Basic Analysis and Graphing
Get the Most from JMP®

Whether you are a first-time or a long-time user, there is always something to learn about JMP.
Visit JMP.com and to find the following:

• live and recorded Webcasts about how to get started with JMP
• video demos and Webcasts of new features and advanced techniques
• schedules for seminars being held in your area
• success stories showing how others use JMP
• a blog with tips, tricks, and stories from JMP staff
• a forum to discuss JMP with other users

http://www.jmp.com/getstarted/
Basic Analysis and Graphing

“The real voyage of discovery consists not in seeking new landscapes, but in having new eyes.”

Marcel Proust
# Preliminaries

## Introducing JMP

- Prerequisites ............................................. 3
- JMP Terminology .......................................... 3

## Learning about JMP

- About JMP Documentation ................................. 3
- Use JMP Help ................................................ 6
- Use Tutorials ............................................... 7
- Access Sample Data Tables .............................. 7
- Learn About Statistical and JSL Terms ................. 7
- Learn JMP Tips and Tricks ............................... 8
- Access Resources on the Web ............................ 8

## Conventions

- Conventions ................................................ 8

## Use JMP Platforms

- Use JMP Platforms ........................................ 8
- Work with Multiple Data Tables and Platforms ....... 8
- How JMP Platforms Are Designed ....................... 9
- Process for Analyzing Data Using Platforms ......... 9

## Common Features Throughout Platforms

- Common Features Throughout Platforms ............... 13
- Launch Window Features .................................. 13
- Script Menus ................................................. 15
- Automatic Recalc Feature ................................. 17

# Performing Univariate Analysis

## Using the Distribution Platform

- Overview of the Distribution Platform ................ 21
- Categorical Variables ...................................... 21
- Continuous Variables .................................... 21

## Example of the Distribution Platform ................ 21

## Launch the Distribution Platform ....................... 23
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Distribution Report</td>
<td>24</td>
</tr>
<tr>
<td>Histograms</td>
<td>26</td>
</tr>
<tr>
<td>- Resize Histogram Bars for Continuous Variables</td>
<td>27</td>
</tr>
<tr>
<td>- Highlight Bars and Select Rows</td>
<td>27</td>
</tr>
<tr>
<td>- Specify Your Selection in Multiple Histograms</td>
<td>28</td>
</tr>
<tr>
<td>Initial Reports</td>
<td>29</td>
</tr>
<tr>
<td>- The Frequencies Report</td>
<td>29</td>
</tr>
<tr>
<td>- The Quantiles Report</td>
<td>31</td>
</tr>
<tr>
<td>- The Moments Report</td>
<td>32</td>
</tr>
<tr>
<td>Distribution Platform Options</td>
<td>34</td>
</tr>
<tr>
<td>Options for Categorical Variables</td>
<td>34</td>
</tr>
<tr>
<td>- Display Options for Categorical Variables</td>
<td>34</td>
</tr>
<tr>
<td>- Histogram Options for Categorical Variables</td>
<td>35</td>
</tr>
<tr>
<td>- Mosaic Plot</td>
<td>35</td>
</tr>
<tr>
<td>- Test Probabilities</td>
<td>36</td>
</tr>
<tr>
<td>- Confidence Intervals for Categorical Variables</td>
<td>39</td>
</tr>
<tr>
<td>- Save Commands for Categorical Variables</td>
<td>39</td>
</tr>
<tr>
<td>Options for Continuous Variables</td>
<td>39</td>
</tr>
<tr>
<td>- Display Options for Continuous Variables</td>
<td>39</td>
</tr>
<tr>
<td>- Histogram Options for Continuous Variables</td>
<td>40</td>
</tr>
<tr>
<td>- Normal Quantile Plot</td>
<td>42</td>
</tr>
<tr>
<td>- Outlier Box Plot</td>
<td>43</td>
</tr>
<tr>
<td>- Quantile Box Plot</td>
<td>45</td>
</tr>
<tr>
<td>- Stem and Leaf</td>
<td>46</td>
</tr>
<tr>
<td>- CDF Plot</td>
<td>47</td>
</tr>
<tr>
<td>- Test Mean</td>
<td>48</td>
</tr>
<tr>
<td>- Test Std Dev</td>
<td>50</td>
</tr>
<tr>
<td>- Confidence Intervals for Continuous Variables</td>
<td>51</td>
</tr>
<tr>
<td>- Save Commands for Continuous Variables</td>
<td>52</td>
</tr>
<tr>
<td>Prediction Intervals</td>
<td>54</td>
</tr>
<tr>
<td>Tolerance Intervals</td>
<td>55</td>
</tr>
<tr>
<td>Capability Analysis</td>
<td>57</td>
</tr>
<tr>
<td>Fit Distributions</td>
<td>61</td>
</tr>
<tr>
<td>- Example of Fitting a Lognormal Distribution</td>
<td>61</td>
</tr>
<tr>
<td>- Continuous Fit</td>
<td>62</td>
</tr>
<tr>
<td>- Discrete Fit</td>
<td>69</td>
</tr>
</tbody>
</table>
3 Introduction to the Fit Y by X Platform
Performing Four Types of Analyses ............................................. 83
Overview of the Fit Y by X Platform ........................................ 85
Launch the Fit Y by X Platform .............................................. 85
Launch Specific Analyses from the JMP Starter Window .......... 86

4 Performing Bivariate Analysis
Using the Fit Y by X or Bivariate Platform ............................... 87
Example of Bivariate Analysis .............................................. 89
Launch the Bivariate Platform .............................................. 89
Example of a Bivariate Report ............................................. 90
Overview of Fitting Commands and General Options ........... 91
Fitting Command Categories .............................................. 92
Fit the Same Command Multiple Times ............................... 93
Fit Mean ................................................................. 93
Fit Mean Menu .......................................................... 94
Fit Mean Report .......................................................... 94
Fit Line and Fit Polynomial .............................................. 95
Linear Fit and Polynomial Fit Menus ................................. 95
Linear Fit and Polynomial Fit Reports ............................... 96
Fit Special .............................................................. 103
Fit Special Reports and Menus ......................................... 105
Fit Spline .............................................................. 106
Smoothing Spline Fit Report ........................................... 107
Smoothing Spline Fit Menu ............................................. 108
Fit Each Value .......................................................... 108
Fit Each Value Report .................................................. 108
Fit Each Value Menu ................................................... 109
The t-test Report .................................................. 140
The Analysis of Variance Report ................................. 142
The Means for OneWay Anova Report ......................... 143
The Block Means Report ......................................... 144
Mean Diamonds and X-Axis Proportional ....................... 144
Mean Lines, Error Bars, and Standard Deviation Lines ....... 145
Analysis of Means Methods ...................................... 146
Compare Means ..................................................... 147
Compare Standard Deviations (or Variances) ................. 147
Analysis of Means Charts ....................................... 148
Analysis of Means Options ..................................... 149
Compare Means ..................................................... 150
Using Comparison Circles ....................................... 151
Each Pair, Student's t ............................................. 153
All Pairs, Tukey HSD ............................................ 155
With Best, Hsu MCB .............................................. 157
With Control, Dunnett's ......................................... 159
Compare the Four Tests .......................................... 160
Compare Means Options ......................................... 161
Nonparametric ...................................................... 162
Nonparametric Report Descriptions ............................ 164
Unequal Variances ................................................ 167
Example of the Unequal Variances Option .................... 168
Equivalence Test .................................................... 172
Example of an Equivalence Test ................................ 172
Power ............................................................... 173
Example of the Power Option .................................. 173
Normal Quantile Plot ............................................. 176
Example of a Normal Quantile Plot .............................. 176
CDF Plot .......................................................... 177
Example of a CDF Plot ......................................... 177
Densities .......................................................... 178
Example of the Densities Options ............................... 178
Matching Column ................................................. 180
Example of the Matching Column Option ...................... 180
<table>
<thead>
<tr>
<th>Topic</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistical Details</td>
<td>182</td>
</tr>
<tr>
<td>Comparison Circles</td>
<td>182</td>
</tr>
<tr>
<td>Power</td>
<td>183</td>
</tr>
<tr>
<td>6 Performing Contingency Analysis</td>
<td>185</td>
</tr>
<tr>
<td>Using the Fit Y by X or Contingency Platform</td>
<td></td>
</tr>
<tr>
<td>Example of Contingency Analysis</td>
<td>187</td>
</tr>
<tr>
<td>Launch the Contingency Platform</td>
<td>188</td>
</tr>
<tr>
<td>The Contingency Report</td>
<td>188</td>
</tr>
<tr>
<td>Contingency Platform Options</td>
<td>189</td>
</tr>
<tr>
<td>Mosaic Plot</td>
<td>191</td>
</tr>
<tr>
<td>Context Menu</td>
<td>192</td>
</tr>
<tr>
<td>Tests</td>
<td>195</td>
</tr>
<tr>
<td>Fisher's Exact Test</td>
<td>197</td>
</tr>
<tr>
<td>Analysis of Means for Proportions</td>
<td>197</td>
</tr>
<tr>
<td>Example of Analysis of Means for Proportions</td>
<td>197</td>
</tr>
<tr>
<td>Correspondence Analysis</td>
<td>199</td>
</tr>
<tr>
<td>Understanding Correspondence Analysis</td>
<td>199</td>
</tr>
<tr>
<td>Example of Correspondence Analysis</td>
<td>199</td>
</tr>
<tr>
<td>Correspondence Analysis Options</td>
<td>201</td>
</tr>
<tr>
<td>The Details Report</td>
<td>202</td>
</tr>
<tr>
<td>Additional Example of Correspondence Analysis</td>
<td>204</td>
</tr>
<tr>
<td>Cochran-Mantel-Haenszel Test</td>
<td>206</td>
</tr>
<tr>
<td>Example of a Cochran Mantel Haenszel Test</td>
<td>206</td>
</tr>
<tr>
<td>Agreement Statistic</td>
<td>207</td>
</tr>
<tr>
<td>Example of the Agreement Statistic</td>
<td>207</td>
</tr>
<tr>
<td>Relative Risk</td>
<td>209</td>
</tr>
<tr>
<td>Example of the Relative Risk Option</td>
<td>209</td>
</tr>
<tr>
<td>Two Sample Test for Proportions</td>
<td>211</td>
</tr>
<tr>
<td>Example of a Two Sample Test for Proportions</td>
<td>211</td>
</tr>
<tr>
<td>Measures of Association</td>
<td>212</td>
</tr>
<tr>
<td>Example of the Measures of Association Option</td>
<td>213</td>
</tr>
<tr>
<td>Cochran Armitage Trend Test</td>
<td>214</td>
</tr>
<tr>
<td>Example of the Cochran Armitage Trend Test</td>
<td>214</td>
</tr>
</tbody>
</table>
7 Performing Simple Logistic Regression
Using the Fit Y by X or Logistic Platform

- Exact Test ................................................................. 215
- Statistical Details for the Agreement Statistic Option .................. 216

8 Comparing Paired Data
Using the Matched Pairs Platform

- Overview of Logistic Regression ........................................ 219
- Nominal Logistic Regression ............................................ 221
- Ordinal Logistic Regression ............................................. 221
- Example of Nominal Logistic Regression .............................. 221
- Launch the Logistic Platform ........................................... 223
- Logistic Report ............................................................ 223
  - Logistic Plot ......................................................... 225
  - Iterations ......................................................... 225
  - Whole Model Test ............................................... 225
  - Parameter Estimates ............................................. 227
- Logistic Platform Options .............................................. 228
  - ROC Curves ....................................................... 229
  - Save Probability Formula ....................................... 231
  - Inverse Prediction ............................................... 231
- Example of Ordinal Logistic Regression ................................ 233
- Additional Example of a Logistic Plot ................................ 235

- Overview of the Matched Pairs Platform ................................ 239
- Example of Comparing Matched Pairs ................................ 241
- Launch the Matched Pairs Platform .................................... 241
  - Multiple Y Columns ............................................... 243
- The Matched Pairs Report ............................................ 243
  - Difference Plot and Report ..................................... 244
  - Across Groups .................................................... 244
- Matched Pairs Options ................................................. 245
- Example of Comparing Matched Pairs Across Groups ............... 246
9 Interactive Data Visualization

Using Graph Builder .................................................. 253

Overview of Graph Builder ............................................ 255
Example Using Graph Builder ......................................... 255
Launch Graph Builder ..................................................... 261
The Graph Builder Window .............................................. 262
Platform Buttons .......................................................... 264
Graph Builder Options .................................................... 264
Graph Builder Right-Click Menus ....................................... 264
Add Variables ............................................................... 268
Example of Adding Variables .......................................... 268
Move Grouping Variable Labels ....................................... 270
Separate Variables into Groups ......................................... 270
Change Variable Roles .................................................... 271
Use the Swap Command .................................................. 271
Use the Clicking and Dragging Method to Change Variable Roles
Remove Variables ........................................................ 272
Use the Remove Command .............................................. 272
Use the Clicking and Dragging Method to Remove Variables ...
Add Multiple Variables to the X or Y Zone ....................... 274
Example of Adding Multiple Variables to the X or Y Zone ...
Merge Variables ........................................................... 276
Order Variables ............................................................. 277
Replace Variables .......................................................... 280
Create a Second Y Axis ................................................... 280
Add Multiple Variables to Grouping Zones ....................... 282
Example of Adding Multiple Variables to Grouping Zones ...
Replace Variables .......................................................... 284
Order Grouping Variables ............................................... 285
Modify the Legend .......................................................... 285
Create Map Shapes .......................................................... 286
Example of Creating Map Shapes .................................................. 286
Built-in Map Files .......................................................................... 287
Create Custom Map Files ................................................................. 288

Additional Examples Using Graph Builder .......................................... 289
Measure Global Oil Consumption and Production ............................... 289
Analyze Popcorn Yield .................................................................... 296
Examine Diamond Characteristics .................................................... 301

10 Creating Summary Charts
Using the Chart Platform .................................................................. 305
Example of the Chart Platform .......................................................... 307
Launch the Chart Platform ................................................................ 309
Plot Statistics for Y Variables ............................................................ 312
Use Categorical Variables ................................................................ 313
Use Grouping Variables ................................................................... 314
Adding Error Bars ........................................................................... 316
The Chart Report ............................................................................... 317
Legends ............................................................................................ 317
Ordering ............................................................................................ 318
Coloring Bars in a Chart ................................................................... 318
Chart Platform Options ..................................................................... 319
General Platform Options ................................................................. 319
Y Options .......................................................................................... 321
Examples of Charts ........................................................................... 321
Plot a Single Statistic ....................................................................... 322
Plot Multiple Statistics .................................................................... 322
Plot Counts of Variable Levels ......................................................... 323
Plot Multiple Statistics with Two X Variables .................................... 325
Create a Stacked Bar Chart ............................................................... 326
Create a Pie Chart ............................................................................ 327
Create a Range Chart ....................................................................... 329
Create a Chart with Ranges and Lines for Statistics ......................... 330

11 Creating Overlay Plots
Using the Overlay Plot Platform ....................................................... 333
Example of an Overlay Plot .............................................................. 335
Launch the Overlay Plot Platform ...................................................... 336
The Overlay Plot Report .................................................. 337
Overlay Plot Options ..................................................... 338
General Platform Options .............................................. 338
Y Options ................................................................. 341
Additional Examples of Overlay Plots .......................... 342
Function Plots ............................................................ 342
Plotting Two or More Variables with a Second Y-axis .... 343
Grouping Variables ....................................................... 345

12 Creating Three-Dimensional Scatterplots
Using the Scatterplot 3D Platform ............................... 347
Example of a 3D Scatterplot ......................................... 349
Launch the Scatterplot 3D Platform .............................. 350
The Scatterplot 3D Report ............................................. 351
Spin the 3D Scatterplot ................................................ 352
Change Variables on the Axes ...................................... 352
Adjust the Axes .......................................................... 353
Assign Colors and Markers to Data Points .................. 354
Assign Colors and Markers in the Data Table ............... 354
Scatterplot 3D Platform Options ................................. 355
Normal Contour Ellipsoids .......................................... 357
Nonparametric Density Contours ................................. 359
Context Menu ........................................................... 364

13 Creating Contour Plots
Using the Contour Plot Platform ................................. 367
Example of a Contour Plot ............................................ 369
Launch the Contour Plot Platform ............................... 369
The Contour Plot Report .............................................. 371
Contour Plot Options .................................................. 371
Fill Areas ................................................................. 372
Contour Specification ................................................. 373
Contour Plot Save Options ......................................... 375
Use Formulas for Specifying Contours ....................... 375
### Creating Bubble Plots

**Using the Bubble Plot Platform**

- Example of a Dynamic Bubble Plot
- Launch the Bubble Plot Platform
- Interact with the Bubble Plot
  - Control Animation for Dynamic Bubble Plots
  - Specify a Time or ID Variable
  - Select Bubbles
  - Use the Brush Tool
- Bubble Plot Platform Options
  - Show Roles
- Additional Examples
  - Example of a Static Bubble Plot
  - Example of a Bubble Plot with a Categorical Y Variable

### Creating Cell-Based Plots

**Using the Parallel and Cell Plot Platforms**

- Example of a Parallel Plot
- Launch the Parallel Plot Platform
- The Parallel Plot Report
  - Interpreting Parallel Plots
- Parallel Plot Platform Options
- Additional Examples of Parallel Plots
  - Examine Iris Measurements
  - Examine Student Measurements
- Example of a Cell Plot
- Launch the Cell Plot Platform
- The Cell Plot Report
  - Context Menu for Cell Plots
- Additional Example of a Cell Plot

### Creating Tree Maps

**Using the Tree Map Platform**

- Example of Tree Maps
Launch the Tree Map Platform .................................................. 419
Categories .............................................................................. 420
Sizes ....................................................................................... 421
Ordering ................................................................................... 422
Coloring ................................................................................... 424
Tree Map Report Window .......................................................... 426
Tree Map Platform Options ....................................................... 427
Context Menu .......................................................................... 428
Additional Tree Map Examples .................................................. 428
Examine Pollution Levels .......................................................... 428
Examine Causes of Failure ........................................................ 430
Examine Patterns in Car Safety .................................................. 431

17 Creating Scatterplot Matrices
Using the Scatterplot Matrix Platform ....................................... 435
Example of a Scatterplot Matrix ................................................ 437
Launch the Scatterplot Matrix Platform ...................................... 437
Change the Matrix Format ....................................................... 439
The Scatterplot Matrix Report ................................................... 440
Scatterplot Matrix Options ...................................................... 441
Example Using a Grouping Variable ......................................... 442
Create a Grouping Variable .................................................... 444

18 Creating Ternary Plots
Using the Ternary Plot Platform ............................................... 445
Example of a Ternary Plot ........................................................ 447
Launch the Ternary Plot Platform .............................................. 449
The Ternary Plot Report .......................................................... 450
Mixtures and Constraints ....................................................... 450
Ternary Plot Platform Options ............................................... 451
Example of Using a Contour Function ....................................... 452

Index
Basic Analysis and Graphing .................................................. 465
Credits and Acknowledgments

Origin

JMP was developed by SAS Institute Inc., Cary, NC. JMP is not a part of the SAS System, though portions of JMP were adapted from routines in the SAS System, particularly for linear algebra and probability calculations. Version 1 of JMP went into production in October 1989.

Credits

JMP was conceived and started by John Sall. Design and development were done by John Sall, Chung-Wei Ng, Michael Hecht, Richard Potter, Brian Corcoran, Annie Dudley Zangi, Bradley Jones, Craig Hales, Chris Gotwalt, Paul Nelson, Xan Gregg, Jianfeng Ding, Eric Hill, John Schroedl, Laura Lancaster, Scott McQuiggan, Melinda Thielbar, Clay Barker, Peng Liu, Dave Barbour, Jeff Polzin, John Ponte, and Steve Amerige.

In the SAS Institute Technical Support division, Duane Hayes, Wendy Murphrey, Rosemary Lucas, Win LeDinh, Bobby Riggs, Glen Grimme, Sue Walsh, Mike Stockstill, Kathleen Kiernan, and Liz Edwards provide technical support.

Nicole Jones, Kyoko Keener, Hui Di, Joseph Morgan, Wenjun Bao, Fang Chen, Susan Shao, Yusuke Ono, Michael Crotty, Jong-Seok Lee, Tonya Mauldin, Audrey Ventura, Ani Eloyan, Bo Meng, and Sequola McNeill provide ongoing quality assurance. Additional testing and technical support are provided by Noriki Inoue, Kyoko Takenaka, and Masakazu Okada from SAS Japan.

Bob Hickey and Jim Borek are the release engineers.

The JMP books were written by Ann Lehman, Lee Creighton, John Sall, Bradley Jones, Erin Vang, Melanie Drake, Meredith Blackwelder, Diane Perhac, Jonathon Gatlin, Susan Conaghan, and Sheila Loring, with contributions from Annie Dudley Zangi and Brian Corcoran. Creative services and production was done by SAS Publications. Melanie Drake implemented the Help system.

Jon Weisz and Jeff Perkinson provided project management. Also thanks to Lou Valente, Ian Cox, Mark Bailey, and Malcolm Moore for technical advice.

Thanks also to Georges Guirguis, Warren Sarle, Gordon Johnston, Duane Hayes, Russell Wolfinger, Randall Tobias, Robert N. Rodriguez, Ying So, Warren Kuhfeld, George MacKensie, Bob Lucas, Warren Kuhfeld, Mike Leonard, and Padraic Neville for statistical R&D support. Thanks are also due to Doug Melzer, Bryan Wolfe, Vincent DelGobbo, Biff Beers, Russell Gonsalves, Mitchel Soltys, Dave Mackie, and Stephanie Smith, who helped us get started with SAS Foundation Services from JMP.

Acknowledgments

We owe special gratitude to the people that encouraged us to start JMP, to the alpha and beta testers of JMP, and to the reviewers of the documentation. In particular we thank Michael Benson, Howard
Yetter (d), Andy Mauromoustakos, Al Best, Stan Young, Robert Muenchen, Lenore Herzenberg, Ramon Leon, Tom Lange, Homer Hedges, Skip Weed, Michael Emptage, Pat Spagan, Paul Wenz, Mike Bowen, Lori Gates, Georgia Morgan, David Tanaka, Zoe Jewell, Sky Alibhai, David Coleman, Linda Blazek, Michael Friendly, Joe Hockman, Frank Shen, J.H. Goodman, David Iklé, Barry Hembree, Dan Obermiller, Jeff Sweeney, Lynn Vanatta, and Kris Ghosh.

Also, we thank Dick DeVeaux, Gray McQuarrie, Robert Stine, George F. R. Thomas, Avigdor Cahaner, José Ramirez, Gudmundur Axelsson, Al Fulmer, Cary Tuckfield, Ron Thisted, Nancy McDermott, Veronica Czitrom, Tom Johnson, Cy Wegman, Paul Dwyer, DaRon Huffaker, Kevin Norwood, Mike Thompson, Jack Reese, Francois Mainville, and John Wass.

We also thank the following individuals for expert advice in their statistical specialties: R. Hocking and P. Spector for advice on effective hypotheses; Robert Mee for screening design generators; Roselinde Kessels for advice on choice experiments; Greg Piepel, Peter Goos, J. Stuart Hunter, Dennis Lin, Doug Montgomery, and Chris Nachtsheim for advice on design of experiments; Jason Hsu for advice on multiple comparisons methods (not all of which we were able to incorporate in JMP); Ralph O’Brien for advice on homogeneity of variance tests; Ralph O’Brien and S. Paul Wright for advice on statistical power; Keith Muller for advice in multivariate methods, Harry Martz, Wayne Nelson, Ramon Leon, Dave Trindade, Paul Tobias, and William Q. Meeker for advice on reliability plots; Lijian Yang and J.S. Marron for bivariate smoothing design; George Milliken and Yurij Bulavski for development of mixed models; Will Potts and Cathy Maahs-Fladung for data mining; Clay Thompson for advice on contour plotting algorithms; and Tom Little, Damon Stoddard, Blanton Godfrey, Tim Clapp, and Joe Ficalora for advice in the area of Six Sigma; and Josef Schmee and Alan Bowman for advice on simulation and tolerance design.

For sample data, thanks to Patrice Strahle for Pareto examples, the Texas air control board for the pollution data, and David Coleman for the pollen (eureka) data.

Translations
Trish O’Grady coordinates localization. Special thanks to Noriki Inoue, Kyoko Takenaka, Masakazu Okada, Naohiro Masukawa and Yusuke Ono (SAS Japan); and Professor Toshiro Haga (retired, Tokyo University of Science) and Professor Hiroshiko Asano (Tokyo Metropolitan University) for reviewing our Japanese translation; Professors Fengshan Bai, Xuan Lu, and Jiaxiu Li at Tsinghua University in Beijing, and their assistants Rui Guo, Shan Jiang, Zhicheng Wan, and Qiang Zhao; and William Zhou (SAS China) and Zhongguo Zheng, professor at Peking University, for reviewing the Simplified Chinese translation; Jacques Goupy (consultant, ReConFor) and Olivier Núñez (professor, Universidad Carlos III de Madrid) for reviewing the French translation; Dr. Byung Chun Kim (professor, Korea Advanced Institute of Science and Technology) and Duk-Hyun Ko (SAS Korea) for reviewing the Korean translation; Bertram Schäfer and David Meintrup (consultants, StatCon) for reviewing the German translation; Patrizia Omodei, Maria Scaccabarozzi, and Letizia Bazzani (SAS Italy) for reviewing the Italian translation. Finally, thanks to all the members of our outstanding translation teams.

Past Support
Many people were important in the evolution of JMP. Special thanks to David DeLong, Mary Cole, Kristin Nauta, Aaron Walker, Ike Walker, Eric Gjertsen, Dave Tilley, Ruth Lee, Annette Sanders, Tim Christensen, Eric Wasserman, Charles Soper, Wenjie Bao, and Junji Kishimoto. Thanks to SAS Institute quality assurance by Jeanne Martin, Fouad Younan, and Frank Lassiter. Additional testing for Versions 3 and 4 was done by Li Yang, Brenda Sun, Katrina Hauser, and Andrea Ritter.
Also thanks to Jenny Kendall, John Hansen, Eddie Routten, David Schlotzhauer, and James Mulherin. Thanks to Steve Shack, Greg Weier, and Maura Stokes for testing JMP Version 1.

Thanks for support from Charles Shipp, Harold Gugel (d), Jim Winters, Matthew Lay, Tim Rey, Rubin Gabriel, Brian Ruff, William Lisowski, David Morganstein, Tom Esposito, Susan West, Chris Fehily, Dan Chilko, Jim Shook, Ken Bodner, Rick Blahunka, Dana C. Aultman, and William Fehlner.

**Technology License Notices**

Scintilla is Copyright 1998-2003 by Neil Hodgson <neilh@scintilla.org>. NEIL HODGSON DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL NEIL HODGSON BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.

XRender is Copyright © 2002 Keith Packard. KEITH PACKARD DISCLAIMS ALL WARRANTIES WITH REGARD TO THIS SOFTWARE, INCLUDING ALL IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS, IN NO EVENT SHALL KEITH PACKARD BE LIABLE FOR ANY SPECIAL, INDIRECT OR CONSEQUENTIAL DAMAGES OR ANY DAMAGES WHATSOEVER RESULTING FROM LOSS OF USE, DATA OR PROFITS, WHETHER IN AN ACTION OF CONTRACT, NEGLIGENCE OR OTHER TORTIOUS ACTION, ARISING OUT OF OR IN CONNECTION WITH THE USE OR PERFORMANCE OF THIS SOFTWARE.

ImageMagick software is Copyright © 1999-2010 ImageMagick Studio LLC, a non-profit organization dedicated to making software imaging solutions freely available.

bzlib software is Copyright © 1991-2009, Thomas G. Lane, Guido Vollbeding. All Rights Reserved.

FreeType software is Copyright © 1996-2002 The FreeType Project (www.freetype.org). All rights reserved.
Using JMP statistical software, you can interact with your graphs and data to do the following:

**Discover**  Using graphics, you can see patterns and relationships in your data, and find data that does not fit identifiable patterns.

**Interact**  Using JMP interactive features, you can follow up on clues and try different approaches. The more approaches you try, the more you likely you are to discover trends in your data.

**Understand**  Using graphics, you can see how the data and the model work together to produce the statistics. Because JMP is a progressively disclosed system, you learn statistics methods in the right context.

Here are just a few of the things you can do with JMP:

- Interact with data tables and reports.
- Compute values using the Formula Editor.
- Design experiments.
- Use scripting features to automate frequently used processes.
- Open SAS data sets, run stored processes, and submit SAS code.

**Figure 1.1 Interacting with JMP**

Clicking on a histogram bar highlights the corresponding data in the associated data table.
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prerequisites</td>
<td>3</td>
</tr>
<tr>
<td>JMP Terminology</td>
<td>3</td>
</tr>
<tr>
<td>Learning about JMP</td>
<td>3</td>
</tr>
<tr>
<td>About JMP Documentation</td>
<td>3</td>
</tr>
<tr>
<td>Use JMP Help</td>
<td>6</td>
</tr>
<tr>
<td>Use Tutorials</td>
<td>7</td>
</tr>
<tr>
<td>Access Sample Data Tables</td>
<td>7</td>
</tr>
<tr>
<td>Learn About Statistical and JSL Terms</td>
<td>7</td>
</tr>
<tr>
<td>Learn JMP Tips and Tricks</td>
<td>8</td>
</tr>
<tr>
<td>Access Resources on the Web</td>
<td>8</td>
</tr>
<tr>
<td>Conventions</td>
<td>8</td>
</tr>
<tr>
<td>Use JMP Platforms</td>
<td>8</td>
</tr>
<tr>
<td>Work with Multiple Data Tables and Platforms</td>
<td>8</td>
</tr>
<tr>
<td>How JMP Platforms Are Designed</td>
<td>9</td>
</tr>
<tr>
<td>Process for Analyzing Data Using Platforms</td>
<td>9</td>
</tr>
<tr>
<td>Common Features Throughout Platforms</td>
<td>13</td>
</tr>
<tr>
<td>Launch Window Features</td>
<td>13</td>
</tr>
<tr>
<td>Script Menus</td>
<td>15</td>
</tr>
<tr>
<td>Automatic Recalc Feature</td>
<td>17</td>
</tr>
</tbody>
</table>
Prerequisites

Before you begin using JMP, note the following information:

- You can use many JMP features, such as data manipulation, graphs, and scripting features, without any statistical knowledge.
- A basic understanding of central statistical concepts, such as mean and variation, is recommended.
- Analytical features require statistical knowledge appropriate for the feature.

JMP Terminology

- You can enter, view, edit, and manipulate data using data tables. In a data table, each variable is a column, and each observation is a row.
- You can access a platform from the Analyze and Graph menus. Platforms contain interactive windows that you can use to analyze data and work with graphs.
- Platforms use these windows:
  - Launch windows where you set up and run your analysis.
  - Report windows showing the output of your analysis.
- Report windows normally contain the following items:
  - A graph of some type (such as a scatterplot or a chart).
  - Specific reports that you can show or hide using the disclosure button.
  - Platform options that are located within red triangle menus.

For more about platforms, see “Use JMP Platforms,” p. 8.

Learning about JMP

JMP provides numerous resources to help you learn about the software. Most of them can be found within the Help menu. You can also access context-sensitive Help from within JMP.

Note: For further information about all of the options in the Help menu, see Using JMP.

About JMP Documentation

You can view the JMP documentation suite by selecting Help > Books.

Table 1.1 describes documents in the JMP documentation suite and the purpose of each document.
Table 1.1 About JMP Documentation

<table>
<thead>
<tr>
<th>Document</th>
<th>Who Should Read This Document</th>
<th>What this Document Covers</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Discovering JMP</em></td>
<td>If you are not familiar with JMP, start here.</td>
<td>Introduces you to JMP and gets you started using JMP</td>
</tr>
<tr>
<td><em>Using JMP</em></td>
<td>If you want to understand JMP data tables and how to perform basic operations in JMP, start here.</td>
<td>• General JMP concepts and features that span across all of JMP</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Material in these JMP Starter categories: File, Tables, and SAS</td>
</tr>
<tr>
<td><em>Basic Analysis and Graphing</em></td>
<td>If you want to perform basic analysis and graphing functions.</td>
<td>• These Analyze platforms:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Distribution</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Fit Y by X</td>
</tr>
<tr>
<td></td>
<td></td>
<td>– Matched Pairs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Many basic graphing platforms</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Material in these JMP Starter categories: Basic and Graph</td>
</tr>
</tbody>
</table>
Table 1.1 About JMP Documentation (Continued)

<table>
<thead>
<tr>
<th>Document</th>
<th>Who Should Read This Document</th>
<th>What this Document Covers</th>
</tr>
</thead>
</table>
| *Modeling and Multivariate Methods* | If you want to perform modeling or multivariate methods | • These Analyze platforms:  
  – Fit Model  
  – Screening  
  – Nonlinear  
  – Neural  
  – Gaussian Process  
  – Partition  
  – Time Series  
  – Categorical  
  – Choice  
  – Multivariate  
  – Cluster  
  – Principal Components  
  – Discriminant  
  – PLS (Partial Least Squares)  
  – Item Analysis  
  • These Graph platforms:  
  – Profilers  
  – Surface Plot  
  • Material in these JMP Starter categories: Model, Multivariate, and Surface |
### Table 1.1 About JMP Documentation (Continued)

<table>
<thead>
<tr>
<th>Document</th>
<th>Who Should Read This Document</th>
<th>What this Document Covers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quality and Reliability Methods</td>
<td>If you want to perform quality control or reliability engineering</td>
<td>• Life Distribution&lt;br&gt;• Fit Life by X&lt;br&gt;• Recurrence Analysis&lt;br&gt;• Degradation&lt;br&gt;• Survival&lt;br&gt;• Fit Parametric Survival&lt;br&gt;• Fit Proportional Hazards&lt;br&gt;• These Graph platforms:&lt;br&gt;  – Variability/Gauge Chart&lt;br&gt;  – Control Charts&lt;br&gt;  – Capability&lt;br&gt;  – Pareto Plot&lt;br&gt;  – Diagram (Ishikawa)&lt;br&gt;• Material in these JMP Starter window categories: Reliability, Measure, and Control</td>
</tr>
<tr>
<td>Design of Experiments</td>
<td>If you want to design experiments</td>
<td>• Everything related to the DOE menu&lt;br&gt;• Material in this JMP Starter window category: DOE</td>
</tr>
<tr>
<td>Scripting Guide</td>
<td>If you want to use the JMP Scripting Language (JSL)</td>
<td>A reference guide for using JSL commands</td>
</tr>
</tbody>
</table>


**Note:** The Books menu also contains two reference cards: The JMP Menu Card describes JMP menus, and the JMP Quick Reference Card describes JMP keyboard shortcuts. You can print these for ease of use.

### Use JMP Help

You can access JMP Help in two ways:

- Access the context-sensitive Help by selecting the Help tool 📚 from the Tools menu. Place the Help tool anywhere in a data table or report window to see the Help for that area.
- Within a window, click on the Help button.
Search and view JMP Help using the **Help > Contents**, **Search**, and **Index** options.

**Use Tutorials**

You can access JMP tutorials by selecting **Help > Tutorials**. The first item on the Tutorials menu is the **Tutorials Directory**. This opens a new window with all the tutorials grouped by category.

If you are not familiar with JMP, then start with the **Beginners Tutorial**. It steps you through the JMP interface and explains the basics of using JMP.

The rest of the tutorials help you with specific aspects of JMP, such as creating a pie chart, using Graph Builder, and so on.

**Access Sample Data Tables**

All of the examples in the JMP documentation suite use sample data. To access JMP's sample data tables, select **Help > Sample Data**. From here, you can do the following:

- Open the sample data directory.
- Open an alphabetized list of all sample data tables.
- Find a sample data table within a category.

Alternatively, the sample data tables are installed in the following directory:

- **On Windows**: `C:\Program Files\SAS\JMP\9\Support Files <language>\Sample Data`
- **On Macintosh**: `\Library\Application Support\JMP\9\<language>\Sample Data`

**Learn About Statistical and JSL Terms**

The **Help** menu contains the following indexes:

**Table 1.2 Descriptions of Help Menu Indexes**

<table>
<thead>
<tr>
<th>Index</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Statistics Index</td>
<td>Provides definitions of statistical terms.</td>
</tr>
<tr>
<td>JSL Functions</td>
<td>Provides definitions of JSL functions.</td>
</tr>
<tr>
<td>Object Scripting</td>
<td>Provides a list of JSL scriptable objects and the messages that can be sent to those objects.</td>
</tr>
<tr>
<td>DisplayBox Scripting</td>
<td>Provides a list of the JSL objects that comprise a JMP report.</td>
</tr>
</tbody>
</table>

For more details about these indexes, see *Using JMP*. 
Learn JMP Tips and Tricks

When you first start JMP, you see the Tip of the Day window.

To turn off the Tip of the Day, clear the Show tips at startup check box. To view it again, select Help > Tip of the Day. Or, you can turn it off using the Preferences window. See the Using JMP book.

You can use the JMP Quick Reference Card to learn more advanced commands in JMP. View this document by selecting Help > Books > JMP Quick Reference Card.

Access Resources on the Web

To access JMP resources on the Web, select Help > JMP.com or Help > JMP User Community.

The JMP.com option takes you to the JMP Web site, and the JMP User Community option takes you to JMP online user forums.

Conventions

The following conventions help you relate written material to information that you see on your screen.

- Most open data table names that are used in examples appear in **Helvetica** font (Animals or Animals.jmp) in this document. References to variable names in data tables and items in reports also appear in **Helvetica** according to the way they appear on the screen or in the documentation.

- **Note:** Special information, warnings, and limitations are noted as such in **boldface**.

- Reference to menu names (File menu) or menu items (Save option) appear in **Helvetica bold**.

- Words or phrases that are important or have definitions specific to JMP are in **italics** the first time they occur in the text. For example, the word platform is in italics the first time you see it. Most words in italics are defined when they are first used unless clear in the context of use.

Use JMP Platforms

JMP uses many statistical methods that are organized and consolidated into the platforms within the Analyze and Graph menus.

Work with Multiple Data Tables and Platforms

As mentioned earlier, platforms use interactive windows that help you analyze data and work with graphs. You can have any number of data tables open at the same time and any number of platforms open for each data table. If you have several data tables open, then the platform that you launch analyzes the current data table. The current data table is either the active data table window or the table corresponding to the current report window.
How JMP Platforms Are Designed

Before you use JMP’s statistical platforms, note the following design aspects of JMP:

- JMP implements general methods that are consistent with its key concepts. Although platforms produce a variety of results that handle many situations, all platforms are consistent in their treatment of data and statistical concepts.
- JMP methods are generic. They adapt themselves to the context at hand. The principal concept that drives analyses is modeling type. For details, see “Assign Modeling Types,” p. 9. For example, JMP automatically analyzes a variable with nominal values differently than it analyzes a variable with continuous values.
- JMP platforms give you almost everything that you need at once. If this is more information than you want, you can conceal the reports or graphs that you do not need. The advantage is that there is less need to search for statistical commands.

Process for Analyzing Data Using Platforms

To begin analyzing your data, proceed as follows:

1. With your data table open, assign or change modeling types, as needed. See “Assign Modeling Types,” p. 9.
2. Choose an analysis and launch the corresponding platform. See “Choose an Analysis and Launching a Platform,” p. 11.

Assign Modeling Types

Note: For more details about working with modeling types, see Using JMP.

The modeling type of a variable can be one of the following types, shown with its corresponding icon:

- Continuous
- Ordinal
- Nominal

When you import data into JMP, it predicts which modeling types to use. Character data is considered nominal, and numeric data is considered continuous. However, you might want to change the modeling type, depending on which type of analysis you are performing.

To change the modeling type, click on the modeling type icon next to the variable and make your selection.
Figure 1.2 Changing Modeling Type

Table 1.3 Descriptions of Modeling Types

<table>
<thead>
<tr>
<th>Continuous</th>
<th>Columns can contain only numeric data types. Continuous values are treated as continuous measurement values. JMP uses the numeric values directly in computations.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ordinal</td>
<td>Columns can contain either numeric or character data types. JMP analyses treat ordinal values as discrete categorical values that have an order. If the values are numbers, the order is the numeric magnitude. If the values are character, the order is the sorting sequence.</td>
</tr>
<tr>
<td>Nominal</td>
<td>Columns can contain either numeric or character data types. All values are treated in JMP analyses as if they are discrete values with no implicit order.</td>
</tr>
</tbody>
</table>
Chapter 1

Choose an Analysis and Launching a Platform

Choose the analysis that you want to perform. You can launch a platform from the Analyze or Graph menu (or from the JMP Starter window).

