads the plug-in's identity and description without executing its analysis code.

Select the installed entry and complete its setup form. The plug-in receives the included active table and validated settings. Results can include at most 64 histogram bins and 5000 scatter points; a large or invalid result is rejected visibly. The footer identifies the plug-in, version and abbreviated code checksum so you can distinguish two different installed versions.

> **Important**
>
> A local plug-in is trusted Python code. It runs with the signed-in user's filesystem, process and network permissions. The separate calculation process supports cancellation and interface stability; it does not confine the plug-in's access. Install and run code only from a source you trust.

HypoNum does not install the plug-in's dependencies, embed its code in projects or run it automatically when a project opens. A run can be cancelled or can time out after two minutes. For plug-in development, consult the API and programming reference's *Local analysis plug-in authoring* chapter.

70.2 Find version, validation and licence information

Help > About HypoNum gives the application version and notices. Include that version when discussing a result or reporting a problem. **Help > Validation** provides local validation and diagnostics entry points. The separate statistical validation report states the methods, comparisons and scope assessed; the existence of a menu item does not establish validation for every possible use.

Help > License... shows the anonymous Installation ID and current trial or paid validity. Copy the ID for the licensing process, paste the supplied key and choose **Activate**. The Individual trial lasts

30 days; a paid licence is machine-bound and normally valid for one year. An operating-system reinstall or a move to another computer requires a replacement key. With no valid entitlement, activation is the available entry point.

70.3 Preserve the project before attempting recovery

A `.hnum` project retains starting data, typed metadata, the recorded workflow, results and a generated Python description. Opening validates the package and restores saved results without rerunning analyses or executing that Python. Earlier development-format projects are rejected rather than silently converted.

If a project member is damaged, keep the original unchanged and make a copy before attempting recovery. If the application offers **Restore recovery snapshot**, inspect the recovered tables and any stale operations, then save under a new project name. A recovery snapshot is something to inspect, not an assurance that every later edit has been recovered.

Refreshing an external source is an explicit action. It creates a new starting revision while retained earlier results keep their original input context. Database credentials are not included in the saved project.

70.4 Inspect, export or reproduce the recorded workflow

File > View project Python... displays the read-only generated description of the recorded workflow. **File > Export script and data...** writes a separate folder with the script and data needed for the selected project or target. Selecting one result includes its prerequisites while omitting unrelated tabs. Choose a new destination if an export already exists.

Exported scripts are independent files. Saving the project does not overwrite a script that you edited, and editing that script does not update the desktop workflow. The optional Python/recipe interface is described in the separate *API and programming reference*; it is not required for the statistical workflows in the analysis chapters.

File > Reproduce project or target... runs recorded commands and writes a separate output directory. Inspect its scope and destination before starting. Exports can contain original data and authored edit values. A missing compatible implementation or required plug-in produces an error. Cancellation or failure does not replace the open project or its saved results.

70.5 Work with scaling and keyboard navigation

The desktop uses bounded dialogs and scrolling areas for 100, 150, and 200 percent scaling. Scroll a tall form to reach the remaining controls; its action buttons stay visible. Tab and Shift+Tab move focus; Space, Enter and Escape follow the relevant control's activation or dismissal behavior. Charts provide text descriptions and equivalent summaries where available. Use these summaries when a small plot or visual distinction is difficult to read.

Text/action contrast is designed toward WCAG AA. Native assistive-technology coverage depends on the supported platform and release evidence; these design goals do not constitute a universal accessibility certification.

70.6 Prepare a useful support request

Record the application and operating-system versions, the smallest sequence that reproduces the problem, and the exact error message. If possible, demonstrate it with non-confidential example data.

A local diagnostics bundle can collect relevant logs and a receipt through the command in the separate *API and programming reference*. Creating it does not upload anything. It excludes project/source/cache payloads and redacts common credentials and private paths. Review the contents before sharing them; do not attach source data or generated Python by default.

Part X

Reference

APPENDIX A

Column formula reference

Use column formulas for a repeatable calculation from existing columns. Choose numeric mode for arithmetic, text mode for labels or parsed text, or date mode for calendar calculations. A formula belongs to one target column and recalculates when its inputs change. The supported expressions below are the formula editor's language.

A.1 Common rules

- A formula may contain at most 2,048 characters, 16 terms, and 32 nested expression levels.
- A simple column name may be written bare, such as `height`. Bracket form is safest for spaces or punctuation: `[partnumber]`.
- Inside brackets, a closing bracket in the name is doubled. Column `a]b` is referenced as `[a]]b]`.
- References must name existing compatible columns. The target cannot depend on itself directly or through another formula column.
- Applying a formula replaces every current target value. The preview shows the first eight current and calculated values.
- Formula columns recalculate when an input changes. Renaming a dependency updates the stored relationship. A dependency cannot be removed in a way that would leave an invalid formula.
- **Remove formula** detaches the definition but retains the latest calculated values and makes the column directly editable.

A.2 Numeric formulas

Numeric mode accepts finite numeric literals, numeric columns, parentheses, arithmetic operators, unary signs, comparisons, logical expressions and functions.

Syntax	Name	Behavior
`a+b`	addition	Adds two numeric expressions.
`a-b`	subtraction	Subtracts the right expression.
`a*b`	multiplication	Multiplies two expressions.
`a/b`	division	Divides; non-finite results become missing.
`a^b`	power	Raises the left expression to the right power.
`-a`, `+a`	unary sign	Negates or preserves an expression.
`sqrt(a)`	square root	Negative-domain results become missing.
`abs(a)`	absolute value	Returns the magnitude.

Precedence from lowest to highest is `or`, `and`, `not`, comparisons, addition/subtraction, multiplication/division, unary sign, and power. Comparisons cannot be chained. Power is right-associative. Parentheses override precedence. Thus `-2^2` means $-(2^2)$, while `(-2)^2` means $(-2)^2$.

```
[revenue] - [cost]
([diameter] / 2) ^ 2 * 3.14159265
4 * sqrt([area])
abs([target] - [measured])
```

Inputs are evaluated as numeric values and the calculated column is represented as floating point where required. Infinite and NaN results, including division by zero and invalid roots, are converted to missing values rather than stored as usable numbers.

A.3 Text formulas

Text mode accepts text-compatible columns, JSON-style double-quoted literals, the + concatenation operator, and the following functions. Positions and part numbers are one-based.

Syntax	Behavior
`text1+text2`	Joins text without an automatic separator. Add a quoted separator explicitly.
`split(text,separator,part)`	Splits at a non-empty quoted literal and returns the requested one-based part. A part beyond the available parts is missing.
`slice(text,start,length)`	Returns `length` characters beginning at the one-based `start`. Length may be zero.
`replace(text,find,replacement)`	Replaces every exact occurrence of non-empty quoted `find`. Matching is literal, not a regular expression.
`trim(text)`	Removes surrounding whitespace.
`upper(text)`	Converts text to uppercase.
`lower(text)`	Converts text to lowercase.

