Analyse-it for Microsoft Excel

The Complete Guide

© 2023 Analyse-it Software, Ltd. All rights reserved.

Under copyright laws, no part of this manual may be copied or reproduced without the prior written consent of Analyse-it Software, Ltd.

The End User License Agreement (EULA) governs your rights to use the software.

Analyse-it and Analyze-it are registered trademarks of Analyse-it Software, Ltd. Use of the Analyse-it logo for commercial purposes without written consent of Analyse-it Software, Ltd. may constitute trademark infringement.

Microsoft and Excel are trademarks or registered trademarks of Microsoft Corporation, Inc. Other company product names mentioned herein are trademarks of their respective companies.

Every effort has been made to ensure that the information in this manual is correct. However, Analyse-it Software, Ltd. is not responsible for any errors or omissions.

Analyse-it Software, Ltd.
The Tannery
91 Kirkstall Road
LEEDS LS3 1HS
United Kingdom

Contents

Preface: Welcome to Analyse-it 13

Part I: Administrator's Guide 15

Chapter 1: Installation 17
System requirements 17
Installing the software 17
Installing the software for concurrent use 18
Transferring the software to another computer 18
Uninstalling the software 19

Chapter 2: Licensing 21
Activating a license 21
Deactivating a license 23
Acquiring a concurrent license 24
Releasing a concurrent license 24
Finding out who's using a concurrent license 24

Chapter 3: Software updates 27
Maintenance 27
Updating the software 27
Checking maintenance expiry and renewing 28

Chapter 4: Troubleshooting 29
Checking the version installed 29
Enabling the add-in 29
Repairing the installation 30
Contacting technical support 31

Part II: User's Guide 33

Chapter 5: Getting to know Analyse-it 35

Chapter 6: Preparing data for analysis 37
Datasets 37
Variables 38
Measurement scales 38
Missing values 40
Case and frequency form data 40
Setting the measurement scale of a variable 41

Setting the minimum, maximum and units of a variable 41
Ordering categorical data 42
Setting the number format of a variable 42
Labeling cases 43
Assigning labels to categories 43
Assigning colors/symbols to categories 43
Transforming variables 44
Transform functions 44
Analyzing a subset of the data 44

Chapter 7: Working with analyses 47
Creating an analysis 47
Analysis reports 47
Editing an analysis 48
Recalculating an analysis 48
Excluding observations from analysis 48
Creating a new analysis based on an existing analysis 48
Printing an analysis report 49
Printing in color / black & white 49
Printing a single chart on a full page 49
Showing the dataset analyzed 49
Changing the user name on analysis reports 50
Numerical accuracy 50
Presentation of numerical data 50
Showing a statistic to full precision 50

Chapter 8: Feedback and crash reports 51
Sending feedback 51
Submitting crash reports 52

Part III: Statistical Reference Guide 55

Chapter 9: Distribution 57
Continuous distributions 57
Univariate descriptive statistics 57
Calculating univariate descriptive statistics 58
Univariate plot 59
Dot plot 59
Box plot 59
Mean plot 60
Creating a univariate plot 60
Frequency distribution 60
Cumulative distribution function plot 61
Creating a CDF plot 61
Histogram 61
Creating a histogram 62
Normality 62
Normal distribution 62
Normal probability (Q-Q) plot 63
Creating a normal probability plot 63
Normality hypothesis test 63
Tests for normality 64
Testing the normality of a distribution 64
Central limit theorem and the normality assumption 64

Inferences about distribution parameters 65
Point and interval estimation 65
Parameter estimate 66
Estimators for the central location parameter of a distribution 67
Estimators for the dispersion parameter of a distribution 67
Estimating the parameters of a distribution 67
Hypothesis testing 67
Parameter hypothesis test 68
Parameter equivalence hypothesis test 68
Tests for the central location parameter of a distribution 69
Tests for the dispersion parameter of a distribution 69
Testing distribution parameters 69
Testing equivalence of distribution parameters 70
Discrete distributions 71
Frequency distribution 71
Frequency table 71
Frequency plot 71
Whole-to-part plot 72
Creating a frequency plot 72
Creating a whole-to-part plot 72
Inferences about Binomial distribution parameters 73
Binomial distribution parameter estimate 73
Estimators for the parameter of a Binomial distribution 73
Binomial distribution parameter hypothesis test 73
Tests for the parameter of a Binomial distribution 73
Testing Binomial distribution parameters 74
Inferences about Multinomial distribution parameters 74
Multinomial distribution parameters hypothesis test 74
Tests for the parameters of a Multinomial distribution 75
Testing Multinomial distribution parameters 75
Study design 76

Chapter 10: Compare groups 77

Calculating univariate descriptive statistics, by group 77
Side-by-side univariate plots 78
Creating side-by-side univariate plots 78
Equality of means/medians hypothesis test 78
Equivalence of means hypothesis test 79
Tests for means/medians 79
Testing equality of means/medians 80
Testing equivalence of means 81
Difference between means/medians effect size 81
Estimators for the difference in means/medians 81
Estimating the difference between means/medians 82
Multiple comparisons 82
Mean-Mean scatter plot 83
Multiple comparison procedures 84
Comparing multiple means/medians 85
Homogeneity of variance hypothesis test 86
Tests for homogeneity of variance 86
Testing homogeneity of variance 86
Study design 88

Chapter 11: Compare pairs 89

Difference plot 89

Creating a Tukey mean-difference plot 90
Equality of means/medians hypothesis test 90
Equivalence of means hypothesis test 90
Tests for means/medians 91
Testing equality of means/medians 92
Testing equivalence of means 92
Difference between means/medians effect size 93
Estimators for the difference in means/medians 93
Estimating the difference between means/medians 93
Study design 95

Chapter 12: Contingency tables 97
Contingency table 97
Creating a contingency table 98
Creating a contingency table (related data) 98
Grouped frequency plot 99
Effect size 99
Estimators 99
Estimating the odds ratio 100
Estimating the odds ratio (related data) 100
Relative risk 101
Inferences about equality of proportions 101
Equality of proportions hypothesis test 101
Exact and asymptotic p-values 102
Wald, Score, Likelihood ratio 102
Tests for equality of proportions (independent samples) 103
Testing equality of proportions (independent samples) 103
Tests for equality of proportions (related samples) 103
Testing equality of proportions (related samples) 104
Inferences about independence 104
Independence hypothesis test 104
Continuity correction 105
Tests for independence 105
Testing independence 105
Mosaic plot 106
Creating a mosaic plot 106
Study design 107

Chapter 13: Correlation and association 109
Scatter plot 109
Creating a scatter plot or scatter plot matrix 109
Covariance 110
Correlation coefficient 110
Color map 111
Calculating a correlation matrix 111
Calculating a covariance matrix 111
Inferences about association 111
Correlation not Causation 111
Parameter estimate 112
Estimators for correlation and association 112
Estimating the correlation/association 112
Correlation/association hypothesis test 113
Testing correlation/association 113
Study design 115

Chapter 14: Principal component analysis (PCA) 117
Principal components 117
Scree plot 118
Calculating principal components 118
Biplot 118
Monoplot 120
Creating a biplot 121
Creating a correlation monoplot 121

Chapter 15: Factor analysis (FA) 123
EFA 123
Factor rotation 123
Matrix rotations 124
Extracting factors 124

Chapter 16: Item reliability 127
Testing item reliability 127

Chapter 17: Fit model 129
Linear fit 129
Simple regression models 129
Fitting a simple linear regression 129
Advanced models 130
Fitting a multiple linear regression 130
Performing ANOVA 130
Performing 2-way or higher factorial ANOVA 131
Performing ANCOVA 132
Fitting an advanced linear model 132
Scatter plot 133
Summary of fit 133
Parameter estimates 134
Effect of model hypothesis test 135
ANOVA table 135
Predicted against actual Y plot 136
Lack of Fit 136
Effect of terms hypothesis test 137
Effect leverage plot 138
Effect means 138
Plotting main effects and interactions 139
Multiple comparisons 139
Multiple comparison procedures 140
Comparing effect means 141
Residual plot 141
Residuals - normality 142
Residuals - independence 143
Plotting residuals 143
Outlier and influence plot 143
Identifying outliers and other influential points 144
Prediction 144
Making predictions 144
Making inverse predictions 145
Saving variables 145
Logistic / Probit fit 146
Generalized linear models 146

Fitting a simple logistic regression 146
Fitting an advanced logistic model 146
Fitting a simple probit regression 147
Parameter estimates 147
Odds ratio estimates 147
Estimating odds ratios 148
Effect of model hypothesis test 148
Effect of term hypothesis test 149
Study design 150

Chapter 18: Method comparison / Agreement 151

Correlation coefficient 151
Scatter plot 152
Fit Y on X 152
Ordinary linear regression 152
Deming regression 153
Passing-Bablok regression 153
Fitting ordinary linear regression 153
Fitting Deming regression 154
Fitting Passing-Bablok regression 155
Linearity 156
Residual plot 157
Checking the assumptions of the fit 157
Average bias 157
Estimating the bias between methods at a decision level 158
Testing commutability of other materials 159
Difference plot (Bland-Altman plot) 159
Fit differences 160
Plotting a difference plot and estimating the average bias 160
Limits of agreement (LoA) 161
Plotting the Bland-Altman limits of agreement 161
Mountain plot (folded CDF plot) 163
Plotting a mountain plot 163
Partitioning and reducing the measuring interval 164
Agreement measures for binary and semi-quantitative data 164
Chance corrected agreement measures for binary and semi-quantitative data 165
Agreement plot 166
Estimating agreement between two binary or semi-quantitative methods 167
Study design 168
Study design for qualitative methods 171

Chapter 19: Measurement systems analysis (MSA) 175

Precision 175
Conditions of measurement 175
Reproducibility 175
Intermediate precision 175
Repeatability 175
Variance components 175
Estimating the precision of a measurement system 176
Testing precision against a performance claim 177
Precision profile plot 178
Plotting a precision profile 178
Variance function 178
Fitting a variance function 179

Trueness 180
Bias 180
Estimating the bias of a measurement system 180
Testing against an assigned value 181
Testing bias against allowable bias 182
Linearity 182
Estimating the linearity of a measurement procedure (EP06-A) 183
Estimating the linearity of a measurement procedure (EP06-Ed2) 184
Interferences 185
Detection capability 186
Limit of blank (LoB) 186
Limit of detection (LoD) 186
Estimating the detection limit of a measurement system 186
Estimating the detection limit using a precision profile 187
Estimating the detection limit using a probit fit 188
Limit of quantitation (LoQ) 188
Study design 189

Chapter 20: Reference interval 193
Reference limits 193
Normal theory quantile 193
Quantile 194
Bootstrap quantile 194
Harrell-Davis quantile 194
Robust bi-weight quantile 194
Estimating a reference limit or interval 194
Histogram 195
Box plot 196
Normal probability (Q-Q) plot 197
Testing the normality of a reference distribution 197
Transforming the reference distribution 197
Transferring reference limits 198
Study design 200

Chapter 21: Diagnostic performance 201
Measures of diagnostic accuracy 201
Sensitivity / Specificity 201
Likelihood ratios 201
Predictive values 202
Youden J 202
Estimating sensitivity and specificity of a diagnostic test 202
Comparing the sensitivity and specificity two diagnostic tests 203
ROC plot 203
Plotting a single ROC curve 204
Comparing two or more ROC curves 205
Area under the curve (AUC) 205
Testing the area under the curve 205
Difference between the areas under two curves 206
Testing the difference between the areas under two curves 207
Decision thresholds 207
Decision plot 208
Finding the optimal decision threshold 208
Predicting the decision threshold 208
Study design 210

Chapter 22: Survival/Reliability 215

Kaplan-Meier survival function 215
Plotting a Kaplan-Meier survival curve 215
Equality of survival functions test 216
Tests for the equality of survival functions 216
Comparing two or more survival functions 216
Proportional hazards fit 218
Fitting a proportional hazard model 218
Hazard ratio estimates 219
Estimating hazard ratios 219
Baseline survival function 220
Effect of model hypothesis test 220
Effect of term hypothesis test 220

Chapter 23: Control charts 221

Shewhart control charts 222
Shewhart variables control charts 222
Xbar-R chart 223
Xbar-S chart 223
Xbar chart 223
R chart 223
S chart 224
I-MR chart 224
I chart 224
MR chart 225
Creating a Xbar-R / S control chart 225
Creating an I-MR control chart 226
Shewhart attributes control charts 227
NP chart 227
P chart 228
C chart 228
U chart 228
Creating an NP / P control chart 229
Creating a C / U control chart 229
Shewhart control chart rules 230
Applying rules to a Shewhart control chart 231
Formatting a control chart 231
Labeling control chart points 232
Displaying stratification in a control chart 232
Applying phases/stages for each change in a process 232
Time-weighted control charts 233
Uniformly Weighted Moving Average (UWMA) chart 233
Exponentially Weighted Moving Average (EWMA) chart 234
Creating a UWMA control chart 235
Creating an EWMA control chart 235
CUmulative SUM (CUSUM) chart 236
Creating a CUSUM control chart 237
Study design 238

Chapter 24: Process capability 241

Capability ratios (Cp/Pp indices) 241
Z benchmark 242
Nonconforming units 243
Estimating process capability 243

Chapter 25: Pareto analysis 245
Pareto chart 245
Creating a Pareto chart 245
Creating comparative Pareto charts 246
Merging Pareto chart categories 246
Varying Pareto chart bar colors 247
Reordering Pareto chart bars 247
Sorting bars on comparative Pareto charts 248
Study design 249

Appendix A: End-User License Agreement (EULA) 251
Bibliography 253

Welcome to Analyse-it

Analyse-it brings powerful statistical analysis and data visualization into Microsoft Excel.

This guide provides information to help you understand important concepts and instructions to accomplish tasks in Analyse-it. In addition to this book, other resources are available to help you:

- Tutorials let you quickly get started using the main features of Analyse-it with real-world examples. To view tutorials anytime, open Analyse-it and on the Analyse-it ribbon tab, choose Analyse-it > Tutorials.
- On-screen help contains all the information in this book at your fingertips. To open on-screen help, open Analyse-it and on the Analyse-it ribbon tab, choose Analyse-it > Help.
- Context-sensitive help provides brief text descriptions for most on-screen items. To see a help tag, hover the mouse pointer over a user interface item for a few seconds.
- Our support website provides all the latest information *www.analyse-it.com/support*.

Part

I

Administrator's Guide

Chapter 1

Installation

You must install the software on your computer before you can use it.

System requirements

Hardware and software requirements to run Analyse-it.

Microsoft Windows Vista, Windows 7, Windows 8, Windows 10, or later.
Microsoft Excel 2007, Excel 2010, Excel 2013, Excel 2016, Excel 2019, or Excel 365 (desktop installed editions).
Internet connection to enable automatic activation, software update checks, and feedback services. If your computer is behind a firewall, ensure that TCP port 443 (HTTPS) on the firewall is open for communication to *secure.analyse-it.com*.

Installing the software

Install the software for a per-user license or free trial.

- Ensure that your computer meets the system requirements.
- Ensure that you have the necessary administrative rights to install the software.

1. To install the software:
   - From a downloaded .EXE file, double-click the file you downloaded.
   - From a CD-ROM, insert the CD into the CD-ROM drive. The installer should automatically start, but if not, click **Start** > **Run**, and then type: `D:\SETUP.EXE` (where `D:` is the drive letter of your CD-ROM drive).

   The **Setup** window opens.
2. Click **Next**.
   The installer checks your computer meets the system requirements, and if necessary installs Microsoft .NET 4.0.
3. Read the End-User License Agreement (EULA) and then:
   - If you agree, select **I accept the terms of the License Agreement**, and then click **Next**.
   - If you have any questions, contact *support@analyse-it.com*. Click **Cancel** to abort the installation.
4. Optional: If you want to change the installation folder, click **Change** and select the folder.
5. Click **Next**.
6. Click **Install**.
   Wait while the software installs on your computer.
7. Click **Start Analyse-it...**.

Analyse-it automatically starts whenever you start Microsoft Excel.

To use the software, you must activate a license.

Installing the software for concurrent use

Install the software for a concurrent-user license.

Note: Installation for concurrent-user licenses is different from versions earlier than 3.00. The software is now installed and activated on each computer, and your internet connection is used to manage concurrent use. Earlier versions installed the software on a network share and used local license control. This new method reduces network traffic, allows installation onto laptops and computers, not on the same network, and depending on policy allows end-users to install the software without systems administrator assistance.

- Ensure that your computer meets the system requirements.
- Ensure that you have the necessary administrative rights to install the software.
- Ensure that TCP port 443 (*HTTPS*) on the firewall is open for communication to *secure.analyse-it.com* on each computer. Analyse-it uses your internet connection to communicate with our license server to manage the concurrent use of the software.

1. To install the software:
   - From a downloaded .EXE file, double-click the file you downloaded.
   - From a CD-ROM, insert the CD into the CD-ROM drive. The installer should automatically start, but if not, click **Start** > **Run**, and then type: `D:\SETUP.EXE` (where `D:` is the drive letter of your CD-ROM drive).

   The **Setup** window opens.
2. Click **Next**.
   The installer checks your computer meets the system requirements, and if necessary installs Microsoft .NET 4.0.
3. Read the End-User License Agreement (EULA) and then:
   - If you agree, select **I accept the terms of the License Agreement**, and then click **Next**.
   - If you have any questions, contact *support@analyse-it.com*. Click **Cancel** to abort the installation.
4. Optional: If you want to change the installation folder, click **Change** and select the folder.
5. Click **Next**.
6. Click **Install**.
   Wait while the software installs on your computer.
7. Click **Close**.
8. Repeat installation on each computer where you want to use the software. There is no limit to the number of installations as our license server manages the concurrent use of the software.

Analyse-it automatically starts whenever you start Microsoft Excel, but a license is only acquired when you use a command on the Analyse-it ribbon tab.

To use the software, you must activate a license.

Transferring the software to another computer

You can move the software to another computer when you get a new computer.

It is important before you move the software that you deactivate the license on your existing computer otherwise you may be unable to activate the license on your new computer.

1. *Deactivate the license on your existing computer* on page 23.
2. *Install the software on your new computer* on page 17.
3. *Activate the license on your new computer* on page 21.

Uninstalling the software

Uninstall the software to remove it from your computer.

Note: You do not need to uninstall the software to install an update.

1. Start the Windows **Control Panel**.
2. Open the list of installed programs:
   - On Windows XP, double-click **Add/Remove Programs**
   - On Windows Vista or Windows 7, and later, click **Programs**, and then click **Programs and Features.**
3. In the list of installed applications, select Analyse-it for Microsoft Excel, and then click **Uninstall**.
   Wait while the software uninstalls from your computer.

Licensing

Chapter 2

License activation helps you meet the terms of the software license agreement and reduces piracy ensuring continued development of the software.

When you request a 30-day trial or purchase a license to use the software you receive a product key. The product key identifies features of the software you can use, the length of time you can use it for, and how many users can use it.

For all products we offer either:

- **A perpetual license** that does not expire and includes 1-year of software maintenance, so you receive all software updates within the year. You can renew maintenance to continue receiving software updates, but it is not mandatory. If you do not renew maintenance you can continue using the software without further cost, although you only have access up to the last version released within your maintenance period.
- **An annual license** that expires 1-year after the initial activation and includes software maintenance, so you receive all software updates within the year. After expiry, you must renew the license to continue using the software.

There are two types of license seats:

- **Fixed, per-user licenses** allow a fixed number of users to install and use the software on their computer, and optionally on their laptop computer for use when out of the office. The number per-user licenses determine the number of users who can install the software. For example, a 2-user license allows two users to install and use the software. The user can transfer the license to another computer when replacing computers.
- **Floating, concurrent-user licenses** allow an unlimited number of users to install the software on their computer or laptop computer, but limits the maximum number of users of the software at any time. For example, a 5 concurrent-user license allows any number of users to install the software, but only allows up to 5 users to use it at the same time. All computers must have internet access so the software can manage concurrent use.

When you activate a license on your computer the product key and a site identifier are sent to our activation server. The site identifier is a non-unique string generated using the computer's hardware configuration. In the request, we also include diagnostic information such as the version of Microsoft Windows and Microsoft Excel you are using, plus some performance metrics of your computer. We use the information to shape the future development of Analyse-it. We do not collect any personal information except the name, organization and e-mail address you enter during activation.

Activating a license

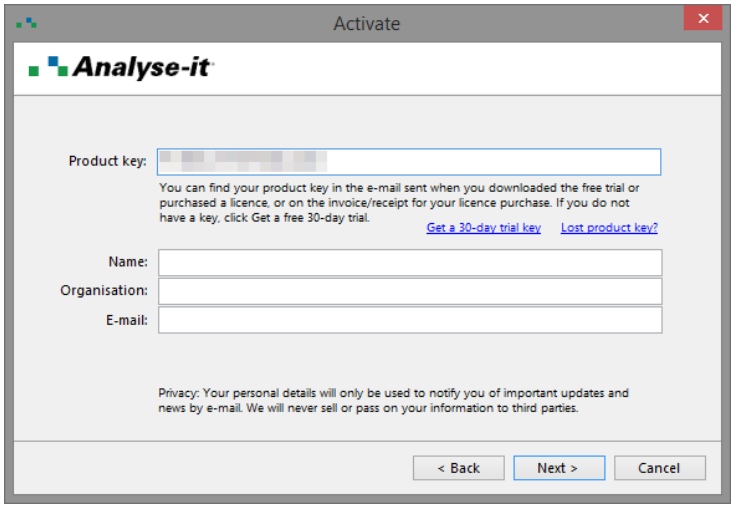
Activate a license to use the software on your computer.

- Ensure that TCP port 443 (*HTTPS*) on the firewall is open for communication to *secure.analyse-it.com* on your computer. Analyse-it uses your internet connection to simplify activation.

1. Start **Microsoft Excel** to start Analyse-it.
2. On the **Analyse-it** ribbon tab, click **Analyse-it**, and then click **Activate**.

The **Activate** window opens.

3. Click **Next**.
   The **Activate** window shows the license details.



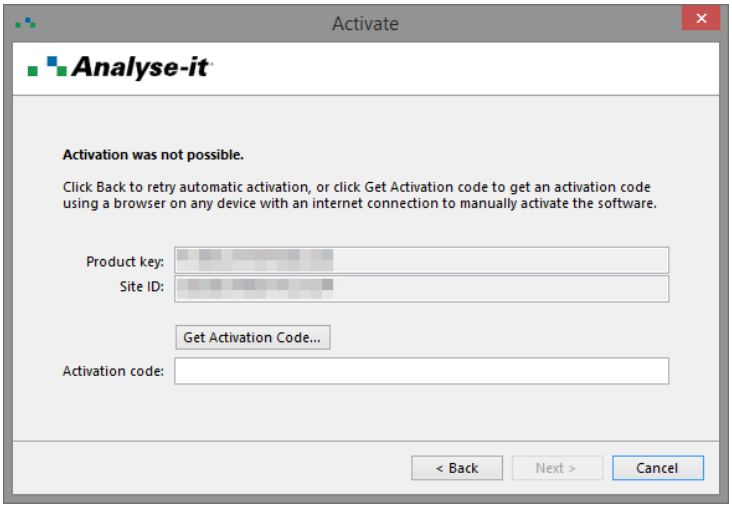
4. Enter the **Product Key** and the **Name**, **Organisation**, and **E-mail** of the registered user of the software.

   The registered user may occasionally receive an e-mail with important information about updates to Analyse-it. They can opt-out via the unsubscribe link in the e-mail they receive.

5. Click **Next**.
   The software uses your internet connection to verify the product key and activates the license on your computer.

6. If activation fails for any reason you must activate manually:
   a) Click **Get Activation Code**.
      Your browser opens the manual activation page at *http://analyse-it.com/activate*. If your computer does not have an internet connection, you can open the page on any computer or device with an internet connection.
   b) On the manual activation page, enter the **Product Key** and **Site ID** from the **Activate** window.

      Note: The Site ID is unique to each computer. Do not contact technical support asking for a Site ID as we cannot provide it. It only shows on the **Activate** window in the software.



c) Enter the **Name**, **Organisation**, and **E-mail** of the registered user of the software.
d) Click **Get Activation Code**.
The activation code shows in your browser window.
e) In the **Activate** window, enter the **Activation code** from your browser window.
f) Click **Next**.

7. Click **Close**.

Deactivating a license

Deactivate the license before you move the software to a different computer.

- Ensure that TCP port 443 (*HTTPS*) on the firewall is open for communication to *secure.analyse-it.com* on your computer. Analyse-it uses your internet connection to simplify deactivation.

1. Start **Microsoft Excel** to start Analyse-it.
2. On the **Analyse-it** ribbon tab, click **Analyse-it**, and then click **Deactivate**.
The **Deactivate** window opens.
3. Click **Next**.
The software uses your internet connection to deactivate the license.
4. If deactivation fails for any reason you must manually deactivate:
   a) Click **Deactivate** to manually deactivate the license.

      Your browser opens the manual deactivation page at *http://analyse-it.com/deactivate*. If your computer does not have an internet connection, you can open the page from any computer or device with an internet connection to complete the deactivation of the license.
   b) On the manual deactivation page, enter the **Product Key** and **Site ID** from the **Deactivate** window.

      Note: The Site ID is unique to each computer. Do not contact technical support asking for a Site ID as we cannot provide it. It only shows on the **Deactivate** window in the software.
   c) In the **Deactivate** window, Click **Next**.
5. Click **Close**.

You can no longer use the software on your computer.

Acquiring a concurrent license

Acquire a concurrent-use license to use the software.

The software acquires a license when you use a command on the Analyse-it ribbon tab, so you do not normally need to acquire a license manually.

- Ensure that you have activated the software with a concurrent-user license.
- Ensure that TCP port 443 (*HTTPS*) on the firewall is open for communication to *secure.analyse-it.com* on each computer. Analyse-it uses your internet connection to communicate with our license server to manage the concurrent use of the software.

1. Start **Microsoft Excel** to start Analyse-it.
2. On the **Analyse-it** ribbon tab, click **Analyse-it**, and then click **Acquire**.
   The software connects to our license server and acquires a license so you can use the software.

If a license is acquired, you can now use the software. If all licenses are currently in use, you can try again later, can ask others using the software to release the license if they are no longer using the software, or you may consider purchasing more concurrent-user licenses to avoid the problem in future.

Releasing a concurrent license

Release a concurrent-use license so others can use the software.

The software releases the license when you close Microsoft Excel, but you can immediately release the license allocated to you if you want to continue using Excel but want to let another user start using Analyse-it.

- Ensure that you have activated the software with a concurrent-user license.
- Ensure that TCP port 443 (*HTTPS*) on the firewall is open for communication to *secure.analyse-it.com* on each computer. Analyse-it uses your internet connection to communicate with our license server to manage the concurrent use of the software.

1. Start **Microsoft Excel** to start Analyse-it.
2. On the **Analyse-it** ribbon tab, click **Analyse-it**, and then click **Release**.
   The software connects to our license server and releases the license allocated to you.

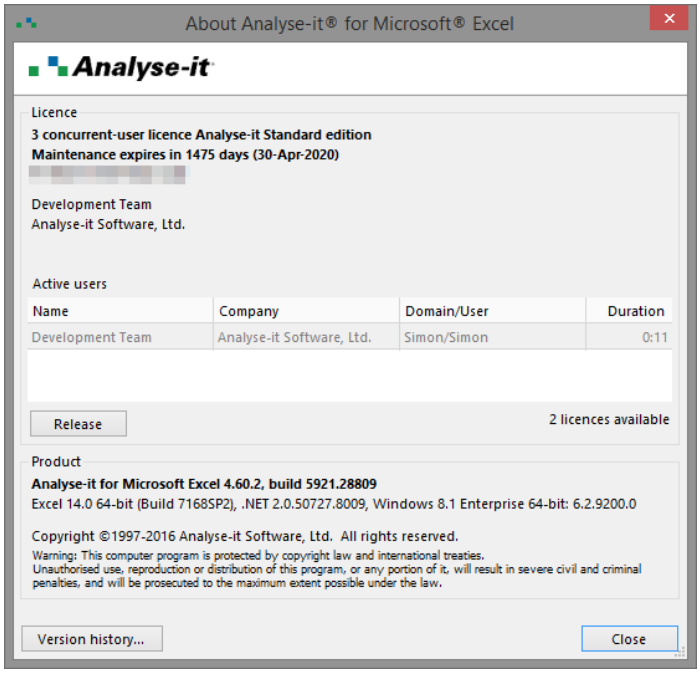
Another user can now use the software. If you subsequently use a command on the Analyse-it ribbon tab, a license is acquired for you again.

Finding out who's using a concurrent license

Find out who is using a concurrent license so that you can manage licenses.

- Ensure that you have activated the software with a concurrent-user license.
- Ensure that TCP port 443 (*HTTPS*) on the firewall is open for communication to *secure.analyse-it.com* on each computer. Analyse-it uses your internet connection to communicate with our license server to manage the concurrent use of the software.

1. Start **Microsoft Excel** to start Analyse-it.
2. On the **Analyse-it** ribbon tab, click **Analyse-it**, and then click **About Analyse-it**.
   The **Active user's** list shows who is currently using the software.



3. Click **Close**.

Chapter 3

Software updates

Software updates keep your software up-to-date with the latest features, bug fixes, and keeps it working when Microsoft release a new version of Microsoft Excel.

We regularly make updates to the software:

- Major updates include new features and important bug fixes.
- Minor updates address bug fixes that only affect a few customers or special situations.

The software periodically uses your internet connection to check for updates and notifies you if an update to the software is available. We do not notify you of minor updates that affect only a few customers; we only notify customers affected directly by e-mail or through the crash reporting system built-in to the software instead.

You must have active maintenance to download and install updates to the software.

Maintenance

Maintenance provides you with access to all software updates released during your maintenance period.

If you purchase a perpetual license, the initial purchase includes 1-year of maintenance. Renewing maintenance is optional. We notify you by e-mail 30 days before maintenance on your licenses expires and again on the day it expires. If maintenance expires you can continue to use the software but you do not receive any updates released after the maintenance expiry date.

If you purchase an annual license, maintenance is included and terminates when the license expires. You must renew the license to continue using the software.

Updating the software

Check that the software is up-to-date and includes all the latest features and bug fixes (including major and minor fixes).

- Ensure that you have an active maintenance.
- Ensure that TCP port 443 (*HTTPS*) on the firewall is open for communication to *secure.analyse-it.com* on your computer. Analyse-it uses your internet connection to check for updates.

1. Start **Microsoft Excel** to start Analyse-it.
2. On the **Analyse-it** ribbon tab, click **Analyse-it**, and then click **Check for Updates**.
   The **Software Update** window opens.
3. Click **Check Now**.
   The software uses your internet connection to check for updates.
4. If updates are available:
   - Click **Download and Install** to download and install the update.
   - Click **Download** to download the update for installation later.

When installing an update do not uninstall the existing version. If you do, the license is deactivated, and you must activate the license again before you can use the software.

5. Click **Close**.

Checking maintenance expiry and renewing

Maintenance provides you with access to software updates.

- Ensure that you have activated the software.

1. Start **Microsoft Excel** to start Analyse-it.
2. On the **Analyse-it** ribbon tab, click **Analyse-it**, and then click **Extend Maintenance**.
   Your browser opens the renew maintenance page at *http://analyse-it.com/store/maintenance* where you can see the maintenance expiry date and the cost to renew maintenance.

Chapter 4

Troubleshooting

If Analyse-it does not start when you start Microsoft Excel, you should troubleshoot the installation.

You can determine if Analyse-it is running by looking for the **Analyse-it** tab in the Excel ribbon. If you cannot see the tab you should follow these steps:

Checking the version installed

Check you are running Analyse-it version 3.00 or later.

1. Start the Windows **Control Panel**.
2. Open the list of installed programs:
   - On Windows XP, double-click **Add/Remove Programs**
   - On Windows Vista or Windows 7 and later, click **Programs**, and then click **Programs and Features**.
3. Check Analyse-it for Microsoft Excel appears in the list of applications and that the version number is 3.00 or later.
   If the version is less than 3.00, or you cannot find Analyse-it in the list, install the latest version.

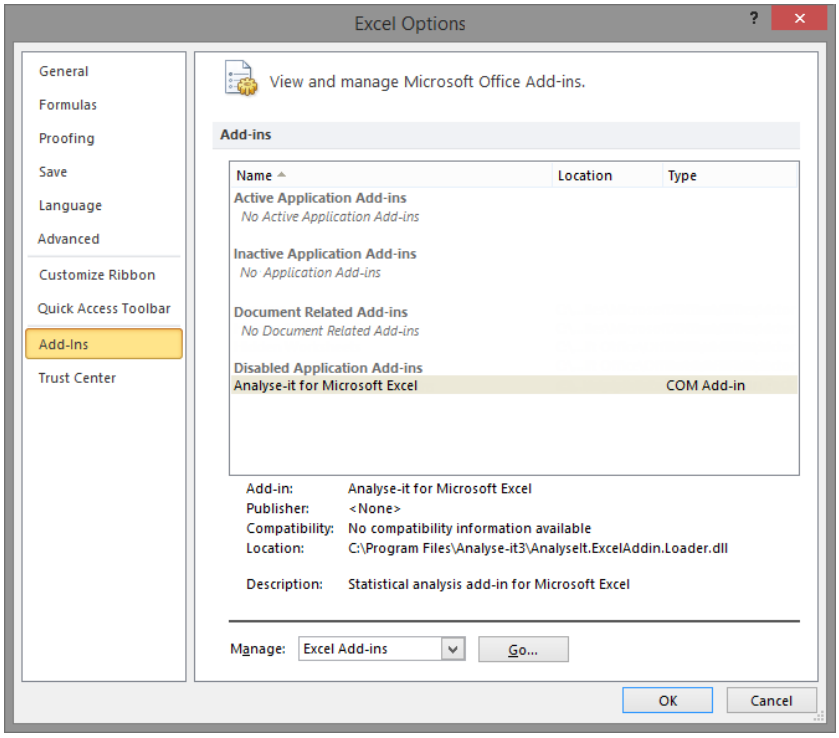
Enabling the add-in

Ensure the Analyse-it add-in is active in Microsoft Excel.

Microsoft Excel disables an add-in if it causes frequent crashes. This behavior prevents an add-in crashing Microsoft Excel so often that you are unable to use it. Sometimes these crashes are caused by interactions between incompatible, or unreliable add-ins installed on your computer.

1. Start **Microsoft Excel**.
2. Open the Microsoft Excel **Options** window:
   - On Excel 2007, click the **Office** button, and then click **Excel Options**.
   - On Excel 2010 and Excel 2013, click **File**, and then click **Options**.

   The **Excel Options** window opens.
3. Click **Add-Ins**.
   The list of installed add-ins opens.



4. If Analyse-it appears in the **Disabled Application Add-ins** section of the Add-ins list:
   a) In the **Manage** drop-down list, select **Disabled Items**, and then click **Go...**.
   The **Disabled Items** window opens.
   b) Select **Add-in: Analyse-it for Microsoft Excel**.
   c) Click **Enable** and then click **Close**.
   The **COM Add-Ins** window closes.
   d) Click **OK**.
   e) Exit and restart **Microsoft Excel**.
   Analyse-it should automatically start.

Repairing the installation

Ensure that all program files are properly installed and registered.

- Ensure that you have the necessary administrative rights to install software on the computer.

1. Start the Windows **Control Panel**.
2. Open the list of installed programs:
   - On Windows XP, double-click **Add/Remove Programs**.
   - On Windows Vista or Windows 7 and later, click **Programs**, and then click **Programs and Features**.
3. In the list of installed programs, select Analyse-it for Microsoft Excel.
4. Click **Repair**.
   The installer repairs the installation to ensure all program files are properly installed and registered.
5. Start **Microsoft Excel**.
   Analyse-it should automatically start.

Contacting technical support

Contact us if you have any problems or questions.

E-mail *support@analyse-it.com*.

Part

II

User's Guide

Chapter

Getting to know Analyse-it

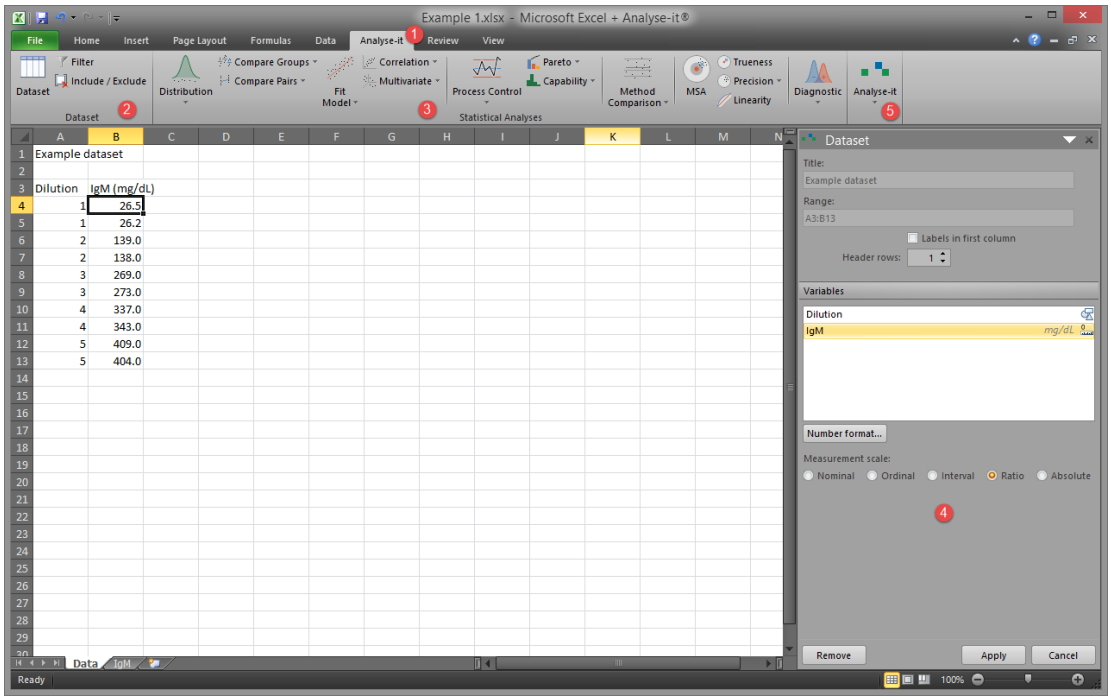
5

Analyse-it provides an intuitive user interface, built on the familiar foundation of Microsoft Excel. Its logical task-based workflow makes sense to those that aren't programmers or full-time statisticians.

All the commands are available on the context-sensitive ribbon tab labeled Analyse-it.

Dataset context

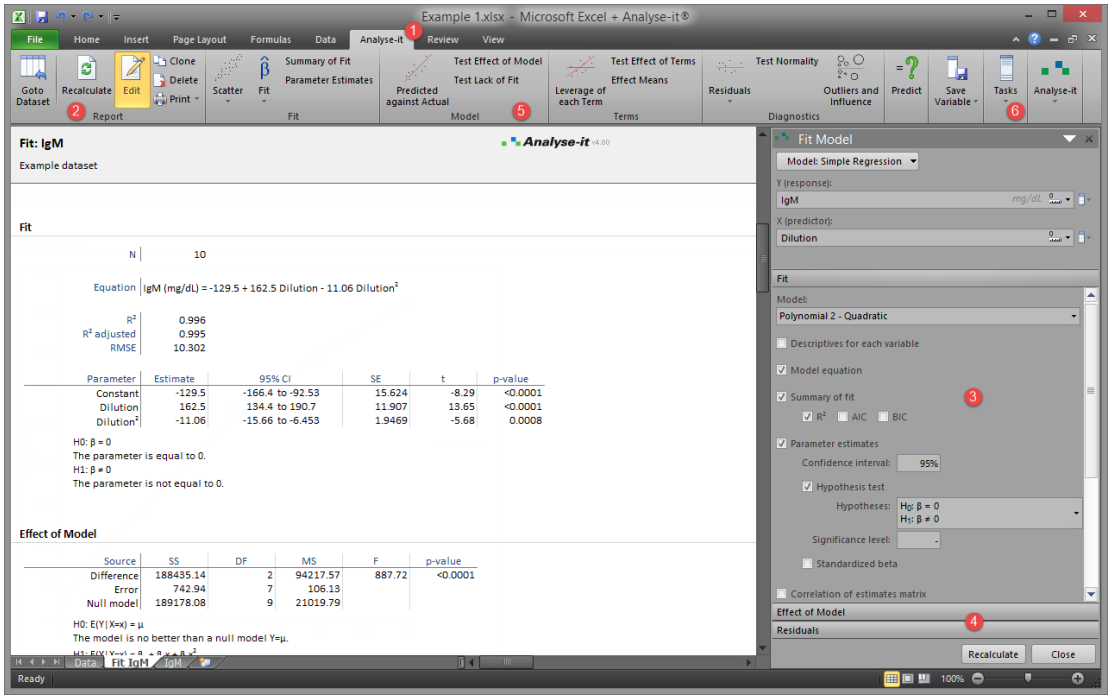
When you are on a standard Excel worksheet the Analyse-it ribbon tab shows commands to manage datasets and create new statistical analyses. The dataset task pane shows the dataset and variables properties.



1. Analyse-it ribbon tab.
2. Dataset commands.
3. Statistical analyses commands.
4. Dataset task pane.
5. Analyse-it command.

Analysis context

When you switch to an analysis report the Analyse-it ribbon tab shows commands to manage the report and perform additional statistical tasks. The analysis task pane shows options for the statistical analysis grouped in panels that facilitate a logical workflow.



1. Analyse-it ribbon tab.
2. Report commands.
3. Analysis task pane.
4. Task panels.
5. Add analysis task commands.
6. Remove analysis tasks.

Chapter

Preparing data for analysis

6

Analyse-it takes your data for analysis directly from an Excel worksheet. It also provides commands to arrange and structure your data ready to analyze. In addition, you can use the built-in Excel commands to transform, sort, and filter data.

Datasets

A dataset is a range of contiguous cells on an Excel worksheet containing data to analyze.

When arranging data on an Excel worksheet you must follow a few simple rules so that Analyse-it works with your data:

Taxes and Home Prices
http://lib.stat.cmu.edu/DASL/Stories/hometax.html

House	Sale price (100$)	Size (sqft)	Age (years)
Avalon	2050	2650	13
Cross Winds	2080	2600	*
The White House	2150	2554	6
The Rectory	2150	2921	3
Larchwood	1999	2580	4
Orchard House	1900	2580	4
Shangri-La	1800	2774	2
The Stables	1560	1920	1
Cobweb Cottage	1450	2150	*
Nairn House	1449	1710	1

1. Title to clearly describe the data. If you do not specify a title, the cell range of the dataset (such as A3:C13) is used to refer to the dataset.
2. A header row containing variable labels. Each variable name should be unique. Units of measurement can be included in the label by enclosing them in brackets after the name.
3. Rows containing the data for each case. The number of rows is only limited by Excel (currently over one million).
4. Columns containing the data for each variable.
5. Optional: Labels in the first column to provide a meaningful name/identifier for each case.

When you use an Analyse-it command, the extent of a dataset is determined by scanning outwards from the active cell to include all surrounding contiguous cells. The extent is known when a blank row or column surrounding the dataset, or the edge of the worksheet, is reached.

Price (100$)	Size (sqft)	Age (years)
2050	2650	13
2080	2600	*
2150	2664	6
2150	2921	3
1999	2580	4
1900	2580	4
1800	2774	2
1560	1920	1
1450	2150	*
1449	1710	1

Most Excel commands (for example, Sort, Pivot Table) use the same technique to determine the range of cells to operate on. It avoids the need to select often-large ranges of cells using the mouse, which is laborious and error-prone. It also ensures that if you add, or remove, cases or variables, that subsequent analyses automatically reflect any changes to the dataset.

When you analyze data, any data in hidden rows on the worksheet are excluded. This feature lets you easily limit analysis to a subset of the cases in the dataset. You can hide rows manually or use a filter to hide them based on criteria.

You can locate a dataset anywhere on a worksheet, and you can keep multiple datasets on a single worksheet provided you separate them from each other by at least one blank row and column. However, we recommend that you use a separate worksheet for each dataset. Using separate worksheets allows you to name datasets using the Excel worksheet tabs, navigate between datasets using the worksheet tabs, and ensures that filtering a dataset does not affect other datasets on the same worksheet.

Variables

A variable is an attribute of an object. The value of a variable can vary from one thing to another.

There are two different types of variables:

- Quantitative (or numeric) - data are numeric values of a quantity.
- Qualitative (or categorical) - data differs only in kind.

Record quantitative data as the numeric value.

Record qualitative data as a numeric coding (for example, 0, 1, 2) or labels (for example, Low, Medium, High). Numeric codings are often preferred as they succinctly represent the values and allow quicker data entry. However, codings can be difficult to understand by anyone other than the person who assigned them. To make their meaning clear, you can also assign a label to each numeric coding.

Data are ordered based on numeric value, numeric coding, or if a numeric coding is not used into alphabetical label order. For some statistical tests, the sort order determines how observations are rank-ordered and can affect the results. Often alphabetic sort order does represent the correct ordering. For example Low, Medium, High is sorted alphabetically into High, Low, Medium which does not represent their correct, natural order. In such cases, you must explicitly set the order before analyzing the data.

Measurement scales

Five different scales are used to classify measurements based on how much information each measurement conveys.

The different levels of measurement involve different properties of the numbers or symbols that constitute the measurements and also an associated set of permissible transformations.

Measurement scale	Properties	Permissible transformation
Nominal	Two things are assigned the same symbol if they have the same value of the attribute. For example, Gender (Male, Female); Religion (coded as 0=None, 1=Christian, 2=Buddhist).	Permissible transformations are any one-to-one or many-to-one transformation, although a many-to-one transformation loses information.
Ordinal	Things are assigned numbers such that the order of the numbers reflects an order relation defined on the attribute. Two things x and y with attribute values a(x) and a(y) are assigned numbers m(x) and m(y) such that if m(x) > m(y), then a(x) > a(y). For example, Moh's scale for the hardness of minerals; academic performance grades (A, B, C, ...).	Permissible transformations are any monotone increasing transformation, although a transformation that is not strictly increasing loses information.
Interval	Things are assigned numbers such that differences between the numbers reflect differences of the attribute. If m(x) - m(y) > m(u) - m(v), then a(x) - a(y) > a(u) - a(v). For example, Temperature measured in degrees Fahrenheit or Celsius.	Permissible transformations are any affine transformation t(m) = c * m + d, where c and d are constants; another way of saying this is that the origin and unit of measurement are arbitrary.
Ratio	Things are assigned numbers such that differences and ratios between the numbers reflect differences and ratios of the attribute. For example, Temperature measured in degrees Kelvin scale; Length in centimeters.	Permissible transformations are any linear (similarity) transformation t(m) = c * m, where c is a constant; another way of saying this is that the unit of measurement is arbitrary.
Absolute	Things are assigned numbers such that all properties of the numbers reflect analogous properties of the attribute. For example, Number of children in a family, Frequency of occurrence.	The only permissible transformation is the identity transformation.

While the measurement scale cannot determine a single best statistical method appropriate for data analysis, it does define which statistical methods are inappropriate. Where possible the use of a variable is restricted when its measurement scale is not appropriate for the analysis. For example, a nominal variable cannot be used in a t-test.

When the measurement scale of a variable is unknown, the scale is inferred from its role in the analysis and the type of data in the variable. If the measurement scale cannot be inferred, you must set the measurement scale.

Missing values

Missing values occur when no data is recorded for an observation; you intended to make an observation but did not. Missing data are a common occurrence and can have a significant effect on the statistical analysis.

Missing values arise due to many reasons. For example, a subject dropped out of the study; a machine fault occurred that prevented a measurement been taken; a subject did not answer a question in a survey; or a researcher made a mistake recording an observation.

A missing value is indicated by an empty cell or a . (full stop) or * (asterisk) in the cell. These values always indicate a missing value, regardless of measurement scale of the variable. Similarly, cell values that are not valid for the measurement scale of the variable are treated as missing. For example, a cell containing text or a #N/A error value is treated as missing for a quantitative measurement scale variable, since neither are valid numeric values.

Some applications, especially older statistical software, use special values such as 99999 to indicate missing values. If you import such data, you should use the Excel Find & Replace command to replace the values with a recognized missing indicator.

In the presence of missing values you often still want to make statistical inferences. Some analyses support techniques such as deletion, imputation, and interpolation to allow the analysis to cope with missing values. However, it is still important before performing an analysis that you understand why the data is missing. When the reason for the missing data is not completely random, the study may be biased, and the statistics may be misleading.

Case and frequency form data

Most data are available in case form with data for each sampling unit. Sometimes data are not available for each unit but are already summarized by counting the frequency of occurrences of each value, called frequency form data.

Whenever possible you should record data in case form. Case form data can be reduced to frequency form, but frequency form data cannot be reconstructed into case form. You may want to use frequency form data when working with a large amount of data from a database, where you can save computing resources by letting the database server tabulate the data.

Case form dataset layout

Each column is a variable (Height, Eye color), each row is a separate case (Subject) with the values of the variables on that case.

Subject (optional)	Height	Eye color
1	175	Blue
2	180	Blue
3	160	Hazel
4	190	Green
5	180	Green
6	150	Brown
7	140	Blue
8	160	Brown

Subject (optional)	Height	Eye color
…	…	…

Frequency form dataset layout

Each column is a variable (Eye color) and a separate column for the number of cases (Frequency), each row is a combination of categories with the frequency count.

Eye color	Frequency
Brown	221
Blue	215
Hazel	93
Green	64

Note: Examples of dataset layouts are included in the Statistical Reference Guide.

Setting the measurement scale of a variable

Set the measurement scale of a variable so data analysis can use the variable correctly.

Note: If you have the analysis task pane open, you can quickly set the measurement scale of a variable. In the analysis task pane, click the drop-down menu button alongside the variable, and then click the measurement scale.

1. Activate the dataset worksheet.
2. On the **Analyse-it** ribbon tab, in the **Dataset** group, click **Dataset**.
   The dataset task pane opens.
3. In the **Variables** list, select the variables:
   - To select a single variable, click the variable.
   - To select multiple variables, click the first variable then hold down the CTRL key and click each additional variable.
   - To select a range of variables, click the first variable then hold down the SHIFT key and click the last variable in the range.

   The cells of the selected variables highlight on the worksheet. You can also select cells on the worksheet to select the corresponding variables in the list.
4. Click **Nominal**, **Ordinal**, **Interval**, **Ratio**, or **Absolute**.
5. Click **Apply**.
   The changes are saved and will be used in future analyses.

Setting the minimum, maximum and units of a variable

Set the minimum/maximum and units so that the axes of charts are scaled to your preferences.

Note: If you do not set a scale then the chart axis scale will be determined automatically by the data values. If you do specify the scale, but the data values are outside the minimum/maximum then axes are scaled to the next unit. If the unit is too small to show with a label on the axis it is shown as a minor tick mark.

Note: If you have the analysis task pane open you can quickly change the scaling for a variable: right-click over the variable, or click the drop-down menu button alongside the variable selector, and then click **Scale**.

1. Activate the dataset worksheet.
2. On the **Analyse-it** ribbon tab, in the **Dataset** group, click **Dataset**.
   The dataset task pane opens.
3. In the **Variable** list, select the variables:
   - To select a single variable, click the variable.
   - To select multiple variables, click the first variable then hold down the CTRL key and click each additional variable.
   - To select a range of variables, click the first variable then hold down the SHIFT key and click the last variable in the range.

   The cells of the selected variables highlight on the worksheet. You can also select cells on the worksheet to select the corresponding variables in the list.
4. If the measurement scale is not already set, click **Interval**, **Ratio**, or **Absolute**.
5. In the **MinimumMaximum** and **Units** edit boxes, optionally, type the value for each. You can provide a minimum value only, a maximum value only, or a minimum, maximum and units. If you provide units then the range (the maximum - minimum) must be a factor of the units (the range must be divisible by the units).
6. Click **OK**.
   Any future or updated analysis will now use the preferred scale where possible.

Ordering categorical data

Set the order of categories for data analysis and output in the analysis report.

1. Activate the dataset worksheet.
2. On the **Analyse-it** ribbon tab, in the **Dataset** group, click **Dataset**.
   The dataset task pane opens.
3. In the **Variables** list, select the variable.
4. Click **Nominal** or **Ordinal** to set the measurement scale.
   The **Categories** grid shows the numeric coding and labels used in the data.
5. Clear the **Sort Categories** check box to allow you to reorder the categories manually.
6. Arrange the categories into order:
   - Using drag and drop, click and hold the mouse button down over a category until the drag insertion cursor appears, drag it to reposition it in the list and then release the mouse button.
   - Using the arrows, click a category to select it and then use the up and down arrow buttons to reposition it in the list.
7. Click **Apply**.
   The changes are saved and will be used in future analyses.

Setting the number format of a variable

Set the precision of the data so that the results of the analysis show to a suitable number of decimal places.

Note: If you do not set the number format (or you use the Excel General number format), the number of decimal places is determined by the data value with the most decimal places.

1. Activate the dataset worksheet.

2. On the **Analyse-it** ribbon tab, in the **Dataset** group, click **Dataset**.
   The dataset task pane opens.
3. In the **Variable** list, select the variables:
   - To select a single variable, click the variable.
   - To select multiple variables, click the first variable then hold down the CTRL key and click each additional variable.
   - To select a range of variables, click the first variable then hold down the SHIFT key and click the last variable in the range.

   The cells of the selected variables highlight on the worksheet. You can also select cells on the worksheet to select the corresponding variables in the list.
4. Click **Number format**.
5. In the **Format Cells** window, click **Number**, set the number of **Decimal places**, and then click **OK**.
6. Click **OK**.
   The data is re-formatted.

Labeling cases

Set a variable to use for labeling data points in plots for easy identification.

- Ensure that the variable containing the labels is the first (leftmost) column in the dataset.

1. Activate the dataset worksheet.
2. On the **Analyse-it** ribbon tab, in the **Dataset** group, click **Dataset**.
   The dataset task pane opens.
3. Select the **Labels in first column** check box.
4. Click **Apply**.
   The changes are saved and will be used in future analyses.

Assigning labels to categories

Assign labels to a numerical coding to make an analysis report more readable.

1. Activate the dataset worksheet.
2. On the **Analyse-it** ribbon tab, in the **Dataset** group, click **Dataset**.
   The dataset task pane opens.
3. In the **Variables** list, select the variable.
4. If the measurement scale is not set, click **Nominal** or **Ordinal**.
   The **Categories** grid shows the numeric coding and labels used in the data.
5. In the **Name** column alongside each coding, type the label.

   Note: Names must be unique.
6. Click **Apply**.
   The changes are saved and will be used in future analyses.

Assigning colors/symbols to categories

Assign colors or symbols to highlight observations on plots.

1. Activate the dataset worksheet.
2. On the **Analyse-it** ribbon tab, in the **Dataset** group, click **Dataset**.
   The dataset task pane opens.
3. In the **Variables** list, select the variable.

4. If the measurement scale is not set, click **Nominal** or **Ordinal**.
   The **Categories** grid shows the codings used in the data.
5. Select the **Assign colors/symbols** check box.
   The **Categories** grid shows the Symbol and Color columns.
6. Optional: Click a cell in the **Color** column and select a preset color.
7. Optional: Click a cell in the **Symbol** column and select a symbol.
8. Repeat steps *6* through *7* for each coding.

   Note: Any categories without an assigned symbol use a square, and any without an assigned color use black. You can assign symbols and colors just to the categories you want to highlight. The others use the default symbol and color.
9. Click **Apply**.
   The changes are saved and will be used in future analyses.

Transforming variables

Transform a variable to normalize, shift, scale or otherwise change the shape of the distribution so that it meets the assumptions of a statistical test.

1. Activate the dataset worksheet.
2. In an empty column adjoining the dataset, enter the transformation function.
   For example =LOG(A1)
3. Double-click on the drag handle at the bottom right of the active cell to copy the formula down the entire column.

Transform functions

Useful Microsoft Excel functions for transforming data.