Table 1.4 provides just a few examples of choosing an analysis and launching the corresponding platform. For a full list of the platforms available within the Analyze and Graph menus, see Using JMP.

Table 1.4 Examples of Choosing an Analysis and Launching a Platform

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analyze a distribution of values</td>
<td>Analyze &gt; Distribution</td>
</tr>
<tr>
<td>Analyze the relationship between two variables</td>
<td>Analyze &gt; Fit Y by X</td>
</tr>
<tr>
<td>Graph continuous X and Y variables</td>
<td>Graph &gt; Overlay Plot</td>
</tr>
<tr>
<td>Animate points on a scatterplot over time</td>
<td>Graph &gt; Bubble Plot</td>
</tr>
</tbody>
</table>

Complete a Launch Window

After you launch a platform, the launch window appears. See Figure 1.3. Use the launch window to set up your analysis by moving columns into roles.

Figure 1.3 The Launch Window (Distribution Platform)

For more details about the launch window, see “Launch Window Features,” p. 13.

View a Report Window

Once you have completed the launch window (which runs the analysis) you see a report window. For example, see Figure 1.4. You can use the output in the report window to interpret your data.
Figure 1.4 Example of a Report Window (Distribution Platform)

- (Windows only) The JMP main menu bar is hidden by default. To see the JMP main menu bar, hover over the bar below the window title. Alternatively, you can always show the main menu bar by changing the Auto-hide menu and toolbars option in File > Preferences > Windows Specific.

For more general information about using report windows, see Using JMP.

- For detailed information about a specific platform’s report window, see the chapter for that platform in this book, or in one of these books:
  - Modeling and Multivariate Methods
  - Quality and Reliability Methods
Common Features Throughout Platforms

All platforms share certain specific features, documented in these sections:

- “Launch Window Features,” p. 13
- “Script Menus,” p. 15
- “Automatic Recalc Feature,” p. 17

Note: Platforms have additional shared features that are not covered here, such as linked graphs, red triangle menus, and hierarchical report organization. See Using JMP.

Launch Window Features

Each analysis platform prompts you with a launch window. Table 1.5 describes three panes that all launch windows have in common.

Table 1.5 Descriptions of Common Panes in Launch Windows

<table>
<thead>
<tr>
<th>Select Columns</th>
<th>Lists all of the variables in your current data table. For details about the red triangle menu, see “Columns Filter Menu,” p. 14.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cast Selected Columns into Roles</td>
<td>Moves selected columns into roles (such as Y, X, and so on.) You cast a column into the role of a variable (like an actor is cast into a role). See “Cast Selected Columns into Roles Buttons,” p. 13. This pane does not exist in the Graph Builder platform.</td>
</tr>
</tbody>
</table>
| Action | **OK** performs the analysis.  
**Cancel** stops the analysis and quits the launch window.  
**Remove** deletes any selected variables from a role.  
**Recall** populates the launch window with the last analysis that you performed.  
**Help** takes you to the Help for the launch window. |

Cast Selected Columns into Roles Buttons

Table 1.6 describes buttons that appear frequently throughout launch windows. Buttons that are specific to certain platforms are described in the chapter for the platform.
Table 1.6 Descriptions of Common Buttons in Launch Windows

<table>
<thead>
<tr>
<th>Button</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>Identifies a column as a response or dependent variable whose distribution is to be studied.</td>
</tr>
<tr>
<td>X</td>
<td>Identifies a column as an independent, classification, or explanatory variable whose values divide the rows into sample groups.</td>
</tr>
<tr>
<td>Weight</td>
<td>Identifies the data table column whose variables assign weight (such as importance or influence) to the data.</td>
</tr>
<tr>
<td>Freq</td>
<td>Identifies the data table column whose values assign a frequency to each row. This option is useful when a frequency is assigned to each row in summarized data.</td>
</tr>
<tr>
<td>By</td>
<td>Identifies a column that creates a report consisting of separate analyses for each level of the variable.</td>
</tr>
</tbody>
</table>

Columns Filter Menu

In most of the platform launch windows there is a Column Filter menu. This menu appears as a red triangle within the Select Columns panel.

Figure 1.5 Example of the Columns Filter Menu

Table 1.7 Options within the Columns Filter Menu

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reset</td>
<td>Resets the columns to its original list.</td>
</tr>
<tr>
<td>Sort by Name</td>
<td>Sorts the columns in alphabetical order by name.</td>
</tr>
<tr>
<td>Continuous</td>
<td>Shows or hides columns whose modeling type is continuous.</td>
</tr>
<tr>
<td>Ordinal</td>
<td>Shows or hides columns whose modeling type is ordinal.</td>
</tr>
<tr>
<td>Nominal</td>
<td>Shows or hides columns whose modeling type is nominal.</td>
</tr>
</tbody>
</table>
Chapter 1

Preliminaries

Use JMP Platforms

Table 1.7 Options within the Columns Filter Menu

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numeric</td>
<td>Shows or hides columns whose data type is numeric.</td>
</tr>
<tr>
<td>Character</td>
<td>Shows or hides columns whose data type is character.</td>
</tr>
<tr>
<td>Match case</td>
<td>(Only applicable to the Name options below) Makes your search case-sensitive.</td>
</tr>
<tr>
<td>Name Contains</td>
<td>Searches for column names containing specified text. To remove the text box, select Reset.</td>
</tr>
<tr>
<td>Name Starts With</td>
<td>Searches for column names that begin with specified text. To remove the text box, select Reset.</td>
</tr>
<tr>
<td>Name Ends With</td>
<td>Searches for column names that end with specified text. To remove the text box, select Reset.</td>
</tr>
</tbody>
</table>

Script Menus

The red triangle menu at the top level of every JMP report contains a Script menu.

Figure 1.6 The Script menu

Most of these options are the same throughout JMP. A few platforms add extra options that are described in the specific platform chapters. Table 1.8 lists the Script menu options that are common to all platforms.

Table 1.8 Descriptions of Script Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Redo Analysis</td>
<td>If the values in the data table that was used to produce the report have changed, this option duplicates the analysis based on the new data. The new analysis appears in a new report window.</td>
</tr>
</tbody>
</table>
If you have specified a By variable in the platform launch window, the Script All By-Groups menu also appears.

**Figure 1.7 Example of the Script All By-Groups Menu**
If you specified a By variable, the script options in Table 1.8 apply to a report for a single level of a By variable. The Script All By-Groups options apply to the reports for all the levels of the By variable.

**Table 1.9** Descriptions of Script All By-Groups Options

<table>
<thead>
<tr>
<th>Redo Analysis</th>
<th>If the values in the data table that was used to produce the reports have changed, this option duplicates the analysis based on the new data and produces new reports.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relaunch Analysis</td>
<td>Opens the platform Launch window and recalls the settings used to create the reports.</td>
</tr>
<tr>
<td>Copy Script</td>
<td>Places the script that reproduces the reports on the clipboard so that it can be pasted elsewhere.</td>
</tr>
<tr>
<td>Save Script to Data Table</td>
<td>Saves the script to the data table that was used to produce the reports.</td>
</tr>
<tr>
<td>Save Script to Journal</td>
<td>Saves a button that runs the script in a journal. The script is added to the current journal.</td>
</tr>
<tr>
<td>Save Script to Script Window</td>
<td>Opens a script editor window and adds the script to it. If you have already saved a script to a script window, additional scripts are added to the bottom of the same script window.</td>
</tr>
</tbody>
</table>

**Automatic Recalc Feature**

The Automatic Recalc feature immediately reflects changes that you make to the data table in the corresponding report window. You can make any of the following data table changes:

- exclude or unexclude data table rows
- delete or add data table rows

This powerful feature immediately reflects these changes to the corresponding analyses, statistics, and graphs that are located in a report window.

To turn on **Automatic Recalc** for a report window, click on the platform red triangle menu and select **Script > Automatic Recalc**. See Figure 1.6. To turn it off, deselect the same option. You can also turn on **Automatic Recalc** using JSL.

Note the following:

- By default, **Automatic Recalc** is turned off for platforms in the **Analyze** menu and turned on for platforms in the **Graph** menu. The exceptions are the **Capability, Variability/Gauge Chart**, and **Control Chart > Run Chart** platforms.
- For some platforms, the **Automatic Recalc** feature is not appropriate, and therefore is not supported. These platforms include the following: DOE, Profilers, Choice, Partition, Nonlinear, Neural, Neural Net, Partial Least Squares, Fit Model (REML, GLM, Log Variance), Gaussian Process, Item Analysis, Cox Proportional Hazard, and Control Charts (except Run Chart).
Preliminaries
Use JMP Platforms
Performing Univariate Analysis
Using the Distribution Platform

The Distribution platform describes the distribution of variables using histograms, additional graphs, and reports. You can examine the distribution of several variables at once. The report content for each variable varies, depending on whether the variable is categorical (nominal or ordinal) or continuous.

The Distribution report window is interactive. Clicking on a histogram bar highlights the corresponding data in any other histograms and in the data table. See Figure 2.1.

**Figure 2.1 Example of the Distribution Platform**
Contents

Overview of the Distribution Platform ................................................. 21
Example of the Distribution Platform ................................................. 21
Launch the Distribution Platform ....................................................... 23
The Distribution Report ................................................................. 24
Histograms ..................................................................................... 26
Initial Reports .................................................................................. 29
  The Frequencies Report ................................................................ 29
  The Quantiles Report .................................................................. 31
  The Moments Report ................................................................. 32
Distribution Platform Options ........................................................ 34
Options for Categorical Variables ..................................................... 34
  Mosaic Plot .................................................................................. 35
  Test Probabilities ....................................................................... 36
  Confidence Intervals for Categorical Variables .............................. 39
Options for Continuous Variables ..................................................... 39
  Normal Quantile Plot .................................................................. 42
  Outlier Box Plot ......................................................................... 43
  Quantile Box Plot ...................................................................... 45
  Stem and Leaf ............................................................................. 46
  CDF Plot ..................................................................................... 47
  Test Mean .................................................................................... 48
  Test Std Dev .............................................................................. 50
  Confidence Intervals for Continuous Variables .............................. 51
Prediction Intervals .......................................................................... 54
Tolerance Intervals .......................................................................... 55
Capability Analysis .......................................................................... 57
Fit Distributions .............................................................................. 61
  Example of Fitting a Lognormal Distribution .............................. 61
Statistical Details ............................................................................. 77
Overview of the Distribution Platform

The treatment of variables in the Distribution platform is different, depending on the modeling type of variable, which can be categorical (nominal or ordinal) or continuous.

Categorical Variables

For categorical variables, the initial graph that appears is a histogram. The histogram shows a bar for each level of the ordinal or nominal variable. You can also add a divided (mosaic) bar chart.

The reports show counts and proportions. You can add confidence intervals and test the probabilities.

Continuous Variables

For numeric continuous variables, the initial graphs show a histogram and an outlier box plot. The histogram shows a bar for grouped values of the continuous variable. The following options are also available:
- quantile box plot
- normal quantile plot
- stem and leaf plot
- CDF plot
- overlaid nonparametric smoothing curve

The reports show selected quantiles and moments. Report options are available for the following:
- saving ranks, probability scores, normal quantile values, and so on, as new columns in the data table
- testing the mean and standard deviation of the column against a constant you specify
- fitting various distributions
- performing a capability analysis for a quality control application
- confidence intervals, prediction intervals, and tolerance intervals

Example of the Distribution Platform

Suppose that you have data on 40 students, and you want to see the distribution of age and height among the students.

1. Open the Big Class.jmp sample data table.
2. Select Analyze > Distribution.
4. Click OK.
From the histograms, you notice the following:

- The ages are not uniformly distributed.
- For height, there are two points with extreme values (that might be outliers).

Click on the bar for 50 in the height histogram to take a closer look at the potential outliers.

- The corresponding ages are highlighted in the age histogram. The potential outliers are age 12.
- The corresponding rows are highlighted in the data table. The names of the potential outliers are Lillie and Robert.

Add labels to the potential outliers in the height histogram.

1. Select both outliers.
2. Right-click on one of the outliers and select Row Label.
   Label icons are added to these rows in the data table.
3. (Optional) Resize the box plot wider to see the full labels.
Launch the Distribution Platform

Launch the Distribution platform by selecting Analyze > Distribution.

Figure 2.4 The Distribution Launch Window

Table 2.1 Description of the Distribution Launch Window

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Y, Columns</strong></td>
<td>Assigns the variables that you want to analyze. A histogram and associated reports appear for each variable.</td>
</tr>
<tr>
<td><strong>Weight</strong></td>
<td>Assigns a variable to give the observations different weights. Any moment that is based on the Sum Wgts is affected by weights (including the Mean, Std Err Mean, and Upper and Lower 95% Mean).</td>
</tr>
</tbody>
</table>
Performing Univariate Analysis

Chapter 2

The Distribution Report

Table 2.1 Description of the Distribution Launch Window (Continued)

<table>
<thead>
<tr>
<th>Role</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freq</td>
<td>Assigns a frequency variable to this role. This is useful if you have summarized data. In this instance, you have one column for the Y values and another column for the frequency of occurrence of the Y values. The sum of this variable is included in the overall count appearing in the Moments report (represented by N). All other moment statistics (mean, standard deviation, and so on) are also affected by the Freq variable.</td>
</tr>
<tr>
<td>By</td>
<td>Produces a separate report for each level of the By variable. If more than one By variables is assigned, a separate report is produced for each possible combination of the levels of the By variables.</td>
</tr>
</tbody>
</table>
| Histograms Only | Removes everything except the histograms from the report window, which includes the following:  
  - Quantiles and Moments reports for continuous variables  
  - Frequencies report for nominal and ordinal variables  
  - Outlier box plot |

The Distribution Report

Follow the instructions in “Example of the Distribution Platform,” p. 21 to produce the report shown in Figure 2.5.

The Distribution report window is divided into variables. For each variable, the initial Distribution report contains a histogram and reports.
Performing Univariate Analysis

The Distribution Report

Figure 2.5 The Initial Distribution Report Window

Table 2.2 Descriptions of Distribution Report Window Objects

<table>
<thead>
<tr>
<th>Platform options</th>
<th>The red triangle menu next to Distributions contains options that affect all of the variables. See “Distribution Platform Options,” p. 34.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable options</td>
<td>The red triangle menu next to each variable contains options that affect only that variable. See “Options for Categorical Variables,” p. 34 or “Options for Continuous Variables,” p. 39.</td>
</tr>
<tr>
<td><strong>Note:</strong> If you hold down the CTRL key and select a variable option, the option applies to all of the variables that have the same modeling type.</td>
<td></td>
</tr>
</tbody>
</table>
**Performing Univariate Analysis**

**Histograms**

Table 2.2 Descriptions of Distribution Report Window Objects (Continued)

| Reports | Each variable contains at least initial reports. Continuous variables contain a Quantiles and a Moments report; categorical variables contain a Frequencies report. See “Initial Reports,” p. 29. Variable options can create additional reports. |

**Histograms**

Histograms visually display your data. For categorical (nominal or ordinal) variables, the histogram shows a bar for each level of the ordinal or nominal variable. For continuous variables, the histogram shows a bar for grouped values of the continuous variable.

Table 2.3 Interacting with the Histogram

| Highlighting data | Click on a histogram bar or an outlying point in the graph. The corresponding rows are highlighted in the data table, and corresponding sections of other histograms are also highlighted, if applicable. See “Highlight Bars and Select Rows,” p. 27. |
| Creating a subset | Double-click on a histogram bar, or right-click on a histogram bar and select **Subset**. A new data table is created that contains only the selected data. |
| Resizing the entire histogram | Hover over the histogram borders until you see a double-sided arrow. Then click and drag the borders. For more details, see the *Using JMP* book. |
| Rescaling the axis | (Continuous variables only) Click and drag on an axis to rescale it. Alternatively, hover over the axis until you see a hand. Then, double-click on the axis and set the parameters in the Axis Specification window. |
| Resizing histogram bars | (Continuous variables only) There are multiple options to resize histogram bars. See “Resize Histogram Bars for Continuous Variables,” p. 27. |
| Specifying your selection | Specify the data that you select in multiple histograms. See “Specify Your Selection in Multiple Histograms,” p. 28. |

To see additional options for the histogram or the associated data table:

- Right-click on a histogram. See the *Using JMP* book.
- Click on the red triangle next to the variable, and select **Histogram Options**. Options are slightly different depending on the variable modeling type. See “Histogram Options for Categorical Variables,” p. 35 or “Histogram Options for Continuous Variables,” p. 40.
Resize Histogram Bars for Continuous Variables

Resize histogram bars for continuous variables by using the following:

- the Grabber (hand) tool
- the Set Bin Width option
- the Increment option

**Use the Grabber Tool**

The Grabber tool is a quick way to explore your data.

1. Select **Tools > Grabber**.

**Note:** (Windows only) To see the menu bar, you might need to hover over the bar below the window title. You can also change this setting in **File > Preferences > Windows Specific**.

2. Place the grabber tool anywhere in the histogram.
3. Click and drag the histogram bars.

Think of each bar as a bin that holds a number of observations:

- Moving the hand to the left increases the bin width and combines intervals. The number of bars decreases as the bar size increases.
- Moving the hand to the right decreases the bin width, producing more bars.
- Moving the hand up or down shifts the bin locations on the axis, which changes the contents and size of each bin.

**Use the Set Bin Width Option**

The Set Bin Width option is a more precise way to set the width for all bars in a histogram. To use the Set Bin Width option, from the red triangle menu for the variable, select **Histogram Options > Set Bin Width**.

Change the bin width value.

**Use the Increment Option**

The Increment option is another precise way to set the bar width. To use the Increment option, double-click on the axis, and change the Increment value.

**Highlight Bars and Select Rows**

Clicking on a histogram bar highlights the bar and selects the corresponding rows in the data table. The appropriate portions of all other graphical displays also highlight the selection. Figure 2.6 shows the results of highlighting a bar in the height histogram. The corresponding rows are selected in the data table.

**Note:** To deselect histogram bars, press the CTRL key and click on the highlighted bars.
Specify Your Selection in Multiple Histograms

Extend or narrow your selection in histograms as follows:

- To extend your selection, hold down the SHIFT key and select another bar. This is the equivalent of using an or operator.
- To narrow your selection, hold down the ALT key and select another bar. This is the equivalent of using an and operator.

Example of Selecting Data in Multiple Histograms

1. Open the Companies.jmp sample data table.
2. Select Analyze > Distribution.
3. Select Type and Size Co and click Y, Columns.
4. Click OK.
   
   You want to see the type distribution of companies that are small.
5. Click on the bar next to small.
   
   You can see that there are more small computer companies than there are pharmaceutical companies. To broaden your selection, add medium companies.
6. Hold down the SHIFT key. In the Size Co histogram, click on the bar next to medium.
   
   You can see the type distribution of small and medium sized companies. See Figure 2.7 at left. To narrow down your selection, you want to see the small and medium pharmaceutical companies only.
7. Hold down the ALT key. In the Type histogram, click in the Pharmaceutical bar.
   
   You can see how many of the small and medium companies are pharmaceutical companies. See Figure 2.7 at right.
Initial Reports

When you complete the Distribution launch window, the following reports appear automatically (unless other preferences are in effect):

- “The Frequencies Report,” p. 29
- “The Quantiles Report,” p. 31
- “The Moments Report,” p. 32

The Frequencies Report

For nominal and ordinal variables, the Frequencies report lists the levels of the variables, along with the associated frequency of occurrence and probabilities.

Example of a Frequencies Report

1. Open the VA Lung Cancer.jmp sample data table.
2. Select Analyze > Distribution.
3. Select Cell Type and click Y, Columns.
4. Click OK.
Performing Univariate Analysis

Chapter 2

Initial Reports

Figure 2.8 Example of a Frequencies Report

For each level of a categorical (nominal or ordinal) variable, the Frequencies report contains the information described in Table 2.4. Missing values are omitted from the analysis.

Table 2.4 Description of the Frequencies Report

<table>
<thead>
<tr>
<th>Level</th>
<th>Lists each value found for a response variable.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>Lists the number of rows found for each level of a response variable. If you use a Freq variable (see Figure 2.4), the Count is the sum of the Freq variables for each level of the response variable.</td>
</tr>
<tr>
<td>Prob</td>
<td>Lists the probability (or proportion) of occurrence for each level of a response variable. The probability is computed as the count divided by the total frequency of the variable, shown at the bottom of the table.</td>
</tr>
<tr>
<td>StdErr Prob</td>
<td>Lists the standard error of the probabilities. This column might be hidden. To show the column, right-click in the table and select Columns &gt; StdErr Prob.</td>
</tr>
<tr>
<td>Cum Prob</td>
<td>Contains the cumulative sum of the column of probabilities. This column might be hidden. To show the column, right-click in the table and select Columns &gt; Cum Prob.</td>
</tr>
</tbody>
</table>
The Quantiles Report

For continuous variables, the Quantiles report lists the values of selected quantiles (sometimes called percentiles).

Example of a Quantiles Report

1. Open the Cities.jmp sample data table.
2. Select Analyze > Distribution.
4. Click OK.

Figure 2.9 Example of a Quantiles Report

For details about how quantiles are computed, see “Statistical Details for Quantiles,” p. 77.
The Moments Report

For continuous variables, the Moments report displays the mean, standard deviation, and other summary statistics.

Example of a Moments Report

1. Open the Cities.jmp sample data table.
2. Select Analyze > Distribution.
4. Click OK.

Table 2.5 Description of the Moments Report

<table>
<thead>
<tr>
<th>Description</th>
<th>Calculation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Estimates the expected value of the underlying distribution for the response variable, which is the arithmetic average of the column’s values. It is the sum of the nonmissing values divided by the number of nonmissing values. If you assigned a Weight or Freq variable, the mean is computed by JMP as follows:</td>
</tr>
<tr>
<td></td>
<td>1. Each column value is multiplied by its corresponding weight or freq</td>
</tr>
<tr>
<td></td>
<td>2. These values are added and divided by the sum of the weights or freqs.</td>
</tr>
</tbody>
</table>
Distribution Platform Options

The red triangle menu next to Distributions contains options that affect all of the reports and graphs in the Distribution platform.

Table 2.6  Descriptions of Distribution Platform Options

<table>
<thead>
<tr>
<th>Uniform Scaling</th>
<th>Scales all axes with the same minimum, maximum, and intervals so that the distributions can be easily compared.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stack</td>
<td>Changes the orientation of the histogram and the reports to horizontal and stacks the individual distribution reports vertically. Deselect this option to return the report window to its original layout.</td>
</tr>
<tr>
<td>Save for Adobe Flash platform (.SWF)</td>
<td>Saves the histograms as .SWF files that are Adobe Flash player compatible. Use these files in presentations and in web pages. An HTML page is also saved that shows you the correct code for using the resulting .SWF file. For more information about this option, go to <a href="http://www.jmp.com/support/swfhelp/en">http://www.jmp.com/support/swfhelp/en</a>.</td>
</tr>
<tr>
<td>Script</td>
<td>This menu contains options that are available to all platforms. They enable you to redo the analysis or save the JSL commands for the analysis to a window or a file. See “Script Menus,” p. 15 in the “Preliminaries” chapter.</td>
</tr>
</tbody>
</table>

Options for Categorical Variables

The red triangle menus next to each variable in the report window contain additional options that apply to the variable. This section describes the options that are available for categorical (nominal or ordinal) variables.

Note: To see the options that are available for continuous variables, see “Options for Continuous Variables,” p. 39.

Display Options for Categorical Variables

Use the Display Options for categorical variables to show or hide the Frequency report and change the orientation of the report.

Table 2.7  Descriptions of Display Options for Categorical Variables

<table>
<thead>
<tr>
<th>Frequencies</th>
<th>Shows or hides the Frequencies report. See “The Frequencies Report,” p. 29.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal Layout</td>
<td>Changes the orientation of the histogram and the reports to vertical or horizontal.</td>
</tr>
</tbody>
</table>
Histogram Options for Categorical Variables

Use the Histogram Options for categorical variables to change various aspects of the histogram, such as orientation, color, size, and so on.

Table 2.8 Descriptions of Histogram Options for Categorical Variables

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vertical</td>
<td>Changes the orientation of the histogram from a vertical to a horizontal orientation.</td>
</tr>
<tr>
<td>Std Error Bars</td>
<td>Draws the standard error bar on each level of the histogram using the standard error $\sqrt{np(1-p)}$ where $p=ni/n$.</td>
</tr>
<tr>
<td>Separate Bars</td>
<td>Separates the histogram bars.</td>
</tr>
<tr>
<td>Histogram Color</td>
<td>Changes the color of the histogram bars.</td>
</tr>
<tr>
<td>Count Axis</td>
<td>Adds an axis that shows the frequency of column values represented by the histogram bars. <strong>Note:</strong> If you resize the histogram bars, the count axis also resizes.</td>
</tr>
<tr>
<td>Prob Axis</td>
<td>Adds an axis that shows the proportion of column values represented by histogram bars. <strong>Note:</strong> If you resize the histogram bars, the probability axis also resizes.</td>
</tr>
<tr>
<td>Density Axis</td>
<td>The density is the length of the bars in the histogram. Both the count and probability are based on the following calculations: $\text{prob} = (\text{bar width}) \times \text{density}$ $\text{count} = (\text{bar width}) \times \text{density} \times (\text{total count})$ <strong>Note:</strong> If you resize the histogram bars, the density axis remains constant.</td>
</tr>
<tr>
<td>Show Percents</td>
<td>Labels the percent of column values represented by each histogram bar.</td>
</tr>
<tr>
<td>Show Counts</td>
<td>Labels the frequency of column values represented by each histogram bar.</td>
</tr>
</tbody>
</table>

Mosaic Plot

The Mosaic Plot option displays a mosaic bar chart for each nominal or ordinal response variable. A mosaic plot is a stacked bar chart where each segment is proportional to its group's frequency count.
Example of a Mosaic Plot

1. Open the VA Lung Cancer.jmp sample data table.
2. Select Analyze > Distribution.
3. Select Cell Type and click Y, Columns.
4. Click OK.
5. From the red triangle menu for Cell Type, select Mosaic Plot.

For a description of the options that appear when you right-click the mosaic plot, see the Using JMP book.

Test Probabilities

Use the Test Probabilities option to enter hypothesized probabilities. The Test Probabilities report contains slightly different options, depending on whether your variable has more than two levels, or exactly two levels. See Figure 2.12.

Examples of the Test Probabilities Option

Initiate a test probability report for a variable with more than two levels:

1. Open the VA Lung Cancer.jmp sample data table.
2. Select Analyze > Distribution.
3. Select Cell Type and click Y, Columns.
4. Click OK.
5. From the red triangle menu next to Cell Type, select Test Probabilities.

See Figure 2.12 at left.

Initiate a test probability report for a variable with exactly two levels:
1. Open the Penicillin.jmp sample data table.
2. Select Analyze > Distribution.
4. Click OK.
5. From the red triangle menu next to Response, select Test Probabilities.

See Figure 2.12 at right.

---

**Figure 2.12 Examples of Test Probabilities Options**

The Test Probabilities option scales the hypothesized values that you type, so that the probabilities sum to one. The easiest way to test that all of the probabilities are equal is to type a one in each field. If you want to test a subset of the probabilities, then do not enter a value for any levels that are not involved. JMP substitutes estimated probabilities.
Performing Univariate Analysis
Options for Categorical Variables

Generate the Test Probabilities Report

To generate a test probabilities report for a variable with more than two levels:
1. Refer to Figure 2.12 at left. Type 0.25 in all four Hypoth Prob fields.
2. Click the **Fix hypothesized values, rescale omitted** button.
3. Click **Done**.

Likelihood Ratio and Pearson Chi-square tests are calculated. See Figure 2.13 at left.

To generate a test probabilities report for a variable with exactly two levels:
1. Refer to Figure 2.12 at right. Type 0.5 in both Hypoth Prob fields.
2. Click the **probability less than hypothesized value** button.
3. Click **Done**.

Exact probabilities are calculated for the binomial test. See Figure 2.13 at right.

---

**Figure 2.13** Examples of Test Probabilities Reports

![Test Probabilities Report Example](image-url)
Confidence Intervals for Categorical Variables

Confidence Interval options compute score confidence intervals. The score confidence interval is recommended because it tends to have better coverage probabilities (1 - \( \alpha \)), especially with smaller sample sizes (Agresti and Coull 1998).

Save Commands for Categorical Variables

Use the Save options for categorical variables to add level numbers to the data table, and use the Script to log command.

Table 2.9 Descriptions of Save Options for Categorical Variables

<table>
<thead>
<tr>
<th>Options</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level Numbers</td>
<td>Creates a new column in the data table called Level &lt;colname&gt;. The level number of each observation corresponds to the histogram bar that contains the observation. The histogram bars are numbered from low to high (or alphabetically) beginning with 1.</td>
</tr>
<tr>
<td>Script to log</td>
<td>In the log window, this option displays the script commands to generate the current report. To see this option, you must have the log window open. Select View &gt; Log to see the log window.</td>
</tr>
</tbody>
</table>

Options for Continuous Variables

The red triangle menus next to each variable in the report window contain additional options that apply to the variable. This section describes the options that are available for continuous variables.

Advanced options for continuous variables are covered in the following sections:

- “Prediction Intervals,” p. 54
- “Tolerance Intervals,” p. 55
- “Capability Analysis,” p. 57
- “Fit Distributions,” p. 61

Note: To see the options that are available for categorical (nominal and ordinal) variables, see “Options for Categorical Variables,” p. 34.

Display Options for Continuous Variables

Use the Display Options for continuous variables to show, hide, or add reports, or to change the orientation of the report.
Performing Univariate Analysis

Chapter 2

Options for Continuous Variables

Table 2.10 Descriptions of Display Options for Continuous Variables

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantiles</td>
<td>Shows or hides the Quantiles report. See “The Quantiles Report,” p. 31.</td>
</tr>
<tr>
<td>Moments</td>
<td>Shows or hides the Moments report. See “The Moments Report,” p. 32.</td>
</tr>
<tr>
<td>More Moments</td>
<td>An extension of the Moments report that adds the following statistics:</td>
</tr>
<tr>
<td></td>
<td>* <strong>Sum Wgt</strong> is the sum of a column assigned to the role of Weight (in the launch window). Sum Wgt is used in the denominator for computations of the mean instead of N.</td>
</tr>
<tr>
<td></td>
<td>* <strong>Sum</strong> is the sum of the response values.</td>
</tr>
<tr>
<td></td>
<td>* <strong>Variance</strong> is the sample variance, and the square of the sample standard deviation.</td>
</tr>
<tr>
<td></td>
<td>* <strong>Skewness</strong> measures sidedness or symmetry. It is based on the third moment about the mean and is computed as follows:</td>
</tr>
<tr>
<td></td>
<td>[ \sum \frac{3}{2} \frac{w_i z_i^3}{(N-1)(N-2)} ] where ( z_i = \frac{x_i - \bar{x}}{s} )</td>
</tr>
<tr>
<td></td>
<td>and ( w_i ) is a weight term (= 1 for equally weighted items)</td>
</tr>
<tr>
<td></td>
<td>* <strong>Kurtosis</strong> measures peakedness or heaviness of tails. It is based on the fourth moment about the mean and is computed as follows:</td>
</tr>
<tr>
<td></td>
<td>[ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_{i=1}^{n} w_i \left( \frac{x_i - \bar{x}}{s} \right)^4 - \frac{3(n-1)^2}{(n-2)(n-3)} ]</td>
</tr>
<tr>
<td></td>
<td>where ( w_i ) is a weight term (= 1 for equally weighted items)</td>
</tr>
<tr>
<td></td>
<td>* <strong>CV</strong> is the percent coefficient of variation. It is computed as the standard deviation divided by the mean and multiplied by 100. The coefficient of variation can be used to assess relative variation, for example when comparing the variation in data measured in different units or with different magnitudes.</td>
</tr>
<tr>
<td></td>
<td>* <strong>N Missing</strong> is the number of missing observations.</td>
</tr>
<tr>
<td>Horizontal Layout</td>
<td>Changes the orientation of the histogram and the reports to vertical or horizontal.</td>
</tr>
</tbody>
</table>

Histogram Options for Continuous Variables

Use the **Histogram Options** for continuous variables to change various aspects of the histogram, such as orientation, color, size, and so on.
Table 2.11  Descriptions of Histogram Options for Continuous Variables

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
</table>
| Shadowgram       | Replaces the histogram with a shadowgram. To understand a shadowgram, consider that if the bin width of a histogram is changed, the appearance of the histogram changes. A shadowgram overlays smoothed histograms across a range of bin widths. Dominant features of a distribution are less transparent on the shadowgram. Note that the following options are not available for shadowgrams:  
  • Std Error Bars
  • Show Counts
  • Show Percents |
| Vertical         | Changes the orientation of the histogram from a vertical to a horizontal orientation.                                                      |
| Std Error Bars   | Draws the standard error bar on each level of the histogram using the standard error. The standard error bar adjusts automatically when you adjust the number of bars with the hand tool (see “Resize Histogram Bars for Continuous Variables,” p. 27. |
| Set Bin Width    | Changes the bin width of the histogram bars. See “Resize Histogram Bars for Continuous Variables,” p. 27.                                    |
| Histogram Color  | Changes the color of the histogram bars.                                                                                                 |
| Count Axis       | Adds an axis that shows the frequency of column values represented by the histogram bars.                                                  |
|                  | **Note:** If you resize the histogram bars, the count axis also resizes.                                                                |
| Prob Axis        | Adds an axis that shows the proportion of column values represented by histogram bars.                                                   |
|                  | **Note:** If you resize the histogram bars, the probability axis also resizes.                                                          |
Performing Univariate Analysis
Options for Continuous Variables

Table 2.11 Descriptions of Histogram Options for Continuous Variables (Continued)

| Density Axis | The density is the length of the bars in the histogram. Both the count and probability are based on the following calculations:
| | prob = (bar width)*density
| | count = (bar width)*density*(total count)
| | When looking at density curves that are added by the Fit Distribution option, the density axis shows the point estimates of the curves.
| | **Note:** If you resize the histogram bars, the density axis remains constant.

| Show Percents | Labels the proportion of column values represented by each histogram bar.
| Show Counts | Labels the frequency of column values represented by each histogram bar.

Normal Quantile Plot

Use the **Normal Quantile Plot** option to visualize the extent to which the variable is normally distributed. If a variable is normally distributed, the normal quantile plot approximates a diagonal straight line. This type of plot is also called a quantile-quantile plot, or Q-Q plot.

The normal quantile plot also shows Lilliefors confidence bounds (Conover 1980) and probability and normal quantile scales. See Figure 2.14.

**Example of a Normal Quantile Plot**

1. Open the Cities.jmp sample data table.
2. Select **Analyze > Distribution**.
3. Select OZONE and click **Y, Columns**.
4. Click **OK**.
5. From the red triangle menu next to OZONE, select **Normal Quantile Plot**.
Note the following information:

- The y-axis shows the column values.
- The x-axis shows the empirical cumulative probability for each value, computed as follows:

\[
\frac{r_i}{N + 1}
\]

where \( r_i \) is the rank of the \( i \)th observation, and \( N \) is the number of nonmissing (and nonexcluded) observations.

- The normal quantile values are computed as follows:

\[
\Phi^{-1}\left( \frac{r_i}{N + 1} \right)
\]

where \( \Phi \) is the cumulative probability distribution function for the normal distribution.

These normal quantile values are Van Der Waerden approximations to the order statistics that are expected for the normal distribution.

### Outlier Box Plot

Use the outlier box plot (also called a Tukey outlier box plot) to see the distribution and identify possible outliers. Generally, box plots show selected quantiles of continuous distributions.
**Example of an Outlier Box Plot**

1. Open the Candy Bars.jmp sample data table.
2. Select **Analyze > Distribution**.
3. Select Total fat g and click **Y, Columns**.
4. Click **OK**.

**Figure 2.15** Example of an Outlier Box Plot

Note the following aspects about outlier box plots (see Figure 2.15):

- The vertical line within the box represents the median sample value.
- The confidence diamond contains the mean and the upper and lower 95% of the mean. If you drew a line through the middle of the diamond, you would have the mean. The top and bottom points of the diamond represent the upper and lower 95% of the mean.
- The ends of the box represent the 25th and 75th quantiles, also expressed as the 1st and 3rd quartile, respectively.
- The difference between the 1st and 3rd quartiles is called the *interquartile range*.
- The box has lines that extend from each end, sometimes called *whiskers*. The whiskers extend from the ends of the box to the outermost data point that falls within the distances computed as follows:
  - 1st quartile - 1.5*(interquartile range)
  - 3rd quartile + 1.5*(interquartile range)
  If the data points do not reach the computed ranges, then the whiskers are determined by the upper and lower data point values (not including outliers).
- The bracket outside of the box identifies the *shortest half*, which is the most dense 50% of the observations (Rousseauw and Leroy 1987).
Options for Continuous Variables

Remove Objects from the Outlier Box Plot

To remove the confidence diamond or the shortest half, proceed as follows:

1. Right-click on the outlier box plot and select **Customize**.
2. Click **Box Plot**.
3. Deselect the check box next to **Confidence Diamond** or **Shortest Half**.

For more details about the Customize Graph window, see the *Using JMP* book.

Quantile Box Plot

The **Quantile Box Plot** displays specific quantiles from the Quantiles report. If the distribution is symmetric, the quantiles in the box plot are approximately equidistant from each other. At a glance, you can see if the distribution is symmetric. For example, if the quantile marks are grouped closely at one end, but have greater spacing at the other end, the distribution is skewed toward the end with more spacing. See Figure 2.16.

**Example of a Quantile Box Plot**

1. Open the Candy Bars.jmp sample data table.
2. Select **Analyze > Distribution**.
3. Select **Total fat g** and click **Y, Columns**.
4. Click **OK**.
5. From the red triangle menu next to **Total fat g**, deselect **Outlier Box Plot**, and select **Quantile Box Plot**.

![Figure 2.16 Example of a Quantile Box Plot](image)
Quantiles are values where the $p$th quantile is larger than $p\%$ of the values. For example, 10% of the data lies below the 10th quantile, and 90% of the data lies below the 90th quantile.

**Stem and Leaf**

The **Stem and Leaf** option constructs a plot that is a variation on the histogram.

**Example of a Stem and Leaf Plot**

1. Open the Candy Bars.jmp sample data table.
2. Select **Analyze > Distribution**.
3. Select Total fat g and click **Y, Columns**.
4. Click **OK**.
5. From the red triangle menu next to Total fat g, select **Stem and Leaf**.

![Figure 2.17 Example of a Stem and Leaf Plot](image-url)
Each line of the plot has a Stem value that is the leading digit of a range of column values. The Leaf values are made from the next-in-line digits of the values. You can see the data point by joining the stem and leaf. In some cases, the numbers on the stem and leaf plot are rounded versions of the actual data in the table. The stem-and-leaf plot actively responds to clicking and the brush tool.

**CDF Plot**

The CDF Plot option creates a plot of the empirical cumulative distribution function. Use the CDF plot to determine the percent of data that is at or below a given value on the x-axis. See Figure 2.18.

**Example of a CDF Plot**

1. Open the Candy Bars.jmp sample data table.
2. Select Analyze > Distribution.
3. Select Total fat g and click Y, Columns.
4. Click OK.
5. From the red triangle menu next to Total fat g, select CDF Plot.

---

**Figure 2.18 Example of a CDF Plot**
Performing Univariate Analysis

Chapter 2

Options for Continuous Variables

For example, use the crosshair tool and position it at 10 on the x-axis. You can see that about 0.2778 percent of the data is at or below the value of 10.

Test Mean

Use the Test Mean option to perform a one-sample test for the mean. If you type a value for the standard deviation, a z-test is performed. Otherwise, the sample standard deviation is used to perform a t-test. You can also request the nonparametric Wilcoxon Signed-Rank test.

Example of Test Mean

1. Open the Candy Bars.jmp sample data table.
2. Select Analyze > Distribution.
3. Select Total fat g and click Y, Columns.
4. Click OK.
5. From the red triangle next to Total fat g, select Test Mean.
6. In the Test Mean window, type 12 next to Specify Hypothesized Mean.
7. Click OK.

Figure 2.19 Example of the Test Mean Option
Use the **Test Mean** option repeatedly to test different values. Each time you test the mean, a new Test Mean report appears.