Quoted text follows JSON string escaping. For example, `"A\tB"` contains a tab. Escaped quotation marks may also appear inside the literal. Function positions are bounded at 1,000,000; a slice length may be zero, while a split part and slice start must be at least one.

```
[first] + " " + [last]
split([code], "+", 2)
slice([serial], 2, 3)
replace(trim([status]), "old", "new")
upper([country])
```

For `abcdefg+505+ns`, the expression `split([code],"+",2)` produces `505`. Missing text inputs propagate as missing through the corresponding expression.

A.4 Follow a calculation through missing values

Suppose **mass** records grams and **volume** records millilitres. For a new numeric density column, enter

```
[mass] / [volume]
```

A row with mass 12 and volume 4 gives 3 grams per millilitre. A row with zero volume produces a missing result because division by zero is not a finite density. Review these missing outputs before using the column in an analysis; replacing them with zero would assign a different physical meaning.

For text **code** equal to `AB-025-Q`, the expression

```
split([code], "-", 2)
```

returns the text `025`. It keeps the leading zero. A request for part 4 is missing because that component does not exist. A formula that splits on a literal separator does not infer a data format from other rows.

A.5 Term counting and validation

Column references and literals count as terms, including literal function arguments. Operators and function names do not add terms by themselves. The validation line reports the current term count and result type. Errors identify the character position for malformed syntax when possible.

A.6 Conditions and missing values

Expressions can contain numbers, text, dates and true/false values. A function reads a column in its required type: `length([name])` reads text even when the resulting formula column is numeric. True/false results in numeric positions become 1/0. Two bare-column comparisons require compatible types.

New formulas read declared user-missing codes as missing. `is_missing(x)` detects empty cells, nonfinite numbers and user-missing codes; `is_user_missing([column])` detects only the stored declared codes. Most operations propagate missing values. `coalesce(a,b,...)` returns the first present value. `if(condition,a,b)` chooses a branch; `case(c1,v1,c2,v2,...,default)` uses the first true condition. A missing condition reached by either leaves its result missing. Three-valued logic still gives false for false-and-missing and true for true-or-missing; other unresolved logical combinations stay missing.

For stored values $(1, 99, 3, \text{empty}, 2, 5)$ with 99 declared missing, `if([q]>2,1,0)` returns $(0, \text{missing}, 1, \text{missing}, 0, 1)$. `is_user_missing([q])` identifies only the second position. Recode and count-value commands can inspect stored values directly; their rules should not be inferred from formula masking.

A.7 Additional numeric and series functions

Function	Meaning
`floor(x)`, `ceil(x)`	Adjacent lower/upper integer.
`exp(x)`, `ln(x)`, `log10(x)`	Exponential and logarithms; nonpositive log arguments are missing.
`round(x,d)`	Round the shortest round-trip decimal with ties away from zero; integer d from -15 to 15, default 0. Negative d rounds to tens, hundreds and so on.

Function	Meaning
`power(x,y)`, `mod(x,y)`	Power and remainder with the sign of the dividend; a zero divisor is missing.
`min`, `max`, `mean`, `sum`	Across the present arguments within a row; suffix a minimum count, as in `mean.2(a,b,c)`.
`count(a,b,...)`	Number of present arguments; an explicit minimum-count suffix can make an inadequately observed result missing.
`lag(x,k)`, `lead(x,k)`	Value k table rows earlier/later, default 1; missing beyond the ends.
`difference(x,k)`	Current value minus its lag.
`cumsum(x)`	Table-order running sum; a missing value adds nothing and its own output remains missing.

Ordinary arithmetic uses stored binary values; for example, `mod(-7,3)` is −1, and current division rounds the quotient once. New formulas use version 4, retaining version 3 rounding: `round(x,d)` rounds the shortest decimal representation that reads back to the same stored value, with ties away from zero. This does not recover the text originally typed. Thus `round(1.005,2)` gives 1.01, `round(2.675,2)` gives 2.68, and `round(125,-1)` gives 130. Earlier saved version-1 and version-2 formulas retain scaled-binary rounding; for them `round(1.005,2)` remains 1.00 and negative d is unavailable. Earlier saved formulas keep their recorded evaluation rules; editing and applying one adopts the current rules. Recheck user-missing handling when updating an older formula.

A.8 Dates, text and conversion

`date(y,m,d)` constructs a valid calendar date or returns missing. `year`, `month`, `day` and `weekday` extract parts; weekday is Monday 1 through Sunday 7. `date_add(d,n,"unit")` steps whole days, weeks, months or years, clamping an excessive day to month end. `date_diff(a,b,"unit")` counts whole such units from a to b, truncated toward zero. Thus 31 January to 29 February 2020 is one month. `today()` records the application date with the formula; reapply it to refresh that date, rather than expecting a saved analysis to change overnight.

`length(t)` counts characters; `substr(t,s,n)` takes a substring from one-based s, to the end if n is omitted. `concat` joins text. `pad(t,w,"c")` left-pads to width w (at most 1000), using a space by default. `contains` and `starts_with` are exact case-sensitive tests. `regex_match(t,"pattern")` searches anywhere unless anchored; `regex_replace(t,"pattern","replacement")` replaces all matches, with `$1` for the first captured group. Patterns are limited to 256 characters and the supported linear-time regular-expression syntax.

`number(t)` parses a finite number or returns missing. `text(x,"0.00")` writes two decimal places. Current formulas round that fixed-decimal text from the shortest round-trip decimal representation, with ties away from zero, consistently with `round`. Thus `text(1.005,"0.00")` is `"1.01"`; a value rounding to zero has no negative sign. Earlier formula versions retain their recorded text formatting. A format changes the output text, not the underlying measurement;

date-format patterns use percent codes, such as day/month/year. Function constants, units, patterns and formats count toward the same 16-term limit. A text conversion can preserve leading zeroes only when they are present in text or explicitly requested by a formatting operation.

APPENDIX B

Defaults and limits at a glance

Use this appendix as a quick lookup after choosing a method. The analysis chapter explains the population, assumptions and interpretation; a default is a starting setting, not a recommendation for every study.

Many table-backed analyses accept up to 1000000 source rows and 20000000 selected cells. Selected cells are rows multiplied by analysis columns, so a wide selection may reach the cell bound first. Some methods have tighter limits because their cost grows faster: pairwise-distance clustering and dense covariance calculations are examples. A method's own documented limit takes precedence over the general ceiling.

Distinguish the calculation population from display limits. A table showing 200 result rows or a plot showing 400 marks may summarize a larger completed calculation. Read the displayed omission and truncation disclosures.