Transform	Excel function
Natural log	=LN(*cell*)
Inverse natural log	=EXP(*cell*)
Reciprocal	=1 / *cell*
Square root	=SQRT(*cell*)
Power	=*cell*^*power*
Log base 10	=LOG(*cell*) or =LOG10(*cell*)
Inverse log base 10	=10^*cell*
Log base 2	=LOG(*cell*, 2)
Inverse log base 2	=2^*cell*

Analyzing a subset of the data

Use an Excel filter to show only the rows that match a value or meet a condition. Rows that do not match the condition are hidden temporarily, and subsequent data analysis excludes them. This feature is extremely useful when you want to see how the analysis changes for a different subset of the data.

Note: Prior to version 5.10 the active filter is not saved with the analysis. When you recalculate an analysis the rows currently visible are analyzed, regardless of the filter you used when creating the analysis. From version 5.10 onwards, by default the filter criteria are saved with the analysis

and re-applied when you recalculate or make other changes to the analysis. The drop-down arrow next to the Filter command on the ribbon of the active analysis lets you change between **Use active filter** which always uses the currently active filter when recalculating (pre version 5.10 behavior) and **Save Filter & Re-apply** (version 5.10 or later behavior) which saves the filter when the analysis is calculated and re-applies it on subsequent recalculation.

1. On the **Analyse-it** ribbon tab, click **Filter** to show filter icons on the active dataset.

   If you are on an analysis report, the saved filter (if any) will be applied to the dataset. You can modify the filter and return to the analysis and click Recalculate to apply the new filter, or you can start a new analysis to reuse the filter criteria.

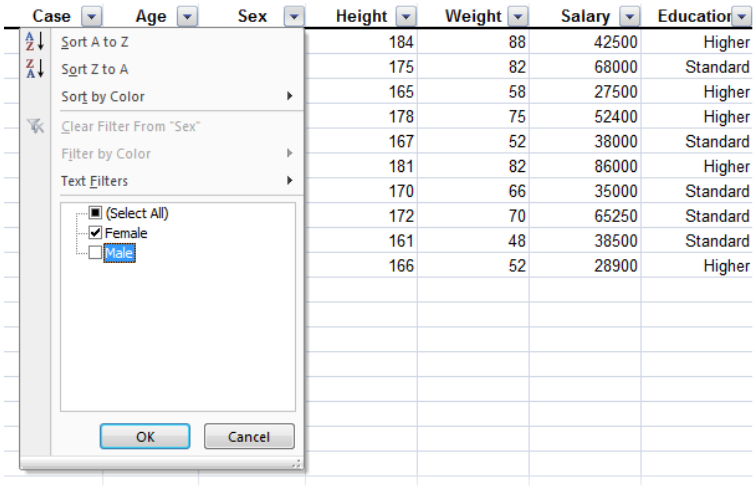
2. Click the drop-down button alongside a variable, and then either select the values to match, or apply other criteria.

   The save filter feature in version 5.10 or later works with almost all filter conditions: filter by value, by condition, by color/fill/icon, above/below average, top/bottom items, and so on. There is one limitation: filtering to multiple dates. If you filter to more than 2 dates, add-ins such as Analyse-it cannot retrieve the list of values to filter to. This is a limitation of Microsoft Excel and unfortunately it hasn't been remedied in the last few years so it is low on Microsoft's priorities. If this is seriously affecting you, please get in touch with Microsoft and ask them to resolve this limitation for add-in developers.

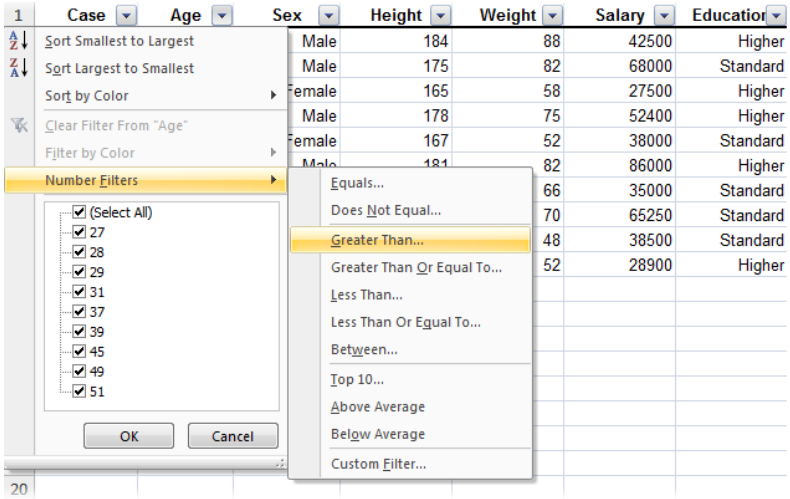
3. Create or recalculate an analysis to see the effect of the filter.

Examples

To filter to Female cases only, click the drop-down button alongside Sex, and then select Female in the list of values to match against:



To filter to cases where the subject is older than 40, click the drop-down button alongside Age, and then click **Number Filters** > **Greater Than**, and then enter 40 as the cut-off point:



Chapter 7

Working with analyses

Statistical analyses describe the observed data, make inferences about a population based on a sample of data, fit models that describe the relationship between variables, reduce the dimensionality of large datasets, and much more.

Creating an analysis

Create a statistical analysis and present the results.

The basic sequence of steps to create an analysis are:

1. Click a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click the analysis command, and then click a specific task command.
   The analysis task pane opens.
3. Select the variables to analyze.
4. Add any additional plots, tests or statistics.
5. Set any options for analysis.
6. Click **Calculate**.
   The analysis report opens.

Analysis reports

An analysis report is a standard Excel worksheet containing statistics and plots calculated by a statistical analysis. You can print, send, move, copy, and save analysis reports like any other worksheet.

Along with the results of the analysis, the report includes a header that describes the analysis, the dataset and variables analyzed, who performed the analysis and when, and the version of Analyse-it used.

Analysis reports are static. If the underlying data changes, the statistics and plots do not update automatically. You must recalculate an analysis to update it. For example, if you exclude or change the value of an observation, add or remove a case, or change the active filter, you must recalculate the analysis to see the effect of the changes.

You can also edit an analysis to make changes such as adding or removing a plot, performing a statistical test, or estimating a parameter.

Note: The analysis report stores a link to the dataset; the link is used to fetch the latest data when you recalculate the analysis. The link to the dataset can become broken if you choose to keep the dataset and analyses in separate workbooks, then move, rename or delete the dataset workbook on disk. When the link is broken, the analysis cannot be changed or updated. To avoid this potential problem we recommend you keep the dataset and associated analyses in the same workbook.

Editing an analysis

Change an analysis and update the report.

1. Activate the report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Report** group, click **Edit**.
   The analysis task pane opens.
3. Change the variables, set the analysis options, or add/remove tasks.
4. Click **Recalculate**.
   The analysis report updates.

Recalculating an analysis

Recalculate an analysis when the data has changed, and you want to update the report to reflect those changes.

Note: Be aware that when you recalculate an analysis, the current analysis report worksheet is deleted and replaced with a new worksheet. If you have made changes to the layout or formatting of the analysis report, the changes are lost.

1. Activate the report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Report** group, click **Recalculate**.

   Note: Prior to version 5.10 the active filter is not saved with the analysis. When you recalculate an analysis the rows currently visible are analyzed, regardless of the filter you used when creating the analysis. From version 5.10 onwards, by default the filter criteria are saved with the analysis and re-applied when you recalculate or make other changes to the analysis. The drop-down arrow next to the Filter command on the ribbon of the active analysis lets you change between **Use active filter** which always uses the currently active filter when recalculating (pre version 5.10 behavior) and **Save Filter & Re-apply** (version 5.10 or later behavior) which saves the filter when the analysis is calculated and re-applies it on subsequent recalculation.

   The analysis recalculates and the report updates.

Excluding observations from analysis

Exclude observations, such as outliers or influential observations, from analysis to see their effect on the results.

Before you exclude observations from the analysis, you should fully investigate them. Rather than a nuisance, outliers can sometimes be the most interesting and insightful observations in the data.

1. Activate the dataset worksheet.
2. Select the cells containing the observations to exclude/include.
3. On the **Analyse-it** ribbon tab, in the **Dataset** group, click **Include / Exclude**.
   The observations will be excluded (or re-included) in future analyses.

Observations to exclude from the analysis are enclosed in square brackets (for example, [65] or [Male]) in the dataset. Excluded observations are treated the same as missing values by an analysis.

We recommend you attach an Excel comment to the cell to document the reason for excluding the observation.

Creating a new analysis based on an existing analysis

Change an analysis and output the results to a new report.

1. Activate the report worksheet.

2. On the **Analyse-it** ribbon tab, in the **Report** group, click **Clone**.
   The analysis task pane opens with the existing analysis options.
3. Change the variables, set the analysis options, or add/remove tasks.
4. Click **Calculate**.
   A new analysis report opens. The original analysis report is left unchanged.

Printing an analysis report

Print a publication-quality report of the analysis.

Note: Reports include page breaks, so tables and plots are not split across multiple pages when printed. The paper size for the default Windows printer determines where to place page breaks. If you are in the US or Canada, the paper size for your printer will most likely be Letter (8.5" x 11"). If the paper size for your printer is not Letter, then the A4 paper size (210mm x 297mm) is used.

1. Activate the report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Report** group, click **Print**.
   The **Print Preview** window opens.
3. Choose the printer options.
4. Click **Print**.

Printing in color / black & white

Set whether to print reports in color or black and white.

Note: By default, reports print in black and white, even if you have a color printer.

1. Activate the report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Report** group, click the **Print** drop-down button, and then click **Color** or **Black and White**.
   Your choice is saved as the default. In future, you can just click **Print** to print using the new default.
3. Choose the printer options.
4. Click **Print**.

Printing a single chart on a full page

Print a single high-resolution chart scaled to fill a single page.

1. Activate the report worksheet.
2. Click on the plot to select it.
   The plot shows a thick border around it to indicate that it is selected.
3. On the **Analyse-it** ribbon tab, in the **Report** group, click **Print**.
   The **Print Preview** window opens.
4. Choose the printer options.
5. Click **Print**.

Showing the dataset analyzed

View the data analyzed and exclude observations, filter, or make other changes to it.

1. Activate the report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Report** group, click **Go to Dataset**.
   The workbook containing the dataset opens.

Changing the user name on analysis reports

Set the Excel user name to show on analysis reports.

1. Open the **Options** window:
   - In Excel 2007, click the **Office** button, and then click **Excel Options**.
   - In Excel 2010, click **File**, and then click **Options**.
2. Change the **User name**.
3. Click **OK**.
   The changes are saved and will be used in future analyses.

Numerical accuracy

All calculations are performed using double-precision IEEE 754 standard, highly reliable numerical methods, and then tested rigorously.

To ensure that Analyse-it produces accurate results it does not use any Excel mathematical, statistical or distribution functions. Many of these functions in early versions of Microsoft Excel used poor algorithms that often produced inaccurate results. Functions like STDEV and LINEST were affected, as were add-ins that used those functions. Microsoft has now addressed most of the problems, though you may still hear warnings against using Microsoft Excel for statistical analysis.

To avoid any of these problems Analyse-it uses reliable numerical algorithms and all calculations are performed using double-precision (IEEE 754 standard, effectively 15 significant digits). We then verify the statistics are correct, and remain so, throughout the development and release process using an in-house library of thousands of validation unit-tests.

Presentation of numerical data

Although all results are computed to high precision, they are presented only to a practical number of decimal places based on the numerical precision of the data.

To decide how many decimal places to present the results to Analyse-it uses the Excel number format you apply to the data. From the format, the precision you are working with is inferred, and therefore the precision you likely expect in the presentation of the statistics. Statistics such as the mean, median, and standard deviation are presented to one more decimal place than the source data.

Although the statistics are presented to only a few decimal places, the Excel cell contains the full precision value. You should use the full precision value to minimize error when using it in another calculation.

If you are preparing an analysis for presentation you can set the Excel number format applied to a cell containing a statistic on the analysis report. Do not format results if you intend to edit or recalculate the analysis as your formatting changes will be lost when you update the report.

Showing a statistic to full precision

Use the full precision of a statistic when you want to use it in further calculations.

1. If the Excel formula bar is not visible, on the **View** ribbon tab, in the **Show** group, select the **Formula bar** check box.
2. Click in the cell containing the statistic.
   The Excel formula bar shows the value with up to 15 significant digits.

Chapter

Feedback and crash reports

8

Suggestions, bug, and crash reports you send to our development team help shape the future of Analyse-it.

The feedback you send helps us to identify, diagnose and fix application crashes and prioritize development time toward the most requested features. When sending feedback try to include as much information as possible to help us understand the issue. We read all your feedback, but we only respond to you if we need more information about your report.

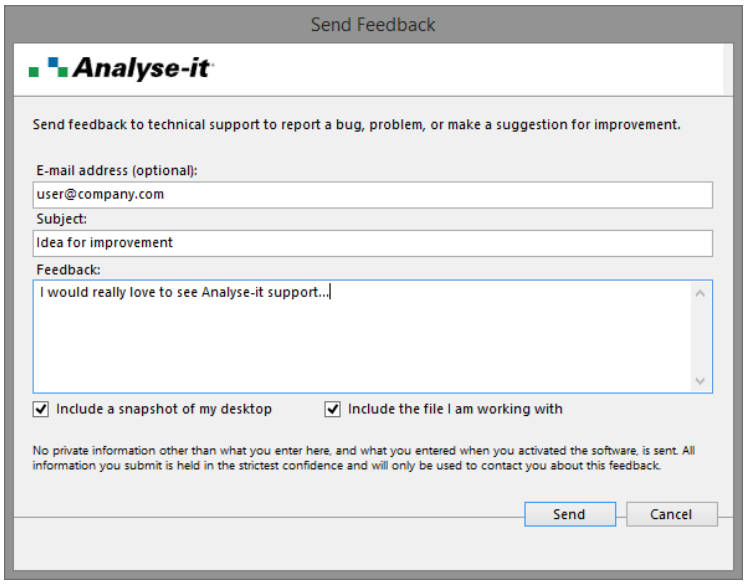
Do not use feedback for support requests. For those, contact us directly *support@analyse-it.com*

Note: Feedback and crash reporting use your internet connection. If your computer is behind a firewall, ensure the TCP port 443 (HTTPS) on the firewall is open for communication to *secure.analyse-it.com*.

Sending feedback

Submit your ideas for improvements, feature requests, and bug reports.

1. On the **Analyse-it** ribbon tab, click **Analyse-it**, and then click **Send Feedback**.
   The **Send Feedback** window opens.



2. Enter an **E-mail** address where we can contact you.

   We respect your privacy at all times. We do not collect personal information, and we only use your e-mail address if we need to contact you about your problem or suggestion.
3. Enter a **Subject** and **Feedback** to describe your comment, suggestion, or bug report. Include as much information as possible in the feedback, including the steps to reproduce the problem if applicable.

4. Select **Include a snapshot of my desktop** if it helps to show the problem or suggestion you are making.
5. Select **Include the file I am working with** to include the active Excel workbook if it contains data that helps to show the problem or suggestion you are making.

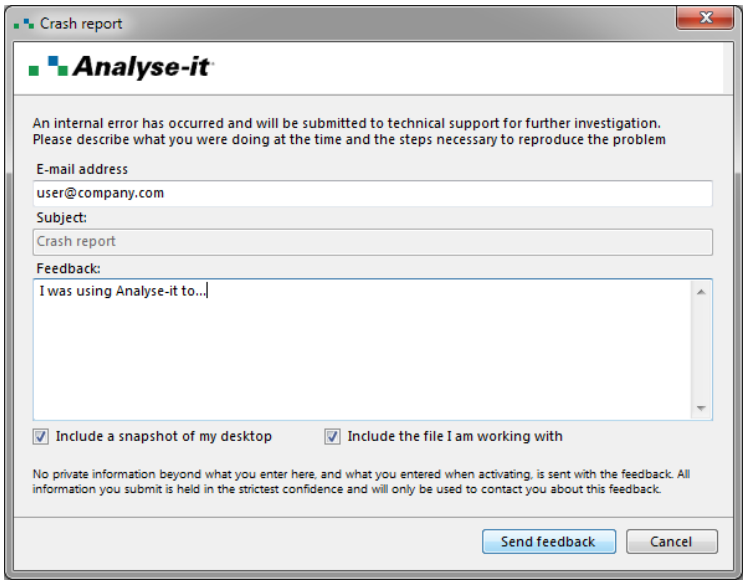
   We treat any data received as private and confidential. It is used only for the purpose you submit it, and we delete it when no longer needed.
6. Click **Send**.
   Your feedback is sent using your internet connection.

Submitting crash reports

Submit a crash report whenever you experience a problem so that we can try to solve it.

If Analyse-it crashes or experiences a problem, it creates a crash report and prompts you to send it to us for investigation. Crash reports include technical information (application state, hardware configuration, and stack traces) to help us diagnose and fix the problem. Before Analyse-it submits the crash report it checks if a fix is available. If so, Analyse-it gives you details on how to fix the problem. If no fix is available, you should submit the crash report to help us improve the application. Even if you experience a crash that seems similar to previous crashes you should still submit it as each report often provides new insight or perspective on a crash.

1. If Analyse-it detects a serious problem, the **Crash Report** window opens.



2. Enter an **E-mail address** where we can contact you.

   We respect your privacy at all times. We do not collect any personal information, and we only use your e-mail address if we need to contact you about the crash report.
3. Enter **Feedback** to describe what you were doing when the crash occurred and, if possible, the steps to reproduce the problem.
4. Select **Include a snapshot of my desktop** if it helps to show the problem.
5. Select **Include the file I am working with** to include the active Excel workbook if it contains data that helps to show the problem.

   We treat any data received as private and confidential. It is used only for the purpose you submit it, and we delete it when no longer needed.

6. Click **Send**.
   Your crash report is sent using your internet connection.

After submitting the crash report, Analyse-it attempts to recover so you can save your work and continue using the software. If the crash was serious and causes repeated subsequent crashes, you should restart Microsoft Excel and Analyse-it.

Part

III

Statistical Reference Guide

Distribution

Chapter 9

Distribution analysis describes the distribution of a variable and makes inferences about population parameters.

Continuous distributions

A continuous distribution describes a variable that can take on any numeric value (for example, weight, age, or height).

Univariate descriptive statistics

Descriptive statistics provide information about the central location (central tendency), dispersion (variability or spread), and shape of the distribution.

There are many measures of central location, dispersion, and shape:

Statistic	Purpose
N	The number of non-missing values in a set of data.
Sum	The sum of the values in a set of data.
Mean	Measure the central tendency using the arithmetic mean.
Harmonic mean	Measure the central tendency using the reciprocal of the arithmetic mean of the reciprocals. Useful when the values are rates and ratios.
Geometric mean	Measure the central tendency using the product of the values. Useful when the values are percentages.
Variance	Measure the amount of variation using the squared deviation of a variable from its mean.
Standard deviation	Measure the amount of variation using the square root of the variance.
CV% / RSD	Measure the spread relative to its expected value (standard deviation divided by the mean). Also known as the coefficient of variation or relative standard deviation.
Skewness	Measure the "sideness" or symmetry of the distribution. Skewness can be positive or negative. Negative skew indicates that the tail on the left side of the distribution is longer or fatter than the right side. Positive skew indicates the converse, the tail on the right side is longer or fatter than the left side. A value of zero indicates the tails on both sides balance; this is the case for symmetric distributions, but also for asymmetric distributions where a short fat tail balances out a long thin tail.

Statistic	Purpose
	Uses the Fisher-Pearson standardized moment coefficient G_1 definition of sample skewness.
Kurtosis	Measure the "tailedness" of the distribution. That is, whether the tails are heavy or light. Excess kurtosis is the kurtosis minus 3 and provides a comparison to the normal distribution. Positive excess kurtosis (called leptokurtic) indicates the distribution has fatter tails than a normal distribution. Negative excess kurtosis (called platykurtic) indicates the distribution has thinner tails than a normal distribution. Uses the G_2 definition of the sample excess kurtosis
Median	Measure the central tendency using the middle value in a set of data.
Minimum	Smallest value in a set of data.
Maximum	Largest value in a set of data.
Range	The difference between the maximum and minimum.
1st quartile	The middle value between the smallest value and median in a set of data.
3rd quartile	The middle value between the median and largest value in a set of data.
Interquartile range	Measure the spread between the 1st and 3rd quartile.
Mode	The value that appears the most in a set of data.
Quantiles	A set of values that divide the range of the distribution into contiguous intervals defined by probabilities.

For normally distributed data, the mean and standard deviation provide the best measures of central location and dispersion.

For data with a non-normal or highly-skewed distribution, or data with extreme values, the median and the first and third quartiles provide better measures of central location and dispersion. When the distribution of the data is symmetric, the inter-quartile range (IQR) is a useful measure of dispersion. Quantiles further describe the distribution of the data, providing an interval containing a specified proportion (for example, 95%) of the data or by breaking the data into intervals each containing a proportion of the data (for example, deciles each containing 10% of the data).

Calculating univariate descriptive statistics

Use univariate descriptive statistics to describe a quantitative variable.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Distribution**, and then click the statistics to show:

Mean and Moments	Show mean, variance, standard deviation, skewness, and kurtosis.
Median and Quartiles	Show minimum, maximum, median, and quartiles.
Quantiles	Show quantiles.

   The analysis task pane opens.
3. In the **Y** drop-down list, select the variable.

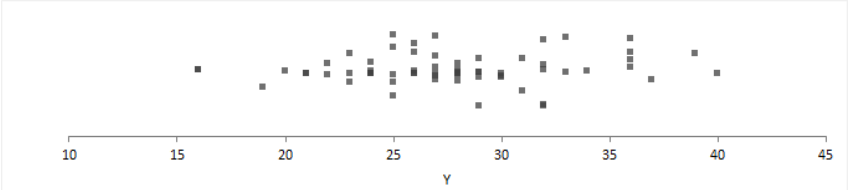
4. Optional: To customize the statistics to show, click **Customize...** and then select or clear the appropriate check boxes. To save the options as the defaults to use for future analyses, click **Save as Defaults**.
5. Click **Calculate**.

Univariate plot

A univariate plot shows the data and summarizes its distribution.

Dot plot

A dot plot, also known as a strip plot, shows the individual observations.

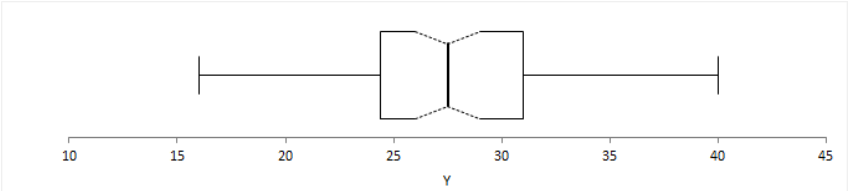


A dot plot gives an indication of the spread of the data and can highlight clustering or extreme values.

Box plot

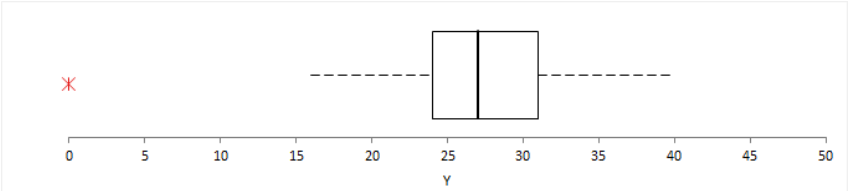
A box plot shows the five-number summary of the data – the minimum, first quartile, median, third quartile, and maximum. An outlier box plot is a variation of the skeletal box plot that also identifies possible outliers.

Skeletal box plot



A skeletal box plot shows the median as a line, a box from the 1st to 3rd quartiles, and whiskers with end caps extending to the minimum and maximum. Optional notches in the box represent the confidence interval around the median.

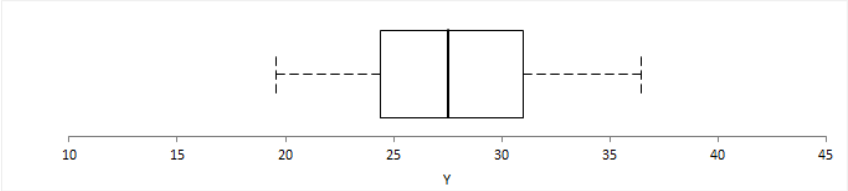
Outlier box plot



An outlier box plot is a variation of the skeletal box plot, but instead of extending to the minimum and maximum, the whiskers extend to the furthest observation within 1.5 x IQR from the quartiles. Possible near outliers (orange plus symbol) are identified as observations further than 1.5 x IQR

from the quartiles, and possible far outliers (red asterisk symbol) as observations further than 3.0 x IQR from the quartiles. You should investigate each possible outlier before deciding whether to exclude it, as even in normally distributed data, an outlier box plot identifies approximately 0.7% of observations as possible outliers.

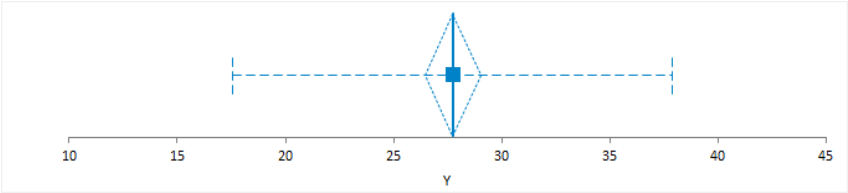
Quantile box plot



A quantile box plot is a variation on the skeletal box plot and shows the whiskers extending to specific quantiles rather than the minimum and maximum value.

Mean plot

A mean plot shows the mean and standard deviation of the data.



A line or dot represents the mean. A standard error or confidence interval measures uncertainty in the mean and is represented as either an error bar or diamond.

An optional error bar or band represents the standard deviation. The standard deviation gives the impression that the data is from a normal distribution centered at the mean value, with most of the data within two standard deviations of the mean. Therefore, the data should be approximately normally distributed. If the distribution is skewed, the plot is likely to mislead.

Creating a univariate plot

Plot a dot plot, box plot, or mean plot to visualize the distribution of a single quantitative variable.

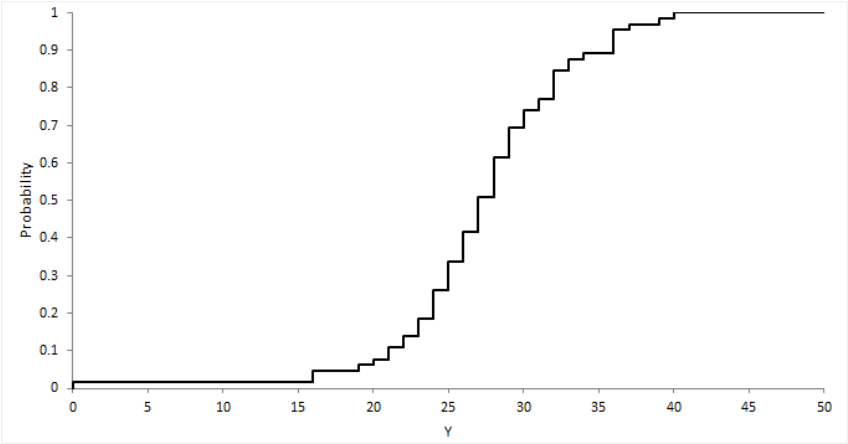
1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Distribution** > **Univariate**, and then click the plot type.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the variable.
4. Optional: To overlay more than one plot type on the same chart, on the **Analyse-it** ribbon tab, in the **Distribution** group, click **Univariate**, and then click an additional plot type.
5. Optional: In the analysis task pane, customize the options for the plot.
6. Click **Calculate**.

Frequency distribution

A frequency distribution reduces a large amount of data into a more easily understandable form.

Cumulative distribution function plot

A cumulative distribution function (CDF) plot shows the empirical cumulative distribution function of the data.



The empirical CDF is the proportion of values less than or equal to X. It is an increasing step function that has a vertical jump of 1/N at each value of X equal to an observed value. CDF plots are useful for comparing the distribution of different sets of data.

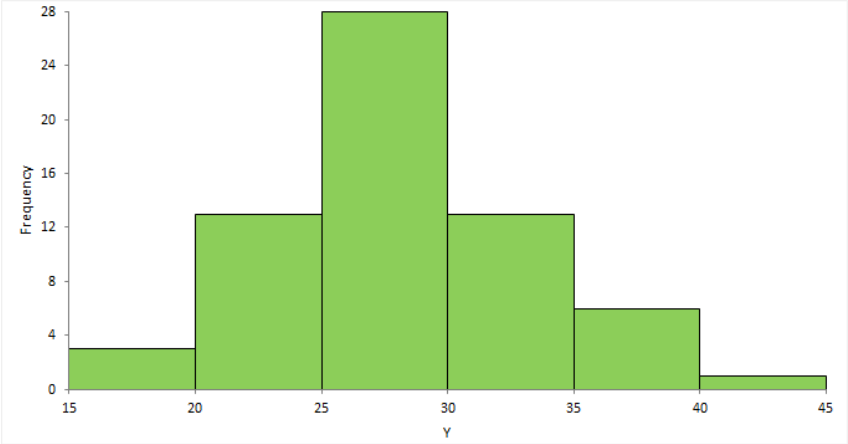
Creating a CDF plot

Plot a CDF to visualize the shape of the distribution of a quantitative variable.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Distribution** > **CDF**, and then click the plot type.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the variable.
4. Optional: To show the confidence band, in the **Confidence band** edit box, type the confidence level.
5. Click **Calculate**.

Histogram

A histogram shows the distribution of the data to assess the central tendency, variability, and shape.



A histogram for a quantitative variable divides the range of the values into discrete classes, and then counts the number of observations falling into each class interval. The area of each bar in the histogram is proportional to the frequency in the class. When the class widths are equal, the height of the bar is also proportional to the frequency in the class.

Choosing the number of classes to use can be difficult as there is no "best," and different class widths can reveal or hide features of the data. Scott's and Freedman-Diaconis' rules provide a default starting point, though sometimes particular class intervals make sense for a particular problem.

The histogram reveals if the distribution of the data is normal, skewed (shifted to the left or right), bimodal (has more than one peak) and so on. Skewed data can sometimes be transformed to normal using a transformation. Bi-modality often indicates that there is more than one underlying population in the data. Individual bars distanced from the bulk of the data can indicate the presence of an outlier.

Creating a histogram

Plot a histogram to visualize the distribution of a quantitative variable.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Distribution** > **Histogram**, and then click the plot type.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the variable.
4. In the **Plot** drop-down list, select the frequency scale.
5. Optional: To label the bars with the frequencies, select the **Label bars** check box.
6. Optional: To show the frequency table, select the **Frequency table** check box.
7. Optional: To customize the classes of the frequency distribution, in the **Start at**, **Width**, and **Classes** edit boxes, type the values to form the classes.
8. Click **Calculate**.

Normality

Normality is the assumption that the underlying random variable is normally distributed, or approximately so.

In some cases, the normality of the data itself may be important in describing a process that generated the data. However, in many cases, it is hypothesis tests and parameter estimators that rely on the assumption of normality, although many are robust against moderate departures in normality due to the central limit theorem.

Normal distribution

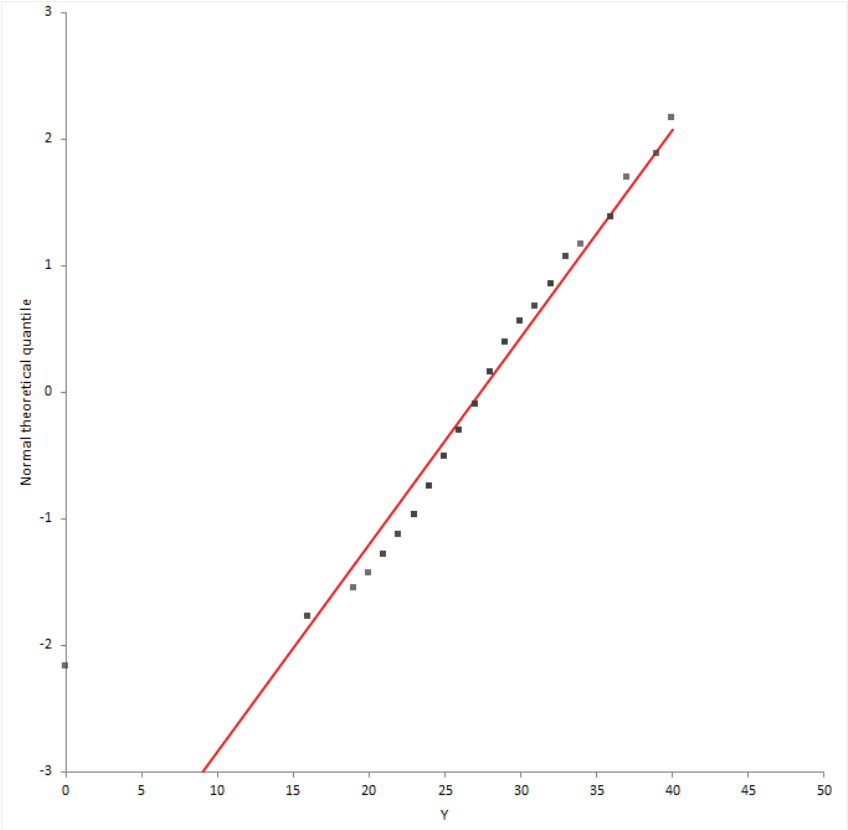
A normal (or Gaussian) distribution is a continuous probability distribution that has a bell-shaped probability density function. It is the most prominent probability distribution in used statistics.

Normal distributions are a family of distributions with the same general symmetric bell-shaped curve, with more values concentrated in the middle than in the tails. Two parameters describe a normal distribution, the mean, and standard deviation. The mean is the central location (the peak), and the standard deviation is the dispersion (the spread). Skewness and excess kurtosis are zero for a normal distribution.

The normal distribution is the basis of much statistical theory. Statistical tests and estimators based on the normal distribution are often more powerful than their non-parametric equivalents. When the distribution assumption can be met they are preferred, as the increased power lets you use a smaller sample size to detect the same difference.

Normal probability (Q-Q) plot

A normal probability plot, or more specifically a quantile-quantile (Q-Q) plot, shows the distribution of the data against the expected normal distribution.



For normally distributed data, observations should lie approximately on a straight line. If the data is non-normal, the points form a curve that deviates markedly from a straight line. Possible outliers are points at the ends of the line, distanced from the bulk of the observations.

Creating a normal probability plot

Plot a Normal (Q-Q) plot to subjectively assess the normality of a quantitative variable.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Distribution** > **Normal**, and then click the plot type.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the variable.
4. Optional: To label the points using the case label, if set, otherwise the observed value, select the **Label points** check box.
5. Optional: To show the normal distribution and confidence band, select the **Normal distribution line** check box, and, if required, in the **Confidence band** edit box, type the level of confidence.
6. Click **Calculate**.

Normality hypothesis test

A hypothesis test formally tests if the population the sample represents is normally-distributed.

The null hypothesis states that the population is normally distributed, against the alternative hypothesis that it is not normally-distributed. If the test p-value is less than the predefined significance level, you can reject the null hypothesis and conclude the data are not from a population with a normal distribution. If the p-value is greater than the predefined significance level, you cannot reject the null hypothesis.

Note that small deviations from normality can produce a statistically significant p-value when the sample size is large, and conversely it can be impossible to detect non-normality with a small sample. You should always examine the normal plot and use your judgment, rather than rely solely on the hypothesis test. Many statistical tests and estimators are robust against moderate departures in normality due to the central limit theorem.

Tests for normality

Tests for normality of a distribution and the type of non-normality detected.

Test	Purpose
Shapiro-Wilk	Test if the distribution is normal. A powerful test that detects most departures from normality when the sample size ≤ 5000.
Anderson-Darling	Test if the distribution is normal. A powerful test that detects most departures from normality.
Kolmogorov-Smirnov	Test if the distribution is normal. Mainly of historical interest and for educational use only.

Testing the normality of a distribution

Test if a distribution is normal.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Distribution**, and then click the hypothesis test.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the variable.
4. In the **Hypotheses** drop-down list, select the null and alternative hypothesis.
5. Optional: To compare the p-value against a predefined significance level, in the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%).
6. Click **Calculate**.

Central limit theorem and the normality assumption

Due to central limit theory, the assumption of normality implied in many statistical tests and estimators is not a problem.

The normal distribution is the basis of much statistical theory. Hypothesis tests and interval estimators based on the normal distribution are often more powerful than their non-parametric equivalents. When the distribution assumption can be met they are preferred as the increased power means a smaller sample size can be used to detect the same difference.

However, violation of the assumption is often not a problem, due to the central limit theorem. The central limit theorem states that the sample means of moderately large samples are often well-

approximated by a normal distribution even if the data are not normally distributed. For many samples, the test statistic often approaches a normal distribution for non-skewed data when the sample size is as small as 30, and for moderately skewed data when the sample size is larger than 100. The downside in such situations is a reduction in statistical power, and there may be more powerful non-parametric tests.

Sometimes a transformation such as a logarithm can remove the skewness and allow you to use powerful tests based on the normality assumption.

Inferences about distribution parameters

Inferences about the parameters of a distribution are made using a random sample of data drawn from the population of interest.

Point and interval estimation

Estimation is the process of making inferences from a sample about an unknown population parameter. An estimator is a statistic that is used to infer the value of an unknown parameter.

A point estimate is the best estimate, in some sense, of the parameter based on a sample. It should be obvious that any point estimate is not absolutely accurate. It is an estimate based on only a single random sample. If repeated random samples were taken from the population, the point estimate would be expected to vary from sample to sample.

A confidence interval is an estimate constructed on the basis that a specified proportion of the confidence intervals include the true parameter in repeated sampling. How frequently the confidence interval contains the parameter is determined by the confidence level. 95% is commonly used and means that in repeated sampling 95% of the confidence intervals include the parameter. 99% is sometimes used when more confidence is needed and means that in repeated sampling 99% of the intervals include the true parameter. It is unusual to use a confidence level of less than 90% as too many intervals would fail to include the parameter. Likewise, confidence levels larger than 99% are not used often because the intervals become wider the higher the confidence level and therefore require large sample sizes to make usable intervals.

Many people misunderstand confidence intervals. A confidence interval does not predict with a given probability that the parameter lies within the interval. The problem arises because the word confidence is misinterpreted as implying probability. In frequentist statistics, probability statements cannot be made about parameters. Parameters are fixed, not random variables, and so a probability statement cannot be made about them. When a confidence interval has been constructed, it either does or does not include the parameter.

In recent years the use of confidence intervals has become more common. A confidence interval provides much more information than just a hypothesis test p-value. It indicates the uncertainty of an estimate of a parameter and allows you to consider the practical importance, rather than just statistical significance.

Confidence intervals and hypothesis tests are closely related. Most introductory textbooks discuss how confidence intervals are equivalent to a hypothesis test. When the 95% confidence interval contains the hypothesized value the hypothesis test is statistically significant at the 5% significance level, and when it does not contain the value the test is not significant. Likewise, with a 99% confidence interval and hypothesis test at the 1% significance level. A confidence interval can be considered as the set of parameter values consistent with the data at some specified level, as assessed by testing each possible value in turn.

Th relationship between hypothesis tests and confidence interval only holds when the estimator and the hypothesis test both use the same underlying evidence function. When a software package uses different evidence functions for the confidence interval and hypothesis test, the results can be inconsistent. The hypothesis test may be statistically significant, but the confidence interval may include the hypothesized value suggesting the result is not significant. Where possible

the same underlying evidence function should be used to form the confidence interval and test the hypotheses. Be aware that not many statistical software packages follow this rule!

The following example illustrates the use of a point estimate, confidence interval and hypothesis test in making inferences.

A scientist might study the difference in blood cholesterol between a new drug treatment and a placebo. Improvements in cholesterol greater than 20mg/dL would be considered practically important, and lead to a change in the treatment of patients, but smaller differences would not. The possible outcomes of the study in terms of a point-estimate, confidence interval estimate, and hypothesis test might be:

Point estimate	**95% confidence interval**	**Hypothesis test (5% significance level)**	**Interpretation**
0 mg/dL	-5 to 5 mg/dL	Not significant	The confidence interval is within the range of no practical difference, 20mg/dL. There is clear evidence the treatment does not produce a difference of practical importance.
10 mg/dL	-5 to 25 mg/dL	Not significant	Part of the confidence interval lies outside the range of no practical difference, 20mg/dL. Although the hypothesis test is not significant, there may be an important practical difference, though a larger sample size is required to make any sharper inferences.
10 mg/dL	5 to 15 mg/dL	Significant	The confidence interval excludes the null (zero) difference but still lies within the range of no practical difference, 20mg/dL. Although the hypothesis test is statistically significant, the difference is of no practical importance.
30 mg/dL	25 to 45 mg/dL	Significant	The confidence interval excludes the null difference and lies outside the range of no practical difference, 20mg/dL. The hypothesis test is statistically significant, and the difference is of practical importance.

Parameter estimate

A parameter estimate is either a point or interval estimate of an unknown population parameter.

A point estimate is a single value that is the best estimate of the true unknown parameter; a confidence interval is a range of values and indicates the uncertainty of the estimate.

Estimators for the central location parameter of a distribution

Estimators for the central location parameter and their properties and assumptions.

Estimator	Purpose
Mean	Estimate the population mean using the sample mean estimator.
Median	Estimate the middle of the distribution using the sample median estimator.
Hodges-Lehmann pseudo-median	Estimate the population pseudo-median using the Hodges-Lehman pseudo-median estimator. The pseudo-median is equivalent to the mean/median when the distribution is symmetric.

Estimators for the dispersion parameter of a distribution

Estimators for the dispersion parameter and their properties and assumptions.

Estimator	Purpose
Standard deviation	Estimate the population standard deviation using the sample standard deviation estimator. The sample standard deviation is a biased estimator of the population standard deviation.
Variance	Estimate the population variance using the sample variance estimator. The sample variance is an unbiased estimator of the population variance.

Estimating the parameters of a distribution

Estimate the mean/median/variance of a distribution.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Distribution**, and then click the parameter estimator.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the variable.
4. In the **Confidence interval** edit box, type the confidence level as a percentage, or type - to suppress the confidence interval, and then in the drop-down list, select the confidence bounds.
5. In the **Method** drop-down list, select the interval estimator.
6. Click **Calculate**.

Hypothesis testing

Hypothesis testing is the formal process of making inferences from a sample whether or not a statement about the population appears to be true.

A hypothesis test is a method of making decisions. You must state a null hypothesis and an alternative hypothesis to perform a hypothesis test. The null hypothesis states what the study is

intending to reject and disprove. The alternative hypothesis is usually the negation of the null and states what the study is trying to prove.

When the hypotheses have been stated a statistical test calculates a test statistic and p-value. The p-value is the probability of obtaining a test statistic at least as extreme as that observed when the null hypothesis is true. It is a measure of evidence against the null hypothesis. When the p-value is small, the data are unlikely to have occurred if the null hypothesis is true so you can reject the null hypothesis and accept the alternative hypothesis. When the p-value is large you cannot reject the null hypothesis; there is insufficient evidence against it. It is not possible to prove the null hypothesis, only disprove it. The p-value does not allow you to make any statements about the probability of the null hypothesis been true, it is a statement based on the observing the data given the null hypothesis is true.

Often a fixed significance level (denoted by the lower case Greek symbol alpha) is used to decide whether the test is statistically significant or not. The significance level is the probability of rejecting the null hypothesis when it is true. When the p-value is less than the significance level, you can declare the test statistically significant. A 5% significance level is typical, which implies there is a 5% chance of wrongly rejecting the null hypothesis when in fact it is true. If more certainty is required, use a 1% significance level. Regardless, you should always report the p-value rather than just a statement of statistically significant, or not.

It is important to remember that a statistically significant test does not imply practically important. The difference might be so small as to be practically useless even though it was statistically significant. Alternatively, the sample size may have been so small that a hypothesis test was not powerful enough to detect anything but a huge difference as statistically significant. It is, therefore, essential that you always interpret the p-value together with a point and interval estimate of the parameter or effect size.

Parameter hypothesis test

A hypothesis test formally tests if a population parameter is different to a hypothesized value.

The null hypothesis states that the parameter is equal to the hypothesized value, against the alternative hypothesis that it is not equal to (or less than, or greater than) the hypothesized value. When the test p-value is small, you can reject the null hypothesis and conclude the sample is not from a population with the parameter equal to the hypothesized value.

Parameter equivalence hypothesis test

An equivalence hypothesis test formally tests if a population parameter is equivalent to a hypothesized value, that is, practically the same.

An typical parameter hypothesis test can never prove that the parameter is equal to the hypothesized value, it can only ever disprove the null hypothesis of equality. By contrast, an equivalence hypothesis test is of interest when the purpose is to prove that the parameter is equivalent to a hypothesized value, that is the difference is less than some small negligible effect size. An equivalence hypothesis test therefore constructs the null hypothesis of non-equivalence and the goal is to prove the parameter is the equivalent to the hypothesized value.

The null hypothesis states that the parameter is not equivalent to the hypothesized value, against the alternative hypothesis that it is equivalent within the equivalence interval. The hypothesis is tested as a composite of two one-sided t-tests (TOST), $H0_1$ tests the hypothesis that parameter is less than the lower bound of the equivalence interval, test $H0_2$ that the parameter is greater than the upper bounds of the equivalence interval. The p-value is the greater of the two one sided t-test p-values. When the test p-value is small, you can reject the null hypothesis and conclude the sample is from a population with the parameter practically equivalent to the hypothesized value.

Tests for the central location parameter of a distribution

Tests for the location parameter of a distribution and their properties and assumptions.

Test	Purpose
Z	Test if the mean is equal to a hypothesized value when the population standard deviation is known.
Student's t	Test if the mean is equal to a hypothesized value. Assumes the population is normally distributed. Due to the central limit theorem, the test may still be useful when this assumption is not true if the sample size is moderate. However, the Wilcoxon test may be more powerful in this situation.
Wilcoxon	Test if the mean/median is equal to a hypothesized value. Assumes the population distribution is symmetric. The mean, median, and pseudo-median are equivalent when the distribution is symmetric. Therefore, the hypotheses can be stated in terms of any of these parameters.
Sign	Test if the median is equal to a hypothesized value. Has few assumptions, but lacks power compared to the Wilcoxon and Student's t test.
TOST (two-one-sided t-test)	Test if the mean is equivalent to a hypothesized value within the equivalence bounds that specify the smallest effect size of interest. Assumes the population is normally distributed. Due to the central limit theorem, the test may still be useful when this assumption is not true if the sample size is moderate.

Tests for the dispersion parameter of a distribution

Tests for the dispersion parameter of a distribution and their properties and assumptions.

Test	Purpose
X^2	Test if the variance/standard deviation is equal to a hypothesized value. Assumes the population is normally distributed. When this assumption is not true, you should be cautious of using the test, as it is extremely sensitive to deviations from normality.

Testing distribution parameters

Test if the mean/median/variance of a distribution is different to the hypothesized value.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Distribution**, and then click the hypothesis test.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the variable.
4. In the **Hypotheses** drop-down list, select the null and alternative hypothesis.
5. In the **Hypothesized value** edit box, type the expected value of the parameter under the null hypothesis.

6. Optional: To compare the p-value against a predefined significance level, in the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%).
7. Click **Calculate**.

Testing equivalence of distribution parameters

Test if the mean of a distribution is equivalent to the hypothesized value.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Distribution**, and then click **TOST**.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the variable.
4. In the **Hypothesized value** edit box, type the expected value of the parameter under the null hypothesis.
5. In the **Hypothesized value** edit box, type the expected value of the parameter, and then in the **Equivalence bounds** edit boxes, type the lower and upper bounds of the equivalence interval. For example, a hypothesized value of 65 with equivalence bounds of ±5. Alternatively, in the **Hypothesized value** edit box, type `-`, and then in the **Equivalence bounds** edit box, type the lower and upper limits of the equivalence interval. For example, 60 to 70.
6. Optional: To compare the p-value against a predefined significance level, in the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%).
7. Click **Calculate**.

Discrete distributions

A discrete distribution describes a variable that can only take discrete values (for example, the number of male and females, or the number of people with a specific eye color).

Frequency distribution

A frequency distribution reduces a large amount of data into a more easily understandable form.

Frequency table

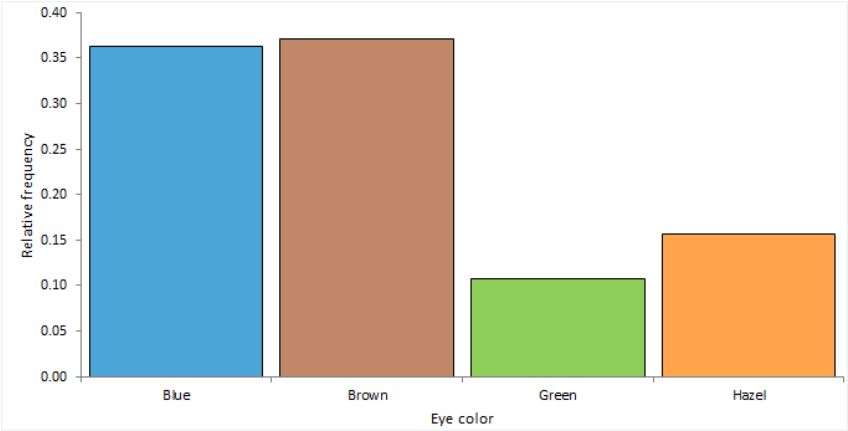
A frequency table is a simple table of the frequencies in each class.

There are many ways of expressing the frequencies:

Statistic	Purpose
Frequency	The number of occurrences in the class.
Relative frequency	The frequency divided by the total.
Frequency density	The relative frequency divided by the width of the class interval. Use with unequal class intervals.
Cumulative frequency	The number of occurrences in the class and all previous classes.
Cumulative relative frequency	The cumulative frequency divided by the total.

Frequency plot

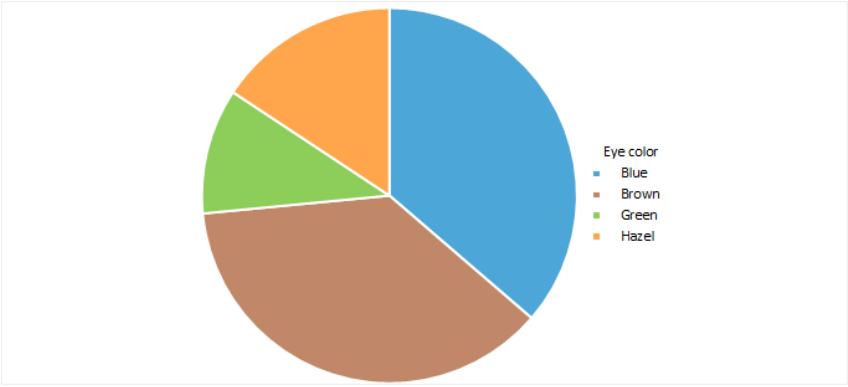
A frequency plot shows the distribution of a qualitative variable.



A frequency plot shows rectangular bars for each class with the height of the bar proportional to the frequency.

Whole-to-part plot

A whole-to-part plot shows how the parts make up the whole.



A pie chart is a common representation which shows a circle divided into sectors for each class, with the angle of each sector proportional to the frequency.

A stacked bar plot shows rectangular bars for each class with the size of each bar proportional to the frequency.

The major problem with the pie chart is that it is difficult to judge the angles and areas of the sectors. It is useful if you want to compare a single class relative to the whole. For comparing classes to each other, we recommend the frequency plot or stacked bar plot for greater clarity and ease of interpretation.

Creating a frequency plot

Plot a frequency plot to visualize the distribution of a qualitative variable.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Distribution** > **Frequency plot**, and then click the plot type.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the categorical variable.
4. If the data are in frequency form, in the **Frequency** drop-down list, select the frequency count variable.
5. Optional: To show the frequency table, select the **Frequency table** check box.
6. Optional: To vary the bar color by the categorical variable, select the **Vary bar color** check box.

   The categories are assigned default colors. To set specific colors for each category, alongside the Y variable, click the drop-down arrow, and then click **Colors / Symbols...**.
7. Optional: To label the bars with the frequencies, select the **Label bars** check box.
8. Optional: To overlay the cumulative frequency distribution line, select the **Cumulative frequency line** check box.
9. Click **Calculate**.

Creating a whole-to-part plot

Plot a pie chart or bar chart to visualize the relative frequencies.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Distribution** > **Whole-to-Part**, and then click the plot type.

The analysis task pane opens.

3. In the **Y** drop-down list, select the categorical variable.
4. If the data are in frequency form, in the **Frequency** drop-down list, select the frequency count variable.
5. Optional: To show the frequency table, select the **Frequency table** check box.
6. Optional: To label the bars/sectors with the frequencies, select the **Label bars/Label sectors** check box.
7. Click **Calculate**.

Inferences about Binomial distribution parameters

Inferences about the parameters of a binomial distribution are made using a random sample of data drawn from the population of interest.

A binomial distribution arises when an experiment consists of a fixed number of repeated trials; each trial has two possible outcomes; the probability of the outcome is the same for each trial; and the trials are independent, that is, the outcome of one trial does not affect the outcome of other trials.

Binomial distribution parameter estimate

A parameter estimate is either a point or interval estimate of an unknown population parameter.

A point estimate is a single value that is the best estimate of the true unknown parameter; a confidence interval is a range of values and indicates the uncertainty of the estimate.

Estimators for the parameter of a Binomial distribution

Estimators for the binomial distribution parameter and their properties and assumptions.

Estimator	Purpose
Proportion	Estimate the population proportion of occurrences of the outcome of interest using the sample proportion estimator.
Odds	Estimate the population odds of the outcome of interest occurring using the sample odds estimator. Odds is an expression of the relative probabilities in favor of an event.

Binomial distribution parameter hypothesis test

A hypothesis test formally tests if a population parameter is equal to a hypothesized value. For a binomial distribution, the parameter is the probability of success, commonly referred to as the proportion with the outcome of interest.

The null hypothesis states that the proportion is equal to the hypothesized value, against the alternative hypothesis that it is not equal to (or less than, or greater than) the hypothesized value. When the test p-value is small, you can reject the null hypothesis and conclude the sample is not from a population with the proportion equal to the hypothesized value.

Tests for the parameter of a Binomial distribution

Tests for the parameter of a binomial distribution and their properties and assumptions.

Test	Purpose
Binomial exact	Test if the proportion with the outcome of interest is equal to a hypothesized value. Uses the binomial distribution and computes an exact p-value. The test is conservative, that is, the type I error is guaranteed to be less than or equal to the desired significance level. Recommended for small sample sizes.
Score Z	Test if the proportion with the outcome of interest is equal to a hypothesized value. Uses the score statistic and computes an asymptotic p-value. Equivalent to Pearson's X^2 test. Recommended for general use.

Testing Binomial distribution parameters

Test if the parameter of a binomial distribution is equal to the hypothesized value.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Distribution**, and then click the hypothesis test.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the categorical variable.
4. If the data are in frequency form, in the **Frequency** drop-down list, select the frequency count variable.
5. In the **Hypotheses** drop-down list, select the null and alternative hypothesis.
6. In the **Hypothesized value** edit box, type the expected value of the parameter under the null hypothesis.
7. Optional: To compare the p-value against a predefined significance level, in the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%).
8. Click **Calculate**.

Inferences about Multinomial distribution parameters

Inferences about the parameters of a multinomial distribution are made using a random sample of data drawn from the population of interest.

A multinomial distribution arises when an experiment consists of a fixed number of repeated trials; each trial has a discrete number of possible outcomes; the probability that a particular outcome occurs is the same for each trial; and the trials are independent, that is, the outcome of one trial does not affect the outcome of other trials.

Multinomial distribution parameters hypothesis test

A hypothesis test formally tests if the population parameters are different from the hypothesized values. For a multinomial distribution, the parameters are the proportions of occurrence of each outcome.

The null hypothesis states that the proportions equal the hypothesized values, against the alternative hypothesis that at least one of the proportions is not equal to its hypothesized value. When the test p-value is small, you can reject the null hypothesis and conclude that at least one proportion is not equal to its hypothesized value.

The test is an omnibus test and does not tell you which proportions differ from the hypothesized values.

Tests for the parameters of a Multinomial distribution

Tests for the parameters of a multinomial distribution and their properties and assumptions.