**Table 2.12 Description of the Test Mean Report**

<table>
<thead>
<tr>
<th>Statistics that are calculated for Test Mean</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>t Test (or z Test)</td>
<td>Lists the value of the test statistic and the <em>p</em>-values for the two-sided and one-sided alternatives.</td>
</tr>
<tr>
<td>Signed-Rank</td>
<td>(Only appears for the Wilcoxon Signed-Rank test) Lists the value of the Wilcoxon signed-rank statistic followed by the <em>p</em>-values for the two-sided and one-sided alternatives. The test assumes only that the distribution is symmetric. The Wilcoxon signed-rank test uses average ranks for ties. The <em>p</em>-values are exact for <em>n</em> ≤ 20 where <em>n</em> is the number of values not equal to the hypothesized value. For <em>n</em> &gt; 20 a Student’s <em>t</em> approximation given by Iman (1974) is used.</td>
</tr>
</tbody>
</table>

**Probability values**

| *Prob>|t| | The probability of obtaining an absolute *t*-value by chance alone that is greater than the observed *t*-value when the population mean is equal to the hypothesized value. This is the *p*-value for observed significance of the two-tailed *t*-test. |
| *Prob>*t | The probability of obtaining a *t*-value greater than the computed sample *t* ratio by chance alone when the population mean is not different from the hypothesized value. This is the *p*-value for an upper-tailed test. |
| *Prob<*>t | The probability of obtaining a *t*-value less than the computed sample *t* ratio by chance alone when the population mean is not different from the hypothesized value. This is the *p*-value for a lower-tailed test. |

**Red triangle options for Test Mean**

| PValue animation | Starts an interactive visual representation of the *p*-value. Enables you to change the hypothesized mean value while watching how the change affects the *p*-value. |
| Power animation  | Starts an interactive visual representation of power and beta. Enables you to change the hypothesized mean and sample mean while watching how the changes affect power and beta. |
| Remove Test      | Removes the mean test. |
Performing Univariate Analysis
Options for Continuous Variables

Test Std Dev

Use the Test Std Dev option to perform a one-sample test for the standard deviation (details in Neter, Wasserman, and Kutner 1990).

Example of Test Std Dev

1. Open the Candy Bars.jmp sample data table.
2. Select Analyze > Distribution.
3. Select Total fat g and click Y, Columns.
4. Click OK.
5. From the red triangle next to Total fat g, select Test Std Dev.
6. In the Test Standard Deviation window, type 4 next to Specify Hypothesized Standard Deviation.
7. Click OK.

Figure 2.20 Example of the Test Std Dev Option

Use the Test Std Dev option repeatedly to test different values. Each time you test the standard deviation, a new Test Standard Deviation report appears.
Options for Continuous Variables

Confidence Intervals for Continuous Variables

The Confidence Interval options display confidence intervals for the mean and standard deviation. The 0.90, 0.95, and 0.99 options compute two-sided confidence intervals for the mean and standard deviation. Use the Confidence Interval > Other option to select a confidence level, and select one-sided or two-sided confidence intervals. You can also type a known sigma. If you use a known sigma, the confidence interval for the mean is based on z-values rather than t-values.

Example of Confidence Intervals for Continuous Variables

1. Open the Candy Bars jmp sample data table.
2. Select Analyze > Distribution.
3. Select Total fat g and click Y, Columns.
4. Click OK.
5. From the red triangle next to Total fat g, select Confidence Interval > 0.95.
Performing Univariate Analysis
Chapter 2
Options for Continuous Variables

Figure 2.21 Example of a Confidence Interval Report

The Confidence Intervals report shows the mean and standard deviation parameter estimates with upper and lower confidence limits for 1 - \( \alpha \).

**Save Commands for Continuous Variables**

Use the Save menu commands to save information about continuous variables. Each Save command generates a new column in the current data table. The new column is named by appending the variable name (denoted <colname> in the following definitions) to the Save command name. See Table 2.14.

Select the Save commands repeatedly to save the same information multiple times under different circumstances, such as before and after combining histogram bars. If you use a Save command multiple times, the column name is numbered (name1, name2, and so on) to ensure unique column names.

**Table 2.14 Descriptions of the Save Commands**

<table>
<thead>
<tr>
<th>Command</th>
<th>Column Added to Data Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level Numbers</td>
<td>Level &lt;colname&gt;</td>
<td>The level number of each observation corresponds to the histogram bar that contains the observation. The histogram bars are numbered from low to high, beginning with 1.</td>
</tr>
<tr>
<td>Level Midpoints</td>
<td>Midpoint &lt;colname&gt;</td>
<td>The midpoint value for each observation is computed by adding half the level width to the lower level bound.</td>
</tr>
</tbody>
</table>
### Table 2.14 Descriptions of the Save Commands (Continued)

<table>
<thead>
<tr>
<th>Command</th>
<th>Column Added to Data Table</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ranks</td>
<td>Ranked &lt;colname&gt;</td>
<td>Provides a ranking for each of the corresponding column’s values starting at 1. Duplicate response values are assigned consecutive ranks in order of their occurrence in the data table.</td>
</tr>
<tr>
<td>Ranks Averaged</td>
<td>RankAvgd &lt;colname&gt;</td>
<td>If a value is unique, then the averaged rank is the same as the rank. If a value occurs ( k ) times, the average rank is computed as the sum of the value’s ranks divided by ( k ).</td>
</tr>
<tr>
<td>Prob Scores</td>
<td>Prob &lt;colname&gt;</td>
<td>For ( N ) nonmissing scores, the probability score of a value is computed as the averaged rank of that value divided by ( N + 1 ). This column is similar to the empirical cumulative distribution function.</td>
</tr>
</tbody>
</table>
| Normal Quantiles | N-Quantile <colname>        | The normal quantile values are computed as follows: \[ \Phi^{-1}\left(\frac{r_i}{N+1}\right) \] where: \[
\begin{align*}
\Phi & \text{ is the cumulative probability distribution function for the normal distribution} \\
r_i & \text{ is the rank of the } i\text{th observation} \\
N & \text{ is the number of nonmissing observations}
\end{align*}
\]
| Standardized     | Std <colname>               | Creates standardized values for numeric columns with the formula editor, using the following formula: \[ \frac{X - \bar{X}}{S_X} \] where: \[
\begin{align*}
X & \text{ is the original column} \\
\bar{X} & \text{ is the mean of column } X \\
S_X & \text{ is the standard deviation of column } X
\end{align*}
\]
| Spec Limits      | (none)                      | Stores the specification limits applied in a capability analysis as a column property of the corresponding column in the current data table. Automatically retrieves and displays the specification limits when you repeat the capability analysis. |
| Script to Log    | (none)                      | Prints the script to the log window. Run the script to recreate the analysis. |
Prediction Intervals

Prediction intervals concern a single observation, or the mean and standard deviation of the next randomly selected sample. The calculations assume that the given sample is selected randomly from a normal distribution. Select one-sided or two-sided prediction intervals.

When you select the **Prediction Interval** option for a variable, the Prediction Intervals window appears. See Figure 2.22.

**Example of Prediction Intervals**

1. Open the Cities.jmp sample data table.
2. Select **Analyze > Distribution**.
3. Select OZONE and click **Y, Columns**.
4. Click **OK**.
5. From the red triangle next to OZONE, select **Prediction Interval**.

![Figure 2.22 The Prediction Intervals Window](image)

6. In the Prediction Intervals window, type 10 next to **Enter number of future samples**.
7. Click **OK**.
Figure 2.23 Example of a Prediction Interval Report

In this example, you can be 95% confident about the following:

- Each of the next 10 observations will be between .013755 and .279995.
- The mean of the next 10 observations will be between .115596 and .178154.
- The standard deviation of the next 10 observations will be between .023975 and .069276.

For details about how prediction intervals are computed, see “Statistical Details for Prediction Intervals,” p. 78.

Tolerance Intervals

A tolerance interval contains at least a specified proportion of the population. It is a confidence interval for a specified proportion of the population, not the mean, or standard deviation. Complete discussions of tolerance intervals are found in Hahn and Meeker (1991) and in Tamhane and Dunlop (2000).

When you select the Tolerance Interval option for a variable, the Tolerance Intervals window appears. See Figure 2.24. The calculations are based on the assumption that the given sample is selected randomly from a normal distribution. JMP can produce one-sided or two-sided tolerance intervals.
Example of Tolerance Intervals

1. Open the Cities.jmp sample data table.
2. Select Analyze > Distribution.
4. Click OK.
5. From the red triangle menu next to OZONE, select Tolerance Interval.

Figure 2.24 The Tolerance Intervals Window

6. Keep the default selections, and click OK.

Figure 2.25 Example of a Tolerance Interval Report
In this example, you can be 95% confident that at least 90% of the population lie between .057035 and .236715, based on the Lower TI (tolerance interval) and Upper TI values.

For details about how tolerance intervals are computed, see “Statistical Details for Tolerance Intervals,” p. 78.

Capability Analysis

The Capability Analysis option measures the conformance of a process to given specification limits. When you select the Capability Analysis option for a variable, the Capability Analysis window appears. See Figure 2.26.

Example of Capability Analysis

1. Open the Cities.jmp sample data table.
2. Select Analyze > Distribution.
4. Click OK.
5. From the red triangle menu next to OZONE, select Capability Analysis.

Figure 2.26 The Capability Analysis Window

6. Type 0 for the Lower Spec Limit.
7. Type 1 for the Target.
8. Type 2 for the Upper Spec Limit.
9. Keep the rest of the default selections, and click OK.
Performing Univariate Analysis
Capability Analysis

Chapter 2

Figure 2.27 Example of the Capability Analysis Report

Note: To save the specification limits to the data table as a column property, select Save > Spec Limits. When you repeat the capability analysis, the saved specification limits are automatically retrieved.

The Capability Analysis report is organized into two sections: Capability Analysis and the distribution type (Long Term Sigma, Specified Sigma, and so on).

Capability Analysis Descriptions

Capability Analysis descriptions appear as follows:

- Table 2.15 describes the Capability Analysis window.
- Table 2.16 describes the Capability Analysis report.
- Table 2.17 describes the Capability Analysis red triangle options.
Table 2.15 Description of the Capability Analysis Window

<table>
<thead>
<tr>
<th>Distribution type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>By default, the normal distribution is assumed when calculating the capability statistics and the percent out of the specification limits. To perform a capability analysis on non-normal distributions, see the description of Spec Limits under “Fit Distribution Options,” p. 71.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Capability Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long Term Sigma</td>
<td>Calculates capability indices using different estimates for sigma (σ). See “Statistical Details for Capability Analysis,” p. 79.</td>
</tr>
<tr>
<td>Specified Sigma</td>
<td>Motion Range, Range Span</td>
</tr>
<tr>
<td>Short Term Sigma</td>
<td>Group by Fixed Subgroup Size</td>
</tr>
</tbody>
</table>

Table 2.16 Description of the Capability Analysis Report

<table>
<thead>
<tr>
<th>Capability Analysis</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specification</td>
<td>Lists the specification limits.</td>
</tr>
<tr>
<td>Value</td>
<td>Lists the values that you specified for each specification limit and the target.</td>
</tr>
<tr>
<td>Portion and % Actual</td>
<td>Portion labels describe the numbers in the % Actual column, as follows:</td>
</tr>
<tr>
<td></td>
<td>• Below LSL gives the percentage of the data that is below the lower specification limit.</td>
</tr>
<tr>
<td></td>
<td>• Above USL gives the percentage of the data that is above the upper specification limit.</td>
</tr>
<tr>
<td></td>
<td>• Total Outside gives the total percentage of the data that is either below LSL or above USL.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Distribution type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Capability</td>
<td>Type of process capability indices. See Table 2.22 “Descriptions of Capability Index Names and Formulas,” p. 80.</td>
</tr>
<tr>
<td>Note: There is a preference for Capability called Ppk Capability Labeling that labels the long-term capability output with Ppk labels. Open the Preference window (File &gt; Preferences), then select Platforms &gt; Distribution to see this preference.</td>
<td></td>
</tr>
<tr>
<td>Index</td>
<td>Process capability index values.</td>
</tr>
<tr>
<td>Upper CI</td>
<td>Upper confidence interval.</td>
</tr>
<tr>
<td>Lower CI</td>
<td>Lower confidence interval.</td>
</tr>
</tbody>
</table>
Portion and Percent labels describe the numbers in the Percent column, as follows:

- **Below LSL** gives the percentage of the fitted distribution that is below the lower specification limit.
- **Above USL** gives the percentage of the fitted distribution that is above the upper specification limit.
- **Total Outside** gives the total percentage of the fitted distribution that is either below LSL or above USL.

**PPM (parts per million)** The PPM value is the **Percent** column multiplied by 10,000.

**Sigma Quality** Sigma Quality is frequently used in Six Sigma methods, and is also referred to as the *process sigma*.

\[
\text{Sigma Quality} = \text{Normal Quantile} \left( 1 - \frac{\text{Expected (# defects)}}{n} \right) + 1.5
\]

For example, if there are 3 defects in \(n=1,000,000\) observations, the formula yields 6.03, or a 6.03 sigma process. The above and below columns do not sum to the total column because Sigma Quality uses values from the normal distribution, and is therefore not additive.

**Table 2.17** Descriptions of the Capability Analysis Options

| **Z Bench** | Shows the values (represented by Index) of the Benchmark Z statistics. According to the AIAG Statistical Process Control manual, \(Z\) represents the number of standard deviation units from the process average to a value of interest such as an engineering specification. When used in capability assessment, \(Z\) USL is the distance to the upper specification limit and \(Z\) LSL is the distance to the lower specification limit. Here are the Benchmark Z formulas:

\[
\begin{align*}
Z \text{ USL} &= \frac{\text{USL} - \text{Xbar}}{\sigma} = 3 \times \text{CPU} \\
Z \text{ LSL} &= \frac{\text{Xbar} - \text{LSL}}{\sigma} = 3 \times \text{CPL} \\
Z \text{ Bench} &= \text{Inverse Cumulative Prob}(1 - \text{P(LSL)} - \text{P(USL)}) \\
\text{where} \\
\text{P(LSL)} &= \text{Prob}(X < \text{LSL}) = 1 - \text{Cum Prob}(Z \text{ LSL}) \\
\text{P(USL)} &= \text{Prob}(X > \text{USL}) = 1 - \text{Cum Prob}(Z \text{ USL})
\end{align*}
\]

| **Capability Animation** | Interactively change the specification limits and the process mean to see the effects on the capability statistics. This option is available only for capability analyses based on the Normal distribution. |
Fit Distributions

Use the Continuous or Discrete Fit options to fit a distribution to a continuous or discrete variable.

Example of Fitting a Lognormal Distribution

1. Open the Lipid Data.jmp sample data table.
2. Select Analyze > Distribution.
4. Click OK.
5. (Optional) Rotate the histogram vertically. From the red triangle menu next to Cholesterol, select Histogram Options > Vertical.
6. From the red triangle menu next to Cholesterol, select Continuous Fit > LogNormal.

Figure 2.28 Example of a Continuous Fit

A curve is overlaid on the histogram, and a Parameter Estimates report is added to the report window. A red triangle menu contains additional options. See “Fit Distribution Options,” p. 71.

Note: The Life Distribution platform also contains options for distribution fitting that might use different parameterizations and allow for censoring. See the Quality and Reliability Methods book.
Continuous Fit

Use the Continuous Fit options to fit a distribution (such as Normal, Lognormal, or Weibull) to a continuous variable.

Normal

The Normal fitting option estimates the parameters of the normal distribution. The normal distribution is often used to model measures that are symmetric with most of the values falling in the middle of the curve. Select the Normal fitting for any set of data and test how well a normal distribution fits your data.

The parameters for the normal distribution are as follows:

- \( \mu \) (the mean) defines the location of the distribution on the \( x \)-axis
- \( \sigma \) (standard deviation) defines the dispersion or spread of the distribution

The standard normal distribution occurs when \( \mu = 0 \) and \( \sigma = 1 \). The Parameter Estimates table shows estimates of \( \mu \) and \( \sigma \), with upper and lower 95% confidence limits.

\[
\text{pdf: } \frac{1}{\sqrt{2\pi}\sigma^2} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad \text{for} \quad -\infty < x < \infty; \quad -\infty < \mu < \infty; \quad 0 < \sigma
\]

\[E(x) = \mu\]
\[\text{Var}(x) = \sigma^2\]

LogNormal

The LogNormal fitting option estimates the parameters \( \mu \) (scale) and \( \sigma \) (shape) for the two-parameter lognormal distribution. A variable \( Y \) is lognormal if and only if \( X = \ln(Y) \) is normal. The data must be greater than zero.

\[
\text{pdf: } \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right) \quad \text{for} \quad 0 \leq x; \quad -\infty < \mu < \infty; \quad 0 < \sigma
\]

\[E(x) = \exp(\mu + \sigma^2/2)\]
\[\text{Var}(x) = \exp(2(\mu + \sigma^2)) - \exp(2\mu + \sigma^2)\]

Weibull, Weibull with threshold, and Extreme Value

The Weibull distribution has different shapes depending on the values of \( \alpha \) (scale) and \( \beta \) (shape). It often provides a good model for estimating the length of life, especially for mechanical devices and in biology. The Weibull option is the same as the Weibull with threshold option, with a threshold (\( \theta \)) parameter of zero. For the Weibull with threshold option, JMP estimates the threshold as the minimum value. If you know what the threshold should be, set it by using the Fix Parameters option. See “Fit Distribution Options,” p. 71.

The pdf for the Weibull with threshold is as follows:
Performing Univariate Analysis

Chapter 2

63

Fit Distributions

pdf: \( \frac{\beta}{\alpha \beta} (x - \theta)^{\beta - 1} \exp \left( -\frac{(x - \theta)^{\beta}}{\alpha} \right) \) for \( \alpha, \beta > 0; \ \theta < x \)

E(x) = \theta + \alpha \Gamma \left( 1 + \frac{1}{\beta} \right)

Var(x) = \alpha^2 \left\{ \Gamma \left( 1 + \frac{2}{\beta} \right) - \Gamma^2 \left( 1 + \frac{1}{\beta} \right) \right\}

where \( \Gamma(\cdot) \) is the Gamma function.

The Extreme Value distribution is a two parameter Weibull \((\alpha, \beta)\) distribution with the transformed parameters \( \delta = 1 / \beta \) and \( \lambda = \ln(\alpha) \).

**Exponential**

The exponential distribution is especially useful for describing events that randomly occur over time, such as survival data. The exponential distribution might also be useful for modeling elapsed time between the occurrence of non-overlapping events, such as the time between a user’s computer query and response of the server, the arrival of customers at a service desk, or calls coming in at a switchboard.

The Exponential distribution is a special case of the two-parameter Weibull when \( \beta = 1 \) and \( \alpha = \sigma \), and also a special case of the Gamma distribution when \( \alpha = 1 \).

pdf: \( \frac{1}{\sigma} \exp(-x/\sigma) \) for \( 0 < \sigma; \ 0 \leq x \)

E(x) = \sigma

Var(x) = \sigma^2

Devore (1995) notes that an exponential distribution is **memoryless**. Memoryless means that if you check a component after \( t \) hours and it is still working, the distribution of additional lifetime (the conditional probability of additional life given that the component has lived until \( t \)) is the same as the original distribution.

**Gamma**

The **Gamma** fitting option estimates the gamma distribution parameters, \( \alpha > 0 \) and \( \sigma > 0 \). The parameter \( \alpha \), called alpha in the fitted gamma report, describes shape or curvature. The parameter \( \sigma \), called sigma, is the scale parameter of the distribution. A third parameter, \( \theta \), called the **Threshold**, is the lower endpoint parameter. It is set to zero by default, unless there are negative values. You can also set its value by using the **Fix Parameters** option. See “Fit Distribution Options,” p. 71.

pdf: \( \frac{1}{\Gamma(\alpha)\sigma^\alpha} (x - \theta)^{\alpha - 1} \exp\left( -(x - \theta)/\sigma \right) \) for \( \theta \leq x; \ 0 < \alpha, \sigma \)

E(x) = \alpha \sigma + \theta

Var(x) = \alpha \sigma^2
Performing Univariate Analysis

Chapter 2

Fit Distributions

- The standard gamma distribution has $\sigma = 1$. Sigma is called the scale parameter because values other than 1 stretch or compress the distribution along the x-axis.
- The Chi-square $\chi^2_v$ distribution occurs when $\sigma = 2$, $\alpha = v/2$, and $\theta = 0$.
- The exponential distribution is the family of gamma curves that occur when $\alpha = 1$ and $\theta = 0$.

The standard gamma density function is strictly decreasing when $\alpha \leq 1$. When $\alpha > 1$, the density function begins at zero, increases to a maximum, and then decreases.

**Beta**

The standard beta distribution is useful for modeling the behavior of random variables that are constrained to fall in the interval 0, 1. For example, proportions always fall between 0 and 1. The Beta fitting option estimates two shape parameters, $\alpha > 0$ and $\beta > 0$. There are also $\theta$ and $\sigma$, which are used to define the lower threshold as $\theta$, and the upper threshold as $\theta + \sigma$. The beta distribution has values only for the interval defined by $\theta \leq x \leq (\theta + \sigma)$. The $\theta$ is estimated as the minimum value, and $\sigma$ is estimated as the range. The standard beta distribution occurs when $\theta = 0$ and $\sigma = 1$.

Set parameters to fixed values by using the Fix Parameters option. The upper threshold must be greater than or equal to the maximum data value, and the lower threshold must be less than or equal to the minimum data value. For details about the Fix Parameters option, see “Fit Distribution Options,” p. 71.

pdf: $\frac{1}{B(\alpha, \beta)\sigma^{\alpha + \beta}}(x - \theta)^{\alpha - 1}(\theta + \sigma - x)^{\beta - 1}$ for $\theta \leq x \leq \theta + \sigma; \ 0 < \sigma, \alpha, \beta$

$E(x) = \theta + \sigma \frac{\alpha}{\alpha + \beta}$

$Var(x) = \frac{\sigma^2 \alpha \beta}{(\alpha + \beta)^2 (\alpha + \beta + 1)}$

where $B(\cdot) \ is \ the \ Beta \ function.$

**Normal Mixtures**

The Normal Mixtures option fits a mixture of normal distributions. This flexible distribution is capable of fitting multi-modal data.

Fit a mixture of two or three normal distributions by selecting the Normal 2 Mixture or Normal 3 Mixture options. Alternatively, you can fit a mixture of $k$ normal distributions by selecting the Other option. A separate mean, standard deviation, and proportion of the whole is estimated for each group.

pdf: $\sum_{i=1}^{k} \frac{\pi_i \phi \left( \frac{x - \mu_i}{\sigma_i} \right)}{\sigma_i}$

$E(x) = \sum_{i=1}^{k} \pi_i \mu_i$
Var(x) = \sum_{i=1}^{k} \pi_i \left(\mu_i^2 + \sigma_i^2\right) - \left(\sum_{i=1}^{k} \pi_i \mu_i\right)^2

where \(\mu_i, \sigma_i, \) and \(\pi_i\) are the respective mean, standard deviation, and proportion for the \(i^{th}\) group, and \(\phi(\cdot)\) is the standard normal pdf.

**Smooth Curve**

The **Smooth Curve** option fits a smooth curve using nonparametric density estimation (kernel density estimation). The smooth curve is overlaid on the histogram and a slider appears beneath the plot. Control the amount of smoothing by changing the kernel standard deviation with the slider. The initial Kernel Std estimate is formed by summing the normal densities of the kernel standard deviation located at each data point.

**Johnson Su, Johnson Sb, Johnson Sl**

The Johnson system of distributions contains three distributions that are all based on a transformed normal distribution. These three distributions are the Johnson Su, which is unbounded for \(Y\); the Johnson Sb, which is bounded on both tails (0 < \(Y\) < 1); and the Johnson Sl, leading to the lognormal family of distributions.

**Note:** The S refers to system, the subscript of the range. Although we implement a different method, information about selection criteria for a particular Johnson system can be found in Slifker and Shapiro (1980).

Johnson distributions are popular because of their flexibility. In particular, the Johnson distribution system is noted for its data-fitting capabilities because it supports every possible combination of skewness and kurtosis.

If \(Z\) is a standard normal variate, then the system is defined as follows:

\[ Z = \gamma + \delta f(Y) \]

where, for the Johnson Su:

\[ f(Y) = \ln\left(Y + \sqrt{1 + Y^2}\right) = \sinh^{-1}Y \]

\[ Y = \frac{X - \theta}{\sigma} \quad -\infty < X < \infty \]

where, for the Johnson Sb:

\[ f(Y) = \ln\left(\frac{Y}{1-Y}\right) \]

\[ Y = \frac{X - \theta}{\sigma} \quad 0 < X < \theta + \sigma \]
Performing Univariate Analysis

Chapter 2

Fit Distributions

and for the Johnson Sl, where \( \sigma = \pm 1 \).

\[
Y = \ln(Y)
\]

\[
f(Y) = 1
\]

\[
y = \frac{x - \theta}{\sigma}
\]

\[
0 < x < \infty \quad \text{if } \sigma = 1
\]

\[
-\infty < x < \theta \quad \text{if } \sigma = -1
\]

Johnson Su

pdf: \[
\frac{\delta}{\sigma} \left[ 1 + \left( \frac{x - \theta}{\sigma} \right)^2 \right] \phi \left[ \gamma + \delta \sinh^{-1} \left( \frac{x - \theta}{\sigma} \right) \right]
\]

\[
\text{for } -\infty < x, \theta, \gamma < \infty; \quad 0 < \sigma, \delta
\]

Johnson Sb

pdf: \[
\phi \left[ \gamma + \delta \ln \left( \frac{x - \theta}{\sigma - (x - \theta)} \right) \right] \left( \frac{\delta \sigma}{(x - \theta)(\sigma - (x - \theta))} \right)
\]

\[
\text{for } \theta < x < \theta + \sigma; \quad 0 < \sigma
\]

Johnson Sl

pdf: \[
\frac{\delta}{|x - \theta|} \phi \left[ \gamma + \delta \ln \left( \frac{x - \theta}{\sigma} \right) \right]
\]

\[
\text{for } \theta < x \text{ if } \sigma = 1; \quad \theta > x \text{ if } \sigma = -1
\]

where \( \phi(\cdot) \) is the standard normal pdf. The \( \mathbb{E}(x) \) and \( \text{Var}(x) \) are not given for any of the Johnson distributions due to the complexity of finding a closed-form solution.

Note: The parameter confidence intervals are hidden in the default report. Parameter confidence intervals are not very meaningful for Johnson distributions, because they are transformations to normality. To show parameter confidence intervals, right-click in the report and select **Columns > Lower 95% and Upper 95%**.

Generalized Log (Glog)

This distribution is useful for fitting data that are rarely normally distributed and often have non-constant variance, like biological assay data. The Glog distribution is described with the parameters \( \mu \) (location), \( \sigma \) (scale), and \( \lambda \) (shape).

pdf: \[
\phi \left[ \frac{1}{\sigma} \log \left( \frac{x + \sqrt{x^2 + \lambda^2}}{2} - \mu \right) \right] \frac{x + \sqrt{x^2 + \lambda^2}}{\sigma (x^2 + \lambda^2 + x \sqrt{x^2 + \lambda^2})}
\]

\[
\text{for } 0 \leq \lambda; \quad 0 < \sigma; \quad -\infty < \mu < \infty
\]

The \( \mathbb{E}(x) \) and \( \text{Var}(x) \) are not given due to the complexity of finding a closed-form solution.
The Glog distribution is a transformation to normality, and comes from the following relationship:

\[ z = \frac{1}{\sigma} \left[ \log\left( \frac{x + \sqrt{x^2 + \lambda^2}}{2} \right) - \mu \right] - N(0,1), \text{ then } x \sim \text{Glog}(\mu, \sigma, \lambda). \]

When \( \lambda = 0 \), the Glog reduces to the LogNormal \((\mu, \sigma)\).

**Note:** The parameter confidence intervals are hidden in the default report. Parameter confidence intervals are not very meaningful for the GLog distribution, because it is a transformation to normality. To show parameter confidence intervals, right-click in the report and select **Columns > Lower 95%** and **Upper 95%**.

**All**

The **All** option fits all applicable continuous distributions to a variable. The Compare Distributions report contains statistics about each fitted distribution. Use the check boxes to show or hide a fit report and overlay curve for the selected distribution. By default, the best fit distribution is selected. See Figure 2.29.

**Example of the All Option**

1. Open the Lipid Data.jmp sample data table.
2. Select **Analyze > Distribution**.
3. Select **Cholesterol** and click **Y, Columns**.
4. Click **OK**.
5. From the red triangle menu next to Cholesterol, select **Continuous Fit > All**.
The ShowDistribution list is sorted by AICc in ascending order. The formula for AICc is as follows:

$$AICc = -2 \log L + \frac{2\nu (\nu + 1)}{n - (\nu + 1)}$$

where:
- \( \log L \) is the logLikelihood
- \( n \) is the sample size
- \( \nu \) is the number of parameters
If your data has negative values, the ShowDistribution list does not include those distributions that require data with positive values. If your data has non-integer values, the list of distributions does not include discrete distributions. Distributions with threshold parameters, like Beta and Johnson Sb, are not included in the list of possible distributions.

**Discrete Fit**

Use the Discrete Fit options to fit a distribution (such as Poisson or Binomial) to a discrete variable.

**Poisson**

The Poisson distribution has a single scale parameter \( \lambda > 0 \).

\[
\text{pdf: } \frac{e^{-\lambda} \lambda^x}{x!} \quad \text{for} \quad 0 \leq \lambda < \infty; \quad x = 0,1,2,...
\]

\[E(x) = \lambda\]

\[\text{Var}(x) = \lambda\]

Since the Poisson distribution is a discrete distribution, the overlaid curve is a step function, with jumps occurring at every integer.

**Binomial**

The Binomial option accepts data in two formats: a constant sample size, or a column containing sample sizes.

\[
\text{pdf: } \binom{n}{x} p^x (1-p)^{n-x} \quad \text{for} \quad 0 \leq p \leq 1; \quad x = 0,1,2,...,n
\]

\[E(x) = np\]

\[\text{Var}(x) = np(1-p)\]

where \( n \) is the number of independent trials.

---

**Note:** The confidence interval for the binomial parameter is a Score interval. See Agresti (1998).

**Gamma Poisson**

This distribution is useful when the data is a combination of several \( \text{Poisson}(\lambda) \) distributions, each with a different \( \lambda \). One example is the overall number of accidents combined from multiple intersections, when the mean number of accidents \( \lambda \) varies between the intersections.

The Gamma Poisson distribution results from assuming that \( x|\lambda \) follows a Poisson distribution, while \( \lambda \) follows a Gamma(\( \alpha, \beta \)). The Gamma Poisson has parameters \( \lambda = \alpha \beta \) and \( \sigma = \beta + 1 \). The \( \sigma \) is a dispersion parameter. If \( \sigma > 1 \), there is overdispersion, meaning there is more variation in \( x \) than explained by the Poisson alone. If \( \sigma = 1 \), \( x \) reduces to \( \text{Poisson}(\lambda) \).
Performing Univariate Analysis
Chapter 2

Fit Distributions

pdf: \( \frac{\Gamma(x + \frac{\lambda}{\sigma - 1})}{\Gamma(x + 1) \Gamma\left(\frac{\lambda}{\sigma - 1}\right)} \left(\frac{\sigma - 1}{\sigma}\right)^x \frac{\lambda}{\sigma - 1} \) for \( 0 < \lambda; \ 1 \leq \sigma; \ x = 0, 1, 2, \ldots \)

\[ E(x) = \lambda \]
\[ \text{Var}(x) = \lambda \sigma \]

where \( \Gamma(\cdot) \) is the Gamma function.

Remember that \( x|\lambda \sim \text{Poisson}(\lambda) \), while \( \lambda \sim \text{Gamma}(\alpha, \beta) \). The platform estimates \( \lambda = \alpha \beta \) and \( \sigma = \beta + 1 \). To obtain estimates for \( \alpha \) and \( \beta \), use the following formulas:

\[ \hat{\beta} = \sigma - 1 \]
\[ \hat{\alpha} = \frac{\lambda}{\sigma} \]

If the estimate of \( \sigma \) is 1, the formulas do not work. In that case, the Gamma Poisson has reduced to the Poisson(\( \lambda \)), and \( \lambda \) is the estimate of \( \lambda \).

If the estimate for \( \alpha \) is an integer, the Gamma Poisson is equivalent to a Negative Binomial with the following pdf:

\[ p(y) = \binom{y + r - 1}{y} p^r (1 - p)^y \quad \text{for} \quad 0 \leq y \]

with \( r = \alpha \) and \( (1-p)/p = \beta \).

**Beta Binomial**

This distribution is useful when the data is a combination of several Binomial(p) distributions, each with a different p. One example is the overall number of defects combined from multiple manufacturing lines, when the mean number of defects (p) varies between the lines.

The Beta Binomial distribution results from assuming that \( x|p \) follows a Binomial(n,p) distribution, while p follows a Beta(\( \alpha, \beta \)). The Beta Binomial has parameters \( p = \alpha/(\alpha + \beta) \) and \( \delta = 1/(\alpha + \beta + 1) \). The \( \delta \) is a dispersion parameter. When \( \delta > 0 \), there is overdispersion, meaning there is more variation in \( x \) than explained by the Binomial alone. When \( \delta < 0 \), there is underdispersion. When \( \delta = 0 \), \( x \) is distributed as Binomial(n,p). The Beta Binomial only exists when \( n \geq 2 \).

pdf: \( \binom{n}{x} \frac{\Gamma\left(\frac{1}{\delta} - 1\right)}{\Gamma\left(\frac{1}{\delta} - 1\right) \Gamma\left[\frac{x + p\left(\frac{1}{\delta} - 1\right)}{\delta}\right] \Gamma\left(n - x + (1-p)\left(\frac{1}{\delta} - 1\right)\right)} \)

for \( 0 \leq p \leq 1; \ \max(-\frac{p}{n-p-1}, \frac{1-p}{n-2+p}) \leq \delta \leq 1; \ x = 0, 1, 2, \ldots, n \)

\[ E(x) = np \]
Chapter 2
Performing Univariate Analysis

Fit Distributions

\[ \text{Var}(x) = np(1-p)[1+(n-1)\delta] \]
where \( \Gamma(\cdot) \) is the Gamma function.

Remember that \( x|p \sim \text{Binomial}(p) \), while \( p \sim \text{Beta}(\alpha, \beta) \). The parameters \( p = \alpha/(\alpha+\beta) \) and \( \delta = 1/(\alpha+\beta+1) \) are estimated by the platform. To obtain estimates of \( \alpha \) and \( \beta \), use the following formulas:

\[
\hat{\alpha} = \hat{p}(1 - \hat{\delta}) \\
\hat{\beta} = (1 - \hat{p})(1 - \frac{1}{\hat{\delta}})
\]

If the estimate of \( \hat{\delta} \) is 0, the formulas do not work. In that case, the Beta Binomial has reduced to the Binomial(\(p\)), and \( \hat{p} \) is the estimate of \( p \).

The confidence intervals for the Beta Binomial parameters are profile likelihood intervals.

**Fit Distribution Options**

Each fitted distribution report has a red triangle menu that contains additional options.

**Table 2.18 Descriptions of Fitted Distribution Options**

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnostic Plot</td>
<td>Creates a quantile or a probability plot. See “Diagnostic Plot,” p. 73.</td>
</tr>
<tr>
<td>Density Curve</td>
<td>Uses the estimated parameters of the distribution to overlay a density curve</td>
</tr>
<tr>
<td>Goodness of Fit</td>
<td>Computes the goodness of fit test for the fitted distribution. See “</td>
</tr>
<tr>
<td>Fix Parameters</td>
<td>Enables you to enter parameters. The Fixed Parameters report reflects the</td>
</tr>
<tr>
<td>Quantiles</td>
<td>Returns the unscaled and uncentered quantiles for the specific upper and</td>
</tr>
<tr>
<td></td>
<td>lower probability values that you specify.</td>
</tr>
</tbody>
</table>
Table 2.18 Descriptions of Fitted Distribution Options *(Continued)*

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Set Spec Limits for K Sigma</strong></td>
<td>Use this option when you do not know the specification limits for a process and you want to use its distribution as a guideline for setting specification limits. Usually specification limits are derived using engineering considerations. If there are no engineering considerations, and if the data represents a trusted benchmark (well behaved process), then quantiles from a fitted distribution are often used to help set specification limits. Type a K value and select one-sided or two-sided for your capability analysis. Tail probabilities corresponding to K standard deviations are computed from the Normal distribution. The probabilities are converted to quantiles for the specific distribution that you have fitted. The resulting quantiles are used for specification limits in the capability analysis. This option is similar to the <em>Quantiles</em> option, but you provide K instead of probabilities. K corresponds to the number of standard deviations that the specification limits are away from the mean. For example, for a Normal distribution, where K=3, the 3 standard deviations below and above the mean correspond to the 0.00135th quantile and 0.99865th quantile, respectively. The lower specification limit is set at the 0.00135th quantile, and the upper specification limit is set at the 0.99865th quantile of the fitted distribution. A capability analysis is returned based on those specification limits.</td>
</tr>
<tr>
<td><strong>Spec Limits</strong></td>
<td>Computes generalizations of the standard capability indices, based on the specification limits and target you specify. See “Spec Limits,” p. 76.</td>
</tr>
<tr>
<td><strong>Save Fitted Quantiles</strong></td>
<td>Saves the fitted quantile values as a new column in the current data table. For details about how the quantiles are computed, see the Diagnostic Plot section on p. 74.</td>
</tr>
<tr>
<td><strong>Save Density Formula</strong></td>
<td>Creates a new column in the current data table that contains fitted values that have been computed by the density formula. The density formula uses the estimated parameter values.</td>
</tr>
<tr>
<td><strong>Save Spec Limits</strong></td>
<td>Saves the specification limits as a column property.</td>
</tr>
<tr>
<td><strong>Saved Transformed</strong></td>
<td>Creates a new column and saves a formula. The formula can transform the column to normality using the fitted distribution. This option is available only when one of the Johnson distributions or the Glog distribution is fit.</td>
</tr>
<tr>
<td><strong>Remove Fit</strong></td>
<td>Removes the distribution fit from the report window.</td>
</tr>
</tbody>
</table>
**Diagnostic Plot**

The **Diagnostic Plot** option creates a quantile or a probability plot. Depending on the fitted distribution, the plot is one of four formats.

**Table 2.19 Descriptions of Plot Formats**

<table>
<thead>
<tr>
<th>Plot Format</th>
<th>Applicable Distributions</th>
</tr>
</thead>
<tbody>
<tr>
<td>The fitted quantiles versus the data</td>
<td>• Weibull with threshold</td>
</tr>
<tr>
<td></td>
<td>• Gamma (see Figure 2.30)</td>
</tr>
<tr>
<td></td>
<td>• Beta</td>
</tr>
<tr>
<td></td>
<td>• Poisson</td>
</tr>
<tr>
<td></td>
<td>• GammaPoisson</td>
</tr>
<tr>
<td></td>
<td>• Binominal</td>
</tr>
<tr>
<td></td>
<td>• BetaBinomial</td>
</tr>
<tr>
<td>The fitted probability versus the data</td>
<td>• Normal (see Figure 2.30)</td>
</tr>
<tr>
<td></td>
<td>• Normal Mixtures</td>
</tr>
<tr>
<td></td>
<td>• Exponential</td>
</tr>
<tr>
<td>The fitted probability versus the data on log scale</td>
<td>• Weibull (see Figure 2.30)</td>
</tr>
<tr>
<td></td>
<td>• LogNormal</td>
</tr>
<tr>
<td></td>
<td>• Extreme Value</td>
</tr>
<tr>
<td>The fitted probability versus the standard normal quantile</td>
<td>• Johnson Sl</td>
</tr>
<tr>
<td></td>
<td>• Johnson Sh (see Figure 2.30)</td>
</tr>
<tr>
<td></td>
<td>• Johnson Su</td>
</tr>
<tr>
<td></td>
<td>• Glog</td>
</tr>
</tbody>
</table>

**Examples of Diagnostic Plots**

1. Open the Lipid Data.jmp sample data table.
2. Select **Analyze > Distribution**.
3. Select Cholesterol and click **Y, Columns**.
4. Click **OK**.

To create the Gamma Quantile plot in Figure 2.30

1. From the red triangle menu next to Cholesterol, select **Continuous Fit > All**.
2. From the red triangle menu next to Fitted Gamma, select **Diagnostic Plot**.
Performing Univariate Analysis

Fit Distributions

To create the Normal Probability plot in Figure 2.30:
1. From the red triangle menu next to Cholesterol, select Continuous Fit > Normal.
2. From the red triangle menu next to Fitted Normal, select Diagnostic Plot.

To create the Weibull 2 Parameter Probability plot in Figure 2.30
1. From the red triangle menu next to Cholesterol, select Continuous Fit > Weibull.
2. From the red triangle menu next to Fitted 2 parameter Weibull, select Diagnostic Plot.

To create the Johnson Sb Probability plot in Figure 2.30
1. From the red triangle menu next to Cholesterol, select Continuous Fit > Johnson Sb.
2. From the red triangle menu next to Fitted Johnson Sb, select Diagnostic Plot.