Feature	Default	Exposed limit or rule
Delimited import	comma, UTF-8, period decimal	Header 1; first data row 2; row controls up to 100,000
Excel worksheet import	header 1; data row 2	One new tab per checked sheet; bounded by remaining tab capacity
New column type	Continuous numeric	Name must be non-empty and unique
Formula mode	Numeric for a new formula	2,048 characters; 16 terms; depth 32; preview 8 rows
Text positions	one-based	Maximum 1,000,000; slice length may be zero
Cleaning population	Included rows only	Excluded rows never match, supply values or change; see Section 6.1
Duplicate detection	All columns; whole-row comparison; keep first; remove rows	Per-column repeats ignore blanks; all column checks use the original included rows
Missing-value selection	All columns; all chosen columns blank; remove rows	Any-blank selection and filling are separate choices
Missing-value fill	Median	Only selected blank cells; no substitute method for unsupported column types; preserve types with stated rounding

Feature	Default	Exposed limit or rule
KNN missing fill	Up to four donors; compare all numeric columns	Numeric targets only; target omitted from comparisons; no comparable donor leaves a blank
Outlier check	All numeric columns; Median/MAD, factor 3.5; any chosen column; remove rows	Per-column limits; strict comparison; non-finite values not flagged
IQR outlier factor	1.5	Beyond $Q_1 - kIQR$ or $Q_3 + kIQR$
Mean/SD outlier factor	3	Uses sample standard deviation
Absolute outlier factor	5; median statistic	Mean, median, or sample SD
Random split	50 percent; seed 1729	Percent between 0 and 100; seed is a non-negative integer
Categorical split	first eligible column	At most 8 output tabs; missing gets its own tab
Concatenate	Common columns only	At least 2 input tabs
Merge	One-key inner match	Two distinct tabs; first duplicate key used
Plot color and facet	Missing categories shown	Numeric X only; at most 36 categories per role and 36 combinations; 5,000 drawn points shared by all panels; no fit
Correlation	Pearson; scatter plots on; no primary	At least 2 numeric columns; plot at most 10 columns and 400 points per panel
Fit Y by X	Drop rows; OLS; Linear; no interactions	Degree 2 for Polynomial; holdout 0 percent; CV 5; confidence 95 percent
Fit Y by X imputation	Local average uses 4 positions; KNN uses 5 neighbors	Performed before validation splitting
Fit Y by X standard errors	Classical	HC3 and CR1 for OLS only; classical and HC3 use $t(N-K)$, CR1 uses $t(G-1)$ with correction $\frac{G}{G-1} \cdot \frac{N-1}{N-K}$; CR1 needs a cluster column and at least 2 clusters and warns below 30; HC3 stops when $1 - h_i \leq 10^{-10}$
Binary logistic	Threshold 0.5	At most 20 predictors and 20 levels per categorical predictor

Feature	Default	Exposed limit or rule
Panel fixed effects	Entity effects; entity-clustered CR1; 95 percent $t(G-1)$ intervals	Unique entity/time keys; 1–32 numeric predictors; 1,000,000 source rows; 20,000,000 projected cells; 10,000 complete-case entities before singleton removal; singletons and invariant predictors disclosed
Survey descriptives	Mean; 95 percent t intervals	Explicit positive weights, strata and PSU; 2 PSUs per stratum; 1,000,000 source rows; 20,000,000 projected cells; 10,000 PSUs; no domain exclusions or clipped proportion intervals
Grouped summaries/cross-tabs	Numeric summary; no groups/frequency; 95 percent $t(N-1)$ mean interval	Up to 3 groups; nonnegative integer replication frequencies; missing N separate; 1,000,000 source rows; 20,000,000 selected cells; 2,000,000 output cells; first 200 display/report rows; complete CSV export
Bars, heatmap and ECDF	Bars; counts; alphabetical; missing shown	36 categories per axis; one declared percentage denominator for counted bars only; ECDF 5,000 drawn steps with exact cumulative values and selection only when steps identify one row; 1,000,000 source rows; 20,000,000 selected cells; 2,000,000 output cells; first 200 display/report rows; complete CSV export
Distribution diagnostics	Shapiro–Wilk; alpha 0.05	Shapiro 3–5,000 complete cases; source 1,000,000 rows; 400 plot points; two-sided single-outlier Grubbs; fitted Poisson with expected-count pooling at 5 and infinite tail
Z, variance and Poisson-rate inference	Known-sigma Z; raw rows; two-sided; 95 percent confidence	Sigma entered explicitly; normality for variance; nonnegative event counts and positive exposure; exact rate intervals; source-independent entered summaries; 1,000,000 raw source rows

Feature	Default	Exposed limit or rule
Probability distributions	Normal$(0, 1)$; example x; $p = 0.95$; no interval; both endpoints included; 100 rows; column Value; fresh seed	No dataset used; 200 values and 200 probabilities; plot 201 points (maximum 512); random tables at most 20 columns and 1,000,000 cells; binomial trials and Poisson mean at most 10^9
MSA study worksheets	Crossed; 10 parts; 3 operators; 2 trials; seed 2026; no blinding; no standard values	No dataset used; 10,000 parts; 1,000 operators; 100 trials per cell; at most 1,000,000 rows; nested designs need at least 2 operators and 2 parts per operator; seed at most $2^{63} - 1$; algorithm numpy-pcg64-operator-block-shuffle-v1
Friedman and runs tests	Friedman; asymptotic; two-sided; numeric runs use median and source row order	Complete block design; average ranks and tie correction; numeric reference ties omitted and counted; 1,000,000 source rows; 10,000 blocks; exact runs at most 200 retained observations
Dunnett/Games–Howell	Dunnett; explicit control; alpha 0.05; seed 12345	Dunnett compares treatments with the control; Games–Howell compares all pairs; simultaneous familywise intervals; 1,000,000 source rows; 32 eligible groups
Equivalence	Independent Welch TOST; alpha 0.05; bounds $[-1, 1]$	One-sample and paired t TOST available; paired log-ratio requires positive inputs and ratio bounds; matching interval $1 - 2\alpha$ (90 percent at 0.05); strict interior decision; row-aligned pairs; no crossover
MANOVA	Type III; sum coding; individual 95 percent intervals	2–8 numeric responses; joint complete cases; Pillai and Wilks; no weights; 1,000,000 source rows and 20,000,000 projected/design cells
Extended mixed models	Random intercept/slope; REML; 95 percent normal intervals	One correlated intercept/slope group or 2–3 declared independent components; 10,000 slope rows or 1,000 component rows; optimizer maximum 200; ML-only AIC