Test	Purpose
Pearson X^2	Tests if the proportions are equal to the hypothesized values. Uses the score statistic and computes an asymptotic p-value.
Likelihood ratio G^2	Tests if the proportions are equal to the hypothesized values. Uses the likelihood ratio statistic and computes an asymptotic p-value. Pearson X^2 usually converges to the chi-squared distribution more quickly than G^2. The likelihood ratio test is commonly used in statistical modeling as the G^2 statistic is easier to compare between different models.

Testing Multinomial distribution parameters

Test if the parameters of a multinomial distribution are equal to the hypothesized values.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Distribution**, and then click the hypothesis test.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the categorical variable.
4. If the data are in frequency form, in the **Frequency** drop-down list, select the frequency count variable.
5. In the **Hypotheses** drop-down list, select the null and alternative hypothesis.
6. In the **Hypothesized values** group, select:
   - **Are all equal** to set the hypothesized values equal for each category.
   - **Are equal to specified**, and then in the grid, under the **Value** column alongside each category, type values as either probabilities/proportions or number of occurrences.
7. Optional: To compare the p-value against a predefined significance level, in the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%).
8. Click **Calculate**.

Study design

Distribution analysis study requirements and dataset layout.

Requirements

- A categorical or quantitative variable.

Dataset layout

Use a column for each variable (Height, Eye color); each row has the values of the variables for a case (Subject).

Subject (optional)	Height	Eye color
1	175	Blue
2	180	Blue
3	160	Hazel
4	190	Green
5	180	Green
6	150	Brown
7	140	Blue
8	160	Brown
9	165	Green
10	180	Hazel
…	…	…

Frequency form dataset layout

Use a column for the variable (Eye color) and a column for the number of cases (Frequency); each row has the values of the variables and the frequency count.

Eye color	Frequency
Brown	221
Blue	215
Hazel	93
Green	64

Chapter

Compare groups

10

Compare Groups examines independent samples and makes inferences about the differences between them.

Independent samples occur when observations are made on different sets of items or subjects. If the values in one sample do not tell you anything about the values in the other sample, then the samples are independent. If the knowing the values in one sample could tell you something about the values in the other sample, then the samples are related.

Calculating univariate descriptive statistics, by group

Use univariate descriptive statistics to describe a quantitative variable that is classified by a factor variable.

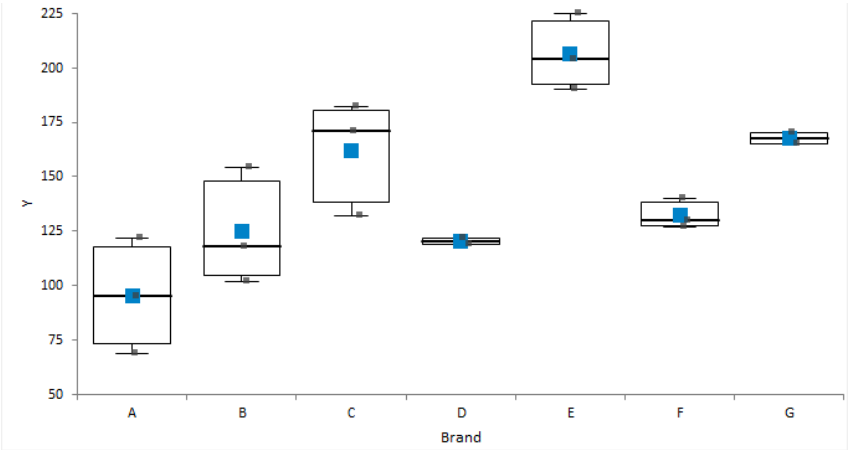
1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Compare Groups**, and then click the statistics to show:

Mean and Moments	Show mean, variance, standard deviation.
Median and Quartiles	Show minimum, maximum, median, and quartiles.

   The analysis task pane opens.
3. In the **Y** drop-down list, select the quantitative response variable.
4. In the **X** drop-down list, select the categorical factor variable identifying the groups.
5. Optional: To customize the statistics to show, click **Customize...** and then select or clear the appropriate check boxes. To save the options as the defaults to use for future analyses, click **Save as Defaults**.
6. Click **Calculate**.

Side-by-side univariate plots

Side-by-side univariate plots summarize the distribution of data stratified into groups.



Creating side-by-side univariate plots

Summarize the distribution of a single quantitative variable stratified into groups.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Compare Groups** > **Univariate**, and then click the plot type.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the quantitative response variable.
4. In the **X** drop-down list, select the categorical factor variable identifying the groups.
5. Optional: To overlay more than one plot type on the same chart, on the **Analyse-it** ribbon tab, in the **Compare Groups** group, click **Univariate**, and then click an additional plot type.
6. Optional: In the analysis task pane, customize the options for the plot.
7. Click **Calculate**.

Equality of means/medians hypothesis test

An equality hypothesis test formally tests if two or more population means/medians are different.

The hypotheses to test depends on the number of samples:

- For two samples, the null hypothesis states that the difference between the mean/medians of the populations is equal to a hypothesized value (0 indicating no difference), against the alternative hypothesis that it is not equal to (or less than, or greater than) the hypothesized value.
- For more than two samples, the null hypothesis states that the means/medians of the populations are equal, against the alternative hypothesis that at least one population mean/median is different.

When the test p-value is small, you can reject the null hypothesis and conclude that the populations differ in means/medians.

Tests for more than two samples are omnibus tests and do not tell you which groups differ from each other. You should use multiple comparisons to make these inferences.

Equivalence of means hypothesis test

An equivalence hypothesis test formally tests if two population means are equivalent, that is, practically the same.

An equality hypothesis test can never prove that the means are equal, it can only ever disprove the null hypothesis of equality. It is therefore of interest when comparing say a new treatment against a placebo, where the null hypothesis (assumption of what is true without evidence to the contrary) is that the treatment has no effect, and you want to prove the treatment produces a useful effect. By contrast, an equivalence hypothesis test is of interest when comparing say a generic treatment to an existing treatment where the aim is to prove that they are equivalent, that is the difference is less than some small negligible effect size. A equivalence hypothesis test therefore constructs the null hypothesis of non-equivalence and the goal is to prove the means are equivalent.

The null hypothesis states that the means are not equivalent, against the alternative hypothesis that the difference between the means is within the bounds of the equivalence interval, that is, the effect size is less than some small difference that is considered practically zero. The hypothesis is tested as a composite of two one-sided t-tests (TOST), $H0_1$ tests the hypothesis that mean difference is less than the lower bound of the equivalence interval, test $H0_2$ that the mean difference is greater than the upper bounds of the equivalence interval. The p-value is the greater of the two one sided t-test p-values. When the test p-value is small, you can reject the null hypothesis and conclude the samples are from populations with practically equivalent means.

Tests for means/medians

Tests for the equality/equivalence of means/medians of independent samples and their properties and assumptions.

Test	Purpose
Z	Test if the difference between means is equal to a hypothesized value when the population standard deviation is known.
Student's t	Test if the difference between means is equal to a hypothesized value. Assumes the populations are normally distributed. Due to the central limit theorem, the test may still be useful when this assumption is not true if the sample sizes are equal, moderate size, and the distributions have a similar shape. However, in this situation the Wilcoxon-Mann-Whitney test may be more powerful. Assumes the population variances are equal. This assumption can be tested using the Levene test. The test may still be useful when this assumption is not true if the sample sizes are equal. However, in this situation, the Welch t-test may be preferred.
Welch t	Test if the difference between means is equal to a hypothesized value. Assumes the populations are normally distributed. Due to the central limit theorem, the test may still be useful when this assumption is not true if the sample sizes are equal, moderate size, and the distributions have a similar shape. However, in this situation the Wilcoxon-Mann-Whitney test may be more powerful. Does not assume the population variances are equal.
TOST (two-one-sided t-tests)	Test if the means are equivalent. Assumes the populations are normally distributed. Due to the central limit theorem, the test may still be useful when this assumption is not

Test	Purpose
	true if the sample sizes are equal, moderate size, and the distributions have a similar shape. **Note:** The TOST can be performed using either a Student's t-test or Welch t-test.
ANOVA	Test if two or means are equal. Assumes the populations are normally distributed. Due to the central limit theorem, the test may still be useful when this assumption is not true if the sample sizes are equal and moderate size. However, in this situation the Kruskal-Wallis test is may be more powerful. Assumes the population variances are equal. This assumption can be tested using the Levene test. The test may still be useful when this assumption is not true if the sample sizes are equal. However, in this situation, the Welch ANOVA may be preferred.
Welch ANOVA	Test if two or more means are equal. Assumes the populations are normally distributed. Due to the central limit theorem, the test may still be useful when this assumption is not true when the sample sizes are equal and moderate size. The Kruskal-Wallis test may be preferable as it is more powerful than Welch's ANOVA. Does not assume the population variances are equal.
Wilcoxon-Mann-Whitney	Test if there is a shift in location. When the population distributions are identically shaped, except for a possible shift in central location, the hypotheses can be stated in terms of a difference between means/medians. When the population distributions are not identically shaped, the hypotheses can be stated as a test of whether the samples come from populations such that the probability is 0.5 that a random observation from one group is greater than a random observation from another group.
Kruskal-Wallis	Test if two or more medians are equal. Assumes the population distributions are identically shaped, except for a possible shift in central location.

Testing equality of means/medians

Test if there is a difference between the means/medians of two or more independent samples.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Compare Groups**, and then click the hypothesis test.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the quantitative response variable.
4. In the **X** drop-down list, select the categorical factor variable identifying the groups.
5. If performing a 2 sample hypothesis test:
   a) In the **Hypotheses** drop-down list, select the null and directional alternative hypotheses.

b) In the **Hypothesized difference** edit box, type the expected difference under the null hypothesis.

6. If performing a k sample hypothesis test:
   a) In the **Hypothesis** drop-down list, select the null and alternative hypotheses.
7. Optional: To compare the p-value against a predefined significance level, in the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%).
8. Click **Calculate**.

Testing equivalence of means

Test if there is equivalence of the means of two independent samples.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Compare Groups**, and then click **TOST** hypothesis test.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the quantitative response variable.
4. In the **X** drop-down list, select the categorical factor variable identifying the groups.
5. In the **Test** drop-down list, select either **Student's t** or **Welch t** depending on the homogeneity of variances.

   The TOST equivalence test is actually specified through the Hypotheses drop-down list, the Test drop-down list indicates the type of t-test to use.
6. In the **Equivalence interval** edit boxes, type the lower and upper bounds of the equivalence interval. For example, type-5 to +5 if differences between the means of ±5 is the effect size considered to be practically the same.
7. Optional: To compare the p-value against a predefined significance level, in the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%).
8. Click **Calculate**.

Difference between means/medians effect size

An effect size estimates the magnitude of the difference between two means/medians.

The term effect size relates to both unstandardized measures (for example, the difference between group means) and standardized measures (such as Cohen's d; the standardized difference between group means). Standardized measures are more appropriate than unstandardized measures when the metrics of the variables do not have intrinsic meaning, when combining results from multiple studies, or when comparing results from studies with different measurement scales.

A point estimate is a single value that is the best estimate of the true unknown parameter; a confidence interval is a range of values and indicates the uncertainty of the estimate.

Estimators for the difference in means/medians

Estimators for the difference in means/medians of independent samples and their properties and assumptions.

Estimator	Purpose
Mean difference	Estimate the difference between the means.
Standardized mean difference	Estimate the standardized difference between the means.

Estimator	Purpose
	Cohen's d is the most popular estimator using the difference between the means divided by the pooled sample standard deviation. Cohen's d is a biased estimator of the population standardized mean difference (although the bias is very small and disappears for moderate to large samples) whereas Hedge's g applies an unbiasing constant to correct for the bias.
Hodges-Lehmann location shift	Estimate the shift in location. A shift in location is equivalent to a difference between means/medians when the distributions are identically shaped.

Estimating the difference between means/medians

Estimate the difference between the means/medians of 2 independent samples.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Compare Groups**, and then click the effect size estimator.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the quantitative response variable.
4. In the **X** drop-down list, select the categorical factor variable identifying the groups.
5. In the **Difference** drop-down list, select the direction of the difference.
6. In the **Confidence interval** edit box, type the confidence level as a percentage, or type - to suppress the confidence interval, and then in the drop-down list, select the confidence bounds.
7. In the **Method** drop-down list, select the interval estimator.
8. Click **Calculate**.

Multiple comparisons

Multiple comparisons make simultaneous inferences about a set of parameters.

When making inferences about more than one parameter (such as comparing many means, or the differences between many means), you must use multiple comparison procedures to make inferences about the parameters of interest. The problem when making multiple comparisons using individual tests such as Student's t-test applied to each comparison is the chance of a type I error increases with the number of comparisons. If you use a 5% significance level with a hypothesis test to decide if two groups are significantly different, there is a 5% probability of observing a significant difference that is simply due to chance (a type I error). If you made 20 such comparisons, the probability that one or more of the comparisons is statistically significant simply due to chance increases to 64%. With 50 comparisons, the chance increases to 92%. Another problem is the dependencies among the parameters of interest also alter the significance level. Therefore, you must use multiple comparison procedures to maintain the simultaneous probability close to the nominal significance level (typically 5%).

Multiple comparison procedures are classified by the strength of inference that can be made and the error rate controlled. A test of homogeneity controls the probability of falsely declaring any pair to be different when in fact all are the same. A stronger level of inference is confident inequalities and confident directions which control the probability of falsely declaring any pair to be different regardless of the values of the others. An even stronger level is a set of simultaneous confidence intervals that guarantees that the simultaneous coverage probability of the intervals is at least 100(1-alpha)% and also have the advantage of quantifying the possible effect size rather than producing just a p-value. A higher strength inference can be used to make an inference of a lesser strength but not vice-versa. Therefore, a confidence interval can be used to perform a

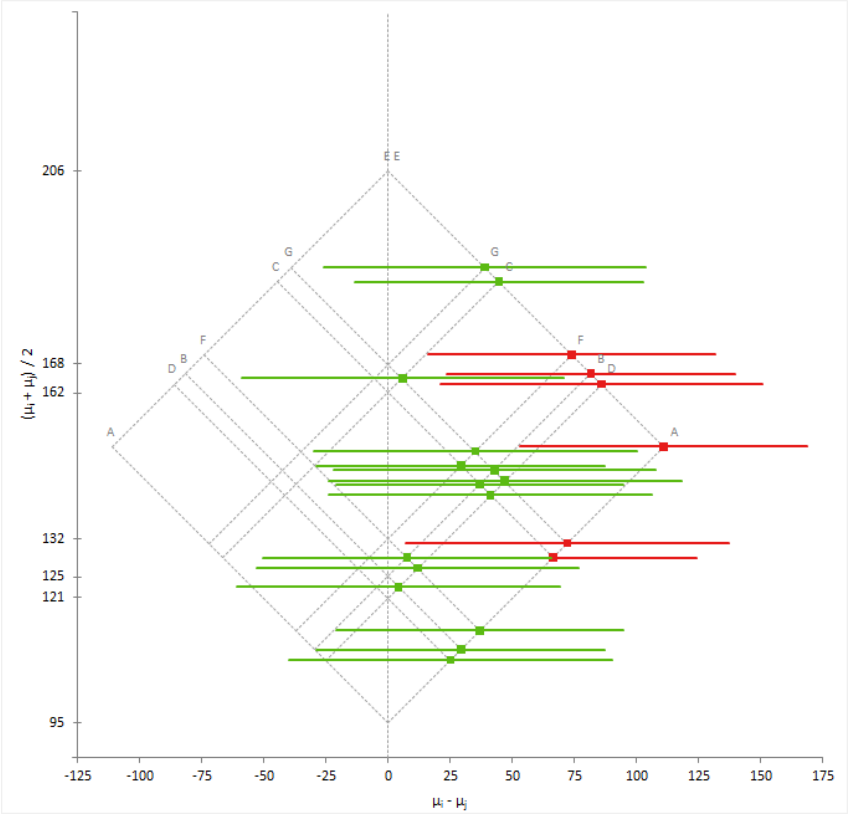
confidence directions/inequalities inference, but a test of homogeneity cannot make a confidence direction/inequalities inference.

The most well known multiple comparison procedures, Bonferroni and Šidák, are not multiple comparison procedures per se. Rather they are an inequality useful in producing easy to compute multiple comparison methods of various types. In most scenarios, there are more powerful procedures available. A useful application of Bonferroni inequality is when there are a small number of pre-planned comparisons. In these cases, you can use the standard hypothesis test or confidence interval with the significance level (alpha) set to the Bonferroni inequality (alpha divided by the number of comparisons).

A side effect of maintaining the significance level is a lowering of the power of the test. Different procedures have been developed to maintain the power as high as possible depending on the strength of inference required and the number of comparisons to be made. All contrasts comparisons allow for any possible contrast; all pairs forms the k*(k-1)/2 pairwise contrasts, whereas with best forms k contrasts each with the best of the others, and against control forms k-1 contrasts each against the control group. You should choose the appropriate contrasts of interest before you perform the analysis, if you decide after inspecting the data, then you should only use all contrasts comparison procedures.

Mean-Mean scatter plot

A mean-mean scatter plot shows a 2-dimensional representation of the differences between many means.



The mean-mean scatter plot shows the mean of a group on the horizontal axis against the mean of the other group on the vertical axis with a dot at the intersection. A vector centered at the intersection with a slope of -1 and a length proportional to the width of the confidence interval represents the confidence interval. A gray identity line represents equality of means; that is the

difference is equal to zero. If the vector does not cross the identity line, you can conclude there is a significant difference between the means.

To make interpretation easier, a 45-degree rotated version of the plot shows the difference between means and its confidence interval on the horizontal axis against average of the means on the vertical axis.

Multiple comparison procedures

Methods of controlling the Type I error and dependencies between parameters when making multiple comparisons between many means/medians of independent samples.

Procedure	Purpose
Student's t (Fisher's LSD)	Compare the means of each pair of groups using the Student's t method. When making all pairwise comparisons this procedure is also known as unprotected Fisher's LSD, or when only performed following significant ANOVA F -test known as protected Fisher's LSD. Control the type I error rate is for each contrast.
Wilcoxon-Mann-Whitney	Compare the median/means each pair of groups using the Wilcoxon nonparametric method. Control the type I error rate for each contrast.
Tukey-Kramer	Compare the means of all pairs of groups using the Tukey-Kramer method. Controls the error rate simultaneously for all k(k+1)/2 contrasts.
Steel-Dwass-Critchlow-Fligner	Compare the median/means of all pairs of groups using the Steel-Dwass-Critchlow-Fligner pairwise ranking nonparametric method. Controls the error rate simultaneously for all k(k+1)/2 contrasts.
Hsu	Compare the means of all groups against the best of the other groups using the Hsu method. Controls the error rate simultaneously for all k contrasts.
Dunnett	Compare the means of all groups against a control using the Dunnett method. Controls the error rate simultaneously for all k-1 contrasts.
Steel	Compare the medians/means of all groups against a control using the Steel pairwise ranking nonparametric method. Controls the error rate simultaneously for all k-1 comparisons.
Scheffé	Compare the means of all groups against all other groups using the Scheffé F method. Controls the error rate simultaneously for all possible contrasts.
Student-Newman-Keuls (SNK)	Not implemented. As discussed in *Hsu (1996)* it is not a confident inequalities method and cannot be recommended.

Procedure	Purpose
Duncan	Not implemented. As discussed in *Hsu (1996)* it is not a confident inequalities method and cannot be recommended.
Dunn	Not implemented in Analyse-it version 3 onwards. Implemented in version 1. A joint-ranking nonparametric method but as discussed in *Hsu (1996)* it is not a confident inequalities method and cannot be recommended. Use the Dwass-Steel-Critchlow-Fligner or Steel test instead.
Conover-Iman	Not implemented in Analyse-it version 3 onwards. Implemented in version 2. As discussed in *Hsu (1996)* applying parametric methods to the rank-transformed data may not produce a confident inequalities method even if the parametric equivalent it mimics is. Use the Dwass-Steel-Critchlow-Fligner or Steel test instead.
Bonferroni	Not a multiple comparisons method. It is an inequality useful in producing easy to compute multiple comparisons. In most scenarios, there are more powerful procedures such as Tukey, Dunnett, Hsu. A useful application of Bonferroni inequality is when there are a small number of pre-planned comparisons. In this case, use the Student's t (LSD) method with the significance level (alpha) set to the Bonferroni inequality (alpha divided by the number of comparisons). In this scenario, it is usually less conservative than using Scheffé all contrast comparisons.

Comparing multiple means/medians

Compare the means/medians of many independent samples.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Compare Groups**, and then click the multiple comparisons procedure.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the quantitative response variable.
4. In the **X** drop-down list, select the categorical factor variable identifying the groups.
5. In the **Type** drop-down list, select the comparisons, and then:
   - If comparing with best, in the **Best** drop-down list, select the "best" group.
   - If comparing against a control, in the **Control** drop-down list, select the control group.
6. Optional: To sort the contrasts by the magnitude of the difference, select the **Ordered** check box.
7. Optional: In the **Method** drop-down list, select the multiple comparisons procedure, and in the **Confidence interval** edit box, type the confidence level as a percentage, or type - to suppress the confidence interval, and then in the **Confidence bounds** drop-down list, select the limits.
8. Optional: To perform a hypothesis test to test if the difference is equal to 0, select the **Hypothesis test** check box, and then to compare the p-value, in the **Significance level** edit box, type the predefined significance level.
9. Optional: To show a mean-mean scatter plot, select the **Mean-Mean** scatter plot check box.

10. Click **Calculate**.

Homogeneity of variance hypothesis test

A homogeneity hypothesis test formally tests if the populations have equal variances.

Many statistical hypothesis tests and estimators of effect size assume that the variances of the populations are equal. This assumption allows the variances of each group to be pooled together to provide a better estimate of the population variance. A better estimate of the variance increases the statistical power of the test meaning you can use a smaller sample size to detect the same difference, or detect smaller differences and make sharper inferences with the same sample size.

The hypotheses to test depends on the number of samples:

- For two samples, the null hypothesis states that the ratio of the variances of the populations is equal to a hypothesized value (usually 1 indicating equality), against the alternative hypothesis that it is not equal to (or less than, or greater than) the hypothesized value.
- For more than two samples, the null hypothesis states that the variances of the populations are equal, against the alternative hypothesis that at least one population variance is different.

When the test p-value is small, you can reject the null hypothesis and conclude that the populations differ in variance.

Tests for homogeneity of variance

Tests for the homogeneity of variance of independent samples and their properties and assumptions.

Test	Purpose
F	Test if the ratio of the variances is equal to a hypothesized value. Assumes the populations are normally distributed and is extremely sensitive to departures from normality. The Levene and Brown-Forsythe tests are more robust against violations of the normality assumption than the F-test.
Bartlett	Test if the variances are equal. Assumes the populations are normally distributed and is extremely sensitive to departures from normality. The Levene and Brown-Forsythe tests are more robust against violations of the normality assumption than Bartlett's test.
Levene	Test if the variances are equal. Most powerful test for symmetric, moderate-tailed, distributions.
Brown-Forsythe	Test if the variances are equal. Robust against many types of non-normality.

Testing homogeneity of variance

Test if the variances of two or more independent samples are equal.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Compare Groups**, and then click the hypothesis test.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the quantitative response variable.

4. In the **X** drop-down list, select the categorical factor variable identifying the groups.
5. In the **Hypotheses** drop-down list, select the null and alternative hypothesis.
6. Optional: To compare the p-value against a predefined significance level, in the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%).
7. Click **Calculate**.

Study design

Compare groups analysis study requirements and dataset layout.

Requirements

- A quantitative response variable.
- A categorical factor variable with 2 or more levels.

Dataset layout

Use a column for the response variable (Height) and a column for the factor variable (Sex); each row has the values of the variables for a case (Subject).

Subject (optional)	Sex	Height
1	Male	175
2	Male	180
3	Male	160
4	Male	190
5	Male	180
6	Female	150
7	Female	140
8	Female	160
9	Female	165
10	Female	180
…	…	…

Note: Do not split groups into separate columns. For example, you should not record the heights of males in one column and heights of females in another.

Chapter

Compare pairs

11

Compare Pairs examines related samples and makes inferences about the differences between them.

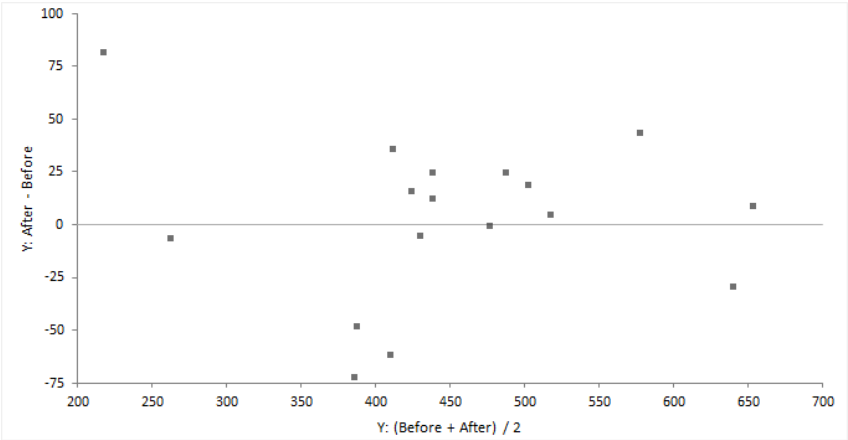
Related samples occur when observations are made on the same set of items or subjects at different times, or when another form of matching has occurred. If the knowing the values in one sample could tell you something about the values in the other sample, then the samples are related.

There are a few different study designs that produce related data. A paired study design takes individual observations from a pair of related subjects. A repeat measures study design takes multiple observations on the same subject. A matched pair study design takes individual observations on multiple subjects that are matched on other covariates. The purpose of matching similar subjects is often to reduce or eliminate the effects of a confounding factor.

Difference plot

A difference plot shows the differences between two observations on the same sampling unit.

The difference plot shows the difference between two observations on the vertical axis against the average of the two observations on the horizontal axis. A gray identity line represents equality; no difference.



If the second observation is always greater than the first the points lie above the line of equality, or vice-versa. If differences are not related to the magnitude the points will form a horizontal band. If the points form an increasing, decreasing, or non-constant width band, then the variance is not constant.

It is common to combine the difference plot with a histogram and a normality plot of the differences to check if the differences are normally distributed, which is an assumption of some statistical tests and estimators.

Creating a Tukey mean-difference plot

Summarize the differences between two related observations.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Compare Pairs**, and then click **Difference**.
   The analysis task pane opens.
3. If the data consists of paired/repeated measurements in separate variables:
   a) In the **Model** drop-down menu, select **Paired/Repeated**.
   b) In the **Y** list, select the quantitative variables.
4. If the data consist of matched pairs with a separate response variable, factor variable, and blocking variable:
   a) In the **Model** drop-down menu, select **Matched**.
   b) In the **Y** drop-down list, select the quantitative response variable.
   c) In the **X (factor)** drop-down list, select the categorical factor variable identifying the groups.
   d) In the **Block** drop-down list, select the blocking variable identifying the matching.
5. Optional: To show a histogram of the distribution of the differences, select the **Histogram** check box.
6. Optional: To show a normality plot of the differences, select the **Normal plot** check box.
7. Click **Calculate**.

Equality of means/medians hypothesis test

An equality hypothesis test formally tests if two or more population means/medians are different.

Inferences about related samples are complicated by the fact that the observations are correlated. Therefore the tests for independent samples are of no use. Instead, tests for related samples focus on the differences within each sampling unit.

The hypotheses to test depends on the number of samples:

- For two samples, the null hypothesis states that the difference between the mean/medians of the populations is equal to a hypothesized value (0 indicating no difference), against the alternative hypothesis that it is not equal to (or less than, or greater than) the hypothesized value.
- For more than two samples, the null hypothesis states that the means/medians of the populations are equal, against the alternative hypothesis that at least one population mean/median is different.

When the test p-value is small, you can reject the null hypothesis and conclude that the populations differ in means/medians.

Equivalence of means hypothesis test

An equivalence hypothesis test formally tests if two population means are equivalent, that is, practically the same.

An equality hypothesis test can never prove that the means are equal, it can only ever disprove the null hypothesis of equality. It is therefore of interest when comparing say a new treatment against a placebo, where the null hypothesis (assumption of what is true without evidence to the contrary) is that the treatment has no effect, and you want to prove the treatment produces a useful effect. By contrast, an equivalence hypothesis test is of interest when comparing say a generic treatment to an existing treatment where the aim is to prove that they are equivalent, that is the difference is

less than some small negligible effect size. A equivalence hypothesis test therefore constructs the null hypothesis of non-equivalence and the goal is to prove the means are equivalent.

The null hypothesis states that the means are not equivalent, against the alternative hypothesis that the difference between the means is within the bounds of the equivalence interval, that is, the effect size is less than some small difference that is considered practically zero. The hypothesis is tested as a composite of two one-sided t-tests (TOST), $H0_1$ tests the hypothesis that mean difference is less than the lower bound of the equivalence interval, test $H0_2$ that the mean difference is greater than the upper bounds of the equivalence interval. The p-value is the greater of the two one-sided t-test p-values. When the test p-value is small, you can reject the null hypothesis and conclude the samples are from populations with practically equivalent means.

Tests for means/medians

Tests for the equality of means/medians of related samples and their properties and assumptions.

Test	Purpose
Z	Test if the difference between means is equal to a hypothesized value when the population standard deviation is known. Assumes the population differences are normally distributed. Due to the central limit theorem, the test may still be useful when this assumption is not true if the sample size is moderate. However, in this case, the Wilcoxon test may be more powerful.
Student's t	Test if the difference between means is equal to a hypothesized value. Assumes the population differences are normally distributed. Due to the central limit theorem, the test may still be useful when this assumption is not true if the sample size is moderate. However, in this case, the Wilcoxon test may be more powerful.
Wilcoxon	Test if there is a shift in location equal to the hypothesized value. Under the assumption that the population distribution of the differences is symmetric, the hypotheses can be stated in terms of a difference between means/medians. Under the less strict hypotheses, requiring no distributional assumptions, the hypotheses can be stated as the probability that the sum of a randomly chosen pair of differences exceeds zero is 0.5.
Sign	Test if the median of the differences is equal to a hypothesized value. Under the more general hypotheses, tests if given a random pair of observations (xi, yi), that xi and yi are equally likely to be larger than the other. Has few assumptions, but lacks power compared to the Wilcoxon and Student's t test.
TOST (two-one-sided t-tests)	Test if the means are equivalent. Assumes the populations are normally distributed. Due to the central limit theorem, the test may still be useful when this assumption is not true if the sample sizes are equal, moderate size, and the distributions have a similar shape.
ANOVA	Test if the two or means are equal.

Test	Purpose
	Assumes the populations are normally distributed. Due to the central limit theorem, the test may still be useful when the assumption is violated if the sample sizes are equal and moderate size. However, in this situation the Friedman test is may be more powerful.
Friedman	Test if two or medians are equal. Has few assumptions, and is equivalent to a two-sided Sign test in the case of two samples.

Testing equality of means/medians

Test if there is a difference between the means/medians of two or more related samples.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Compare Pairs**, and then click the hypothesis test.
   The analysis task pane opens.
3. If the data consists of paired/repeated measurements in separate variables:
   a) In the **Model** drop-down menu, select **Paired/Repeated**.
   b) In the **Y** list, select the quantitative variables.
4. If the data consist of matched pairs with a separate response variable, factor variable, and blocking variable:
   a) In the **Model** drop-down menu, select **Matched**.
   b) In the **Y** drop-down list, select the quantitative response variable.
   c) In the **X (factor)** drop-down list, select the categorical factor variable identifying the groups.
   d) In the **Block** drop-down list, select the blocking variable identifying the matching.
5. If performing a 2 sample hypothesis test:
   a) In the **Hypotheses** drop-down list, select the null and directional alternative hypotheses.
   b) In the **Hypothesized difference** edit box, type the expected difference under the null hypothesis.
6. If performing a k sample hypothesis test:
   a) In the **Hypothesis** drop-down list, select the null and alternative hypotheses.
7. Optional: To compare the p-value against a predefined significance level, in the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%).
8. Click **Calculate**.

Testing equivalence of means

Test if there is equivalence of means of two related samples.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Compare Pairs**, and then click **TOST** hypothesis test.
   The analysis task pane opens.
3. If the data consists of paired/repeated measurements in separate variables:
   a) In the **Model** drop-down menu, select **Paired/Repeated**.
   b) In the **Y** list, select the quantitative variables.

4. If the data consist of matched pairs with a separate response variable, factor variable, and blocking variable:
   a) In the **Model** drop-down menu, select **Matched**.
   b) In the **Y** drop-down list, select the quantitative response variable.
   c) In the **X (factor)** drop-down list, select the categorical factor variable identifying the groups.
   d) In the **Block** drop-down list, select the blocking variable identifying the matching.
5. In the **Equivalence interval** edit boxes, type the lower and upper bounds of the equivalence interval. For example, type-5 to +5 if differences between the means of ±5 is the effect size considered to be practically the same.
6. Optional: To compare the p-value against a predefined significance level, in the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%).
7. Click **Calculate**.

Difference between means/medians effect size

An effect size estimates the magnitude of the difference between two means/medians.

The term effect size relates to both unstandardized measures (for example, the difference between group means) and standardized measures (such as Cohen's d; the standardized difference between group means). Standardized measures are more appropriate than unstandardized measures when the metrics of the variables do not have intrinsic meaning, when combining results from multiple studies, or when comparing results from studies with different measurement scales.

A point estimate is a single value that is the best estimate of the true unknown parameter; a confidence interval is a range of values and indicates the uncertainty of the estimate.

Estimators for the difference in means/medians

Estimators for the difference in means/medians of related samples and their properties and assumptions.

Estimator	Purpose
Mean difference	Estimate the difference between the means.
Standardized mean difference	Estimate the standardized difference between the means. Cohen's d is the most popular estimator using the difference between the means divided by the pooled sample standard deviation. Cohen's d is a biased estimator of the population standardized mean difference (although the bias is very small and disappears for moderate to large samples) whereas Hedge's g applies an unbiasing constant to correct for the bias.
Median difference	Estimate the median of the differences.
Hodges-Lehmann location shift	Estimate the shift in location. A shift in location is equivalent to a difference between means/medians when the distribution of the differences is symmetric. Symmetry is quite often inherently satisfied for paired data.

Estimating the difference between means/medians

Estimate the difference between the means/medians of 2 related samples.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Compare Pairs**, and then click the effect size estimator.
   The analysis task pane opens.
3. If the data consists of paired/repeated measurements in separate variables:
   a) In the **Model** drop-down menu, select **Paired/Repeated**.
   b) In the **Y** list, select the quantitative variables.
4. If the data consist of matched pairs with a separate response variable, factor variable, and blocking variable:
   a) In the **Model** drop-down menu, select **Matched**.
   b) In the **Y** drop-down list, select the quantitative response variable.
   c) In the **X (factor)** drop-down list, select the categorical factor variable identifying the groups.
   d) In the **Block** drop-down list, select the blocking variable identifying the matching.
5. In the **Difference** drop-down list, select the direction of the difference.
6. In the **Confidence interval** edit box, type the confidence level as a percentage, or type - to suppress the confidence interval, and then in the drop-down list, select the confidence bounds.
7. In the **Method** drop-down list, select the interval estimator.
8. Click **Calculate**.

Study design

Compare pairs analysis study requirements and dataset layout.

Requirements

- 2 or more repeated measurements on a quantitative response variable.

Paired/Repeated dataset layout

Use a column for each response variable (Blood pressure - Before, Blood pressure - After); each row has the values of the variables for a case (Subject).

	Blood pressure	
Subject (optional)	Before	After
1	123	124
2	109	97
3	112	113
4	102	105
5	98	95
6	114	119
7	119	114
8	112	114
9	110	121
10	117	118
…	…	…

Matched dataset layout

Use a column the response variable (Blood pressure), a column for the factor variable (Intervention), and a column for the blocking variable (Pair); each row has the values of the variables for a case.

Pair	Intervention	Blood pressure
1	Before	123
2	Before	109
3	Before	112
4	Before	102
5	Before	98
6	Before	114
7	Before	119
8	Before	112
9	Before	110
10	Before	117
…	Before	…
1	After	124
2	After	97
3	After	113
4	After	105
5	After	95
6	After	119
7	After	114
8	After	114
9	After	121
10	After	118
…	After	…

Chapter

Contingency tables

12

Contingency analysis describes and visualizes the distribution of categorical variables, and makes inferences about the equality of proportions, independence of the variables, or agreement between variables.

Contingency table

A contingency table, also known as a cross-classification table, describes the relationships between two or more categorical variables.

A table cross-classifying two variables is called a 2-way contingency table and forms a rectangular table with rows for the R categories of the X variable and columns for the C categories of a Y variable. Each intersection is called a cell and represents the possible outcomes. The cells contain the frequency of the joint occurrences of the X, Y outcomes. A contingency table having R rows and C columns is called an R x C table.

A variable having only two categories is called a binary variable. When both variables are binary, the resulting contingency table is a 2 x 2 table. Also, commonly known as a four-fold table because there are four cells.

	Smoke		
Alcohol consumption	Yes	No	Total
Low	10	80	90
High	50	40	90
Total	60	120	180

A contingency table can summarize three probability distributions – joint, marginal, and conditional.

- The joint distribution describes the proportion of the subjects jointly classified by a category of X and a category of Y. The cells of the contingency table divided by the total provides the joint distribution. The sum of the joint distribution is 1.
- The marginal distributions describe the distribution of the X (row) or Y (column) variable alone. The row and column totals of the contingency table provide the marginal distributions. The sum of a marginal distribution is 1.
- The conditional distributions describe the distribution of one variable given the levels of the other variable. The cells of the contingency table divided by the row or column totals provide the conditional distributions. The sum of a conditional distribution is 1.

When both variables are random, you can describe the data using the joint distribution, the conditional distribution of Y given X, or the conditional distribution of X given Y.

When one variable is and explanatory variable (X, fixed) and the other a response variable (Y, random), the notion of a joint distribution is meaningless, and you should describe the data using the conditional distribution of Y given X. Likewise, if Y is a fixed variable and X random, you should describe the data using the conditional distribution of X given Y.

When the variables are matched-pairs or repeated measurements on the same sampling unit, the table is square R=C, with the same categories on both the rows and columns. For these tables, the cells may exhibit a symmetric pattern about the main diagonal of the table, or the two marginal distributions may differ in some systematic way.

	After 6 months		
Before	Approve	Disapprove	Total
Approve	794	150	944
Disapprove	86	570	656
Total	880	720	1600

Creating a contingency table

Cross-classify data into a 2-way contingency table.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Compare Groups**, and then click **Contingency table**.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the first categorical variable.
4. In the **X** drop-down list, select the second categorical variable.
5. If the data are in frequency form, in the **Frequency** drop-down list, select the frequency count variable.
6. Optional: To customize the statistics to show, click **Customize...** and then select or clear the appropriate check boxes. To save the options as the default to show for all new analyses, click **Save as Defaults**.
7. Click **Calculate**.

Creating a contingency table (related data)

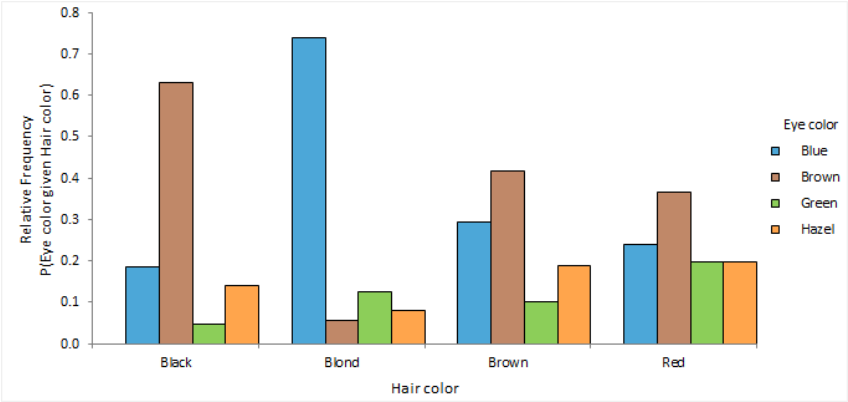
Cross-classify related data into a 2-way contingency table.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Compare Pairs**, and then click **Contingency Table**.
   The analysis task pane opens.
3. If the data consists of paired/repeated measurements in separate variables:
   a) In the **Model** drop-down menu, select **Paired / Repeated**.
   b) In the **Y** lists, select two categorical variables.
4. If the data consist of matched pairs with a separate response variable, factor variable, and blocking variable:
   a) In the **Model** drop-down menu, select **Matched**.
   b) In the **Y** drop-down list, select the categorical response variable.
   c) In the **X (factor)** drop-down list, select the categorical factor variable identifying the groups.
   d) In the **Block** drop-down list, select the blocking variable identifying the matching.
5. If the data are in frequency form, in the **Frequency** drop-down list, select the frequency count variable.
6. Optional: To customize the statistics to show, click **Customize...** and then select or clear the appropriate check boxes. To save the options as the default to show for all new analyses, click **Save as Defaults**.

7. Click **Calculate**.

Grouped frequency plot

A grouped or stacked frequency plot shows the joint or conditional distributions.



Effect size

An effect size estimates the magnitude of the difference of proportions or the association between two categorical variables.

The effect size that best describes a 2 x 2 contingency table depends on the study design that produced the data:

- When both variables are random variables, the odds ratio provides the best measure of association between the variables.
- When one variable is an explanatory variable (a fixed variable) and the other a response variable (a random variable), the effect size between the two groups of the response variable can be expressed as the odds ratio, the difference between proportions, or ratio of proportions.
- When the variables are matched-pairs or repeated measurements, the odds ratio or the difference between proportions are appropriate. The ratio of proportions is meaningless in this scenario.

A point estimate is a single value that is the best estimate of the true unknown parameter; a confidence interval is a range of values and indicates the uncertainty of the estimate.

For tables larger than 2 x 2, you must partition the contingency table into a series of 2 x 2 sub-tables to estimate the effect size.

Estimators

Estimators for contingency tables and their properties and assumptions.

Estimator	Purpose
Proportion difference	Estimate the difference of proportions (also known as Risk Difference).
Proportion ratio	Estimate the ratio of proportions (also known as Risk Ratio).
Odds ratio	Estimate the odds ratio. Values of the odds ratio range from zero to infinity. Values further from 1 in either direction represent a stronger association. The inverse value of an odds ratio represents the same degree of association in opposite directions.

Estimator	Purpose
	The odds-ratio does not change value when the orientation of a contingency table is reversed so that rows become columns and columns become rows. Therefore it is useful in situations where both variables are random.

Estimating the odds ratio

Estimate the odds ratio.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Compare Groups**, and then click **Odds Ratio**.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the first categorical variable.
4. In the **X** drop-down list, select the second categorical variable.
5. If the data are in frequency form, in the **Frequency** drop-down list, select the frequency count variable.
6. In the **Ratio** drop-down list, select the ratio of interest, and in the **Event** drop-down list, select the outcome of interest.
7. In the **Confidence interval** edit box, type the confidence level as a percentage, or type - to suppress the confidence interval, and then in the drop-down list, select the confidence bounds.
8. In the **Method** drop-down list, select the interval estimator.
9. Click **Calculate**.

Estimating the odds ratio (related data)

Estimate the odds ratio for related data.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Compare Pairs**, and then click **Odds Ratio**.
   The analysis task pane opens.
3. If the data consists of paired/repeated measurements in separate variables:
   a) In the **Model** drop-down menu, select **Paired / Repeated**.
   b) In the **Y** lists, select two categorical variables.
4. If the data consist of matched pairs with a separate response variable, factor variable, and blocking variable:
   a) In the **Model** drop-down menu, select **Matched**.
   b) In the **Y** drop-down list, select the categorical response variable.
   c) In the **X (factor)** drop-down list, select the categorical factor variable identifying the groups.
   d) In the **Block** drop-down list, select the blocking variable identifying the matching.
5. If the data are in frequency form, in the **Frequency** drop-down list, select the frequency count variable.
6. In the **Ratio** drop-down list, select the ratio of interest, and in the **Event** drop-down list, select the outcome of interest.
7. In the **Confidence interval** edit box, type the confidence level as a percentage, or type - to suppress the confidence interval, and then in the drop-down list, select the confidence bounds.
8. In the **Method** drop-down list, select the interval estimator.

9. Click **Calculate**.

Relative risk

Relative risk is the ratio of the probability of the event occurring in one group versus another group.

Medical statistics uses categorical data analysis extensively to describe the association between the occurrence of a disease and exposure to a risk factor. In this application, the terminology is often used to frame the statistics in terms of risk. Risk difference is equivalent to a difference between proportions, and a risk ratio is equivalent to a ratio of proportions.

The term relative risk is the ratio of the probability of the event occurring in the exposed group versus a non-exposed group. It is the same as the risk ratio. Although a relative risk is different from odds ratio, in some circumstances, such as the low prevalence of the disease, the odds ratio asymptotically approaches the relative risk and is, therefore, an estimate of the relative risk. It is better to avoid confusion and use the terms risk ratio and odds ratio for the effect size estimates and use the term relative risk for the population parameter.

A prospective cohort study or a clinical trial can estimate the risk ratio or the odds ratio of the occurrence of a disease given exposure to a risk factor. Whereas, a retrospective case-control study can estimate the odds ratio of the risk factor given the disease, which is equivalent to the odds ratio of the occurrence of the disease given the risk factor. A risk ratio is not useful because it refers to the risk factor given the disease, which is not normally of interest. However, as stated above the odds ratio can be used as an estimate of the relative risk in most circumstances.

Practitioners often prefer the risk ratio due to its more direct interpretation. Statisticians tend to prefer the odds ratio as it applies to a wide range of study designs, allowing comparison between different studies and meta-analysis based on many studies. It also forms the basis of logistic regression.

Inferences about equality of proportions

Inferences about the proportions in populations are made using random samples of data drawn from each population of interest.

Equality of proportions hypothesis test

A hypothesis test formally tests if the proportions in two or more populations are equal.

When one variable is an explanatory variable (X, fixed) and the other a response variable (Y, random), the hypothesis of interest is whether the populations have the same or different proportions in each category.

The hypotheses to test depends on the number of samples:

- For two samples, the null hypothesis states that the parameter of interest is equal to the hypothesized value, against the alternative hypothesis it is not equal to (or less than, or greater than) the hypothesized value.

  You can formulate the hypotheses in terms of the parameter of interest: odds ratio = 1, the ratio of proportions = 1, or the difference of proportions = 0 depending on the desired effect size estimate.
- For more than two samples, the null hypothesis states that the proportions in each category are equal for all populations, against the alternative hypothesis that the proportions in a category are not equal for at least 2 populations.

When the test p-value is small, you can reject the null hypothesis and conclude that the populations differ in the proportions in at least one category.

Tests for contingency tables larger than 2 x 2 are omnibus tests and do not tell you which groups differ from each other or in which categories. You should use the mosaic plot to examine the association, or partition the contingency table into a series of 2 x 2 sub-tables and test each table.

Exact and asymptotic p-values

Asymptotic p-values are useful for large sample sizes when the calculation of an exact p-value is too computer-intensive.

Many test statistics follow a discrete probability distribution. However, hand calculation of the true probability distributions of many test statistics is too tedious except for small samples. It can be computationally difficult and time intensive even for a powerful computer. Instead, many statistical tests use an approximation to the true distribution. Approximations assume the sample size is large enough so that the test statistic converges to an appropriate limiting normal or chi-square distribution.

A p-value that is calculated using an approximation to the true distribution is called an asymptotic p-value. A p-value calculated using the true distribution is called an exact p-value. For large sample sizes, the exact and asymptotic p-values are very similar. For small sample sizes or sparse data, the exact and asymptotic p-values can be quite different and can lead to different conclusions about the hypothesis of interest.

When using the true distribution, due to the discreteness of the distribution, the p-value and confidence intervals are conservative. A conservative interval guarantees that the actual coverage level is at least as large as the nominal confidence level, though it can be much larger. For a hypothesis test, it guarantees protection from type I error at the nominal significance level. Some statisticians prefer asymptotic p-values and intervals, as they tend to achieve an average coverage closer to the nominal level. The problem is the actual coverage of a specific test or interval may be less than the nominal level.

Although a statistical test might commonly use an approximation, it does not mean it cannot be calculated using the true probability distribution. For example, an asymptotic p-value for the Pearson X^2 test uses the chi-squared approximation, but the test could also compute an exact p-value using the true probability distribution. Similarly, although the Fisher test is often called the Fisher exact test because it computes an exact p-value using the hypergeometric probability distribution, the test could also compute an asymptotic p-value. Likewise, the Wilcoxon-Mann-Whitney test often computes an exact p-value for small sample sizes and reverts to an asymptotic p-value for large sample sizes.

Wald, Score, Likelihood ratio

Three different ways are used to make large sample inferences all using a X^2 statistic, so it is important to understand the differences to avoid confusion.

The score, likelihood ratio, and Wald evidence functions are useful when analyzing categorical data. All have approximately chi-squared (X^2) distribution when the sample size is sufficiently large.

Because of their underlying use of the X^2 distribution, mixing of the techniques often occurs when testing hypotheses and constructing interval estimates. Introductory textbooks often use a score evidence function for the hypothesis test then use a Wald evidence function for the confidence interval. When you use different evidence functions the results can be inconsistent. The hypothesis test may be statistically significant, but the confidence interval may include the hypothesized value suggesting the result is not significant. Where possible you should use the same underlying evidence function to form the confidence interval and test the hypotheses.

Note that when the degrees of freedom is 1 the Pearson X^2 statistic is equivalent to a squared score Z statistic because they are both based on the score evidence function. A 2-tailed score Z test produces the same p-value as a Pearson X^2 test.

Many Wald-based interval estimates have poor coverage and defy conventional wisdom. Modern computing power allows the use of computationally intensive score evidence functions for interval estimates, such that we recommended them for general use.

Tests for equality of proportions (independent samples)

Tests for the equality of proportions of independent samples and their properties and assumptions.

Test	Purpose
Fisher exact	Test if the two proportions are equal. Assumes fixed marginal distributions. Although not naturally fixed in most studies, this test still applies by conditioning on the marginal totals. Uses hypergeometric distribution and computes an exact p-value. The exact p-value is conservative, that is, the actual rejection rate is below the nominal significance level. Recommended for small sample sizes or sparse data.
Score Z	Test if the two proportions are equal. Uses the score statistic and computes an asymptotic p-value. Note that the two-sided Score Z test is equivalent to Pearson's X^2 test.
Pearson X^2	Test if the proportions are equal in each category. Uses the score statistic and computes an asymptotic p-value. Note that the asymptotic distribution of the X^2 test is the same regardless of whether only the total is fixed, both marginal distributions are fixed, or only one marginal distribution is fixed.

Testing equality of proportions (independent samples)

Test if there is a difference between the proportions of two or more independent samples.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Compare Groups**, and then click the hypothesis test.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the categorical response variable.
4. In the **X** drop-down list, select the categorical factor variable that identifies the groups.
5. If the data are in frequency form, in the **Frequency** drop-down list, select the frequency count variable.
6. In the **Hypotheses** drop-down list, select the null and alternative hypothesis.
7. Optional: To compare the p-value against a predefined significance level, in the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%).
8. Click **Calculate**.

Tests for equality of proportions (related samples)

Tests for the equality of proportions of related samples and their properties and assumptions.

Test	Purpose
McNemar-Mosteller Exact	Test if two proportions are equal. Uses a binomial distribution and computes an exact p-value. The exact p-value is conservative, that is, the actual rejection rate is below the nominal significance level. Recommended for small sample sizes or sparse data.
Score Z	Test if two proportions are equal. Uses the score statistic and computes an asymptotic p-value. Note that the two-sided Score Z test is equivalent to McNemar's X^2 test.

Testing equality of proportions (related samples)

Test if there is a difference between the proportions of two or more related samples.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Compare Pairs**, and then click the hypothesis test.
   The analysis task pane opens.
3. If the data consists of paired/repeated measurements in separate variables:
   a) In the **Model** drop-down menu, select **Paired / Repeated**.
   b) In the **Y** lists, select two categorical variables.
4. If the data consist of matched pairs with a separate response variable, factor variable, and blocking variable:
   a) In the **Model** drop-down menu, select **Matched**.
   b) In the **Y** drop-down list, select the categorical response variable.
   c) In the **X (factor)** drop-down list, select the categorical factor variable identifying the groups.
   d) In the **Block** drop-down list, select the blocking variable identifying the matching.
5. If the data are in frequency form, in the **Frequency** drop-down list, select the frequency count variable.
6. In the **Hypotheses** drop-down list, select the null and alternative hypothesis.
7. Optional: To compare the p-value against a predefined significance level, in the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%).
8. Click **Calculate**.

Inferences about independence

Inferences about the independence between categorical variables are made using a random bivariate sample of data drawn from the population of interest.

Independence hypothesis test

A hypothesis test formally tests whether 2 categorical variables are independent.

When both variables are random, the hypothesis of interest is whether the variables are independent.

The null hypothesis states that the variables are independent, against the alternative hypothesis that the variables are not independent. That is, in terms of the contingency table, the null

hypothesis states the event "an observation is in row i" is independent of the event "that same observation is in column j," for all i and j, against the alternative hypothesis that is the negation of the null hypothesis.

When the test p-value is small, you can reject the null hypothesis and conclude that the variables are not independent.

Continuity correction

Continuity corrections such as Yates $X^2$are no longer needed with modern computing power.

Continuity corrections have historically been used to make adjustments to the p-value when a continuous distribution approximates a discrete distribution. Yates correction for the Pearson chi-square (X^2) test is probably the most well-known continuity correction. In some cases, the continuity correction may adjust the p-value too far, and the test then becomes overly conservative.

Modern computing power makes such corrections unnecessary, as exact tests that use the discrete distributions are available for moderate and in many cases even large sample sizes. Hirji *(Hirji 2005)*, states “An applied statistician today, in our view, may regard such corrections as interesting historical curiosities.”

Tests for independence

Tests for the independence of 2 categorical variables and their properties and assumptions.

Test	Purpose
Pearson X^2	Test if the variables are independent. Uses the score statistic and computes an asymptotic p-value.
Likelihood ratio G^2	Test if the variables are independent. Uses the likelihood ratio statistic and computes an asymptotic p-value. Pearson X^2 usually converges to the chi-squared distribution more quickly than G^2. The likelihood ratio test is often used in statistical modeling as the G^2 statistic is easier to compare between different models.

Testing independence

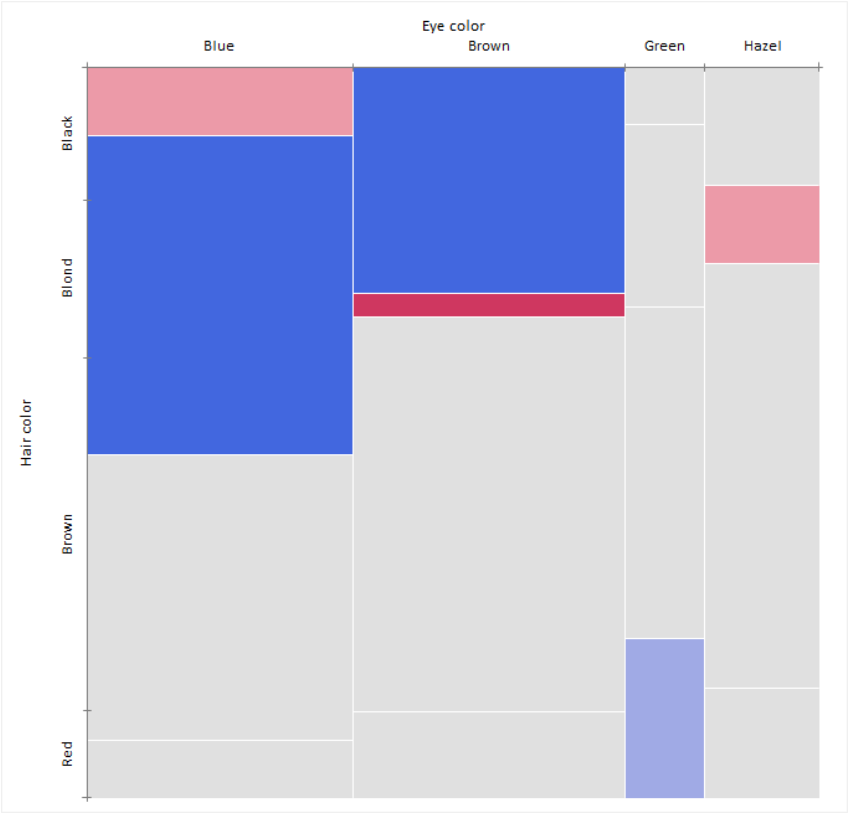
Test if two categorical variables are independent in a bivariate population represented by the sample.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Compare Groups**, and then click the hypothesis test.
   The analysis task pane opens.
3. In the **X** and **Y** drop-down lists, select the categorical variables.
4. In the **Hypotheses** drop-down list, select the independence null and alternative hypotheses (the 2nd in the list).
5. Optional: To compare the p-value against a predefined significance level, in the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%).
6. Click **Calculate**.