Figure 2.30 Examples of Diagnostic Plots

The fitted quantiles in the Diagnostic Plot and the fitted quantiles saved with the Save Fitted Quantiles command are formed using the following method:
1. The data are sorted and ranked. Ties are assigned different ranks.
2. Compute the percentile \( p_i = \frac{\text{rank}_i}{n+1} \).
3. Compute the quantile \( q_i = \text{Quantile}_d(p_i) \) where \( \text{Quantile}_d \) is the quantile function for the specific fitted distribution, and \( i = 1, 2, \ldots, n \).

Table 2.20 describes the options in the red triangle menu next to Diagnostic Plot.

**Table 2.20 Descriptions of Diagnostic Plot Options**

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rotate</td>
<td>Reverses the x- and y-axes.</td>
</tr>
<tr>
<td>Confidence Limits</td>
<td>Draws Lilliefors 95% confidence limits for the Normal Quantile plot, and 95% equal precision bands with ( a = 0.001 ) and ( b = 0.99 ) for all other quantile plots (Meeker and Escobar (1998)).</td>
</tr>
<tr>
<td>Line of Fit</td>
<td>Draws the straight diagonal reference line. If a variable fits the selected distribution, the values fall approximately on the reference line.</td>
</tr>
<tr>
<td>Median Reference Line</td>
<td>Draws a horizontal line at the median of the response.</td>
</tr>
</tbody>
</table>

**Goodness of Fit**

The Goodness of Fit option computes the goodness of fit test for the fitted distribution. The goodness of fit tests are not Chi-square tests, but are EDF (Empirical Distribution Function) tests. EDF tests offer advantages over the Chi-square tests, including improved power and invariance with respect to histogram midpoints.

- For Normal distributions, the Shapiro-Wilk test for normality is reported when the sample size is less than or equal to 2000, and the KSL test is computed for samples that are greater than 2000.
- For Poisson distributions that have sample sizes less than or equal to 30, the Goodness of Fit test is formed using two one-sided exact Kolmogorov tests combined to form a near exact test. For details, see Conover 1972. For sample sizes greater than 30, a Pearson Chi-squared goodness of fit test is performed.

**Table 2.21 Descriptions of JMP Goodness of Fit Tests**

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Parameters</th>
<th>Goodness of Fit Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal(^a)</td>
<td>( \mu ) and ( \sigma ) are unknown</td>
<td>Shapiro-Wilk (for ( n \leq 2000 )) Kolmogorov-Smirnov-Lillefors (for ( n &gt; 2000 ))</td>
</tr>
<tr>
<td></td>
<td>( \mu ) and ( \sigma ) are both known</td>
<td>Kolmogorov-Smirnov-Lillefors</td>
</tr>
<tr>
<td></td>
<td>either ( \mu ) or ( \sigma ) is known</td>
<td>(none)</td>
</tr>
<tr>
<td>LogNormal</td>
<td>( \mu ) and ( \sigma ) are known or unknown</td>
<td>Kolmogorov’s ( D )</td>
</tr>
</tbody>
</table>
Performing Univariate Analysis  

Fit Distributions  

Table 2.21 Descriptions of JMP Goodness of Fit Tests (Continued)

<table>
<thead>
<tr>
<th>Distribution</th>
<th>Parameters</th>
<th>Goodness of Fit Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weibull</td>
<td>$\alpha$ and $\beta$ known or unknown</td>
<td>Cramer von Mises $W^2$</td>
</tr>
<tr>
<td>Weibull with threshold</td>
<td>$\alpha$, $\beta$ and $\theta$ known or unknown</td>
<td>Cramer von Mises $W^2$</td>
</tr>
<tr>
<td>Extreme Value</td>
<td>$\alpha$ and $\beta$ known or unknown</td>
<td>Cramer von Mises $W^2$</td>
</tr>
<tr>
<td>Exponential</td>
<td>$\sigma$ is known or unknown</td>
<td>Kolmogorov’s $D$</td>
</tr>
<tr>
<td>Gamma</td>
<td>$\alpha$ and $\sigma$ are known</td>
<td>Cramer von Mises $W^2$</td>
</tr>
<tr>
<td>Beta</td>
<td>$\alpha$ and $\beta$ are known</td>
<td>Kolmogorov’s $D$</td>
</tr>
<tr>
<td>Binomial</td>
<td>$\rho$ is known or unknown and $n$ is known</td>
<td>Pearson $\chi^2$</td>
</tr>
<tr>
<td>Beta Binomial</td>
<td>$\rho$ and $\delta$ known or unknown</td>
<td>Kolmogorov’s $D$ (for $n \leq 30$) or Pearson $\chi^2$ (for $n &gt; 30$)</td>
</tr>
<tr>
<td>Poisson</td>
<td>$\lambda$ known or unknown</td>
<td>Kolmogorov’s $D$ (for $n \leq 30$) or Pearson $\chi^2$ (for $n &gt; 30$)</td>
</tr>
<tr>
<td>Gamma Poisson</td>
<td>$\lambda$ or $\sigma$ known or unknown</td>
<td>Kolmogorov’s $D$ (for $n \leq 30$) or Pearson $\chi^2$ (for $n &gt; 30$)</td>
</tr>
</tbody>
</table>

- For the three Johnson distributions and the Glog distribution, the data are transformed to Normal, then the appropriate test of normality is performed.

Spec Limits

The **Spec Limits** option launches a window requesting specification limits and target, and then computes generalizations of the standard capability indices. This is done using the fact that for the normal distribution, $3\sigma$ is both the distance from the lower 0.135 percentile to median (or mean) and the distance from the median (or mean) to the upper 99.865 percentile. These percentiles are estimated from the fitted distribution, and the appropriate percentile-to-median distances are substituted for $3\sigma$ in the standard formulas.

Writing $T$ for the target, $LSL$, and $USL$ for the lower and upper specification limits, and $P_\alpha$ for the 100th percentile, the generalized capability indices are as follows:

$$C_{pl} = \frac{P_{0.5} - LSL}{P_{0.5} - P_{0.00135}}$$
Statistical Details

This section contains the following information:

- “Statistical Details for Quantiles,” p. 77
- “Statistical Details for Prediction Intervals,” p. 78
- “Statistical Details for Tolerance Intervals,” p. 78
- “Statistical Details for Capability Analysis,” p. 79

Statistical Details for Quantiles

This section describes how quantiles are computed.

To compute the $p$th quantile of $N$ nonmissing values in a column, arrange the $N$ values in ascending order and call these column values $y_1, y_2, \ldots, y_N$. Compute the rank number for the $p$th quantile as $p / 100(N+1)$.

- If the result is an integer, the $p$th quantile is that rank’s corresponding value.
- If the result is not an integer, the $p$th quantile is found by interpolation. The $p$th quantile, denoted $q_p$, is computed as follows:

\[ q_p = (1 - f) y_i + f y_{i+1} \]

where:

- $n$ is the number of nonmissing values for a variable

\[
C_{pu} = \frac{USL - P_{0.5}}{P_{0.99865} - P_{0.5}}
\]

\[
C_p = \frac{USL - LSL}{P_{0.99865} - P_{0.00135}}
\]

\[
C_{pk} = \min\left(\frac{P_{0.5} - LSL}{P_{0.5} - P_{0.00135}}, \frac{USL - P_{0.5}}{P_{0.99865} - P_{0.5}}\right)
\]

\[
K = 2 \times \left[ \frac{1}{2} \frac{USL + LSL - P_{0.5}}{USL - LSL} \right]
\]

\[
C_{pm} = \frac{\min\left(\frac{T - LSL}{P_{0.5} - P_{0.00135}}, \frac{USL - T}{P_{0.99865} - P_{0.5}}\right)}{\sqrt{1 + \left(\frac{\mu - T}{\sigma}\right)^2}}
\]

If the data are normally distributed, these formulas reduce to the formulas for standard capability indices. See Table 2.22.
Performing Univariate Analysis

Chapter 2

Statistical Details

- \( y_1, y_2, ..., y_n \) represents the ordered values of the variable
- \( y_{n+1} \) is taken to be \( y_n \)
- \( i \) is the integer part and \( f \) is the fractional part of \( (n+1)p \).
- \( (n + 1)p = i + f \)

For example, suppose a data table has 15 rows and you want to find the 75th and 90th quantile values of a continuous column. After the column is arranged in ascending order, the ranks that contain these quantiles are computed as follows:

\[
\frac{75}{100} (15 + 1) = 12 \quad \text{and} \quad \frac{90}{100} (15 + 1) = 14.4
\]

The value \( y_{12} \) is the 75th quantile. The 90th quantile is interpolated by computing a weighted average of the 14th and 15th ranked values as \( y_{90} = 0.6y_{14} + 0.4y_{15} \).

Statistical Details for Prediction Intervals

The formulas that JMP uses for computing prediction intervals are as follows:

- For \( m \) future observations:
  \[
  [\hat{y}_m, \hat{y}_m] = \bar{x} \pm \frac{1}{\sqrt{n}} t_{(1 - \alpha/2, n-1)} \sqrt{s} \times \hat{f} \text{ for } m \geq 1
  \]

- For the mean of \( m \) future observations:
  \[
  [\hat{Y}_m, \hat{Y}_m] = \bar{X} \pm t_{(1 - \alpha/2, n-1)} \frac{1}{\sqrt{n}} \sqrt{s} \times \hat{f} \text{ for } m \geq 1.
  \]

- For the standard deviation of \( m \) future observations:
  \[
  [\hat{s}_m, \hat{s}_m] = \left[ t_{(1 - \alpha/2, (m-1, n-1))} \times \sqrt{1 + \frac{1}{n} \times \hat{f}}, t_{(1 - \alpha/2, (m-1, n-1))} \times \sqrt{1 + \frac{1}{n} \times \hat{f}} \right] \text{ for } m \geq 2
  \]

where \( m \) = number of future observations, and \( n \) = number of points in current analysis sample.

- The one-sided intervals are formed by using \( 1 - \alpha \) in the quantile functions.

For references, see Hahn and Meeker (1991), pages 61-64.

Statistical Details for Tolerance Intervals

One-Sided Interval

The one-sided interval is computed as follows:

Upper Limit = \( \bar{x} + gs \)

Lower Limit = \( \bar{x} - gs \)

where
Performing Univariate Analysis

Statistical Details

\[ g' = t(1 - \alpha, n - 1, \Phi^{-1}(p) \cdot \sqrt{n}) / \sqrt{n} \]
from Table 1 of Odeh and Owen (1980).

\( t \) is the quantile from the non-central t-distribution, and \( \Phi^{-1} \) is the standard normal quantile.

Two-Sided Interval

The two-sided interval is computed as follows:

\[
\left[ T_{p_L}, T_{p_U} \right] = \left[ \bar{x} - g(1 - \alpha/2; \bar{x}, n)^2, \bar{x} + g(1 - \alpha/2; \bar{x}, n)^2 \right]
\]

where

\( s = \text{standard deviation} \) and \( g(1 - \alpha/2; \bar{x}, n) \) is a constant that can be found in Table 4 of Odeh and Owen (1980).

To determine \( g \), consider the fraction of the population captured by the tolerance interval. Tamhane and Dunlop (2000) give this fraction as follows:

\[
\Phi\left( \frac{\bar{x} + g - \mu}{\sigma} \right) - \Phi\left( \frac{\bar{x} - g - \mu}{\sigma} \right)
\]

where \( \Phi \) denotes the standard normal c.d.f. (cumulative distribution function). Therefore \( g \) solves the following equation:

\[
P\left\{ \Phi\left( \frac{\bar{x} + g - \mu}{\sigma} \right) - \Phi\left( \frac{\bar{x} - g - \mu}{\sigma} \right) \geq 1 - \gamma \right\} = 1 - \alpha
\]

where \( 1 - \gamma \) is the fraction of all future observations contained in the tolerance interval.


Statistical Details for Capability Analysis

All capability analyses use the same formulas. Options differ in how \( \sigma \) is computed:

- **Long-term** uses the overall \( \sigma \). This option is used for \( P_{pk} \) statistics, and computes as follows:

\[
\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n - 1}}
\]

**Note:** There is a preference for Distribution called **Ppk Capability Labeling** that labels the long-term capability output with \( P_{pk} \) labels. This option is found using **File > Preferences**, then select **Platforms > Distribution**.

- **Specified Sigma** enables you to type a specific, known \( \sigma \) used for computing capability analyses. Sigma is user-specified, and is therefore not computed.
• **Moving Range** enables you to enter a range span, which computes sigma as follows:

\[ \sigma = \frac{\bar{R}}{d_2(n)} \]

where

\[ \bar{R} \] is the average of the moving ranges

\[ d_2(n) \] is the expected value of the range of \( n \) independent normally distributed variables with unit standard deviation.

• **Short Term Sigma, Group by Fixed Subgroup Size** if \( r \) is the number of subgroups and each \( i \)th subgroup is defined by the order of the data, sigma is computed as follows:

\[ \sigma = \sqrt{\frac{1}{n-r} \sum_{i=1}^{n} (x_{ij} - \bar{x}_i)^2} \]

This formula is commonly referred to as the Root Mean Square Error, or RMSE.

Table 2.22 describes capability indices and gives computational formulas.

### Table 2.22 Descriptions of Capability Index Names and Formulas

<table>
<thead>
<tr>
<th>Index</th>
<th>Index Name</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>CP</td>
<td>process capability</td>
<td>(USL - LSL)/6( \sigma ) where:</td>
</tr>
<tr>
<td></td>
<td>ratio, ( C_p )</td>
<td>• USL is the upper spec limit</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• LSL is the lower spec limit</td>
</tr>
<tr>
<td>CIs for CP</td>
<td>Lower CI on CP</td>
<td>( CP \sqrt{n} \frac{\chi^2_{\alpha/2, n-1}}{n-1} )</td>
</tr>
<tr>
<td>CIs for CP</td>
<td>Upper CI on CP</td>
<td>( CP \sqrt{n} \frac{\chi^2_{1-\alpha/2, n-1}}{n-1} )</td>
</tr>
<tr>
<td>CPK (PPK for</td>
<td></td>
<td>min(CPL, CPU)</td>
</tr>
<tr>
<td>AIAG)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIs for CPK</td>
<td>Lower CI</td>
<td>( \hat{C}<em>{pk} \left[ 1 - \Phi^{-1}(1 - \alpha/2) \left{ \frac{1}{9 n C</em>{pk}^2} + \frac{1}{2(n-1)} \right} \right] )</td>
</tr>
<tr>
<td>See Bissell (1990)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIs for CPK</td>
<td>Upper CI</td>
<td>( \hat{C}<em>{pk} \left[ 1 + \Phi^{-1}(1 - \alpha/2) \left{ \frac{1}{9 n C</em>{pk}^2} + \frac{1}{2(n-1)} \right} \right] )</td>
</tr>
</tbody>
</table>
A capability index of 1.33 is considered to be the minimum acceptable. For a normal distribution, this gives an expected number of nonconforming units of about 6 per 100,000.

Exact 100(1 - \(\alpha\))% lower and upper confidence limits for CPL are computed using a generalization of the method of Chou et al. (1990), who point out that the 100(1 - \(\alpha\)) lower confidence limit for CPL (denoted by CPL\(_{\text{LCL}}\)) satisfies the following equation:

\[
\Pr\{T_{n-1}(\delta = 3\sqrt{n}) \leq \text{CPL\(_{\text{LCL}}\}) \} = 1 - \alpha
\]

where \(T_{n-1}(\delta)\) has a non-central \(t\)-distribution with \(n - 1\) degrees of freedom and noncentrality parameter \(\delta\).

A capability index of 1.33 is considered to be the minimum acceptable. For a normal distribution, this gives an expected number of nonconforming units of about 6 per 100,000.

Exact 100(1 - \(\alpha\))% lower and upper confidence limits for CPL are computed using a generalization of the method of Chou et al. (1990), who point out that the 100(1 - \(\alpha\)) lower confidence limit for CPL (denoted by CPL\(_{\text{LCL}}\)) satisfies the following equation:

\[
\Pr\{T_{n-1}(\delta = 3\sqrt{n}) \leq \text{CPL\(_{\text{LCL}}\}) \} = 1 - \alpha
\]

where \(T_{n-1}(\delta)\) has a non-central \(t\)-distribution with \(n - 1\) degrees of freedom and noncentrality parameter \(\delta\).

- A capability index of 1.33 is considered to be the minimum acceptable. For a normal distribution, this gives an expected number of nonconforming units of about 6 per 100,000.

Table 2.22 Descriptions of Capability Index Names and Formulas (Continued)

<table>
<thead>
<tr>
<th>Index</th>
<th>Index Name</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPM</td>
<td>process capability index, (C_{\text{pm}})</td>
<td>(\frac{\min(\text{target} - \text{LSL}, \text{USL} - \text{target})}{3\sqrt{\frac{\hat{x}^2}{s^2} + \frac{(\text{mean} - \text{target})^2}{s^2}}})</td>
</tr>
<tr>
<td>CIs for CPM</td>
<td>Lower CI on CPM</td>
<td>(C_{\text{pm}}\frac{\chi^2_{\alpha/2, \gamma}}{\gamma}), where (\gamma = n\left(1 + \frac{(\bar{x} - \text{Target})^2}{s^2}\right)^2)</td>
</tr>
<tr>
<td></td>
<td>Upper CI on CPM</td>
<td>(C_{\text{pm}}\frac{\chi^2_{1 - \alpha/2, \gamma}}{\gamma}) where (\gamma = \text{same as above.})</td>
</tr>
<tr>
<td>CPL</td>
<td>process capability ratio of one-sided lower spec</td>
<td>((\text{mean} - \text{LSL})/3s)</td>
</tr>
<tr>
<td>CPU</td>
<td>process capability ratio of one-sided upper spec</td>
<td>((\text{USL} - \text{mean})/3s)</td>
</tr>
</tbody>
</table>
Note: Because of a lack of supporting research at the time of this writing, computing confidence intervals for capability indices is not recommended, except for cases when the capability indices are based on the standard deviation.
Introduction to the Fit Y by X Platform
Performing Four Types of Analyses

The Fit Y by X platform analyzes the pair of X and Y variables that you specify, by context, based on modeling type. See Figure 3.1.

Here are the four types of analyses:

- Bivariate fitting
- One-way analysis of variance
- Logistic regression
- Contingency table analysis

Figure 3.1 Examples of Four Types of Analyses
Contents

Overview of the Fit Y by X Platform ......................................................... 85
Launch the Fit Y by X Platform ................................................................. 85
  Launch Specific Analyses from the JMP Starter Window ...................... 86
Overview of the Fit Y by X Platform

As shown in Table 3.1, this platform is a collection of four specific platforms (or types of analyses).

**Table 3.1 Descriptions of the Fit Y by X Specific Platforms**

<table>
<thead>
<tr>
<th>Specific Platform</th>
<th>Modeling Types</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bivariate</td>
<td>Continuous Y by continuous X</td>
<td>Analyzes the relationship between two continuous variables</td>
<td>See Performing Bivariate Analysis.</td>
</tr>
<tr>
<td>Oneway</td>
<td>Continuous Y by nominal or ordinal X</td>
<td>Analyzes how the distribution of a continuous Y variable differs across groups defined by a categorical X variable</td>
<td>See Performing Oneway Analysis.</td>
</tr>
<tr>
<td>Logistic</td>
<td>Nominal or ordinal Y by continuous X</td>
<td>Fits the probabilities for response categories to a continuous X predictor</td>
<td>See Performing Simple Logistic Regression.</td>
</tr>
<tr>
<td>Contingency</td>
<td>Nominal or ordinal Y by nominal or ordinal X</td>
<td>Analyzes the distribution of a categorical response variable Y as conditioned by the values of a categorical X factor</td>
<td>See Performing Contingency Analysis.</td>
</tr>
</tbody>
</table>

Launch the Fit Y by X Platform

Launch the Fit Y by X platform by selecting **Analyze > Fit Y by X**.
Introduction to the Fit Y by X Platform

Overview of the Fit Y by X Platform

Chapter 3

Overview of the Fit Y by X Platform

For more information, see “Launch Window Features,” p. 13 in the “Preliminaries” chapter.

Launch Specific Analyses from the JMP Starter Window

From the JMP Starter window, you can launch a specific analysis (Bivariate, Oneway, Logistic, or Contingency). If you select this option, specify the correct modeling types for the analysis. See Table 3.3.

To launch a specific analysis from the JMP Starter Window, proceed as follows:

1. Click on the Basic category.
2. Click on the specific analysis that you want to launch.

Most of the platform launch options are the same, but the naming for some of the Y and X platform buttons is tailored for the specific analysis that you are performing.

Table 3.2 Description of the Fit Y by X Launch Window

| Bivariate, Oneway, Logistic, Contingency | This grid shows which analysis results from the different combinations of data types. Once you have assigned your columns, the applicable platform appears as a label above the grid. |
| Block | (Optional, for Oneway and Contingency only) Identifies a second factor, which forms a two-way analysis without interaction. If you specify a block factor, the data should be balanced and have equal counts in each block by group cell. |
| Y, Response | See “Cast Selected Columns into Roles Buttons,” p. 13 in the “Preliminaries” chapter. |
| X, Factor | |
| Weight | |
| Freq | |
| By | |

Table 3.3 Descriptions of the Y and X Platform Buttons

<table>
<thead>
<tr>
<th>Platform or Analysis</th>
<th>Y Button</th>
<th>X Button</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fit Y by X</td>
<td>Y, Response</td>
<td>X, Factor</td>
</tr>
<tr>
<td>Bivariate</td>
<td>X, Regressor</td>
<td></td>
</tr>
<tr>
<td>Oneway</td>
<td>X, Grouping</td>
<td></td>
</tr>
<tr>
<td>Logistic</td>
<td>Y, Categorical Response</td>
<td>X, Continuous Regressor</td>
</tr>
<tr>
<td>Contingency</td>
<td>Y, Response Category</td>
<td>X, Grouping Category</td>
</tr>
</tbody>
</table>
Performing Bivariate Analysis
Using the Fit Y by X or Bivariate Platform

The Bivariate platform is the continuous by continuous personality of the Fit Y by X platform. Bivariate analysis shows the relationship between two continuous variables. The analysis results appear in a scatterplot. You can interactively add other types of fits, such as simple linear regression, polynomial regression, and so on.

Figure 4.1 Example of Bivariate Analysis
## Contents

- Example of Bivariate Analysis ................................................................. 89
- Launch the Bivariate Platform ............................................................... 89
- Example of a Bivariate Report ............................................................... 90
- Overview of Fitting Commands and General Options .............................. 91
  - Fitting Command Categories ............................................................... 92
  - Fit the Same Command Multiple Times ................................................ 93
- Fit Mean ........................................................................................................ 93
- Fit Line and Fit Polynomial ......................................................................... 95
- Fit Special ..................................................................................................... 103
- Fit Spline ....................................................................................................... 106
- Fit Each Value .............................................................................................. 108
- Fit Orthogonal .............................................................................................. 109
- Density Ellipse ............................................................................................. 112
- Nonpar Density ............................................................................................. 114
- Histogram Borders ....................................................................................... 116
- Group By ....................................................................................................... 116
  - Example of Group By Using Density Ellipses ......................................... 116
  - Example of Group By Using Regression Lines ......................................... 117
- Fitting Menus ............................................................................................... 118
- Fitting Menu Options ................................................................................... 119
- Statistical Details .......................................................................................... 122
Example of Bivariate Analysis

This example uses the SAT.jmp sample data table. SAT test scores for students in the 50 U.S. states, plus the District of Columbia, are divided into two areas: verbal and math. You want to find out how the percentage of students taking the SAT tests is related to verbal test scores for 2004.

1. Open the SAT.jmp sample data table.
2. Select Analyze > Fit Y by X.
4. Select % Taking (2004) and click X, Factor.
5. Click OK.

Figure 4.2 Example of SAT Scores by Percent Taking

You can see that the verbal scores were higher when a smaller percentage of the population took the test.

Launch the Bivariate Platform

You can perform a bivariate analysis using either the Fit Y by X platform or the Bivariate platform. The two approaches give equivalent results.

- To launch the Fit Y by X platform, select Analyze > Fit Y by X. The Fit Y by X launch window appears. See Figure 3.2 in the “Introduction to the Fit Y by X Platform” chapter.
- To launch the Bivariate platform, from the JMP Starter window, click on the Basic category and click Bivariate. The Bivariate launch window appears.
Performing Bivariate Analysis

Example of a Bivariate Report

Figure 4.3 The Bivariate Launch Window

The distribution of the response as $X$ varies continuously.

Select Columns
- State
- % Taking (2004)
- 2002 Verbal
- 2002 Math
- 2003 Verbal
- 2003 Math
- 2004 Verbal
- 2004 Math
- 2001 Verbal
- 2001 Math
- 1999 Verbal
- 1999 Math
- 1998 Verbal
- 1998 Math
- 1997 Verbal

Cast Selected Columns into Roles
- Y, Response
  - % Taking (2004)
- X, Regressor
  - Required numeric
- Weight
  - Optional numeric
- Freq
  - Optional numeric
- By
  - Optional numeric

For details about this window, see the “Introduction to the Fit Y by X Platform” chapter.

Example of a Bivariate Report

To produce the plot shown in Figure 4.4, follow the instructions in “Example of Bivariate Analysis,” p. 89.

Figure 4.4 The Bivariate Report Window

The Bivariate report begins with a scatterplot for each pair of $X$ and $Y$ variables. You can interact with this scatterplot just as you can with other JMP plots (for example, resizing the plot, highlighting points with the arrow or brush tool, and labeling points). For details about these features, see the Using JMP book.

You can fit curves on the scatterplot and view statistical reports and additional menus using the fitting commands that are located within the red triangle menu. See “Overview of Fitting Commands and General Options,” p. 91.
Overview of Fitting Commands and General Options

Here are the options in the Bivariate Fit red triangle menu.

Figure 4.5 Fitting Commands and General Options

Table 4.1 Descriptions of General Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show Points</td>
<td>Hides or shows the points in the scatterplot. A check mark indicates that points are shown.</td>
</tr>
<tr>
<td>Histogram Borders</td>
<td>Attaches histograms to the x- and y-axes of the scatterplot. A check mark indicates that histogram borders are turned on. See “Histogram Borders,” p. 116.</td>
</tr>
<tr>
<td>Group By</td>
<td>Lets you select a classification (or grouping) variable. A separate analysis is computed for each level of the grouping variable, and regression curves or ellipses are overlaid on the scatterplot. See “Group By,” p. 116.</td>
</tr>
<tr>
<td>Script</td>
<td>Contains options that are available to all platforms. These options enable you to redo the analysis or save the JSL commands for the analysis to a window or a file. See “Script Menus,” p. 15 in the “Preliminaries” chapter.</td>
</tr>
</tbody>
</table>

Table 4.2 describes the fitting commands. Each fitting command adds the following:

- a curve or distribution to the scatterplot
- a red triangle menu to the report window
- a specific report to the report window

To see an example of the elements that are added by a fitting command, see Figure 4.6.
Performing Bivariate Analysis
Overview of Fitting Commands and General Options

Table 4.2 Descriptions of Fitting Commands

<table>
<thead>
<tr>
<th>Fitting Command</th>
<th>Description</th>
<th>See</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fit Mean</td>
<td>Adds a horizontal line to the scatterplot that represents the mean of the Y response variable.</td>
<td>“Fit Mean,” p. 93.</td>
</tr>
<tr>
<td>Fit Line</td>
<td>Adds straight line fits to your scatterplot using least squares regression.</td>
<td>“Fit Line and Fit Polynomial,” p. 95.</td>
</tr>
<tr>
<td>Fit Polynomial</td>
<td>Fits polynomial curves of a certain degree using least squares regression.</td>
<td></td>
</tr>
<tr>
<td>Fit Special</td>
<td>Transforms $Y$ and $X$. Transformations include: log, square root, square, reciprocal, and exponential. You can also turn off center polynomials, constrain the intercept and the slope, and fit polynomial models.</td>
<td>“Fit Special,” p. 103.</td>
</tr>
<tr>
<td>Fit Spline</td>
<td>Fits a smoothing spline that varies in smoothness (or flexibility) according to the lambda ($\lambda$) value. The $\lambda$ value is a tuning parameter in the spline formula.</td>
<td>“Fit Spline,” p. 106.</td>
</tr>
<tr>
<td>Fit Each Value</td>
<td>Fits a value to each unique $X$ value and compares it to other fitted lines, showing the concept of lack of fit.</td>
<td>“Fit Each Value,” p. 108.</td>
</tr>
<tr>
<td>Fit Orthogonal</td>
<td>Fits lines that adjust for variability in $X$ as well as $Y$.</td>
<td>“Fit Orthogonal,” p. 109.</td>
</tr>
<tr>
<td>Nonpar Density</td>
<td>Shows patterns in the point density, which is useful when the scatterplot is so darkened by points that it is difficult to distinguish patterns.</td>
<td>“Nonpar Density,” p. 114.</td>
</tr>
</tbody>
</table>

**Note:** You can remove a fit using the **Remove Fit** command. For details, see “Fitting Menu Options,” p. 119.

**Fitting Command Categories**

Fitting command categories include regression fits and density estimation.
Performing Bivariate Analysis

Chapter 4

Fit Mean

You can select the same fitting command multiple times, and each new fit is overlaid on the scatterplot. You can try fits, exclude points and refit, and you can compare them on the same scatterplot.

To apply a fitting command to multiple analyses in your report window, hold down the CTRL key and select a fitting option.

Fit Mean

Using the **Fit Mean** command, you can add a horizontal line to the scatterplot that represents the mean of the \( Y \) response variable. You can start by fitting the mean and then use the mean line as a reference for other fits (such as straight lines, confidence curves, polynomial curves, and so on).

---

**Table 4.3** Descriptions of Fitting Commands by Category

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Fitting Commands</th>
</tr>
</thead>
<tbody>
<tr>
<td>Regression Fits</td>
<td>Regression methods fit a curve through the points. The curve is an equation (a model) that is estimated using least squares, which minimizes the sum of squared differences from each point to the line (or curve). Regression fits assume that the ( Y ) variable is distributed as a random scatter above and below a line of fit.</td>
<td>Fit Mean, Fit Line, Fit Polynomial, Fit Special, Fit Spline, Fit Each Value, Fit Orthogonal</td>
</tr>
<tr>
<td>Density Estimation</td>
<td>Density estimation fits a bivariate distribution to the points. You can either select a bivariate normal density, characterized by elliptical contours, or a general nonparametric density.</td>
<td>Fit Density Ellipse, Nonpar Density</td>
</tr>
</tbody>
</table>
Performing Bivariate Analysis

Chapter 4

Fit Mean

Figure 4.6 Example of Fit Mean

![Bivariate Fit of 2004 Verbal By % Taking (2004)]

Fit Mean line

Fit Mean menu

Fit Mean report

<table>
<thead>
<tr>
<th>Fit Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean 535.5922</td>
</tr>
<tr>
<td>Std Dev [RMSE] 33.84617</td>
</tr>
<tr>
<td>Std Error 4.790412</td>
</tr>
<tr>
<td>SSE 57.278.19</td>
</tr>
</tbody>
</table>

Fit Mean Menu

For details about the Fit Mean menu options, see “Fitting Menu Options,” p. 119.

Fit Mean Report

The following table describes the Fit Mean Report options.

Table 4.4 Description of Fit Mean Report

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>Mean of the response variable. The predicted response when there are no specified effects in the model.</td>
</tr>
<tr>
<td>Std Dev [RMSE]</td>
<td>Standard deviation of the response variable. Square root of the mean square error, also called the root mean square error (or RMSE).</td>
</tr>
<tr>
<td>Std Error</td>
<td>Standard deviation of the response mean. Calculated by dividing the RMSE by the square root of the number of values.</td>
</tr>
<tr>
<td>SSE</td>
<td>Error sum of squares for the simple mean model. Appears as the sum of squares for Error in the analysis of variance tables for each model fit.</td>
</tr>
</tbody>
</table>
Fit Line and Fit Polynomial

Using the Fit Line command, you can add straight line fits to your scatterplot using least squares regression. Using the Fit Polynomial command, you can fit polynomial curves of a certain degree using least squares regression.

Figure 4.7 Example of Fit Line and Fit Polynomial

Figure 4.7 shows an example that compares a linear fit to the mean line and to a degree 2 polynomial fit. Note the following information:

- The Fit Line output is equivalent to a polynomial fit of degree 1.
- The Fit Mean output is equivalent to a polynomial fit of degree 0.

Note: For statistical details about Fit Line, see “Statistical Details,” p. 122.

Linear Fit and Polynomial Fit Menus

For details about the Linear Fit and Polynomial Fit Degree=X menu options, see “Fitting Menu Options,” p. 119.
Linear Fit and Polynomial Fit Reports

The Linear Fit and Polynomial Fit reports begin with the equation of fit.

**Summary of Fit Report**

The Summary of Fit reports in Figure 4.9 show the numeric summaries of the response for the linear fit and polynomial fit of degree 2 for the same data. You can compare multiple Summary of Fit reports to see the improvement of one model over another, indicated by a larger Rsquare value and smaller Root Mean Square Error.

<table>
<thead>
<tr>
<th>Linear Fit</th>
<th>Polynomial Fit Degree=2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rsquare</td>
<td>Rsquare</td>
</tr>
<tr>
<td>0.798274</td>
<td>0.879413</td>
</tr>
<tr>
<td>R_square</td>
<td>0.795994</td>
</tr>
<tr>
<td>0.874389</td>
<td></td>
</tr>
<tr>
<td>Root Mean Square Error</td>
<td>Root Mean Square Error</td>
</tr>
<tr>
<td>15.85752</td>
<td>11.99688</td>
</tr>
<tr>
<td>Mean of Response</td>
<td>Mean of Response</td>
</tr>
<tr>
<td>535.3922</td>
<td>535.3622</td>
</tr>
<tr>
<td>Observations (or Sum Wt)</td>
<td>Observations (or Sum Wt)</td>
</tr>
<tr>
<td>51</td>
<td>51</td>
</tr>
</tbody>
</table>
Performing Bivariate Analysis

Chapter 4

Table 4.5 Description of Summary of Fit Report

<table>
<thead>
<tr>
<th>Description</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>RSquare</strong></td>
<td>Measures the proportion of the variation explained by the model. The remaining variation is not explained by the model and attributed to random error.</td>
</tr>
<tr>
<td></td>
<td>• The $R^2$ is 1 if the model fits perfectly.</td>
</tr>
<tr>
<td></td>
<td>Using quantities from the corresponding analysis of variance table, the $R^2$ for any continuous response fit is calculated as follows:</td>
</tr>
<tr>
<td></td>
<td>$\frac{\text{Sum of Squares for Model}}{\text{Sum of Squares for C. Total}}$</td>
</tr>
<tr>
<td></td>
<td>The $R^2$ values in Figure 4.9 indicate that the polynomial fit of degree 2 gives a small improvement over the linear fit.</td>
</tr>
<tr>
<td><strong>RSquare Adj</strong></td>
<td>Adjusts the $R^2$ value to make it more comparable over models with different numbers of parameters by using the degrees of freedom in its computation. It is a ratio of mean squares instead of sums of squares and is calculated as follows:</td>
</tr>
<tr>
<td></td>
<td>$1 - \frac{\text{Mean Square for Error}}{\text{Mean Square for C. Total}}$</td>
</tr>
<tr>
<td></td>
<td>The mean square for Error is in the Analysis of Variance report. See Figure 4.11. You can compute the mean square for C. Total as the Sum of Squares for C. Total divided by its respective degrees of freedom.</td>
</tr>
<tr>
<td><strong>Root Mean Square Error</strong></td>
<td>Estimates the standard deviation of the random error. It is the square root of the mean square for Error in the Analysis of Variance report. See Figure 4.11.</td>
</tr>
<tr>
<td><strong>Mean of Response</strong></td>
<td>Provides the sample mean (arithmetic average) of the response variable. This is the predicted response when no model effects are specified.</td>
</tr>
<tr>
<td><strong>Observations</strong></td>
<td>Provides the number of observations used to estimate the fit. If there is a weight variable, this is the sum of the weights.</td>
</tr>
</tbody>
</table>

**Lack of Fit Report**

**Note:** The Lack of Fit report appears only if there are multiple rows that have the same $x$ value.

Using the Lack of Fit report, you can estimate the error, regardless of whether you have the right form of the model. This occurs when multiple observations occur at the same $x$ value. The error that you measure for these exact replicates is called *pure error*. This is the portion of the sample error that cannot be explained or
predicted no matter what form of model is used. However, a lack of fit test might not be of much use if it has only a few degrees of freedom for it (few replicated $x$ values).

The difference between the residual error from the model and the pure error is called the **lack of fit error**. The lack of fit error can be significantly greater than the pure error if you have the wrong functional form of the regressor. In that case, you should try a different type of model fit. The Lack of Fit report tests whether the lack of fit error is zero.

### Table 4.6 Description of Lack of Fit Report

<table>
<thead>
<tr>
<th>Source</th>
<th>The three sources of variation: Lack of Fit, Pure Error, and Total Error.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DF</td>
<td>The degrees of freedom (DF) for each source of error.</td>
</tr>
<tr>
<td></td>
<td>• The <strong>Total Error</strong> DF is the degrees of freedom found on the Error line of the Analysis of Variance table (shown under the “Analysis of Variance Report,” p. 100). It is the difference between the <strong>Total</strong> DF and the <strong>Model</strong> DF found in that table. The <strong>Error</strong> DF is partitioned into degrees of freedom for lack of fit and for pure error.</td>
</tr>
<tr>
<td></td>
<td>• The <strong>Pure Error</strong> DF is pooled from each group where there are multiple rows with the same values for each effect. For example, there are multiple instances in the Big Class.jmp sample data table where more than one subject has the same value of height. In general, if there are $g$ groups having multiple rows with identical values for each effect, the pooled DF, denoted $DF_p$, is as follows: $DF_p = \sum_{i=1}^{g}(n_i-1)$ where $n_i$ is the number of subjects in the $i$th group.</td>
</tr>
<tr>
<td></td>
<td>• The <strong>Lack of Fit</strong> DF is the difference between the <strong>Total Error</strong> and <strong>Pure Error</strong> DF.</td>
</tr>
</tbody>
</table>
Chapter 4
Performing Bivariate Analysis
Fit Line and Fit Polynomial

Table 4.6 Description of Lack of Fit Report (Continued)

<table>
<thead>
<tr>
<th>Sum of Squares</th>
<th>The sum of squares (SS for short) for each source of error.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• The Total Error SS is the sum of squares found on the Error line of the corresponding Analysis of Variance table, shown under “Analysis of Variance Report,” p. 100.</td>
</tr>
<tr>
<td></td>
<td>• The Pure Error SS is pooled from each group where there are multiple rows with the same value for the x variable. This estimates the portion of the true random error that is not explained by model x effect. In general, if there are $g$ groups having multiple rows with the same x value, the pooled SS, denoted $SS_p$, is written as follows:</td>
</tr>
<tr>
<td></td>
<td>$SS_p = \sum_{i=1}^{g} SS_i$</td>
</tr>
<tr>
<td></td>
<td>where $SS_i$ is the sum of squares for the $i$th group corrected for its mean.</td>
</tr>
<tr>
<td></td>
<td>• The Lack of Fit SS is the difference between the Total Error and Pure Error sum of squares. If the lack of fit SS is large, the model might not be appropriate for the data. The $F$-ratio described below tests whether the variation due to lack of fit is small enough to be accepted as a negligible portion of the pure error.</td>
</tr>
</tbody>
</table>

| Mean Square | The sum of squares divided by its associated degrees of freedom. This computation converts the sum of squares to an average (mean square). $F$-ratios for statistical tests are the ratios of mean squares. |

| F Ratio | The ratio of mean square for lack of fit to mean square for Pure Error. It tests the hypothesis that the lack of fit error is zero. |

| Prob > F | The probability of obtaining a greater $F$-value by chance alone if the variation due to lack of fit variance and the pure error variance are the same. A high $p$ value means that there is not a significant lack of fit. |

| Max RSq | The maximum $R^2$ that can be achieved by a model using only the variables in the model. Because Pure Error is invariant to the form of the model and is the minimum possible variance, Max RSq is calculated as follows: |

$$Max \, RSq = 1 - \frac{SS(Pure \, error)}{SS(Total \, for \, whole \, model)}$$
Analysis of Variance Report

Analysis of variance (ANOVA) for a regression partitions the total variation of a sample into components. These components are used to compute an $F$-ratio that evaluates the effectiveness of the model. If the probability associated with the $F$-ratio is small, then the model is considered a better statistical fit for the data than the response mean alone. The Analysis of Variance reports in Figure 4.11 compare a linear fit (Fit Line) and a second degree (Fit Polynomial). Both fits are statistically better from a horizontal line at the mean.