Feature	Default	Exposed limit or rule
Classification tree	Gini; no class weighting; test 25 percent; CV 5; seed 0; maximum depth 8; minimum leaf 1	Binary positive class explicit; 2–20 classes; 64 source predictors; 32 training levels per categorical predictor; 128 encoded features; 1,000,000 source rows; 20,000,000 source predictor cells
Distribution fit	All 8 families selected	Positive-only families skipped for nonpositive observations; goodness-of-fit uses 9,999 seeded refits
Hypothesis test	One-sample t; target 0; alpha 0.05; two-sided	Welch is the two-sample default; Brown-Forsythe is the equal-variance default
Power planning	Required N; effect 0.5; alpha 0.05; power 0.80	Sample N 30, 3 ANOVA groups, allocation ratio 1
Precision/tolerance sample size	Mean absolute half-width 1; assumed SD 1; confidence 0.95	Student-t mean and Wilson proportion; assumed mean 1, proportion 0.5; tolerance content 0.90 and two sides; integer bounds 2–100,000, maximum allowed 1,000,000; conventions in Chapter 64
Reliability demonstration and estimation plans	Binomial demonstration; reliability 0.90; mission time 1000; confidence 0.95; zero allowed failures	Exact binomial and exponential chi-square rules; declared Weibull shape only (never estimated); failure-terminated is exponential-only; precision simulation seed 42, 100–5000 replicates, relative margin 0.10, coverage 0.90, censor factor 5, percentile 0.10, bounds 1–2000; conventions in Section 57.4
Extended power	Raw mean/t; required sample size; one sample; two-sided; mean difference 1 and SD 1; alpha 0.05 and power 0.80	$n = 30$, allocation 1; Desktop search 2–1,000; effect magnitude 0.001–10; 41 curve points (maximum 201); other family inputs and limits in Chapter 65
Stored regression scoring	OLS; first response; no predictors preselected; mean response interval; fixed GUI confidence 0.95	Score column `model_prediction`; 32 predictors, 32 levels per categorical predictor, 256 encoded features and 1,000,000 rows; tables show 200 rows and plots 400 marks; artifact and model conditions in Chapter 66

Feature	Default	Exposed limit or rule
Stored-model profiler	Goals maximize on $[0, 1]$, weight 1; confidence 0.95; seed 42; 5,000 samples; 1,024 local evaluations	1–8 frozen models, 1–16 shared predictors, 128 categorical combinations; samples 128–10,000 and local evaluations 0–4,096; explicit bounds and permitted levels; pointwise mean-response intervals; Chapter 67
Deming comparison	First two numeric columns; Y-error/X-error variance ratio 1; identity line off; jackknife t intervals	Confidence 0.95; jackknife needs at least four pairs; optional 200–5,000 paired percentile draws (1,000 when chosen), seed 0; 1,000,000 source rows, 200 displayed rows and 400 marks; conventions and limits in Section 12.11
Correspondence analysis	Categorical input; roles selected explicitly; two displayed dimensions	Integer replication counts or explicit count table; 200 categories per axis, 40,000 dense cells; positive rank can reduce displayed dimensions; normalization and limits in Section 15.6
Short-run Z–MR	Explicit product centers and positive sigmas; no stage; cross-product MR included	Source order; gaps and stage transitions break MR; fixed Z limits ± 3; known-normal MR limits; 1,000,000 source rows, 200 displayed rows and 400 chart marks; Section 48.9
Apply spec	No transpose; ignore whitespace and case	Preview at most 10 by 10; at least one of LSL or USL
Continuous capability	Normal, individuals, 95 percent confidence, minimum Ppk lower bound 1.33	Within sigma uses the selected production order and subgroup method; overall sigma uses sample SD; stability and MSA required by default; fitted-family passing bounds are conditional on the model; Section 56.8
Fitted capability	Explicit family; confidence 0.95; minimum lower-bound criterion 1.33	Positive data for Lognormal, Weibull and Gamma; Logistic unrestricted; 399 refits from at least 20 usable rows; fitted-data p-values descriptive; Section 56.8

Feature	Default	Exposed limit or rule
Within/between batch capability	Confidence 0.95; minimum Cpk 1.33	One or two specification limits; at least five complete observations and two retained batches with two observations each; conservative Gaussian simultaneous bounds; Section 56.9
Censored life families	Event `1`; censored `0`; confidence 0.95	Exponential, Weibull, lognormal and loglogistic on a common population; three positive finite times and two failures; Section 57.1
Mixed censored life	Confidence 0.95	Exact/right/left/interval bounds; three records and two failure-indicative records; endpoint semantics and validation in Section 57.2
Physical accelerated life	Arrhenius; Celsius; use stress 25; reference stress 120; confidence 0.95	Five observations, three failures and informative stress; positive absolute temperature; family-specific predictions and extrapolation disclosure; Section 57.3
Attribute capability	Binomial defectives; exposure on for Poisson	Fixed 95 percent rate interval
SPC	Variables; Individuals	Constant subgroup size 5 with range 2–25
Attribute SPC	p chart; constant-denominator mode, value blank	np requires constant whole-unit sample; p and Laney p′ require whole inspected-unit denominators; u-type exposure may be fractional; p/u allow varying denominators; Laney p′/u′ and rare-event g/t charts available
CUSUM	$k = 0.5, h = 5$	Target and sigma automatic unless supplied
EWMA	$\lambda = 0.2, L = 3$	$0 < \lambda \leq 1; L > 0$
Finite-window moving average	Width 3; moving-range sigma; expanding startup	Width at least 1; frozen Phase-I or historical parameters; stage boundaries reset windows; the desktop uses expanding startup; Section 48.10
SPC rules	Rules 1–4 on; Rules 5–8 off	Rule 1 cannot be disabled

Feature	Default	Exposed limit or rule
Subgroup Hotelling T-squared	$\alpha = 0.0027$; subgroup size pre-filled with the most common subgroup size	2–64 variables; fixed n from 2 to 100; $m(n-1) \geq p$; generalized-variance chart needs $n > p$ and uses three-sigma limits; 1,000,000 rows; 2,000 displayed subgroups; 200 report rows
Multivariate EWMA	$\lambda = 0.2$; no default decision limit h; uncalibrated trial limit	2–64 variables; $0 < \lambda \leq 1$; $h > 0$; $m > p$; a user-calibrated limit needs a source note; 1,000,000 rows; 2,000 displayed points; 200 report rows; chapter 51
Crossed Gage R&R	Study variation 6; initial AIAG MSA 4th ed. interaction alpha 0.25 (HypoNum extended: 0.05); confidence 95 percent	Balanced ANOVA uses component-specific F/chi-square intervals; unbalanced REML uses actual-design profiles with a boundary-bootstrap fallback; Repeatability bounds use pooled within-cell SSE and $N - c$ degrees of freedom; aggregate bound sums are labeled; at least 2 trials per observed cell
Expanded gage R&R	Classic crossed (delegates to the existing engine); fully crossed; site nested in part; operator nested in site; multiplier 6; alpha 0.05; confidence 0.95	REML components with explicit indicator design matrices; declared interactions require at least 2 trials per cell and are never pooled silently; repeatability interval uses the pooled chi-square rule, other REML intervals are marked unsupported; 1,000,000 rows; 128-indicator budget
Extended gage	Nested study	Type-1 uses fixed 6-sigma variation and 20 percent of tolerance; stability needs at least 10 observations
Attribute agreement	Nominal exact; trial and standard absent	Fully crossed, balanced; exact-agreement intervals are Wilson 95 percent

Feature	Default	Exposed limit or rule
Attribute gage response	Pass modeled; transition points 5 and 95 percent; confidence 95 percent; no default threshold	Grouped binomial probit maximum likelihood; at least 3 distinct reference values; separated or all-pass/all-fail data rejected; Fieller bounded intervals when supported; other confidence-set shapes are reported as unbounded; no overdispersion rescaling; 1,000,000 rows

These values describe the current desktop controls. Library defaults that the application overrides are intentionally not presented as user choices.