Mosaic plot

A mosaic plot is a visual representation of the association between two variables.

A mosaic plot is a square subdivided into rectangular tiles the area of which represents the conditional relative frequency for a cell in the contingency table. Each tile is colored to show the deviation from the expected frequency (residual) from a Pearson X^2 or Likelihood Ratio G^2 test.



You can use the mosaic plot to discover the association between two variables. Red tiles indicate significant negative residuals, where the frequency is less than expected. Blue tiles indicate significant positive residuals, where the frequency is greater than expected. The intensity of the color represents the magnitude of the residual.

Creating a mosaic plot

Visualize the association between 2 categorical variables.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Compare Groups**, and then click **Mosaic**.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the categorical response variable.
4. In the **X** drop-down list, select the categorical factor variable identifying the groups.
5. In the **Color by** drop-down list, select **Category** to color by category, or **Residual** to color by the Pearson X^2 residual.
6. To label the tiles with the conditional relative frequency, click **Label tiles**.
7. Click **Calculate**.

Study design

Contingency analysis study requirements and dataset layout.

Requirements

- 2 categorical variables.

or:

- A qualitative response variable.
- A categorical factor variable with 2 or more levels.

or:

- 2 or more repeated measurements on a qualitative response variable.

Dataset layout

Use a column for each variable (Eye color, Hair color); each row has the values of the variables.

Subject (optional)	Eye color	Hair color
1	Brown	Black
2	Blue	Red
3	Blue	Blonde
4	Brown	Red
5	Green	Blonde
6	Hazel	Brown
7	Blue	Blonde
8	Brown	Black
9	Green	Red
10	Green	Blonde
…	…	…

Frequency form dataset layout

Use a column for each variable (Eye color, Hair color) and a column for the number of cases (Frequency); each row has the values of the variables and the frequency count.

Eye color	Hair color	Frequency
Brown	Black	69
Brown	Brown	119
Brown	Red	26
Brown	Blonde	7
Blue	Black	20
Blue	Brown	84
Blue	Red	17
Blue	Blonde	94
Hazel	Black	15
Hazel	Brown	54
Hazel	Red	14
Hazel	Blonde	10
Green	Black	5
Green	Brown	29
Green	Red	14
Green	Blonde	16

Chapter

Correlation and association

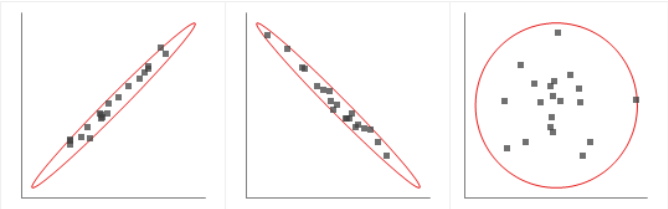
13

Correlation analysis explores the association between two or more variables and makes inferences about the strength of the relationship.

Note: It is common to use the terms correlation and association interchangeably. Technically, association refers to any relationship between two variables, whereas correlation is often used to refer only to a linear relationship between two variables. The terms are used interchangeably in this guide, as is common in most statistics texts.

Scatter plot

A scatter plot shows the association between two variables. A scatter plot matrix shows all pairwise scatter plots for many variables.



If the variables tend to increase and decrease together, the association is positive. If one variable tends to increase as the other decreases, the association is negative. If there is no pattern, the association is zero.

When a straight line describes the relationship between the variables, the association is linear. When a constantly increasing or decreasing nonlinear function describes the relationship, the association is monotonic. Other relationships may be nonlinear or non-monotonic.

The type of relationship determines the statistical measures and tests of association that are appropriate.

If the association is a linear relationship, a bivariate normal density ellipse summarizes the correlation between variables. The narrower the ellipse, the greater the correlation between the variables. The wider and more round it is, the more the variables are uncorrelated. If the association is nonlinear, it is often worth trying to transform the data to make the relationship linear as there are more statistics for analyzing linear relationships and their interpretation is easier than nonlinear relationships.

An observation that appears detached from the bulk of observations may be an outlier requiring further investigation. An individual observation on each of the variables may be perfectly reasonable on its own but appear as an outlier when plotted on a scatter plot. Outliers can badly affect the product-moment correlation coefficient, whereas other correlation coefficients are more robust to them.

Creating a scatter plot or scatter plot matrix

Visualize the relationships between two or more variables.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Correlation**, and then click **Scatterplot** or **Scatterplot Matrix**, and then click the plot type.
   The analysis task pane opens.
3. If the model is bivariate, in the **X** and **Y** drop-down lists, select the variables.
4. If the model is multivariate, in the **Variables** list, select the variables.
5. Optional: To show stratification of the observations, select **Vary point color/symbol**, and then in the **Group / Color / Symbol** drop-down list, select a variable.
6. Optional: To label the observations, select the **Label observations** check box.
7. Click **Calculate**.

Covariance

Covariance is a measure of how much two variables change together. A covariance matrix measures the covariance between many pairs of variables.

When the variables tend to show similar behavior, the covariance is positive. That is when greater values of one variable mainly correspond with the greater values of the other variable, or lesser values of one variable correspond with lesser values of the other variable. When the variables tend to show opposite behavior, the covariance is negative. That is when the greater values of one variable mainly correspond to the lesser values of the other and vice-versa.

The magnitude of the covariance is not meaningful to interpret. However, the standardized version of the covariance, the correlation coefficient, indicates by its magnitude the strength of the relationship.

A covariance matrix measures the covariance between many variables. Because the covariance of a variable with itself is that variable's variance, the diagonal of the covariance matrix is simply the variance of each variable.

Correlation coefficient

A correlation coefficient measures the association between two variables. A correlation matrix measures the correlation between many pairs of variables.

The type of relationship between the variables determines the best measure of association:

- When the association between the variables is linear, the product-moment correlation coefficient describes the strength of the linear relationship.

  The correlation coefficient ranges from -1 to +1. +1 indicates a perfect positive linear relationship, and -1 indicates a perfect negative linear relationship. Zero indicates the variables are uncorrelated and there is no linear relationship. Normally the correlation coefficient lies somewhere between these values.
- When the association between the variables is not linear, a rank correlation coefficient describes the strength of association.

  Rank correlation coefficients range from -1 to +1. A positive rank correlation coefficient describes the extent to which as one variable increases the other variable also tends to increase, without requiring that increase to be linear. If one variable increases, as the other tends to decrease, the rank correlation coefficient is negative.

It is best to use a scatter plot to identify the type of association between the variables and then use an appropriate measure of association for the relationship. Do not be tempted just to look for the highest correlation coefficient.

A correlation matrix measures the correlation between many variables. It is equivalent to a covariance matrix of the standardized variables.

Color map

A color map highlights the sign and magnitude of coefficients in a patterned matrix.

	P2	N1	N2	Z1	P1	Z2
P2	-	0.163	-0.144	0.012	0.990	-0.209
N1	0.163	-	-0.990	-0.094	0.106	0.119
N2	-0.144	-0.990	-	0.043	-0.086	-0.157
Z1	0.012	-0.094	0.043	-	-0.003	0.000
P1	0.990	0.106	-0.086	-0.003	-	-0.221
Z2	-0.209	0.119	-0.157	0.000	-0.221	-

Red indicate negative values, blue indicates positive values. Intensity of the color represents the magnitude of the value, the darker more extreme.

Calculating a correlation matrix

Measure the correlation / association between 2 or more variables.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Correlation** or **Multivariate**, and then click the statistics to show:

Correlation Matrix	Show the product-moment correlation coefficients.
Rank Correlation Matrix	Show the rank correlation coefficients.

   The analysis task pane opens.
3. In the **Variables** list, select the variables.
4. Optional: To highlight the coefficients depending on their sign and magnitude, select the **Color maps** check box.
5. Optional: To show univariate descriptive statistics for each variable, select the **Descriptives for each variable** check box.
6. Click **Calculate**.

Calculating a covariance matrix

Measure the covariance between 2 or more variables.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Multivariate**, and then click **Covariance Matrix**.
   The analysis task pane opens.
3. In the **Variables** list, select the variables.
4. Click **Calculate**.

Inferences about association

Inferences about the strength of association between variables are made using a random bivariate sample of data drawn from the population of interest.

Correlation not Causation

Do not be fooled into thinking a significant correlation implies you to make statements about causation.

Correlation does not imply causation is a reminder that although a statistically significant correlation might exist between two variables, it does not imply that one causes the other.

There are many reasons why correlation does not imply causation: reverse causation where causation is actually in the opposite direction; bidirectional causation where a change in one variable causes a change in the other and vice-versa; a third unobserved variable that is the actual cause of the correlation; or simply a coincidence.

Correlation is a hint to investigate further, where possible using a controlled experiment. In some fields controlled experiments are impossible. Instead, cause and effect can be established only by observing the same thing in a wide variety of scenarios until all but the suspected cause can be ruled out. Even then the causal relationship can only be confirmed when you discover the underlying mechanism of the cause and effect.

Parameter estimate

A parameter estimate is either a point or interval estimate of the unknown population correlation coefficient.

A point estimate is a single value that is the best estimate of the true unknown population correlation coefficient; a confidence interval is a range of values and indicates the uncertainty of the estimate.

Estimators for correlation and association

Estimators for correlation/association and their properties and assumptions.

Estimator	Purpose
Pearson r	Estimate the Pearson rho correlation coefficient using the sample Pearson correlation coefficient r. Use when a linear function best describes the relationship between the variables. Susceptible to outliers. Assumes a bivariate normal distribution.
Spearman rs	Estimate the Spearman rho rank correlation coefficient using the sample Spearman rs correlation coefficient. Use when a monotonic function best describes the relationship between variables.
Kendall tau	Estimate the Kendall tau rank correlation coefficient. Use when a monotonic function best describes the relationship between the variables. Preferred instead of Spearman correlation because Kendall's tau is the difference between the probability that the two variables are in the same order, versus the probability that the two variables are in different orders. Robust to outliers.

Estimating the correlation/association

Estimate the correlation/association between pairs of variables.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Correlation**, and then the parameter estimator.
   The analysis task pane opens.
3. If analyzing two variables:

a) In the **Model** drop-down menu, select **Bivariate**.
b) In the **X** drop-down list, select the first variable.
c) In the **Y** drop-down list, select the second variable.

4. If analyzing more than two variables:
   a) In the **Model** drop-down menu, select **Multivariate**.
   b) In the **Variables** list, select the variables.
5. In the **Confidence interval** edit box, type the confidence level as a percentage, or type - to suppress the confidence interval, and then in the drop-down list, select the confidence bounds.
6. In the **Method** drop-down list, select the interval estimator.
7. Click **Calculate**.

Correlation/association hypothesis test

A hypothesis test formally tests if there is correlation/association between two variables in a population.

The hypotheses to test depends on the type of association:

- For a product-moment correlation, the null hypothesis states that the population correlation coefficient is equal to a hypothesized value (usually 0 indicating no linear correlation), against the alternative hypothesis that it is not equal (or less than, or greater than) the hypothesized value.
- For rank correlation methods, the hypotheses are restricted and more general. The null hypothesis states the variables are independent, against the alternative hypothesis that there is an association, such as a monotonic function.

When the test p-value is small, you can reject the null hypothesis and conclude that the population correlation coefficient is not equal to the hypothesized value, or for rank correlation that the variables are not independent. It is important to remember that a statistically significant test may not have any practical importance if the correlation coefficient is very small.

Pearson's and Kendall's tests are preferred as both have associated estimators of the population correlation coefficient (rho and tau respectively). Although the Spearman test is popular due to the ease of computation, the Spearman correlation coefficient is a measure of the linear association between the ranks of the variables and not the measure of association linked with the Spearman test.

Testing correlation/association

Test if the correlation between 2 variables is equal to a hypothesized value, or test for independence.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Correlation**, and then the parameter estimator.
   The analysis task pane opens.
3. If analyzing two variables:
   a) In the **Model** drop-down menu, select **Bivariate**.
   b) In the **X** drop-down list, select the first variable.
   c) In the **Y** drop-down list, select the second variable.
4. If analyzing more than two variables:
   a) In the **Model** drop-down menu, select **Multivariate**.
   b) In the **Variables** list, select the variables.
5. Select the **Hypothesis test** check box.

6. In the **Hypotheses** drop-down list, select the null and alternative hypothesis.
7. Optional: To compare the p-value against a predefined significance level, in the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%).
8. Click **Calculate**.

Study design

Multivariate analysis study requirements and dataset layout.

Requirements

- 2 or more quantitative variables.

Dataset layout

Use a a column for each variable (Height, Weight); each row has the values of the variables for a case (Subject).

Subject (optional)	Height	Weight
1	175	65
2	180	70
3	160	90
4	190	55
5	180	100
6	150	55
7	140	75
8	160	80
9	165	80
10	180	95
…	…	…

Chapter 14

Principal component analysis (PCA)

Principal component analysis (PCA) reduces the dimensionality of a dataset with a large number of interrelated variables while retaining as much of the variation in the dataset as possible.

PCA is a mathematical technique that reduces dimensionality by creating a new set of variables called principal components. The first principal component is a linear combination of the original variables and explains as much variation as possible in the original data. Each subsequent component explains as much of the remaining variation as possible under the condition that it is uncorrelated with the previous components.

The first few principal components provide a simpler picture of the data than trying to understand all the original variables. Sometimes, it is desirable to try and name and interpret the principal components, a process call *reification*, although this should not be confused with the purpose of factor analysis.

Principal components

Principal components are the linear combinations of the original variables.

Variances

Variances of each principal component show how much of the original variation in the dataset is explained by the principal component.

When the data is standardized, a component with a variance of 1 indicates that the principal component accounts for the variation equivalent to one of the original variables. Also, the sum of all the variances is equal to the number of original variables.

Coefficients

Coefficients are the linear combinations of the original variables that make up the principal component. The coefficients for each principal component can sometimes reveal the structure of the data. Absolute values near zero indicate that a variable contributes little to the component, whereas larger absolute values indicate variables that contribute more to the component.

Often, when the data is centered and standardized, the coefficients are normalized so that the sum of the squares of the coefficients of a component is equal to the variance of the component. In this normalization, the coefficients can be interpreted as the correlation between the original variable and the principal component, and are often called loadings (a term borrowed from factor analysis).

Scores

Scores are new variables that are the value of the linear combination of the original variables. The scores are normalized so that the sum of squares equals the variance of the principal component.

Scree plot

A scree plot visualizes the dimensionality of the data.

The scree plot shows the cumulative variance explained by each principal component. You can make decision on the number of components to keep to adequately describe a dataset using ad-hoc rules such as components with a variance > 0.7 or where the cumulative proportion of variation is > 80% or > 90% *(Jolliffe 2002)*

Calculating principal components

Reduce the dimensionality of multidimensional data using PCA.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Multivariate**, and then click **Principal Components**.
   The analysis task pane opens.
3. In the **Variables** list, select the variables.
4. Optional: To highlight the coefficients depending on their sign and magnitude, select the **Color maps** check box.
5. Optional: To show a plot of the cumulative variance explained by each principal component, select the **Scree plot** check box.
6. Click **Calculate**.
7. To save the computed principal component scores back to the dataset for further analysis, on the **Analyse-it** ribbon tab, click **Save Variable** > **Scores** and select the number of dimensions to store.

Biplot

A biplot simultaneously plots information on the observations and the variables in a multidimensional dataset.

A biplot can optimally represent any two of the following characteristics:

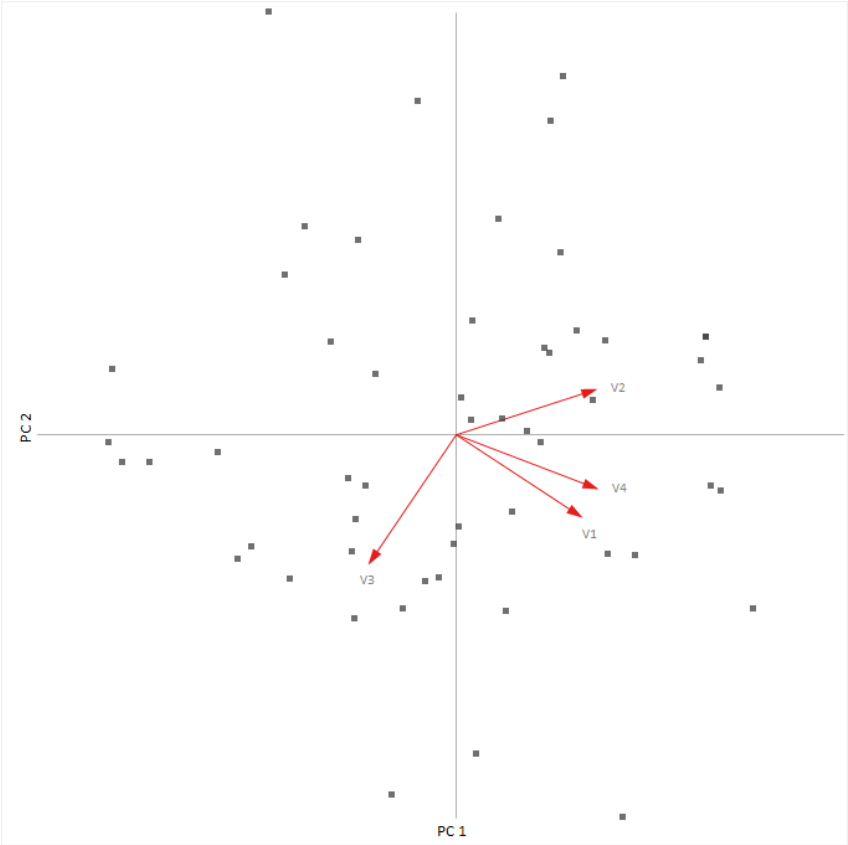
- distances between observations
- relationships between variables
- inner products between observations and variables

There are 3 types of biplot based on which of these characteristics they represent:

Type	Characteristics
PCA	Distances between the observations and also the inner products between observations and variables.
Covariance / Correlation	Relationships between the variables and the inner products between observations and variables.
Joint	Distances between observations and also the relationship between variables.

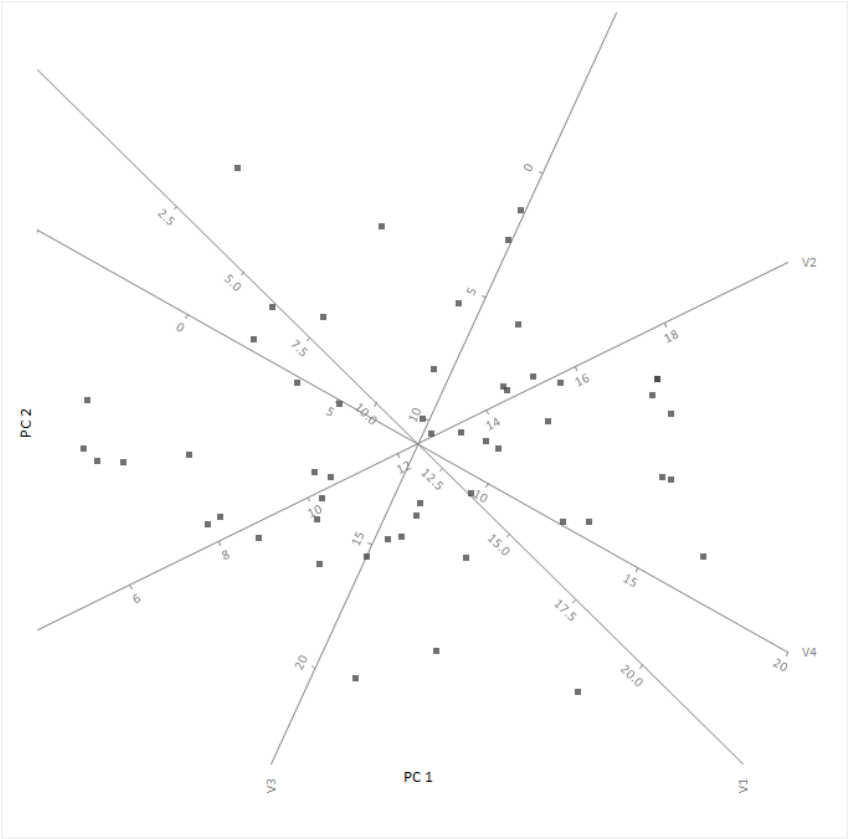
A 2-dimensional biplot represents the information contained in two of the principal components. It is an approximation of the original multidimensional space.

Classic biplot



The classical biplot *(Gabriel 1971)* plots points representing the observations and vectors representing the variables.

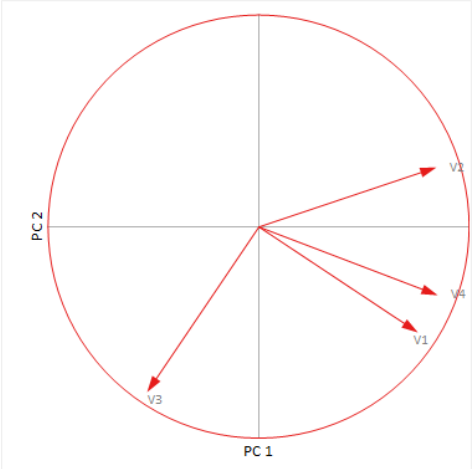
PCA biplot



A more recent innovation, the PCA biplot *(Gower & Hand 1996)*, represents the variables with calibrated axes and observations as points allowing you to project the observations onto the axes to make an approximation of the original values of the variables.

Monoplot

A monoplot plots information on the observations or the variables in a multidimensional dataset.



A monoplot can only represent either of the following characteristics - distance between observations, or relationships between variables. In contrast, a joint biplot can represent both characteristics.

A 2-dimensional monoplot represents the information contained in two of the principal components. It is an approximation of the original multidimensional space.

The correlation monoplot plots vectors pointing away from the origin to represent the original variables. The angle between the vectors is an approximation of the correlation between the variables. A small angle indicates the variables are positively correlated, an angle of 90 degrees indicates the variables are not correlated, and an angle close to 180 degrees indicates the variables are negatively correlated. The length of the line and its closeness to the circle indicate how well the plot represents the variable. It is, therefore, unwise to make inferences about relationships involving variables with poor representation.

The covariance monoplot plots vectors pointing away from the origin to represent the original variables. The length of the line represents the variance of the variable, and the inner product between the vectors represents the covariance.

Creating a biplot

A biplot simultaneously shows information on the observations and the variables in a multidimensional dataset.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Multivariate** > **Biplot / Monoplot**, and then click the plot type.
   The analysis task pane opens.
3. In the **Variables** list, select the variables.
4. Optional: To label the observations, select the **Label points** check box.
5. Optional: To label the variables, select the **Label vectors** check box.
6. Optional: To show stratification, select **Vary point color/symbol**, and then in the **Group / Color / Symbol** drop-down list, select a variable.
7. Click **Calculate**.

Creating a correlation monoplot

A correlation monoplot shows the relationship between many variables in a multidimensional dataset.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Multivariate** > **Biplot / Monoplot**, and then click **Correlation Monoplot**.
   The analysis task pane opens.
3. In the **Variables** list, select the variables.
4. Click **Calculate**.

Chapter 15

Factor analysis (FA)

Factor analysis (FA) explores the underlying factors (latent variables) among a large number of interrelated variables.

Factor analysis is related to principal component analysis (PCA), but the two are not identical and have different purposes. FA makes the assumption that an underlying model exists, whereas PCA is a descriptive variable reduction technique.

EFA

Exploratory factor analysis (EFA) identifies the underlying relationships between a large number of interrelated variables when there are no prior hypotheses about factors or patterns amongst the variables.

EFA is a technique based on the common factor model which describes the measured variables by a function of the common factors, unique factors, and error of measurements. Common factors are those that influence two or more measured variables, while unique factors influence only one measured variable.

Pattern matrix

The factor pattern matrix loadings are the linear combinations of the factors that make up the original standardized variables.

Structure matrix

The factor structure matrix loadings are the correlation coefficients between the factors and the variables.

Correlation matrix

The factor correlation matrix coefficients are the correlation coefficients between the factors.

Note: When the factors are not rotated, or the rotation is orthogonal, there is no correlation between the factors and the correlation matrix is equal to the identity matrix. Also, the loadings in the pattern matrix and structure matrix are identical, although it can be useful to remember the different interpretations - as linear coefficients or correlation coefficients.

Factor rotation

Rotations minimize the complexity of the factor loadings to make the structure simpler to interpret.

Factor loading matrices are not unique, for any solution involving two or more factors there are an infinite number of orientations of the factors that explain the original data equally well. Rotation of the factor loading matrices attempts to give a solution with the best simple structure.

There are two types of rotation:

- Orthogonal rotations constrain the factors to be uncorrelated. Although often favored, in many cases it is unrealistic to expect the factors to be uncorrelated, and forcing them to be uncorrelated makes it less likely that the rotation produces a solution with a simple structure.
- Oblique rotations permit the factors to be correlated with one another. Often produces solutions with a simpler structure.

Matrix rotations

Orthogonal and oblique matrix rotations.

p = number of variables, m = number of factors.

Method	Parameters	Purpose
Varimax		Orthogonal only. A computational faster equivalent to CF-Varimax.
Promax	power 0 ... 4	Oblique only.
Crawford-Ferguson	kappa 0 ... 1	Smaller kappa minimizes variables complexity and larger kappa minimizes factor complexity.
CF-Varimax	Crawford-Ferguson kappa = 1/p.	Spread variances across factors. Each factor tends to have either large or small loadings on a particular variable making it easy to identify each variable with a single factor.
CF-Quartimax	Crawford-Ferguson kappa = 0.	Minimizes variable complexity. Works well with distinct clusters without cross-loadings.
CF-Equamax	Crawford-Ferguson kappa = m/(2p).	
CF-Parsimax	Crawford-Ferguson kappa = (m-1) / (p+m-2).	
CF-Factor Parsimony	Crawford-Ferguson kappa = 1.	Minimizes factor complexity. Primarily of theoretical interest.
Oblimin	gamma 0 ... 1	
O-Quartimin	Oblimin gamma = 0.	Equivalent to quartimax.
O-Biquartimin	Oblimin gamma = 0.5.	
O-Covarimin	Oblimin gamma = 1.	Equivalent to varimax.
Geomin	delta > 0	Minimizes variable complexity.

Extracting factors

Extract the underlying factors (latent variables) among a large number of interrelated variables in a multidimensional dataset.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Multivariate**, and then click **Common Factors**.
   The analysis task pane opens.
3. In the **Variables** list, select the variables.
4. In the **Factors to extract** edit box, type the number of underlying factors to attempt to extract.

5. Optional: To highlight the coefficients depending on their sign and magnitude, select the **Color maps** check box.
6. Optional: To rotate the extracted factors to allow easier interpretation, in the **Rotation** drop-down list, select the type of rotation, and then in the **Method** drop-down list, select the rotation and type any additional required parameters.
7. Click **Calculate**.

Chapter 16

Item reliability

Item reliability is the consistency of a set of items (variables); that is to what extent they measure the same thing. When a set of items are consistent, they can make a measurement scale such as a sum scale.

Cronbach's alpha is the most popular measure of item reliability; it is the average correlation of items in a measurement scale. If the items have variances that significantly differ, standardized alpha is preferred.

When all items are consistent and measure the same thing, then the coefficient alpha is equal to 1. A high value for alpha does not imply that the measure is unidimensional. To prove that a scale is unidimensional, you can use factor analysis to check the dimensionality.

It is possible to see the effect of an individual item on the overall alpha value by recomputing Cronbach's alpha excluding that item. If alpha increases when you exclude an item, that item does not highly correlate with the other items in the scale. If the alpha decreases, that item does correlate with the other items in the scale.

Testing item reliability

Test the consistency of a set of items (variables).

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Multivariate**, and then click **Cronbach's Alpha**.
   The analysis task pane opens.
3. In the **Variables** list, select the variables.
4. If the variables have widely varying variances, select the **Standardized** check box.
5. To see the effect of each variable on the overall alpha, select the **Effect of excluding each variable** check box.
6. Click **Calculate**.

Chapter 17

Fit model

Fit model describes the relationship between a response variable and one or more predictor variables.

There are many different models that you can fit including simple linear regression, multiple linear regression, analysis of variance (ANOVA), analysis of covariance (ANCOVA), and binary logistic regression.

Linear fit

A linear model describes the relationship between a continuous response variable and one or more explanatory variables using a linear function.

Simple regression models

Simple regression models describe the relationship between a single predictor variable and a response variable.

Name	Model	Purpose
Line	$Y=b_0+b_1x$	Fit a straight line.
Polynomial	$Y=b_0+b_1x+b_1x2....$	Fit a polynomial curve. Polynomials are useful when the function is smooth, a polynomial of a high enough degree estimates any smooth function. Polynomials are used as approximations and rarely represent a physical model.
Logarithmic	$Y=b_0+ b_1Log(x)$	Fit a logarithmic function curve.
Exponential	$Y=a*b_1x$	Fit an exponential function curve.
Power	$Y=a*x^{b_1}$	Fit a power function curve.

Fitting a simple linear regression

Fit a simple linear regression model to describe the relationship between single a single predictor variable and a response variable.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Fit Model**, and then click the simple regression model.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the response variable.
4. In the **X** drop-down list, select the predictor variable.
5. Optional: To assign a weight to each row for the analysis, in the **Weight** drop-down list, select the weight variable.
6. Click **Calculate**.

Advanced models

Advanced models describe the relationship between a response variable and multiple predictor terms.

An advanced model is built from simple terms, polynomial terms and their interactions.

Some common forms of advanced models are:

Type	Description
Simple regression	A single quantitative simple term (X).
Multiple regression	Two or more quantitative simple terms (X, Z)
Polynomial	A series of polynomial terms of the nth degree (X, X^2, X^3...).
One-way ANOVA	A single categorical term (A).
Two-way ANOVA	Two categorical terms (A, B).
Two-way ANOVA with interaction	Two categorical terms and their interaction (A, B, A*B).
Fully factorial ANOVA	All the simple categorical terms and all the crossed interaction terms (A, B, C, A*B, A*C, B*C, A*B*C).
ANCOVA	One or more categorical terms and one or more quantitative terms (A, X) or (A, X, A*X).

Fitting a multiple linear regression

Fit a multiple linear regression model to describe the relationship between many quantitative predictor variables and a response variable.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Fit Model**, and then click **Multiple Regression**.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the response variable.
4. In the **Available variables** list, select the predictor variables:
   - To select a single variable, click the variable.
   - To select multiple variables, click the first variable then hold down the CTRL key and click each additional variable.
   - To select a range of variables, click the first variable then hold down the SHIFT key and click the last variable in the range.
5. Click **Add Predictor**.
6. Repeat steps *4* through *5* for each additional set of variables.
7. Optional: To assign a weight to each row for the analysis, in the **Weight** drop-down list, select the weight variable.
8. Click **Calculate**.

Performing ANOVA

Test if there is a difference between population means when a response variable is classified by one or more categorical variables (factors).

Note: Although the Fit Model analysis can be used to perform a simple 1-way ANOVA it is more intuitive to use **Compare Groups** > **ANOVA** analysis. Also, see the related tasks section for performing specific ANOVA models.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Fit Model**, and then click **ANOVA**.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the response variable.
4. In the **Available variables** list, select the factor variable(s):
   - To select a single variable, click the variable.
   - To select multiple variables, click the first variable then hold down the CTRL key and click each additional variable.
   - To select a range of variables, click the first variable then hold down the SHIFT key and click the last variable in the range.
5. Click **Add Factor** to add the simple factors.
6. To add the interaction between two or more factors, in the **Terms** list box, click the first term then hold down the CTRL key and click each additional term to include in the interaction, and then click **Cross**.
7. To remove a term, in the **Terms** list box, click the term and then click **Remove**.

   Note: It is sometimes quicker to use the **Factorial** command to build a factorial model and then remove any unnecessary terms.
8. To remove a factor and all terms using the factor, in the **Factor** list box, click the factor and then click **Remove**.

   When you select a factor, all terms that include the factor are selected. If you hold down the CTRL key and click additional factors, then all terms which include all the selected factors are selected.
9. Repeat steps *4* through *8* to build the model.
10. Click **Calculate**.

Performing 2-way or higher factorial ANOVA

Test if there is a difference between population means when a response variable is classified by two or more categorical variables (factors).

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Fit Model**, and then click **ANOVA**.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the response variable.
4. In the **Available variables** list, select the factor variables:
   - To select a single variable, click the variable.
   - To select multiple variables, click the first variable then hold down the CTRL key and click each additional variable.
   - To select a range of variables, click the first variable then hold down the SHIFT key and click the last variable in the range.
5. Click **Factorial**, and then click **Full** for a full factorial model, or click the highest order of interaction terms to include in the model.

6. Optional: To sort the terms by order (that is, all simple terms, all two-way interactions, then all three-way interactions...) rather than in the order the variables were added, click **Sort**.
7. Click **Calculate**.

Performing ANCOVA

Test if there is a difference between population means when a response variable is classified by one or more categorical variables (factors) while adjusting for the effect of one or more quantitative variables (covariates).

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Fit Model**, and then click **ANCOVA**.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the response variable.
4. In the **Available variables** list, select the variable(s):
   - To select a single variable, click the variable.
   - To select multiple variables, click the first variable then hold down the CTRL key and click each additional variable.
   - To select a range of variables, click the first variable then hold down the SHIFT key and click the last variable in the range.
5. To add the variables to the model as categorical factors, click **Add Factor**.
6. To add the variables to the model as quantitative covariates, click **Add Covariate**.
7. To add the interaction between two or more factors/covariates, in the **Terms** list box, click the first term then hold down the CTRL key and click each additional term to include in the interaction, and then click **Cross**.
8. To remove a term, in the **Terms** list box, click the term and then click **Remove**.

   Note: It is sometimes quicker to use the **Factorial** command to build a model and then remove any unnecessary terms.
9. To remove a factor/covariate and all terms using the factor/covariate, in the **Factors** or **Covariates** list box, click the factor/covariate and then click **Remove**.

   When you select a factor or covariate, all terms that include the factor or covariate are selected. If you hold down the CTRL key and click additional factors or covariates, then all terms which include all the selected factors or covariates are selected.
10. Repeat steps *4* through *9* to build the model.
11. Click **Calculate**.

Fitting an advanced linear model

Fit an advanced linear model to describe the relationship between many predictor variables and a response variable.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Fit Model**, and then click **Advanced**.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the response variable.
4. In the **Available variables** list, select the predictor variable(s):
   - To select a single variable, click the variable.
   - To select multiple variables, click the first variable then hold down the CTRL key and click each additional variable.

- To select a range of variables, click the first variable then hold down the SHIFT key and click the last variable in the range.

5. Optional: Click the drop-down menu arrow next to the **Available variables** list, and then select the measurement scale of the variable.

   If the measurement scale is not set, Analyse-it tries to determine if the variable is categorical or continuous scale. After you add a variable to a model, the icon alongside the variable indicates the assumed scale.
6. Click **Add** to add a variable, or click **Factorial** to add the factorial terms, or click **Polynomial** to add a polynomial terms.
7. To add the interaction between two or more variables, in the **Terms** list box, click the first term then hold down the CTRL key and click each additional term to include in the interaction, and then click **Cross**.
8. To remove a term, in the **Terms** list box, click the term and then click **Remove**.
9. Repeat steps *4* through *8* to build the model.
10. Optional: To assign a weight to each row for the analysis, in the **Weight** drop-down list, select the weight variable.
11. Click **Calculate**.

Scatter plot

A scatter plot shows the relationship between variables.

The scatter plot identifies the relationship that best describes the data, whether a straight line, polynomial or some other function.

A scatter plot matrix shows the relationship between each predictor and the response, and the relationship between each pair of predictors. You can use the matrix to identify the relationship between variables, to identify where additional terms such as polynomials or interactions are needed, and to see if transformations are needed to make the predictors or response linear. The scatter plot matrix does not convey the joint relationship between each predictor and the response since it does not take into account the effect on the response of the other variables in the model. Effect leverage and residual plots fulfill this purpose after fitting the model.

Summary of fit

R^2 and similar statistics measure how much variability is explained by the model.

R^2 is the proportion of variability in the response explained by the model. It is 1 when the model fits the data perfectly, though it can only attain this value when all sets of predictors are different. Zero indicates the model fits no better than the mean of the response. You should not use R^2 when the model does not include a constant term, as the interpretation is undefined.

For models with more than a single term, R^2 can be deceptive as it increases as you add more parameters to the model, eventually reaching saturation at 1 when the number of parameters equals the number of observations. Adjusted R^2 is a modification of R^2 that adjusts for the number of parameters in the model. It only increases when the terms added to the model improve the fit more than would be expected by chance. It is preferred when building and comparing models with a different number of parameters.

For example, if you fit a straight-line model, and then add a quadratic term to the model, the value of R^2 increases. If you continued to add more the polynomial terms until there are as many parameters as the number of observations, then the R^2 value would be 1. The adjusted R^2 statistic is designed to take into account the number of parameters in the model and ensures that adding the new term has some useful purpose rather than simply due to the number of parameters approaching saturation.

In cases where each set of predictor values are not unique, it may be impossible for the R^2 statistic to reach 1. A statistic called the maximum attainable R^2 indicates the maximum value that R^2 can achieve even if the model fitted perfectly. It is related to the pure error discussed in the lack of fit test.

The root mean square error (RMSE) of the fit, is an estimate of the standard deviation of the true unknown random error (it is the square root of the residual mean square). If the model fitted is not the correct model, the estimate is larger than the true random error, as it includes the error due to lack of fit of the model as well as the random errors.

Parameter estimates

Parameter estimates (also called coefficients) are the change in the response associated with a one-unit change of the predictor, all other predictors being held constant.

The unknown model parameters are estimated using least-squares estimation.

A coefficient describes the size of the contribution of that predictor; a near-zero coefficient indicates that variable has little influence on the response. The sign of the coefficient indicates the direction of the relationship, although the sign can change if more terms are added to the model, so the interpretation is not particularly useful. A confidence interval expresses the uncertainty in the estimate, under the assumption of normally distributed errors. Due to the central limit theorem, violation of the normality assumption is not a problem if the sample size is moderate.

- For quantitative terms, the coefficient represents the rate of change in the response per 1 unit change of the predictor, assuming all other predictors are held constant. The units of measurement for the coefficient are the units of response per unit of the predictor.

  For example, a coefficient for Height of 0.75, in a simple model for the response Weight (kg) with predictor Height (cm), could be expressed as 0.75 *kg per cm* which indicates a 0.75 kg weight increase per 1 cm in height.

  When a predictor is a logarithm transformation of the original variable, the coefficient is the rate of change in the response per 1 unit change in the log of the predictor. Commonly *base 2* log and *base 10* log are used as transforms. For *base 2* log the coefficient can be interpreted as the rate of change in the response when for a doubling of the predictor value. For *base 10* log the coefficient can be interpreted as the rate of change in the response when the predictor is multiplied by 10, or as the % change in the response per % change in the predictor.
- For categorical terms, there is a coefficient for each level:
  - For nominal predictors the coefficients represent the difference between the level mean and the grand mean.

    Analyse-it uses *effect coding* for nominal terms (also known as the *mean deviation coding*). The sum of the parameter estimates for a categorical term using effect coding is equal to 0.
  - For ordinal predictors, the coefficients represent the difference between the level mean and the baseline mean.

    Analyse-it uses *reference coding* for ordinal terms. The first level is used as the baseline or reference level.
- For the constant term, the coefficient is the response when all predictors are 0, and the units of measurement are the same as the response variable.

A standardized parameter estimate (commonly known as standardized beta coefficient) removes the unit of measurement of predictor and response variables. They represent the change in standard deviations of the response for 1 standard deviation change of the predictor. You can use them to compare the relative effects of predictors measured on different scales.

VIF, the variance inflation factor, represents the increase in the variance of the parameter estimate due to correlation (collinearity) between predictors. Collinearity between the predictors can lead to

unstable parameter estimates. As a rule of thumb, VIF should be close to the minimum value of 1, indicating no collinearity. When VIF is greater than 5, there is high collinearity between predictors.

A t-test formally tests the null hypothesis that the parameter is equal to 0, against the alternative hypothesis that it is not equal to 0. When the p-value is small, you can reject the null hypothesis and conclude that the parameter is not equal to 0 and it does contribute to the model.

When a parameter is not deemed to contribute statistically to the model, you can consider removing it. However, you should be cautious of removing terms that are known to contribute by some underlying mechanism, regardless of the statistical significance of a hypothesis test, and recognize that removing a term can alter the effect of other terms.

Effect of model hypothesis test

An F-test formally tests the hypothesis of whether the model fits the data better than no model.

It is common to test whether the model fits the data better than a null model that just fits the mean of the response. An analysis of variance table partitions the total variance in the response variable into the variation after fitting the full model (called the model error, or residual), the variation after fitting the null model, and the reduction by fitting the full model compared to the null model.

An F-test formally tests whether the reduction is statistically significant. The null hypothesis states that all the parameters except the intercept are zero against the alternative that at least one parameter is not equal to zero. When the p-value is small, you can reject the null hypothesis and conclude that at least one parameter is not zero.

ANOVA table

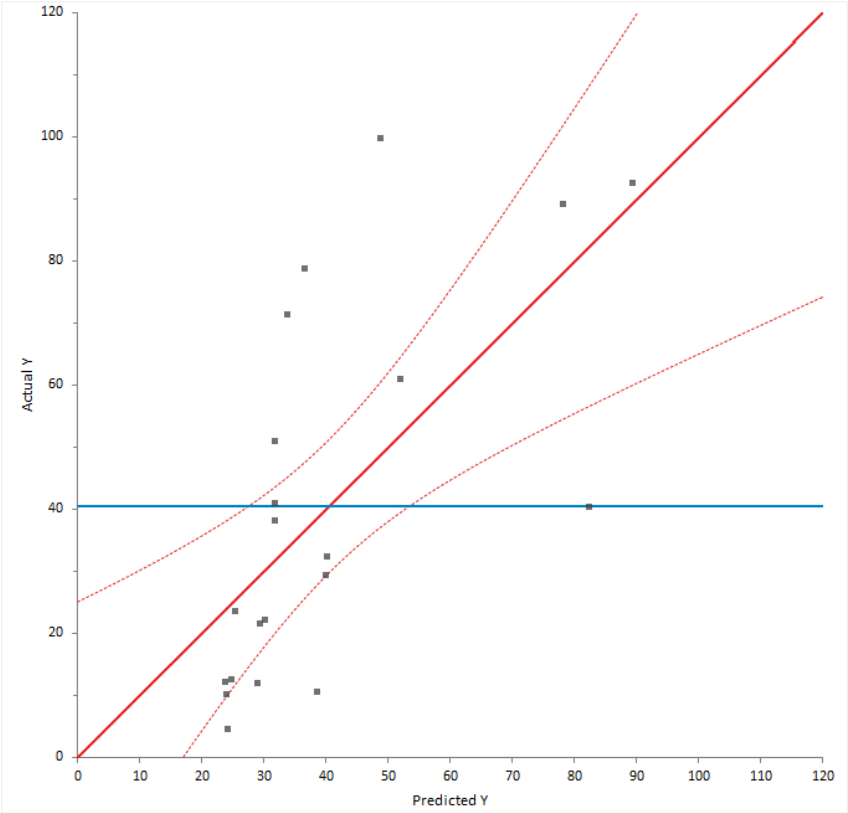
An analysis of variance (ANOVA) table shows the sources of variation.

An ANOVA table uses various abbreviations:

Abbreviation	Description
SS	Sum of squares, the sum of the squared deviations from the expected value.
DF	Degrees of freedom, the number of values that are free to vary.
MS	Mean square, the sum of squares divided by the degrees of freedom.
F	F-statistic, the ratio of two mean squares that forms the basis of a hypothesis test.
p-value	p-value, the probability of obtaining an F statistic at least as extreme as that observed when the null hypothesis is true.

Predicted against actual Y plot

A predicted against actual plot shows the effect of the model and compares it against the null model.



For a good fit, the points should be close to the fitted line, with narrow confidence bands. Points on the left or right of the plot, furthest from the mean, have the most leverage and effectively try to pull the fitted line toward the point. Points that are vertically distant from the line represent possible outliers. Both types of points can adversely affect the fit.

The plot is also a visualization of the ANOVA table, except each observation is shown so you can gain much more insight than just the hypothesis test. The confidence interval around the full model line portrays the F-test that all the parameters except the intercept are zero. When the confidence interval does not include the horizontal null model line, the hypothesis test is significant.

Lack of Fit

An F-test or X^2-test formally tests how well the model fits the data.

When the model fitted is correct the residual (model error) mean square provides and unbiased estimate of the true variance. If the model is wrong, then the mean square is larger than the true variance. It is possible to test for lack of fit by comparing the model error mean square to the true variance.

When the true variance is known, a X^2-squared test formally tests whether the model error is equal to the hypothesized value.

When the true variance is unknown, and there are multiple observations for each set of predictor values, an F-test formally tests whether there is a difference between pure error and the model error. The pure error is the pooled variance calculated for all unique sets of predictor values. It

makes no assumptions about the model but does assume that the variance is the same for each set of predictor values. An analysis of variance table shows the pure error, model error, and the difference between them called the lack of fit.

The null hypothesis states that the model error mean square is equal to the hypothesized value/ pure error, against the alternative that it is greater than. When the test p-value is small, you can reject the null hypothesis and conclude that there is a lack of fit.

Effect of terms hypothesis test

An F-test formally tests whether a term contributes to the model.

In most modeling analyses the aim is a model that describes the relationship using as few terms as possible. It is therefore of interest to look at each term in the model to decide if the term is providing any useful information.

An analysis of variance table shows the reduction in error by including each term. An F-test for each term is a formal hypothesis test to determine if the term provides useful information to the model. The null hypothesis states that the term does not contribute to the model, against the alternative hypothesis that it does. When the p-value is small, you can reject the null hypothesis and conclude that the term does contribute to the model.

When a term is not deemed to contribute statistically to the model, you may consider removing it. However, you should be cautious of removing terms that are known to contribute by some underlying mechanism, regardless of the statistical significance of a hypothesis test, and recognize that removing a term can alter the effect of other terms.

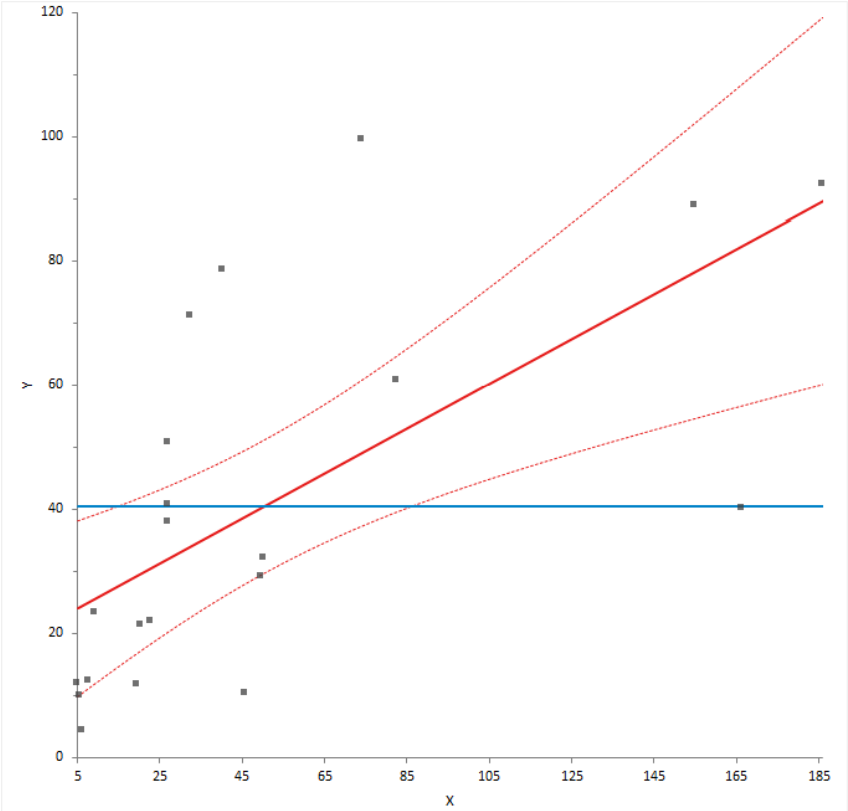
You should choose the type of F-test based on what other terms you want the test to consider.

Test	Purpose
Partial	Measures the effect of the term adjusted for all other terms in the model. The sum of squares in the ANOVA table is known as the Type III sum of squares. A partial F-test is equivalent to testing if all the parameter estimates for the term are equal to zero.
Sequential	Measures the effect of the term adjusting only for the previous terms included in the model. The sum of squares in the ANOVA table is known as the Type I sum of squares. A sequential F-test is often useful when fitting a polynomial regression.

Note: The squared t-statistic for a coefficient t-test is equivalent to the F statistic when using the partial F-test. The t-test is not suitable when the model includes categorical variables coded as dummy predictor variables as each term consists multiple coefficient t-tests.

Effect leverage plot

An effect leverage plot, also known as added variable plot or partial regression leverage plot, shows the unique effect of a term in the model.



A horizontal line shows the constrained model without the term; a slanted line shows the unconstrained model with the term. The plot shows the unique effect of adding a term to a model assuming the model contains all the other terms and the influence of each point on the effect of term hypothesis test. Points further from the horizontal line than the slanted line effectively try to make the hypothesis test more significant, and those closer to the horizontal than the slanted line try to make the hypothesis test less significant. When the confidence band fully encompasses the horizontal line, you can conclude the term does not contribute to the model. It is equivalent to a non-significant F-test for the partial effect of the term.

You can also use the plot to identify cases that are likely to influence the parameter estimates. Points furthest from the intersection of the horizontal and slanted lines have high leverage, and effectively try to pull the line towards them. A high leverage point that is distant from the bulk of the points can have a large influence the parameter estimates for the term.

Finally, you can use the plot to spot near collinearity between terms. When the terms are collinear, the points collapse toward a vertical line.

Effect means

Effect means are least-squares estimates predicted by the model for each combination of levels in a categorical term, adjusted for the other model effects.

Least-squares means are predictions made at specific values of the predictors. For each categorical variable in the term, the value is the combination of levels. For all other variables that are not part of the term, the values are neutral values. The neutral value for a quantitative variable is the mean. The neutral value of a nominal variable is the average over all levels which is equivalent to values

of 0 for each dummy predictor variable in an effect coding scheme. The neutral value of an ordinal variable is the reference level.

Plotting main effects and interactions

Plot the main effects and interactions to interpret the means.

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Terms** group, click **Effect Means**.
   The analysis task pane **Effect Means** panel opens.
3. To show main effect plots for each factor, click the **Main effect plots** check box.
4. To show interaction effect plots for each 2-way interaction and up to a single 3-way interaction, click the **Interaction effect plots** check box.
5. To show the means plot for individual terms, in the **Individual effects** list box, select the individual term check boxes, and then select the **Means plot** check box.
6. To show the means table for individual terms, in the **Individual effects** list box, select the individual term check boxes, and then select the **Means** check box.

   Note: For convenience, the **Means** and **Means plot** options apply to all selected **Individual effects**.
7. Click **Recalculate**.

Multiple comparisons

Multiple comparisons make simultaneous inferences about a set of parameters.

When making inferences about more than one parameter (such as comparing many means, or the differences between many means), you must use multiple comparison procedures to make inferences about the parameters of interest. The problem when making multiple comparisons using individual tests such as Student's t-test applied to each comparison is the chance of a type I error increases with the number of comparisons. If you use a 5% significance level with a hypothesis test to decide if two groups are significantly different, there is a 5% probability of observing a significant difference that is simply due to chance (a type I error). If you made 20 such comparisons, the probability that one or more of the comparisons is statistically significant simply due to chance increases to 64%. With 50 comparisons, the chance increases to 92%. Another problem is the dependencies among the parameters of interest also alter the significance level. Therefore, you must use multiple comparison procedures to maintain the simultaneous probability close to the nominal significance level (typically 5%).

Multiple comparison procedures are classified by the strength of inference that can be made and the error rate controlled. A test of homogeneity controls the probability of falsely declaring any pair to be different when in fact all are the same. A stronger level of inference is confident inequalities and confident directions which control the probability of falsely declaring any pair to be different regardless of the values of the others. An even stronger level is a set of simultaneous confidence intervals that guarantees that the simultaneous coverage probability of the intervals is at least 100(1-alpha)% and also have the advantage of quantifying the possible effect size rather than producing just a p-value. A higher strength inference can be used to make an inference of a lesser strength but not vice-versa. Therefore, a confidence interval can be used to perform a confidence directions/inequalities inference, but a test of homogeneity cannot make a confidence direction/inequalities inference.

The most well known multiple comparison procedures, Bonferroni and Šidák, are not multiple comparison procedures per se. Rather they are an inequality useful in producing easy to compute multiple comparison methods of various types. In most scenarios, there are more powerful procedures available. A useful application of Bonferroni inequality is when there are a small number of pre-planned comparisons. In these cases, you can use the standard hypothesis test or

confidence interval with the significance level (alpha) set to the Bonferroni inequality (alpha divided by the number of comparisons).

A side effect of maintaining the significance level is a lowering of the power of the test. Different procedures have been developed to maintain the power as high as possible depending on the strength of inference required and the number of comparisons to be made. All contrasts comparisons allow for any possible contrast; all pairs forms the k*(k-1)/2 pairwise contrasts, whereas with best forms k contrasts each with the best of the others, and against control forms k-1 contrasts each against the control group. You should choose the appropriate contrasts of interest before you perform the analysis, if you decide after inspecting the data, then you should only use all contrasts comparison procedures.

Multiple comparison procedures

Methods of controlling the Type I error and dependencies between point estimates when making multiple comparisons between effect means in a linear model.

Procedure	Purpose / Assumptions
Student's t (Fisher's LSD)	Compare the means of each pair of groups using the Student's t method. When making all pairwise comparisons this procedure is also known as unprotected Fisher's LSD, or when only performed following significant ANOVA F -test known as protected Fisher's LSD. Controls the Type I error rate individually for each contrast.
Tukey-Kramer	Compare the means of all pairs of groups using the Tukey-Kramer method. Controls the error rate simultaneously for all k(k+1)/2 contrasts.
Hsu	Compare the means of all groups against the best of the other groups using the Hsu method. Controls the error rate simultaneously for all k contrasts.
Dunnett	Compare the means of all groups against a control using the Dunnett method. Controls the error rate simultaneously for all k-1 contrasts.
Scheffé	Compare the means of all groups against all other groups using the Scheffé F method. Controls the error rate simultaneously for all possible contrasts.
Bonferroni	Not a multiple comparisons method. It is an inequality useful in producing easy to compute multiple comparisons. In most scenarios, there are more powerful procedures such as Tukey, Dunnett, Hsu. A useful application of Bonferroni inequality is when there are a small number of pre-planned comparisons. In this case, use the Student's t (LSD) method with the significance level (alpha) set to the Bonferroni inequality (alpha divided by the number of comparisons). In this scenario, it is usually less conservative than using Scheffé all contrast comparisons.

Comparing effect means

Compare the effect means to make inferences about the differences between them.