Table 4.7 Description of Analysis of Variance Report

| Source | DF | The three sources of variation: Model, Error, and C. Total.
|--------|----|---------------------------------------------------------------------
|        |    | The degrees of freedom (DF) for each source of variation:
|        |    | - A degree of freedom is subtracted from the total number of nonmissing values ($N$) for each parameter estimate used in the computation. The computation of the total sample variation uses an estimate of the mean. Therefore, one degree of freedom is subtracted from the total, leaving 39. The total corrected degrees of freedom are partitioned into the Model and Error terms.
|        |    | - One degree of freedom from the total (shown on the Model line) is used to estimate a single regression parameter (the slope) for the linear fit. Two degrees of freedom are used to estimate the parameters ($\beta_1$ and $\beta_2$) for a polynomial fit of degree 2.
|        |    | - The Error degrees of freedom is the difference between C. Total df and Model df. |
Performing Bivariate Analysis
Fit Line and Fit Polynomial

Chapter 4

Parameter Estimates Report

The terms in the Parameter Estimates report for a linear fit (Figure 4.12) are the intercept and the single $x$ variable. Also shown in Figure 4.12 are the parameter estimates for a polynomial model of degree 2. For a polynomial fit of order $k$, there is an estimate for the model intercept and a parameter estimate for each of the $k$ powers of the $X$ variable.

Table 4.7 Description of Analysis of Variance Report (Continued)

<table>
<thead>
<tr>
<th>Sum of Squares</th>
<th>The sum of squares (SS for short) for each source of variation:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• In this example, the total (C. Total) sum of squared distances of each response from the sample mean is 57,258.157, as shown in Figure 4.11. That is the sum of squares for the base model (or simple mean model) used for comparison with all other models.</td>
</tr>
<tr>
<td></td>
<td>• For the linear regression, the sum of squared distances from each point to the line of fit reduces from 12,012.733. This is the residual or unexplained (Error) SS after fitting the model. The residual SS for a second degree polynomial fit is 6,906.997, accounting for slightly more variation than the linear fit. That is, the model accounts for more variation because the model SS are higher for the second degree polynomial than the linear fit. The C. total SS less the Error SS gives the sum of squares attributed to the model.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Mean Square</th>
<th>The sum of squares divided by its associated degrees of freedom. The $F$-ratio for a statistical test is the ratio of the following mean squares:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• The Model mean square for the linear fit is 45,265.424. This value estimates the error variance, but only under the hypothesis that the model parameters are zero.</td>
</tr>
<tr>
<td></td>
<td>• The Error mean square is 245.2. This value estimates the error variance.</td>
</tr>
</tbody>
</table>

| F Ratio | The model mean square divided by the error mean square. The underlying hypothesis of the fit is that all the regression parameters (except the intercept) are zero. If this hypothesis is true, then both the mean square for error and the mean square for model estimate the error variance, and their ratio has an $F$-distribution. If a parameter is a significant model effect, the $F$-ratio is usually higher than expected by chance alone. |

| Prob > F | The observed significance probability ($p$-value) of obtaining a greater $F$-value by chance alone if the specified model fits no better than the overall response mean. Observed significance probabilities of 0.05 or less are often considered evidence of a regression effect. |
Figure 4.12  Examples of Parameter Estimates Reports for Linear and Polynomial Fits

Table 4.8  Description of Parameter Estimates Report

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term</td>
<td>Lists the name of each parameter in the requested model. The intercept is a constant term in all models.</td>
</tr>
<tr>
<td>Estimate</td>
<td>Lists the parameter estimates of the linear model. The prediction formula is the linear combination of these estimates with the values of their corresponding variables.</td>
</tr>
<tr>
<td>Std Error</td>
<td>Lists the estimates of the standard errors of the parameter estimates. They are used in constructing tests and confidence intervals.</td>
</tr>
<tr>
<td>t Ratio</td>
<td>Lists the test statistics for the hypothesis that each parameter is zero. It is the ratio of the parameter estimate to its standard error. If the hypothesis is true, then this statistic has a Student's $t$-distribution.</td>
</tr>
<tr>
<td>Prob&gt;</td>
<td>t</td>
</tr>
</tbody>
</table>

To reveal additional statistics, right-click in the report and select the Columns menu. Statistics not shown by default are as follows:

- **Lower 95%** is the lower endpoint of the 95% confidence interval for the parameter estimate.
- **Upper 95%** is the upper endpoint of the 95% confidence interval for the parameter estimate.
- **Std Beta** is the standardized parameter estimate. It is useful for comparing the effect of $X$ variables that are measured on different scales. It is calculated as follows:

$$\hat{\beta} \left( \frac{s_X}{s_Y} \right),$$

where $\hat{\beta}$ is the estimated parameter, $s_X$ and $s_Y$ are the standard deviations of the $X$ and $Y$ variables.
• VIF is the variance inflation factor.
• Design Std Error is the design standard error for the parameter estimate. It is calculated as the standard error of the parameter estimate divided by the RMSE.

**Fit Special**

Using the Fit Special command, you can transform Y and X. Transformations include the following: log, square root, square, reciprocal, and exponential. You can also constrain the slope and intercept, fit a polynomial of specific degree, and center the polynomial.

To transform Y as log and X as square root, proceed as follows:

1. Open the SAT.jmp sample data table.
2. Select Analyze > Fit Y by X.
4. Select % Taking (2004) and click X, Factor.
5. Click OK.

![Figure 4.13 Example of SAT Scores by Percent Taking](image)

6. From the red triangle menu for Bivariate Fit, select Fit Special. The Specify Transformation or Constraint window appears. (For a description of this window, see Table 4.9.)
Performing Bivariate Analysis

Chapter 4

Figure 4.14 The Specify Transformation or Constraint Window

7. Within Y Transformation, select Natural Logarithm: log(y).
8. Within X Transformation, select Square Root: sqrt(x).
9. Click OK.

Figure 4.15 Example of Fit Special Report
Figure 4.15 shows the fitted line plotted on the original scale. The model appears to fit the data well, as the plotted line goes through the center of the points.

Table 4.9 Description of Specify Transformation or Constraint Window

<table>
<thead>
<tr>
<th>Description</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y Transformation</td>
<td>Use these options to transform the Y variable.</td>
</tr>
<tr>
<td>X Transformation</td>
<td>Use these options to transform the X variable.</td>
</tr>
<tr>
<td>Degree</td>
<td>Use this option to fit a polynomial of the specified degree.</td>
</tr>
<tr>
<td>Centered Polynomial</td>
<td>To turn off polynomial centering, deselect the Centered Polynomial check box. See Figure 4.14. Note that for transformations of the X variable, polynomial centering is not performed. Polynomial centering can make main effect tests meaningful in the presence of higher order effects (quadratic, cubic, and so on).</td>
</tr>
<tr>
<td>Constrain Intercept to</td>
<td>Select this check box to constrain the model intercept to be the specified value.</td>
</tr>
<tr>
<td>Constrain Slope to</td>
<td>Select this check box to constrain the model slope to be the specified value.</td>
</tr>
</tbody>
</table>

**Fit Special Reports and Menus**

Depending on your selections in the Fit Special window, you see different reports and menus. The flowchart in Figure 4.16 shows you what reports and menus you see depending on your choices.

---

**Figure 4.16 Example of Fit Special Flowchart**

```
Transformation?  
  Yes  
  > Transformed Fit Report and menu  
  No  
  > Degree?  
  > 1  
  > Linear Fit Report and menu  
  > 2-5  
  > Polynomial Fit Report and menu
```
Performing Bivariate Analysis

Chapter 4

Fit Spline

• For details about the Transformed Fit Report, see “Transformed Fit Report,” p. 106.
• For details about Linear and Polynomial Fit Reports, see “Linear Fit and Polynomial Fit Reports,” p. 96.
  • Transformed Fit Report

The Transformed Fit report contains the reports described in “Linear Fit and Polynomial Fit Reports,” p. 96.

However, if you transformed Y, the Fit Measured on Original Scale report appears. This shows the measures of fit based on the original Y variables, and the fitted model transformed back to the original scale.

Transformed Fit Menu

For details about the Transformed Fit menu options, see “Fitting Menu Options,” p. 119.

Fit Spline

Using the Fit Spline command, you can fit a smoothing spline that varies in smoothness (or flexibility) according to the lambda (λ) value. The lambda value is a tuning parameter in the spline formula. As the value of λ decreases, the error term of the spline model has more weight and the fit becomes more flexible and curved. As the value of λ increases, the fit becomes stiff (less curved), approaching a straight line.

For example, you want to compare two splines with high and low λ values. Proceed as follows:

1. Complete the steps that precede Figure 4.13 in “Fit Special,” p. 103.
2. From the red triangle menu for Bivariate Fit, select Fit Spline > 1000000, stiff. Then select Fit Spline > 0.01, flexible. See Figure 4.17.

Note: If you want to use a lambda value that is not listed on the menu, select Fit Spline > Other. You can also specify whether the x values should be standardized.
Note the following information:

- The smoothing spline can help you see the expected value of the distribution of $Y$ across $X$.
- The points closest to each piece of the fitted curve have the most influence on it. The influence increases as you lower the value of $\lambda$, producing a highly flexible curve.
- You might find it helpful to try several $\lambda$ values. You can use the Lambda slider beneath the Smoothing Spline report to experiment with different $\lambda$ values. However, $\lambda$ is not invariant to the scaling of the data. For example, the $\lambda$ value for an $X$ measured in inches, is not the same as the $\lambda$ value for an $X$ measured in centimeters.

**Note:** For statistical details about Fit Spline, see "Statistical Details," p. 122.

### Smoothing Spline Fit Report

The Smoothing Spline Fit report contains the **R-Square** for the spline fit and the **Sum of Squares Error**. You can use these values to compare the spline fit to other fits, or to compare different spline fits to each other.

#### Table 4.10 Descriptions of Smoothing Spline Fit Report

<table>
<thead>
<tr>
<th>Description</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>R-Square</strong></td>
<td>Measures the proportion of variation accounted for by the smoothing spline model. This is equal to 1-(SSE/C.Total SS), where C.Total SS is available in the Fit Line ANOVA report.</td>
</tr>
<tr>
<td><strong>Sum of Squares Error</strong></td>
<td>Sum of squared distances from each point to the fitted spline. It is the unexplained error (residual) after fitting the spline model.</td>
</tr>
</tbody>
</table>
Performing Bivariate Analysis

Chapter 4

Fit Each Value

Table 4.10 Descriptions of Smoothing Spline Fit Report (Continued)

| Change Lambda | Enables you to change the $\lambda$ value, either by entering a number, or by moving the slider. |

Smoothing Spline Fit Menu

For details about the Smoothing Spline Fit menu options, see “Fitting Menu Options,” p. 119.

Fit Each Value

The Fit Each Value command fits a value to each unique $X$ value. The fitted values are the means of the response for each unique $X$ value.

Figure 4.18 Example of Fit Each Value

Fit Each Value Report

The following table describes the Fit Each Value Report.

Table 4.11 Description of Fit Each Value Report

<table>
<thead>
<tr>
<th>Number of Observations</th>
<th>Gives the total number of observations.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Unique Values</td>
<td>Gives the number of unique $X$ values.</td>
</tr>
</tbody>
</table>
Chapter 4
Performing Bivariate Analysis

Fit Orthogonal

The **Fit Orthogonal** command fits lines that adjust for variability in \( X \) as well as \( Y \).

This section contains the following information:

- “Fit Orthogonal Options,” p. 109
- “Example of a Scenario Using the Fit Orthogonal Command,” p. 111
- “Orthogonal Regression Report,” p. 112
- “Orthogonal Fit Ratio Menu,” p. 112

**Note:** For statistical details about Fit Orthogonal, see “Statistical Details,” p. 122.

### Fit Orthogonal Options

The following table describes the available options to specify a variance ratio.

**Table 4.12** Descriptions of Fit Orthogonal Options

| Univariate Variances, Prin Comp | Uses the univariate variance estimates computed from the samples of \( X \) and \( Y \). This turns out to be the standardized first principal component. This option is not a good choice in a measurement systems application since the error variances are not likely to be proportional to the population variances. |
| Equal Variances                   | Uses 1 as the variance ratio, which assumes that the error variances are the same. Using equal variances is equivalent to the non-standardized first principal component line. If the scatterplot is scaled the same in the \( X \) and \( Y \) directions, and you show a normal density ellipse, you can see that this line is the longest axis of the ellipse (Figure 4.19). |
Performing Bivariate Analysis

Chapter 4

Fit Orthogonal

Table 4.12 Descriptions of Fit Orthogonal Options (Continued)

<table>
<thead>
<tr>
<th>Fit X to Y</th>
<th>Uses a variance ratio of zero, which indicates that $Y$ effectively has no variance. See Figure 4.19.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specified Variance Ratio</td>
<td>Lets you enter any ratio that you want, giving you the ability to make use of known information about the measurement error in $X$ and response error in $Y$.</td>
</tr>
</tbody>
</table>

Illustrate the Fit Orthogonal Command Options

This example involves two parts. First, standardize the variables using the Distribution platform. Then, use the standardized variables to fit the orthogonal model.

Standardize the Variables
1. Open the Big Class.jmp sample data table.
2. Select Analyze > Distribution.
4. Click OK.
5. Hold down the CTRL key. On the red triangle menu next to height, select Save > Standardized.
   Holding down the CTRL key broadcasts the operation to all variables in the report window. Notice that in the Big Class.jmp sample data table, two new columns have been added.

Use the Standardized Variables to Fit the Orthogonal Model
1. From the Big Class.jmp sample data table, select Analyze > Fit Y by X.
2. Select Std weight and click Y, Response.
3. Select Std height and click X, Factor.
4. Click OK.
5. From the red triangle menu, select Fit Line.
6. From the red triangle menu, select Fit Orthogonal, then select each of the following:
   - Equal Variances
   - Fit X to Y
     - Specified Variance Ratio and type 0.2
     - Specified Variance Ratio and type 5
The scatterplot in Figure 4.19 shows the standardized height and weight values with various line fits that illustrate the behavior of the orthogonal variance ratio selections. The standard linear regression (Fit Line) occurs when the variance of the $X$ variable is considered to be very small. Fit $X$ to $Y$ is the opposite extreme, when the variation of the $Y$ variable is ignored. All other lines fall between these two extremes and shift as the variance ratio changes. As the variance ratio increases, the variation in the $Y$ response dominates and the slope of the fitted line shifts closer to the $Y$ by $X$ fit. Likewise, when you decrease the ratio, the slope of the line shifts closer to the $X$ by $Y$ fit.

**Example of a Scenario Using the Fit Orthogonal Command**

Suppose you have two weighing scales, a good one called A, and a bad one called B. Scale A is known to be unbiased and has a standard deviation in repeated measurements of 21 grams. Scale B is known to be biased (but linearly so) and has a standard deviation of 23 grams in repeated measurements. You have weight measurements of your experimental responses from both scales. It would be tempting to just use the measurements from Scale A, the good scale. But you can do better. The readings from Scale B still have value. Do an orthogonal regression of the B on the A weight measurements and use a variance ratio of $(23/21)^2$. Then you get predicted values for both A and B. Because A is unbiased, you can use the predicted values of A as your new-improved measurement, with a measurement error variance of nearly half of what you would have gotten with just Scale A.
Orthogonal Regression Report

The following table describes the Orthogonal Regression report.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable</td>
<td>Gives the names of the variables used to fit the line.</td>
</tr>
<tr>
<td>Mean</td>
<td>Gives the mean of each variable.</td>
</tr>
<tr>
<td>Std Dev</td>
<td>Gives the standard deviation of each variable.</td>
</tr>
<tr>
<td>Variance Ratio</td>
<td>Gives the variance ratio used to fit the line.</td>
</tr>
<tr>
<td>Correlation</td>
<td>Gives the correlation between the two variables.</td>
</tr>
<tr>
<td>Intercept</td>
<td>Gives the intercept of the fitted line.</td>
</tr>
<tr>
<td>Slope</td>
<td>Gives the slope of the fitted line.</td>
</tr>
<tr>
<td>LowerCL</td>
<td>Gives the lower confidence limit for the slope.</td>
</tr>
<tr>
<td>UpperCL</td>
<td>Gives the upper confidence limit for the slope.</td>
</tr>
<tr>
<td>Alpha</td>
<td>Enter the alpha level used in computing the confidence interval.</td>
</tr>
</tbody>
</table>

Orthogonal Fit Ratio Menu

For details about the Orthogonal Fit menu options, see “Fitting Menu Options,” p. 119.

Density Ellipse

Using the **Density Ellipse** option, you can draw an ellipse that contains the specified mass of points (determined by the probability that you select from the **Density Ellipse** menu). The density ellipsoid is computed from the bivariate normal distribution fit to the \( X \) and \( Y \) variables. The bivariate normal density is a function of the means and standard deviations of the \( X \) and \( Y \) variables and the correlation between them. The **Other** selection lets you specify any probability greater than zero and less than or equal to one. These ellipses are both density contours and confidence curves. As confidence curves, they show where a given percentage of the data is expected to lie, assuming the bivariate normal distribution.

For example, to add a variety of different ellipses, proceed as follows:

1. Complete the steps that precede Figure 4.13 in “Fit Special,” p. 103.
2. From the red triangle menu for Bivariate Fit, select **Density Ellipse**, then select each of the following:
   - 0.99
   - 0.95
   - 0.90
Figure 4.20 compares ellipses with different probabilities.

**Figure 4.20 Example of Density Ellipses**

The density ellipsoid is a good graphical indicator of the correlation between two variables. The ellipsoid collapses diagonally as the correlation between the two variables approaches either 1 or –1. The ellipsoid is more circular (less diagonally oriented) if the two variables are less correlated.

**Correlation Report**

The Correlation report that accompanies each Density Ellipse fit shows the correlation coefficient for the X and Y variables.

**Note:** To see a matrix of ellipses and correlations for many pairs of variables, use the Multivariate command in the Analyze > Multivariate Methods menu.

**Table 4.14 Description of Correlation Report**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Correlation</th>
<th>Signif. Prob</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>% Taking (2004)</td>
<td>9.397009</td>
<td>0.26694</td>
<td>-0.88897</td>
<td>.0001*</td>
<td>51</td>
</tr>
<tr>
<td>2004 Verbal</td>
<td>525.3022</td>
<td>33.8517</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A discussion of the mean and standard deviation are in the section “The Moments Report,” p. 32 in the “Performing Univariate Analysis” chapter.
Performing Bivariate Analysis

Chapter 4

Nonpar Density

Bivariate Normal Ellipse Menu

For details about the Bivariate Normal Ellipse menu options, see “Fitting Menu Options,” p. 119.

Nonpar Density

Use the Nonpar Density (nonparametric density) command when you have thousands of points and the scatterplot is so darkened by points that it is difficult to see patterns in the point density.

Bivariate density estimation models a smooth surface that describes how dense the data points are at each point in that surface. The plot adds a set of contour lines showing the density (Figure 4.21). The contour lines are quantile contours in 5% intervals. This means that about 5% of the points are below the lowest contour, 10% are below the next contour, and so forth. The highest contour has about 95% of the points below it.

Table 4.14 Description of Correlation Report (Continued)

| Correlation | The Pearson correlation coefficient, denoted r, and is computed as follows:
|-------------|---------------------------------------------------------------------------------------------------
|             | \[ r_{xy} = \frac{s_{xy}^2}{s_x^2 s_y^2} \text{ where } s_{xy}^2 = \frac{\sum w_i (x_i - \overline{x})(y_i - \overline{y})}{df} \]  
|             | Where \( w_i \) is either the weight of the \( i \)th observation if a weight column is specified, or 1 if no weight column is assigned. If there is an exact linear relationship between two variables, the correlation is 1 or –1 depending on whether the variables are positively or negatively related. If there is no relationship, the correlation tends toward zero. |
| Signif. Prob | Probability of obtaining, by chance alone, a correlation with greater absolute value than the computed value if no linear relationship exists between the \( X \) and \( Y \) variables. |
| Number | Gives the number of observations used in the calculations. |
Nonparametric Bivariate Density Report

The nonparametric bivariate density report shows the kernel standard deviations used in creating the nonparametric density.

Quantile Density Contours Menu

For details about the Quantile Density Contours menu options, see “Fitting Menu Options,” p. 119.
Histogram Borders

The Histogram Borders option appends histograms to the x- and y-axes of the scatterplot. You can use the histograms to visualize the marginal distributions of the X and Y variables.

Figure 4.22 Example of Histogram Borders

Group By

Using the Group By option, you can select a classification (grouping) variable. When a grouping variable is in effect, the Bivariate platform computes a separate analysis for each level of the grouping variable and overlays the regression curves or ellipses on the scatterplot. The fit for each level of the grouping variable is identified beneath the scatterplot, with individual popup menus to save or remove fitting information.

The Group By option is checked in the Fitting menu when a grouping variable is in effect. You can change the grouping variable by first selecting the Group By option to remove (uncheck) the existing variable. Then, select the Group By option again and respond to its window as before.

You might use the Group By option in these different ways:

- An overlay of linear regression lines lets you compare slopes visually.
- An overlay of density ellipses can show clusters of points by levels of a grouping variable.

Example of Group By Using Density Ellipses

This example uses the Hot Dogs.jmp sample data table. The Type column identifies three different types of hot dogs: beef, meat, or poultry. You want to group the three types of hot dogs according to their cost variables.

1. Open the Hot Dogs.jmp sample data table.
2. Select Analyze > Fit Y by X.
3. Select $/oz and click Y, Response.
4. Select $/lb Protein and click X, Factor.
5. Click OK.
6. From the red triangle menu, select Group By.
7. From the list, select Type.
8. Click OK. If you look at the Group By option again, you will see it has a check mark next to it.
9. From the red triangle menu, select Density Ellipse > 0.90.

To color the points according to Type, proceed as follows:
10. Right-click on the scatterplot and select Row Legend.
11. Select Type in the column list and click OK.

**Figure 4.23 Example of Group By**

The ellipses in Figure 4.23 show clearly how the different types of hot dogs cluster with respect to the cost variables.

**Example of Group By Using Regression Lines**

Another use for grouped regression is overlaying lines to compare slopes of different groups.

1. Open the Big Class.jmp sample data table.
2. Select Analyze > Fit Y by X.
3. Select weight and click Y, Response.
4. Select height and click X, Factor.
5. Click OK.
6. To create the example on the left in Figure 4.24, select Fit Line from the red triangle menu.
Performing Bivariate Analysis

Fitting Menus

To create the example on the right, instead of step 6, proceed as follows:

7. From the red triangle menu, select **Group By**.
8. From the list, select **sex**.
9. Click **OK**.
10. Select **Fit Line** from the red triangle menu.

---

**Figure 4.24** Example of Regression Analysis for Whole Sample and Grouped Sample

The scatterplot to the left in Figure 4.24 has a single regression line that relates weight to height. The scatterplot to the right shows separate regression lines for males and females.

---

Fitting Menus

In addition to a report, each fitting command adds a fitting menu to the report window. Table 4.15 shows the fitting menus that correspond to each fitting command.

**Table 4.15** Descriptions of Fitting Menus

<table>
<thead>
<tr>
<th>Fitting Command</th>
<th>Fitting Menu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fit Mean</td>
<td>Fit Mean</td>
</tr>
<tr>
<td>Fit Line</td>
<td>Linear Fit</td>
</tr>
<tr>
<td>Fit Polynomial</td>
<td>Polynomial Fit Degree=X*</td>
</tr>
<tr>
<td>Fit Special</td>
<td>Linear Fit</td>
</tr>
<tr>
<td></td>
<td>Polynomial Fit Degree=X*</td>
</tr>
<tr>
<td></td>
<td>Transformed Fit X*</td>
</tr>
<tr>
<td></td>
<td>Constrained Fits</td>
</tr>
</tbody>
</table>
Fitting Menus

Table 4.15 Descriptions of Fitting Menus  (Continued)

<table>
<thead>
<tr>
<th>Fitting Command</th>
<th>Fitting Menu</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fit Spline</td>
<td>Smoothing Spline Fit, lambda=X*</td>
</tr>
<tr>
<td>Fit Each Value</td>
<td>Fit Each Value</td>
</tr>
<tr>
<td>Fit Orthogonal</td>
<td>Orthogonal Fit Ratio=X*</td>
</tr>
<tr>
<td>Density Ellipse</td>
<td>Bivariate Normal Ellipse=X*</td>
</tr>
<tr>
<td>Nonpar Density</td>
<td>Quantile Density Colors</td>
</tr>
</tbody>
</table>

*X=variable character or number

Fitting Menu Options

The following table describes the Fitting menu options:

Table 4.16 Descriptions of Fitting Menu Options

<table>
<thead>
<tr>
<th>Confid Curves Fit</th>
<th>Displays or hides confidence curves for mean, line, and polynomial fits. This option is not available for the Fit Spline, Density Ellipse, Fit Each Value, and Fit Orthogonal fits and is dimmed on those menus.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confid Curves Indiv</td>
<td>Displays the upper and lower confidence limits for an individual predicted value. The confidence limits reflect variation in the error and variation in the parameter estimates. This option is not available for the Fit Mean, Fit Spline, Density Ellipse, Fit Each Value, and Fit Orthogonal fits and is dimmed on those menus.</td>
</tr>
<tr>
<td>Line Color</td>
<td>Lets you select from a palette of colors for assigning a color to each fit.</td>
</tr>
<tr>
<td>Line of Fit</td>
<td>Displays or hides the line of fit.</td>
</tr>
<tr>
<td>Line Style</td>
<td>Lets you select from the palette of line styles for each fit.</td>
</tr>
<tr>
<td>Line Width</td>
<td>Gives three line widths for the line of fit. The default line width is the thinnest line.</td>
</tr>
<tr>
<td>Report</td>
<td>Turns the fit’s text report on and off.</td>
</tr>
<tr>
<td>Table 4.16 Descriptions of Fitting Menu Options (Continued)</td>
<td></td>
</tr>
<tr>
<td>------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td><strong>Save Predicteds</strong></td>
<td>Creates a new column in the current data table called Predicted colname where colname is the name of the Y variable. This column includes the prediction formula and the computed sample predicted values. The prediction formula computes values automatically for rows that you add to the table. This option is not available for the Fit Each Value and Density Ellipse fits and is dimmed on those menus.</td>
</tr>
<tr>
<td><strong>Note:</strong> You can use the Save Predicteds and Save Residuals commands for each fit. If you use these commands multiple times or with a grouping variable, it is best to rename the resulting columns in the data table to reflect each fit.</td>
<td></td>
</tr>
<tr>
<td><strong>Save Residuals</strong></td>
<td>Creates a new column in the current data table called Residuals colname where colname is the name of the Y variable. Each value is the difference between the actual (observed) value and its predicted value. Unlike the Save Predicteds command, this command does not create a formula in the new column. This option is not available for the Fit Each Value and Density Ellipse fits and is dimmed on those menus.</td>
</tr>
<tr>
<td><strong>Note:</strong> You can use the Save Predicteds and Save Residuals commands for each fit. If you use these commands multiple times or with a grouping variable, it is best to rename the resulting columns in the data table to reflect each fit.</td>
<td></td>
</tr>
<tr>
<td><strong>Remove Fit</strong></td>
<td>Removes the fit from the graph and removes its text report.</td>
</tr>
<tr>
<td><strong>Linear Fits, Polynomial Fits, and Fit Special Only:</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Confid Shaded Fit</strong></td>
<td>Draws the same curves as the Confid Curves Fit command and shades the area between the curves.</td>
</tr>
<tr>
<td><strong>Confid Shaded Indiv</strong></td>
<td>Draws the same curves as the Confid Curves Indiv command and shades the area between the curves.</td>
</tr>
<tr>
<td><strong>Plot Residuals</strong></td>
<td>Produces four diagnostic plots: residual by predicted (with a histogram of the residuals), actual by predicted, residual by row, and a normal quantile plot of the residuals.</td>
</tr>
<tr>
<td><strong>Set Alpha Level</strong></td>
<td>Prompts you to enter the alpha level to compute and display confidence levels for line fits, polynomial fits, and special fits.</td>
</tr>
</tbody>
</table>
Table 4.16 Descriptions of Fitting Menu Options  (Continued)

<table>
<thead>
<tr>
<th><strong>Smoothing Spline Fit and Local Smoother Only:</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Save Coefficients</strong></td>
<td>Saves the spline coefficients as a new data table, with columns called X, A, B, C, and D. The X column gives the knot points. A, B, C, and D are the intercept, linear, quadratic, and cubic coefficients of the third-degree polynomial, spanning from the corresponding value in the X column to the next highest value.</td>
</tr>
<tr>
<td><strong>Save Prediction Formula</strong></td>
<td>Saves the prediction formula to a new column in the data table.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Bivariate Normal Ellipse Only:</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Shaded Contour</strong></td>
<td>Shades the area inside the density ellipse.</td>
</tr>
<tr>
<td><strong>Select Points Inside</strong></td>
<td>Selects the points inside the ellipse.</td>
</tr>
<tr>
<td><strong>Select Points Outside</strong></td>
<td>Selects the points outside the ellipse.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Quantile Density Contours Only:</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Kernel Control</strong></td>
<td>Displays a slider for each variable, where you can change the kernel standard deviation that defines the range of X and Y values for determining the density of contour lines.</td>
</tr>
<tr>
<td><strong>5% Contours</strong></td>
<td>Shows or hides the 5% contour lines.</td>
</tr>
<tr>
<td><strong>Contour Lines</strong></td>
<td>Shows or hides the contour lines.</td>
</tr>
<tr>
<td><strong>Contour Fill</strong></td>
<td>Fills the areas between the contour lines.</td>
</tr>
<tr>
<td><strong>Select Points by Density</strong></td>
<td>Selects points that fall in a user-specified quantile range.</td>
</tr>
<tr>
<td><strong>Color by Density Quantile</strong></td>
<td>Colors the points according to density.</td>
</tr>
<tr>
<td><strong>Save Density Quantile</strong></td>
<td>Creates a new column containing the density quantile each point is in.</td>
</tr>
<tr>
<td><strong>Mesh Plot</strong></td>
<td>Is a three-dimensional plot of the density over a grid of the two analysis variables. See Figure 4.25.</td>
</tr>
<tr>
<td><strong>Model Clustering</strong></td>
<td>Creates a new column in the current data table and fills it with cluster values.</td>
</tr>
</tbody>
</table>

**Note:** If you save the modal clustering values first and then save the density grid, the grid table also contains the cluster values. The cluster values are useful for coloring and marking points in plots.
Performing Bivariate Analysis
Chapter 4

Statistical Details

Table 4.16 Descriptions of Fitting Menu Options  (Continued)

<table>
<thead>
<tr>
<th>Save Density Grid</th>
</tr>
</thead>
<tbody>
<tr>
<td>Saves the density estimates and the quantiles associated with them in a new data table. The grid data can be used to visualize the density in other ways, such as with the Scatterplot 3D or the Contour Plot platforms.</td>
</tr>
</tbody>
</table>

Figure 4.25 Example of a Mesh Plot

![Mesh Plot](image)

Statistical Details

This section contains statistical details for the Fit Line, Fit Spline, and Fit Orthogonal commands.

Fit Line

The **Fit Line** command finds the parameters $\beta_0$ and $\beta_1$ for the straight line that fits the points to minimize the residual sum of squares. The model for the $i$th row is written $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$.

A polynomial of degree 2 is a parabola; a polynomial of degree 3 is a cubic curve. For degree $k$, the model for the $i$th observation is as follows:

$$ y_i = \sum_{j=0}^{k} \beta_j x_i^j + \epsilon_i $$
Chapter 4
Performing Bivariate Analysis
Statistical Details

Fit Spline

The cubic spline method uses a set of third-degree polynomials spliced together such that the resulting curve is continuous and smooth at the splices (knot points). The estimation is done by minimizing an objective function that is a combination of the sum of squares error and a penalty for curvature integrated over the curve extent. See the paper by Reinsch (1967) or the text by Eubank (1988) for a description of this method.

Fit Orthogonal

Standard least square fitting assumes that the \( X \) variable is fixed and the \( Y \) variable is a function of \( X \) plus error. If there is random variation in the measurement of \( X \), then instead of fitting a standard regression line that minimizes the sum of the squared vertical distances, you should fit a line that minimizes the sum of the squared perpendicular differences. See Figure 4.26. However, the perpendicular distance depends on how \( X \) and \( Y \) are scaled, and the scaling for the perpendicular is reserved as a statistical issue, not a graphical one.

![Figure 4.26 Line Perpendicular to the Line of Fit](image)

The fit requires that you specify the ratio of the variance of the error in \( X \) to the error in \( Y \). This is the variance of the error, not the variance of the sample points, so you must choose carefully. The ratio \((\sigma_Y^2)/(\sigma_X^2)\) is infinite in standard least squares because \(\sigma_X^2\) is zero. If you do an orthogonal fit with a large error ratio, the fitted line approaches the standard least squares line of fit. If you specify a ratio of zero, the fit is equivalent to the regression of \( Y \) on \( X \), instead of \( X \) on \( Y \).

The most common use of this technique is in comparing two measurement systems that both have errors in measuring the same value. Thus, the \( Y \) response error and the \( X \) measurement error are both the same type of measurement error. Where do you get the measurement error variances? You cannot get them from bivariate data because you cannot tell which measurement system produces what proportion of the error. So, you either must blindly assume some ratio like 1, or you must rely on separate repeated measurements of the same unit by the two measurement systems.

An advantage to this approach is that the computations give you predicted values for both \( Y \) and \( X \); the predicted values are the point on the line that is closest to the data point, where closeness is relative to the variance ratio.

Confidence limits are calculated as described in Tan and Iglewicz (1999).
Performing Oneway Analysis
Using the Fit Y by X or Oneway Platform

Using the Oneway or Fit Y by X platform, you can explore how the distribution of a continuous Y variable differs across groups defined by a single categorical X variable. For example, you might want to find out how different categories of the same type of drug (X) affect patient pain levels on a numbered scale (Y).

The Oneway platform is the *continuous by nominal or ordinal* personality of the Fit Y by X platform. The analysis results appear in a plot, and you can interactively add additional analyses, such as the following:

- a one-way analysis of variance to fit means and to test that they are equal
- nonparametric tests
- a test for homogeneity of variance
- multiple-comparison tests on means, with means comparison circles
- outlier box plots overlaid on each group
- power details for the one-way layout

**Figure 5.1** Oneway Analysis
Contents

Overview of Oneway Analysis ................................................................. 127
Example of Oneway Analysis ................................................................. 127
Launch the Oneway Platform ................................................................. 129
The Oneway Report .............................................................................. 129
Oneway Platform Options ................................................................. 130
Quantiles ............................................................................................... 136
Means/Anova and Means/Anova/Pooled t ........................................... 138
Analysis of Means Methods ................................................................. 146
Compare Means ....................................................................................... 150
Nonparametric ........................................................................................ 162
Unequal Variances ................................................................................. 167
Equivalence Test .................................................................................... 172
Power ....................................................................................................... 173
Normal Quantile Plot ............................................................................ 176
CDF Plot ................................................................................................. 177
Densities ................................................................................................. 178
Matching Column .................................................................................. 180
Statistical Details ................................................................................... 182
Overview of Oneway Analysis

A one-way analysis of variance tests for differences between group means. The total variability in the response is partitioned into two parts: within-group variability and between-group variability. If the between-group variability is large relative to the within-group variability, then the differences between the group means are considered to be significant.

Example of Oneway Analysis

This example uses the Analgesics.jmp sample data table. Thirty-three subjects were administered three different types of analgesics (A, B, and C). The subjects were asked to rate their pain levels on a sliding scale. You want to find out if the means for A, B, and C are significantly different.

1. Open the Analgesics.jmp sample data table.
2. Select Analyze > Fit Y by X.
4. Select drug and click X, Factor.
5. Click OK.

You notice that one drug (A) has consistently lower scores than the other drugs. You also notice that the x-axis ticks are unequally spaced. The length between the ticks is proportional to the number of scores (observations) for each drug.

Perform an analysis of variance on the data.

6. From the red triangle menu for Oneway Analysis, select Means/Anova.

**Note:** If the X factor has only two levels, the Means/Anova option appears as Means/Anova/Pooled t, and adds a pooled t-test report to the report window.
Performing One-way Analysis

Chapter 5

Example of One-way Analysis

Figure 5.3 Example of the Means/Anova Option

Note the following observations:

- Mean diamonds representing confidence intervals appear.
  - The line near the center of each diamond represents the group mean. At a glance, you can see that the mean for each drug looks significantly different.
  - The vertical span of each diamond represents the 95% confidence interval for the mean of each group.

For more details, see “Mean Diamonds and X-Axis Proportional,” p. 144.

- The Summary of Fit table provides overall summary information about the analysis.
- The Analysis of Variance report shows the standard ANOVA information. You notice that the Prob > F (the p-value) is 0.0053, which supports your visual conclusion that there are significant differences between the drugs.
- The Means for One-way Anova report shows the mean, sample size, and standard error for each level of the categorical factor.
Launch the Oneway Platform

You can perform a Oneway analysis using either the Fit Y by X platform or the Oneway platform. The two approaches are equivalent.

- To launch the Fit Y by X platform, select Analyze > Fit Y by X. The Fit Y by X launch window appears. See Figure 3.2 in the “Introduction to the Fit Y by X Platform” chapter.
- or

- To launch the Oneway platform, from the JMP Starter window, click on the Basic category and click Oneway. The Oneway launch window appears.

**Figure 5.4 The Oneway Launch Window**

You can specify a Block variable to add a second factor, which forms a two-way analysis without interaction. If you specify a Block variable, the data should be balanced and have equal counts in each block by group cell.

For details about this window, see the “Introduction to the Fit Y by X Platform” chapter.

The Oneway Report

To produce the plot shown in Figure 5.5, follow the instructions in “Example of Oneway Analysis,” p. 127.
Performing Oneway Analysis

Oneway Platform Options

Figure 5.5 The Oneway Report Window

Figure 5.5 shows the response points for each X factor value. Using the plot, you can compare the distribution of the response across the levels of the X factor. The distinct values of X are sometimes called levels.

You can add reports, plots, and tests to the report window using the options in the red triangle menu for Oneway Analysis. See "Oneway Platform Options," p. 130.

Oneway Platform Options

Figure 5.6 shows an example of the platform options in the red triangle menu for Oneway Analysis. For most options, when you select the option, a check mark appears next to the option. Deselect an option to remove it from the report window.

Figure 5.6 Example of Oneway Platform Options
When you select a platform option, objects might be added to the plot, and a report is added to the report window.

**Table 5.1 Examples of Options and Elements**

<table>
<thead>
<tr>
<th>Platform Option</th>
<th>Object Added to Plot</th>
<th>Report Added to Report Window</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quantiles</td>
<td>Box plots</td>
<td>Quantiles report</td>
</tr>
<tr>
<td>Means/Anova</td>
<td>Mean diamonds</td>
<td>Oneway ANOVA reports</td>
</tr>
<tr>
<td>Means and Std Dev</td>
<td>Mean lines, error bars, and standard deviation lines</td>
<td>Means and Std Devations report</td>
</tr>
<tr>
<td>Compare Means</td>
<td>Comparison circles (except Nonparametric Multiple Comparisons option)</td>
<td>Means Comparison reports</td>
</tr>
</tbody>
</table>

Table 5.2 briefly describes all of the platform options in the red triangle menu for Oneway Analysis. Some options might not appear unless specific conditions are met.