APPENDIX C

Menu command index

The current desktop organizes commands by task. The quick reference gives the exact setup page for every command below; the preceding manual chapters explain the methods and interpretation. Cross-listed commands share one primary explanation. Context-sensitive commands become available when their required table, selection, saved model or design is present.

C.1 File

Import project...
File.

Import data...
File.

Import multiple files...
File.

Import database...
File.

Import from database server...
File.

Search...
File.

Export Excel workbook...
File.

Export Parquet file...
File.

Export data...
File.

Refresh database dataset
File.

Save
File.

Save as...
File.

View project Python...
File.

Export script and data...
File.

Reproduce project or target...
File.

Close project
File.

C.2 Edit

Undo
Edit.

Redo
Edit.

Settings...
Edit.

Rename columns...
Edit.

Show value labels
Edit.

Invert selection
Edit > Select.

Exclude selected rows
Edit > Select.

Include selected rows
Edit > Select.

Delete selected rows
Edit > Select.

Delete selected columns
Edit > Select.

C.3 Data

Column properties...
Data.

Formula column...
Data.

Recode into new column...
Data > Transform.

Automatic recode...
Data > Transform.

Rank cases...
Data > Transform.

Standardize...
Data > Transform.

Count values...
Data > Transform.

Visual binning...
Data > Transform.

Missing values...
Data.

Outlier values...
Data.

Remove duplicates...
Data.

Weight cases...
Data.

Split file...
Data.

Define multiple response sets...
Data.

Ascending
Data > Sort.

Descending
Data > Sort.

Stack measurements
Data.

Pivot measurements
Data.

Split tab...
Data.

Concatenate tabs...
Data.

Merge tabs...
Data.

C.4 Explore

Plot...
Explore.

Multi-Plot...
Explore.

Remove quick plots...
Explore > Quick plot.

Variation charts...
Explore.

Pareto chart...
Explore.

Bars, pie, heatmap and ECDF...
Explore.

Matrix plot...
Explore > Exploration graphs.

Stem-and-leaf display...
Explore > Exploration graphs.

Probability plot...
Explore > Exploration graphs.

Contour and surface plot...
Explore > Exploration graphs.

3D scatter plot...
Explore > Exploration graphs.

Multimodality...
Explore.

Explain multimodality...
Explore.

C.5 Statistics

Guided comparison of groups...
Statistics > Guided study.

Guided association or prediction...
Statistics > Guided study.

Guided missing-data analysis...
Statistics > Guided study.

Frequencies...
Statistics > Descriptives & tables.

Explore...
Statistics > Descriptives & tables.

Descriptive summary...
Statistics > Descriptives & tables.

Codebook...
Statistics > Descriptives & tables.

Case summaries...
Statistics > Descriptives & tables.

Grouped summaries and cross-tabs...
Statistics > Descriptives & tables.

Crosstab statistics...
Statistics > Descriptives & tables.

Custom tables...
Statistics > Descriptives & tables.

Multiple response frequencies...
Statistics > Descriptives & tables.

Multiple response crosstab...
Statistics > Descriptives & tables.

Distribution diagnostics...
Statistics > Descriptives & tables.

Distribution fit...
Statistics > Descriptives & tables.

Hypothesis tests...
Statistics > Compare means & ANOVA.

Z, variance and Poisson-rate tests...
Statistics > Compare means & ANOVA.

Equivalence tests...
Statistics > Compare means & ANOVA.

One-way ANOVA with post hoc and contrasts...
Statistics > Compare means & ANOVA.

Repeated-measures ANOVA...
Statistics > Compare means & ANOVA.

Linear effects (ANOVA / ANCOVA)...
Statistics > Compare means & ANOVA.

General linear model...
Statistics > Compare means & ANOVA.

Dunnett and Games-Howell comparisons...
Statistics > Compare means & ANOVA.

Analysis of means (ANOM)...
Statistics > Compare means & ANOVA.

Friedman and runs tests...
Statistics > Compare means & ANOVA.

Permutation tests...
Statistics > Compare means & ANOVA.

Bootstrap intervals...
Statistics > Compare means & ANOVA.

Bootstrap an estimate...
Statistics > Compare means & ANOVA.

Correlation...
Statistics > Correlation & regression.

Kendall's tau-b...
Statistics > Correlation & regression.

Partial correlation...
Statistics > Correlation & regression.

Regression...
Statistics > Correlation & regression.

Linear regression (blocks and diagnostics)...
Statistics > Correlation & regression.

Curve estimation...
Statistics > Correlation & regression.

Nonlinear regression...
Statistics > Correlation & regression.

Deming measurement-method comparison...
Statistics > Correlation & regression.

Weight estimation...
Statistics > Correlation & regression.

Two-stage least squares...
Statistics > Correlation & regression.

Quantile regression...
Statistics > Correlation & regression.

Model selection...
Statistics > Correlation & regression.

Validated regression...
Statistics > Correlation & regression.

Partial least squares...
Statistics > Correlation & regression.

Generalized linear models...
Statistics > Generalized & mixed models.

Binary logistic...
Statistics > Generalized & mixed models.

Penalized logistic (elastic net)...
Statistics > Generalized & mixed models.

Ordinal logistic...
Statistics > Generalized & mixed models.

Nominal logistic...
Statistics > Generalized & mixed models.

Poisson (count)...
Statistics > Generalized & mixed models.

Negative binomial (count)...
Statistics > Generalized & mixed models.

Binomial and Poisson mixed models...
Statistics > Generalized & mixed models.

Population-average GEE (clustered)...
Statistics > Generalized & mixed models.

Mixed models (random intercept)...
Statistics > Generalized & mixed models.

Mixed models (slopes / nested / crossed)...
Statistics > Generalized & mixed models.

Repeated measures (residual covariance)...
Statistics > Generalized & mixed models.

Panel fixed-effects regression...
Statistics > Generalized & mixed models.

General...
Statistics > Generalized & mixed models > Loglinear.

Logit...
Statistics > Generalized & mixed models > Loglinear.

Model selection...
Statistics > Generalized & mixed models > Loglinear.

Principal component analysis...
Statistics > Multivariate.

Exploratory factor analysis...
Statistics > Multivariate.

Multidimensional scaling...
Statistics > Multivariate.

Canonical correlation...
Statistics > Multivariate.

Correspondence analysis...
Statistics > Multivariate.

Multiple correspondence analysis...
Statistics > Multivariate.

Categorical principal components...
Statistics > Multivariate.

Hierarchical clustering...
Statistics > Multivariate.

K-means clustering...
Statistics > Multivariate.

Model-based clustering...
Statistics > Multivariate.

Discriminant analysis...
Statistics > Multivariate.

Nearest neighbours...
Statistics > Multivariate.

Classification tree...
Statistics > Multivariate.

Regression tree...
Statistics > Multivariate.

CHAID tree...
Statistics > Multivariate.