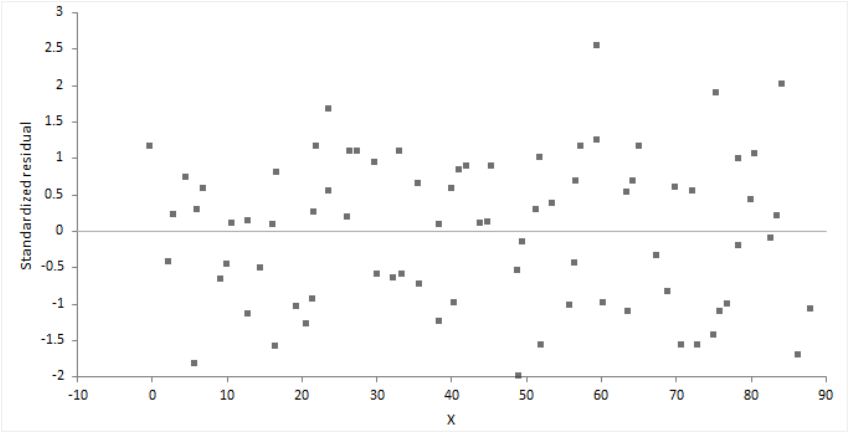
1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Terms** group, click **Effect Means**.
   The analysis task pane **Effect Means** panel opens.
3. In the **Individual effects** list box, select the individual term check box, and then select the **Multiple Comparisons** check box.
4. In the **Type** drop-down list, select the comparisons, and then:
   - If comparing with best, in the **Best** drop-down list, select the "best" group.
   - If comparing against a control, in the **Control** drop-down list, select the control group.
5. Optional: To sort the contrasts by the magnitude of the difference, select the **Ordered** check box.
6. Optional: In the **Method** drop-down list, select the multiple comparisons procedure, and in the **Confidence interval** edit box, type the confidence level as a percentage, or type - to suppress the confidence interval, and then in the drop-down list, select the confidence bounds.
7. Optional: To perform a hypothesis test to test if the difference is equal to 0, select the **Hypothesis test** check box, and then in the **Significance level** edit box, type the predefined significance level against.
8. Optional: To show a mean-mean scatter plot, select the **Mean-Mean** scatter plot check box.

   Note: For convenience the confidence interval, hypothesis test, and mean-mean scatter plot options apply to all selected **Individual effects**.
9. Click **Calculate**.

Residual plot

A residual plot shows the difference between the observed response and the fitted response values.

The ideal residual plot, called the null residual plot, shows a random scatter of points forming an approximately constant width band around the identity line.



It is important to check the fit of the model and assumptions – constant variance, normality, and independence of the errors, using the residual plot, along with normal, sequence, and lag plot.

Assumption	How to check
Model function is linear	The points form a pattern when the model function is incorrect. You might be able to transform variables or add polynomial and interaction terms to remove the pattern.
Constant variance	If the points tend to form an increasing, decreasing or non-constant width band, then the variance is not constant. You should consider transforming the response variable or incorporating weights into the model. When variance increases as a percentage of the response, you can use a log transform, although you should ensure it does not produce a poorly fitting model. Even with non-constant variance, the parameter estimates remain unbiased if somewhat inefficient. However, the hypothesis tests and confidence intervals are inaccurate.
Normality	Examine the normal plot of the residuals to identify non-normality. Violation of the normality assumption only becomes an issue with small sample sizes. For large sample sizes, the assumption is less important due to the central limit theorem, and the fact that the F- and t-tests used for hypothesis tests and forming confidence intervals are quite robust to modest departures from normality.
Independence	When the order of the cases in the dataset is the order in which they occurred: Examine a sequence plot of the residuals against the order to identify any dependency between the residual and time. Examine a lag-1 plot of each residual against the previous residual to identify a serial correlation, where observations are not independent, and there is a correlation between an observation and the previous observation. Time-series analysis may be more suitable to model data where serial correlation is present.

For a model with many terms, it can be difficult to identify specific problems using the residual plot. A non-null residual plot indicates that there are problems with the model, but not necessarily what these are.

Residuals - normality

Normality is the assumption that the underlying residuals are normally distributed, or approximately so.

While a residual plot, or normal plot of the residuals can identify non-normality, you can formally test the hypothesis using the Shapiro-Wilk or similar test.

The null hypothesis states that the residuals are normally distributed, against the alternative hypothesis that they are not normally-distributed. If the test p-value is less than the predefined significance level, you can reject the null hypothesis and conclude the residuals are not from a normal distribution. If the p-value is greater than the predefined significance level, you cannot reject the null hypothesis.

Violation of the normality assumption only becomes an issue with small sample sizes. For large sample sizes, the assumption is less important due to the central limit theorem, and the fact that

the F- and t-tests used for hypothesis tests and forming confidence intervals are quite robust to modest departures from normality.

Residuals - independence

Autocorrelation occurs when the residuals are not independent of each other. That is, when the value of e[i+1] is not independent from e[i].

While a residual plot, or lag-1 plot allows you to visually check for autocorrelation, you can formally test the hypothesis using the Durbin-Watson test. The Durbin-Watson statistic is used to detect the presence of autocorrelation at lag 1 (or higher) in the residuals from a regression. The value of the test statistic lies between 0 and 4, small values indicate successive residuals are positively correlated. If the Durbin-Watson statistic is much less than 2, there is evidence of positive autocorrelation, if much greater than 2 evidence of negative autocorrelation.

The null hypothesis states that the residuals are not autocorrelated, against the alternative hypothesis that they are. If the test p-value is less than the predefined significance level, you can reject the null hypothesis and conclude the residuals are correlated. If the p-value is greater than the predefined significance level, you cannot reject the null hypothesis.

Note: The p-value is computed using the bootstrap method and can take a long time to compute.

Plotting residuals

Plot the residuals to check the fit and assumptions of the model.

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Diagnostics** group, click **Residuals**, and then click:

Residual	Plot the residuals.
3-up	Plot the residuals and a histogram and normal plot.
5-up	Plot the residuals, a histogram and normal plot, a sequence plot, and a lag plot.

   The analysis task pane, the **Residuals** panel opens.
3. In the **Residuals** drop-down list, select:

Raw	Plot the raw residual Y - Y fitted.
Standardized	Plot the standardized residual (Y - Y fitted) / SE.
4. Optional: To test the normality of the residuals, On the **Analyse-it** ribbon tab, in the **Diagnostics** group, click **Test Normality**.
5. Optional: To test for autocorrelation amongst the residuals, On the **Analyse-it** ribbon tab, in the **Diagnostics** group, click **Test Autocorrelation**.
6. Click **Recalculate**.

Outlier and influence plot

An influence plot shows the outlyingness, leverage, and influence of each case.

The plot shows the residual on the vertical axis, leverage on the horizontal axis, and the point size is the square root of Cook's D statistic, a measure of the influence of the point.

Outliers are cases that do not correspond to the model fitted to the bulk of the data. You can identify outliers as those cases with a large residual (usually greater than approximately +/- 2), though not all cases with a large residual are outliers and not all outliers are bad. Some of the most interesting cases may be outliers.

Leverage is the potential for a case to have an influence on the model. You can identify points with high leverage as those furthest to the right. A point with high leverage may not have much influence on the model if it fits the overall model without that case.

Influence combines the leverage and residual of a case to measure how the parameter estimates would change if that case were excluded. Points with a large residual and high leverage have the most influence. They can have an adverse effect on (perturb) the model if they are changed or excluded, making the model less robust. Sometimes a small group of influential points can have an unduly large impact on the fit of the model.

Identifying outliers and other influential points

Plot measures to identify cases with large outliers, high leverage, or major influence on the fitted model.

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Diagnostics** group, click **Outliers and Influence**.
   The analysis task pane **Outliers, Leverage, Influence** panel opens.
3. Optional: To highlight and label cases that are large outliers, have high leverage or major influence on the fitted model, select the **Identify influential cases** check box.
4. Optional: To individually plot the Studentized residual, leverage, and Cook's D influence measures by observation number, select the **Index plots** check box.
5. Click **Recalculate**.

Points highlighted in red are deemed to be potentially influential to the fitted model.

Prediction

Prediction is the use of the model to predict the population mean or value of an individual future observation, at specific values of the predictors. Inverse prediction deals with the problem of predicting the value of a predictor for a given value of the response variable.

When making predictions, it is important that the data used to fit the model is similar to future populations to which you want to apply the prediction. You should be careful of making predictions outside the range of the observed data. Assumptions met for the observed data may not be met outside the range. Non-constant variance can cause confidence intervals for the predicted values to become unrealistically narrow or so large as to be useless. Alternatively, a different fit function may better describe the unobserved data outside the range.

When making multiple predictions of the population mean at different sets of predictor values the confidence intervals can be simultaneous or individual. A simultaneous interval ensures you achieve the confidence level simultaneously for all predictions, whereas individual intervals only ensure confidence for the individual prediction. With individual inferences, the chance of at least one interval not including the true value increases with the number of predictions.

Making predictions

Predict the value of the mean at specific values of the predictors, or the value of an individual future observation.

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, click **Predict**, and then click **Y**.
   The analysis task pane **Predict Y** panel opens.
3. In the **Terms** list box, under the #1 column, type the value for each predictor.

   You can leave a value blank in which case the average value is used. Leaving the value blank is useful in ANOVA models where you can predict the mean at specific combinations averaged over other effects.

4. To predict the expected mean, select the **Mean** check box, and in the **Confidence interval** edit box, type the confidence level. Finally, in the **Method** drop-down list, select **Simultaneous** for intervals that ensure you achieve the confidence level simultaneously for all predictions, or **Individual** for intervals only that ensure confidence for each individual prediction.
5. To predict the future value of an individual observation, select the **Individual** check box, and in the **Prediction interval** edit box, type the confidence level.
6. Optional: On the **Fit** panel, in the **Predicted values** drop-down list, select the style to plot the predicated values on the scatter plot.
7. Click **Recalculate**.

Making inverse predictions

Predict the values of the predictors at a specific value or values of the response variable.

You can only make inverse predictions when fitting a simple regression model.

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, click **Predict**, and then click **X given Y**.
   The analysis task pane **Inverse Predict X given Y** panel opens.
3. In the **Predictions** list box, under the Y column, type the values for the response variable.
4. Optional: In the **Confidence interval** edit box, type the confidence level, and then in the drop-down list, select the confidence bounds. Finally, in the **Method** drop-down list, select **Simultaneous** for intervals that ensure you achieve the confidence level simultaneously for all predictions, or **Individual** for intervals only that ensure confidence for each individual prediction.
5. Optional: On the **Fit** panel, in the **Predicted values** drop-down list, select the style to plot the predicated values on the scatter plot.
6. Click **Recalculate**.

Saving variables

Save the values computed by the analysis back to the dataset for further analysis.

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, click **Save Variable**, and then select the variable to store to the dataset.

   Note: If the **Save Variable** command is shaded, On the **Analyse-it** ribbon tab, in the **Report group**, click **Recalculate** to update the analysis.

Logistic / Probit fit

A model that describes the relationship between a categorical response variable and one or more explanatory variables using a logit or probit function.

Generalized linear models

A generalized linear model (GLM) is a generalization of the linear models by allowing the linear model to be related to the response variable via a link function and an error distribution other than a normal distribution. The unknown model parameters are estimated using maximum-likelihood estimation.

Type	Description
Logit / Logistic	Fit a model to a binary response variable expressed by the logit link function (log odds ratio) and binomial error distribution. **Note:** This model is very common as the parameter estimates can be interpreted as the log-odds or back transformed into an odds ratio.
Probit	Fit a model to a binary response variable expressed by the probit function and binomial error distribution.

Fitting a simple logistic regression

Fit a simple logistic regression model to describe the relationship between a single predictor variable and a binary response variable.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Fit Model**, and then click **Logit / Logistic**.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the binary response variable.
4. In the **Event** drop-down list, select the outcome of interest.
5. In the **X** drop-down list, select the predictor variable.
6. Click **Calculate**.

Fitting an advanced logistic model

Fit an advanced linear model to describe the relationship between many predictor variables and a binary response variable.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Fit Model**, and then click **Logit / Logistic**.
   The analysis task pane opens.
3. In the **Model** drop-down menu, select **Advanced**.
4. In the **Y** drop-down list, select the response variable.
5. In the **Event** drop-down list, select the outcome of interest.
6. In the **Available variables** list, select the predictor variable(s):
   - To select a single variable, click the variable.
   - To select multiple variables, click the first variable then hold down the CTRL key and click each additional variable.

- To select a range of variables, click the first variable then hold down the SHIFT key and click the last variable in the range.

7. Optional: Click the drop-down menu arrow next to the **Available variables** list, and then select the measurement scale of the variable.

   If the measurement scale is not set, Analyse-it tries to determine if the variable is categorical or continuous scale. After you add a variable to a model, the icon alongside the variable indicates the assumed scale.
8. Click **Add** to add a variable, or click **Factorial** to add the factorial terms, or click **Polynomial** to add a polynomial terms.
9. To add the interaction between two or more variables, in the **Terms** list box, click the first term then hold down the CTRL key and click each additional term to include in the interaction, and then click **Cross**.
10. To remove a term, in the **Terms** list box, click the term and then click **Remove**.
11. Repeat steps *6* through *10* to build the model.
12. Click **Calculate**.

Fitting a simple probit regression

Fit a simple probit regression model to describe the relationship between a single predictor variable and a binary response variable.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Fit Model**, and then click **Probit**.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the binary response variable.
4. In the **Event** drop-down list, select the outcome of interest.
5. In the **X** drop-down list, select the predictor variable.
6. Click **Calculate**.

Parameter estimates

Parameter estimates (also called coefficients) are associated with a one-unit change of the predictor, all other predictors being held constant.

A coefficient describes the size of the contribution of that predictor; a large coefficient indicates that the variable strongly influences the probability of that outcome, while a near-zero coefficient indicates that variable has little influence on the probability of that outcome. A positive sign indicates that the explanatory variable increases the probability of the outcome, while a negative sign indicates that the variable decreases the probability of that outcome. A confidence interval for each parameter shows the uncertainty in the estimate.

When the model contains categorical variables, the interpretation of the coefficients is more complex. For each term involving a categorical variable, a number of dummy predictor variables are created to predict the effect of each different level. There are different ways to code the predictors for a categorical variable, the most common method in logii/probit regression is called *reference cell coding* or *dummy coding*. In reference cell coding, the first category acts as a baseline, and you can interpret the other coefficients as an increase or decrease over the baseline category.

Odds ratio estimates

Odds ratios are the increase or decrease in odds associated with a change of the predictor, all other predictors been held constant.

Although mathematically the log odds ratios are easier to work with, when it comes to interpretation, it is more natural to use the odds ratios. Unlike the log odds ratio, the odds ratio is always positive. A value of 1 indicates no change. Values between 0 and less than 1 indicate a decrease in the probability of the outcome event. Values greater than 1 indicate an increase in the probability of the outcome event.

When the predictor variable is a categorical variable, the odds ratio is the increase or decrease in odds over the baseline category.

When a predictor variable is a continuous variable, the odds ratio is the increase or decrease in odds for a change in the predictor variable. The default is for a 1 unit change in the predictor, although it may be more appropriate to use a larger unit, such as for a change of 10 units of the predictor variable.

When a term includes an interaction, the odds ratio will depend on the interacting variables' values, so you will need to use multiple odd ratios to describe the behavior at different levels of interest.

Estimating odds ratios

Estimate the odds ratios using a logistic model.

Ensure you have followed the steps in *Fitting a simple logistic regression* on page 146 or *Fitting an advanced logistic model* on page 146.

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Fit Model** group, click **Odds Ratios**.
   The analysis task pane **Odds Ratios** panel opens.
3. In the **Odds Ratios** grid, in the **Comparison** column.: For a continuous variable, type the change in units of the predictor for the odd ratio, the default is for 1 unit change in the predictor, but for some predictors the odds ratio for a 10 or 100 unit change in the predictor maybe more interpretable. For a categorical variable, select the type of comparison:

All pairs	Compare all pairs of categories against each other. For example, with 3 categories the comparisons, are 1v2,1v3,2v1,2v3,3v1,3v2.
All distinct pairs	Compare all unique pairs of categories against each other. For example, with 3 categories the comparisons as 2v1,3v1,3v2).
Against reference	Compare all categories against the reference category (the first category in the categorical variable). For example, with 3 categories the comparisons are 3v1, 2v1.

4. If the model has interaction terms, in the **At** column for each continuous variable involved in an interaction, type the levels of interest as a comma (or your list delimiter) separated list. For example, if Age is a variable involved in an interaction with Gender, you may use 16, 25, 30, 40, and 60 as the levels of interest at which the Gender Female/Male odds ratios will be computed. For a continuous variable involved in an interaction with a categorical variable, the unit change for each level of the categorical variable will be computed. For example, if the Age unit is 10 years and Age is involved in an interaction with Gender, the odds ratio for a change of 10 years will be computed for Females and Males.
5. In the **Confidence interval** edit box, type the confidence level.
6. Click **Recalculate**.

Effect of model hypothesis test

A likelihood ratio or Wald X^2 test formally tests the hypothesis of whether the model fits the data better than no model.

It is common to test whether the model fits the data better than the null model with no parameters.

A X^2 test formally tests whether the reduction is statistically significant. The null hypothesis states that all the parameters for the covariates are zero against the alternative that at least one parameter is not equal to zero. When the p-value is small, you can reject the null hypothesis and conclude that at least one parameter is not zero.

Effect of term hypothesis test

A likelihood ratio or Wald X^2 test formally tests the hypothesis of whether a term contributes to the model.

In most modeling analyses the aim is a model that describes the relationship using as few terms as possible. It is therefore of interest to look at each term in the model to decide if the term is providing any useful information.

A X^2 test for each term is a formal hypothesis test to determine if the term provides useful information to the model. The null hypothesis states that the term does not contribute to the model, against the alternative hypothesis that it does. When the p-value is small, you can reject the null hypothesis and conclude that the term does contribute to the model.

When a term is not deemed to contribute statistically to the model, you may consider removing it. However, you should be cautious of removing terms that are known to contribute by some underlying mechanism, regardless of the statistical significance of a hypothesis test, and recognize that removing a term can alter the effect of other terms.

Study design

Fit model analysis study requirements and dataset layout.

Requirements

- 1 or more categorical or quantitative predictor variables.
- A categorical or quantitative response variable.

Dataset layout

Use a column for each predictor variable (Height, Sex) and a column for the response variable (Weight); each row has the values of the variables for a case (Subject).

Subject (optional)	Height	Sex	Weight
1	175	M	65
2	180	M	70
3	160	F	90
4	190	F	55
5	180	M	100
6	150	F	55
7	140	M	75
8	160	M	80
9	165	F	80
10	180	M	95
…	…	…	…

Chapter

Method comparison / Agreement

18

Method comparison measures the closeness of agreement between the measured values of two methods.

Note: The term method is used as a generic term and can include different measurement procedures, measurement systems, laboratories, or any other variable that you want to if there are differences between measurements.

Correlation coefficient

A correlation coefficient measures the association between two methods.

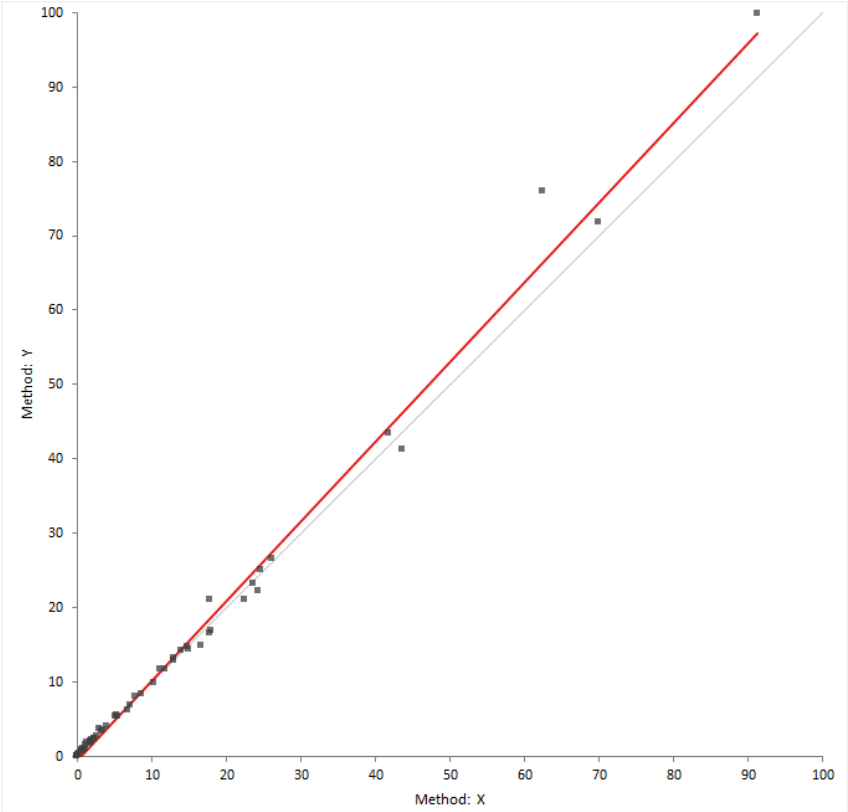
The correlation coefficient is probably the most commonly reported statistic in method comparison studies. However, it is irrelevant for a number of reasons *(Bland & Altman, 1986)*:

- It is a measure of the strength of linear association between two methods, the extent to which as one variable increases the other variable also tends to increase, not the agreement between them.
- A change in the scale of measurement does not affect the correlation, even though it affects the agreement. For example, if one method reports double the value of another method the correlation coefficient would still be high even though the agreement between the methods is poor.
- It simply represents the ratio of variation between the subjects relative to the measurement variation. The measuring interval chosen in study design can affect the correlation coefficient.

The correlation coefficient is sometimes re-purposed as an adequate range test *(CLSI, 2002)* on the basis that the ratio of variation between subjects, relative to measurement variation, is an indicator of the quality of the data *(Stöckl & Thienpont, 1998)*. When the correlation coefficient is greater than 0.975 the parameters of an ordinary linear regression are not significantly biased by the error in the X variable, and so linear regression is sometimes recommended. However, with the wide range of proper regression procedures available for analyzing method comparison studies, there is little need to use inappropriate models.

Scatter plot

A scatter plot shows the relationship between two methods.



The scatter plot shows measured values of the reference or comparison method on the horizontal axis, against the test method on the vertical axis.

The relationship between the methods may indicate a constant, or proportional bias, and the variability in the measurements across the measuring interval. If the points form a constant-width band, the method has a constant standard deviation (constant SD). If the points form a band that is narrower at small values and wider at large values, there is a constant relationship between the standard deviation and value, and the method has constant a coefficient of variation (CV). Some measurement procedures exhibit constant SD in the low range and constant CV in the high range.

If both methods measure on the same scale, a gray identity line shows ideal agreement and is useful for comparing the relationship against.

Fit Y on X

Regression of Y on X describes the linear relationship between the methods.

Ordinary linear regression

Ordinary linear regression fits a line describing the relationship between two variables assuming the X variable is measured without error.

Ordinary linear regression finds the line of best fit by minimizing the sum of the vertical distances between the measured values and the regression line. It, therefore, assumes that the X variable is measured without error.

Weighted linear regression is similar to ordinary linear regression but weights each item to take account of the fact that some values have more measurement error than others. Typically, for

methods that exhibit constant CV the higher values are less precise and so weights equal to $1 / X^2$ are given to each observation.

Although the assumption that there is no error in the X variable is rarely the case in method comparison studies, ordinary and weighted linear regression are often used when the X method is a reference method. In such cases the slope and intercept estimates have little bias, but hypothesis tests and confidence intervals are inaccurate due to an underestimation of the standard errors *(Linnet, 1993)*. We recommend the use of correct methods such as Deming regression.

Deming regression

Deming regression is an errors-in-variables model that fits a line describing the relationship between two variables. Unlike ordinary linear regression, it is suitable when there is measurement error in both variables.

Deming regression *(Cornbleet & Gochman, 1979)* finds the line of best fit by minimizing the sum of the distances between the measured values and the regression line, at an angle specified by the variance ratio. It assume both variables are measured with error. The variance ratio of the errors in the X / Y variable is required and assumed to be constant across the measuring interval. If you measure the items in replicate, the measurement error of each method is estimated from the data and the variance ratio calculated.

In the case where the variance ratio is equal to 1, Deming regression is equivalent to orthogonal regression. When only single measurements are made by each method and the ratio of variances is unknown, a variance ratio of 1 is sometimes used as a default. When the range of measurements is large compared to the measurement error this can be acceptable. However in cases where the range is small the estimates are biased and the standard error underestimated, leading to incorrect hypothesis tests and confidence intervals *(Linnet, 1998)*.

Weighted Deming regression *(Linnet, 1990)* is a modification of Deming regression that assumes the ratio of the coefficient of variation (CV), rather than the ratio of variances, is constant across the measuring interval.

Confidence intervals for parameter estimates use a t-distribution and the standard errors are computed using a Jackknife procedure *(Linnet, 1993)*.

Passing-Bablok regression

Passing-Bablok regression fits a line describing the relationship between two variables. It is robust, non-parametric, and is not sensitive to outliers or the distribution of errors.

Passing-Bablok regression *(Passing & Bablok, 1983)* finds the line of best fit using a shifted median of all possible pairwise slopes between points. It does not make assumptions of the distribution of the measurement errors, and the variance of the measurement errors need not remain constant over the measuring interval though their ratio should remain proportional to $\#^2$ (the slope squared, in many cases $\# \approx 1$).

There is a modification to the procedure for use when the methods measure on different scales, or where the purpose is to transform results from one method to another rather than to compare two methods for equality *(Passing & Bablok, 1988)*.

Confidence intervals for parameter estimates are based on a normal approximation or bootstrap. Confidence curves around the regression line and at specific points on the line are obtained by bootstrap.

Fitting ordinary linear regression

Fit a ordinary least square regression to estimate the relationship between a test method and a reference method when the reference method is measured without error.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Method Comparison**, and then click the fit:

Ordinary Least Square	Fit an ordinary regression where the test method measurement error SD is constant throughout the measuring interval.
Weighted Least Squares	Fit an ordinary regression where the test method measurement error CV is constant throughout the measuring interval.

   The analysis task pane opens.
3. If the data are in 2 variables:
   a) In the **X** drop-down list, select the comparative or reference measurement procedure variable.
   b) In the **Y** drop-down list, select the test measurement procedure variable.

   Note: If the variables consist of replicate measurements, select the variable name that spans all the replicate columns.
4. If the data are in 2 variables with a separate variable identifying replicates of each item:
   a) In the **X** drop-down list, select the comparative or reference measurement procedure variable.
   b) In the **Y** drop-down list, select the test measurement procedure variable.
   c) In the **Item** drop-down list, select the variable identifying each item.
5. If the data are in a single variable with a separate variable matching each item and a variable identifying the method:
   a) In the **Model** drop-down menu, select **Matched Pairs**.
   b) In the **Y** drop-down list, select the measurement variable.
   c) In the **Item** drop-down list, select the item variable that identifies each item.
   d) In the **Method** drop-down list, select the method variable.
6. If the items are measured in replicate, in the **Replicates** group, select:

1st X, 1st Y	Uses only the 1st X replicate and 1st Y replicate in the regression.
Mean X, 1st Y	Uses the Mean of X replicates and the 1st Y replicate in the regression.
Mean X, Mean Y	Uses the Mean of the X replicates and the Mean of the Y replicates in the regression.

   Note: All items must have the same number of replicates. Items that do not are excluded from analysis.
7. Click **Calculate**.

Fitting Deming regression

Fit a Deming regression to estimate the relationship between a test method and a reference or comparative method when both variables are measured with error.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Method Comparison**, and then click the fit:

Ordinary Deming	Fit a Deming regression where the ratio of measurement error SD is constant throughout the measuring interval.

Weighted Deming	Fit a Deming regression where the ratio of measurement error CV is constant throughout the measuring interval.

The analysis task pane opens.

3. If the data are in 2 variables:
   a) In the **X** drop-down list, select the comparative or reference measurement procedure variable.
   b) In the **Y** drop-down list, select the test measurement procedure variable.

   Note: If the variables consist of replicate measurements, select the variable name that spans all the replicate columns.
4. If the data are in 2 variables with a separate variable identifying replicates of each item:
   a) In the **X** drop-down list, select the comparative or reference measurement procedure variable.
   b) In the **Y** drop-down list, select the test measurement procedure variable.
   c) In the **Item** drop-down list, select the variable identifying each item.
5. If the data are in a single variable with a separate variable matching each item and a variable identifying the method:
   a) In the **Model** drop-down menu, select **Matched Pairs**.
   b) In the **Y** drop-down list, select the measurement variable.
   c) In the **Item** drop-down list, select the item variable that identifies each item.
   d) In the **Method** drop-down list, select the method variable.
6. If the items are measured in replicate, in the **Replicates** group, select:

1st X, 1st Y	Uses only the 1st X replicate and 1st Y replicate in the regression.
Mean X, 1st Y	Uses the Mean of X replicates and the 1st Y replicate in the regression.
Mean X, Mean Y	Uses the Mean of the X replicates and the Mean of the Y replicates in the regression.

   Note: All items must have the same number of replicates. Items that do not are excluded from analysis.
7. If the items are not measured in replicate, in the **Variance ratio**, edit box, type the ratio of the variances, or in the **SD/CV X** and **SD/CV Y** edit boxes, type the standard deviation or CV of the measurement error for each method.

   Note: The regression procedure uses all replicate measurements to estimate the precision of each method.
8. Click **Calculate**.

Fitting Passing-Bablok regression

Fit a Passing-Bablok regression to estimate the relationship between a test method and a reference or comparative method.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Method Comparison**, and then click **Passing-Bablok**.
   The analysis task pane opens.
3. If the data are in 2 variables:

a) In the **X** drop-down list, select the comparative or reference measurement procedure variable.
b) In the **Y** drop-down list, select the test measurement procedure variable.

Note: If the variables consist of replicate measurements, select the variable name that spans all the replicate columns.

4. If the data are in 2 variables with a separate variable identifying replicates of each item:
   a) In the **X** drop-down list, select the comparative or reference measurement procedure variable.
   b) In the **Y** drop-down list, select the test measurement procedure variable.
   c) In the **Item** drop-down list, select the variable identifying each item.
5. If the data are in a single variable with a separate variable matching each item and a variable identifying the method:
   a) In the **Model** drop-down menu, select **Matched Pairs**.
   b) In the **Y** drop-down list, select the measurement variable.
   c) In the **Item** drop-down list, select the item variable that identifies each item.
   d) In the **Method** drop-down list, select the method variable.
6. If the items are measured in replicate, in the **Replicates** group, select:

1st X, 1st Y	Uses only the 1st X replicate and 1st Y replicate in the regression.
Mean X, 1st Y	Uses the Mean of X replicates and the 1st Y replicate in the regression.
Mean X, Mean Y	Uses the Mean of the X replicates and the Mean of the Y replicates in the regression.

Note: All items must have the same number of replicates. Items that do not are excluded from analysis.

7. In the **Method** drop-down list, select:

Part I	When the methods are measured on the same scale and the purpose is to compare if the methods are equal.
Part III	When the methods are measured on different scales, or the purpose is to convert the values from one method to the other using the regression formula.

8. Click **Calculate**.

Linearity

Linearity is the assumption that the relationship between the methods is linear.

The regression procedures used in method comparison studies assume the relationship between the methods is linear. A CUSUM is a measure of the linearity, defined as a running sum of the number of observations above and below the fitted regression line. When the relationship is linear it is expected the points above and below the line are randomly scattered, and the CUSUM statistic is small. Clusters of points on one side of the regression line produce a large CUSUM statistic.

A formal hypothesis test for linearity is based on the largest CUSUM statistic and the Kolmogorov-Smirnov test. The null hypothesis states that the relationship is linear, against the alternative hypothesis that it is not linear. When the test p-value is small, you can reject the null hypothesis and conclude that the relationship is nonlinear.

Residual plot

A residual plot shows the difference between the measured values and the predicted values against the true values.

The residual plot shows disagreement between the data and the fitted model. The ideal residual plot (called the null residual plot) shows a random scatter of points forming an approximately constant width band around the identity line.

It is important to check the fit of the model and the assumptions:

Assumption	How to check
Model function is linear	The points will form a pattern when the model function is not linear.
Constant variance	If the points tend to form an increasing, decreasing, or non-constant width band, the variance is not constant and you should consider using weighted regression.
Normality	A histogram of the residuals should form a normal distribution. This is an assumption of linear regression. Deming regression with Jacknife standard errors is robust to this assumption. Passing-Bablok regression is non-parametric and this assumption does not apply.

Checking the assumptions of the fit

Inspect the residual and linearity plots to ensure the assumptions of the regression procedure are met before using the analysis results to make inferences.

You must have already completed either of the tasks:

- *Fitting ordinary linear regression* on page 153
- *Fitting Deming regression* on page 154
- *Fitting Passing-Bablok regression* on page 155

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Method Comparison** group, click **Residuals**.
3. On the **Analyse-it** ribbon tab, in the **Method Comparison** group, click **Linearity**.
4. Click **Recalculate**.

Average bias

Bias is a measure of a systematic measurement error, the component of measurement error that remains constant in replicate measurements on the same item. When measuring a method against a reference method using many items the average bias is an estimate of bias that is averaged over all the items.

Bias is the term used when a method is compared against a reference method. When the comparison is not against a reference method but instead another routine comparative laboratory method, it is simply an average difference between methods rather than an average bias. For clarify of writing we will use the term average bias.

The average bias is usually expressed as the constant and proportional bias from a regression procedure, or as a constant or proportional bias from the mean of the differences or relative differences. If there are other sources systematic errors present, such as nonlinearity or interferences, the average bias will be incorrect.

The average bias is an estimate of the true unknown average bias in a single study. If the study were repeated, the estimate would be expected to vary from study to study. Therefore, if a

single estimate is compared directly to 0 or compared to the allowable bias the statement is only applicable to the single study. To make inferences about the true unknown bias you must perform a hypothesis test:

There are two common hypotheses of interest that can be tested:

- Equality test

  The null hypothesis states that the bias is equal to 0, against the alternative hypothesis that it is not equal zero. When the test p-value is small, you can reject the null hypothesis and conclude that the bias is different to zero.

  It is important to remember that a statistically significant p-value tells you nothing about the practical importance of what was observed. For a large sample, the bias for a statistically significant hypothesis test may be so small as to be practically useless. Conversely, although there may some evidence of bias, the sample size may be too small for the test to reach statistical significance, and you may miss an opportunity to discover a true meaningful bias. Lack of evidence against the null hypothesis does not mean it has been proven to be true, the belief before you perform the study is that the null hypothesis is true and the purpose is to look for evidence against it. An equality test at the 5% significance level is equivalent to testing if the 95% confidence interval includes zero.
- Equivalence test

  The null hypothesis states that the bias is outside an interval of practical equivalence, against the alternative hypothesis that the bias is within the interval considered practically equivalent. When the test p-value is small, you can reject the null hypothesis and conclude that the bias is practically equivalent, and within the specified interval.

  An equivalence test is used to prove a bias requirement can be met. The null hypothesis states the methods are not equivalent and looks for evidence that they are in fact equivalent. An equivalence hypothesis test at the 5% significance level is the same as testing if the 90% confidence interval lies within the allowable bias interval.

Estimating the bias between methods at a decision level

Estimate the average bias (or average difference) at a decision level using the regression fit.

You must have already completed either of the tasks:

- *Fitting ordinary linear regression* on page 153
- *Fitting Deming regression* on page 154
- *Fitting Passing-Bablok regression* on page 155

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Method Comparison** group, click **Predict At**.
   The analysis task pane shows the **Comparability** task.
3. In the **Decision level** grid, under the **Id** column, type an identifier for the decision level, and then under the **X** column, type the value of the decision level
4. Optional: To test if the two methods are equal at each decision level:
   a) On the **Analyse-it** ribbon tab, in the **Method Comparison** group, click **Test Equality**.
      The analysis task pane shows the **Comparability** task.
   b) To compare the p-value against a predefined significance level, in the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%).
5. Optional: To test if two methods are practically equivalent at each decision level:
   a) On the **Analyse-it** ribbon tab, in the **Method Comparison** group, click **Test Equivalence**.
      The analysis task pane shows the **Comparability** task.

b) If the allowable difference is a constant or proportional value across the measuring interval, in the **Allowable difference** group, select **Across measuring interval**, and then in the **Absolute** edit box, type the bias in measurement units, and/or in the **Relative** edit box, type the bias as a percentage (suffix with % symbol).
c) If the allowable difference varies for each level, in the **Allowable difference** group, select **Each level** and then in the **Decision level** grid, under the **Allowable difference** column, alongside each level, type the absolute bias in measurement units, or the relative bias as a percentage (suffix with % symbol).
d) To compare the p-value against a predefined significance level, in the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%).

6. Click **Recalculate**.

Testing commutability of other materials

Test the commutability of reference materials or processed samples with unprocessed samples.

You must fit a Deming regression to perform this task.

1. On the **Analyse-it** ribbon tab, in the **Method Comparison** group, click **Restrict to Group**.
2. In the **Group / Color / Symbol** drop-down list, select the variable indicating the type of sample.
3. In the **Restrict fit to group** drop-down list, select the group of unprocessed samples.
4. On the **Fit Y on X** task pane, in the **Prediction band** edit box, type 95%, and if required select the **Familywise coverage** check box.
5. Click **Recalculate**.

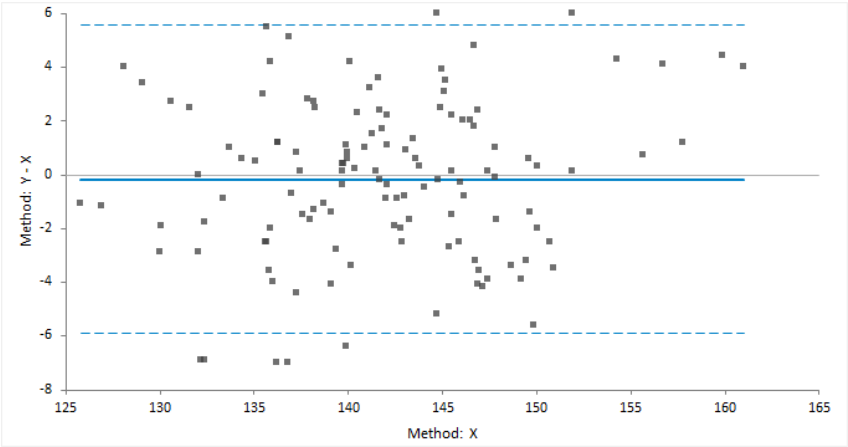
The scatter plot shows the points in different colors depending on the type of sample. The fit is restricted to a unprocessed samples and the prediction band includes any additional samples that are commutable with the unprocessed samples. A table lists the additional samples and if they are commutable with the unprocessed samples.

Difference plot (Bland-Altman plot)

A difference plot shows the differences in measurements between two methods, and any relationship between the differences and true values.

Bland and Altman (1983) popularized the use of the difference plot for assessing agreement between two methods. Although their use was to show the limits of agreement between two

methods, the plot is also widely used for assessing the average bias (average difference) between two methods.



The classic difference plot shows the difference between the methods on the vertical axis, against the best estimate of the true value on the horizontal axis. When one method is a reference method, it is used as the best estimate of the true value and plotted on the horizontal axis *(Krouwer, 2008)*. In other cases, using the average of the methods as the best estimate of the true value, to avoid an artificial relationship between the difference and magnitude *(Bland & Altman, 1995)*.

A relative difference plot *(Pollock et al., 1993)* shows the relative differences on the vertical axis, against the best estimate of the true value on the horizontal axis. It is useful when the methods show variability related to increasing magnitude, that is where the points on a difference plot form a band starting narrow and becoming wider as X increases. Another alternative is to plot the ratio of the methods on the vertical axis, against the best estimate of the true value on the horizontal axis.

Fit differences

Regression of the differences on the true value describes the relationship between the methods.

Mean difference measures the constant relationship between the variables. An assumption is that the difference is not relative to magnitude across the measuring interval. If the differences are related to the magnitude, the relationship should be modeled by using relative differences or regressing the differences on the true value.

Plotting a difference plot and estimating the average bias

Plot a difference plot and estimate the average bias (average difference) between the methods.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Method Comparison**, and then click:

Mean difference	Estimate the average difference using the mean of the differences.
Median difference	Estimate the average difference using the median of the differences. Useful when the distribution of the differences is not normally distributed

   The analysis task pane opens.
3. If the data are in 2 variables:

a) In the **X** drop-down list, select the comparative or reference measurement procedure variable.
b) In the **Y** drop-down list, select the test measurement procedure variable.

Note: If the variables consist of replicate measurements, select the variable name that spans all the replicate columns.

4. If the data are in 2 variables with a separate variable identifying replicates of each item:
   a) In the **X** drop-down list, select the comparative or reference measurement procedure variable.
   b) In the **Y** drop-down list, select the test measurement procedure variable.
   c) In the **Item** drop-down list, select the variable identifying each item.
5. If the data are in a single variable with a separate variable matching each item and a variable identifying the method:
   a) In the **Model** drop-down menu, select **Matched Pairs**.
   b) In the **Y** drop-down list, select the measurement variable.
   c) In the **Item** drop-down list, select the item variable that identifies each item.
   d) In the **Method** drop-down list, select the method variable.
6. In the **D** drop-down list, select:

Difference	The differences are not related to magnitude.
Relative Difference	The differences are related to magnitude.

7. In the **Z** drop-down list, select:

X	X is a reference method.
(X+Y)/2	Neither X or Y are reference methods.

8. Click **Calculate**.

Limits of agreement (LoA)

Limits of agreement estimate the interval within which a proportion of the differences between measurements lie.

The limits of agreement includes both systematic (bias) and random error (precision), and provide a useful measure for comparing the likely differences between individual results measured by two methods. When one of the methods is a reference method, the limits of agreement can be used as a measure of the total error of a measurement procedure *(Krouwer, 2002)*.

Limits of agreement can be derived using parametric method given the assumption of normality of the differences. or using non-parametric percentiles when such assumptions do not hold.

Plotting the Bland-Altman limits of agreement

Plot a difference plot and limits of agreement.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Method Comparison**(or **Agreement** depending on the edition), and then click **Bland-Altman**. The analysis task pane opens.
3. If the data are in 2 variables:
   a) In the **X** drop-down list, select the comparative or reference measurement procedure variable.
   b) In the **Y** drop-down list, select the test measurement procedure variable.

Note: If the variables consist of replicate measurements, select the variable name that spans all the replicate columns.

4. If the data are in 2 variables with a separate variable identifying replicates of each item:
   a) In the **X** drop-down list, select the comparative or reference measurement procedure variable.
   b) In the **Y** drop-down list, select the test measurement procedure variable.
   c) In the **Item** drop-down list, select the variable identifying each item.
5. If the data are in a single variable with a separate variable matching each item and a variable identifying the method:
   a) In the **Model** drop-down menu, select **Matched Pairs**.
   b) In the **Y** drop-down list, select the measurement variable.
   c) In the **Item** drop-down list, select the item variable that identifies each item.
   d) In the **Method** drop-down list, select the method variable.
6. In the **D** drop-down list, select:

Difference	Plot the differences on the vertical axis.
Relative Difference	Plot the differences relative to Z on the vertical axis.
Ratio	Plot the ratio of the values on the vertical axis.

7. In the **on Z** drop-down list, select:

X	Plot X as a reference method on the horizontal axis.
(X+Y)/2	Plot the average on the horizontal axis. Neither X or Y are reference methods.

8. In the **Mean function** drop-down list, select:

Constant	There is a constant relationship between D and Z.
Linear function	There is a linear relationship between D and Z.

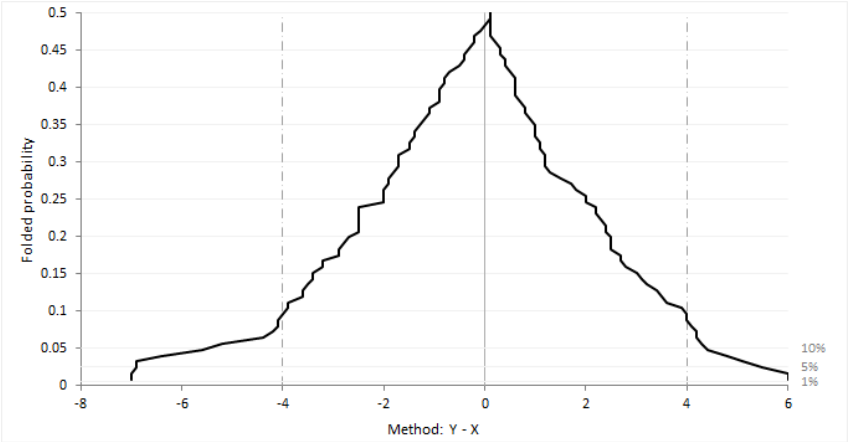
9. In the **SD** drop-down list, select:

Constant	The measurement error is constant throughout the measuring interval.
Linear function	The measurement error is linear across the measuring interval.

10. Click **Calculate**.

Mountain plot (folded CDF plot)

A mountain plot shows the distribution of the differences between two methods. It is a complementary plot to the difference plot.



Krouwer and Monti (1995) devised the mountain plot (also known as a folded empirical cumulative distribution plot) as a complementary representation of the difference plot. It shows the distribution of the differences with an emphasis on the center and the tails of the distribution. You can use the plot to estimate the median of the differences, the central 95% interval, the range, and the percentage of observations outside the total allowable error bands.

The plot is simply the empirical cumulative distribution function of the differences folded around the median (that is, the plotted function = p where p < 0.5 otherwise 1-p). Unlike the histogram it is unaffected by choice of class intervals, however it should be noted that although the mountain plot looks like a frequency polygon it does not display the density function. It has recently been proven that the area under the plot is equal to the mean absolute deviation from the median (Xue and Titterington, 2010).

Plotting a mountain plot

Plot a mountain plot to see the distribution of the differences between two methods.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Method Comparison** (or **Agreement** depending on the edition), and then click **Mountain**.
   The analysis task pane opens.
3. If the data are in 2 variables:
   a) In the **X** drop-down list, select the comparative or reference measurement procedure variable.
   b) In the **Y** drop-down list, select the test measurement procedure variable.

   Note: If the variables consist of replicate measurements, select the variable name that spans all the replicate columns.
4. If the data are in 2 variables with a separate variable identifying replicates of each item:
   a) In the **X** drop-down list, select the comparative or reference measurement procedure variable.
   b) In the **Y** drop-down list, select the test measurement procedure variable.
   c) In the **Item** drop-down list, select the variable identifying each item.
5. If the data are in a single variable with a separate variable matching each item and a variable identifying the method:

a) In the **Model** drop-down menu, select **Matched Pairs**.
b) In the **Y** drop-down list, select the measurement variable.
c) In the **Item** drop-down list, select the item variable that identifies each item.
d) In the **Method** drop-down list, select the method variable.

6. If the data are measured in replicate and the X method is a reference method and the Y method is a comparative method, select the **Mean X, 1st Y replicate** check box so that the differences represent the difference between an individual test result by the Y method and the average of the replicates for the X method. These differences are the best representation of the total error in the test method, by using the mean of the X method it reduces the amount of random error in the X result so it reflects the true value of the reference method to compare the test method against. If you select **Mean X, Mean Y**, the differences will only represent the some of the error, namely the systematic error due to bias with some of the random error removed due to averaging of replicates. Likewise, if you select **1st X, replicate, 1st Y replicate** the differences will include any random error in the X method so not reflect the error just in the test method. There may be occasions when these other sources of error are of interest, but generally the interest is in the total error of the test method.
7. Click **Calculate**.

Partitioning and reducing the measuring interval

Partition or reduce the measuring interval to fit different functions.

You may need to reduce the measuring interval if a shows nonlinearity at the ends of the measuring interval. Or, you may need to partition the measuring interval if the differences show constant relationship at small magnitude and proportional relationship at higher values.

1. Activate the analysis report worksheet.
2. To reduce the range of the measuring interval, On the **Analyse-it** ribbon tab, in the **Method Comparison** group, click **Adjust Interval**.
   The analysis task pane shows the **Comparability** task.
3. To partition the measuring interval, On the **Analyse-it** ribbon tab, in the **Method Comparison** group, click **Partition**.
4. In the **Measuring interval(s)** grid, under either or both the **Lower** and **Upper** column, in the edit boxes, type the lower and upper measuring interval or partition.

   Note: If you edit the upper/lower limit of a partition, the lower/upper limit of the next partition is set to the same value respectively.
5. To only plot points in each interval on the analysis reports, clear the **Show data points outside interval** check box. If you select this option, the points are shaded gray on plots.
6. Change the options on the tasks.

   Note: Each partition has its own task settings. If you want to use the same settings for a number of partitions, it is best to configure the settings before you partition the measuring interval.
7. Repeat steps *3* through *6* for each partition.
8. Click **Recalculate**.

The analysis creates a separate analysis report for each partition. You can quickly switch between analysis reports for each partition by selecting the partition in the partition grid.

Agreement measures for binary and semi-quantitative data

Agreement measures summarize the similarity of the results of two binary or semi-quantitative methods.

Asymmetric agreement measures are dependent on the assignment of X and Y variables. They are often useful due to the natural interpretation as the proportion of the comparative method (X) results in which the new method (Y) results are the same.

For a binary test, with positive/negative results, the results can be expressed as a 2x2 contingency table:

	Y, Test		
X, Comparative	+	-	Total
+	a	b	a+b
-	c	d	c+d
Total	a+c	b+d	n

The positive agreement (PPA) and negative agreement (NPA) have a natural interpretation:

- Positive agreement is the proportion of comparative/reference method positive results in which the test method result is positive.

  $$\frac{a}{a+b}$$

- Negative agreement is the proportion of comparative/reference method negative results in which the test method result is negative.

  $$\frac{d}{c+d}$$

Symmetric agreement measures are not affected by interchanging the X and Y variable. These are useful in many other cases, such as comparing observers, laboratories, or other factors where neither is a natural comparator. There are various measures based on the mean of the proportions to which X agrees with Y, and Y agrees with X. The Kulczynski, Dice-Sørensen, and Ochiai are three such measures that use arithmetic, harmonic, and geometric mean of the proportions, respectively.

The harmonic mean weights the smaller proportion more heavily and produces the smallest value amongst the three measures. The geometric mean is the square root of Bangdiwala's B statistic, which is the ratio of the observed agreement to maximum possible agreement in the agreement plot. The arithmetic mean has the greatest value. In most cases, of moderate to high agreement, there is very little to choose between the measures.

The average positive agreement and average negative agreement using the Dice-Sørensen measure are:

- Average positive agreement is the number of positive matches as a proportion of the average of the number of positive results by X, Y.

  $$\frac{a}{\big((a+b)+(a+c)\big)/2}$$

- Average negative agreement is the number of negative matches as a proportion of the average of the number of negative results by X, Y.

  $$\frac{d}{\big((d+b)+(d+c)\big)/2}$$

The overall proportion of agreement is the sum of the diagonal entries divided by the total.

Chance corrected agreement measures for binary and semi-quantitative data

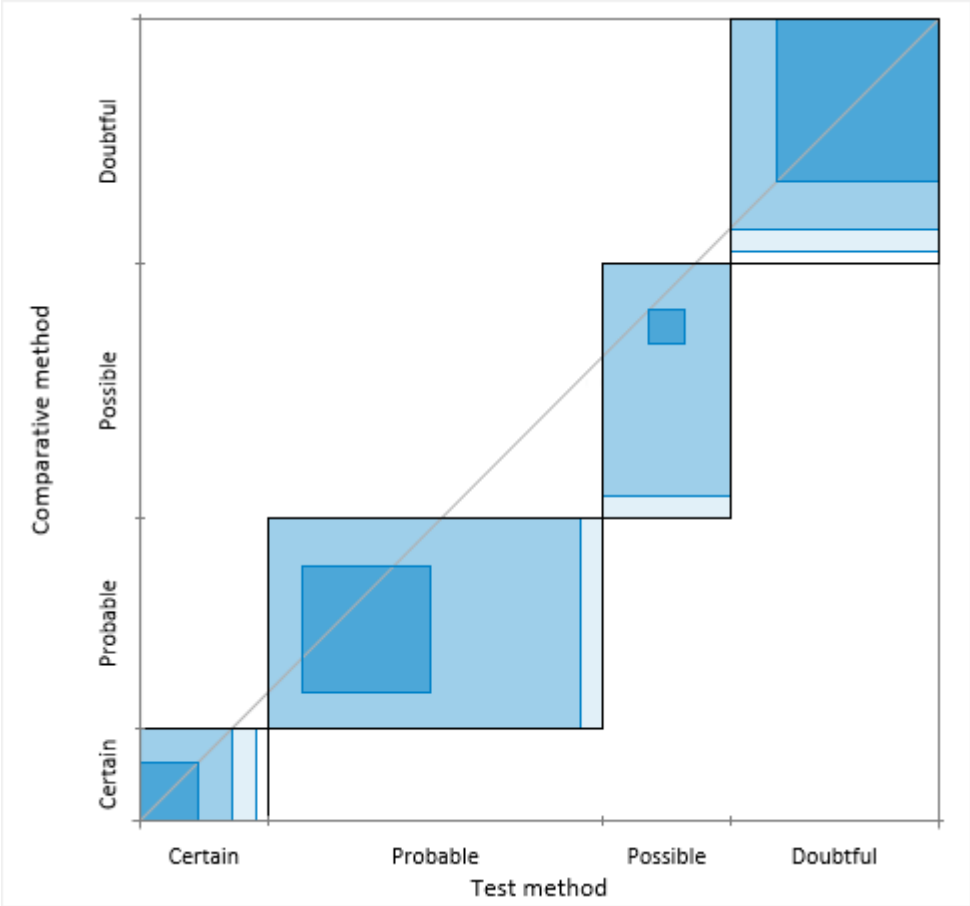
A chance corrected agreement measure takes into account the possibility of agreement occurring by chance.

The Kappa coefficient is the most popular measure for chance corrected agreement between qualitative variables. It is the overall observed agreement corrected for the possibility of agreement occurring by chance. A weighted version of the statistic is useful for ordinal variables as it weights disagreements dependent on the degree of disagreement between observers. As the Kappa coefficient is an overall summary statistic, it should be accompanied by an agreement plot which can show more insight than an overall summary statistic.

Interpretation of the Kappa coefficient is difficult. The most popular set of criteria for assessing agreement *Landis and Koch (1977)*, who characterized values < 0 as indicating no agreement and 0–0.20 as slight, 0.21–0.40 as fair, 0.41–0.60 as moderate, 0.61–0.80 as substantial, and 0.81–1 as almost perfect agreement. This set of guidelines is, however, by no means universally accepted. A substantial imbalance in the contingency table's marginal totals, either horizontally or vertically, results in a lower Kappa coefficient. And, it will be higher if the imbalance in the corresponding marginal totals is asymmetrical rather than symmetrical, or imperfectly rather than perfectly symmetrical.

Agreement plot

An agreement plot shows the agreement between two binary or semi-quantitative methods.



Bangdiwala (2013) devised the agreement plot as a complement to the kappa or B-statistics. It is invaluable for assessing agreement as it gives a visual impression that no summary statistic can convey.

The agreement plot is a visual representation of a k by k square contingency table. Each black rectangle represents the marginal totals of the rows and columns. Shaded boxes represent the agreement based on the diagonal cell frequencies; they are positioned inside the rectangles

using the sum of the off-diagonal cell frequencies from the same row and column. The partial agreement in the off-diagonal cells can be represented similarly with decreased shading based on the distance from the diagonal. The visualization is affected by the order of the categories, and so the plot is only useful for ordinal or binary data. The plot can have the origin at the bottom-left corner, or at the top-left where it more clearly mimics the contingency table.

Perfect agreement is represented by rectangles that are all perfect squares, with corners on the diagonal identity line, and with shaded boxes equal to the rectangle. Lesser agreement is represented by the area of shaded boxes compared to the area of rectangles. The path of the rectangles, how they deviate from the 45-degree identity line, represents bias in the marginal totals.