**Table 5.2 Descriptions of Oneway Analysis Platform Options**

<table>
<thead>
<tr>
<th>Quantiles</th>
<th>Lists the following quantiles for each group:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• 0% (Minimum)</td>
</tr>
<tr>
<td></td>
<td>• 10%</td>
</tr>
<tr>
<td></td>
<td>• 25%</td>
</tr>
<tr>
<td></td>
<td>• 50% (Median)</td>
</tr>
<tr>
<td></td>
<td>• 75%</td>
</tr>
<tr>
<td></td>
<td>• 90%</td>
</tr>
<tr>
<td></td>
<td>• 100% (Maximum)</td>
</tr>
<tr>
<td></td>
<td>Activates <strong>Box Plots</strong> from the <strong>Display Options</strong> menu.</td>
</tr>
</tbody>
</table>

See “Quantiles,” p. 136.
Table 5.2 Descriptions of Oneway Analysis Platform Options (Continued)

| Means/Anova or Means/Anova/Pooled t | Fits means for each group and performs a one-way analysis of variance to test if there are differences among the means. The plot now contains mean diamonds. For a brief description of mean diamonds, see “Display Options,” p. 134. For more details about mean diamonds, see “Mean Diamonds and X-Axis Proportional,” p. 144. The following reports appear: a Summary report, a one-way Analysis of Variance report, and a Means report that lists group frequencies, means, and standard errors computed with the pooled estimate of the error variance. See “The Summary of Fit Report,” p. 139, “The Analysis of Variance Report,” p. 142, and “The Means for Oneway Anova Report,” p. 143. The Means/Anova/Pooled t option appears only if the X factor has two levels. This option adds an additional t-Test report to the report window. See “The t-test Report,” p. 140. If you have specified a Block variable in the launch window, the Block Means report appears. See “The Block Means Report,” p. 144. |
| Means and Std Dev | Gives summary statistics for each group. The standard errors for the means use individual group standard deviations rather than the pooled estimate of the standard deviation. The plot now contains mean lines, error bars, and standard deviation lines. For a brief description of these elements, see “Display Options,” p. 134. For more details about these elements, see “Mean Lines, Error Bars, and Standard Deviation Lines,” p. 145. |
| t test | This option appears only if the X factor has two levels. Produces a t-test report assuming that the variances are not equal. See “The t-test Report,” p. 140. |
| Analysis of Means Methods | Provides four commands for performing Analysis of Means (ANOM) procedures. There are commands for comparing both means and variances. See “Analysis of Means Methods,” p. 146. |
| Nonparametric | Provides nonparametric comparisons of group means. See “Nonparametric,” p. 162. |
| Unequal Variances | Performs four tests for equality of group variances. Also gives the Welch test, which is an ANOVA test for comparing means when the variances within groups are not equal. See “Unequal Variances,” p. 167. |
Table 5.2 Descriptions of Oneway Analysis Platform Options  (Continued)

<table>
<thead>
<tr>
<th><strong>Equivalence Test</strong></th>
<th>Tests that a difference is less than a threshold value. See “Equivalence Test,” p. 172.</th>
</tr>
</thead>
</table>
| **Power**            | • Provides calculations of statistical power and other details about a given hypothesis test. See “Power,” p. 173.  
|                      | • The Power Details window and reports also appear within the Fit Model platform. For further discussion and examples of power calculations, see the *Modeling and Multivariate Methods* book. |
| **Set a Level**      | You can select an option from the most common alpha levels or specify any level with the **Other** selection. Changing the alpha level results in the following actions:  
|                      | • recalculates confidence limits  
|                      | • adjusts the mean diamonds on the plot (if they are showing)  
|                      | • modifies the upper and lower confidence level values in reports  
|                      | • changes the critical number and comparison circles for all Compare Means reports  
|                      | • changes the critical number for all Nonparametric Multiple Comparison reports |
| **Normal Quantile Plot** | Provides the following options for plotting the quantiles of the data in each group:  
|                      | • **Plot Actual by Quantile** generates a quantile plot with the response variable on the y-axis and quantiles on the x-axis. The plot shows quantiles computed within each level of the categorical X factor.  
|                      | • **Plot Quantile by Actual** reverses the x- and y-axes.  
|                      | • **Line of Fit** draws straight diagonal reference lines on the plot for each level of the X variable. This option is available only once you have created a plot (Actual by Quantile or Quantile by Actual). |
| **CDF Plot**         | Plots the cumulative distribution function for all of the groups in the Oneway report. See “CDF Plot,” p. 177. |
| **Densities**        | Compares densities across groups. See “Densities,” p. 178. |
Matching Column

- Specify a matching variable to perform a matching model analysis. Use this option when the data in your Oneway analysis comes from matched (paired) data, such as when observations in different groups come from the same subject.
- The plot now contains matching lines that connect the matching points.
- For further details and an example, see “Matching Column,” p. 180.

Save

Saves the following quantities as new columns in the current data table:

- **Save Residuals** saves values computed as the response variable minus the mean of the response variable within each level of the factor variable.
- **Save Standardized** saves standardized values of the response variable computed within each level of the factor variable. This is the centered response divided by the standard deviation within each level.
- **Save Normal Quantiles** saves normal quantile values computed within each level of the categorical factor variable.
- **Save Predicted** saves the predicted mean of the response variable for each level of the factor variable.

Display Options

Adds or removes elements from the plot. See “Display Options,” p. 134

Script

This menu contains options that are available to all platforms. They enable you to redo the analysis or save the JSL commands for the analysis to a window or a file. See “Script Menus,” p. 15 in the “Preliminaries” chapter.

### Display Options

Using Display Options, you can add or remove elements from a plot. Some options might not appear unless they are relevant.

**Table 5.3** Descriptions of Display Options

<p>| All Graphs | Shows or hides all graphs. |
| Points     | Shows or hides data points on the plot. |
| Box Plots  | Shows or hides outlier box plots for each group. |</p>
<table>
<thead>
<tr>
<th><strong>Table 5.3</strong> Descriptions of Display Options (Continued)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mean Diamonds</strong></td>
</tr>
<tr>
<td><strong>Mean Lines</strong></td>
</tr>
<tr>
<td><strong>Mean CI Lines</strong></td>
</tr>
<tr>
<td><strong>Mean Error Bars</strong></td>
</tr>
<tr>
<td><strong>Grand Mean</strong></td>
</tr>
<tr>
<td><strong>Std Dev Lines</strong></td>
</tr>
<tr>
<td><strong>Comparison Circles</strong></td>
</tr>
<tr>
<td><strong>Connect Means</strong></td>
</tr>
<tr>
<td><strong>Mean of Means</strong></td>
</tr>
<tr>
<td><strong>X-Axis proportional</strong></td>
</tr>
<tr>
<td><strong>Points Spread</strong></td>
</tr>
<tr>
<td><strong>Points Jittered</strong></td>
</tr>
<tr>
<td><strong>Matching Lines</strong></td>
</tr>
<tr>
<td><strong>Matching Dotted Lines</strong></td>
</tr>
<tr>
<td><strong>Histograms</strong></td>
</tr>
</tbody>
</table>
Quantiles

When you select the **Quantiles** option from the red triangle menu for Oneway Analysis, the Quantiles report appears, and the following elements are added to the plot:

- the grand mean representing the overall mean of the Y variable
- outlier box plots summarizing the distribution of points at each factor level

**Note:** To turn off outlier box plots, click the red triangle next to Oneway Analysis and select Display Options > Box Plots.

Example of the Quantiles Option

To see an example of the Quantiles option, proceed as follows:

1. Open the *Analgesics.jmp* sample data table.
2. Select Analyze > Fit Y by X.
4. Select drug and click X, Factor.
5. Click OK.
6. From the red triangle menu next to Oneway Analysis, select Quantiles.

**Figure 5.7 Example of the Quantiles Option**

```
<table>
<thead>
<tr>
<th>Level</th>
<th>Minimum</th>
<th>10%</th>
<th>25%</th>
<th>Median</th>
<th>75%</th>
<th>90%</th>
<th>Maximum</th>
</tr>
</thead>
</table>
```
Chapter 5
Performing Oneway Analysis

The report window shows the outlier box plots and the grand mean in the plot. The Quantiles report window also appears.

The Quantiles report lists selected percentiles for each level of the X factor variable. The median is the 50th percentile, and the 25th and 75th percentiles are called the quartiles.

Note the following aspects about outlier box plots (see Figure 5.8):

- The vertical line within the box represents the median sample value.
- The ends of the box represent the 75th and 25th quantiles, also expressed as the 3rd and 1st quartile, respectively.
- The difference between the 1st and 3rd quartiles is called the interquartile range.
- Each box has lines, sometimes called whiskers, that extend from each end. The whiskers extend from the ends of the box to the outermost data point that falls within the distances computed as follows:

\[
3\text{rd quartile} + 1.5 \times (\text{interquartile range}) \\
1\text{st quartile} - 1.5 \times (\text{interquartile range})
\]

If the data points do not reach the computed ranges, then the whiskers are determined by the upper and lower data point values (not including outliers).

Figure 5.8 Examples of Outlier Box Plots
Means/Anova and Means/Anova/Pooled t

The **Means/Anova** option performs an analysis of variance. If the X factor contains exactly two levels, this option appears as **Means/Anova/Pooled t**. In addition to the other reports, a *t*-test report assuming pooled (or equal) variances appears.

**Examples of the Means/Anova and the Means/Anova/Pooled t Options**

To see an example of the **Means/Anova** option, proceed as follows:
1. Open the Analgesics.jmp sample data table.
2. Select **Analyze > Fit Y by X**.
3. Select **pain** and click **Y, Response**.
4. Select **drug** and click **X, Factor**.
5. Click **OK**.
6. From the red triangle menu next to Oneway Analysis, select **Means/Anova**.
   See Figure 5.9 at left.

To see an example of the **Means/Anova/Pooled t** option, proceed as follows:
1. Open the Popcorn.jmp sample data table.
2. Select **Analyze > Fit Y by X**.
3. Select **yield** and click **Y, Response**.
4. Select **batch** and click **X, Factor**.
5. Click **OK**.
6. From the red triangle menu next to Oneway Analysis, select **Means/Anova/Pooled t**.
   See Figure 5.9 at right.
Chapter 5
Performing One-way Analysis
Means/Anova and Means/Anova/Pooled t

Figure 5.9 Examples of the Means/Anova Report and the Means/Anova/Pooled t Report

Both options contain the following reports:

- “The Summary of Fit Report,” p. 139
- “The Analysis of Variance Report,” p. 142
- “The Means for One-way Anova Report,” p. 143

The Means/Anova/Pooled t report window also contains a t Test report. See “The t-test Report,” p. 140.

The Summary of Fit Report

The Summary of Fit report shows a summary for a one-way analysis of variance, listing the quantities in Table 5.4.
Performing Oneway Analysis
Chapter 5
Means/Anova and Means/Anova/Pooled t

Table 5.4 Description of the Summary of Fit Report

<table>
<thead>
<tr>
<th>Description</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rsquare</td>
<td>Measures the proportion of the variation accounted for by fitting means to each factor level. The remaining variation is attributed to random error. The $R^2$ value is 1 if fitting the group means accounts for all the variation with no error. An $R^2$ of 0 indicates that the fit serves no better as a prediction model than the overall response mean. Using quantities from the Analysis of Variance report for the model, the $R^2$ for any continuous response fit is always calculated as follows: $R^2 = \frac{\text{Sum of Squares (Model)}}{\text{Sum of Squares (C Total)}}$. Note: $R^2$ is also called the coefficient of determination.</td>
</tr>
<tr>
<td>Adj Rsquare</td>
<td>Adjusts $R^2$ to make it more comparable over models with different numbers of parameters by using the degrees of freedom in its computation. It is a ratio of mean squares instead of sums of squares and is calculated as follows: $R^2_{adj} = 1 - \frac{\text{Mean Square (Error)}}{\text{Mean Square (C Total)}}$. The mean square for Error is found in the Analysis of Variance report and the mean square for C. Total can be computed as the C. Total Sum of Squares divided by its respective degrees of freedom (see “The Analysis of Variance Report,” p. 142).</td>
</tr>
<tr>
<td>Root Mean Square Error</td>
<td>Estimates the standard deviation of the random error. It is the square root of the mean square for Error found in the Analysis of Variance report.</td>
</tr>
<tr>
<td>Mean of Response</td>
<td>Overall mean (arithmetic average) of the response variable.</td>
</tr>
<tr>
<td>Observations (or Sum Wgts)</td>
<td>Number of observations used in estimating the fit. If weights are used, this is the sum of the weights.</td>
</tr>
</tbody>
</table>

The t-test Report

Note: This option is applicable only for the Means/Anova/Pooled t option.

There are two types of t-Tests:

- Equal variances. If you select the Means/Anova/Pooled t option, a t-Test report appears. This t-Test assumes equal variances.
• Unequal variances. If you select the t-Test option from the red triangle menu, a t-Test report appears. This t-Test assumes unequal variances.

**Example of T-test Reports**

1. Open the Popcorn.jmp sample data table.
2. Select Analyze > Fit Y by X.
4. Select batch and click X, Factor.
5. Click OK.
6. From the red triangle menu next to Oneway Analysis, select Means/Anova/Pooled t.
7. From the red triangle menu next to Oneway Analysis, select t Test.

In Figure 5.10, the first t Test assumes equal variances, and the second t Test assumes unequal variances.

**Figure 5.10 Examples of t-test Reports for Equal and Unequal Variances**
Performing Oneway Analysis
Means/Anova and Means/Anova/Pooled t

Table 5.5 Description of the \( t \)-Test Report

<table>
<thead>
<tr>
<th>Difference</th>
<th>Shows the estimated difference between the two X levels. In the plots, the Difference value appears as a red line that compares the two levels.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Std Err Dif</td>
<td>Shows the standard error of the difference.</td>
</tr>
<tr>
<td>Upper CL Dif</td>
<td>Shows the upper confidence limit for the difference.</td>
</tr>
<tr>
<td>Lower CL Dif</td>
<td>Shows the lower confidence limit for the difference.</td>
</tr>
<tr>
<td>Confidence</td>
<td>Shows the level of confidence (1-( \alpha )). To change the level of confidence, select a new alpha level from the <strong>Set a Level</strong> command from the platform red triangle menu.</td>
</tr>
<tr>
<td>( t ) Ratio</td>
<td>Value of the ( t )-statistic.</td>
</tr>
<tr>
<td>DF</td>
<td>The degrees of freedom used in the ( t )-test.</td>
</tr>
<tr>
<td>( \text{Prob} &gt;</td>
<td>t</td>
</tr>
<tr>
<td>( \text{Prob} &gt; t )</td>
<td>The ( p )-value associated with a lower-tailed test.</td>
</tr>
<tr>
<td>( \text{Prob} &lt; t )</td>
<td>The ( p )-value associated with an upper-tailed test.</td>
</tr>
</tbody>
</table>

The Analysis of Variance Report

The Analysis of Variance report partitions the total variation of a sample into two components. The ratio of the two mean squares forms the \( F \)-ratio. If the probability associated with the \( F \)-ratio is small, then the model is a better fit statistically than the overall response mean.

**Note:** If you specified a Block column, then the Analysis of Variance report includes the Block variable.

Table 5.6 Description of the Analysis of Variance Report

<table>
<thead>
<tr>
<th>Source</th>
<th>Lists the three sources of variation, which are the model source, <strong>Error</strong>, and <strong>C. Total</strong> (corrected total).</th>
</tr>
</thead>
<tbody>
<tr>
<td>DF</td>
<td>Records an associated degrees of freedom (DF for short) for each source of variation:</td>
</tr>
<tr>
<td></td>
<td>• The degrees of freedom for <strong>C. Total</strong> are ( N - 1 ), where ( N ) is the total number of observations used in the analysis.</td>
</tr>
<tr>
<td></td>
<td>• If the X factor has ( k ) levels, then the model has ( k - 1 ) degrees of freedom.</td>
</tr>
<tr>
<td></td>
<td>• The <strong>Error</strong> degrees of freedom is the difference between the <strong>C. Total</strong> degrees of freedom and the <strong>Model</strong> degrees of freedom (in other words, ( N - k )).</td>
</tr>
</tbody>
</table>
The Means for Oneway Anova Report

The Means for Oneway Anova report summarizes response information for each level of the nominal or ordinal factor.

<table>
<thead>
<tr>
<th>Table 5.7 Description of the Means for Oneway Anova Report</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Level</strong></td>
</tr>
<tr>
<td><strong>Number</strong></td>
</tr>
<tr>
<td><strong>Mean</strong></td>
</tr>
</tbody>
</table>
The Block Means Report

If you have specified a Block variable on the launch window, the Means/Anova and Means/Anova/Pooled t commands produce a Block Means report. This report shows the means for each block, and the number of observations in each block.

Mean Diamonds and X-Axis Proportional

A mean diamond illustrates a sample mean and confidence interval.

Example of Mean Diamonds and X-Axis Proportional Options

To create the plot at left in Figure 5.11, proceed as follows:
1. Open the Analgesics.jmp sample data table.
2. Select Analyze > Fit Y by X.
4. Select drug and click X, Factor.
5. Click OK.
6. To create the plot at right, from the red triangle menu, select Display Options.
7. Deselect X Axis proportional.
Chapter 5

Performing One-Way Analysis

Means/Anova and Means/Anova/Pooled t

Figure 5.11 Examples of Mean Diamonds and X-Axis Proportional Options

Note the following observations:

- The top and bottom of each diamond represent the (1-alpha)x100 confidence interval for each group. The confidence interval computation assumes that the variances are equal across observations. Therefore, the height of the diamond is proportional to the reciprocal of the square root of the number of observations in the group.

- If the X-Axis proportional option is selected, the horizontal extent of each group along the x-axis (the horizontal size of the diamond) is proportional to the sample size for each level of the X variable. Therefore, the narrower diamonds are usually taller, because fewer data points result in a wider confidence interval.

- The mean line across the middle of each diamond represents the group mean.

- Overlap marks appear as lines above and below the group mean. For groups with equal sample sizes, overlapping marks indicate that the two group means are not significantly different at the given confidence level. Overlap marks are computed as group mean ± (\sqrt{2}/2 × CI/2). Overlap marks in one diamond that are closer to the mean of another diamond than that diamond's overlap marks indicate that those two groups are not different at the given confidence level.

- The mean diamonds automatically appear when you select the Means/Anova/Pooled t or Means/Anova option from the platform menu. However, you can show or hide them at any time by selecting Display Options > Mean Diamonds from the red triangle menu.

Mean Lines, Error Bars, and Standard Deviation Lines

You can show mean lines by selecting Display Options > Mean Lines. Mean lines show the mean of the response for each level of the X variable.

Mean error bars and standard deviation lines appear when you select the Means and Std Dev option from the red triangle menu. See Figure 5.12. To turn each option on or off singly, select Display Options > Mean Error Bars or Std Dev Lines.
Performing Oneway Analysis

Analysis of Means Methods

Example of Mean Lines, Mean Error Bars, and Standard Deviation Lines

1. Open the Analgesics.jmp sample data table.
2. Select Analyze > Fit Y by X.
4. Select drug and click X, Factor.
5. Click OK.
6. From the red triangle menu, select Display Options > Mean Lines.
7. From the red triangle menu, select Means and Std Dev.

Figure 5.12 Examples of Mean Lines, Mean Error Bars, and Std Dev Lines

Note: For a description of ANOM methods, see the document by Nelson, Wludyka, and Copeland (2005). For a description of the specific ANOM for Variances method, see the paper by Wludyka and Nelson (1997).

Analysis of Means Methods

Note: For a description of ANOM methods, see the document by Nelson, Wludyka, and Copeland (2005). For a description of the specific ANOM for Variances method, see the paper by Wludyka and Nelson (1997).

Analysis of means (ANOM) methods compare means and variances across several groups. You might want to use these methods under these circumstances:

- if you want to test whether any of the group means are statistically different from the overall mean
- if you want to test whether any of the group standard deviations are statistically different from the root mean square error (RMSE)
Note: Within the Contingency platform, you can use the Analysis of Means for Proportions when the response has two categories. For details, see the “Performing Contingency Analysis” chapter.

**Compare Means**

Use the ANOM and ANOM with Transformed Ranks options to compare group means to the overall mean.

<table>
<thead>
<tr>
<th>Table 5.8 Descriptions of Methods for Comparing Means</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ANOM</strong></td>
</tr>
<tr>
<td>Compares group means to the overall mean. This method assumes that your data is approximately normally distributed.</td>
</tr>
<tr>
<td><strong>ANOM with Transformed Ranks</strong></td>
</tr>
<tr>
<td>This is the nonparametric version of the ANOM analysis. Use this method if your data is clearly non-normal and cannot be transformed to normality. Compares each group mean transformed rank to the overall mean transformed rank.</td>
</tr>
</tbody>
</table>

**Compare Standard Deviations (or Variances)**

Use the ANOM for Variances and ANOM for Variances with Levene (ADM) options to compare group standard deviations to the root mean squared error. This is a type of variance heterogeneity test.

<table>
<thead>
<tr>
<th>Table 5.9 Descriptions of Methods for Comparing Standard Deviations (or Variances)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ANOM for Variances</strong></td>
</tr>
<tr>
<td>Compares group standard deviations to the root mean squared error. This method assumes that your data is approximately normally distributed. To use this method, each group must have at least four observations.</td>
</tr>
<tr>
<td><strong>ANOM for Variances with Levene (ADM)</strong></td>
</tr>
<tr>
<td>This is the nonparametric version of the ANOM for Variances analysis. Use this method if you suspect your data is non-normal and cannot be transformed to normality. Compares the group means of the absolute deviation from the median (ADM) to the overall mean ADM.</td>
</tr>
</tbody>
</table>

**Example of an Analysis of Means for Variances Chart**

This example uses the Spring Data.jmp sample data table. Four different brands of springs were tested to see what weight is required to extend a spring 0.10 inches. Six springs of each brand were tested. The data was checked for normality, since the ANOMV test is not robust to non-normality. Examine the brands to determine whether the variability is significantly different between brands.

1. Open the Spring Data.jmp sample data table.
2. Select Analyze > Fit Y by X.
3. Select Weight and click Y, Response.
4. Select Brand and click X, Factor.
5. From the red triangle menu, select **Analysis of Means Methods > ANOM for Variances**.
6. From the red triangle menu next to Analysis of Means for Variances, select **Show Summary Report**.

**Figure 5.13** Example of Analysis of Means for Variances Chart

From Figure 5.13, notice that the standard deviation for Brand 2 exceeds the lower decision limit. Therefore, Brand 2 has significantly lower variance than the other brands.

**Analysis of Means Charts**

Each Analysis of Means Methods option adds a chart to the report window, that shows the following:

- a center line indicating the overall mean or root mean squared error (or MSE when in variance scale)
- upper decision limits (UDL)
- lower decision limits (LDL)

If a group mean falls outside of the decision limits, then that mean is significantly different from the overall mean. If a group standard deviation falls outside of the decision limits, then that standard deviation is significantly different from the root mean squared error.

**Example of an Analysis of Means Chart**

1. Open the Analgesics.jmp sample data table.
2. Select Analyze > Fit Y by X.
4. Select drug and click X, Factor.
5. Click OK.
6. From the red triangle menu, select Analysis of Means Methods > ANOM.

Figure 5.14 Example of Analysis of Means Chart

For the example in Figure 5.14, the means for drug A and C are statistically different from the overall mean. The drug A mean is lower and the drug C mean is higher. Note the decision limits for the drug types are not the same, due to different sample sizes.

Analysis of Means Options

Each Analysis of Means Methods option adds an Analysis of Means red triangle menu to the report window.

Table 5.10 Descriptions of Analysis of Means Options

<table>
<thead>
<tr>
<th>Set Alpha Level</th>
<th>Select an option from the most common alpha levels or specify any level with the Other selection. Changing the alpha level modifies the upper and lower decision limits.</th>
</tr>
</thead>
</table>
| Show Summary Report | • For ANOM, creates a report showing group means and decision limits.  
• For ANOM with Transformed Ranks, creates a report showing group mean transformed ranks and decision limits.  
• For ANOM for Variances, creates a report showing group standard deviations (or variances) and decision limits.  
• For ANOM for Variances with Levene (ADM), creates a report showing group mean ADMs and decision limits. |
Performing Oneway Analysis

Chapter 5

Compare Means

Note: Another method for comparing means is ANOM. See "Analysis of Means Methods," p. 146.

Use the Compare Means options to perform multiple comparisons of group means. Each Compare Means option adds comparison circles next to the plot, and specific reports to the report window.

Table 5.10 Descriptions of Analysis of Means Options (Continued)

<table>
<thead>
<tr>
<th>Graph in Variance Scale</th>
<th>(Only for ANOM for Variances) Changes the scale of the y-axis from standard deviations to variances.</th>
</tr>
</thead>
</table>

<table>
<thead>
<tr>
<th>Display Options</th>
<th>Display options include the following:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• Show Decision Limits shows or hides decision limit lines.</td>
</tr>
<tr>
<td></td>
<td>• Show Decision Limit Shading shows or hides decision limit shading.</td>
</tr>
<tr>
<td></td>
<td>• Show Center Line shows or hides the center line statistic.</td>
</tr>
<tr>
<td></td>
<td>• Show Needles shows or hides the needles.</td>
</tr>
</tbody>
</table>

Table 5.11 Descriptions of Compare Means Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
<th>Reference</th>
<th>Nonparametric Menu Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>Each Pair, Student’s t</td>
<td>Computes individual pairwise comparisons using Student’s t-tests. If you make many pairwise tests, there is no protection across the inferences. Therefore, the alpha-size (Type I error rate) across the hypothesis tests is higher than that for individual tests.</td>
<td>See “Using Comparison Circles,” p. 151 and “Each Pair, Student’s t,” p. 153.</td>
<td>Nonparametric &gt; Nonparametric Multiple Comparisons &gt; Wilcoxon Each Pair</td>
</tr>
</tbody>
</table>


Table 5.11 Descriptions of Compare Means Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
<th>Reference</th>
<th>Nonparametric Menu Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Pairs, Tukey HSD</td>
<td>Shows a test that is sized for all differences among the means. This is the Tukey or Tukey-Kramer HSD (honestly significant difference) test. (Tukey 1953, Kramer 1956). This test is an exact alpha-level test if the sample sizes are the same, and conservative if the sample sizes are different (Hayter 1984).</td>
<td>See “Using Comparison Circles,” p. 151 and “All Pairs, Tukey HSD,” p. 155.</td>
<td>Nonparametric &gt; Nonparametric Multiple Comparisons &gt; Steel-Dwass All Pairs</td>
</tr>
<tr>
<td>With Best, Hsu MCB</td>
<td>Tests whether the means are less than the unknown maximum or greater than the unknown minimum. This is the Hsu MCB test (Hsu 1981).</td>
<td>See “Using Comparison Circles,” p. 151 and “With Best, Hsu MCB,” p. 157.</td>
<td>none</td>
</tr>
<tr>
<td>With Control, Dunnett's</td>
<td>Tests whether the means are different from the mean of a control group. This is Dunnett's test (Dunnett 1955).</td>
<td>See “Using Comparison Circles,” p. 151 and “With Control, Dunnett’s,” p. 159.</td>
<td>Nonparametric &gt; Nonparametric Multiple Comparisons &gt; Steel With Control</td>
</tr>
</tbody>
</table>

**Note:** If you have specified a Block column, then the multiple comparison methods are performed on data that has been adjusted for the Block means.

**Using Comparison Circles**

Each multiple comparison test begins with a comparison circles plot, which is a visual representation of group mean comparisons. Figure 5.15 shows the comparison circles for the All Pairs, Tukey HSD method. Other comparison tests lengthen or shorten the radii of the circles.

**Example of Comparison Circles**

1. Open the Big Class.jmp sample data table.
2. Select **Analyze** > **Fit Y by X**.
3. Select **height** and click **Y, Response**.
4. Select **age** and click **X, Factor**.
Performing Oneway Analysis

Chapter 5

Compare Means

5. Click **OK**.

6. From the red triangle menu, select **Compare Means > All Pairs, Tukey HSD**.

---

**Figure 5.15** Visual Comparison of Group Means

Compare each pair of group means visually by examining the intersection of the comparison circles. The outside angle of intersection tells you whether the group means are significantly different. See Figure 5.16.

- Circles for means that are significantly different either do not intersect, or intersect slightly, so that the outside angle of intersection is less than 90 degrees.
- If the circles intersect by an angle of more than 90 degrees, or if they are nested, the means are not significantly different.

**Figure 5.16** Angles of Intersection and Significance

- **Angle greater than 90 degrees**
- **Angle equal to 90 degrees**
- **Angle less than 90 degrees**

- **Not significantly different**
- **Borderline significantly different**
- **Significantly different**
If the intersection angle is close to 90 degrees, you can verify whether the means are significantly different by clicking on the comparison circle to select it. See Figure 5.17. To deselect circles, click in the white space outside the circles.

**Figure 5.17** Highlighting Comparison Circles

For more information about comparison circles, see “Statistical Details,” p. 182.

**Each Pair, Student's t**

The **Each Pair, Student's t** test shows the Student's $t$-test for each pair of group levels and tests only individual comparisons.

**Example of the Each Pair, Student's t Test**

This example uses the Big Class.jmp sample data table. It shows a one-way layout of weight by age, and shows the group comparison using comparison circles that illustrate all possible $t$-tests.

1. Open the Big Class.jmp sample data table.
2. Select Analyze > Fit Y by X.
3. Select weight and click Y, Response.
4. Select age and click X, Factor.
5. Click OK.
6. From the red triangle menu, select Compare Means > Each Pair, Student's t.
The means comparison method can be thought of as seeing if the actual difference in the means is greater than the difference that would be significant. This difference is called the LSD (least significant difference). The LSD term is used for Student’s $t$ intervals and in context with intervals for other tests. In the comparison circles graph, the distance between the circles’ centers represent the actual difference. The LSD is what the distance would be if the circles intersected at right angles.
Performing Oneway Analysis

Chapter 5
Compare Means

155

Figure 5.19 Example of Means Comparisons Report for Each Pair, Student’s t

In Figure 5.19, the LSD threshold table shows the difference between the absolute difference in the means and the LSD (least significant difference). If the values are positive, the difference in the two means is larger than the LSD, and the two groups are significantly different.

All Pairs, Tukey HSD

The All Pairs, Tukey HSD test (also called Tukey-Kramer) protects the significance tests of all combinations of pairs, and the HSD intervals become greater than the Student’s t pairwise LSDs. Graphically, the comparison circles become larger and differences are less significant.

Example of the All Pairs, Tukey HSD Test

1. Open the Big Class.jmp sample data table.
2. Select Analyze > Fit Y by X.
3. Select weight and click Y, Response.
4. Select age and click X, Factor.
5. Click OK.
Performing Oneway Analysis

Chapter 5

Compare Means

6. From the red triangle menu, select Compare Means > All Pairs, Tukey HSD.

---

Figure 5.20 Example of All Pairs, Tukey HSD Comparison Circles

Figure 5.21 Example of Means Comparisons Report for All Pairs, Tukey HSD

---

### Figure 5.20 Example of All Pairs, Tukey HSD Comparison Circles

### Figure 5.21 Example of Means Comparisons Report for All Pairs, Tukey HSD

---

### Means Comparisons

**Comparisons for all pairs using Tukey-Kramer HSD**

<table>
<thead>
<tr>
<th>Level</th>
<th>Level</th>
<th>Difference</th>
<th>StdErr Diff</th>
<th>Lower CL</th>
<th>Upper CL</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>16</td>
<td>4.9523</td>
<td>4.3651</td>
<td>9.5862</td>
<td>0.0233</td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>15</td>
<td>9.2181</td>
<td>9.2579</td>
<td>18.4760</td>
<td>0.0233</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>15</td>
<td>9.2640</td>
<td>9.2579</td>
<td>18.4760</td>
<td>0.0233</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>14</td>
<td>4.9523</td>
<td>4.3651</td>
<td>9.5862</td>
<td>0.0233</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>13</td>
<td>9.2181</td>
<td>9.2579</td>
<td>18.4760</td>
<td>0.0233</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>13</td>
<td>4.9523</td>
<td>4.3651</td>
<td>9.5862</td>
<td>0.0233</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>14</td>
<td>9.2181</td>
<td>9.2579</td>
<td>18.4760</td>
<td>0.0233</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>13</td>
<td>4.9523</td>
<td>4.3651</td>
<td>9.5862</td>
<td>0.0233</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>12</td>
<td>9.2181</td>
<td>9.2579</td>
<td>18.4760</td>
<td>0.0233</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>12</td>
<td>4.9523</td>
<td>4.3651</td>
<td>9.5862</td>
<td>0.0233</td>
<td></td>
</tr>
</tbody>
</table>

Positive values show pairs of means that are significantly different.
In Figure 5.21, the Tukey-Kramer HSD Threshold matrix shows the actual absolute difference in the means minus the HSD, which is the difference that would be significant. Pairs with a positive value are significantly different. Because the borders of the table are sorted by the mean, the more significant differences tend toward the upper right, or symmetrically, the lower left. The band near the diagonal tends to be negative because the means are closer. The q* (shown above the HSD Threshold Matrix table) is the quantile that is used to scale the HSDs. It has a computational role comparable to a Student’s $t$.

**With Best, Hsu MCB**

The With Best, Hsu MCB test determines whether a mean can be rejected as the maximum or minimum among the means. The Hsu’s MCB test is given for each level versus the maximum, and each level versus the minimum.

The quantile that scales the LSD is not the same for each level. Therefore, the comparison circles technique does not work as well, because the circles must be rescaled according to the level being tested (unless the sample sizes are equal). The comparison circles plot uses the largest quantiles value shown to make the circles. Use the p-values of the tests to obtain precise assessments of significant differences.

**Example of the With Best, Hsu MCB Test**

1. Open the Big Class.jmp sample data table.
2. Select Analyze > Fit Y by X.
3. Select weight and click Y, Response.
4. Select age and click X, Factor.
5. Click OK.
6. From the red triangle menu, select Compare Means > With Best, Hsu MCB.

![Figure 5.22 Examples of With Best, Hsu MCB Comparison Circles](image-url)
Performing One-Way Analysis

Chapter 5

Compare Means

Note the following:

- If a mean has means significantly separated above it, it is not regarded as the maximum.
- If a mean has means significantly separated below it, it is not regarded as the minimum.
- If a mean is significantly separated above all other means, it is regarded as the maximum.
- If a mean is significantly separated below all other means, it is regarded as the minimum.

**Note:** Means that are not regarded as the maximum or the minimum by MCB are also the means that are not contained in the selected subset of Gupta (1965) of potential maximums or minimum means.

**Figure 5.23** Example of Means Comparisons Report for With Best, Hsu MCB

<table>
<thead>
<tr>
<th>Means Comparisons</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Comparisons with the best using Hsu’s MCB</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>vs. Max</th>
<th>vs. Min</th>
</tr>
</thead>
<tbody>
<tr>
<td>p-Value</td>
<td>p-Value</td>
</tr>
<tr>
<td>17</td>
<td>0.0207*</td>
</tr>
<tr>
<td>16</td>
<td>0.1208</td>
</tr>
<tr>
<td>15</td>
<td>0.0241*</td>
</tr>
<tr>
<td>14</td>
<td>0.0150*</td>
</tr>
<tr>
<td>13</td>
<td>0.0080*</td>
</tr>
</tbody>
</table>

LSD threshold matrix for maximum

LSD threshold matrix for minimum

For the **maximum** report, a column shows the row mean minus the column mean minus the LSD. If a value is positive, the row mean is significantly higher than the mean for the column, and the mean for the column is not the maximum.

For the **minimum** report, a column shows the row mean minus the column mean plus the LSD. If a value is negative, the row mean is significantly less than the mean for the column, and the mean for the column is not the minimum.
With Control, Dunnett’s

The With Control, Dunnett’s test compares a set of means against the mean of a control group. The LSDs that it produces are between the Student’s $t$ and Tukey-Kramer LSDs, because they are sized to guard against an intermediate number of comparisons.

Example of the With Control, Dunnett’s Test

1. Open the Big Class.jmp sample data table.
2. Select Analyze > Fit Y by X.
3. Select weight and click Y, Response.
4. Select age and click X, Factor.
5. Click OK.
6. From the red triangle menu, select Compare Means > With Control, Dunnett’s.
7. Select the group to use as the control group. In this example, select age 12.
   Alternatively, click on a row to highlight it in the scatterplot before selecting the Compare Means > With Control, Dunnett’s option. The test uses the selected row as the control group.
8. Click OK.

Figure 5.24 Example of With Control, Dunnett’s Comparison Circles

Figure 5.24 shows the comparisons circles plot of Dunnett’s test, with the control group of age 12 highlighted. Circles that are gray represent groups that are different from the control group.
In the Dunnett’s report, the $|d|$ quantile appears, and can be used in a manner similar to a Student’s $t$-statistic. The LSD threshold matrix shows the absolute value of the difference minus the LSD. If a value is positive, its mean is more than the LSD apart from the control group mean and is therefore significantly different.

**Compare the Four Tests**

1. Open the Big Class.jmp sample data table.
2. Select Analyze > Fit Y by X.
3. Select weight and click Y, Response.
4. Select age and click X, Factor.
5. Click OK.
6. From the red triangle menu, select each one of the Compare Means options.

Although the four methods all test differences between group means, different results can occur. Figure 5.26 shows the comparison circles for all four tests, with the age 17 group as the control group.
From Figure 5.26, notice that for the Student’s t and Hsu methods, age group 15 (the third circle from the top) is significantly different from the control group and appears gray. But, for the Tukey and Dunnett method, age group 15 is not significantly different, and appears red.

**Compare Means Options**

The Means Comparison reports for all four tests contain a red triangle menu with customization options. All of these methods use pooled variance estimates for the means.
Performing Oneway Analysis
Nonparametric

Nonparametric tests are useful for testing whether group means or medians are located the same across groups. However, the usual analysis of variance assumption of normality is not made. Nonparametric tests use functions of the response ranks, called rank scores (Hajek 1969).

Table 5.12 Descriptions of Compare Means Options

<table>
<thead>
<tr>
<th>Description</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference Matrix</td>
<td>Shows a table of all differences of means.</td>
</tr>
<tr>
<td>Confid Quantile</td>
<td>Shows the t-value or other corresponding quantiles used for confidence intervals.</td>
</tr>
<tr>
<td>LSD Threshold Matrix</td>
<td>Shows a matrix showing if a difference exceeds the least significant difference for all comparisons.</td>
</tr>
<tr>
<td>Connecting Letters Report</td>
<td>Shows the traditional letter-coded report where means that are not sharing a letter are significantly different.</td>
</tr>
<tr>
<td>Ordered Differences Report</td>
<td>Shows all the positive-side differences with their confidence interval in sorted order. For the Student’s t and Tukey-Kramer comparisons, an Ordered Difference report appears below the text reports. This report shows the ranked differences, from highest to lowest, with a confidence interval band overlaid on the plot. Confidence intervals that do not fully contain their corresponding bar are significantly different from each other.</td>
</tr>
<tr>
<td>Detailed Comparisons Report</td>
<td>Shows a detailed report for each comparison. Each section shows the difference between the levels, standard error and confidence intervals, t-ratios, p-values, and degrees of freedom. A plot illustrating the comparison appears on the right of each report.</td>
</tr>
</tbody>
</table>

a. For Hsu’s MCB and Dunnett’s test, only Difference Matrix, Confid Quantile, and LSD Threshold Matrix are applicable.

b. The Detailed Comparisons Report option is not available for All Pairs, Tukey’s HSD and Nonparametric Multiple Comparisons.

Nonparametric tests are useful for testing whether group means or medians are located the same across groups. However, the usual analysis of variance assumption of normality is not made. Nonparametric tests use functions of the response ranks, called rank scores (Hajek 1969).