Regression random forest...
Statistics > Multivariate.

Multivariate analysis of variance...
Statistics > Multivariate.

Survival curves (Kaplan-Meier)...
Statistics > Survival.

Life tables...
Statistics > Survival.

Cox proportional hazards...
Statistics > Survival.

Cox regression with time-dependent covariates...
Statistics > Survival.

Missing value analysis...
Statistics > Missing data.

Multiple imputation (pooled regression)...
Statistics > Missing data.

Meta-analysis...
Statistics.

t tests...
Statistics > Bayesian statistics.

One-way ANOVA...
Statistics > Bayesian statistics.

Linear regression...
Statistics > Bayesian statistics.

Pearson correlation...
Statistics > Bayesian statistics.

Binomial proportion...
Statistics > Bayesian statistics.

Poisson rate...
Statistics > Bayesian statistics.

Normal mean (known SD)...
Statistics > Bayesian statistics.

C.6 Survey

Define design...
Survey.

Generate replicate weights...
Survey.

Calibrate weights...
Survey.

Means, totals and proportions...
Survey.

Frequencies and crosstabs...
Survey.

Ratios...
Survey.

Linear regression...
Survey.

Logistic regression...
Survey.

Ordinal regression...
Survey.

Multinomial regression...
Survey.

Cox regression...
Survey.

Count regression...
Survey.

Survey means and proportions...
Survey.

Item consistency (Cronbach's alpha)...
Survey.

Scale reliability (split-half, Guttman, ICC)...
Survey.

Rater agreement...
Survey.

C.7 Time Series

Guided time-series study...
Time Series.

Trend analysis...
Time Series.

ACF & PACF...
Time Series.

Cross-correlation...
Time Series.

Stationarity & Differencing...
Time Series.

Seasonal Decomposition & Adjustment...
Time Series.

Fit ARMA / ARIMA...
Time Series.

Forecast...
Time Series.

Dynamic Regression...
Time Series.

Multivariate Models — VAR / VECM...
Time Series.

Volatility Models — GARCH...
Time Series.

Compare & Validate...
Time Series.

Saved Studies...
Time Series.

Unit Root with a Break — Zivot–Andrews...
Time Series.

C.8 Quality

Guided measurement-system study...
Quality > Guided study.

Guided measurement → stability → capability study...
Quality > Guided study.

Guided reliability study...
Quality > Guided study.

Define characteristic / apply spec...
Quality.

Assess stability / control charts...
Quality > Statistical process control.

Short-run standardized Z-MR...
Quality > Statistical process control.

Between/within I-MR-R/S chart...
Quality > Statistical process control.

Zone chart...
Quality > Statistical process control.

Finite-window moving-average control chart...
Quality > Statistical process control.

Hotelling T-squared (multivariate)...
Quality > Statistical process control.

Hotelling T-squared subgroups and generalized variance...
Quality > Statistical process control.

Multivariate EWMA (MEWMA)...
Quality > Statistical process control.

Assess measurement / Gage R&R...
Quality > Measurement systems.

Gage run chart...
Quality > Measurement systems.

Gage R repeatability screening...
Quality > Measurement systems.

Expanded gage R&R (multi-factor)...
Quality > Measurement systems.

Extended gage studies (MSA)...
Quality > Measurement systems.

Attribute Agreement (MSA)...
Quality > Measurement systems.

Attribute gage response curve (MSA)...
Quality > Measurement systems.

Assess capability...
Quality > Capability.

Capability sixpack...
Quality > Capability.

Compare capability of populations...
Quality > Capability.

Capability of several characteristics...
Quality > Capability.

Within/between batch capability...
Quality > Capability.

Acceptance sampling by variables...
Quality.

Censored life distribution families...
Quality > Reliability engineering.

Parametric life distribution...
Quality > Reliability engineering.

Distribution ID plot...
Quality > Reliability engineering.

Nonparametric life estimate (Kaplan-Meier or Turnbull)...
Quality > Reliability engineering.

Mixed censored life...
Quality > Reliability engineering.

Regression with life data...
Quality > Reliability engineering.

Weibull accelerated failure time...
Quality > Reliability engineering.

Physical accelerated life...
Quality > Reliability engineering.

Repairable systems (recurrent events)...
Quality > Reliability engineering.

Degradation-based lifetime...
Quality > Reliability engineering.

Warranty data pre-processing...
Quality > Reliability engineering.

Warranty prediction...
Quality > Reliability engineering.

Stability study...
Quality > Reliability engineering.

Manufacturing decision tools...
Quality.

C.9 Design

Guided experiment planning...
Design > Guided study.

Guided analysis of a designed experiment...
Design > Guided study.

Design of experiments (DOE)...
Design > Design of experiments (DOE).

Taguchi design...
Design > Design of experiments (DOE).

Analyze Taguchi design...
Design > Design of experiments (DOE).

Robust parameter design...
Design > Design of experiments (DOE).

Sample size for precision and tolerance...
Design > Power & sample size.

Extended power and power curves...
Design > Power & sample size.

Two-level factorial and Plackett-Burman designs...
Design > Power & sample size.

Reliability demonstration and estimation plans...
Design > Power & sample size.

Probability distributions...
Design > Simulation.

Simulate model...
Design > Simulation.

Create MSA study worksheet...
Design.

C.10 Predict

Stored regression scoring and influence...
Predict.

Stored-model profiler and desirability...
Predict.

Overlaid contour of profiled models...
Predict.

C.11 Results

Open results workspace
Results.

New report
Results.

C.12 Plug-ins

Reload plug-ins
Plug-ins.

C.13 Help

Validation
Help.

License...
Help.

About HypoNum
Help.

APPENDIX D

Statistical references and example data

The references below provide background for the statistical ideas and named conventions used in the book. A method's chapter states the settings and conventions used by HypoNum; a reference to a broader method does not imply that every option in the cited source is implemented. The separate statistical validation report identifies retained numerical comparisons and their assessed scope.

D.1 General methods, design and reliability

1. NIST/SEMATECH. *e-Handbook of Statistical Methods*. https://www.itl.nist.gov/div898/handbook/. The exploratory analysis, process improvement, process monitoring and reliability chapters provide worked explanations across these topics. Specific sections are linked where their conventions are used.

2. NIST/SEMATECH. *Maximum likelihood estimation*, Section 8.4.1.2. https://www.itl.nist.gov/div898/handbook/apr/section4/apr412.htm. Censored lifetime likelihoods and estimation.

3. NIST/SEMATECH. *How do you project reliability at use conditions?*, Section 8.4.3. https://www.itl.nist.gov/div898/handbook/apr/section4/apr43.htm. The importance of the declared lifetime and acceleration models.

4. NIST/SEMATECH. *Single sampling plans*, Section 6.2.3.2. https://www.itl.nist.gov/div898/handbook/pmc/section2/pmc232.htm. Binomial operating characteristics and sampling-plan design.