Estimating agreement between two binary or semi-quantitative methods

Measure the agreement between a test method and a reference or comparative method when both variables are binary or semi-quantitative.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Method Comparison** (or **Agreement** depending on the edition), and then click **Binary (PPA/NPA)** or **Semi-Quantitative**.
   The analysis task pane opens.
3. In the **X** drop-down list, select the comparative or reference measurement procedure, or the first observer/factor variable.
4. In the **Y** drop-down list, select the test measurement procedure, or the second observer/factor variable.
5. If the data are in frequency form, in the **Frequency** drop-down list, select the frequency count variable.
6. In the **Measure** drop-down list, select the measure of agreement:

Reference	Compute the proportion of the Y (Test method) that agree with the X (Reference / Comparative method). Useful to make comparisons against a reference method or the current laboratory method. Note: This method is asymmetric and the agreement is different depending on the assignment of X and Y.
Average	Compute the Dice-Sorenson measure of average agreement. Useful when comparing say 2 observers or laboratories where neither is a natural comparator. Note: This method is symmetric and doesn't depend on assignment of X and Y.

7. On the **Analyse-it** ribbon tab, in the **Method Comparison** group, click **Estimate** and then select an overall agreement estimator.
8. Click **Calculate**.

Study design

Method comparison study requirements and dataset layout.

Requirements

- 2 quantitative variables.
- A recommended minimum of at least 40 cases.
- Each measurement in singlicate or replicate.

Paired dataset layout

Use a column for each method (Method 1, Method 2); each row has the measurement by each method for an item (Subject).

Subject (optional)	Method 1	Method 2
1	120	121
2	113	118
3	167	150
4	185	181
5	122	122
…	…	…

Paired dataset layout with replicates

Use multiple columns for the replicates of each method (Method 1, Method 2); each row has the replicate measurements by each method for an item (Subject).

Subject (optional)	Method 1		Method 2	
1	120	122	121	120
2	113	110	118	119
3	167	170	150	155
4	185	188	181	180
5	122	123	122	130
…	…	…	…	…

Paired dataset layout with matched replicates

Use a column for each method (Method 1, Method 2); each row has the measurements by each methods for an item (Subject), with multiple rows for replicates of each item.

Subject	Method 1	Method 2
1	120	121
1	122	120
2	113	118
2	110	119
3	167	150
3	170	155
4	185	181
4	188	180
5	122	122
5	123	130
…	…	…

Matched dataset layout

Use a column for the measurement variable (Measured value), a column for the method variable (Method), and a column for the item variable (Subject); each row has the measurement for a method for an item (Subject), with multiple rows for replicates of each item.

Subject	Method	Measured value
1	X	120
1	X	122
1	Y	121
1	Y	120
2	X	113
2	X	110
2	Y	118
2	Y	119
3	X	167
3	X	170
3	Y	150
3	Y	155
4	X	185
4	X	188
4	Y	181
4	Y	180
5	X	122
5	X	123
5	Y	122
5	Y	130
…	…	…

Study design for qualitative methods

Method comparison study requirements and dataset layout for qualitative methods.

Requirements

- 2 qualitative (binary / semi-quantitative) variables.

Dataset layout for a two qualitative tests

Use a column for the reference/comparative method, and a column for the test method; each row has the values of the variables for a case (Subject).

Subject (optional)	Reference/ Comparative method	Test method
1	+	+
2	-	-
3	+	+
4	-	-
5	-	-
6	-	-
7	+	+
8	+	-
9	-	-
10	-	-
…	…	…

Frequency form dataset layout for two qualitative tests

Use a column for the reference/comparative method, and a column for the test method and a column for the number of cases (Frequency); each row has the values of the variables and the frequency count.

Reference / Comparative method	Test method	Count
+	+	32
+	-	1
-	+	3
-	-	38

Chapter 19

Measurement systems analysis (MSA)

Measurement systems analysis is the study of a measurement system to ensure it can meet requirements.

Precision

Precision is the closeness of agreement between measured quantity values obtained by replicate measurements on the same or similar objects under specified conditions.

Precision is not a quantity and therefore it is not expressed numerically. Rather, it is expressed by measures such as the variance, standard deviation, or coefficient of variation under the specified conditions of measurement

Conditions of measurement

Many different factors may contribute to the variability between replicate measurements, including the operator; the equipment used; the calibration of the equipment; the environment; the time elapsed between measurements.

Two conditions of precision, termed repeatability and reproducibility conditions are useful for describing the variability of a measurement procedure. Other intermediate conditions between these two extreme conditions of precision are also conceivable and useful, such as conditions within a single laboratory.

Reproducibility

Reproducibility is a measure of precision under a defined set of conditions: different locations, operators, measuring systems, and replicate measurements on the same or similar objects.

Intermediate precision

Intermediate precision (also called within-laboratory or within-device) is a measure of precision under a defined set of conditions: same measurement procedure, same measuring system, same location, and replicate measurements on the same or similar objects over an extended period of time. It may include changes to other conditions such as new calibrations, operators, or reagent lots.

Repeatability

Repeatability (also called within-run precision) is a measure of precision under a defined set of conditions: same measurement procedure, same operators, same measuring system, same operating conditions and same location, and replicate measurements on the same or similar objects over a short period of time.

Variance components

Variance components are estimates of a part of the total variability accounted for by a specified source of variability.

Random factors are factors where a number of levels are randomly sampled from the population, and the intention is to make inferences about the population. For example, a study might examine the precision of a measurement procedure in different laboratories. In this case, there are 2 variance components: variation within an individual laboratory and the variation among all laboratories. When performing the study it is impractical to study all laboratories, so instead a random sample of laboratories are used. More complex studies may examine the precision within a single run, within a single laboratory, and across laboratories.

Most precision studies use a nested (or hierarchical) model where each level of a nested factor is unique amongst each level of the outer factor. The basis for estimating the variance components is the nested analysis of variance (ANOVA). Estimates of the variance components are extracted from the ANOVA by equating the mean squares to the expected mean squares. If the variance is negative, usually due to a small sample size, it is set to zero. Variance components are combined by summing them to estimate the precision under different conditions of measurements.

The variance components can be expressed as a variance, standard deviation (SD), or coefficient of variation (CV). A point estimate is a single value that is the best estimate of the true unknown parameter; a confidence interval is a range of values and indicates the uncertainty of the estimate. A larger estimate reflects less precision.

There are numerous methods for constructing confidence intervals for variance components:

Estimator	Description
Exact	Used to form intervals on the inner-most nested variance components. Based on the F-distribution.
Satterthwaite	Used to form intervals on the sum of the variances. Based on the F-distribution, as above, but uses modified degrees of freedom. Works well when the factors have equal or a large number of levels, though when the differences between them are large it can produce unacceptably liberal confidence intervals.
Modified Large Sample (MLS)	Used to form intervals on the sum of variances or individual components. A modification of the large sample normal theory approach to constructing confidence intervals. Provides good coverage close to the nominal level in a wide range of cases.

Estimating the precision of a measurement system

Estimate the precision of a measurement system or procedure.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Precision**, and then click:

Simple	Measurements are made under the same conditions.
1 Factor	Measurements vary by a single random factor. Use for performing CLSI EP5 or EP15 in a single laboratory with Run as a factor.
2 Factor Nested	Measurements vary by 2 random factors. Use for performing CLSI EP5 in a single laboratory with Day and Run as nested factors.
3 Factor Nested	Measurements vary by 3 random factors.

Use for performing CLSI EP5 in multiple laboratories with Laboratory, Day, Run as nested factors.

The analysis task pane opens.

3. In the **Y** drop-down list, select the measured quantity variable.
4. If required: In the **Factor A** drop-down list, select the first random factor.
5. If required: In the **Factor B nested within Factor A** drop-down list, select the second random factor that is nested within the first factor.
6. If required: In the **Factor C nested within Factor B** drop-down list, select the third random factor that is nested within the second factor.
7. Optional: If there are multiple levels of measurement, in the **By** drop-down list, select the level variable.
8. Optional: To label the conditions of measurement for the innermost and outermost nested factors, in the **Conditions** drop-down list, select the labels.
9. Optional: To customize the detail of the output, select or clear the **Abbreviated** and **Detail** check boxes as appropriate.
10. Optional: To compare the precision against performance requirements:
    - If the allowable imprecision is a constant or proportional value across the measuring interval, under the **Allowable imprecision** group, select **Across measuring interval**, and then in the **SD / CV** edit box, type the SD in measurement units, or the CV as a percentage (suffix with % symbol).
    - If the allowable imprecision varies for each level, under the **Allowable imprecision** group, select **Each level** and then in the **Allowable imprecision** grid, under the **SD / CV** column, alongside each level, type the SD in measurement units, or the CV as a percentage (suffix with % symbol).

    Note: To compare precision against a performance claim, see *Testing precision against a performance claim* on page 177.
11. Click **Calculate**.

Testing precision against a performance claim

Test if the precision matches a performance claim.

You must have already completed the task:

- *Estimating the precision of a measurement system* on page 176

You should use this procedure when you already have a performance claim from a manufacturer's package insert, and you want to test whether your precision is significantly greater than the claim. It is possible for the precision from a study to be greater than the claim, but for the difference to be due to the random error in your study.

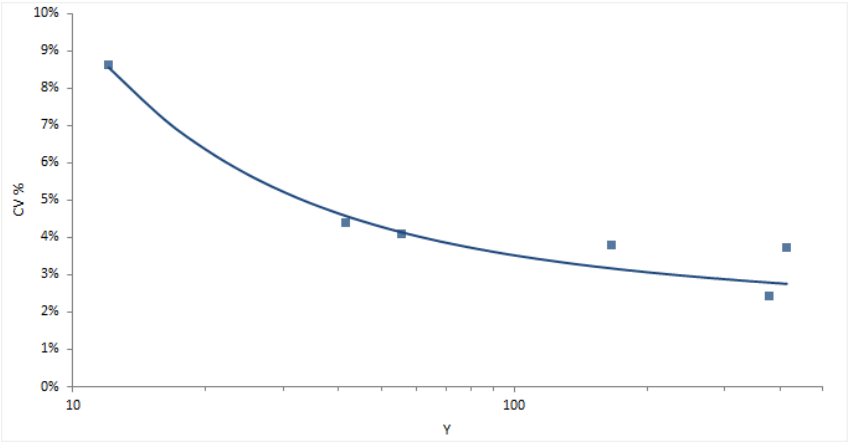
1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **MSA** group, click **Test**.
   The analysis task pane **Precision** panel opens.
3. In the **Hypothesized imprecision** grid, alongside each level, enter the performance claim as the SD in measurement units or the CV as a percentage (suffix with % symbol).
4. Optional: In the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%) and if required select the **Familywise error rate** check box.

Note: CLSI EP15 uses a 5% familywise significance level so the overall significance level is a maximum of 5% regardless of the number of comparisons (for example, for 1 level the significance level is 5%, for 2 levels the significance level is 5%/2 = 2.5% for each level, for 3 levels the significance level is 5%/3 = 1.6% for each level).

5. Click **Recalculate**.

Precision profile plot

A precision profile plot shows precision against the measured quantity value over the measuring interval.



Plotting a precision profile

Plot a precision profile to observe the performance over the entire measuring interval.

You must have already completed the task:

- *Estimating the precision of a measurement system* on page 176

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **MSA** group, click **Profile**.
   The analysis task pane **Precision Profile** panel opens.
3. In the **Precision profile** drop-down list, select the measure of precision to plot.
4. Optional: Select the required variance component check boxes.
5. Optional: If the measurements spread over a wide interval with many concentrated near zero, select the **Logarithmic X-axis scale** check box, so they are more clearly visible.
6. Click **Recalculate**.

Variance function

A variance function describes the relationship between the variance and the measured quantity value.

The are numerous models that describe the relationship between the variance and measured quantity value across the measuring interval:

Fit	Description
Constant variance	Fit constant variance across the measuring interval.
Constant CV	Fit constant coefficient of variation across the measuring interval.

Fit	Description
Mixed constant / proportional variance	Fit constant variance at low levels with constant coefficient of variation at high levels.
2-parameter	Fir a 2-parameter linear variance function.
3-parameter	Fit Sadler 3-parameter variance function – a monotone relationship (either increasing or decreasing) between the variance and level of measurement.
3-parameter alternative	Fit Sadler alternate 3-parameter variance function – gives more flexibility than the Sadler standard 3-parameter variance function especially near zero.
4-parameter	Fit a Sadler 4-parameter variance function - allows for a turning point near the detection limit.

A variance function can be useful to estimate the limit of detection or limit of quantitation.

Fitting a variance function

Plot a precision profile to observe the performance over the entire measuring interval.

You must have already completed the task:

- *Estimating the precision of a measurement system* on page 176

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **MSA** group, click **Profile**.
   The analysis task pane **Precision Profile** panel opens.
3. Select the **Fit variance function** check box.
4. In the **Model** drop-down list, select the variance function.
5. Optional: To estimate the limit of quantitation using functional sensitivity, select the **Predict X given CV** check box, and then in the edit box, type the CV as a percentage (for example 20%).
6. Click **Recalculate**.

Trueness

Trueness is the closeness of agreement between the average of an infinite number of replicate measured quantity values and a reference quantity value.

Trueness is not a quantity and therefore cannot be expressed numerically. Rather it is expressed as bias.

Bias

Bias is a measure of a systematic measurement error, the component of measurement error that remains constant in replicate measurements.

The bias can be expressed in absolute measurement units or as a percentage relative to the known value. A point estimate is a single value that is the best estimate of the true unknown parameter; a confidence interval is a range of values and indicates the uncertainty of the estimate.

The bias is an estimate of the true unknown bias in a single study. If the study were repeated, the estimate would be expected to vary from study to study. Therefore, if a single estimate is compared directly to 0 or compared to the allowable bias the statement is only applicable to the single study. To make inferences about the true unknown bias you must perform a hypothesis test:

There are two common hypotheses of interest that can be tested:

- Equality test

  The null hypothesis states that the bias is equal to 0, against the alternative hypothesis that it is not equal zero. When the test p-value is small, you can reject the null hypothesis and conclude that the bias is different to zero.

  It is important to remember that a statistically significant p-value tells you nothing about the practical importance of what was observed. For a large sample, the bias for a statistically significant hypothesis test may be so small as to be practically useless. Conversely, although there may some evidence of bias, the sample size may be too small for the test to reach statistical significance, and you may miss an opportunity to discover a true meaningful bias. Lack of evidence against the null hypothesis does not mean it has been proven to be true, the belief before you perform the study is that the null hypothesis is true and the purpose is to look for evidence against it. An equality test at the 5% significance level is equivalent to comparing a 95% confidence interval to see if it includes zero.
- Equivalence test

  The null hypothesis states that the bias is outside an interval of practical equivalence, against the alternative hypothesis that the bias is within the interval considered practically equivalent. When the test p-value is small, you can reject the null hypothesis and conclude that the bias is practically equivalent, and within the specified interval.

  An equivalence test is used to prove a bias requirement can be met. The null hypothesis states the methods are not equivalent and looks for evidence that they are in fact equivalent. An equivalence hypothesis test at the 5% significance level is the same as comparing the 90% confidence interval to the allowable bias interval.

Estimating the bias of a measurement system

Estimate the trueness of a measurement system or measurement procedure.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Trueness**.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the measured quantity variable.

4. If there are multiple levels of measurement, in the **By** drop-down list, select the level variable, and then:
   - If the values are identifiers, select the **Identifier**, and then in the **Assigned values** grid, under the **Value** column alongside each level, type the value.
   - If the values are dilutions made by diluting a high pool or mixing high and low pools, select the **Relationship**, and then select **Mixture**, **Dilution**, or **Addition** based on how the levels were prepared. In the **Assigned values** grid, under the **Value** column for the first and/ or last level, type the value (intermediate values are automatically calculated using relative values).
   - If the values are known/expected/assigned values, select the **Known values**.
5. If the reference materials package insert includes an estimate of the uncertainty of the assigned values, select the **Uncertainty in assigned values** check box, and then in the **Assigned values** grid, under the **SE** column for each level, type the value:
   - If the estimate is expressed as standard uncertainty or combined standard uncertainty, use the value.
   - If the estimate is expressed as a 95% confidence interval (CI) use the value CI/2, or if the confidence interval is expressed as upper and lower limits use the value (CI upper limit- CI lower limit)/2.
   - If the estimate is expressed as expanded uncertainty (U) with a coverage factor (% or k) use the value U/k. If coverage is 95%, use the value U/2; if coverage is 99%, use the value U/3.
6. Optional: To compare the bias against performance requirements:
   - If the allowable bias is a constant or proportional value across the measuring interval, select **Across measuring interval**, and then in the **Absolute** edit box, type the bias in measurement units, and/or in the **Relative** edit box, type the bias as a percentage (suffix with % symbol).

     Note: The allowable bias is the greater of the absolute bias and the relative bias for each level. Therefore, with a absolute bias of 5mg/dL and a relative bias of 10%, the allowable bias will be set at 5mg/dL for all values 0 mg/dL up to 50mg/dL and then at 10% of assigned value for values above 50mg/dL.
   - If the allowable bias varies for each level, select **Each level** and then in the **Allowable bias** grid, under the **Absolute / Relative** column alongside each level, type the bias in measurement units, or the bias as a percentage (suffix with % symbol).

   Note: To compare if the value of the reference material is equal to the assigned value, see *Testing against an assigned value* on page 181.
7. Click **Calculate**.

Testing against an assigned value

Test if the value of the reference material is equal to the assigned value; that is the bias is zero.

You must have already completed the task:

- *Estimating the bias of a measurement system* on page 180

It is possible for the bias from a study to be different to 0, but for the difference to be due to the random error in your study. The smaller your study, the larger the bias has to be to be declared statistically significant. Also, just because bias is statistically significant does not mean that it is of concern, it may still meet your allowable bias performance requirements.

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Bias** group, click **Test Equality**.
   The analysis task pane **Trueness** panel opens.

3. Optional: In the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%) and if required select the **Familywise error rate** check box.

   Note: CLSI EP15 uses a 5% familywise significance level so the overall significance level is a maximum of 5% regardless of the number of comparisons (for example, for 1 level the significance level is 5%, for 2 levels the significance level is 5%/2 = 2.5% for each level, for 3 levels the significance level is 5%/3 = 1.6% for each level).
4. Click **Recalculate**.

Testing bias against allowable bias

Test if the bias meets performance requirements; that is the bias is less than allowable bias.

You must have already completed the task:

- *Estimating the bias of a measurement system* on page 180

If the bias from a study is less than the allowable bias, it meets the performance requirements, but you cannot make any statements relating to probability. An equivalence test allows you to make statements with a given level of confidence about what you observed.

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Bias** group, click **Test Equivalence**.
   The analysis task pane **Trueness** panel opens.
3. Optional: In the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%) and if required select the **Familywise error rate** check box.

   Note: CLSI EP15 uses a 5% familywise significance level so the overall significance level is a maximum of 5% regardless of the number of comparisons (for example, for 1 level the significance level is 5%, for 2 levels the significance level is 5%/2 = 2.5% for each level, for 3 levels the significance level is 5%/3 = 1.6% for each level).
4. Enter the performance requirements:
   - If the allowable bias is a constant or proportional value across the measuring interval, select **Across measuring interval**, and then in the **Absolute** edit box, type the bias in measurement units, and/or in the **Relative** edit box, type the bias as a percentage (suffix with % symbol).

     Note: The allowable bias is the greater of the absolute bias and the relative bias for each level. Therefore, with a absolute bias of 5mg/dL and a relative bias of 10%, the allowable bias will be set at 5mg/dL for all values 0 mg/dL up to 50mg/dL and then at 10% of assigned value for values above 50mg/dL.
   - If the allowable bias varies for each level, select **Each level** and then in the **Allowable bias** grid, under the **Absolute / Relative** column, alongside each level, type the absolute bias in measurement units, or the relative bias as a percentage (suffix with % symbol).
5. Click **Recalculate**.

Linearity

Nonlinear bias is a component of bias that cannot be represented by a linear relationship between the measured and true values.

A measurement procedure is linear when there is a mathematically verified straight-line relationship between the measured and true values. It is an important parameter as it allows linear interpolation of results between points.

A linear (straight-line) fit describes a relationship where the measuring system is linear. A polynomial fit describes a relationship where the measuring system is nonlinear. In evaluating linearity, a nonlinear polynomial fit is compared against a linear fit. Typically the 2nd and 3rd order polynomial fits are of interest:

- 2nd-order polynomial fit (curved line), where there is nonlinearity due to increasing or decreasing recovery at higher levels.
- 3rd-order polynomial fit (slanted S-shaped curve), where there is nonlinearity at the ends of the measuring interval due to changing recovery across levels.

Bias due to nonlinearity is measured as the difference between the linear fit and the best fitting polynomial fit.

Estimating the linearity of a measurement procedure (EP06-A)

Determine whether a measurement system or procedure provides measured quantity values (within a measuring interval) that are directly proportional to the true value.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Linearity**.
3. In the **Y** drop-down list, select the measured variable.
4. In the **By** drop-down list, select the level variable, and then:
   - If the values are identifiers, select the **Identifier**, and then in the **Assigned values** grid, under the **Value** column alongside each level, type the value.
   - If the values are dilutions made by diluting a high pool or mixing high and low pools, select the **Relationship**, and then select **Mixture**, **Dilution**, or **Addition** based on how the levels were prepared. In the **Assigned values** grid, under the **Value** column for the first and/or last level, type the value (intermediate values are automatically calculated using relative values).
   - If the values are known/expected/assigned values, select the **Known values**.

   Note: Computation of linearity only requires the relationship between the levels, you do not need to enter the assigned values if they are unknown.
5. Optional: To compare the nonlinearity bias against performance requirements:
   - If the allowable nonlinearity bias is a constant or proportional value across the measuring interval, select **Across measuring interval**, and then in the **Absolute** edit box, type the bias in measurement units, and/or in the **Relative** edit box, type the bias as a percentage (suffix with % symbol).

     Note: The allowable bias is the greater of the absolute bias and the relative bias for each level. Therefore, with a absolute bias of 5mg/dL and a relative bias of 10%, the allowable bias will be set at 5mg/dL for all values 0 mg/dL up to 50mg/dL and then at 10% of assigned value for values above 50mg/dL.
   - If the allowable nonlinearity bias varies for each level, select **Each level** and then in the **Allowable nonlinearity** grid, under the **Absolute / Relative** column alongside each level, type the bias in measurement units, or the bias as a percentage (suffix with % symbol).
6. Optional: On the **Fit Model** panel, in the **Fit** drop-down list, select:

Best significant term polynomial (CLSI EP6-A)	Fits a 2nd- and 3rd-order polynomial fit, then determines if either are better than the linear fit by testing whether the nonlinear terms (polynomial fit coefficients) are statistically significant. The 3rd-order polynomial is used if it has significant nonlinear terms, the 2nd-order polynomial is used if it has significant nonlinear terms, otherwise the measurement procedure is assumed to be linear. Recommended by CLSI EP6.

Forward stepwise polynomial LoF	Fits a linear fit and performs a lack-of-fit (LoF) significance test to determine how well the model fits the data. If the fit is significantly worse than expected, a 2nd-order polynomial is fit and the LoF test is repeated. If the 2nd-order polynomial fit is not significantly worse, it is used, otherwise a 3rd-order polynomial is fit. Recommended by Liu and others.
Polynomial	Fits a polynomial model of given order.

Note: If the precision of the measurement procedure is poor nonlinearity can be difficult to detect as neither of the polynomial fits will be significantly better than the linear fit, due to the amount of random error in the measurements.

7. Click **Calculate**.

Estimating the linearity of a measurement procedure (EP06-Ed2)

Determine whether a measurement system or procedure provides measured quantity values (within a measuring interval) that are directly proportional.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Linearity**.
3. In the **Y** drop-down list, select the measured variable.
4. In the **By** drop-down list, select the level variable, and then:
   - If the values are identifiers, select the **Identifier**, and then in the **Assigned values** grid, under the **Value** column alongside each level, type the value.
   - If the values are dilutions made by diluting a high pool or mixing high and low pools, select the **Relationship**, and then select **Mixture**, **Dilution**, or **Addition** based on how the levels were prepared. In the **Assigned values** grid, under the **Value** column for the first and/or last level, type the value (intermediate values are automatically calculated using relative values).
   - If the values are known/expected/assigned values, select the **Known values**.

   Note: Computation of linearity only requires the relationship between the levels, you do not need to enter the assigned values if they are unknown.
5. Optional: To compare the nonlinearity bias against performance requirements:
   - If the allowable nonlinearity bias is a constant or proportional value across the measuring interval, select **Across measuring interval**, and then in the **Absolute** edit box, type the bias in measurement units, and/or in the **Relative** edit box, type the bias as a percentage (suffix with % symbol).

     Note: The allowable bias is the greater of the absolute bias and the relative bias for each level. Therefore, with a absolute bias of 5mg/dL and a relative bias of 10%, the allowable bias will be set at 5mg/dL for all values 0 mg/dL up to 50mg/dL and then at 10% of assigned value for values above 50mg/dL.
   - If the allowable nonlinearity bias varies for each level, select **Each level** and then in the **Allowable nonlinearity** grid, under the **Absolute / Relative** column alongside each level, type the bias in measurement units, or the bias as a percentage (suffix with % symbol).
6. On the **Fit Model** panel, in the **Fit** drop-down list, select **Linear**.
7. In the **X** drop-down list, select **Expected values** and in the **Y** down list, select **Mean**.
8. In the **Weights** drop down list, select

None	Fit an ordinary regression. Use when measurement procedure exhibits constant SD over the measuring interval.

Var(Y level)	Fit a weighted regression with weight based on the variance at each level. Recommended when number of replicates per level is 4 or more.
Var(Y level pooled)	Fit a weighted regression with weight based on the pooled variance over a subinterval of levels. Recommended when number of replicates per level is 2 or 3.
VarFn(Y)	Fit a weighted regression with weight based on the variance function at the mean of each level. Recommended when the precision can be modelled by a variance function.

9. If the levels are made by a dilution of a high level, select **Force through zero** check box. Otherwise, if the levels are produced by mixing a high and low level, clear the **Force though origin** check box.
10. Click **Calculate**.

Interferences

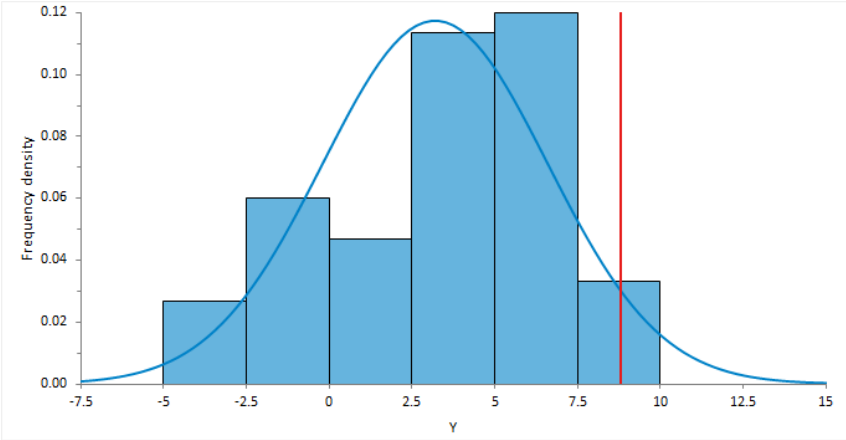
Interference bias is a component of bias caused by nonspecificity attributable to the presence of a specific interfering substance.

Detection capability

Detection capability describes the performance of a measurement system at small measured quantity values.

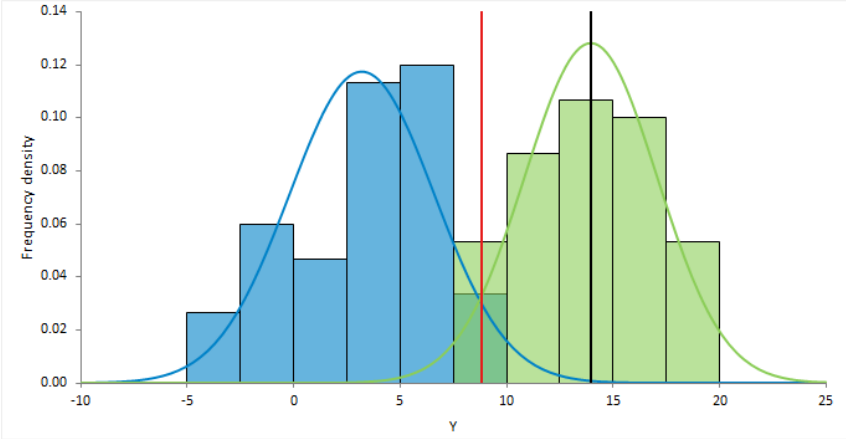
Limit of blank (LoB)

Limit of blank (also known as the critical value) is the highest quantity value that is likely to be observed, with a stated probability, for a blank material.



Limit of detection (LoD)

Limit of detection (also known as the minimum detectable value) is the quantity value, for which the probability of falsely claiming the absence of a measurand in a material is beta, given probability (alpha) of falsely claiming its presence.



Estimating the detection limit of a measurement system

Estimate the limit of detection of a measurement system or procedure.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Precision** and then click **Detection Capability**.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the measured variable.

4. In the **By** drop-down list, select the level variable, and then:
   - If the values are identifiers, select the **Identifier**, and then in the **Assigned values** grid, under the **Value** column alongside each level, type the value.
   - If the values are dilutions made by diluting a high pool or mixing high and low pools, select the **Relationship**, and then select **Mixture**, **Dilution**, or **Addition** based on how the levels were prepared. In the **Assigned values** grid, under the **Value** column for the first and/or last level, type the value (intermediate values are automatically calculated using relative values).
   - If the values are known/expected/assigned values, select the **Known values**.

   Note: Computation of detection limits only requires the assigned value of 0 for blank materials. You do not need to assign values to other materials.
5. In the **LoB Estimator** drop-down list, select:

Normal quantile	Estimate the LoB using the 5% upper of the distribution of blank values (levels with assigned value 0). Use when values are normally distributed and not truncated at 0.
Quantile	Estimate the LoB using the 95th percentile of the distribution of blank values. Use when values are truncated at zero.
Known value	Use specific value as the LoB.

6. In the **Alpha** edit box, type the probability that a blank material gives a result greater than the critical value (LoB), when in fact the substance is not present (typically 5%).
7. In the **Beta** edit box, type the probability that a non-blank material gives a result less than the critical value (LoB) when in fact the substance is present (typically 5%).
8. Click **Calculate**.

Estimating the detection limit using a precision profile

Estimate the limit of detection of a measurement system or procedure using a precision profile variance function.

You must have already completed the task:

- *Fitting a variance function* on page 179

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **MSA** group, click **Detection Capability**.
   The analysis task pane **Detection Capability** panel opens.
3. In the **LoB Estimator** drop-down list, select:

Normal quantile	Estimate the LoB using the 5% upper of the distribution of blank values (levels with assigned value 0). Use when values are normally distributed and not truncated at 0.
Quantile	Estimate the LoB using the 95th percentile of the distribution of blank values. Use when values are truncated at zero.
Known value	Use specific value as the LoB.

4. If you selected Normal quantile, in the **SD** drop-down list, select:

Blank material	Estimate based on material assigned a value of 0.
Precision profile at blank material value	Estimate based on the precision profile variance function at the mean value of the material assigned a value of 0.
Precision profile at zero	Estimate based on the precision profile variance function at the value 0. Use when it is not possible to obtain blank material, but the profile function is representative of values near zero.

5. In the **Alpha** edit box, type the probability that a blank material gives a result greater than the critical value (LoB), when in fact the substance is not present (typically 5%).
6. In the **SD** drop-down list, select **Precision profile function**.
7. In the **Beta** edit box, type the probability that a non-blank material gives a result less than the critical value (LoB) when in fact the substance is present (typically 5%).
8. Click **Calculate**.

Estimating the detection limit using a probit fit

Fit a probit regression model to estimate the limit of detection.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Fit Model**, and then click **Probit**.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the binary response variable.
4. In the **Event** drop-down list, select the positive outcome.
5. In the **X** drop-down list, select the level variable.
6. Next to the **X** drop-down list, click the drop-down menu, and then in the **Transform** drop-down list, select **Log 10**
7. In the **Plot** drop-down list, select **Scaled - X against Y**.
8. Clear the **Individual** check box, and select the **Aggregate** check box.
9. On the **Analyse-it** ribbon tab, in the **Fit Model** group, click **Predict**, and then click **X given Probability**.
10. On the analysis task pane, in the **Fit** panel, in the **Predict X given Probability** edit box, type `0.95`
11. Click **Calculate**.

On the analysis report, in the **Predict X for Given Probability** section, the predicted level at 0.95 probability is the limit of detection.

Limit of quantitation (LoQ)

Limit of quantitation is the smallest quantity value that meets the requirements for intended use.

There is no set procedure to determine the limit of quantitation. Some authors suggest setting separate goals for bias and imprecision. Others such as Westgard use a model to combine the bias and imprecision into "total error" and compare it to a total allowable error goal. Others prefer to avoid the use of models and estimate total error directly using difference between the results of a method and a reference method *(Krouwer, 2002)*. In other areas such as immunoassays the limit of quantitation is often defined as functional sensitivity where the precision profile function CV is equal 20%.

Study design

Measurement systems analysis study requirements and dataset layout.

Requirements

- A quantitative variable.
- 1 or more optional factor variables indicating the random effects of interest.
- At least 2 replicates at each level.

Dataset layout

Use a column for the measured variable (Measured value) and optionally a by variable (Level); each row is a separate measurement.

Level (optional)	Measured value
120	121
120	118
120	124
120	120
120	116
…	…
240	240
240	246
240	232
240	241
240	240
…	…

Dataset layout for 1 random factor

Use a column for the measured variable (Measured value), and 1 columns for the random factor (Run), and optionally a by variable (Level); each row is a separate measurement.

Level (optional)	Run	Measured value
120	1	121
120	1	118
120	1	…
120	2	120
120	2	116
120	2	…
120	3	…
120	…	…
…	…	…
240	1	240
240	1	242
240	1	…
240	2	260
240	2	238
240	2	…
240	3	…
240	…	…
…	…	…

Dataset layout for 2 random factors

Use a column for the measured variable (Measured value), and 2 columns for the random factors (Day, Run), and optionally a by variable (Level); each row is a separate measurement.

Level (optional)	Day	Run	Measured value
120	1	1	121
120	1	1	118
120	1	1	…
120	1	2	120
120	1	2	116
120	1	2	…
120	1	3	…
120	1	…	…
120	2	1	124
120	2	1	119
120	2	1	…
120	2	2	118
120	2	2	121
120	2	2	…
120	2	3	…
120	2	…	…
…		…	…

Alternate dataset layout

Use multiple columns for the replicates of the measured variable (Measured value) and a by variable (Level), each row is a separate combination of level and factors.

Level	Measured value		
1	0.7	0.9	…
2	4.6	4.1	…
3	6.5	6.9	…
4	11	12.2	…

Note: All the above dataset layouts can arrange replicate measurements in a single row rather than in multiple rows.

Chapter

Reference interval

20

A reference interval (sometimes called a reference range or normal range) describes the range of values of a measured quantity in healthy individuals.

Reference limits

A reference limit defines a value where a given proportion of reference values are less than or equal to. A reference interval defines the interval between a lower and upper reference limit that includes a given proportion of the reference values.

The process of defining a reference interval is that reference individuals compromise a reference population from which is selected a reference sample group on which are determined reference values on which is observed a reference distribution from which are calculated reference limits that define a reference interval.

A point estimate of a reference limit is a single value that is the best estimate of the true unknown parameter; a confidence interval is a range of values and indicates the uncertainty of the estimate. It is important to remember that the width of the confidence interval is dependent upon sample size and that large sample sizes are required to produce narrow confidence intervals, particularly for skewed distributions *(Linnet, 2000)*.

There are numerous quantile estimators that may define the reference limits. The best estimator depends heavily on the shape of the distribution.

Normal theory quantile

A quantile estimator derived from a normally distributed population with unknown mean and standard deviation.

A normal theory quantile is most powerful when a random sample is from a population with a normal distribution.

Estimator	Description
MVUE	The uniformly minimum unbiased variance quantile estimator uses the sample mean and unbiased standard deviation as the best estimate of the population parameters (Xbar ± Z(alpha) * (s / c4(n))). The factor c4(n) is applied to the sample standard deviation to account for the bias in the estimate for small sample sizes.
t-based	The t-based prediction interval quantile estimator uses the Student's t distribution for a prediction interval (Xbar ± t(alpha, n-1) * s * sqrt(1+1/n)) for a single future observation *(Horn, 2005)*. Note that this method produces a wider interval than the MVUE estimator for small samples.

In many cases, the data are skewed to the right and do not follow a normal distribution. A Box-Cox (or logarithmic) transform can correct the skewness, allowing you to use the Normal theory quantile. If not, a distribution-free estimator may be more powerful *(IFCC, 1987)*.

Quantile

A distribution-free (non-parametric) quantile estimator based on the order statistics (the sorted values in the sample).

There are several definitions for the quantile estimator useful in defining reference limits.

Formula (N is the sample size and p is the quantile)	Description
Np+1/2	Piecewise linear function where the knots are the values midway through the steps of the empirical distribution function. Recommended by *Linnet (2000)* as having the lowest root-mean-squared-error .
(N+1)p	Linear interpolation of the expectations for the order statistics of the uniform distribution on [0,1]. Recommended by *IFCC (1987)* and *CLSI (2010)*.
(N+1/3)p+1/3	Linear interpolation of the approximate medians for the order statistics. Has many excellent properties and is approximately median-unbiased regardless of the distribution.

A 95% reference interval (0.025 and 0.975 quantiles) requires a minimum sample size of 39. A 90% confidence interval for a 95% reference interval requires a minimum sample size of 119.

Bootstrap quantile

A distribution-free (non-parametric) quantile estimator that is the median of a set of quantiles calculated by re-sampling the original sample a large number of times and computing a quantile for each sample.

The bootstrap quantile *(Linnet, 2000)*as providing the lowest root-mean-squared-error (an estimate of the bias and precision in the estimate) for both normal and skewed distributions. Another advantage is that confidence intervals can be computed for smaller sample sizes, although Linnet still recommends a sample size of at least 100.

Harrell-Davis quantile

A distribution-free (non-parametric) estimator that is a weighted linear combination of order statistics. It is substantially more efficient than the traditional estimator based on one or two order statistics.

Robust bi-weight quantile

A quantile estimator that uses robust estimators of location and spread, which resist the effects of extreme observations.

The robust bi-weight estimator *(Horn & Pesce 1998)* uses robust estimators of location and spread, which resist the effects of extreme observations. It requires a symmetric distribution. If the distribution is skewed, you should apply a transform to achieve symmetry, or create a new sample by mirroring the sample around the median to form a symmetric distribution *(Horn, 1990)*, before applying the bi-weight estimator.

Estimating a reference limit or interval

Estimate the reference limit or reference interval using a quantile estimator.

1. Select a cell in the dataset.

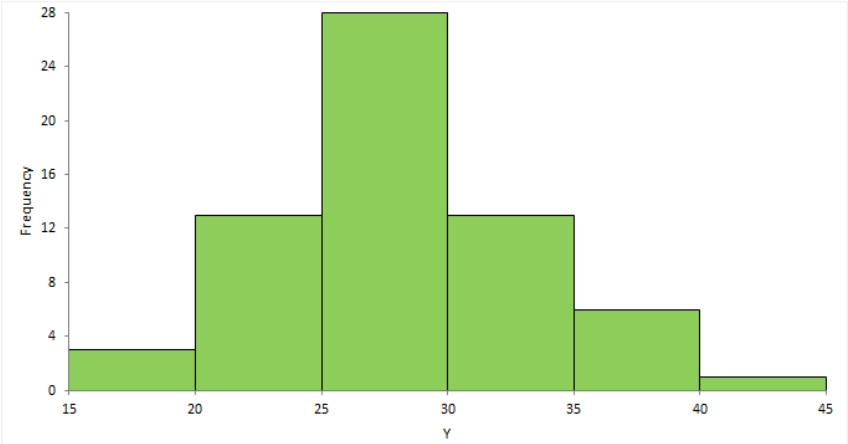
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Diagnostic**, and then under the **Reference Interval** heading, click the estimator.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the variable.
4. In the **Type** drop-down list, select:

Reference interval	Estimate lower and upper reference limits that a stated proportion of values are between.
Reference limit	Estimate a single reference limit that a stated proportion of values are less than.

5. In the **Proportion inside interval/Proportion less than** edit box, type the reference level as a percentage.
6. In the **Method** drop-down list, select the computation method.
7. In the **Confidence interval** edit box, type the confidence level as a percentage, or type - to suppress the confidence interval, and then in the drop-down list, select the confidence bounds.
8. Optional: To split the reference values by factors such as gender or age:
   a) On the **Analyse-it** ribbon tab, in the **Reference Interval** group, click **Partition**
   b) In the **Factors** list, select the factor variables.
   c) In the **Partitions** list, select a partition.
      The tasks change to the analysis options for that partition.
   d) Optional: By default, the analysis options are the same for each partition, set the specific analysis options for the partition as required.
   e) Repeat steps *8.c* through *8.d* for each partition.
9. Click **Calculate**.

Histogram

A histogram shows the distribution of the data to assess the central tendency, variability, and shape.



A histogram for a quantitative variable divides the range of the values into discrete classes, and then counts the number of observations falling into each class interval. The area of each bar in the histogram is proportional to the frequency in the class. When the class widths are equal, the height of the bar is also proportional to the frequency in the class.

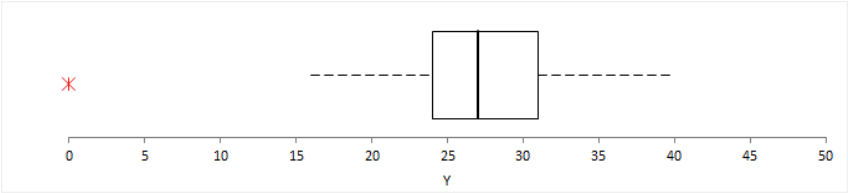
Choosing the number of classes to use can be difficult as there is no "best," and different class widths can reveal or hide features of the data. Scott's and Freedman-Diaconis' rules provide a

default starting point, though sometimes particular class intervals make sense for a particular problem.

The histogram reveals if the distribution of the data is normal, skewed (shifted to the left or right), bimodal (has more than one peak) and so on. Skewed data can sometimes be transformed to normal using a transformation. Bi-modality often indicates that there is more than one underlying population in the data. Individual bars distanced from the bulk of the data can indicate the presence of an outlier.

Box plot

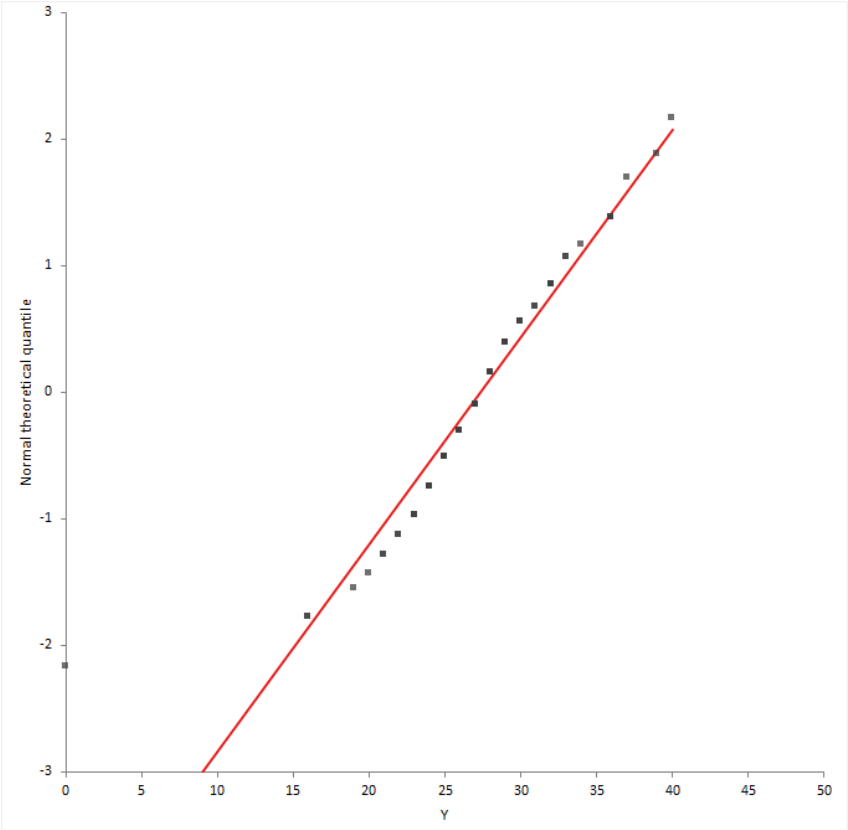
A box plot shows the five-number summary of the data – the minimum, first quartile, median, third quartile, and maximum. An outlier box plot is a variation of the skeletal box plot that also identifies possible outliers.



An outlier box plot is a variation of the skeletal box plot, but instead of extending to the minimum and maximum, the whiskers extend to the furthest observation within 1.5 x IQR from the quartiles. Possible near outliers (orange plus symbol) are identified as observations further than 1.5 x IQR from the quartiles, and possible far outliers (red asterisk symbol) as observations further than 3.0 x IQR from the quartiles. You should investigate each possible outlier before deciding whether to exclude it, as even in normally distributed data, an outlier box plot identifies approximately 0.7% of observations as possible outliers.

Normal probability (Q-Q) plot

A normal probability plot, or more specifically a quantile-quantile (Q-Q) plot, shows the distribution of the data against the expected normal distribution.



For normally distributed data, observations should lie approximately on a straight line. If the data is non-normal, the points form a curve that deviates markedly from a straight line. Possible outliers are points at the ends of the line, distanced from the bulk of the observations.

Testing the normality of a reference distribution

Test if a reference distribution is normal.

You must have already completed the task:

- *Estimating a reference limit or interval* on page 194

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Reference Interval** group, click **Normality**.
   The analysis task pane **Normality** panel opens with the **Normal plot** check box selected.
3. Optional: To perform a formal normality hypothesis test, select the **Normality test** check box.
4. Click **Recalculate**.

Transforming the reference distribution

Transform the reference values to achieve a normal distribution.

You must have already completed the task:

- *Estimating a reference limit or interval* on page 194

You should transform the data when the distribution does not match the assumptions of the quantile estimator. If you cannot transform the distribution to meet the assumptions of the estimator you should consider using another estimator.

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Reference interval** group, click **Transform**, and then click the transformation.
   The analysis task pane **Transform** panel opens.
3. Set the transform parameters:
   - If the transform is the logarithm, in the **Base** edit box, type the base of the logarithm or enter - for the natural logarithm.
   - If the transform is Box-Cox, in **Shift** and **Lambda** edit boxes, type the parameters, or to estimate them from the data, click **Estimate unknown parameter**.
   - If the transform is exponential, in the **Gamma** edit box, type parameter.
   - If the transform is 2-stage exponential/modulus, in the **Gamma** and **Kappa** edit boxes, type parameters.
4. Click **Recalculate**.

Transferring reference limits

Transfer a reference limit or reference interval from one method or laboratory to another.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Diagnostic**, and then under the **Reference Interval** heading, click **Transference**.
   The analysis task pane opens.
3. In the **Y** drop-down list, select the variable.
4. Optional: To split the reference values by factors such as gender or age:
   a) On the **Analyse-it** ribbon tab, in the **Reference Interval** group, click **Partition**
   b) In the **Factors** list, select the factor variables.
   c) In the **Partitions** list, select a partition.
      The tasks change to the analysis options for that partition.
   d) Optional: By default, the analysis options are the same for each partition, set the specific analysis options for the partition as required.
   e) Repeat steps *4.c* through *4.d* for each partition.
5. In the **Reference level** and **Reference interval** edit boxes, type the reference level as a percentage and the upper/lower reference limits.
6. Optional: If transferring from another measurement procedure, under the **Linear transformation** heading, in the **Constant** and **Proportional** edit boxes, type the constant and proportional average bias (available from the method comparison study) to translate the old reference limits to the new measurement procedure.
7. Select the **Hypothesis test** check box.
8. In the **Hypotheses** drop-down list, select the null and alternative hypotheses.

   The default null hypothesis is that the proportion of observations in the reference interval is equal to the reference level (95% default) against the alternative hypothesis that there are more or fewer observations in the interval. CLSI EP28-A3 performs a one-sided hypothesis test that proportion inside the interval is greater than or equal to 95% against the alternative that less than 95% are inside the interval.
9. In the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%).

You need at least 20 samples for a one-sided hypothesis test at the 5% significance level. More samples are required to be able to detect a significant difference for the two-sided hypothesis test, or when a smaller significance level is required.

10. If you selected partition factors, in the **Partitions** list, select another partition, and then repeat steps *5* through *9* for all partitions.
11. Click **Calculate**.

Study design

Reference interval study requirements and dataset layout.

Requirements

- A quantitative variable of the reference values.
- 1 or more factor variables.

Dataset layout

Use a column for the variable (Reference value), and optionally additional columns for each partition factor (Sex); each row has the values of the variables for a case (Subject).

Subject (optional)	Sex (optional)	Reference value
1	Male	121
2	Male	118
3	Female	124
4	Female	120
5	Male	116
6	Male	…
7	Female	100
8	Male	115
9	Female	102
10	Female	98
11	Male	118
…	…	…

Chapter

Diagnostic performance

21

Diagnostic performance evaluates the ability of a qualitative or quantitative test to discriminate between two subclasses of subjects.

Measures of diagnostic accuracy

Diagnostic accuracy measures the ability of a test to detect a condition when it is present and detect the absence of a condition when it is absent.

Comparison of the result of a diagnostic test to the true known condition of each subject classifies each outcome as:

True positive (TP)	Test result correctly identifies the presence of the condition.
False positive (FP)	Test result incorrectly identifies the presence of the condition when it was absent.
True negative (TN)	Test result correctly identifies the absence of the condition.
False negative (FN)	Test result incorrectly identifies the absence of the condition when it was present.

A perfect diagnostic test can discriminate all subjects with and without the condition and results in no false positive or false negatives. However, this is rarely achievable, as misdiagnosis of some subjects is inevitable. Measures of diagnostic accuracy quantify the discriminative ability of a test.

Sensitivity / Specificity

Sensitivity and specificity are the probability of a correct test result in subjects with and without a condition, respectively.

Sensitivity (true positive fraction, TPF) measures the ability of a test to detect the condition when it is present. It is the probability that the test result is positive when the condition is present.

Specificity (true negative fraction, TNF) measures the ability of a test to detect the absence of the condition when it is not present. It is the probability that the test result is negative when the condition is absent.

Likelihood ratios

Likelihood ratios are the ratio of the probability of a specific test result for subjects with the condition against the probability of the same test result for subjects without the condition.

A likelihood ratio of 1 indicates that the test result is equally likely in subjects with and without the condition. A ratio > 1 indicates that the test result is more likely in subjects with the condition than without the condition, and conversely, a ratio < 1 indicates that the test result is more likely in subjects without the condition. The larger the ratio, the more likely the test result is in subjects with the condition than without; likewise, the smaller the ratio, the more likely the test result is in subjects without than with the condition.

The likelihood ratio of a positive test result is the ratio of the probability of a positive test result in a subject with the condition (true positive fraction) against the probability of a positive test result in a subject without the condition (false positive fraction). The likelihood ratio of a negative test result is the ratio of the probability of a negative test result in a subject with the condition (false negative fraction) to the probability of a negative test result in a subject without the condition (true negative fraction).

Predictive values

Predictive values are the probability of correctly identifying a subject's condition given the test result.

Predictive values use Bayes' theorem along with a pre-test prior probability (such as the prevalence of the condition in the population) and the sensitivity and specificity of the test to compute the post-test probability (predictive value).

The positive predictive value is the probability that a subject has the condition given a positive test result; the negative predictive value is the probability that a subject does not have the condition given a negative test result.

Youden J

Youden's J is the likelihood of a positive test result in subjects with the condition versus those without the condition. It is also the probability of an informed decision (as opposed to a random guess).

Youden's J index combines sensitivity and specificity into a single measure (Sensitivity + Specificity - 1) and has a value between 0 and 1. In a perfect test, Youden's index equals 1. It is also equivalent to the vertical distance above the diagonal no discrimination (chance) line to the ROC curve for a single decision threshold.

Estimating sensitivity and specificity of a diagnostic test

Estimate the sensitivity/specificity and other measures of the accuracy of a binary diagnostic test (or a quantitative test at a specific decision threshold).

You can use this task with up to two binary tests or quantitative tests at a specified decision threshold. To show the sensitivity/specificity for all possible thresholds of a quantitative test you should use ROC curves.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Diagnostic**, and then click **Binary (Sensitivity / Specificity)**.
   The analysis task pane opens.
3. In the **True state** drop-down list, select the true condition variable.
4. In the **Positive event** drop-down list, select the state that indicates the presence of the condition/event of interest.
5. In the **Y** drop-down list(s), select the test variable(s).
6. If the Y variable is binary, in the **Positive event** drop-down list, select the state that indicates a positive test.
7. If the Y variable is quantitative, click the menu next to the variable and select **Cut-off**, and then enter the criteria to recode the variable into a binary variable.
8. If the data are in frequency form, in the **Frequency** drop-down list, select the frequency count variable.
9. Optional: Select the appropriate **TP/TN/FP/FN**, **Sensitivity / Specificity**, **Likelihood ratio**, **Predictive value** check boxes.

10. Click **Calculate**.

Comparing the sensitivity and specificity two diagnostic tests

Compare the sensitivity and specificity of two diagnostic tests and make inferences about the differences.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Diagnostic**, and then click **Binary (Sensitivity / Specificity)**.
   The analysis task pane opens.
3. In the **Model** drop-down menu, select the number of tests and the type of study design.

2 paired tests	Compare 2 tests where the results are on the same subjects.
2 independent tests / groups	Compare 2 tests where the results are on different subjects, or 1 test applied to 2 different groups of subjects.

4. In the **True state** drop-down list, select the true condition variable.
5. In the **Positive event** drop-down list, select the state that indicates the presence of the condition/event of interest.
6. If comparing 2 independent tests, in the **Y** drop-down list, select the diagnostic test variable, and then in the **Factor** drop-down list, select the factor variable (test or group indicator).
7. If comparing 2 paired tests, in the **Y** drop-down lists, select the diagnostic test variables.
8. If the data are in frequency form, in the **Frequency** drop-down list, select the frequency count variable.
9. On the **Analyse-it** ribbon tab, in the **Diagnostic Accuracy** group, click **Test**, and then click:

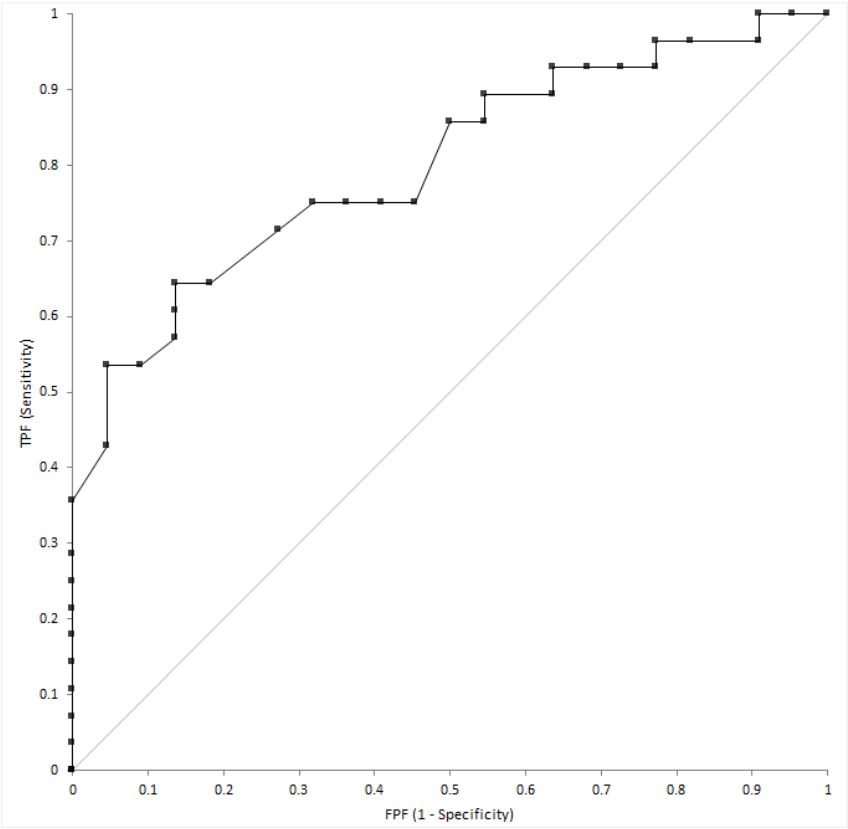
Equality	Test if the sensitivities/specificities of two tests are equal.
Equivalence	Test if the sensitivities/specificities of two tests are equivalent within a practical difference.
Non-inferiority	Test if the sensitivity/specificity of a new test is not inferior to a standard test.