Table 5.13 Types of Nonparametric Tests

<table>
<thead>
<tr>
<th>Test</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wilcoxon Test</td>
<td>Performs the test based on Wilcoxon rank scores. The Wilcoxon rank scores are the simple ranks of the data. The Wilcoxon test is the most powerful rank test for errors with logistic distributions. If the factor has two or more levels, the Kruskal-Wallis test is performed. The Wilcoxon test is also called the Mann-Whitney test.</td>
</tr>
</tbody>
</table>
Table 5.13  Types of Nonparametric Tests  *(Continued)*

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median Test</td>
<td>Performs the test based on Median rank scores. The Median rank scores are either 1 or 0, depending on whether a rank is above or below the median rank. The Median test is the most powerful rank test for errors with double-exponential distributions.</td>
</tr>
<tr>
<td>van der Waerden Test</td>
<td>Performs the test based on Van der Waerden rank scores. The Van der Waerden rank scores are the ranks of the data divided by one, plus the number of observations transformed to a normal score by applying the inverse of the normal distribution function. The Van der Waerden test is the most powerful rank test for errors with normal distributions.</td>
</tr>
<tr>
<td>Kolmogorov Smirnov Test</td>
<td>Performs the test based on the empirical distribution function, which tests whether the distribution of the response is the same across the groups. Both an approximate and an exact test are given. This test is available only when the X factor has two levels.</td>
</tr>
<tr>
<td>Exact Test</td>
<td>Provides options for performing exact versions of the Wilcoxon, Median, van der Waerden, and Kolmogorov Smirnov tests. These options are available only when the X factor has two levels, and after the approximate test is requested.</td>
</tr>
</tbody>
</table>
Nonparametric Report Descriptions

All nonparametric reports are described in the following tables:

- Table 5.14 “Descriptions of the Wilcoxon, Median, and Van der Waerden Tests,” p. 164
- Table 5.15 “Description of the Kolmogorov Smirnov Test,” p. 165
- Table 5.16 “Description of the Nonparametric Multiple Comparisons Tests,” p. 166

### Table 5.14 Descriptions of the Wilcoxon, Median, and Van der Waerden Tests

<table>
<thead>
<tr>
<th>Level</th>
<th>Lists the factor levels occurring in the data.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>Records the frequencies of each level.</td>
</tr>
</tbody>
</table>
Table 5.14 Descriptions of the Wilcoxon, Median, and Van der Waerden Tests

<table>
<thead>
<tr>
<th>Description</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Score Sum</td>
<td>Records the sum of the rank score for each level.</td>
</tr>
<tr>
<td>Expected Score</td>
<td>Records the expected score under the null hypothesis that there is no difference among class levels.</td>
</tr>
<tr>
<td>Score Mean</td>
<td>Records the mean rank score for each level.</td>
</tr>
<tr>
<td>(Mean-Mean0)/Std0</td>
<td>Records the standardized score. Mean0 is the mean score expected under the null hypothesis. Std0 is the standard deviation of the score sum expected under the null hypothesis. The null hypothesis is that the group means or medians are in the same location across groups.</td>
</tr>
<tr>
<td>ChiSquare</td>
<td>Gives the values of the chi-square test statistic.</td>
</tr>
<tr>
<td>DF</td>
<td>Gives the degrees of freedom for the test.</td>
</tr>
<tr>
<td>Prob&gt;</td>
<td>ChiSq</td>
</tr>
<tr>
<td>S</td>
<td>Gives the sum of the rank scores. This is reported only when the X factor has two levels.</td>
</tr>
<tr>
<td>Z</td>
<td>Gives the test statistic for the normal approximation test. This is reported only when the X factor has two levels.</td>
</tr>
<tr>
<td>Prob&gt;</td>
<td></td>
</tr>
<tr>
<td>Prob&gt;</td>
<td></td>
</tr>
<tr>
<td>Note: Exact tests are available only in JMP Pro.</td>
<td></td>
</tr>
<tr>
<td>Prob&gt;</td>
<td></td>
</tr>
<tr>
<td>Note: Exact tests are available only in JMP Pro.</td>
<td></td>
</tr>
</tbody>
</table>

Table 5.15 Description of the Kolmogorov Smirnov Test

<table>
<thead>
<tr>
<th>Description</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
<td>Lists the factor levels occurring in the data.</td>
</tr>
<tr>
<td>Count</td>
<td>Records the frequencies of each level.</td>
</tr>
<tr>
<td>EDF at Maximum</td>
<td>Lists the value at which the maximum deviation from the empirical distribution function (EDF) of each level and the overall EDF occurs.</td>
</tr>
<tr>
<td>Deviation from Mean at Maximum</td>
<td>Lists the value of the EDF of a sample at the maximum deviation from the mean of the EDF for the overall sample.</td>
</tr>
</tbody>
</table>
Performing Oneway Analysis

Chapter 5

Nonparametric

Table 5.15 Description of the Kolmogorov Smirnov Test

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>KS</td>
<td>Measures the maximum deviation of the EDF within the classes from the pooled EDF. A Kolmogorov-Smirnov statistic.</td>
</tr>
<tr>
<td>KSa</td>
<td>Lists the asymptotic statistics that measure the maximum deviation of the EDF within the classes from pooled EDF. An asymptotic Kolmogorov-Smirnov statistic.</td>
</tr>
<tr>
<td>(D = \max</td>
<td>F_1 - F_2</td>
</tr>
<tr>
<td>Prob &gt; D</td>
<td>Lists the p-value for the test. In other words, the probability that (D) is greater than the observed value (d), under the null hypothesis of no difference between class levels or samples.</td>
</tr>
<tr>
<td>(D^+ = \max(F_1 - F_2))</td>
<td>Lists a one-sided test statistic that max deviation between the EDF of two class levels is positive.</td>
</tr>
<tr>
<td>Prob &gt; D+</td>
<td>Lists the probability that (D^+) is greater than the observed value (d^+), under the null hypothesis of no difference between the two class levels.</td>
</tr>
<tr>
<td>(D^- = \max(F_2 - F_1))</td>
<td>Lists a one-sided test statistic that max deviation between the EDF of two class levels is negative.</td>
</tr>
<tr>
<td>Prob &gt; D-</td>
<td>Lists the probability that (D^-) is greater than the observed value for (d^-).</td>
</tr>
</tbody>
</table>

Table 5.16 Description of the Nonparametric Multiple Comparisons Tests

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(q^*)</td>
<td>Gives the quantile value used in the confidence intervals.</td>
</tr>
<tr>
<td>Alpha</td>
<td>Gives the alpha level used in the confidence intervals</td>
</tr>
<tr>
<td>Level</td>
<td>Gives the pair used in the current comparison</td>
</tr>
<tr>
<td>Score Mean Diff</td>
<td>Gives the difference of the score means.</td>
</tr>
<tr>
<td>Std Err Dif</td>
<td>Gives the standard error of the difference between the score means.</td>
</tr>
<tr>
<td>Z</td>
<td>Gives the standardized test statistic, which has an asymptotic standard normal deviation under the null hypothesis.</td>
</tr>
<tr>
<td>p-Value</td>
<td>Gives the asymptotic two-sided p-value for (Z).</td>
</tr>
<tr>
<td>Hodges-Lehmann</td>
<td>Gives the Hodges-Lehmann estimator of location shift. It is the median of all paired differences between observations in the two samples.</td>
</tr>
<tr>
<td>Lower CL</td>
<td>Gives the lower confidence limit for the Hodges-Lehmann statistic.</td>
</tr>
<tr>
<td>Upper CL</td>
<td>Gives the upper confidence limit for the Hodges-Lehmann statistic.</td>
</tr>
</tbody>
</table>
Unequal Variances

When the variances across groups are not equal, the usual analysis of variance assumptions are not satisfied and the ANOVA $F$-test is not valid. JMP provides four tests for equality of group variances and an ANOVA that is valid when the group sample variances are unequal. The concept behind the first three tests of equal variances is to perform an analysis of variance on a new response variable constructed to measure the spread in each group. The fourth test is Bartlett’s test, which is similar to the likelihood-ratio test under normal distributions.

Note: Another method to test for unequal variances is ANOMV. See “Analysis of Means Methods,” p. 146.

Table 5.17 Descriptions of Unequal Variance Tests

<table>
<thead>
<tr>
<th>Test</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>O’Brien</td>
<td>Constructs a dependent variable so that the group means of the new variable equal the group sample variances of the original response. An ANOVA on the O’Brien variable is actually an ANOVA on the group sample variances (O’Brien 1979, Olejnik, and Algina 1987).</td>
</tr>
<tr>
<td>Brown-Forsythe</td>
<td>Shows the $F$-test from an ANOVA where the response is the absolute value of the difference of each observation and the group median (Brown and Forsythe 1974a).</td>
</tr>
<tr>
<td>Levene</td>
<td>Shows the $F$-test from an ANOVA where the response is the absolute value of the difference of each observation and the group mean (Levene 1960). The spread is measured as $z_{ij} =</td>
</tr>
<tr>
<td>Bartlett</td>
<td>Compares the weighted arithmetic average of the sample variances to the weighted geometric average of the sample variances. The geometric average is always less than or equal to the arithmetic average with equality holding only when all sample variances are equal. The more variation there is among the group variances, the more these two averages differ. A function of these two averages is created, which approximates a $\chi^2$-distribution (or, in fact, an $F$-distribution under a certain formulation). Large values correspond to large values of the arithmetic or geometric ratio, and therefore to widely varying group variances. Dividing the Bartlett Chi-square test statistic by the degrees of freedom gives the $F$-value shown in the table. Bartlett’s test is not very robust to violations of the normality assumption (Bartlett and Kendall 1946).</td>
</tr>
</tbody>
</table>

If there are only two groups tested, then a standard $F$-test for unequal variances is also performed. The $F$-test is the ratio of the larger to the smaller variance estimate. The $p$-value from the $F$-distribution is doubled to make it a two-sided test.
Example of the Unequal Variances Option

1. Open the Big Class.jmp sample data table.
2. Select Analyze > Fit Y by X.
4. Select sex and click X, Factor.
5. Click OK.
6. From the red triangle menu, select Unequal Variances.

![Figure 5.27 Example of the Unequal Variances Report]

The Tests that the Variances are Equal report shows the differences between group means to the grand mean and to the median, and gives a summary of testing procedures. See Table 5.18.
If the equal variances test reveals that the group variances are significantly different, use Welch's test instead of the regular ANOVA test. The Welch statistic is based on the usual ANOVA $F$-test. However, the means are weighted by the reciprocal of the group mean variances (Welch 1951; Brown and Forsythe 1974b; Asiribo, Osebekwin, and Gurland 1990). If there are only two levels, the Welch ANOVA is equivalent to an unequal variance $t$-test. See Table 5.19.

Table 5.18 Descriptions of the Tests that the Variances are Equal Report

<table>
<thead>
<tr>
<th>Level</th>
<th>Lists the factor levels occurring in the data.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>Records the frequencies of each level.</td>
</tr>
<tr>
<td>Std Dev</td>
<td>Records the standard deviations of the response for each factor level. The standard deviations are equal to the means of the O’Brien variable. If a level occurs only once in the data, no standard deviation is calculated.</td>
</tr>
<tr>
<td>MeanAbsDif to Mean</td>
<td>Records the mean absolute difference of the response and group mean. The mean absolute differences are equal to the group means of the Levene variable.</td>
</tr>
<tr>
<td>MeanAbsDif to Median</td>
<td>Records the absolute difference of the response and group median. The mean absolute differences are equal to the group means of the Brown-Forsythe variable.</td>
</tr>
<tr>
<td>Test</td>
<td>Lists the names of the tests performed.</td>
</tr>
</tbody>
</table>
### Table 5.18 Descriptions of the Tests that the Variances are Equal Report (Continued)

<table>
<thead>
<tr>
<th>Test</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>F Ratio</strong></td>
<td>Records the following calculated F-statistic for each test:</td>
</tr>
<tr>
<td></td>
<td>O’Brien’s test constructs a dependent variable so that the group means of the new variable equal the group sample variances of the original response. The O’Brien variable is computed as follows:</td>
</tr>
<tr>
<td></td>
<td>$$r_{ijk} = \frac{(n_{ij} - 1.5)n_{ij}(y_{ijk} - \bar{y}<em>{ij})^2 - 0.5r</em>{ij}^2(n_{ij} - 1)}{(n_{ij} - 1)(n_{ij} - 2)}$$</td>
</tr>
<tr>
<td></td>
<td>where $n$ represents the number of $y_{ijk}$ observations.</td>
</tr>
<tr>
<td></td>
<td>Brown-Forsythe is the model F-statistic from an ANOVA on $z_{ij} =</td>
</tr>
<tr>
<td></td>
<td>The Levene $F$ is the model F-statistic from an ANOVA on $z_{ij} =</td>
</tr>
<tr>
<td></td>
<td>Bartlett’s test is calculated as follows:</td>
</tr>
<tr>
<td></td>
<td>$$T = \frac{\sum_i v_i \log \left( \frac{1}{v_i} \sum_j \frac{s_{ij}}{v_i} \right) - \sum_i v_i \log (s_{ij}^2)}{1 + \left( \frac{\sum_i \frac{1}{v_i} - 1}{3(k-1)} \right)}$$</td>
</tr>
<tr>
<td></td>
<td>where $v_i = n_i - 1$ and $v = \sum v_i$</td>
</tr>
<tr>
<td></td>
<td>and $n_i$ is the count on the $i$th level and $s_{ij}^2$ is the response sample variance on the $i$th level. The Bartlett statistic has a $\chi^2$-distribution. Dividing the Chi-square test statistic by the degrees of freedom results in the reported $F$-value.</td>
</tr>
<tr>
<td><strong>DFNum</strong></td>
<td>Records the degrees of freedom in the numerator for each test. If a factor has $k$ levels, the numerator has $k - 1$ degrees of freedom. Levels occurring only once in the data are not used in calculating test statistics for O’Brien, Brown-Forsythe, or Levene. The numerator degrees of freedom in this situation is the number of levels used in calculations minus one.</td>
</tr>
<tr>
<td><strong>DFDen</strong></td>
<td>Records the degrees of freedom used in the denominator for each test. For O’Brien, Brown-Forsythe, and Levene, a degree of freedom is subtracted for each factor level used in calculating the test statistic. One more degree of freedom is subtracted for the overall mean. If a factor has $k$ levels, the denominator degrees of freedom is $n - k - 1$.</td>
</tr>
<tr>
<td><strong>p-Value</strong></td>
<td>Probability of obtaining, by chance alone, an $F$-value larger than the one calculated if in reality the variances are equal across all levels.</td>
</tr>
</tbody>
</table>
Table 5.19 Description of Welch’s Test Report

| F Ratio | Computed as follows:
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$F = \frac{\sum_{i} w_i (\bar{y}_i - \bar{y})^2}{1 + \frac{2(k - 2)}{k^2 - 1} \left( \sum \frac{w_i^2}{u} \right) - \frac{3}{k^2 - 1} \sum \frac{w_i^2}{n_i - 1}}$</td>
<td>where $w_i = \frac{n_i}{2}$, $u = \sum w_i$, $\bar{y}_i$ is the mean response for the $i$th level, and $s_i^2$ is the response sample variance for the $i$th level.</td>
</tr>
</tbody>
</table>

| DFNum | Records the degrees of freedom in the numerator of the test. If a factor has $k$ levels, the numerator has $k - 1$ degrees of freedom. Levels occurring only once in the data are not used in calculating the Welch ANOVA. The numerator degrees of freedom in this situation is the number of levels used in calculations minus one. |

| DFDen | Records the degrees of freedom in the denominator of the test. The Welch approximation for the denominator degrees of freedom is as follows: $df = \frac{1}{\left( \frac{3}{k^2 - 1} \right) \sum \frac{w_i^2}{u} - \frac{3}{k^2 - 1} \sum \frac{w_i^2}{n_i - 1}}$ where $w_i$, $n_i$ and $u$ are defined as in the $F$-ratio formula. |

| Prob>F | Probability of obtaining, by chance alone, an $F$-value larger than the one calculated if in reality the means are equal across all levels. Observed significance probabilities of 0.05 or less are considered evidence of unequal means across the levels. |

| t Test | Shows the relationship between the $F$ ratio and the $t$ Test. Calculated as the square root of the $F$ ratio. Appears only if the $X$ factor has two levels. |

**Note:** If you have specified a Block column, then the variance tests are performed on data after it has been adjusted for the Block means.
Equivalence Test

Equivalence tests assess whether there is a practical difference in means. You must pick a threshold difference for which smaller differences are considered practically equivalent. The most straightforward test to construct uses two one-sided $t$-tests from both sides of the difference interval. If both tests reject (or conclude that the difference in the means differs significantly from the threshold), then the groups are practically equivalent. The **Equivalence Test** option uses the Two One-Sided Tests (TOST) approach.

**Example of an Equivalence Test**

This example uses the *Big Class.jmp* sample data table. Examine if the difference in height between males and females is less than 6 inches.

1. Open the *Big Class.jmp* sample data table.
2. Select **Analyze > Fit Y by X**.
3. Select **height** and click **Y, Response**.
4. Select **sex** and click **X, Factor**.
5. Click **OK**.
6. From the red triangle menu, select **Equivalence Test**.
7. Type 6 as the difference considered practically zero.
8. Click **OK**.

**Figure 5.28 Example of an Equivalence Test**
From Figure 5.28, notice the following:

- The Upper Threshold test compares the actual difference to 6.
- The Lower Threshold test compares the actual difference to -6.
- For both tests, the p-value is small, therefore you can conclude that the actual difference in means (3.02) is significantly different from 6 and -6. For your purposes, you can declare the means to be practically equivalent.

**Power**

The Power option calculates statistical power and other details about a given hypothesis test. (Note: Examples use Typing Data.jmp)

- **LSV** (the Least Significant Value) is the value of some parameter or function of parameters that would produce a certain $p$-value alpha. Said another way, you want to know how small an effect would be declared significant at some $p$-value alpha. The LSV provides a measuring stick for significance on the scale of the parameter, rather than on a probability scale. It shows how sensitive the design and data are.

- **LSN** (the Least Significant Number) is the total number of observations that would produce a specified $p$-value alpha given that the data has the same form. The LSN is defined as the number of observations needed to reduce the variance of the estimates enough to achieve a significant result with the given values of alpha, sigma, and delta (the significance level, the standard deviation of the error, and the effect size). If you need more data to achieve significance, the LSN helps tell you how many more. The LSN is the total number of observations that yields approximately 50% power.

- **Power** is the probability of getting significance ($p$-value < alpha) when a real difference exists between groups. It is a function of the sample size, the effect size, the standard deviation of the error, and the significance level. The power tells you how likely your experiment is to detect a difference (effect size), at a given alpha level.

**Note:** When there are only two groups in a one-way layout, the LSV computed by the power facility is the same as the least significant difference (LSD) shown in the multiple-comparison tables.

**Example of the Power Option**

1. Open the Typing Data.jmp sample data table.
2. Select **Analyze > Fit Y by X.**
3. Select **speed** and click **Y, Response.**
4. Select **brand** and click **X, Factor.**
5. Click **OK.**
6. From the red triangle menu, select **Power.**
7. Within the From row, type 6 for Delta (the third box).
8. Within the To row, type 2 for Delta, and 11 in the Number box.
Performing Oneway Analysis

Chapter 5

Power

9. Within the By row, type 2 for both Delta and Number.

10. Select the Solve for Power check box.

---

**Figure 5.29** Example of the Power Details Window

11. Click **Done**.

12. From the floating red triangle menu, select **Power Plot**.

This option plots power by the Delta and Number values that you entered.
Figure 5.30 Example of the Power Report

Note that Power is computed and plotted for each combination of Delta and Number. As you might expect, the power rises for larger Number values and for larger Delta values.

For each of four columns Alpha, Sigma, Delta, and Number, fill in a single value, two values, or the start, stop, and increment for a sequence of values. See Figure 5.29. Power calculations are performed on all possible combinations of the values that you specify.

Table 5.20 Description of the Power Details Window

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alpha ($\alpha$)</td>
<td>Significance level, between 0 and 1 (usually 0.05, 0.01, or 0.10). Initially, a value of 0.05 shows.</td>
</tr>
<tr>
<td>Sigma ($\sigma$)</td>
<td>Standard error of the residual error in the model. Initially, RMSE, the estimate from the square root of the mean square error is supplied here.</td>
</tr>
<tr>
<td>Delta ($\delta$)</td>
<td>Raw effect size. For details about effect size computations, see the Modeling and Multivariate Methods book. The first position is initially set to the square root of the sums of squares for the hypothesis divided by $n$; that is, $\delta = \sqrt{\text{SS}/n}$</td>
</tr>
<tr>
<td>Number ($n$)</td>
<td>Total sample size across all groups. Initially, the actual sample size is put in the first position.</td>
</tr>
<tr>
<td>Solve for Power</td>
<td>Solves for the power (the probability of a significant result) as a function of all four values: $\alpha$, $\sigma$, $\delta$, and $n$.</td>
</tr>
</tbody>
</table>
Normal Quantile Plot

You can create two types of normal quantile plots:

- **Plot Actual by Quantile** creates a plot of the response values versus the normal quantile values. The quantiles are computed and plotted separately for each level of the X variable.

- **Plot Quantile by Actual** creates a plot of the normal quantile values versus the response values. The quantiles are computed and plotted separately for each level of the X variable.

The **Line of Fit** option shows or hides the lines of fit on the quantile plots.

### Example of a Normal Quantile Plot

1. Open the *Big Class.jmp* sample data table.
2. Select **Analyze > Fit Y by X**.
3. Select **height** and click **Y, Response**.
4. Select **sex** and click **X, Factor**.
5. Click **OK**.
6. From the red triangle menu, select **Normal Quantile Plot > Plot Actual by Quantile**.

---

**Table 5.20 Description of the Power Details Window**

<table>
<thead>
<tr>
<th>Solve for Least Significant Number</th>
<th>Solves for the number of observations needed to achieve approximately 50% power given $\alpha$, $\sigma$, and $\delta$.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solve for Least Significant Value</td>
<td>Solves for the value of the parameter or linear test that produces a $p$-value of $\alpha$. This is a function of $\alpha$, $\sigma$, $n$, and the standard error of the estimate. This feature is available only when the X factor has two levels and is usually used for individual parameters.</td>
</tr>
</tbody>
</table>
| Adjusted Power and Confidence Interval | When you look at power retrospectively, you use estimates of the standard error and the test parameters.  
• Adjusted power is the power calculated from a more unbiased estimate of the noncentrality parameter.  
• The confidence interval for the adjusted power is based on the confidence interval for the noncentrality estimate.  
Adjusted power and confidence limits are computed only for the original Delta, because that is where the random variation is. |
Chapter 5
Performing One-way Analysis

CDF Plot

Figure 5.31 Example of a Normal Quantile Plot

From Figure 5.31, notice the following:

- The Line of Fit appears by default.
- The data points track very closely to the line of fit, indicating a normal distribution.

CDF Plot

A CDF plot shows the cumulative distribution function for all of the groups in the One-way report. CDF plots are useful if you want to compare the distributions of the response across levels of the X factor.

Example of a CDF Plot

1. Open the Analgesics.jmp sample data table.
2. Select Analyze > Fit Y by X.
4. Select drug and click X, Factor.
5. Click OK.
6. From the red triangle menu, select CDF Plot.
The levels of the X variables in the initial Oneway analysis appear in the CDF plot as different curves. The horizontal axis of the CDF plot uses the \( y \) value in the initial Oneway analysis.

**Densities**

The **Densities** options provide several ways to compare the distribution and composition of the response across the levels of the X factor. There are three density options:

- **Compare Densities** shows a smooth curve estimating the density of each group. The smooth curve is the kernel density estimate for each group.

- **Composition of Densities** shows the summed densities, weighted by each group’s counts. At each X value, the Composition of Densities plot shows how each group contributes to the total.

- **Proportion of Densities** shows the contribution of the group as a proportion of the total at each X level.

**Example of the Densities Options**

1. Open the Big Class.jmp sample data table.
2. Select **Analyze > Fit Y by X**.
3. Select height and click **Y, Response**.
4. Select sex and click **X, Factor**.
5. Click **OK**.
6. From the red triangle menu, select **Densities > Compare Densities**, **Densities > Composition of Densities**, and **Densities > Proportion of Densities**.

---

**Figure 5.33 Example of the Densities Options**

- **Compare Densities**
- **Composition of Densities**: contribution of males to height density, contribution of females to height density
- **Proportion of Densities**: females contributed about 61% to height density at this X level
Matching Column

Use the Matching Column option to specify a matching (ID) variable for a matching model analysis. The Matching Column option addresses the case when the data in a one-way analysis come from matched (paired) data, such as when observations in different groups come from the same subject.

**Note:** A special case of matching leads to the paired t-test. The Matched Pairs platform handles this type of data, but the data must be organized with the pairs in different columns, not in different rows.

The Matching Column option performs two primary actions:

- It fits an additive model (using an iterative proportional fitting algorithm) that includes both the grouping variable (the X variable in the Fit Y by X analysis) and the matching variable that you select. The iterative proportional fitting algorithm makes a difference if there are hundreds of subjects, because the equivalent linear model would be very slow and would require huge memory resources.
- It draws lines between the points that match across the groups. If there are multiple observations with the same matching ID value, lines are drawn from the mean of the group of observations.

The Matching Column option automatically activates the Matching Lines option connecting the matching points. To turn the lines off, select Display Options > Matching Lines.

**Example of the Matching Column Option**

This example uses the Matching.jmp sample data table, which contains data on six animals and the miles that they travel during different seasons.

1. Open the Matching.jmp sample data table.
2. Select Analyze > Fit Y by X.
3. Select miles and click Y, Response.
4. Select season and click X, Factor.
5. Click OK.
6. From the red triangle menu, select Matching Column.
7. Select subject as the matching column.
8. Click OK.
Figure 5.34 Example of the Matching Column Report

The plot graphs the miles traveled by season, with subject as the matching variable. The labels next to the first measurement for each subject on the graph are determined by the species and subject variables.

The Matching Fit report shows the season and subject effects with $F$-tests. These are equivalent to the tests that you get with the Fit Model platform if you run two models, one with the interaction term and one without. If there are only two levels, then the $F$-test is equivalent to the paired $t$-test.

Note: For details about the Fit Model platform, see the Modeling and Multivariate Methods book.
Statistical Details

The following sections provide statistical details for selected options.

Comparison Circles

One approach to comparing two means is to determine whether their actual difference is greater than their least significant difference (LSD). This least significant difference is a Student’s t-statistic multiplied by the standard error of the difference of the two means and is written as follows:

\[
\text{LSD} = t_{\alpha/2} \frac{\text{std}(\mu_1 - \mu_2)}{}
\]

The standard error of the difference of two independent means is calculated from the following relationship:

\[
\left(\text{std}(\hat{\mu}_1 - \hat{\mu}_2)\right)^2 = \left(\text{std}(\hat{\mu}_1)\right)^2 + \left(\text{std}(\hat{\mu}_2)\right)^2
\]

When the means are uncorrelated, these quantities have the following relationship:

\[
\text{LSD}^2 = \left(t_{\alpha/2} \frac{\text{std}(\hat{\mu}_1 - \hat{\mu}_2)}{2}\right)^2 = \left(t_{\alpha/2} \frac{\text{std}(\hat{\mu}_1)}{2}\right)^2 + \left(t_{\alpha/2} \frac{\text{std}(\hat{\mu}_2)}{2}\right)^2
\]

These squared values form a Pythagorean relationship, illustrated graphically by the right triangle shown in Figure 5.35.

**Figure 5.35** Relationship of the Difference between Two Means

The hypotenuse of this triangle is a measuring stick for comparing means. The means are significantly different if and only if the actual difference is greater than the hypotenuse (LSD).

Suppose that you have two means that are exactly on the borderline, where the actual difference is the same as the least significant difference. Draw the triangle with vertices at the means measured on a vertical scale. Also, draw circles around each mean so that the diameter of each is equal to the confidence interval for that mean.
The radius of each circle is the length of the corresponding leg of the triangle, which is \( t_{\alpha/2} \cdot \text{std}(\mu_j) \).

The circles must intersect at the same right angle as the triangle legs, giving the following relationship:

- If the means differ exactly by their least significant difference, then the confidence interval circles around each mean intersect at a right angle. That is, the angle of the tangents is a right angle.

Now, consider the way that these circles must intersect if the means are different by greater than or less than the least significant difference:

- If the circles intersect so that the outside angle is greater than a right angle, then the means are not significantly different. If the circles intersect so that the outside angle is less than a right angle, then the means are significantly different. An outside angle of less than 90 degrees indicates that the means are farther apart than the least significant difference.

- If the circles do not intersect, then they are significantly different. If they nest, they are not significantly different. See Figure 5.16.

The same graphical technique works for many multiple-comparison tests, substituting a different probability quantile value for the Student’s \( t \).

**Power**

To compute power, you make use of the noncentral \( F \) distribution. The formula (O’Brien and Lohr 1984) is given as follows:

\[
\text{Power} = \text{Prob}(F > F_{\text{crit}}, v_1, v_2, nc)
\]

where:

- \( F \) is distributed as the noncentral \( F(nc, v_1, v_2) \) and \( F_{\text{crit}} = F(1 - \alpha, v_1, v_2) \) is the 1 - \( \alpha \) quantile of the \( F \) distribution with \( v_1 \) and \( v_2 \) degrees of freedom.
- \( v_1 = r - 1 \) is the numerator df.
- \( v_2 = r(n - 1) \) is the denominator df.
- \( n \) is the number per group.
• $r$ is the number of groups.
• $nc = n(CSS)/\sigma^2$ is the noncentrality parameter.

$$CSS = \sum_{g=1}^{r} (\mu_g - \mu)^2$$

is the corrected sum of squares.
• $\mu_g$ is the mean of the $g^{th}$ group.
• $\mu$ is the overall mean.
• $\sigma^2$ is estimated by the mean squared error (MSE).
Performing Contingency Analysis
Using the Fit Y by X or Contingency Platform

Using the Contingency or Fit Y by X platform, you can explore the distribution of a categorical (nominal or ordinal) variable Y across the levels of a second categorical variable X. The Contingency platform is the categorical by categorical personality of the Fit Y by X platform. The analysis results include a mosaic plot, frequency counts, and proportions. You can interactively perform additional analyses and tests on your data, such as an Analysis of Means for Proportions, a correspondence analysis plot, and so on.

**Figure 6.1 Example of Contingency Analysis**
Contents

Example of Contingency Analysis .............................................................. 187
Launch the Contingency Platform ............................................................... 188
The Contingency Report ........................................................................... 188
Contingency Platform Options ................................................................. 189
Mosaic Plot ............................................................................................... 191
Context Menu ........................................................................................... 192
Contingency Table .................................................................................... 193
Tests ........................................................................................................... 195
Fisher's Exact Test ..................................................................................... 197
Analysis of Means for Proportions .......................................................... 197
Correspondence Analysis ........................................................................ 199
Understanding Correspondence Analysis Plots ......................................... 199
The Details Report .................................................................................... 202
Cochran-Mantel-Haenszel Test ................................................................. 206
Agreement Statistic .................................................................................. 207
Relative Risk ............................................................................................. 209
Two Sample Test for Proportions ............................................................... 211
Measures of Association .......................................................................... 212
Cochran Armitage Trend Test ................................................................. 214
Exact Test .................................................................................................. 215
Statistical Details for the Agreement Statistic Option .............................. 216
Example of Contingency Analysis

This example uses the Car Poll.jmp sample data table, which contains data collected from car polls. The data includes aspects about the individual polled, such as their sex, marital status, and age. The data also includes aspects about the car that they own, such as the country of origin, the size, and the type. Examine the relationship between car sizes (small, medium, and large) and the cars’ country of origin.

1. Open the Car Poll.jmp sample data table.
2. Select Analyze > Fit Y by X.
3. Select size and click Y, Response.
4. Select country and click X, Factor.
5. Click OK.

Figure 6.2 Example of Contingency Analysis

<table>
<thead>
<tr>
<th>Count</th>
<th>Large</th>
<th>Medium</th>
<th>Small</th>
<th>Total %</th>
<th>Col %</th>
<th>Row %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>American</td>
<td>56</td>
<td>53</td>
<td>36</td>
<td>115</td>
<td>11.89</td>
<td>47.49</td>
</tr>
<tr>
<td></td>
<td>1.32</td>
<td>5.61</td>
<td>6.37</td>
<td>13.20</td>
<td>0.09</td>
<td>52.50</td>
</tr>
<tr>
<td>European</td>
<td>4</td>
<td>17</td>
<td>19</td>
<td>40</td>
<td>1.52</td>
<td>13.87</td>
</tr>
<tr>
<td></td>
<td>10.09</td>
<td>42.50</td>
<td>47.90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japanese</td>
<td>2</td>
<td>54</td>
<td>92</td>
<td>148</td>
<td>0.66</td>
<td>17.82</td>
</tr>
<tr>
<td></td>
<td>4.78</td>
<td>43.56</td>
<td>67.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1.35</td>
<td>36.49</td>
<td>62.16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>42</td>
<td>124</td>
<td>137</td>
<td>303</td>
<td>1.35</td>
<td>36.49</td>
</tr>
<tr>
<td></td>
<td>43.38</td>
<td>40.32</td>
<td>45.21</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>N</th>
<th>DF</th>
<th>-LogLfe</th>
<th>P-Square</th>
<th>ChiSquare</th>
<th>Prob=ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>303</td>
<td>4</td>
<td>38.00069</td>
<td>0.1200</td>
<td>72.019</td>
<td>&lt;0.0001*</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>72.019</td>
<td>&lt;0.0001*</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pearson</td>
<td>66.313</td>
<td>&lt;0.0001*</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Performing Contingency Analysis
The Contingency Report

From the mosaic plot and legend in Figure 6.2, notice the following:

- Very few Japanese cars fall into the Large size category.
- The majority of the European cars fall into the Small and Medium size categories.
- The majority of the American cars fall into the Large and Medium size categories.

Launch the Contingency Platform

You can perform a contingency analysis using either the Fit Y by X platform or the Contingency platform. The two approaches are equivalent.

- To launch the Fit Y by X platform, select Analyze > Fit Y by X. The Fit Y by X launch window appears. See Figure 3.2 in the “Introduction to the Fit Y by X Platform” chapter.
- To launch the Contingency platform, from the JMP Starter window, click on the Basic category and click Contingency. The Contingency launch window appears.

Figure 6.3 The Contingency Launch Window

For details about this window, see the “Introduction to the Fit Y by X Platform” chapter.

The Contingency Report

To produce the plot shown in Figure 6.4, follow the instructions in “Example of Contingency Analysis,” p. 187.
Figure 6.4 Example of a Contingency Report

The Contingency report initially shows a Mosaic Plot, a Contingency Table, and a Tests report. You can add other analyses and tests using the options that are located within the red triangle menu. For details about all of these reports and options, see “Contingency Platform Options,” p. 189.

Contingency Platform Options

Use the platform options within the red triangle menu next to Contingency Analysis to perform additional analyses and tests on your data.

Table 6.1 Descriptions of Contingency Platform Options

| Mosaic Plot | A graphical representation of the data in the Contingency Table. See "Mosaic Plot," p. 191. |
Performing Contingency Analysis

Chapter 6

Contingency Platform Options

<table>
<thead>
<tr>
<th>Contingency Table</th>
<th>A two-way frequency table. There is a row for each factor level and a column for each response level. See “Contingency Table,” p. 193.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tests</td>
<td>Analogous to the Analysis of Variance table for continuous data. The tests show that the response level rates are the same across X levels. See “Tests,” p. 195.</td>
</tr>
<tr>
<td>Set α level</td>
<td>Changes the alpha level used in confidence intervals. Select one of the common values (0.10, 0.05, 0.01) or select a specific value using the Other option.</td>
</tr>
<tr>
<td>Analysis of Means for Proportions</td>
<td>Only appears if the response has exactly two levels. Compares response proportions for the X levels to the overall response proportion. See “Analysis of Means for Proportions,” p. 197.</td>
</tr>
<tr>
<td>Correspondence Analysis</td>
<td>Shows which rows or columns of a frequency table have similar patterns of counts. In the correspondence analysis plot, there is a point for each row and for each column of the contingency table. See “Correspondence Analysis,” p. 199.</td>
</tr>
<tr>
<td>Cochran Mantel Haenszel</td>
<td>Tests if there is a relationship between two categorical variables after blocking across a third classification. See “Cochran-Mantel-Haenszel Test,” p. 206.</td>
</tr>
<tr>
<td>Agreement Statistic</td>
<td>Only appears when both the X and Y variables have the same levels. Displays the Kappa statistic (Agresti 1990), its standard error, confidence interval, hypothesis test, and Bowker’s test of symmetry, also known as McNemar’s test. See “Agreement Statistic,” p. 207.</td>
</tr>
<tr>
<td>Relative Risk</td>
<td>Calculates risk ratios. Appears only when both the X and Y variables have only two levels. See “Relative Risk,” p. 209.</td>
</tr>
</tbody>
</table>
| Odds Ratio        | Appears only when there are exactly two levels for each variable. Produces a report of the odds ratio, calculated as follows:  
\[
\frac{p_{11} \times p_{22}}{p_{12} \times p_{21}}
\]
where \( p_{ij} \) is the count in the \( i^{th} \) row and \( j^{th} \) column of the 2x2 table. The report also gives a confidence interval for this ratio. You can change the alpha level using the Set α Level option. |
Mosaic Plot

The mosaic plot is a graphical representation of the two-way frequency table or Contingency Table. A mosaic plot is divided into rectangles, so that the area of each rectangle is proportional to the proportions of the Y variable in each level of the X variable. The mosaic plot was introduced by Hartigan and Kleiner in 1981 and refined by Friendly (1994).

To produce the plot shown in Figure 6.5, follow the instructions in “Example of Contingency Analysis,” p. 187.

Figure 6.5 Example of a Mosaic Plot
Performing Contingency Analysis

Chapter 6

Mosaic Plot

Note the following about the mosaic plot in Figure 6.5:

- The proportions on the x-axis represent the number of observations for each level of the X variable, which is country.
- The proportions on the y-axis at right represent the overall proportions of Small, Medium, and Large cars for the combined levels (American, European, and Japanese).
- The scale of the y-axis at left shows the response probability, with the whole axis being a probability of one (representing the total sample).

Clicking on a rectangle in the mosaic plot highlights the selection and highlights the corresponding data in the associated data table.

Context Menu

Right-click on the mosaic plot to change colors and label the cells.

Table 6.2 Descriptions of Right-Click Menu Options

<table>
<thead>
<tr>
<th>Set Colors</th>
<th>Shows the current assignment of colors to levels. See “Set Colors,” p. 192.</th>
</tr>
</thead>
</table>
| Cell Labeling | Specify a label to be drawn in the mosaic plot. Select one of the following options:  
  Unlabeled | shows no labels, and removes any of the other options.  
  Show Counts | Shows the number of observations in each cell.  
  Show Percents | Shows the percent of observations in each cell.  
  Show Labels | Shows the levels of the Y variable corresponding to each cell.  
  Show Row Labels | Shows the row labels for all of the rows represented by the cell. |

For descriptions of the remainder of the right-click options, see the Using JMP book.

Set Colors

When you select the Set Colors option, the Select Colors for Values window appears.

Figure 6.6 Select Colors for Values Window
The default mosaic colors depend on whether the response column is ordinal or nominal, and whether there is an existing Value Colors column property. To change the color for any level, click on the oval in the second column of colors and pick a new color.

**Table 6.3** Description of Select Colors for Values Window

<table>
<thead>
<tr>
<th>Macros</th>
<th>Computes a color gradient between any two levels, as follows:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• If you select a range of levels (by dragging the mouse over the levels that you want to select, or pressing the SHIFT key and clicking the first and last level), the <strong>Gradient Between Selected Points</strong> option applies a color gradient to the levels that you have selected.</td>
</tr>
<tr>
<td></td>
<td>• The <strong>Gradient Between Ends</strong> option applies a gradient to all levels of the variable.</td>
</tr>
<tr>
<td></td>
<td>• Undo any of your changes by selecting <strong>Revert to Old Colors</strong>.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Color Theme</th>
<th>Changes the colors for each value based on a color theme.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Save Colors to Column</strong></td>
<td>If you select this check box, a new column property (Value Colors) is added to the column in the associated data table. To edit this property from the data table, select <strong>Cols &gt; Column Info</strong>.</td>
</tr>
</tbody>
</table>

- **Contingency Table**

The Contingency Table is a two-way frequency table. There is a row for each factor level and a column for each response level.

To produce the plot shown in Figure 6.7, follow the instructions in “Example of Contingency Analysis,” p. 187.

**Figure 6.7** Example of a Contingency Table
Performing Contingency Analysis

Chapter 6

Mosaic Plot

Note the following about Contingency Tables:

- The Count, Total %, Col %, and Row % correspond to the data within each cell that has row and column headings (such as the cell under American and Large).
- The last column contains the total counts for each row and percentages for each row.
- The bottom row contains total counts for each column and percentages for each column.

For example, in Figure 6.7, focus on the cars that are large and come from America. Table 6.4 explains the conclusions that you can make about these cars using the Contingency Table.

**Table 6.4** Cars that are Large and Come From America

<table>
<thead>
<tr>
<th>Number</th>
<th>Description</th>
<th>Label in Table</th>
</tr>
</thead>
<tbody>
<tr>
<td>36</td>
<td>Number of cars that are both large and come from America</td>
<td>Count</td>
</tr>
<tr>
<td>11.88%</td>
<td>Percentage of all cars that are both large and come from America (36/303).</td>
<td>Total %</td>
</tr>
<tr>
<td>85.71%</td>
<td>Percentage of large cars that come from America (36/42).</td>
<td>Col %</td>
</tr>
<tr>
<td>31.30%</td>
<td>Percentage of American cars that are large (36/115).</td>
<td>Row %</td>
</tr>
<tr>
<td>37.95%</td>
<td>Percentage of all cars that come from America (115/303).</td>
<td>(none)</td>
</tr>
<tr>
<td>13.86%</td>
<td>Percentage of all cars that are large (42/303).</td>
<td>(none)</td>
</tr>
</tbody>
</table>

a. 303 is the total number of cars in the poll.
b. 42 is the total number of large cars in the poll.
c. 115 is the total number of American cars in the poll.

To show or hide data in the Contingency Table, from the red triangle menu next to Contingency Table, select the option that you want to show or hide.