5. Howe, W. G. (1969). Two-Sided Tolerance Limits for Normal Populations—Some Improvements. *Journal of the American Statistical Association*, 64(326), 610–620. Equation 2.13 is the simple HE2 approximation used here; it is distinct from the paper's improved and recommended forms.

6. Derringer, G. C. (1994). A Balancing Act: Optimizing a Product's Properties. *Quality Progress*, 27(6), 51–58. Importance-weighted composite desirability.

7. Stephens, M. A. (1974). EDF Statistics for Goodness of Fit and Some Comparisons. *Journal of the American Statistical Association*, 69(347), 730–737. https://doi.org/10.1080/01621459.1974.10480196. Distinguish the original fitted-normal calibration from the later adjustment named in the distribution-diagnostics chapter.

D.2 Time series

1. Hyndman, R. J., and Athanasopoulos, G. *Forecasting: Principles and Practice*, 3rd edition. https://otexts.com/fpp3/. Forecast benchmarks, evaluation, ARIMA and regression with time-series errors.

2. Box, G. E. P., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. *Time Series Analysis: Forecasting and Control*, 5th edition. Wiley. The ARIMA modelling framework.

3. Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer. VAR, VECM and multivariate dynamic interpretation.

4. Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics*, 31, 307–327. https://doi.org/10.1016/0304-4076(86)90063-1.

5. Zivot, E., and Andrews, D. W. K. (1992). Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis. *Journal of Business & Economic Statistics*, 10, 251–270. The chapter identifies the lag-selection approximation and row convention used in HypoNum separately from the original procedure.

D.3 Process monitoring and capability

1. Montgomery, D. C. (2009). *Introduction to Statistical Quality Control*, 6th edition. Wiley. Examples 9.2–9.3 develop EWMA and moving-average calculations; Section 11.3 treats multivariate monitoring.

2. Nelson, L. S. (1984). The Shewhart Control Chart—Tests for Special Causes. *Journal of Quality Technology*, 16(4), 237–239. The eight tests and their intended use.

3. Lucas, J. M., and Saccucci, M. S. (1990). Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements. *Technometrics*, 32(1), 1–12.

4. Laney, D. B. (2002). Improved Control Charts for Attributes. *Quality Engineering*, 14(4), 531–537. Dispersion-adjusted attribute charts.

5. Benneyan, J. C. (2001). Number-Between g-Type Statistical Quality Control Charts for Monitoring Adverse Events. *Health Care Management Science*, 4, 305–318. Count support and the distinction between rare-event chart conventions matter when comparing calculations.

6. Lowry, C. A., Woodall, W. H., Champ, C. W., and Rigdon, S. E. (1992). A Multivariate Exponentially Weighted Moving Average Control Chart. *Technometrics*, 34(1), 46–53. Equations 2.3–2.5 define the finite-time covariance used in MEWMA.

7. Alt, F. B., and Grimshaw, S. D. (2013). Multivariate Quality Control. In S. I. Gass and M. C. Fu (eds.), *Encyclopedia of Operations Research and Management Science*, 1014–1017. Subgroup monitoring and generalized-variance moments.

8. Bissell, A. F. (1990). How Reliable is Your Capability Index? *Applied Statistics*, 39(3), 331–340. Approximate normal-index uncertainty and estimator-specific effective degrees of freedom.

9. Slifker, J. F., and Shapiro, S. S. (1980). The Johnson System: Selection and Parameter Estimation. *Technometrics*, 22(2), 239–246. The family definitions and percentile-estimation treatment are useful background; HypoNum's explicit family selection and likelihood fit must be distinguished from that procedure.

D.4 Measurement systems and method comparison

1. Automotive Industry Action Group (2010). *Measurement Systems Analysis Reference Manual*, 4th edition. The Gage R&R chapter identifies profile conventions and applicable worksheet/page references.

2. Wheeler, D. J., and Lyday, R. W. (1989). *Evaluating the Measurement Process*, 2nd edition. SPC Press. Background for the Chapter IV, Section E measurement worksheet.

3. ASTM E691. *Standard Practice for Conducting an Interlaboratory Study to Determine the Precision of a Test Method*. Pooled standard deviation and the h and k statistics. A reference to the method does not constitute conformity with every requirement of a particular edition of the standard.

4. Duncan, A. J. (1986). *Quality Control and Industrial Statistics*, 5th edition. Irwin. Background for the d_2^* approximation discussed in the measurement conventions.

5. NCSS. *Deming Regression*, Chapter 303. `https://www.ncss.com/wp-content/themes/ncss/pdf/Procedures/NCSS/Deming_Regression.pdf`. Error-variance direction, point estimates, jackknife intervals and the ten-pair example used in this manual.

6. Franz, V. H. (2007). Ratios: A Short Guide to Confidence Limits and Proper Use. `https://arxiv.org/abs/0710.2024`. Fieller confidence sets and the distinction between bounded, exclusive and unbounded outcomes.

D.5 Additional inference and modelling references

These sources explain the statistical methods. The chapters state the implemented options, defaults and limitations; a source citation alone is not a claim that every option or result in the source is reproduced.

1. Breiman, L. (2001). Random Forests. *Machine Learning* 45, 5–32. `https://doi.org/10.1023/A:1010933404324`.
2. Friedman, J., Hastie, T. and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software* 33(1). `https://doi.org/10.18637/jss.v033.i01`.
3. Koenker, R. and Bassett, G. (1978). Regression Quantiles. *Econometrica* 46(1), 33–50. `https://people.eecs.berkeley.edu/~jordan/sail/readings/koenker-bassett.pdf`.
4. Efron, B. and Tibshirani, R. (1993). *An Introduction to the Bootstrap*. Chapman & Hall. Resampling units, percentile intervals and bias-corrected acceleration.
5. Viechtbauer, W. (2010). Conducting Meta-Analyses in R with the metafor Package. *Journal of Statistical Software* 36(3). `https://doi.org/10.18637/jss.v036.i03`.
6. Little, R. J. A. (1988). A Test of Missing Completely at Random for Multivariate Data with Missing Values. *Journal of the American Statistical Association* 83, 1198–1202. `https://doi.org/10.1080/01621459.1988.10478722`.
7. Shrout, P. E. and Fleiss, J. L. (1979). Intraclass correlations: uses in assessing rater reliability. *Psychological Bulletin* 86, 420–428.
8. McGraw, K. O. and Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods* 1, 30–46; read with its published correction.
9. Krippendorff, K. (2018). *Content Analysis: An Introduction to Its Methodology*, 4th edition. Sage. Agreement based on coincidence and the selected disagreement metric.
10. Conover, W. J. (1999). *Practical Nonparametric Statistics*, 3rd edition. Wiley. Rank, runs and complete-block comparison procedures.
11. Marsaglia, G., Tsang, W. W. and Wang, J. (2003). Evaluating Kolmogorov's Distribution. *Journal of Statistical Software* 8(18). `https://doi.org/10.18637/jss.v008.i18`.