   The default options use a Miettinen-Nurminen or Tango score confidence interval, and a Score Z test. The Newcombe score confidence interval is approximate not based on the same evidence function, and rejection of a hypothesis based on the confidence limits and hypothesis test may give conflicting results.
10. In the analysis task pane, under the hypotheses drop-down list:
    - If testing a hypothesis of equivalence, in the **Equivalent difference**, type the smallest practical difference that would be considered the same.
    - If testing a hypothesis of non-inferiority, in the **Standard** drop-down list, select the standard diagnostic test, and then in the **Smallest difference** edit box, type the smallest difference that would be considered inferior.
11. Optional: To compare the p-value against a predefined significance level, in the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%).
12. Click **Calculate**.

ROC plot

ROC (receiver operating characteristic) curves show the ability of a quantitative diagnostic test to classify subjects correctly as the decision threshold is varied.

The ROC plot shows sensitivity (true positive fraction) on the horizontal axis against 1-specificity (false positive fraction) on the vertical axis over all possible decision thresholds.



A diagnostic test able to perfectly identify subjects with and without the condition produces a curve that passes through the upper left corner (0, 1) of the plot. A diagnostic test with no ability to discriminate better than chance produces a diagonal line from the origin (0, 0) to the top right corner (1, 1) of the plot. Most tests lie somewhere between these extremes. If a curve lies below the diagonal line (0, 0 to 1, 1), you can invert it by swapping the decision criteria to produce a curve above the line.

An empirical ROC curve is the simplest to construct. Sensitivity and specificity use the empirical distributions for the subjects with and without the condition. This method is nonparametric because no parameters are needed to model the behavior of the curve, and it makes no assumptions about the underlying distribution of the two groups of subjects.

Plotting a single ROC curve

Plot the receiver-operator characteristic (ROC) curve to visualize the accuracy of a diagnostic test.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Diagnostic**, and then under the **Accuracy** heading, click **ROC Curve**.
   The analysis task pane opens.
3. In the **True state** drop-down list, select the true condition variable.
4. In the **Positive event** drop-down list, select the state that indicates the presence of the condition/event of interest.
5. In the **Y** drop-down list, select the diagnostic test variable.

6. Optional: To show a bi-histogram of the distributions of the diagnostic test by the true condition, On the **Analyse-it** ribbon tab, in the **Diagnostic Accuracy** group, click **Bi-Histogram**
7. Optional: To show the diagnostic accuracy for each decision threshold, On the **Analyse-it** ribbon tab, in the **Diagnostic Accuracy** group, click **Sensitivity / Specificity**, **Likelihood ratio**, **Predictive value**.
8. Click **Calculate**.

Comparing two or more ROC curves

Plot multiple receiver-operator characteristics (ROC) curves to make comparisons between them.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Diagnostic**, and then under the **Accuracy** heading, click:

Compare ROC Curves	Compare 2 or more independent ROC curves.
Compare Correlated ROC Curves	Compare 2 or more paired/correlated ROC curves.

   The analysis task pane opens.
3. In the **True state** drop-down list, select the true condition variable.
4. In the **Positive event** drop-down list, select the state that indicates the presence of the condition/event of interest.
5. If comparing 2 or more independent ROC curves, in the **Y** drop-down list, select the diagnostic test variable, and then in the **Factor** drop-down list, select the grouping variable.
6. If comparing 2 or more paired/correlated ROC curves, in the **Y** list, select the diagnostic test variables.
7. Click **Calculate**.

Area under the curve (AUC)

The area under (a ROC) curve is a measure of the accuracy of a quantitative diagnostic test.

A point estimate of the AUC of the empirical ROC curve is the Mann-Whitney U estimator *(DeLong et. al., 1988)*. The confidence interval for AUC indicates the uncertainty of the estimate and uses the Wald Z large sample normal approximation *(DeLong et al., 1998)*. A test with no better accuracy than chance has an AUC of 0.5, a test with perfect accuracy has an AUC of 1.

The interpretation of the AUC is:

- The average value of sensitivity for all possible values of specificity *(Zhou, Obuchowski, McClish, 2001)*.
- The average value of specificity for all possible values of sensitivity *(Zhou, Obuchowski, McClish, 2001)*.
- The probability that a randomly selected subject with the condition has a test result indicating greater suspicion than that of a randomly chosen subject without the condition *(Hanley & McNeil, 1982)*.

AUC can be misleading as it gives equal weight to the full range of sensitivity and specificity values even though a limited range, or specific threshold, may be of practical interest.

Testing the area under the curve

Test if the area under the curve is better than chance, or some other value.

You must have already completed either of the tasks:

- *Plotting a single ROC curve* on page 204
- *Comparing two or more ROC curves* on page 205

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Diagnostic Accuracy** group, click **Test**, and then click **No Discrimination**.

   The default null hypothesis is that the AUC is less than or equal to 0.5 against the alternative hypothesis that the AUC is greater than the 0.5. This tests if the AUC is better than chance alone.

   The analysis task pane **ROC Curve** panel opens.
3. Optional: In the **Hypotheses** drop-down list, select the null and alternative hypothesis.
4. Optional: In the **Hypothesized value** edit box, type the expected value of the parameter under the null hypothesis.
5. Optional: To compare the p-value against a predefined significance level, in the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%).
6. Click **Recalculate**.

Difference between the areas under two curves

The difference between areas under the ROC curves compares the accuracy of two or more diagnostic tests.

It is imperative when comparing tests that you choose the correct type of analysis dependent on how you collect the data. If the tests are performed on the same subjects (paired design), the test results are usually correlated. Less commonly, you may perform the different tests on different groups of subjects or the same test on different groups of subjects, and the test results are independent. A paired design is more efficient and is preferred whenever possible.

A point estimate of the difference between the area under two curves is a single value that is the best estimate of the true unknown parameter; a confidence interval indicates the uncertainty of the estimate. If the tests are independent, the confidence interval uses the combined variance of the curves and a large sample Wald approximation. If the tests are paired, the standard error incorporates the covariance *(DeLong et al., 1998)* and uses a large sample Wald approximation.

A hypothesis test for the difference in AUC can test equality, equivalence, or non-inferiority of the diagnostic tests. Inferences about the difference between AUC are made using a Z test. The three hypotheses of interest are:

- Equality

  The null hypothesis states that the difference is equal to zero, against the alternative hypothesis that it is not equal to zero. When the test p-value is small, you can reject the null hypothesis and conclude the difference is not equal to zero, the tests are different.
- Equivalence

  The null hypothesis states that the difference is less than a lower bound of practical equivalence or greater than an upper bound of practical equivalence, against the alternative hypothesis that the difference is within an interval considered practically equivalent. When the test p-value is small, you can reject the null hypothesis and conclude that the tests are equivalent.
- Non-inferiority

  The null hypothesis states that the difference from a standard test is greater than the smallest practical difference against the alternative hypothesis that the difference from the standard test

is less than the smallest practical difference. When the test p-value is small, you can reject the null hypothesis and conclude that the test is not inferior to the standard test.

When the ROC curves cross, the difference between the AUC does not provide much useful information.

Testing the difference between the areas under two curves

Test if the area under curves is equal, equivalent, or not inferior to another.

You must have already completed the task:

- *Comparing two or more ROC curves* on page 205

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Diagnostic Accuracy** group, click **Test**, and then click:

Equality	Test if the AUCs are equal.
Equivalence	Test if the AUCs are equivalent within a practical difference.
Non-inferiority	Test if the AUC is not inferior to another AUC.

   The analysis task pane **Comparisons** panel opens.
3. In the analysis task pane, under the hypotheses drop-down list:
   - If testing a hypothesis of equivalence, in the **Contrast** drop-down list, select the comparisons, and then in the **Equivalent difference**, type the smallest practical difference that would be considered the same.
   - If testing a hypothesis of non-inferiority, in the **Standard** drop-down list, select the standard diagnostic test, and then in the **Smallest difference** edit box, type the smallest difference that would be considered inferior.
4. Optional: To compare the p-value against a predefined significance level, in the **Significance level** edit box, type the maximum probability of rejecting the null hypothesis when in fact it is true (typically 5% or 1%).
5. Click **Recalculate**.

Decision thresholds

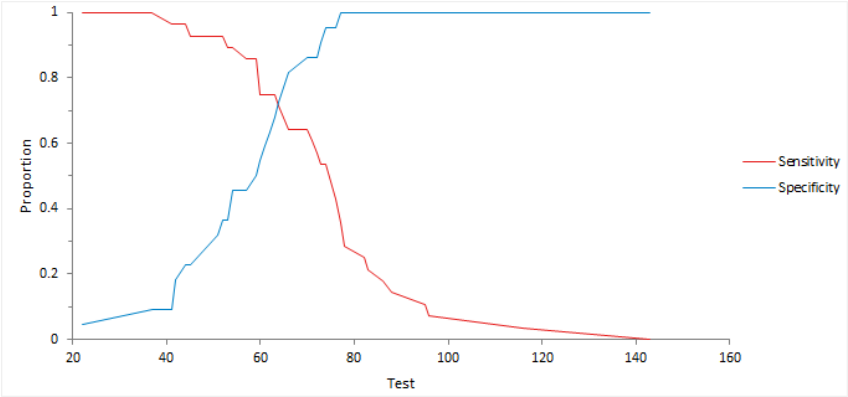
A decision threshold is a value that dichotomizes the result of a quantitative test to a simple binary decision.

The test result of a quantitative diagnostic test is dichotomized by treating the values above or equal to a threshold as positive, and those below as negative, or vice-versa.

There are many ways to choose a decision threshold for a diagnostic test. For a simple screening test, the decision threshold is often chosen to incur a fixed, true positive, or false positive rate. In more complex cases, the optimal decision threshold depends on both the cost of performing the test and the cost of the consequences of the test result. The costs may include both financial and health-related costs. A simple formula to determine the optimal decision threshold *(Zweig & Campbell, 1993)* maximizes: Sensitivity - m * (1- Specificity) where m = (Cost FP - Cost TN) / (Cost FN - Cost TP).

Decision plot

A decision plot shows a measure of performance (such as sensitivity, specificity, likelihood ratios, or predictive values) against all decision thresholds to help identify optimal decision threshold.



Finding the optimal decision threshold

Find the optimal decision threshold using information on the relative costs of making correct and incorrect decisions.

You must have already completed either of the tasks:

- *Plotting a single ROC curve* on page 204
- *Comparing two or more ROC curves* on page 205

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Diagnostic Accuracy** group, click **Find Optimal**.
   The analysis task pane **Decision** panel opens.
3. If the sample does not represent the prevalence of the condition in the target population, then, in the **Prior probability** edit box, type the **Prevalence** of the condition in the target population.
4. In the **TP**, **FP**, **TN**, **FN** edit boxes, enter the cost of each type of decision.

   The cost depends on the difference in costs between FPs and TNs relative to the difference in costs between FNs and TPs. Therefore, the costs can be expressed in any unit and need not be purely financial costs. The default costs are equal.
5. If you want a plot of the cost for each possible decision threshold:
   a) On the **Analyse-it** ribbon tab, in the **Diagnostic Accuracy** group, click **Decision Threshold**.
      The analysis task pane **Decision** panel opens.
   b) In the **Plot** drop-down list, select **Cost**.
6. Click **Recalculate**.

Predicting the decision threshold

Predict the threshold for a fixed true/false positive rate.

You must have already completed either of the tasks:

- *Plotting a single ROC curve* on page 204
- *Comparing two or more ROC curves* on page 205

1. Activate the analysis report worksheet.

2. On the **Analyse-it** ribbon tab, in the **Diagnostic Accuracy** group, click **Predict**, and then click:

Fixed FPF	Predict the TPF (Sensitivity) and Threshold at a fixed FPF
Fixed TPF (Sensitivity)	Predicts the FPF and Threshold at a fixed TPF.
Threshold	Predict the TPF (Sensitivity) and FPF at a decision threshold.

   The analysis task pane **Decision** panel opens.

3. In the **TPF / FPF / Threshold** edit box, type the value at which to determine the other parameters.
4. If you want a plot of the sensitivity and specificity for each decision threshold:
   a) On the **Analyse-it** ribbon tab, in the **Diagnostic Accuracy** group, click **Decision Threshold**.
      The analysis task pane **Decision** panel opens.
   b) In the **Plot** drop-down list, select **Sensitivity / Specificity**.
5. Click **Recalculate**.

Study design

Diagnostic performance study requirements and dataset layout.

Requirements

- A qualitative or quantitative variable of the measured values or indications (positive/negative) of the diagnostic test.
- A qualitative variable indicating the true state (positive/negative) of each subject.

Dataset layout for a single test

Use a column for the true state variable (State), and a column for the test measured values (Test); each row has the values of the variables for a case (Subject).

Subject (optional)	State	Test
1	Diseased	121
2	Diseased	118
3	Diseased	124
4	Diseased	120
5	Diseased	116
6	…	…
7	Healthy	100
8	Healthy	115
9	Healthy	102
10	Healthy	98
11	Healthy	118
…	…	…

Dataset layout for paired tests

Use a column for the true state variable (State), and a column for each of the test measured values (Test 1, Test 2); each row has the values of the variables for a case (Subject).

Subject (optional)	State	Test 1	Test 2
1	Diseased	121	86
2	Diseased	118	90
3	Diseased	124	91
4	Diseased	120	99
5	Diseased	116	89
6	…	…	…
7	Healthy	100	70
8	Healthy	115	80
9	Healthy	102	79
10	Healthy	98	87
11	Healthy	118	90
…	…	…	

Dataset layout for independent tests

Use a column for the true state variable (State), a column for the factor (test or group indicator), and a column for the test measured values (Test); each row has the values of the variables for a case (Subject).

Subject (optional)	State	Test/Group ID	Test result
1	Diseased	1	121
2	Diseased	1	118
3	Diseased	1	124
4	Diseased	1	120
5	Diseased	1	116
6	Healthy	1	100
7	Healthy	1	115
8	Healthy	1	102
9	Healthy	1	98
10	Healthy	1	118
11	…	…	…
12	Diseased	2	86
13	Diseased	2	90
14	Diseased	2	91
15	Diseased	2	99
16	Diseased	2	89
17	Healthy	2	70
18	Healthy	2	80
19	Healthy	2	79
20	Healthy	2	87
21	Healthy	2	90
…	…	…	…

Dataset layout for a single binary test

Use a column for the true state variable (State), and a column for the test indications (Test); each row has the values of the variables for a case (Subject).

Subject (optional)	State	Test
1	Diseased	P
2	Diseased	P
3	Diseased	P
4	Diseased	N
5	Diseased	P
6	Healthy	N
7	Healthy	N
8	Healthy	N
9	Healthy	N
10	Healthy	N
…	…	…

Frequency form dataset layout for a binary test

Use a column for the true state variable (State), and a column for the test indications (Test) and a column for the number of cases (Frequency); each row has the values of the variables and the frequency count.

State	Test	Count
Diseased	Positive	3
Diseased	Negative	1
Healthy	Positive	0
Healthy	Negative	5

Frequency form dataset layout for 2 independent binary tests

Use a column for the true state variable (State), a column for the factor (test or group indicator), a column for the test indications (Test) and a column for the number of cases (Frequency); each row has the values of the variables and the frequency count.

State	Test	Group/Test ID	Count
Diseased	Positive	1	3
Diseased	Negative	1	1
Healthy	Positive	1	0
Healthy	Negative	1	5

State	Test	Group/Test ID	Count
Diseased	Positive	2	24
Diseased	Negative	2	2
Healthy	Positive	2	2
Healthy	Negative	2	30

Frequency form dataset layout for 2 paired binary tests

Use a column for the true state variable (State), a column for each of the test indications (Test 1, Test 2) and a column for the number of cases (Frequency); each row has the values of the variables and the frequency count.

State	Test 1	Test 2	Count
Diseased	Positive	Positive	3
Diseased	Negative	Positive	1
Diseased	Positive	Negative	24
Diseased	Negative	Negative	2
Healthy	Positive	Positive	0
Healthy	Negative	Positive	5
Healthy	Positive	Negative	2
Healthy	Negative	Negative	30

Survival/Reliability

Chapter 22

Survival/reliability analyzes the time until an event of interest occurs.

In survival analysis, the variable of interest is the time it takes until an event occurs. Time is the length of time in hours, days, months, weeks, years or some other unit. The event of interest is often death, recovery from surgery, product failure or some other characteristic. A key feature of survival analysis is that the event of interest may not occur during the study's time frame, or contact may be lost with the participant part way through the study. In these cases, the observation is said to be censored, and we only have partial but still important information about the survival time. Most survival data is right-censored because the true survival time has been cut off at the right side of the observed time, giving us an observed survival time shorter than the true survival time.

Kaplan-Meier survival function

The Kaplan-Meier estimator (also known as the product-limit estimator) is a non-parametric statistic used to estimate the empirical survival function.

The Kaplan-Meier survival probability at failure time t(f) is the probability of surviving past the previous failure time t(f-1) multiplied by the conditional probability of surviving past time t(f) given survival to at least time t(f).

A plot of the survival probabilities at each ordered failure time produces an empirical survival function (or survival curve) - a step-function starting with a horizontal line at a survival probability of 1 and a step down at each failure time. A censored observation does not produce any step in the survival function, but it is sometimes useful to denote such observations.

A key statistic for describing the survival function are the quartiles (25th, 50th, and 75th quantiles of the survival function). Another statistic is the mean survival time, the area under the survival curve. It is often less useful as many survival functions do not drop to 0 (due to some observations being censored at the end of the study). A workaround, particularly when comparing multiple survival functions, is to restrict the area under the curve to the interval 0, t, which is common to all the survival curves.

Plotting a Kaplan-Meier survival curve

Plot the Kaplan-Meier survival function to visualize the survival probability over time.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Survival/Reliability**, and then click **Kaplan-Meier Survival Function**.
   The analysis task pane opens.
3. In the **Y (time to event)** drop-down list, select the time until the event of interest variable.
4. Optional: In the **Censor** drop-down list, select the state variable that indicates whether the time was censored or if the event of interest occured, and then in the **Event** drop-down list, select the state that indicates the time was censored.
5. Optional: In the **X (factor)** drop-down list, select the grouping variable.

6. Optional: To show the survival probabilities for each failure time, On the **Analyse-it** ribbon tab, in the **Survival/Reliability** group, click **S(t)**.
7. Click **Calculate**.

Equality of survival functions test

An equality hypothesis test formally tests if two or more population survival functions are different.

A common goal of survival analysis is to compare the survival function of two or more groups. A formal hypothesis test tests whether two or more survival functions are statistically equal in some overall sense. That is, there is no evidence to suggest that the true (population) survival functions are different. The most popular test is the Log-rank test, although various other tests differ in how much weight they give to the survival probabilities at the start or end of the distribution.

The null hypothesis states that there is no difference between the survival functions, against the alternative hypothesis that at least one of the survival functions is different. When the test p-value is small, you can reject the null hypothesis and conclude the population survival functions differ.

Tests for the equality of survival functions

Tests for the equality of survival functions and their properties and assumptions.

Test	Purpose
Log-rank	Test if the survival functions are equal (uses equal weights). The most popular test. Equivalent to the Mantel-Haenszel test of the hypothesis that the stratum specific odds-ratio is equal to one.
Wilcoxon	Test if the survival functions are equal with more weight on differences between the survival functions at smaller values of time (higher weights on early death). This test is appropriate when hazard functions vary in ways other than proportionally and when censoring patterns are similar across groups.
Tarone-Ware	Test if the survival functions are equal using weights equal to the sqrt(number at risk) rather than the number at risk used in Wilcoxon. Like the Wilcoxon test, it gives larger weights (although not as large) to earlier failure times. Although less susceptible to the censoring pattern in the data than Wilcoxon's test, this could remain a problem if large differences in these patterns exist between groups.
Fleming-Harrington	Test if the survival functions are equal using weights km^p multiplied by $(1\text{-}km)^q$. When $p > q$, the test weights earlier failures more than later ones. When $p < q$, the opposite is true, and more weight is given to later failures than to earlier ones. The Fleming–Harrington test reduces to the log-rank test when $p = 1$ and $q = 0$.

Comparing two or more survival functions

Plot multiple survival curves to make comparisons between them.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Survival/Reliability**, and then click **Compare Survival Functions**.
   The analysis task pane opens.
3. In the **Y (time to event)** drop-down list, select the time until the event of interest variable.

4. Optional: In the **Censor** drop-down list, select the state variable that indicates whether the time was censored or if the event of interest occured, and then in the **Event** drop-down list, select the state that indicates the time was censored.
5. In the **X (factor)** drop-down list, select the grouping variable.
6. In the **Survival function** task panel, under the **Equality of survival functions hypothesis test** check box, in the **Test** drop-down list, select the statistical test to perform.
7. Click **Calculate**.

Proportional hazards fit

A semi-parametric model that describes the relationship between a time to event response variable and one or more explanatory variables using a hazard function.

The Cox proportional hazard model is expressed as a hazard function $h_0(t) + \exp(\sum \#_i X_i)$. The formula expresses the hazard at time t for an individual with a given set of explanatory variables as the product of two quantities. The first quantity is $h_0(t)$, the baseline hazard function. The second quantity is the exponent of the linear sum $\#_i X_i$ where the sum is over the p explanatory variables. An important feature of the proportional hazards model is that the baseline hazard function is a function of t but does not include the Xs. Therefore, it is unnecessary to specify the form of the baseline function in estimating the parameters. Also, the parameters in the second quantity are dependent only on the Xs and are independent of time. The unknown model parameters are estimated using maximum-likelihood estimation.

Fitting a proportional hazard model

Fit a proportional hazards model to describe the relationship between a time to event response variable and one or more predictor variables.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Fit Model**, and then click **Proportional Hazard**.
   The analysis task pane opens.
3. In the **Model** drop-down menu, select **Advanced**.
4. In the **Y** drop-down list, select the response variable.
5. Optional. In the **Censor** drop-down list, select the censor variable, and then in the **Event** drop-down list, select the censor event. A censored event is an individual was lost to follow up or the event of interest had not occured at the end of the study.
6. In the **Available variables** list, select the predictor variable(s):
   - To select a single variable, click the variable.
   - To select multiple variables, click the first variable then hold down the CTRL key and click each additional variable.
   - To select a range of variables, click the first variable then hold down the SHIFT key and click the last variable in the range.
7. Optional: Click the drop-down menu arrow next to the **Available variables** list, and then select the measurement scale of the variable.

   If the measurement scale is not set, Analyse-it tries to determine if the variable is categorical or continuous scale. After you add a variable to a model, the icon alongside the variable indicates the assumed scale.
8. Click **Add** to add a variable, or click **Factorial** to add the factorial terms, or click **Polynomial** to add a polynomial terms.
9. To add the interaction between two or more variables, in the **Terms** list box, click the first term then hold down the CTRL key and click each additional term to include in the interaction, and then click **Cross**.
10. To remove a term, in the **Terms** list box, click the term and then click **Remove**.
11. Repeat steps *6* through *10* to build the model.
12. Click **Calculate**.

Hazard ratio estimates

Hazard ratios are the increase or decrease in the hazard associated with a change of the predictor, all other predictors been held constant.

Although mathematically the log hazard ratios are easier to work with when fitting models, when it comes to interpretation, it is more natural to use the hazard ratios. A hazard ratio is defined as the hazard for one individual (or group) divided by the hazard for another individual or group. A value of 1 indicates no change in the hazard. Values between 0 and less than 1 indicate a decrease in the hazard . Values greater than 1 indicate an increase in the hazard.

When the predictor variable is a categorical variable, the hazard ratio is the increase or decrease in hazard over the baseline category.

When a predictor variable is a continuous variable, the hazard ratio is the increase or decrease in hazard for a change in the predictor variable. The default is for a 1 unit change in the predictor, although it may be more appropriate to use a larger unit, such as for a change of 10 units of the predictor variable.

When a term includes an interaction, the hazard ratio will depend on the interacting variables' values, so you will need to use multiple hazard ratios to describe the behavior at different levels of interest.

Estimating hazard ratios

Estimate the hazard ratios using a proportional hazards model.

Ensure you have followed the steps in *Fitting a proportional hazard model* on page 218.

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Fit Model** group, click **Hazard Ratios**.
   The analysis task pane **Hazard Ratios** panel opens.
3. In the **Hazard Ratios** grid, in the **Comparison** column.: For a continuous variable, type the change in units of the predictor for the hazard ratio, the default is for 1 unit change in the predictor, but for some predictors the odds ratio for a 10 or 100 unit change in the predictor maybe more interpretable. For a categorical variable, select the type of comparison:

All pairs	Compare all pairs of categories against each other. For example, with 3 categories the comparisons, are 1v2,1v3,2v1,2v3,3v1,3v2.
All distinct pairs	Compare all unique pairs of categories against each other. For example, with 3 categories the comparisons as 2v1,3v1,3v2).
Against reference	Compare all categories against the reference category (the first category in the categorical variable). For example, with 3 categories the comparisons are 3v1, 2v1.

4. If the model has interaction terms, in the **At** column for each continuous variable involved in an interaction, type the levels of interest as a comma (or your list delimiter) separated list. For example, if Age is a variable involved in an interaction with Gender, you may use 16, 25, 30, 40, and 60 as the levels of interest at which the Gender Female/Male hazard ratios will be computed. For a continuous variable involved in an interaction with a categorical variable, the unit change for each level of the categorical variable will be computed. For example, if the Age unit is 10 years and Age is involved in an interaction with Gender, the hazard ratio for a change of 10 years will be computed for Females and Males.
5. In the **Confidence interval** edit box, type the confidence level.
6. Click **Recalculate**.

Baseline survival function

A baseline survivual function is an estimate of $h_0(t)$ in a proportional hazards model.

A major advantage of the proportional hazard model is that it is semi-parametric, and baseline survival function h0(t) does not need to be specified to estimate the parameters. However, sometimes the baseline survival function is useful to understand the change in survival over time and combined with the estimated coefficients and specific covariate values to estimate the survival experience of subgroups of subjects of particular interest. The baseline survival function is estimated using maximum-likelihood at #=0 for the coefficients. In many cases, the baseline survival function for #=0 isn't easily visualized, and some recommend that the mean be subtracted before continuous variables are analyzed. In these cases, the reference survival function at # =0 for categorical variables (equivalent to the reference group) and # =mean for continuous variables is usually used.

Effect of model hypothesis test

A likelihood ratio or Wald X^2 test formally tests the hypothesis of whether the model fits the data better than no model.

It is common to test whether the model fits the data better than the null model with no parameters.

A X^2 test formally tests whether the reduction is statistically significant. The null hypothesis states that all the parameters for the covariates are zero against the alternative that at least one parameter is not equal to zero. When the p-value is small, you can reject the null hypothesis and conclude that at least one parameter is not zero.

Effect of term hypothesis test

A likelihood ratio or Wald X^2 test formally tests the hypothesis of whether a term contributes to the model.

In most modeling analyses the aim is a model that describes the relationship using as few terms as possible. It is therefore of interest to look at each term in the model to decide if the term is providing any useful information.

A X^2 test for each term is a formal hypothesis test to determine if the term provides useful information to the model. The null hypothesis states that the term does not contribute to the model, against the alternative hypothesis that it does. When the p-value is small, you can reject the null hypothesis and conclude that the term does contribute to the model.

When a term is not deemed to contribute statistically to the model, you may consider removing it. However, you should be cautious of removing terms that are known to contribute by some underlying mechanism, regardless of the statistical significance of a hypothesis test, and recognize that removing a term can alter the effect of other terms.

Chapter

Control charts

23

Control charts determine if a process is in a state of statistical control.

A control chart plots a quality characteristic statistic in a time-ordered sequence. A center line indicates the process average, and two other horizontal lines called the lower and upper control limits represent process variation.

All processes have some natural degree of variation. A control chart for a process that is in-control has points randomly distributed within the control limits. That is it has variation only from sources common to the process (called common-cause variation). An out-of-control process has points falling outside the control limits or non-random patterns of points (called special-cause variation).

If the process is in-control, no corrections or changes to the process are needed.

If the process is out-of-control, the control chart can help determine the sources of variation in need of further investigation. It is appropriate to determine if the results with the special-cause are better than or worse than results from common causes alone. If worse, then that cause should be eliminated if possible. If better, it may be appropriate to investigate the system further as it may lead to improvements in the process.

Typically control limits are defined as a multiple of the process sigma. For a Shewhart control chart with 3-sigma control limits and assuming normality, the probability of exceeding the upper control limit is 0.00135 and the probability of falling below the lower control limit is also 0.00135. Their sum is 0.0027 (0.27%). Therefore the probability of a point between the control limits for an in-control process is 0.9973 (99.73%). An alternative is to define the control limits as probability limits based on a specified distribution rather than assuming a normal distribution.

Another way to look at the performance of a control chart is the average run length (ARL). An average in control run length is the number of observations when a process is in-control before a false alarm occurs. An average out of control run length is the number of observations that a process is out-of-control before a shift is detected, and depends on the size of the shift to be detected. The Shewhart control chart described above has an ARL = 1/0.0027 = 370.37. That is, when a process is in control, you should expect a false alarm out-of-control signal approximately once every 371 runs.

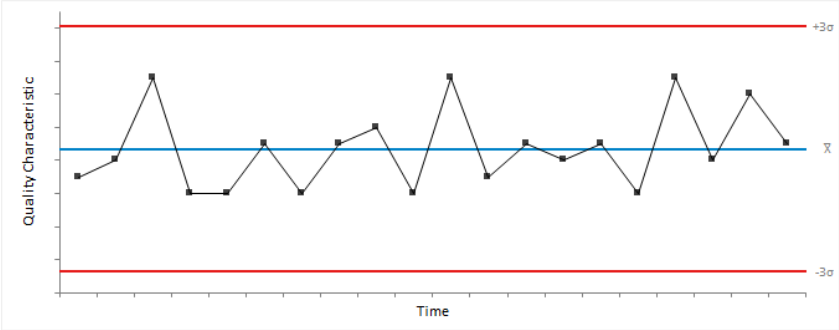
Although most examples of control charts show quality characteristics that are of interest to the end-user (such as length, diameter, or weight) they are most beneficial applied to process variables further upstream (such as the temperature of the furnace or content of tin in the raw material).

Note: It is important not to confuse control limits used in control charts with specification limits used in process capability. Natural variation in a process defines the control limits. Whereas, customer requirements define the specification limits. Likewise, the center line should not be confused with a target value.

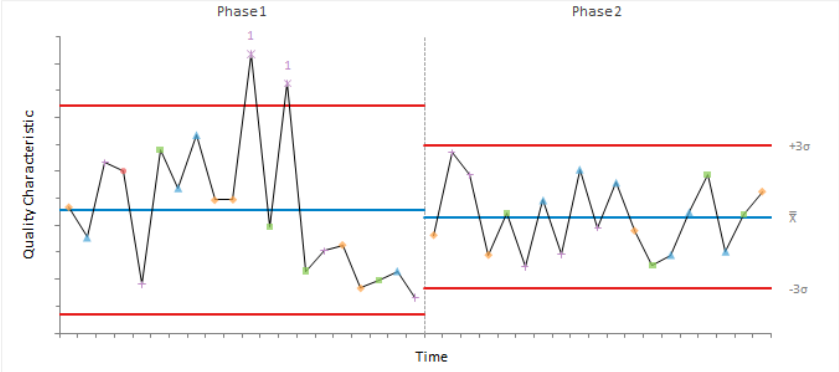
Shewhart control charts

A Shewhart control chart detects changes in a process.

Basic Shewhart control chart



Advanced Shewhart control chart with phases and stratification



Shewhart variables control charts

Variables control charts plot quality characteristics that are numerical (for example, weight, the diameter of a bearing, or temperature of the furnace).

There are two types of variables control charts: charts for data collected in subgroups, and charts for individual measurements. For sub-grouped data, the points represent a statistic of subgroups such as the mean, range, or standard deviation. For individual measurements, the points represent the individual observations or a statistic such as the moving range between successive observations.

An important concept for data collected in subgroups is that of a rational subgroup. A rational subgroup is a number of units produced under the same set of conditions. The measurements in each subgroup should be taken within a short period and should minimize the chances of special-cause variation occurring between the observations. That is, it should represent a snapshot of the process and the inherent common-cause variation.

Subgroups are often preferable when learning about a process as they can better estimate the process variability or detect smaller shifts in the process average. However, when monitoring a process, if the amount of time taken to collect enough observations to form a subgroup is too long, or the cost of sampling is expensive or destructive then the benefits may be negated.

Xbar-R chart

An Xbar-R chart is a combination of control charts used to monitor the process variability (as the range) and average (as the mean) when measuring subgroups at regular intervals from a process.

Xbar-S chart

An Xbar-S chart is a combination of control charts used to monitor the process variability (as the standard deviation) and average (as the mean) when measuring subgroups at regular intervals from a process.

Xbar chart

An Xbar-chart is a type of control chart used to monitor the process mean when measuring subgroups at regular intervals from a process.

Each point on the chart represents the value of a subgroup mean.

The center line is the process mean. If unspecified, the process mean is the weighted mean of the subgroup means.

The control limits are either:

- A multiple (k) of sigma above and below the center line. Default k=3.
- Probability limits, defined as the probability (alpha) of a point exceeding the limits. Default alpha=0.27%.

If unspecified, the process sigma is the pooled standard deviation of the subgroups, unless the chart is combined with an R- or S- chart where it is estimated as described for the respective chart.

The observations are assumed to be independent, and the means normally distributed. Individual observations need not be normally distributed. Due to the central limit theorem, the subgroup means are often approximately normally distributed for subgroup sizes larger than 4 or 5 regardless of the distribution of the individual observations.

For data with different subgroup sizes, the control limits vary. A standardized version of the control chart plots the points in standard deviation units. Such a control chart has a constant center line at 0, and upper and lower control limits of -3 and +3 respectively making patterns in the data easier to see.

It is important to ensure process variability is in a state of statistical control before using the Xbar-chart to investigate if the process mean is in control. Therefore a Xbar-chart is often combined with an R- or S-chart to monitor process variability. If the variability is not under control, the control limits may be too wide leading to an inability to detect special causes of variation affecting the process mean.

R chart

An R-chart is a type of control chart used to monitor the process variability (as the range) when measuring small subgroups ($n \leq 10$) at regular intervals from a process.

Each point on the chart represents the value of a subgroup range.

The center line for each subgroup is the expected value of the range statistic. Note that the center line varies when the subgroup sizes are unequal.

The control limits are either:

- A multiple (k) of sigma above and below the center line. Default k=3.
- Probability limits, defined as the probability (alpha) of a point exceeding the limits. Default alpha=0.27%.

If unspecified, the process sigma is the weighted average of the unbiased subgroup estimates of sigma based on the range statistics.

For small subgroup sizes with k-sigma control limits the lower limit is zero in many situations due to the asymmetric distribution of the range statistic. In these cases, probability limits may be more appropriate.

S chart

An S-chart is a type of control chart used to monitor the process variability (as the standard deviation) when measuring subgroups ($n \geq 5$) at regular intervals from a process.

Each point on the chart represents the value of a subgroup standard deviation.

The center line for each subgroup is the expected value of the standard deviation statistic. Note that the center line varies when the subgroup sizes are unequal.

The control limits are either:

- A multiple (k) of sigma above and below the center line. Default k=3.
- Probability limits, defined as the probability (alpha) of a point exceeding the limits. Default alpha=0.27%.

If unspecified, the process sigma is the weighted average of the unbiased subgroup estimates of sigma based on the standard deviation statistics.

The individual observations are assumed to be independent and normally distributed. Although the chart is fairly robust and non-normality is not a serious problem unless there is a considerable deviation from normality *(Burr, 1967)*.

For subgroup sizes less than 6 with k-sigma control limits, the lower limit is zero. In these cases, probability limits may be more appropriate.

I-MR chart

An I-MR chart is a combination of control charts used to monitor the process variability (as the moving range between successive observations) and average (as the mean) when measuring individuals at regular intervals from a process.

I chart

An I-chart is a type of control chart used to monitor the process mean when measuring individuals at regular intervals from a process.

It is typical to use individual observations instead of rational subgroups when there is no basis for forming subgroups, when there is a long interval between observations becoming available, when testing is destructive or expensive, or for many other reasons. Other charts such as the exponentially weighted moving average and cumulative sum may be more appropriate to detect smaller shifts more quickly.

Each point on the chart represents the value of an individual observation.

The center line is the process mean. If unspecified, the process mean is the mean of the individual observations.

The control limits are either:

- A multiple (k) of sigma above and below the center line. Default k=3.
- Probability limits, defined as the probability (alpha) of a point exceeding the limits. Default alpha=0.27%.

If unspecified, the process sigma is the standard deviation of the individual observations, unless the chart is combined with an MR-chart where it is estimated as described for the respective chart.

For an in-control process, both estimators are unbiased, and the standard deviation is a more efficient estimator. For a process with special causes of variation, both estimators are biased, and it may be preferable to use the median moving range estimator.

The observations are assumed to be independent and normally distributed. If the process shows a moderate departure from normality, it may be preferable to transform the variable so that it is approximately normally distributed.

It is important to ensure process variability is in a state of statistical control before using the I-chart to investigate if the process mean is in control. Therefore an I-chart is often combined with an MR-chart to monitor process variability, although some authors question the benefit of an additional MR-chart.

MR chart

An MR-chart is a type of control chart used to process variability (as the moving range of successive observations) when measuring individuals at regular intervals from a process.

When data are individual observations, it is not possible to use the standard deviation or range of subgroups to assess the variability and instead the moving range is an estimate the variability.

Given a series of observations and a fixed subset size, the first element of the moving range is the range of the initial subset of the number series. Then the subset is modified by "shifting forward"; that is, excluding the first number of the series and including the next number following the subset in the series. The next element of the moving range is the range of this subset. This process is repeated over the entire series creating a moving range statistic.

The moving range requires:

- A fixed subset size, the number of successive observations (span) in the moving range. Span must satisfy $1 < \text{span} \leq n$. Default span=2.

  Larger values have a dampening effect on the statistic and may be preferable when the data are cyclical.

Each point on the chart represents the value of a moving range.

The center line is the expected value of the range statistic.

The control limits are either:

- A multiple (k) of sigma above and below the center line. Default k=3.
- Probability limits, defined as the probability (alpha) of a point exceeding the limits. Default alpha=0.27%.

If unspecified, the process sigma is the weighted average of the unbiased moving range estimates of sigma based on the range statistics.

You should be careful when interpreting a moving range chart because the values of the statistic are correlated. Correlation may appear as a pattern of runs or cycles on the chart. Some authors *(Rigdon et al., 1994)* recommend not plotting a moving range chart as moving range does not provide any useful information about shifts in process variability beyond the I-chart.

Creating a Xbar-R / S control chart

Plot a Shewhart control chart for data collected in rational subgroups to determine if a process is in a state of statistical control.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Process Control**, and then click the type of control chart:

Xbar-R	Plot Shewhart control charts of the mean and range of subgroups.
Xbar-S	Plot Shewhart control charts of the mean and standard deviation of subgroups.
Xbar	Plot a Shewhart control chart of the mean of subgroups.
R	Plot a Shewhart control chart of the range of subgroups.
S	Plot a Shewhart control chart of the standard deviation of subgroups.

The analysis task pane opens.

3. In the **Process** drop-down list, select the quality characteristic variable.
4. In the **Sample size** group:
   - If the subgroups are identified by a variable, select **Subgroup by identifier variable**, and in the **Id** drop-down list, select the subgroup identifier variable.
   - If the subgroups are a constant size and formed by consecutive number of observations of the process variable, select **Subgroup size constant**, and in the edit box, enter the number of consecutive observations per subgroup.
5. In the **Control limits** drop-down list:
   - To define control limits based on a multiple of the process sigma, select **Sigma**, and then in the edit box, type the multiple of the process sigma to define the control limits.
   - To define control limits based on the probability distribution, select **Probability**, and then in the edit box, type the probability that a single statistic exceeds the control limits, assuming the process is in control.
6. Optional: To specify known process parameters, select the **Known process parameters** check box, and then in either or both the **Mean** and **Standard deviation** edit boxes, type the known value of the parameter.
7. Optional: To show the process mean and sigma statistics, select the **Show process statistics** check box.
8. Optional: To show the center and control limit statistics, select the **Show process control statistics** check box.
9. Click **Calculate**.

Creating an I-MR control chart

Plot a Shewhart control chart for individual observations to determine if a process is in a state of statistical control.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Process Control**, and then click the type of control chart:

I-MR	Plot Shewhart control charts of the individual observations and moving range of consecutive observations.
I	Plot a Shewhart control chart of the individual observations.
MR	Plot a Shewhart control chart of the moving range of consecutive observations.

The analysis task pane opens.

3. In the **Process** drop-down list, select the quality characteristic variable.
4. Optional: In the **Id** drop-down list, select the identifier variable.
5. In the **Control limits** drop-down list:

- To define control limits based on a multiple of the process sigma, select **Sigma**, and then in the edit box, type the multiple of the process sigma to define the control limits.
- To define control limits based on the probability distribution, select **Probability**, and then in the edit box, type the probability that a single statistic exceeds the control limits, assuming the process is in control.

6. Optional: To specify known process parameters, select the **Known process parameters** check box, and then in either or both the **Mean** and **Standard deviation** edit boxes, type the known value of the parameter.
7. Optional: To show the process mean and sigma statistics, select the **Show process statistics** check box.
8. Optional: To show the center and control limit statistics, select the **Show process control statistics** check box.
9. Click **Calculate**.

Shewhart attributes control charts

Attributes control charts plot quality characteristics that are not numerical (for example, the number of defective units, or the number of scratches on a painted panel).

It is sometimes necessary to simply classify each unit as either conforming or not conforming when a numerical measurement of a quality characteristic is not possible. In other cases, it is convenient to count the number of nonconformities rather than the number of nonconforming units. A unit may have a number of nonconformities without classing the unit as nonconforming. For example, a scratch on a painted panel may be a nonconformity but only if several such scratches exist would the entire panel be classed as a nonconforming unit.

NP chart

A np-chart is a type of control chart used to monitor the number of nonconforming units when measuring subgroups at regular intervals from a process.

Each point on the chart represents the number of nonconforming units in a subgroup.

The center line is the average number of nonconforming units. If unspecified, the process average proportion of nonconforming units is the total nonconforming units divided by the sum of subgroup sizes. Note that the center line varies when the subgroup sizes are unequal.

The control limits are either:

- A multiple (k) of sigma above and below the center line. Default k=3.
- Probability limits, defined as the probability (alpha) of a point exceeding the limits. Default alpha=0.27%.

A Binomial distribution is assumed. That is, units are either conforming or nonconforming, and that nonconformities are independent; the occurrence of a nonconforming unit at a particular point in time does not affect the probability of a nonconforming unit in the periods that immediately follow. Violation of this assumption can cause overdispersion; the presence of greater variance that would be expected based on the distribution. When k-sigma limits are used the normal approximation to the binomial distribution is assumed adequate which may require large subgroup sizes when the proportion of nonconforming units is small.

A np-chart is useful when the number of units in each subgroup is constant as interpretation is easier than a p-chart. For data with different subgroup sizes the center line and control limits both vary making interpretation difficult. In this case, you should use a p-chart which has a constant center line but varying control limits.

For small subgroup sizes, the lower control limit is zero in many situations. The lack of a lower limit is troublesome if the charts use is for quality improvement as the lower limit is desirable as points

appearing below that value may reflect a significant reduction in the number of nonconforming units. It is necessary to increase the subgroup size to overcome this issue.

P chart

A p-chart is a type of control chart used to monitor the proportion of nonconforming units when measuring subgroups at regular intervals from a process.

A p-chart is a scaled version of the np-chart representing a proportion of nonconforming units rather than the number of nonconforming units. The same assumptions and recommendations apply.

For data with different subgroup sizes, the control limits vary although the center line is constant. A standardized version of the control chart plots the points in standard deviation units. Such a control chart has a constant center line at 0, and upper and lower control limits of -3 and +3 respectively making it easier to see patterns.

C chart

A c-chart is a type of control chart used to monitor the total number of nonconformities when measuring subgroups at regular intervals from a process.

Each point on the chart represents the total number of nonconformities in a subgroup.

The center line is the average number of nonconformities. If unspecified, the process average number of nonconformities per unit is the total number of nonconformities divided by the sum of subgroup sizes. Note that the center line varies when the subgroup sizes are unequal.

The control limits are either:

- A multiple (k) of sigma above and below the center line. Default k=3.
- Probability limits, defined as the probability (alpha) of a point exceeding the limits. Default alpha=0.27%.

A Poisson distribution is assumed. That is, the probability of observing a nonconformity in the inspection unit should be small, but a large number of nonconformities should be theoretically possible, and the size of an inspection unit should also be constant over time. When k-sigma limits are used the normal approximation to the Poisson distribution is assumed adequate which usually requires the average number of nonconformities to be at least 5.

A c-chart is useful when the number of units in each subgroup is constant as interpretation is easier than a u-chart. For data with different subgroup sizes, the center line and control limits both vary making interpretation difficult. In this case, you should use a u-chart which has a constant center line but varying control limits.

U chart

A u-chart is a type of control chart used to monitor the average number of nonconformities per unit when measuring subgroups at regular intervals from a process.

A u-chart is a scaled version of the c-chart representing the average number of nonconformities per unit rather than the number of nonconformities. The same assumptions and recommendations apply.

For data with different subgroup sizes, the control limits vary although the center line is constant. A standardized version of the control chart plots the points in standard deviation units. Such a control chart has a constant center line at 0, and upper and lower control limits of -3 and +3 respectively making it easier to see patterns.

Creating an NP / P control chart

Plot a Shewhart control chart for the number/proportion of nonconforming units to determine if a process is in a state of statistical control.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Process Control**, and then click the type of control chart:

np	Plot a Shewhart control chart of the number of nonconforming units.
p	Plot a Shewhart control chart of the proportion of nonconforming units.

   The analysis task pane opens.
3. In the **Process** drop-down list, select the number of nonconforming units variable.
4. In the **Sample size** group:
   - If the subgroup sizes are variable, select **Sample size variable**, and in the **Size** drop-down list, select the corresponding sample sizes variable.
   - If the subgroups are a constant size, select **Sample size constant**, and in the edit box, enter sample size.
5. Optional: In the **Id** drop-down list, select the identifier variable.
6. In the **Control limits** drop-down list:
   - To define control limits based on a multiple of the process sigma, select **Sigma**, and then in the edit box, type the multiple of the process sigma to define the control limits.
   - To define control limits based on the probability distribution, select **Probability**, and then in the edit box, type the probability that a single statistic exceeds the control limits, assuming the process is in control.
7. Optional: To specify known process parameters, select the **Known process parameters** check box, and then in the **Proportion** edit box, type the known proportion of nonconforming units.
8. Optional: To show the process mean and sigma statistics, select the **Show process statistics** check box.
9. Optional: To show the center and control limit statistics, select the **Show process control statistics** check box.
10. Click **Calculate**.

Creating a C / U control chart

Plot a Shewhart control chart for the total number of nonconformities or the average number of nonconformities per unit to determine if a process is in a state of statistical control.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Process Control**, and then click the type of control chart:

c	Plot a Shewhart control chart of the total number of nonconformities.
u	Plot a Shewhart control chart of the average number of nonconformities per unit.

   The analysis task pane opens.
3. In the **Process** drop-down list, select the number of nonconformities variable.
4. In the **Sample size** group:

- If the subgroup sizes are variable, select **Sample size variable**, and in the **Size** drop-down list, select the corresponding sample sizes variable.
- If the subgroups are a constant size, select **Sample size constant**, and in the edit box, enter sample size.

5. Optional: In the **Id** drop-down list, select the identifier variable.
6. In the **Control limits** drop-down list:
   - To define control limits based on a multiple of the process sigma, select **Sigma**, and then in the edit box, type the multiple of the process sigma to define the control limits.
   - To define control limits based on the probability distribution, select **Probability**, and then in the edit box, type the probability that a single statistic exceeds the control limits, assuming the process is in control.
7. Optional: To specify known process parameters, select the **Known process parameters** check box, and then in the **Mean** edit box, type the known average number of nonconformities per unit.
8. Optional: To show the process mean and sigma statistics, select the **Show process statistics** check box.
9. Optional: To show the center and control limit statistics, select the **Show process control statistics** check box.
10. Click **Calculate**.

Shewhart control chart rules

Tests for special-cause variation determine when a process needs further investigation.

There are various tests that can be used in conjunction with a control chart to identify special-cause variation:

Test	Rule	Problem indicated
1	1 point is outside the control limits.	A large shift.
2	8/9 points on the same size of the center line.	A small sustained shift.
3	6 consecutive points are steadily increasing or decreasing.	A trend or drift up or down.
4	14 consecutive points are alternating up and down.	Non-random systematic variation.
5	2 out of 3 consecutive points are more than 2 sigmas from the center line in the same direction.	A medium shift.
6	4 out of 5 consecutive points are more than 1 sigma from the center line in the same direction.	A small shift.
7	15 consecutive points are within 1 sigma of the center line.	Stratification.
8	8 consecutive points on either side of the center line with not within 1 sigma.	A mixture pattern.

You should choose tests in advance of looking at the control chart based on your knowledge of the process. Applying test 1 to a Shewhart control chart for an in-control process with observations from a normal distribution leads to a false alarm once every 370 observations on average. Additional tests make the chart more sensitive to detecting special-cause variation, but also increases the chance of false alarms. For example, applying tests 1, 2, 5, 6 raises the false alarm rate to once every 91.75 observations.

Most tests beyond test 1 are only appropriate when trying to bring a process under control. Tests 2, 3, 5, and 6 detect small shifts once a process is under control although it is often preferable to using a combination of a Shewhart chart with test 1 for detection of large shifts and an EWMA or CUSUM chart for detecting smaller shifts and trends.

It can be useful to add zones at ±1, 2, and 3 sigmas to the control chart to help interpret patterns. For control charts with unequal subgroup sizes, the center line, control limits and zones may vary. In this case, the tests apply to a standardized control chart where the points are the number of standard deviation units from the center line. Such a control chart has a constant center line at 0, and upper and lower control limits of +3 and -3 respectively making patterns easier to spot.

Tests 1, 5, 6, 2 are defined by the *Western Electric CO (1958)* as the original 4 rules. Tests 1-8, with the modification of test 2 to from 8 to 9 points, are defined by *Lloyd S. Nelson (1984)*. Variations of these eight rules with different run lengths and rule ordering are recommended by various other authors, one of the most popular in the book by *(Douglas C. Montgomery, 2012)*.

Applying rules to a Shewhart control chart

Apply advanced rules to increase the sensitivity of a Shewhart control chart to detect small shifts, drift, and other patterns.

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Process Control** group, click **Apply Rules**, and then click:

Standard	Apply the standard action rule only.
WECO	Apply the original 4 rules defined by the Western Electric CO.
Nelson	Apply the 8 rules suggested by Lloyd S. Nelson.
Montgomery	Apply the 8 Nelson rules as modified by Douglas C. Montgomery.
Custom	Apply a user-defined set of rules.

   The analysis task pane **Control Chart (Shewhart)** panel opens.
3. Optional: To customize which rules to apply, select or clear the check boxes alongside the rules.
4. Optional: To change either or both the **Run** length, and **Sigma** in a custom ruleset, type the new value in the edit box(es) alongside the rules.
5. Optional: To standardize the quality characteristic statistic, so the control limits are constant regardless of the size of each subgroup, select the **Standardized plot** check box.
6. Optional: To show ±1,2,3 sigma zones for interpreting rules, select the **Zones** check box.
7. Click **Recalculate**.

A red asterisk marks points on the control chart that break a rule. A label indicates the broken rule number. Line segments between the points that form part of the broken rule are colored orange.

Note: Rule numbering depends on the ruleset applied. Only rule 1 is the same regardless of the ruleset.

Formatting a control chart

Change the style of the control chart.

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Process Control** group, under the **Format** heading, click **Style**, and then click:

Point	Show only points.
Line	Show only connecting lines, except for out-of-control points.
Point+Line	Show points and connecting lines.

The analysis task pane **Control Chart** panel opens.

3. To show the individual subgroup observations where applicable to the control chart, select the **Individual observations** check box, and then in the drop-down list select:

Dot	Show individual observations
High-Low	Show a box spanning the highest to lowest observations.
Boxplot	Show a boxplot of the observations.

4. Click **Recalculate**.

Labeling control chart points

Label control chart points to identify them or document causes of special-cause variation.

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Process Control** group, under the **Format** heading, click **Label**.
   The analysis task pane **Control Chart** panel opens.
3. Alongside the **Label points** check box, select:

All	Label all points.
Out-of-control	Label only points that are out-of-control.

4. In the **Label** drop-down list, select a variable.
5. Click **Recalculate**.

Displaying stratification in a control chart

Stratify the points on a control chart using colors and symbols to uncover causes of variation the process.

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Process Control** group, click **Stratification**.
   The analysis task pane **Control Chart** panel opens, and the **Vary point color/symbol** check box is selected.
3. In the **Color / Symbol** drop-down list, select a variable.
4. Click **Recalculate**.

Applying phases/stages for each change in a process

Use separate process statistics for different phases/stages or when other changes occur in a process that affects the process quality characteristic statistic.

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Process Control** group, click **Phases**.
3. In the **By** drop-down list, select the phase/stage, or another change variable.
4. Click **Recalculate**.

Time-weighted control charts

A time-weighted control chart detects small shifts in a process mean.

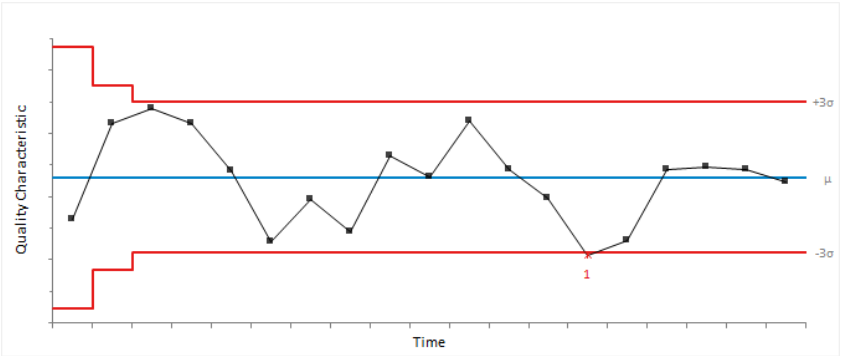
Time-weighted control charts combine information from the current and past observations. They are more sensitive to detecting small shifts in the process mean than a Shewhart chart, but slower at detecting large shifts.

The charts only monitor the process mean; monitoring the process variability requires the use of some other technique.

Uniformly Weighted Moving Average (UWMA) chart

A uniformly weighted moving average (UWMA) chart is a type of control chart used to monitor small shifts in the process mean. It uses the average of a number of consecutive observations.

The UWMA chart plots the moving average for individual measurements or subgroup means.



Given a series of observations and a fixed subset size, the first element of the moving average is the average of the initial subset of the number series. Then the subset is modified by "shifting forward"; that is, excluding the first number of the series and including the next number following the subset in the series. The next element of the moving average is the average of this subset. This process is repeated over the entire series creating the moving average statistic.