12. Nelsen, R. B. (2006). *An Introduction to Copulas*, 2nd edition. Springer. Gaussian-copula dependence and continuous-margin rank correlation.
13. Kass, G. V. (1980). An exploratory technique for investigating large quantities of categorical data. *Applied Statistics* 29(2), 119–127. https://doi.org/10.2307/2986296. Algorithm step 3 and its convergence condition are on p. 121; the available table example is on pp. 124–125.
14. Bonett, D. G. and Wright, T. A. (2000). Sample size requirements for estimating Pearson, Kendall and Spearman correlations. *Psychometrika* 65, 23–28. https://doi.org/10.1007/BF02294183.
15. Fieller, E. C., Hartley, H. O. and Pearson, E. S. (1957). Tests for rank correlation coefficients. I. *Biometrika* 44(3–4), 470–481. https://doi.org/10.1093/biomet/44.3-4.470.
16. Connor, R. J. (1987). Sample size for testing differences in proportions for the paired-sample design. *Biometrics* 43(1), 207–211. https://doi.org/10.2307/2531961. Original worked-example numerical qualification is not claimed in this edition.

D.6 Reproduce an example with its own data

The supplied `sample-data/manual/` directory contains the complete CSV files named by the worked examples. Import the named file, choose the printed column roles, then enter the stated settings. A different rule set, baseline, confidence level or handling of missing data can change the result.

Most datasets are deliberately small constructed teaching examples. The Deming ten-pair dataset is attributed to the published NCSS example. The accompanying sample-data catalogue records the origin of each file, including constructed datasets reused for consistent examples across chapters. Synthetic data do not establish an operating process's stability, a model's real-world adequacy or a method's general coverage.

Printed numerical values are rounded for reading. The settings and full CSV precision define the example; do not round the input to match the display before calculating. Explanatory figures use those calculations and are labeled as illustrations, so their typography can differ from the application's plots.

For software interfaces and links to upstream library documentation, see the separate *API and programming reference*.

APPENDIX E

Glossary

These definitions distinguish ideas used throughout the book. Read them with the conditions stated for the chosen procedure. Chapter 8 explains how the question, design and uncertainty fit together; appendix C locates the corresponding desktop commands.

Analysis population
The observations to which a calculation applies after user exclusions and the procedure's own eligibility and missing-data rules. Compare the included, omitted and analyzed counts when changing columns or settings.

Association
A relationship between variables. Association alone does not establish that changing one variable would cause the other to change.

Bias
The difference between an estimator's repeated-sampling mean and its target. In measurement studies, the estimated difference from a reference value is also called bias; its uncertainty and the reference's suitability matter.

Bootstrap
A resampling procedure used to approximate a statistic's sampling behavior. The resampling unit must respect the design; individual rows are not always independent units.

Capability
A comparison of a process distribution with stated specification limits. An index does not establish process stability or repair an inappropriate distribution or variation estimate.

Censoring
Partial information about an event time: for example, that failure occurs after the last observation, or within an interval. A censored time is not an observed failure time.

Confidence interval
An interval from a procedure designed to cover a fixed target at its stated long-run rate under the procedure's assumptions. It does not assign that probability to the fixed target after the interval has been observed.

Confounding
An inability to separate effects using the available design or information. In an experiment, aliased effects may produce the same fitted pattern; in an observational study, another variable may mix with the association of interest.

Contrast
A specified comparison of model quantities, often a linear combination of means or coefficients. Its interpretation depends on the coding, scale and population being compared.

Control limits
Limits used to assess variation relative to a specified process reference. They differ from specification limits, which express requirements.

Credible interval
An interval containing a stated posterior probability under a Bayesian model, likelihood and prior. Its meaning differs from repeated-sampling confidence coverage.

Cross-validation
Assessment using observations held out from fitting. Selection, preprocessing and tuning must respect the training folds if the held-out score is to assess the complete prediction procedure.

Effect size
The magnitude of a comparison or relationship on a stated scale. Its units, reference category and standardization determine its practical interpretation.

Equivalence
Agreement within prespecified, practically acceptable bounds. Failure to reject a zero-effect null hypothesis does not establish equivalence.

Estimand
The population quantity an analysis seeks to estimate, including its target population, outcome scale and comparison.

Exclusion
A recorded decision to omit a row from analysis. Exclusion differs from deleting the row and from a method omitting it because required values are missing.

Experimental unit
The unit to which a treatment is independently assigned. Repeated readings on one unit generally do not supply additional independent treatment replicates.

Frequency weight
A count of represented observations. It has a different meaning from a survey weight or a precision weight; follow the selected method's supported weighting rules.

Interaction
A model term allowing an effect to depend on another variable. A main-effect coefficient in such a model must be read at the specified reference levels or values.

Missing value
An absent or unavailable observation. A displayed missing-code label may retain the reason for absence; it is not a measured numeric value. Each procedure states how missing values affect its analysis population.

Multiplicity
The issue created by considering several inferential comparisons together. An adjustment's protection applies to its declared family of comparisons, not automatically to every analysis tried during a project.

Null hypothesis
The assertion against which a test calibrates its evidence. A test's null may concern a difference, distribution, association or other model property.

Paired observations
Measurements linked by the same unit or a justified matching design. Their within-pair relationship is part of the analysis, rather than an accidental ordering of two independent samples.

Parameter
A quantity defining a statistical model or population. An estimate is a data-based value for that quantity; neither is an individual observation.

p-value
Under the null model and its assumptions, the probability of a test statistic at least as extreme as observed according to the specified test. It is not the probability that the null hypothesis is true.

Power
The probability that a specified test rejects its null under a specified alternative and design. A planning calculation is conditional on its inputs, including effect, variability, sample allocation and significance level.

Prediction interval
An interval for a declared future outcome. Its uncertainty includes the variation appropriate to that outcome, rather than only uncertainty in a fitted mean.

Precision weight
A weight reflecting information about an observation's variability, often proportional to inverse variance. Its interpretation and scale follow the model.

Random seed
A value used to initialize a random-number procedure. Reproduction also depends on the data, settings, software implementation and relevant versions.

Repeatability and reproducibility
In measurement-system analysis, repeatability concerns repeated readings under the same specified conditions; reproducibility concerns declared changes such as operators. Project reproduction instead concerns rerunning recorded work from its data and settings.

Residual
The discrepancy between an observation and its fitted value on a declared scale. Standardized, studentized, Pearson and deviance residuals have different definitions; read the plot's label and method description.

Standard deviation and standard error
A standard deviation describes variation in a distribution. A standard error describes the sampling variability of an estimator. They answer different questions.

Statistical significance
A test decision relative to a chosen significance level. It does not by itself establish a practically important effect, model adequacy or causation.

Survey weight
A weight representing a sampling or survey-adjustment design. Valid uncertainty may also require strata, primary sampling units and other design information; weights alone do not describe every survey.

Tolerance interval
An interval intended to cover a specified population fraction with a specified confidence. Both numbers must be reported.

Validation
An assessment against stated requirements and evidence. The companion statistical validation report names its assessed methods, settings and limitations. Availability in a menu and successful execution alone do not establish independent statistical qualification.