The UWMA requires:

- A fixed subset size, the number of successive observations (span) in the moving average. Span must satisfy 1 < span ≤ n. Default span=3.

  A small span reduces the influence of older observations; a large span slows the response to large shifts. In general, the magnitude of the shift to detect and the span are inversely related.

Each point on the chart represents the value of a moving average.

The center line is the process mean. If unspecified, the process mean is the weighted mean of the subgroup means or the mean of the individual observations.

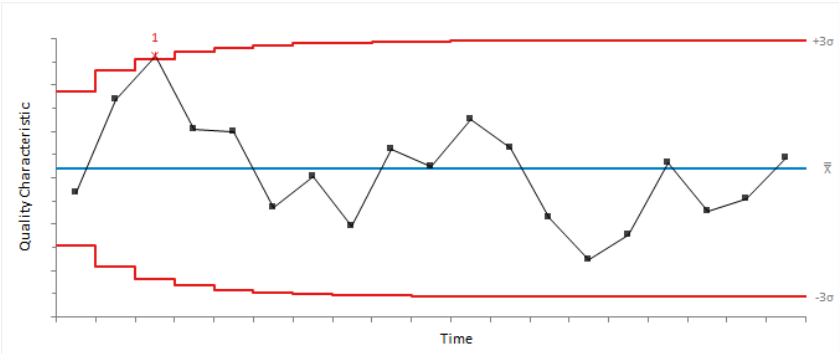
The control limits are a multiple (L) of sigma above and below the center line. Default L=3. If unspecified, the process sigma is the pooled standard deviation of the subgroups, or the standard deviation of the individual observations, unless the chart is combined with an R-, S-, or MR- chart where it is estimated as described for the respective chart.

UWMA is sensitive to small shifts in the process mean, but is not as effective as either the CUSUM or EWMA *(Montgomery 2012)*.

Exponentially Weighted Moving Average (EWMA) chart

An exponentially weighted moving average (EWMA) chart is a type of control chart used to monitor small shifts in the process mean. It weights observations in geometrically decreasing order so that the most recent observations contribute highly while the oldest observations contribute very little.

The EWMA chart plots the exponentially weighted moving average of individual measurements or subgroup means.



Given a series of observations and a fixed weight, the first element of the exponentially weighted moving average is computed by taking the (1-weight) * previous EWMA + (weight * current observation). Then the current observation is modified by "shifting forward"; and repeating the calculation. This process is repeated over the entire series creating the exponentially weighted moving average statistic.

The EWMA requires:

- A weight for the most recent observation. Weight must satisfy 0 < weight ≤ 1. Default weight=0.2.

  The "best" value is a matter of personal preference and experience. A small weight reduces the influence of the most recent sample; a large value increases the influence of the most recent sample. A value of 1 reduces the chart to a Shewhart Xbar chart. Recommendations suggest a weight between 0.05 and 0.25 *(Montgomery 2012)*.

When designing an EWMA chart it is necessary to consider the average run length and shift to be detected. Extensive guidance is available on suitable parameters *(Montgomery 2012)*.

It is possible to modify the EWMA, so it responds more quickly to detect a process that is out-of-control at start-up. This modification is done using a further exponentially decreasing adjustment to narrow the limits of the first few observations *(Montgomery 2012)*.

Each point on the chart represents the value of the exponentially weighted moving average.

The center line is the process mean. If unspecified, the process mean is the weighted mean of the subgroup means or the mean of the individual observations.

The control limits are a multiple (L) of sigma above and below the center line. Default L=3. If unspecified, the process sigma is the pooled standard deviation of the subgroups, or the standard deviation of the individual observations, unless the chart is combined with an R-, S-, or MR- chart where it is estimated as described for the respective chart.

Because the EWMA is a weighted average of all past and the current observations, it is very insensitive to the assumption of normality. It is, therefore, an ideal replacement for a Shewhart I-chart when normality cannot be assumed.

Like the CUSUM, EWMA is sensitive to small shifts in the process mean but does not match the ability of a Shewhart chart to detect larger shifts. For this reason, it is sometimes used together with a Shewhart chart *(Montgomery 2012)*.

Creating a UWMA control chart

Plot a UWMA control chart to determine if a process is in a state of statistical control.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Process Control**, and then click **UWMA**
   The analysis task pane opens.
3. In the **Model** drop-down menu:
   - If the data are individual observations, select **Process variable** .
   - If the data are rational subgroups, select **Process variable with Subgroups**.
4. In the **Process** drop-down list, select the quality characteristic variable.
5. If the data are collected in subgroups, in the **Sample size** group:
   - If the subgroups are identified by a variable, select **Subgroup by identifier variable**, and in the **Id** drop-down list, select the subgroup identifier variable.
   - If the subgroups are a constant size and formed by consecutive number of observations of the process variable, **Subgroup size constant**, and in the edit box, enter the number of consecutive observations per subgroup.
6. Optional: If the data are individual observations, in the **Id** drop-down list, select an identifier variable
7. In the **Span** edit box, type the number of consecutive values to compute the UWMA statistic.
8. In the **L** edit box, type the multiple of the process sigma to define the control limits.
9. Optional: To reset the value of the UWMA statistic after each out-of-control signal, select the **Reset after out-of-control signal** check box.
10. Optional: To specify known process parameters, select the **Known process parameters** check box, and then in either or both the **Mean** and **Standard deviation** edit boxes, type the known value of the parameter.
11. Optional: To show the process mean and sigma statistics, select the **Show process statistics** check box.
12. Optional: To show the center and control limit statistics, select the **Show process control statistics** check box.
13. Click **Calculate**.

Creating an EWMA control chart

Plot an EWMA control chart to determine if a process is in a state of statistical control.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Process Control**, and then click **EWMA**
   The analysis task pane opens.
3. In the **Model** drop-down menu:
   - If the data are individual observations, select **Process variable** .
   - If the data are rational subgroups, select **Process variable with Subgroups**.
4. In the **Process** drop-down list, select the quality characteristic variable.
5. If the data are collected in subgroups, in the **Sample size** group:

- If the subgroups are identified by a variable, select **Subgroup by identifier variable**, and in the **Id** drop-down list, select the subgroup identifier variable.
- If the subgroups are a constant size and formed by consecutive number of observations of the process variable, **Subgroup size constant**, and in the edit box, enter the number of consecutive observations per subgroup.

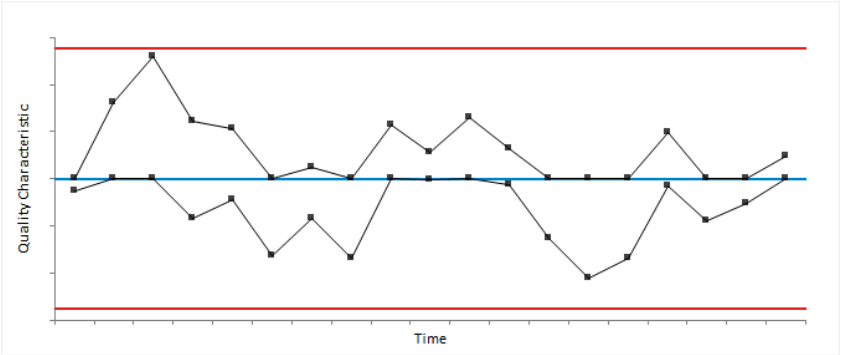
6. Optional: If the data are individual observations, in the **Id** drop-down list, select an identifier variable
7. In the **Weight** edit box, type the weight to be assigned to the current sample in the EWMA statistic.
8. In the **L** edit box, type the multiple of the process sigma to define the control limits.
9. Optional: To detect existing out-of-control situations quickly, select the **Headstart** check box, and in the **a** and **f** edit boxes, type the parameters.

   A value of a=0.3 and f=0.5 mimics a 50% headstart.
10. Optional: To reset the value of the EWMA statistic after each out-of-control signal, select the **Reset after out-of-control signal** check box.
11. Optional: To specify known process parameters, select the **Known process parameters** check box, and then in either or both the **Mean** and **Standard deviation** edit boxes, type the known value of the parameter.
12. Optional: To show the process mean and sigma statistics, select the **Show process statistics** check box.
13. Optional: To show the center and control limit statistics, select the **Show process control statistics** check box.
14. Click **Calculate**.

CUmulative SUM (CUSUM) chart

A cumulative sum (CUSUM) chart is a type of control chart used to monitor small shifts in the process mean. It uses the cumulative sum of deviations from a target.

The CUSUM chart plots the cumulative sum of deviations from the target for individual measurements or subgroup means.



The CUSUM requires two parameters:

- A reference value (k) specified in sigma units. k is often set to half the shift to be detected, in sigma units. Default k=0.5 which is equal to detecting a shift of 1 sigma.
- The decision limit (h) specified in sigma units. Default h=5.

When designing a CUSUM chart it is necessary to consider the average run length and shift to be detected. Extensive guidance is available on suitable parameters *(NIST 2012, Montgomery 2012)*.

It is possible to modify the CUSUM, so it responds more quickly to detect a process that is out-of-control at start-up. This modification is done using an initial CUSUM equal to half of the h parameter.

Like the EWMA, CUSUM is sensitive to small shifts in the process mean but does not match the ability of a Shewhart chart to detect larger shifts. For this reason, it is sometimes used together with a Shewhart chart *(Montgomery 2012)*.

Creating a CUSUM control chart

Plot a CUSUM control chart to determine if a process is in a state of statistical control.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Process Control**, and then click **CUSUM**
   The analysis task pane opens.
3. In the **Model** drop-down menu:
   - If the data are individual observations, select **Process variable** .
   - If the data are rational subgroups, select **Process variable with Subgroups**.
4. In the **Process** drop-down list, select the quality characteristic variable.
5. If the data are collected in subgroups, in the **Sample size** group:
   - If the subgroups are identified by a variable, select **Subgroup by identifier variable**, and in the **Id** drop-down list, select the subgroup identifier variable.
   - If the subgroups are a constant size and formed by consecutive number of observations of the process variable, **Subgroup size constant**, and in the edit box, enter the number of consecutive observations per subgroup.
6. Optional: If the data are individual observations, in the **Id** drop-down list, select an identifier variable
7. In the **k** and **h** edit boxes, type the design parameters for the CUSUM chart.

   Note: k and h are specified in multiples of sigma and not in the data measurement units.
8. Optional: To detect existing out-of-control situations quickly, select the **Headstart** check box, and in the **Headstart h** edit box, type the headstart value.

   Typically the headstart h is set to half the h value.
9. Optional: To standardize the CUSUM statistic, so the control limits are constant regardless of the size of each subgroup, select **Standardized plot**.
10. Optional: To reset the value of the CUSUM statistic after each out-of-control signal, select the **Reset after out-of-control signal** check box.
11. Optional: To specify known process parameters, select the **Known process parameters** check box, and then in either or both the **Mean** and **Standard deviation** edit boxes, type the known value of the parameter.
12. Optional: To show the process mean and sigma statistics, select the **Show process statistics** check box.
13. Optional: To show the center and control limit statistics, select the **Show process control statistics** check box.
14. Click **Calculate**.

Study design

Process control and capability analysis study requirements and dataset layout.

Requirements

- A categorical or quantitative variable.
- A stratification variable.
- A subgroup variable.
- A phase/stage variable.

Dataset layout for individuals or fixed subgroup size

Use a column for each variable (Copper); each row has a single measurement.

Copper
7.96
8.52
9.24
7.96
10.04
8.68
7.46
8.84
8.9
9.28
…

Advanced dataset layout with stratification and phases

Use a column for each variable (Copper), and optionally columns for subgroup (Day), stratification (Operator), phase/stage (Phase); each row has a single measurement.

Day (optional)	Operator (optional)	Phase (optional)	Copper	Comments (optional)
1	SNH	IQ	7.96	
1	SNH	IQ	8.52	
2	JDH	IQ	9.24	
2	GMH	IQ	7.96	
3	SNH	IQ	10.04	Electrode failure
3	SNH	IQ	8.68	
…	…	IQ	…	
20	JDH	OQ	8.68	
20	SNH	OQ	8.76	
21	JDH	OQ	8.02	
21	GMH	OQ	8.7	
…	…	OQ	…	

Frequency form dataset layout

Use a column for each subgroup (Sample), a column for the subgroup size (Sample size), and a column for the number of cases (Frequency); each row has a the frequency count for each sample.

Sample	Sample size	Nonconforming units
1	100	7
2	80	8
3	80	12
4	100	6
5	110	10
6	110	12
7	100	16
8	90	10
9	90	6
10	120	20
…	…	…

Chapter

Process capability

24

Capability analysis measures the ability of a process to meet specifications when the process is in statistical control.

A process must be in control before attempting to assess the capability. An out-of-control process is unpredictable and not capable of been characterized by a probability distribution.

Most process capability indices assume a normally distributed quality characteristic. If the distribution is non-normal, it may be possible to transform the data to be normally distributed. The process mean and process sigma define the normal distribution.

Capability indices are either "long-term" or "short-term" depending on the definition of the process sigma:

- Long-term indices measure the process performance and represent the quality the end-user experiences. They are computed using the process sigma that includes both within-subgroup and between-subgroup variation (the standard deviation of the individual measurements).
- Short-term indices measure the potential process performance ignoring differences between subgroups. They are computed using the process sigma that includes only within-subgroup variation (the Xbar-, R-, S-, or MR- control chart process sigma).

If the process is stable over time, the estimates of short-term sigma and long-term sigma are very similar. They are both estimates of the same parameter, although statistically speaking the long-term sigma is a slightly more efficient estimator.

However, if there are any changes in the process mean over time, the estimate of long-term sigma is greater than that of short-term sigma. The larger the difference between the values of long-term and short-term indices, the more opportunity there is to improve the process by eliminating drift, shifts and other sources of variation.

Note: There is much confusion over the meaning of the phrases long-term and short-term. It is important not to confuse them with the collection period of the sample data.

Capability ratios (Cp/Pp indices)

Capability ratios (Cp/Pp) describe the variability of a process relative to the specification limits.

A capability ratio is a unit-less value describing the ratio of process distribution spread to specification limits spread. A value of less than 1 is unacceptable, with values greater than 1.33 (1.25 for one-sided specification limits) widely accepted as the minimum acceptable value, and values greater than 1.50 (1.45 for one-sided specification limits) for critical parameters *(Montgomery, 2012)*. The higher the value, the more capable the process of meeting specifications. A value of 2 or higher is required to achieve Six Sigma capability which is defined as the process mean not closer than six standard deviations from the nearest specification limit.

All of the indices assume a normally distributed process quality characteristic with the parameters specified by the process mean and sigma. The process sigma is either the short-term or long-term sigma estimate.

Indices computed using the short-term sigma estimate are called Cp indices (Cp, Cpl, Cpu, Cpk, Cpm). While those using long-term sigma estimate are called Pp indices (Pp, Ppl, Ppu, Ppk, Ppm). If

the Cp indices are much smaller than the Pp indices, it indicates that there are improvements you could make by eliminating shifts and drifts in the process mean.

Various indices measure how the process is performing against the specification limits:

Index	Purpose
Cp/Pp	Estimates the capability of a process if the process mean were to be centered between the specification limits. **Note:** If the process mean is not centered between the specification limits the value is only the potential capability, and you should not report it as the capability of the process.
Cpl/Ppl	Estimates the capability of a process to meet the lower specification limit. Defined as how close the process mean is to the lower specification limit.
Cpu/Ppu	Estimates the capability of a process to meet the upper specification limit. Defined as how close the process mean is to the upper specification limit.
Cpk/Ppk	Estimates the capability of a process, considering that the process mean may not be centered between the specification limits. Defined as the lesser of Cpl and Cpu. **Note:** If Cpk is equal to Cp, then the process is centered at the midpoint of the specification limits. The magnitude of Cpk relative to Cp is a measure of how off center the process is and the potential improvement possible by centering the process.
Cpm/Ppm	Estimates the capability of a process, and is dependent on the deviation of the process mean from the target. **Note:** Cpm increases as the process mean moves towards the target. Cpm, Cpk, and Cp all coincide when the target is the center of the specification limits and the process mean is centered.

Note: There is some confusion between terms "Cp" and "Pp" as some authors suggest the use of Pp indices when a process is not-in-control and Cp indices when a process is in control. However, it is nonsense to interpret the indices when the process is not-in-control as no probability distribution can describe the process performance. We make the distinction between Pp and Cp indices on the estimate of sigma used not on the state of the process.

Z benchmark

Z benchmark describes the sigma capability of a process.

Z benchmark indices are an alternative to Cp and Pp indices. They are the definition of the sigma capability of a Six Sigma process.

All of the indices assume a normally distributed process quality characteristic with the parameters specified by the process mean and sigma. The process sigma is either the short-term or long-term sigma estimate.

Various indices measure how the process is performing against the specification limits:

Index	Description
< LSL	The number of sigma units from the process mean to the lower specification limit.
> USL	The number of sigma units from the process mean to the upper specification limit.

Index	Description
< SL >	The number of sigma units from the process to mean to the point if all nonconforming units are put in one tail of the distribution.

Z shift is the difference between the short-term and long-term indices. The larger the Z shift, the more scope there is to improve the process by eliminating shifts and drifts in the process mean. Some industries define the sigma capability of a process as the long-term Z benchmark + a 1.5 Z shift. Meaning a process with a long-term Z benchmark of 4.5 is quoted as a Six Sigma process. It is best to avoid such rules and directly measure the short-term and long-term Z benchmark capability.

Nonconforming units

Nonconforming units describe the number of nonconforming units a process produces, expressed in parts per million.

The number of nonconforming units is an alternative to traditional Cp and Pp indices. They are easily understandable by end-users.

The number of nonconforming units is either:

- The actual number of nonconforming units in the sample.
- The expected number of nonconforming units, assuming a normally distributed process quality characteristic with the parameters specified by the process mean and sigma. The process sigma is either the short-term or long-term sigma estimate.

Various indices measure how the process is performing against the specification limits:

Index	Description
< LSL	The number of nonconforming units that are less than the lower specification limit.
> USL	Thr number of nonconforming units that are greater than the upper specification limit.
< SL >	The total number of nonconforming units that are outside the specification limits.

Estimating process capability

Plot a process capability histogram and compute indices to determine if a process can meet specifications.

This task creates a new process capability analysis report. You can also estimate the process capability directly from a process control analysis report using the **Capability** command in the **Process Capability** group on the **Analyse-it** ribbon tab.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Capability**, and then click:

   3-up Compute capability ratios and plot a histogram and univariate plot.

   4-up Compute capability ratios and plot a histogram, univariate plot, and normal probability plot.

   6-up Compute capability ratios and plot a histogram, univariate plot, normal probability plot and control charts.

   The analysis task pane opens.

3. In the **Model** drop-down menu:
   - If the data are individual observations, select **Process variable** .
   - If the data are rational subgroups, select **Process variable with Subgroups**.
4. In the **Process** drop-down list, select the quality characteristic variable.
5. If the data are collected in subgroups, in the **Sample size** group:
   - If the subgroups are identified by a variable, select **Subgroup by identifier variable**, and in the **Id** drop-down list, select the subgroup identifier variable.
   - If the subgroups are a constant size and formed by consecutive number of observations of the process variable, **Subgroup size constant**, and in the edit box, enter the number of consecutive observations per subgroup.
6. Optional: If the data are individual observations, in the **Id** drop-down list, select an identifier variable
7. In either or both the **LSL** and **USL** edit boxes, type the lower and upper specification limits.
8. Optional: In the **Target** edit box, type the target value.
9. Optional: If you are using control charts, in the **Sigma estimator** drop-down list, select:

Long-term	Use the overall standard deviation of the individual measurements.
Short-term	Use the within-subgroup sigma of the process (using the process control R-, S-, or MR- chart).

10. Select the **Capability ratios**, **Z benchmark**, or **Nonconforming units** check boxes.
11. Click **Calculate**.

Chapter

Pareto analysis

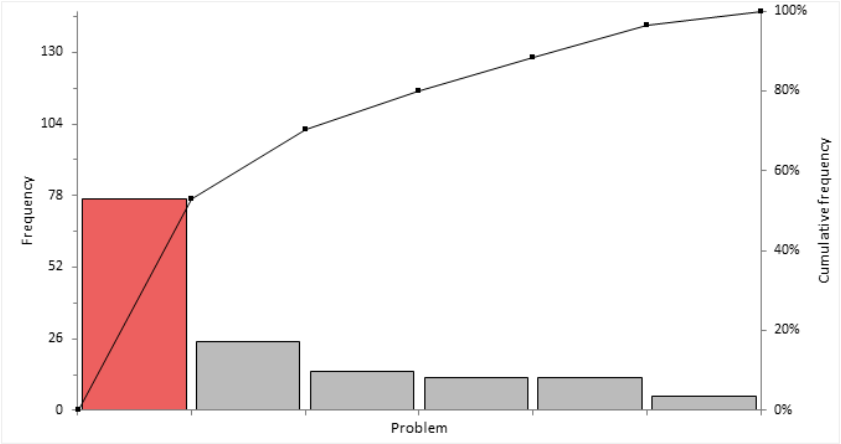
25

Pareto analysis identifies the most important quality-related problems to resolve in a process.

The Pareto principle (also known as the 80/20 rule) states that for many events, roughly 80% of the problems come from 20% of the causes. This statement is merely a rule of thumb and is not an immutable law of nature. More generally, a small subset of issues tend to cause most problems, and it is useful to identify those issues as they have the most impact on the process.

Pareto chart

A Pareto chart shows the frequency of occurrences of quality-related problems to highlight those that need the most attention.



Bars represent the individual values ordered by decreasing magnitude. A line represents the cumulative total. The left vertical axis is the frequency of occurrence or some other unit of measurement (such as cost or time). The right vertical axis is the cumulative frequency expressed as a percentage or total of the unit of measurement (such as total cost, total time).

Note: In user-interface controls and documentation, it is typical to refer to the count/frequency for usability and clarity rather than more generic terms such as measure/unit of measurement.

Creating a Pareto chart

Plot a Pareto chart to identify the most frequently occurring quality-related problems.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Pareto**, and then click **Single Pareto Chart**.
   The analysis task pane opens.
3. In the **Variable** drop-down list, select the categorical variable.
4. If the data are in frequency form, in the **Frequency** drop-down list, select the frequency count variable.

5. In the **Primary axis** drop-down list, select:

Count	Show the frequency of occurrence or unit of measurement.
Percent	Show the percentage of the total frequency or a total of the unit of measurement.

6. Optional: To force the primary and secondary axis so that 1 unit on the primary axis is equal to the percentage of the total on the secondary axis, select the **Lock axes scales** check box.
7. Optional: To label the bars, select the **Label bars** check box.
8. Optional: To label the line, select the **Label cumulative line** check box
9. Click **Calculate**.

Creating comparative Pareto charts

Plot a series of Pareto charts to identify the most frequently occurring quality-related problems stratified by up to 2 factors.

1. Select a cell in the dataset.
2. On the **Analyse-it** ribbon tab, in the **Statistical Analyses** group, click **Pareto**, and then click **Comparative Pareto Chart**.
   The analysis task pane opens.
3. In the **Variable** drop-down list, select the categorical variable.
4. If the data are in frequency form, in the **Frequency** drop-down list, select the frequency count variable.
5. In either or both the **Factor (column)** and **Factor (row)** drop-down lists, select the categorical stratification variables.
6. In the **Primary axis** drop-down list, select:

Count	Show the frequency of occurrence or unit of measurement.
Percent	Show the percentage of the total frequency or a total of the unit of measurement.

7. Optional: To force the primary and secondary axis so that 1 unit on the primary axis is equal to the percentage of the total on the secondary axis, select the **Lock axes scales** check box.
8. Optional: To label the bars, select the **Label bars** check box.
9. Optional: To label the line, select the **Label cumulative line** check box
10. Optional: In the **Layout** drop-down list, select:

Flow	Layout the plots across the width of the page, before flowing onto a new row.
Matrix	Layout the plots in a matrix as defined by the row and column factors.

11. Click **Calculate**.

Merging Pareto chart categories

Merge categories to reduce clutter and see the vital few rather than the trivial many.

You must have already completed either of the tasks:

- *Creating a Pareto chart* on page 245
- *Creating comparative Pareto charts* on page 246

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Pareto** group, click **Combine**.
   The analysis task pane, the **Categories** panel opens.

3. To merge all the categories with a frequency less than a threshold:
   a) In the **Combine** group, select the **Threshold** option.
   b) In the **Less than value / %** edit box, type the frequency count or relative frequency % threshold.
   c) In the **Name** edit box, type the name of the newly merged category.
4. To merge a fixed number of smallest categories:
   a) In the **Combine** group, select the **Last** option.
   b) In the **Categories / %** edit box, type the number of categories to merge, or the % of categories.
   c) In the **Name** edit box, type the name of the newly merged category.
5. To merge specific categories:
   a) In the **Combine** group, select the **Custom** option.
   b) In the **Categories** grid, click to select the first category, and then hold down SHIFT and click to select the last category to make a contiguous selection, or hold down CTRL and click to select additional categories to make a discontiguous selection.
   c) Click **Merge**.
   d) In the **Categories** grid, under the **Name** column, alongside the category in the edit box, type the name of the newly merged category.
6. Click **Recalculate**.

Varying Pareto chart bar colors

Color bars to highlight the importance of the vital few.

You must have already completed either of the tasks:

- *Creating a Pareto chart* on page 245
- *Creating comparative Pareto charts* on page 246

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Pareto** group, click **Vary Colors**.
   Each bar is varied in color using the color assigned to the categories of the variable in the dataset.
3. To color the bars to highlight some other attribute such as importance:
   a) Select the **Assign Colors** check box.
   b) In the **Categories** grid, under the **Color** column alongside the category, in the drop-down list, select the color. If you do not assign a color, the bar is colored gray.
   c) Repeat step *3.b* for each category.
4. Click **Recalculate**.

Reordering Pareto chart bars

Reorder bars to put a specific category at the end (for example, Miscellaneous).

You must have already completed either of the tasks:

- *Creating a Pareto chart* on page 245
- *Creating comparative Pareto charts* on page 246

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Pareto** group, click **Reorder**.
   The analysis task pane **Categories** panel opens.
3. Select the **Keep specific category at end** check box.

4. In the drop-down list, select the category to place at the end.
5. Click **Recalculate**.

Sorting bars on comparative Pareto charts

Sort the bars into the same order to make comparisons easier.

You must have already completed either of the tasks:

- *Creating a Pareto chart* on page 245
- *Creating comparative Pareto charts* on page 246

1. Activate the analysis report worksheet.
2. On the **Analyse-it** ribbon tab, in the **Pareto** group, click **Reorder**.
   The analysis task pane **Categories** panel opens.
3. Select the **Sort all by same key** check box.
4. In the drop-down list(s), select the plot to use as the sort order.
5. Click **Recalculate**.

Study design

Pareto analysis study requirements and dataset layout.

Requirements

- A categorical variable of failures.
- 1 or 2 factor variables.

Dataset layout

Use a column for the variable (Failute), an optional column for the factor variable (Operator); each row has the values of the variables for a failure (ID).

ID (optional)	Operator (optional)	Failure
1	SNH	Colorimeter drift
2	SNH	Miscellaneous
3	JDH	Electrode failure
4	GMH	Deformed tubing
5	SNH	Reagents
6	SNH	Light failure
7	JDH	Deformed tubing
8	SNH	Colorimeter drift
9	JDH	Colorimeter drift
10	GMH	Deformed tubing
…	…	…

Frequency form dataset layout

Use a column for the variable (Failute), an optional column for the factor variable (Operator) and a column for the number of cases (Frequency); each row has the values of the variables and the frequency count.

Failure	Operator (optional)	Frequency
Colorimeter drift	SNH	14
Colorimeter drift	JDH	15
Colorimeter drift	GMH	22
Deformed tubing	SNH	4
Deformed tubing	JDH	5
Deformed tubing	GMH	5
Electrode failure	SNH	1
Electrode failure	JDH	2
Electrode failure	GMH	3
…	…	…

Appendix

End-User License Agreement (EULA)

A

THIS END-USER LICENSE AGREEMENT ("EULA") IS A LEGAL AGREEMENT BETWEEN YOU (EITHER AN INDIVIDUAL OR A SINGLE ENTITY) AND ANALYSE-IT SOFTWARE LIMITED FOR THE SOFTWARE PRODUCT ANALYSE-IT FOR MICROSOFT EXCEL, WHICH INCLUDES COMPUTER SOFTWARE, PRINTED MATERIALS, AND ELECTRONIC DOCUMENTATION ("SOFTWARE"). BY INSTALLING, COPYING, OR OTHERWISE USING THE SOFTWARE, YOU AGREE TO BE BOUND BY THE TERMS OF THIS EULA. IF YOU DO NOT AGREE TO THE TERMS OF THIS EULA, DO NOT INSTALL OR USE THE SOFTWARE, AND DESTROY ALL COPIES.

The SOFTWARE is protected by copyright and international copyright treaties, as well as other intellectual property laws and treaties. The SOFTWARE is licensed, not sold.

GRANT OF LICENSE

ANALYSE-IT SOFTWARE LIMITED grants you a non-exclusive license to use the SOFTWARE subject to restrictions:

A time-limited license permits use of the SOFTWARE during the specified time period. After the time period expires you must either (a) destroy all copies of the SOFTWARE, or (b) purchase a license to use the SOFTWARE.

A fixed per-user user license permits a fixed number of users to install and use the software on their computer, and optionally on their laptop or home computer. The SOFTWARE may be installed and used only by the number of users as you have fixed per-user licenses for.

A floating concurrent-user license permits an unlimited number of users to install the software on their computer or laptop computer, but limits the number of users of the software at any time. The SOFTWARE may be used concurrently by as many users as you have floating concurrent-user licenses for. Use of the SOFTWARE requires an active Internet connection to our license server to control concurrent usage.

OTHER RIGHTS AND LIMITATIONS

You may not reverse engineer, decompile, or disassemble the SOFTWARE, except and only to the extent that such activity is expressly permitted by applicable law notwithstanding this limitation.

You may not attempt to bypass any license control restrictions imposed by the SOFTWARE.

You may not rent, lease, or loan the SOFTWARE.

You agree to hold the SOFTWARE in confidence and to take reasonable steps to prevent unauthorised copying.

You may permanently transfer all your rights under this EULA, provided you retain no copies, you transfer all of the SOFTWARE, and the recipient agrees to the terms of this EULA. If the SOFTWARE is an upgrade, any transfer must include all prior versions of the SOFTWARE.

LICENSE VIOLATION

We collect information about your computer, including a unique computer identifier to determine the number of installations of the software, and also collect information about your version of Windows, Excel, your IP address, and your name / company / e-mail address. This information is held in the strictest confidence and is used to inform you of updates and also to detect, prevent and prosecute violation of the license terms and conditions.

DUAL VERSIONS

You may receive multiple versions of the SOFTWARE for different versions of Microsoft Excel. Regardless you may only install or use the SOFTWARE on a single computer system located at your site. You may not use or install the other versions on another computer, nor loan, rent, lease or otherwise transfer them except as part of a permanent transfer as provided above.

UPGRADES

If the SOFTWARE is labelled as an upgrade, you must be properly licensed to use a product identified by ANALYSE-IT SOFTWARE LIMITED as being eligible for the upgrade. SOFTWARE labelled as an upgrade replaces and/or supplements the product that formed your eligibility for the upgrade. You may use the resulting updated SOFTWARE only in accordance with the terms of this EULA.

COPYRIGHT

All title and copyrights in and to the SOFTWARE, the accompanying printed materials, and any copies of the SOFTWARE, are owned by ANALYSE-IT SOFTWARE LIMITED. The SOFTWARE is protected by copyright laws and international treaty provisions. Therefore, you must treat the SOFTWARE like any other copyrighted material except that you may either (a) make one copy of the SOFTWARE solely

for backup or archival purposes, or (b) install the SOFTWARE in accordance with the Grant of License provided you keep the original solely for backup or archival purposes. You may not copy the printed materials accompanying the SOFTWARE.

LIMITED WARRANTY

ANALYSE-IT SOFTWARE LIMITED makes no warranty, expressed or implied, about the merchantability or fitness for a particular purpose, of the SOFTWARE during the evaluation period. You install it, and evaluate it, entirely at your own risk.

Only when a license has been purchased does ANALYSE-IT SOFTWARE LIMITED warrant that the SOFTWARE will perform substantially in accordance with the accompanying documentation for a period of ninety (90) days from the date of receipt. Any implied warranties on the SOFTWARE are limited to ninety (90) days, or the shortest period permitted by applicable law, whichever is greater.

To the maximum extent permitted by the applicable law, ANALYSE-IT SOFTWARE LIMITED disclaims all other warranties, either express or implied, including but not limited to implied warranties of merchantability and fitness for a particular purpose, with respect to the SOFTWARE, the accompanying product manual(s) and written material. The Limited Warranty contained herein gives you specific legal rights.

CUSTOMER REMEDIES

ANALYSE-IT SOFTWARE LIMITED's entire liability and your exclusive remedy shall be, at ANALYSE-IT SOFTWARE LIMITED's option, either (a) return of the price paid or (b) repair or replacement of the SOFTWARE that does not meet ANALYSE-IT SOFTWARE LIMITED's Limited Warranty and which is returned to ANALYSE-IT SOFTWARE LIMITED with a copy of the receipt. This Limited Warranty is void if failure of the SOFTWARE has resulted from an accident, abuse, or misapplication. Any replacement SOFTWARE will be warranted for the remainder of the original warranty period or thirty (30) days, whichever is longer.

LIMITED LIABILITY

To the maximum extent permitted by applicable law, ANALYSE-IT SOFTWARE LIMITED shall not be liable for any other damages whatsoever (including, without limitation, damages for loss of business profits, business interruption, business information, or other pecuniary loss) arising out of the use or inability to use this product, even if ANALYSE-IT SOFTWARE LIMITED has been advised of the possibility of such damages. In any case, ANALYSE-IT SOFTWARE LIMITED's entire liability under any provision of this agreement shall be limited to the price paid for the SOFTWARE.

TERMINATION

Without prejudice to any other rights, ANALYSE-IT SOFTWARE LIMITED may terminate this EULA if you fail to comply with the terms and conditions herein. In such event, you must destroy all copies of the SOFTWARE.

GOVERNING LAW

This EULA is governed by the laws of England, UK.

Bibliography

List of references to articles and books.

Altman, D. G., & Bland, J. M. (1983). Measurement in medicine: the analysis of method comparison studies. *The statistician*, 307-317.

Bablok, W., Passing, H., Bender, R., & Schneider, B. (1988). A general regression procedure for method transformation. Application of linear regression procedures for method comparison studies in clinical chemistry, Part III. *Clinical Chemistry and Laboratory Medicine*, *26*(11), 783-790.

Bangdiwala, S. I., & Shankar, V. (2013). The agreement chart. *BMC medical research methodology*, 13(1), 97.

Bland, J., & Altman, D. (1986). Statistical methods for assessing agreement between two methods of clinical measurement. *The Lancet*, *327*(8476), 307-310.

Bland, J. M., & Altman, D. G. (1995). Comparing methods of measurement: why plotting difference against standard method is misleading. *The Lancet*, *346*(8982), 1085-1087.

Bland, J. M., & Altman, D. G. (1999). Measuring agreement in method comparison studies. *Statistical methods in medical research*, *8*(2), 135-160.

Burdick, R. K. & Graybill, F. A. (1992). *Confidence intervals on variance components*. CRC Press.

Burr, I. W. (1967). The effect of non-normality on constants for Xbar and R charts. *Industrial Quality Control*, *23*(11), 563-569.

Carstensen, B. (2011). *Comparing clinical measurement methods: a practical guide*. John Wiley & Sons.

CLSI. (2014a). *Evaluation of Precision Performance of Quantitative Measurement Methods (EP5-A3)*. Clinical and Laboratory Standards Institute.

CLSI. (2003). *Evaluation of the Linearity of Quantitative Measurement Procedures (EP6-A)*. Clinical and Laboratory Standards Institute.

CLSI. (2020). *Evaluation of the Linearity of Quantitative Measurement Procedures (EP6-Ed2)*. Clinical and Laboratory Standards Institute.

CLSI. (2002). *Method Comparison and Bias Estimation Using Patient Samples (EP9-A2)*. Clinical and Laboratory Standards Institute.

CLSI. (2014b). *User Verification of Performance for Precision and Trueness (EP15-A3)*. Clinical and Laboratory Standards Institute.

CLSI. (2012). *Evaluation of Detection Capability for Clinical Laboratory Measurement Procedures (EP17-A2)*. Clinical and Laboratory Standards Institute.

CLSI. (2013). *Estimation of Total Analytical Error for Clinical Laboratory Methods (EP21-A)*. Clinical and Laboratory Standards Institute.

CLSI. (2010). *Defining, Establishing, and Verifying Reference Intervals in the Clinical Laboratory (EP28-A3C, formerly C28-A3C)*. Clinical and Laboratory Standards Institute.

Cicchetti, D. V., & Feinstein, A. R. (1990). High agreement but low kappa: II. Resolving the paradoxes. *Journal of clinical epidemiology*, 43(6), 551-558.

Cornbleet, P. J., & Gochman, N. (1979). Incorrect least-squares regression coefficients in method-comparison analysis. *Clinical Chemistry*, *25*(3), 432-438.

Currie, L. A. (1999). Detection and quantification limits: origins and historical overview. *Analytica Chimica Acta*, *391*(2), 127-134.

DeLong, E. R., DeLong, D. M., & Clarke-Pearson, D. L. (1988). Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics*, 837-845.

Gabriel, K. R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, *58*(3), 453-467.

Emancipator, K., & Kroll, M. H. (1993). A quantitative measure of nonlinearity. *Clinical chemistry*, *39*(5), 766-772.

Feinstein, A. R., & Cicchetti, D. V. (1990). High agreement but low kappa: I. The problems of two paradoxes. *Journal of clinical epidemiology*, 43(6), 543-549.

Gower, J. C., Lubbe, S. G., & Le Roux, N. J. (2011). *Understanding Biplots*. John Wiley & Sons.

Harrell, F. E., & Davis, C. E. (1982). A new distribution-free quantile estimator. *Biometrika*, *69*(3), 635-640.

Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29-36.

Hanley, J. A., & McNeil, B. J. (1983). A method of comparing the areas under receiver operating characteristic curves derived from the same cases. *Radiology*, *148*(3), 839-843.

Hsieh, E., & Liu, J. P. (2008). On statistical evaluation of the linearity in assay validation. *Journal of Biopharmaceutical Statistics*, *18*(4), 677-690.

Hsieh, E., Hsiao, C. F., & Liu, J. P. (2009). Statistical methods for evaluating the linearity in assay validation. *Journal of Chemometrics*, *23*(1), 56-63.

Hirji, K. F. (2005). *Exact analysis of discrete data*. CRC Press.

Horn, P. S. (1990). Robust quantile estimators for skewed populations. *Biometrika*, *77*(3), 631-636.

Horn, P. S., Pesce, A. J., & Copeland, B. E. (1998). A robust approach to reference interval estimation and evaluation. *Clinical Chemistry*, *44*(3), 622-631.

Horn, P. S., Pesce, A. J., & Copeland, B. E. (1999). Reference interval computation using robust vs. parametric and nonparametric analyses. *Clinical Chemistry*, *45*(12), 2284-2285.

Horn, P. S., & Pesce, A. J. (2005). *Reference intervals: a user's guide*. American Association for Clinical Chemistry.

Hsu, J. C. (1996). *Multiple Comparisons: Theory and Methods*. CRC Press.

ISO. (1994). *ISO 5725: Accuracy (Trueness and Precision) of Measurement Methods and Results*. International Organization for Standardization.

Jolliffe, I. (2002). *Principal component analysis*. Springer-Verlag, New York, Inc.

Kroll, M. H., & Emancipator, K. (1993). A theoretical evaluation of linearity. *Clinical chemistry*, *39*(3), 405-413.

Krouwer, J. S. (2008). Why Bland–Altman plots should use X, not (Y+ X)/2 when X is a reference method. *Statistics in Medicine*, *27*(5), 778-780.

Krouwer, J. S., & Monti, K. L. (1995). A simple, graphical method to evaluate laboratory assays. *European journal of clinical chemistry and clinical biochemistry*, *33*(8), 525-528.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.

Linnet, K. (1993). Evaluation of regression procedures for methods comparison studies. *Clinical Chemistry*, *39*, 424-424.

Linnet, K. (1990). Estimation of the linear relationship between the measurements of two methods with proportional errors. *Statistics in Medicine*, *9*(12), 1463-1473.

Linnet, K. (1998). Performance of Deming regression analysis in case of misspecified analytical error ratio in method comparison studies. *Clinical chemistry*, *44*(5), 1024-1031.

Linnet, K. (2000). Nonparametric estimation of reference intervals by simple and bootstrap-based procedures. *Clinical Chemistry*, *46*(6), 867-869.

Krouwer, J. S. (2002). Setting performance goals and evaluating total analytical error for diagnostic assays. *Clinical Chemistry*, *48*(6), 919-927.

Linnet, K., & Kondratovich, M. (2004). Partly nonparametric approach for determining the limit of detection. *Clinical chemistry*, *50*(4), 732-740.

Liu, A., & Bandos, A. I. (2012). *Statistical evaluation of diagnostic performance: topics in ROC analysis*. CRC Press.

Monti, K. L. (1995). Folded empirical distribution function curves—mountain plots. *The American Statistician*, *49*(4), 342-345.

Montgomery, D. C. (2012). *Introduction to statistical quality control*. John Wiley & Sons.

Nelson, L. S. (1984). Technical Aids: The Shewhart Control Chart -Tests for Special Causes. *Journal of quality technology*, *16*(4).

Passing, H., & Bablok, W. (1983). A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part I. *Clinical Chemistry and Laboratory Medicine*, *21*(11), 709-720.

Passing, H., & Bablok, W. (1984). Comparison of Several Regression Procedures for Method Comparison Studies and Determination of Sample Sizes Application of linear regression procedures for method comparison studies in Clinical Chemistry, Part II. *Clinical Chemistry and Laboratory Medicine*, *22*(6), 431-445.

Pepe, M. S. (2003). *The statistical evaluation of medical tests for classification and prediction*. Oxford University Press.

Pollock, M. A., Jefferson, S. G., Kane, J. W., Lomax, K., MacKinnon, G., & Winnard, C. B. (1992). Method comparison - a different approach. *Annals of clinical biochemistry*, *29*, 556-560.

Rigdon, S. E., Cruthis, E. N., & Champ, C. W. (1994). Design strategies for individuals and moving range control charts. *Journal of Quality Technology*, *26*(4), 274-287.

Ryan, T. P. (2011). *Statistical methods for quality improvement*. John Wiley & Sons.

Sadler, W. A., & Smith, M. H. (1986). A reliable method of estimating the variance function in immunoassay. *Computational Statistics & Data Analysis*, *3*, 227-239.

Sadler, W. A., Smith, M. H., & Legge, H. M. (1988). A method for direct estimation of imprecision profiles, with reference to immunoassay data. *Clinical chemistry*, *34*(6), 1058-1061.

Sadler, W. A., (2016). Using the variance function to estimate limit of blank, limit of detection and their confidence intervals. *Ann Clin Biochem*, *53*, 141-149.

Strike, P. W. (2014). *Statistical methods in laboratory medicine*. Butterworth-Heinemann.

N. I. S. T. (2012). *e-Handbook of Statistical Methods*.

Sahai, H., & Ojeda, M. M. (2004). *Analysis of Variance for Random Models, Volume 1: Balanced Data*. Springer Science & Business Media.

Sahai, H., & Ojeda, M. M. (2004). *Analysis of Variance for Random Models, Volume 2: Unbalanced Data*. Springer Science & Business Media.

Solberg, H. E. (1987). Approved recommendation on the theory of reference values. Part 5. Statistical treatment of collected reference values. Determination of reference limits. *Clinica Chimica Acta*, *170*(2), S13-S32.

Stöckl, D., Dewitte, K., & Thienpont, L. M. (1998). Validity of linear regression in method comparison studies: is it limited by the statistical model or the quality of the analytical input data?. *Clinical Chemistry*, *44*(11), 2340-2346.

Western Electric Company. (1958). *Statistical Quality Control Handbook*. AT&T.

Xue J., Titterington D. M. (2010, unpublished). The p-folded Cumulative Distribution Function and the Mean Absolute Deviation from the p-quantile.

Zhou, X. H., Obuchowski, N. A., & McClish, D. K. (2011). *Statistical methods in diagnostic medicine*. Wiley-Blackwell.

Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, *39*(4), 561-577.

Index

A

agreement measures *164*, *165*
agreement plot *166*
analysis
    cloning *48*
    creating *47*
    editing *48*
    printing *49*
    printing charts *49*
    recalculating *48*
    report *47*
    task pane *35*
ANCOVA *130*, *132*
Anderson-Darling test
    normality *64*, *64*, *197*
ANOVA
    factorial *130*, *131*
    one factor between subjects *79*, *80*
    repeated measures *91*, *92*
    table *135*
    two factor *130*, *131*
association *109*
AUC *204*, *205*, *205*, *205*, *206*, *207*
autocorrelation *143*

B

Bartlett test
    independent samples *86*, *86*
baseline survival function *220*
beta coefficients *134*, *147*
bias
    average *157*, *158*, *160*
Binomial test
    one sample *73*, *74*
biplot *118*, *121*
Bland-Altman plot *161*
Bonferroni
    multiple comparisons *84*, *85*, *140*, *141*
box plot
    side-by-side *78*
    skeletal *59*
    Tukey outlier *59*, *196*
Brown-Forsythe test
    independent samples *86*, *86*

C

capability indices *241*, *243*
case form data *40*
categories
    color *43*
    labeling *43*
    ordering *42*
    symbol *43*
CDF plot *61*, *61*
central limit theorem *64*
CLSI EP15 *176*, *180*, *181*
CLSI EP17 *186*, *187*, *188*
CLSI EP5 *176*, *177*
CLSI EP6 *183*, *184*
coefficient of variation *57*
Cohen's d
    independent samples *81*, *82*
    related samples *93*, *93*
color map *111*
compare groups *77*
compare pairs *89*
confidence interval *65*
contingency table *97*, *98*, *98*
control chart
    CUSUM *236*, *237*
    EWMA *234*, *235*
    Shewhart *222*
    UWMA *233*, *235*
Cook's D *143*, *144*
correlation
    coefficient *110*
    matrix *110*, *111*
    method comparison *151*
    monoplot *120*, *121*
covariance
    matrix *110*, *111*
    monoplot *120*
Cox's proportional hazards regression *218*
Cronbach's alpha *127*, *127*
CUSUM control chart *236*, *237*

D

dataset
    definition *37*
    filtering *44*
    go to *49*
    task pane *35*

dataset layout
compare groups *88*
compare pairs *95*
contingency table *107*
correlation *115*
diagnostic performance *210*
distribution *76*
fit model *150*
method comparison *168*, *171*
MSA *189*
multivariate *115*
reference interval *200*
decision thresholds *207*, *208*, *208*, *208*
Deming regression *153*, *154*
descriptive statistics
multivariate *110*, *110*
univariate *57*, *58*
univariate, by group *77*
detection capability *186*
detection limit
probit *188*
diagnostic performance *201*
difference between means/medians
independent samples *81*
related samples *93*
difference plot
method comparison *159*, *160*, *161*
distribution
continuous *57*
discrete *71*
dot plot
side-by-side *78*
Dunnett
multiple comparisons *84*, *84*, *85*, *85*, *140*, *140*, *141*, *141*
Durbin-Watson *143*

E

EFA *123*, *124*
eigenvalues *117*
eigenvectors *117*

F

factor analysis *123*, *123*, *124*
false negative *201*
false positive *201*
Fisher's LSD
multiple comparisons *84*, *85*, *140*, *141*
Fisher exact test
independent samples *103*, *103*
fit model
linear *129*
logistic *146*
probit *146*
proportional hazards *218*
Fleming-Harrington test *216*
folded CDF plot *163*, *163*
frequency
count *71*
cumulative *71*
cumulative relative *71*
density *71*
relative *71*
frequency distribution *60*, *71*
frequency form data *40*
frequency plot
bar *71*, *72*
grouped bar *99*
pie *72*, *72*
stacked bar *72*, *72*
Friedman test
related samples *91*, *92*
F test
effect of model *135*
effect of term *137*
independent samples *86*, *86*
lack of fit *136*

G

geometric mean *57*

H

harmonic mean *57*
hazard ratio *219*
Hedge's g
independent samples *81*, *82*
related samples *93*, *93*
histogram *61*, *62*, *195*
Hodges-Lehmann location shift
independent samples *81*, *82*
related samples *93*, *93*
Hodges-Lehmann pseudo-median *67*, *67*
Hsu
multiple comparisons *84*, *85*, *140*, *141*
hypothesis testing *67*
hypothesis tests
association *113*
binomial *73*
correlation *113*
equality of means/medians
independent samples *78*
related samples *90*
equality of proportions *101*
equality of survival functions *216*
equivalence of means
independent samples *79*
related samples *90*
homogeneity of variance *86*
independence *104*
mean *68*, *68*
median *68*
multinomial *74*
normality *63*
variance *68*

I

influence plot *143*
interaction plot *139*
interferences *185*
intermediate precision *175*
interquartile range (IQR) *57*
inverse prediction *145*
ISO 5725 *176*
item reliability *127*, *127*

K

Kaplan-Meier survival function *215*, *215*
kappa *165*
Kendall tau *112*, *112*, *113*
Kolmogorov-Smirnov test
    normality *64*, *64*
Kruskal-Wallis test
    independent samples *79*, *80*
kurtosis *57*

L

lag plot *141*, *143*
least-squares means *138*
Levene test
    independent samples *86*, *86*
leverage *143*, *144*
leverage plot *138*
license
    acquiring *24*
    activating *21*
    agreement *251*
    deactivating *23*
    releasing *24*
    transferring *18*
    who's using *24*
likelihood ratio *201*
likelihood ratio G^2 test
    independence *105*, *105*
    one sample *75*, *75*
likelihood ratio X^2 test
    effect of model *148*, *220*
    effect of term *149*, *220*
limit of blank *186*
limit of detection *186*
limit of quantitation *188*
limits of agreement *161*, *161*
linearity *156*, *157*, *182*, *183*, *184*
logit *146*
log odds ratio *147*, *219*
Log-rank test *216*

M

main effects plot *139*
maintenance
    checking expiry *28*
    renewing *28*
matrix rotation *124*
maximum *57*
McNemar-Mosteller exact test
    related samples *103*, *104*
McNemar X^2 test
    related samples *103*, *104*
mean *57*, *67*, *67*
mean difference
    independent samples *81*, *82*
    related samples *93*, *93*
mean-mean scatter plot *83*
mean plot
    side-by-side *78*
measurement scale *38*
measurement systems analysis *175*
median *57*, *67*, *67*
median difference
    related samples *93*
method comparison *151*
minimum *57*
missing values *40*
mode *57*
monoplot *118*, *120*, *121*
mosaic plot *106*, *106*
mountain plot *163*, *163*
moving average control chart
    EWMA *234*, *235*
    UWMA *233*, *235*
MSA *175*
multiple comparisons *82*, *85*, *139*, *141*

N

nonconforming units *243*, *243*
normal distribution *62*
normality *62*, *64*, *142*
normal probability plot *63*, *63*, *197*, *197*
numerical accuracy *50*

O

odds *73*
odds ratio
    logistic *148*
ordinary linear regression
    method comparison *152*, *153*
outlier *143*
outliers *48*, *59*, *144*

P

Pareto analysis *245*
Pareto chart *245*
Passing-Bablok regression *153*, *155*
pattern matrix *123*
PCA *117*, *118*
Pearson r *112*, *112*, *113*
Pearson X^2 test
    independence *105*, *105*
    independent samples *103*, *103*

one sample *75*, *75*
point estimate *65*
PPA/NPA *164*
precision *175*, *175*, *176*, *177*
precision profile *178*, *178*, *178*, *179*, *187*
predicted against actual Y plot *136*
predictive value *202*
principal components *117*, *117*, *118*
process capability *241*
proportion *73*
proportion difference *99*
proportion ratio *99*
p-value
asymptotic *102*
exact *102*
hypothesis testing *67*

Q

Q-Q plot *63*, *63*, *197*
qualitative variable *38*
quantile
bootstrap *194*
Harrell-Davis *194*
normal *193*
robust bi-weight *194*
quantiles *57*
quantitative variable *38*
quartiles *57*

R

range *57*
reference interval *193*, *193*, *194*, *198*
reference limit *193*, *194*, *198*
regression
advanced *130*, *132*
exponential *129*, *129*
linear *129*, *129*
logarithmic *129*, *129*
logistic *146*, *146*, *146*
multiple linear *130*, *130*
polynomial *129*, *129*
power *129*, *129*
probit *146*, *147*
proportional hazard *218*
relative risk *99*, *101*
relative standard deviation *57*
repeatability *175*
reproducibility *175*
residual plot *141*, *143*, *157*, *157*
ribbon *35*
risk difference *99*
risk ratio *99*
ROC plot *203*, *204*, *205*
R^2 *133*

S

scatter plot
correlation / association *109*, *109*
matrix *109*, *109*
method comparison *152*
regression *133*
Scheffé
multiple comparisons *84*, *85*, *140*, *141*
Score Z test
independent samples *103*, *103*
one sample *73*, *74*
related samples *103*, *104*
scree plot *118*
sensitivity *201*, *202*, *203*
sequence plot *141*, *143*
Shapiro-Wilk test
normality *64*, *64*, *197*
Shewhart control chart
attributes *227*
c *228*, *229*
I *224*, *226*
I-MR *224*, *226*
MR *225*, *226*
np *227*, *229*
p *228*, *229*
R *223*, *225*
rules *230*
S *224*, *225*
u *228*, *229*
variables *222*
Xbar *223*, *225*
Xbar-R *223*, *225*
Xbar-S *223*, *225*
Sign test
one sample *69*, *69*
related samples *91*, *92*
skewness *57*
software
installing *17*
installing for concurrent-use *18*
uninstalling *19*
updating *27*
Spearman rs *112*, *112*
specificity *201*, *202*, *203*
standard deviation *57*, *67*, *67*
standardized beta coefficients *134*
standardized mean difference
independent samples *81*, *82*
related samples *93*, *93*
Steel
multiple comparisons *84*, *85*
Steel-Dwass-Critchlow-Flinger
multiple comparisons *84*, *85*
strip plot *59*
structure matrix *123*
Student's t test
independent samples *79*, *80*
multiple comparisons *84*, *85*, *140*, *141*
one sample *69*, *69*
related samples *91*, *92*

Student-Newman-Keuls
multiple comparisons *84*, *85*
summary of fit *133*
survival analysis *215*
survival curve plot *215*
survival function *216*
system requirements *17*

T

Tarone-Ware test *216*
TOST
one sample *70*
TOST (two-one-sided t-tests)
independent samples *79*, *81*
repeated measures *91*, *92*
troubleshooting *29*
true negative *201*
trueness *180*
true positive *201*
Tukey-Kramer
multiple comparisons *84*, *85*, *140*, *141*
two-one-sided t tests
one sample *70*

U

univariate plot
side-by-side *78*, *78*
single *59*, *60*

V

variable
definition *38*
saving *145*
setting axis scale *41*
setting measurement scale *41*
setting number format *42*
transforming *44*
variance *57*, *67*, *67*
variance components *175*
variance inflation factors *134*

W

Wald X^2 test
effect of model *148*, *220*
effect of term *149*, *220*
weighted Deming regression *153*, *154*
weighted linear regression
method comparison *152*, *153*
Welch ANOVA
independent samples *79*, *80*
Welch t test
independent samples *79*, *80*
Wilcoxon-Gehan test *216*
Wilcoxon-Mann-Whitney test
independent samples *79*, *80*
multiple comparisons *84*, *85*
Wilcoxon test
one sample *69*, *69*
related samples *91*, *92*

X

X^2 test
lack of fit *136*
X^2 variance test
one sample *69*, *69*

Y

Yates correction *105*
Youden J *202*

Z

Z benchmark indices *242*, *243*
Z test
independent samples *79*, *80*
one sample *69*, *69*
related samples *91*, *92*