**Table 6.5** Descriptions of Contingency Table Options

<table>
<thead>
<tr>
<th>Count</th>
<th>Cell frequency, margin total frequencies, and grand total (total sample size).</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total%</td>
<td>Percent of cell counts and margin totals to the grand total.</td>
</tr>
<tr>
<td>Row%</td>
<td>Percent of each cell count to its row total.</td>
</tr>
<tr>
<td>Col%</td>
<td>Percent of each cell count to its column total.</td>
</tr>
<tr>
<td>Expected</td>
<td>Expected frequency ( (E) ) of each cell under the assumption of independence. Computed as the product of the corresponding row total and column total divided by the grand total.</td>
</tr>
<tr>
<td>Deviation</td>
<td>Observed cell frequency ( (O) ) minus the expected cell frequency ( (E) ).</td>
</tr>
<tr>
<td>Cell Chi Square</td>
<td>Chi-square values computed for each cell as ((O - E)^2 / E).</td>
</tr>
</tbody>
</table>
Chapter 6
Performing Contingency Analysis

Tests

The Tests report shows the results for two tests to determine whether the response level rates are the same across X levels.

To produce the report shown in Figure 6.8, follow the instructions in “Example of Contingency Analysis,” p. 187.

Figure 6.8 Example of a Tests Report

Note the following about the Chi-square statistics:

• When both categorical variables are responses (Y variables), the Chi-square statistics test that they are independent.
• If one variable is considered as Y and the X is regarded as fixed, the Chi-square statistics test that the distribution of the Y variable is the same across each X level. This is a test of marginal homogeneity.

Table 6.6 Description of the Tests Table

<table>
<thead>
<tr>
<th>N</th>
<th>Total number of observations.</th>
</tr>
</thead>
<tbody>
<tr>
<td>DF</td>
<td>Records the degrees of freedom associated with the test. The degrees of freedom are equal to ((c - 1)(r - 1)), where (c) is the number of columns and (r) is the number of rows.</td>
</tr>
<tr>
<td>-LogLike</td>
<td>Negative log-likelihood, which measures fit and uncertainty (much like sums of squares in continuous response situations).</td>
</tr>
</tbody>
</table>
Performing Contingency Analysis
Chapter 6
Tests

Table 6.6 Description of the Tests Table (Continued)

| Rsquare (U) | Portion of the total uncertainty attributed to the model fit, computed as follows: 
|---|---|
| \[ \frac{-\log \text{likelihood for Model}}{-\log \text{likelihood for Corrected Total}} \] | The total negative log-likelihood is found by fitting fixed response rates across the total sample.
| • An \( R^2 \) of 1 means that the factors completely predict the categorical response.
| • An \( R^2 \) of 0 means that there is no gain from using the model instead of fixed background response rates. | |

| Test | Lists two Chi-square statistical tests of the hypothesis that the response rates are the same in each sample category. The two Chi-square tests are as follows:
|---|---|
| The Likelihood Ratio Chi-square test is computed as twice the negative log-likelihood for Model in the Tests table. Some books use the notation \( G^2 \) for this statistic. The difference of two negative log-likelihoods, one with whole-population response probabilities and one with each-population response rates, is written as follows:
| \[ G^2 = 2 \left( \sum_{ij} (-n_{ij}) \ln(p_j) - \sum_{ij} -n_{ij} \ln(p_{ij}) \right) \] | where \( p_{ij} = \frac{n_{ij}}{N} \) and \( p_j = \frac{N_j}{N} \)
| This formula can be more compactly written as follows:
| \[ G^2 = 2 \sum_{ij} n_{ij} \ln \left( \frac{p_{ij}}{p_j} \right) \] | |
| The Pearson Chi-square is calculated by summing the squares of the differences between the observed and expected cell counts. The Pearson Chi-square exploits the property that frequency counts tend to a normal distribution in very large samples. The familiar form of this Chi-square statistic is as follows:
| \[ \chi^2 = \sum \frac{(O - E)^2}{E} \] | where \( O \) is the observed cell counts and \( E \) is the expected cell counts. The summation is over all cells. There is no continuity correction done here, as is sometimes done in 2-by-2 tables. |

| Prob>ChiSq | Lists the probability of obtaining, by chance alone, a Chi-square value greater than the one computed if no relationship exists between the response and factor. If both variables have only two levels, Fisher’s exact probabilities for the one-tailed tests and the two-tailed test also appear. |
Chapter 6

Performing Contingency Analysis

Analysis of Means for Proportions

Fisher’s Exact Test

This report gives the results of Fisher’s exact test for a 2x2 table. The results appear automatically for 2x2 tables. For more details about Fisher’s exact test, and for details about the test for $r \times c$ tables, see “Exact Test,” p. 215.

Analysis of Means for Proportions

If the response has two levels, you can use this option to compare response proportions for the X levels to the overall response proportion. This method uses the normal approximation to the binomial. Therefore, if the sample sizes are too small, a warning appears in the results.


Example of Analysis of Means for Proportions

This example uses the Office Visits jmp sample data table, which records late and on-time appointments for six clinics in a geographic region. 60 random appointments were selected from 1 week of records for each of the six clinics. To be considered on-time, the patient must be taken to an exam room within five minutes of their scheduled appointment time. Examine the proportion of patients that arrived on-time to their appointment.

1. Open the Office Visits jmp sample data table.
2. Select Analyze > Fit Y by X.
4. Select Clinic and click X, Factor.
5. Select Frequency and click Freq.
6. Click OK.
7. From the red triangle menu next to Contingency Analysis, select Analysis of Means for Proportions.
8. From the red triangle menu next to Analysis of Means for Proportions, select Show Summary Report.
Figure 6.9 shows the proportion of patients who were on-time from each clinic. From Figure 6.9, notice the following:

- The proportion of on-time arrivals is the highest for clinic F, followed by clinic B.
- Clinic D has the lowest proportion of on-time arrivals, followed by clinic A.
- Clinic E and clinic C are close to the average, and do not exceed the decision limits.

Table 6.7 Descriptions of the Analysis of Means for Proportions Options

<table>
<thead>
<tr>
<th>Set Alpha Level</th>
<th>Selects the alpha level used in the analysis.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show Summary Report</td>
<td>Produces a report that shows the response proportions with decision limits for each level of the X variable. The report indicates whether a limit has been exceeded.</td>
</tr>
<tr>
<td>Switch Response Level for Proportion</td>
<td>Changes the response category used in the analysis.</td>
</tr>
<tr>
<td>Display Options</td>
<td>Shows or hides the decision limits, decision limit shading, center line, and needles.</td>
</tr>
</tbody>
</table>
Correspondence Analysis

Correspondence analysis is a graphical technique to show which rows or columns of a frequency table have similar patterns of counts. In the correspondence analysis plot, there is a point for each row and for each column. Use Correspondence Analysis when you have many levels, making it difficult to derive useful information from the mosaic plot.

Understanding Correspondence Analysis Plots

The row profile can be defined as the set of row-wise rates, or in other words, the counts in a row divided by the total count for that row. If two rows have very similar row profiles, their points in the correspondence analysis plot are close together. Squared distances between row points are approximately proportional to Chi-square distances that test the homogeneity between the pair of rows.

Column and row profiles are alike because the problem is defined symmetrically. The distance between a row point and a column point has no meaning, but the directions of columns and rows from the origin are meaningful, and the relationships help to interpret the plot.

Example of Correspondence Analysis

This example uses the Cheese.jmp sample data table, which is taken from the Newell cheese tasting experiment, reported in McCullagh and Nelder (1989). The experiment records counts over nine different response levels across four different cheese additives.

1. Open the Cheese.jmp sample data table.
2. Select Analyze > Fit Y by X.
   - The Response values range from one to nine, where one is the least liked, and nine is the best liked.
4. Select Cheese and click X, Factor.
   - A, B, C, and D represent four different cheese additives.
5. Select Count and click Freq.
6. Click OK.
Performing Contingency Analysis
Chapter 6

Correspondence Analysis

Figure 6.10 Mosaic Plot for the Cheese Data

From the mosaic plot in Figure 6.10, you notice that the distributions do not appear alike. However, it is challenging to make sense of the mosaic plot across nine levels. A correspondence analysis can help define relationships in this type of situation.

7. To see the correspondence analysis plot, from the red triangle menu next to Contingency Analysis, select Correspondence Analysis.

The Correspondence Analysis Plot

Figure 6.11 Example of a Correspondence Analysis Plot
Figure 6.11 shows the correspondence analysis graphically, with the plot axes labeled c1 and c2. Notice the following:

- c1 seems to correspond to a general satisfaction level. The cheeses on the c1 axis go from least liked at the top to most liked at the bottom.
- c2 seems to capture some quality that makes B and D different from A and C.
- Cheese D is the most liked cheese, with responses of 8 and 9.
- Cheese B is the least liked cheese, with responses of 1, 2, and 3.
- Cheeses C and A are in the middle, with responses of 4, 5, 6, and 7.

### Correspondence Analysis Options

Use the options in the red triangle menu next to Correspondence Analysis to produce a 3-D scatterplot and add column properties to the data table.

**Table 6.8 Descriptions of Correspondence Analysis Options**

<table>
<thead>
<tr>
<th>3D Correspondence Analysis</th>
<th>Produces a 3-D scatterplot.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Save Value Ordering</td>
<td>Takes the order of the levels sorted by the first correspondence score coefficient and makes a column property for both the X and Y columns.</td>
</tr>
</tbody>
</table>

**Example of a 3-D Scatterplot**

1. Open the Cheese.jmp sample data table.
2. Select Analyze > Fit Y by X.
   - The Response values range from one to nine, where one is the least liked, and nine is the best liked.
4. Select Cheese and click X, Factor.
   - A, B, C, and D represent four different cheese additives.
5. Select Count and click Freq.
6. Click OK.
7. To see the correspondence analysis plot, from the red triangle menu next to Contingency Analysis, select Correspondence Analysis.
8. From the red triangle menu next to Correspondence Analysis, select 3D Correspondence Analysis.
From Figure 6.12, notice the following:

- Looking at the c1 axis, responses 1 through 5 appear to the right of 0 (positive). Responses 6 through 9 appear to the left of 0 (negative).
- Looking at the c2 axis, A and C appear to the right of 0 (positive). B and D appear to the left of 0 (negative).
- You can conclude that c1 corresponds to the general satisfaction (least liked to most liked) and c2 corresponds to a quality that makes B and D different from A and C.

The Details Report

The Details report contains statistical information about the correspondence analysis plot.
Figure 6.13 Example of a Details Report

<table>
<thead>
<tr>
<th>Singular Value</th>
<th>Inertia</th>
<th>Portion</th>
<th>Cumulative</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.73409</td>
<td>0.64163</td>
<td>0.6636</td>
<td>0.6636</td>
</tr>
<tr>
<td>0.42010</td>
<td>0.17649</td>
<td>0.2259</td>
<td>0.8915</td>
</tr>
<tr>
<td>0.25970</td>
<td>0.05205</td>
<td>0.0605</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

Table 6.9 Description of the Details Report

| Singular Value | Lists the singular values of the following equation: \( D_r^{-0.5}(P - r^c)D_c^{-0.5} \) where:
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Inertia</td>
<td>Lists the square of the singular values, reflecting the relative variation accounted for in the canonical dimensions.</td>
</tr>
<tr>
<td>Portion</td>
<td>Portion of inertia with respect to the total inertia.</td>
</tr>
<tr>
<td>Cumulative</td>
<td>Shows the cumulative portion of inertia. If the first two singular values capture the bulk of the inertia, then the 2-D correspondence analysis plot is sufficient to show the relationships in the table.</td>
</tr>
<tr>
<td>X variable (Cheese) c1, c2, c3</td>
<td>The values plotted on the Correspondence Analysis plot.</td>
</tr>
<tr>
<td>Y variable (Response) c1, c2, c3</td>
<td>The values plotted on the Correspondence Analysis plot.</td>
</tr>
</tbody>
</table>
Performing Contingency Analysis

Correspondence Analysis

Additional Example of Correspondence Analysis

This example uses the Mail Messages.jmp sample data table, which contains data about e-mail messages that were sent and received. The data includes the time, sender, and receiver. Examine the pattern of e-mail senders and receivers.

1. Open the Mail Messages.jmp sample data table.
2. Select Analyze > Fit Y by X.
3. Select To and click Y, Response.
4. Select From and click X, Factor.
5. Click OK.
6. From the red triangle menu next to Contingency Table, deselect everything except Count.

Figure 6.14 Contingency Analysis for E-mail Data

Looking at the frequency table, you notice the following differences:

- Jeff sends messages to everyone but receives messages only from Michael.
- Michael and John send many more messages than the others.
- Michael sends messages to everyone.
• John sends messages to everyone except Jeff.
• Katherine and Ann send messages only to Michael and John.

Further visualize the results of the contingency table with a correspondence analysis. From the red triangle menu next to Contingency Analysis, select Correspondence Analysis.

Figure 6.15 Correspondence Analysis for E-mail Data

From the Details report in Figure 6.15, you notice the following:

• The Portion column in the table shows that the bulk of the variation (56% + 42%) of the mail sending pattern is summarized by \( c_1 \) and \( c_2 \), for the To and From groups.

• The Correspondence Analysis plot of \( c_1 \) and \( c_2 \) shows the pattern of mail distribution among the mail group, as follows:
  – Katherine and Ann have similar sending and receiving patterns; they both send emails to Michael and John and receive emails from Michael, John, and Jeff.
  – Jeff and Michael lie in a single dimension, but have opposite sending and receiving patterns. Jeff sends emails to everyone and receives emails only from Michael. Michael sends email to everyone and receives email from everyone.
  – John’s patterns differ from the others. He sends email to Ann, Katherine, and Michael, and receives email from everyone.
Performing Contingency Analysis
Chapter 6

Cochran-Mantel-Haenszel Test

The Cochran-Mantel-Haenszel test discovers if there is a relationship between two categorical variables after blocking across a third classification.

Example of a Cochran Mantel Haenszel Test

This example uses the Hot Dogs.jmp sample data table. Examine the relationship between hot dog type and taste.

1. Open the Hot Dogs.jmp sample data table.
2. Select Analyze > Fit Y by X.
3. Select Type and click Y, Response.
4. Select Taste and click X, Factor.
5. Click OK.
6. From the red triangle menu next to Contingency Analysis, select Cochran Mantel Haenszel.
7. Select Protein/Fat as the grouping variable and click OK.

Figure 6.16 Example of a Cochran-Mantel-Haenszel Test
Performing Contingency Analysis

Agreement Statistic

From Figure 6.16, you notice the following:

- The Tests report shows a marginally significant Chi-square probability of about 0.0799, indicating some significance in the relationship between hot dog taste and type.
- However, the Cochran Mantel Haenszel report shows no relationship at all between hot dog taste and type after adjusting for the protein-fat ratio. The Chi-square test for the adjusted correlation has a probability of 0.857, and the Chi-square probability associated with the general association of categories is 0.282.

Table 6.10 Description of the Cochran Mantel Haenszel Tests Report

<table>
<thead>
<tr>
<th>Correlation of Scores</th>
<th>Applicable when either the Y or X is ordinal. The alternative hypothesis is that there is a linear association between Y and X in at least one level of the blocking variable.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row Score by Col Categories</td>
<td>Applicable when Y is ordinal or interval. The alternative hypothesis is that, for at least one level of the blocking variable, the mean scores of the r rows are unequal.</td>
</tr>
<tr>
<td>Col Score by Row Categories</td>
<td>Applicable when X is ordinal or interval. The alternative hypothesis is that, for at least one level of the blocking variable, the mean scores of the c columns are unequal.</td>
</tr>
<tr>
<td>General Assoc. of Categories</td>
<td>Tests that for at least one level of the blocking variable, there is some type of association between X and Y.</td>
</tr>
</tbody>
</table>

Agreement Statistic

When the two variables have the same levels, the Agreement Statistic option is available. This option shows the Kappa statistic (Agresti 1990), its standard error, confidence interval, hypothesis test, and Bowker’s test of symmetry.

Example of the Agreement Statistic Option

This example uses the Attribute Gauge.jmp sample data table. The data gives results from three people (raters) rating fifty parts three times each. Examine the relationship between raters A and B.

1. Open the Attribute Gauge.jmp sample data table.
2. Select Analyze > Fit Y by X.
4. Select B and click X, Factor.
5. Click OK.
6. From the red triangle menu next to Contingency Analysis, select Agreement Statistic.
Performing Contingency Analysis

Agreement Statistic

Chapter 6

Figure 6.17 Example of the Agreement Statistic Report

Viewing the two response variables as two independent ratings of the \( n \) subjects, the Kappa coefficient equals +1 when there is complete agreement of the raters. When the observed agreement exceeds chance agreement, the Kappa coefficient is positive, with its magnitude reflecting the strength of agreement. Although unusual in practice, Kappa is negative when the observed agreement is less than chance agreement. The minimum value of Kappa is between -1 and 0, depending on the marginal proportions.

From Figure 6.17, you notice that the agreement statistic of 0.86 is high (close to 1) and the p-value of <.0001 is small. This reinforces the high agreement seen by looking at the diagonal of the contingency table. Agreement between the raters occurs when both raters give a rating of 0 or both give a rating of 1.

Table 6.11 Description of the Agreement Statistic Report

<table>
<thead>
<tr>
<th>Description</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kappa</td>
<td>Shows the Kappa statistic.</td>
</tr>
<tr>
<td>Std Err</td>
<td>Shows the standard error of the Kappa statistic.</td>
</tr>
<tr>
<td>Lower 95%</td>
<td>Shows the lower endpoint of the confidence interval for Kappa.</td>
</tr>
<tr>
<td>Upper 95%</td>
<td>Shows the upper endpoint of the confidence interval for Kappa.</td>
</tr>
<tr>
<td>Prob&gt;</td>
<td>Z</td>
</tr>
<tr>
<td>Prob&gt;</td>
<td></td>
</tr>
</tbody>
</table>
Chapter 6
Performing Contingency Analysis

Relative Risk

The Kappa statistic and associated p-value given in this section are approximate. An exact version of the agreement test is available. See “Exact Test,” p. 215.

For statistical details, see “Statistical Details for the Agreement Statistic Option,” p. 216.

Table 6.11 Description of the Agreement Statistic Report (Continued)

<table>
<thead>
<tr>
<th>ChiSquare</th>
<th>Shows the test statistic for Bowker’s test. For Bowker’s test of symmetry, the null hypothesis is that the probabilities in the square table satisfy symmetry, or that $p_{ij}=p_{ji}$ for all pairs of table cells. When both X and Y have two levels, this test is equal to McNemar’s test.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prob&gt;ChiSq</td>
<td>Shows the p-value for the Bowker’s test.</td>
</tr>
</tbody>
</table>

The Kappa statistic and associated p-value given in this section are approximate. An exact version of the agreement test is available. See “Exact Test,” p. 215.

For statistical details, see “Statistical Details for the Agreement Statistic Option,” p. 216.

Relative Risk

Calculate risk ratios for 2x2 contingency tables using the Relative Risk option. Confidence intervals also appear in the report. You can find more information about this method in Agresti (1990) section 3.4.2.

Example of the Relative Risk Option

This example uses the Car Poll.jmp sample data table. Examine the relative probabilities of being married and single for the participants in the poll.

1. Open the Car Poll.jmp sample data table.
2. Select Analyze > Fit Y by X.
3. Select marital status and click Y, Response.
4. Select sex and click X, Factor.
5. Click OK.
6. From the red triangle menu next to Contingency Analysis, select Relative Risk.
   The Choose Relative Risk Categories window appears.

Figure 6.18 The Choose Relative Risk Categories Window
Performing Contingency Analysis

Chapter 6

Relative Risk

Note the following about the Choose Relative Risk Categories window:

- If you are interested in only a single response and factor combination, you can select that here. For example, if you clicked OK in the window in Figure 6.18, the calculation would be as follows:

\[
\frac{P(Y = \text{Married}|X = \text{Female})}{P(Y = \text{Married}|X = \text{Male})}
\]

- If you would like to calculate the risk ratios for all \(2 \times 2 = 4\) combinations of response and factor levels, select the Calculate All Combinations check box. See Figure 6.19.

7. Ask for all combinations by selecting the Calculate All Combinations check box. Leave all other default selections as is.

**Figure 6.19 Example of the Risk Ratio Report**

To see how the relative risk is calculated, proceed as follows:

1. Examine the first entry in the Relative Risk report, which is \(P(\text{Married}|\text{Female})/P(\text{Married}|\text{Male})\).

2. You can find these probabilities in the Contingency Table. Since the probabilities are computed based on two levels of sex, which differs across the rows of the table, use the Row% to read the probabilities, as follows:

\[
P(\text{Married}|\text{Female}) = 0.6884
\]

\[
P(\text{Married}|\text{Male}) = 0.6121
\]

Therefore, the calculations are as follows:

\[
\frac{P(\text{Married}|\text{Female})}{P(\text{Married}|\text{Male})} = \frac{0.6884}{0.6121} = 1.1247
\]
Two Sample Test for Proportions

When both the X and Y variables have two levels, you can request a two-sample test for proportions. This test determines whether the chosen response level has equal proportions across the levels of the X variable.

Example of a Two Sample Test for Proportions

This example uses the Car Poll.jmp sample data table. Examine the probability of being married for males and females.

1. Open the Car Poll.jmp sample data table.
2. Select Analyze > Fit Y by X.
3. Select marital status and click Y, Response.
4. Select sex and click X, Factor.
5. Click OK.
6. From the red triangle menu next to Contingency Analysis, select Two Sample Test for Proportions.

![Figure 6.20 Example of the Two Sample Test for Proportions Report](image)
In this example, you are comparing the probability of being married between females and males. See the Row% in the Contingency Table to obtain the following:

\[
P(\text{Married}|\text{Female}) = 0.6884
\]
\[
P(\text{Married}|\text{Male}) = 0.6121
\]

The difference between these two numbers, 0.0763, is the Proportion Difference shown in the report. The two-sided p-value (0.1686) is large, indicating that there is no significant difference between the proportions.

**Table 6.12** Description of the Two Sample Test for Proportions Report

<table>
<thead>
<tr>
<th>Description</th>
<th>Shows the test being performed.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proportion Difference</td>
<td>Shows the difference in the proportions between the levels of the X variable.</td>
</tr>
<tr>
<td>Lower 95%</td>
<td>Shows the lower endpoint of the confidence interval for the difference. Based on the adjusted Wald confidence interval.</td>
</tr>
<tr>
<td>Upper 95%</td>
<td>Shows the upper endpoint of the confidence interval for the difference. Based on the adjusted Wald confidence interval.</td>
</tr>
<tr>
<td>Adjusted Wald Test</td>
<td>Shows two-tailed and one-tailed tests.</td>
</tr>
<tr>
<td>Prob</td>
<td>Shows the p-values for the tests.</td>
</tr>
<tr>
<td>Response &lt;variable&gt;</td>
<td>Select which response level to use in the test.</td>
</tr>
<tr>
<td>category of interest</td>
<td></td>
</tr>
</tbody>
</table>

**Measures of Association**

You can request several statistics that describe the association between the variables in the contingency table by selecting the **Measures of Association** option.

For details about measures of association, see the following references:

- Brown and Benedetti (1977)
- Goodman and Kruskal (1979)
- Kendall and Stuart (1979)
- Snedecor and Cochran (1980)
- Somers (1962)
Example of the Measures of Association Option

This example uses the Car Poll.jmp sample data table. Examine the probability of being married for males and females.

1. Open the Car Poll.jmp sample data table.
2. Select Analyze > Fit Y by X.
3. Select marital status and click Y, Response.
4. Select sex and click X, Factor.
5. Click OK.
6. From the red triangle menu next to Contingency Analysis, select Measures of Association.

Each statistic appears with its standard error and confidence interval. Note the following:

- Gamma, Kendall’s Tau-b, Stuart’s Tau-c, and Somer’s D are measures of ordinal association that consider whether the variable Y tends to increase as X increases. They classify pairs of observations as concordant or discordant. A pair is concordant if an observation with a larger value of X also has a larger value of Y. A pair is discordant if an observation with a larger value of X has a smaller value of Y. These measures are appropriate only when both variables are ordinal.

- The Lambda and Uncertainty measures are appropriate for ordinal and nominal variables.

Table 6.13 Description of the Measures of Association Report

<table>
<thead>
<tr>
<th>Measure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gamma</td>
<td>Based on the number of concordant and discordant pairs and ignores tied pairs. Takes values in the range -1 to 1.</td>
</tr>
<tr>
<td>Kendall’s Tau-b</td>
<td>Similar to Gamma and uses a correction for ties. Takes values in the range -1 to 1.</td>
</tr>
<tr>
<td>Stuart’s Tau-c</td>
<td>Similar to Gamma and uses an adjustment for table size and a correction for ties. Takes values in the range -1 to 1.</td>
</tr>
</tbody>
</table>
Performing Contingency Analysis

Cochran Armitage Trend Test

This Cochran Armitage Trend tests for trends in binomial proportions across the levels of a single variable. This test is appropriate only when one variable has two levels and the other variable is ordinal. The two-level variable represents the response, and the other represents an explanatory variable with ordered levels. The null hypothesis is the hypothesis of no trend, which means that the binomial proportion is the same for all levels of the explanatory variable.

Example of the Cochran Armitage Trend Test

1. Open the Car Poll.jmp sample data table.
2. In the Columns panel, right-click on the icon next to size and select Ordinal.

Table 6.13 Description of the Measures of Association Report (Continued)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Somer's D</td>
<td>An asymmetric modification of Tau-b.</td>
</tr>
<tr>
<td></td>
<td>* The C</td>
</tr>
<tr>
<td></td>
<td>* Similarly, the R</td>
</tr>
<tr>
<td></td>
<td>Somer's D differs from Tau-b in that it uses a correction for ties only when the pair is tied on the independent variable. It takes values in the range -1 to 1.</td>
</tr>
<tr>
<td>Lambda Asymmetric</td>
<td>* For C</td>
</tr>
<tr>
<td></td>
<td>* For R</td>
</tr>
<tr>
<td></td>
<td>Takes values in the range 0 to 1.</td>
</tr>
<tr>
<td>Lambda Symmetric</td>
<td>Loosely interpreted as the average of the two Lambda Asymmetric measures. Takes values in the range 0 to 1.</td>
</tr>
<tr>
<td>Uncertainty Coef</td>
<td>* For C</td>
</tr>
<tr>
<td></td>
<td>* For R</td>
</tr>
<tr>
<td></td>
<td>Takes values in the range 0 to 1.</td>
</tr>
<tr>
<td>Uncertainty Coef Symmetric</td>
<td>Symmetric version of the two Uncertainty Coef measures. Takes values in the range 0 to 1.</td>
</tr>
</tbody>
</table>
3. Select Analyze > Fit Y by X.
4. Select sex and click Y, Response.
5. Select size and click X, Factor.
6. Click OK.
7. From the red triangle menu next to Contingency Analysis, select Cochran Armitage Trend Test.

Figure 6.22 Example of the Cochran Armitage Trend Test Report

The two-sided p-value (0.7094) is large. From this, you can conclude that there is no trend in the proportion of male and females that purchase different sizes of cars.

The test statistic and p-values given in this section are approximate. An exact version of the trend test is available, see “Exact Test,” p. 215.

**Exact Test**

**Note:** These options are available only in JMP Pro.

This section discusses the exact versions of three of the tests available in this platform.
Table 6.14 Descriptions of Exact Tests

<table>
<thead>
<tr>
<th>Statistical Details for the Agreement Statistic Option</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fisher's Exact Test</td>
</tr>
<tr>
<td>( \hat{\kappa} = \frac{p_0 - p_e}{1 - p_e} ) where ( p_0 = \sum p_{ii} ) and ( p_e = \sum p_i p_{.i} )</td>
</tr>
<tr>
<td>( \text{var} = \frac{A + B - C}{(1 - p_e)^2 n} ) where ( A = \sum p_{ij} [1 - (p_{i.} + p_{.j})(1 - \hat{\kappa})]^2 ), ( B = (1 - \hat{\kappa})^2 \sum_{i \neq j} p_{ij} (p_{i.} + p_{.j})^2 )</td>
</tr>
</tbody>
</table>

| Exact Cochran Armitage Trend Test | Performs the exact version of the Cochran Armitage Trend Test. This test is available only when one of the variables has two levels. For more details about the trend test, see “Cochran Armitage Trend Test,” p. 214. |
| Exact Agreement Test | Performs an exact test for testing agreement between variables. This is an exact test for the Kappa statistic. This is available only when the two variables have the same levels. For more details about agreement testing, see “Agreement Statistic,” p. 207. |

| Table Probability (P) | gives the probability for the observed table. This is not the p-value for the test. |
| Two-sided Prob \( \leq P \) | gives the p-value for the two-sided test. |

For 2x2 tables, the Fisher's Exact test is automatically performed. See “Tests,” p. 195.
Statistical Details for the Agreement Statistic Option

For Bowker's test of symmetry, the null hypothesis is that the probabilities in the two-by-two table satisfy symmetry \( p_{ij} = p_{ji} \).

\[
C = \left[ \kappa - P_e (1 - \kappa) \right]^2
\]

Performing Contingency Analysis
Statistical Details for the Agreement Statistic Option
Performing Simple Logistic Regression
Using the Fit Y by X or Logistic Platform

The Logistic platform fits the probabilities for response categories to a continuous $x$ predictor. The fitted model estimates probabilities for each $x$ value. The Logistic platform is the nominal or ordinal by continuous personality of the Fit Y by X platform. There is a distinction between nominal and ordinal responses on this platform:

- Nominal logistic regression estimates a set of curves to partition the probability among the responses.
- Ordinal logistic regression models the probability of being less than or equal to a given response. This has the effect of estimating a single logistic curve, which is shifted horizontally to produce probabilities for the ordered categories. This model is less complex and is recommended for ordered responses.

**Figure 7.1 Examples of Logistic Regression**

![Ordinal Logistic Regression](image1)

![Nominal Logistic Regression](image2)
Contents

Overview of Logistic Regression ................................................................. 221
    Nominal Logistic Regression ............................................................. 221
    Ordinal Logistic Regression ............................................................. 221
Example of Nominal Logistic Regression ................................................. 221
Launch the Logistic Platform ................................................................. 223
Logistic Report ....................................................................................... 223
    Logistic Plot ..................................................................................... 225
    Iterations ......................................................................................... 225
    Whole Model Test ............................................................................ 225
    Parameter Estimates ........................................................................ 227
Logistic Platform Options ....................................................................... 228
    ROC Curves ..................................................................................... 229
    Save Probability Formula ................................................................. 231
    Inverse Prediction ............................................................................ 231
Example of Ordinal Logistic Regression .................................................. 233
Additional Example of a Logistic Plot ....................................................... 235
Overview of Logistic Regression

Logistic regression has a long tradition with widely varying applications such as modeling dose-response data and purchase-choice data. Unfortunately, many introductory statistics courses do not cover this fairly simple method. Many texts in categorical statistics cover it (Agresti 1998), in addition to texts on logistic regression (Hosmer and Lemeshow 1989). Some analysts use the method with a different distribution function, the normal. In that case, it is called probit analysis. Some analysts use discriminant analysis instead of logistic regression because they prefer to think of the continuous variables as Y’s and the categories as X’s and work backwards. However, discriminant analysis assumes that the continuous data are normally distributed random responses, rather than fixed regressors.

Simple logistic regression is a more graphical and simplified version of the general facility for categorical responses in the Fit Model platform. For examples of more complex logistic regression models, see the Modeling and Multivariate Methods book.

Nominal Logistic Regression

Nominal logistic regression estimates the probability of choosing one of the response levels as a smooth function of the x factor. The fitted probabilities must be between 0 and 1, and must sum to 1 across the response levels for a given factor value.

In a logistic probability plot, the y-axis represents probability. For k response levels, k - 1 smooth curves partition the total probability (which equals 1) among the response levels. The fitting principle for a logistic regression minimizes the sum of the negative natural logarithms of the probabilities fitted to the response events that occur—that is, maximum likelihood.

Ordinal Logistic Regression

When Y is ordinal, a modified version of logistic regression is used for fitting. The cumulative probability of being at or below each response level is modeled by a curve. The curves are the same for each level except that they are shifted to the right or left.

The ordinal logistic model fits a different intercept, but the same slope, for each of r - 1 cumulative logistic comparisons, where r is the number of response levels. Each parameter estimate can be examined and tested individually, although this is seldom of much interest.

The ordinal model is preferred to the nominal model when it is appropriate because it has fewer parameters to estimate. In fact, it is practical to fit ordinal responses with hundreds of response levels.

Example of Nominal Logistic Regression

This example uses the Penicillin.jmp sample data table. The data in this example comes from an experiment where 5 groups, each containing 12 rabbits, were injected with streptococcus bacteria. Once the rabbits were confirmed to have the bacteria in their system, they were given different doses of penicillin. You want to find out whether the natural log (ln(dose)) of dosage amounts has any effect on whether the rabbits are cured.
Performing Simple Logistic Regression
Example of Nominal Logistic Regression

1. Open the Penicillin.jmp sample data table.
2. Select Analyze > Fit Y by X.
4. Select ln(Dose) and click X, Factor.
   Notice that JMP automatically fills in Count for Freq. Count was previously assigned the role of Freq.
5. Click OK.

Figure 7.2 Example of Nominal Logistic Report

---

**Logistic Fit of Response By ln(dose)**

![Graph showing logistic fit of response by ln(dose)](image)

**Whole Model Test**

<table>
<thead>
<tr>
<th>Model</th>
<th>LogLikelihood</th>
<th>DF</th>
<th>ChiSquare</th>
<th>Prob&gt;ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difference</td>
<td>18.968285</td>
<td>1</td>
<td>37.63586</td>
<td>&lt; 0.001*</td>
</tr>
<tr>
<td>Full</td>
<td>19.313389</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reduced</td>
<td>37.26164</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-squared (U) 0.5489
AICc 40.162
BIC 44.8047
Observations (or Data Points) 54

**Measures**

Entropy R-Square 0.5605
Generalized R-Square 0.5467
Mean-Lag p 0.3391
RMSE 0.3288
Mean Abs Dev 0.2140
Misclassification Rate 0.1401
N 54

**Parameter Estimates**

<table>
<thead>
<tr>
<th>Term</th>
<th>Estimate</th>
<th>Std Error</th>
<th>ChiSquare</th>
<th>Prob&gt;ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>2.68517103</td>
<td>0.0253384</td>
<td>10.56</td>
<td>&lt; 0.001*</td>
</tr>
<tr>
<td>ln(Dose)</td>
<td>2.9098472</td>
<td>0.7692693</td>
<td>14.36</td>
<td>0.0002*</td>
</tr>
</tbody>
</table>

For log odds of Cure|Dead

---

**Covariance of Estimates**
The plot shows the fitted model, which is the predicted probability of being cured, as a function of ln(dose). The p-value is significant, indicating that the dosage amounts have a significant effect on whether the rabbits are cured.

Launch the Logistic Platform

You perform a logistic analysis using either the Fit Y by X platform or the Logistic platform. The two approaches are equivalent.

- To launch the Fit Y by X platform, select Analyze > Fit Y by X. The Fit Y by X launch window appears. See Figure 3.2 in the “Introduction to the Fit Y by X Platform” chapter.

- or

- To launch the Logistic platform, from the JMP Starter window, click on the Basic category and then click Logistic. The Logistic launch window appears.

Figure 7.3 The Logistic Launch Window

For details about this window, see the “Introduction to the Fit Y by X Platform” chapter.

Logistic Report

To produce the plot shown in Figure 7.4, follow the instructions in “Example of Nominal Logistic Regression,” p. 221.
The Logistic report window contains the following:

- “Logistic Plot,” p. 225
- “Iterations,” p. 225
- “Whole Model Test,” p. 225
- “Parameter Estimates,” p. 227

Note: The red triangle menu provides more options that can add to the initial report window. See “Logistic Platform Options,” p. 228.
Logistic Plot

The logistic probability plot gives a complete picture of what the logistic model is fitting. At each x value, the probability scale in the y direction is divided up (partitioned) into probabilities for each response category. The probabilities are measured as the vertical distance between the curves, with the total across all Y category probabilities summing to 1.

For an additional example of a logistic plot, see “Additional Example of a Logistic Plot,” p. 235.

Iterations

The Iterations report shows each iteration and the evaluated criteria that determine whether the model has converged. Iterations appear only for nominal logistic regression.

Whole Model Test

The Whole Model Test report shows if the model fits better than constant response probabilities. This report is analogous to the Analysis of Variance report for a continuous response model. It is a specific likelihood-ratio Chi-square test that evaluates how well the categorical model fits the data. The negative sum of natural logs of the observed probabilities is called the negative log-likelihood (~LogLikelihood). The negative log-likelihood for categorical data plays the same role as sums of squares in continuous data. Twice the difference in the negative log-likelihood from the model fitted by the data and the model with equal probabilities is a Chi-square statistic. This test statistic examines the hypothesis that the x variable has no effect on the responses.

Values of the Rsquare (U) (sometimes denoted as $R^2$) range from 0 to 1. High $R^2$ values are indicative of a good model fit, and are rare in categorical models.

<table>
<thead>
<tr>
<th>Model (sometimes called Source)</th>
<th>• The Reduced model only contains an intercept.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• The Full model contains all of the effects as well as the intercept.</td>
</tr>
<tr>
<td></td>
<td>• The Difference is the difference of the log likelihoods of the full and reduced models.</td>
</tr>
<tr>
<td>DF</td>
<td>Records the degrees of freedom associated with the model.</td>
</tr>
</tbody>
</table>
Performing Simple Logistic Regression

Chapter 7

Logistic Report

Table 7.1 Description of the Whole Model Test Report (Continued)

<table>
<thead>
<tr>
<th>Measure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>–LogLikelihood</td>
<td>Measures variation, sometimes called uncertainty, in the sample.</td>
</tr>
<tr>
<td></td>
<td><strong>Full</strong> (the full model) is the negative log-likelihood (or uncertainty) calculated after fitting the model. The fitting process involves predicting response rates with a linear model and a logistic response function. This value is minimized by the fitting process.</td>
</tr>
<tr>
<td></td>
<td><strong>Reduced</strong> (the reduced model) is the negative log-likelihood (or uncertainty) for the case when the probabilities are estimated by fixed background rates. This is the background uncertainty when the model has no effects.</td>
</tr>
<tr>
<td></td>
<td>The difference of these two negative log-likelihoods is the reduction due to fitting the model. Two times this value is the likelihood-ratio Chi-square test statistic.</td>
</tr>
<tr>
<td>Chi-Square</td>
<td>The likelihood-ratio Chi-square test of the hypothesis that the model fits no better than fixed response rates across the whole sample. It is twice the –LogLikelihood for the Difference Model. It is two times the difference of two negative log-likelihoods, one with whole-population response probabilities and one with each-population response rates. This statistic is sometimes denoted $G^2$ and is written as follows: $G^2 = 2(\sum -\ln p(\text{background}) - \sum -\ln p(\text{model}))$ where the summations are over all observations instead of all cells.</td>
</tr>
<tr>
<td>Prob&gt;ChiSq</td>
<td>The observed significance probability, often called the $p$ value, for the Chi-square test. It is the probability of getting, by chance alone, a Chi-square value greater than the one computed. Models are often judged significant if this probability is below 0.05.</td>
</tr>
<tr>
<td>Rsquare (U)</td>
<td>The proportion of the total uncertainty that is attributed to the model fit. To test that the factor variable has no effect on the response, look at the difference between the following:</td>
</tr>
<tr>
<td></td>
<td>• the log-likelihood from the fitted model</td>
</tr>
<tr>
<td></td>
<td>• the log-likelihood from the model that uses horizontal lines</td>
</tr>
</tbody>
</table>
|                   | The ratio of this test statistic to the background log-likelihood is subtracted from 1 to calculate $R^2$. More simply, it is computed as follows: \[
-\log \text{likelihood for Difference} -\log \text{likelihood for Reduced}
\]
|                   | using quantities from the Whole Model Test report. |
| AICc              | The corrected Akaike Information Criterion. |
### Parameter Estimates

The nominal logistic model fits a parameter for the intercept and slope for each of \( k - 1 \) logistic comparisons, where \( k \) is the number of response levels. The Parameter Estimates report, shown next, lists these estimates. Each parameter estimate can be examined and tested individually, although this is seldom of much interest.

#### Table 7.2 Description of the Parameter Estimates Report

<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term</td>
<td>Lists each parameter in the logistic model. There is an intercept and a slope term for the factor at each level of the response variable, except the last level.</td>
</tr>
</tbody>
</table>
Performing Simple Logistic Regression

Chapter 7

Logistic Platform Options

<table>
<thead>
<tr>
<th>Covariance of Estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reports the estimated variances of the parameter estimates, and the estimated covariances between the parameter estimates. The square root of the variance estimates is the same as those given in the Std Error section.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 7.2 Description of the Parameter Estimates Report (Continued)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Estimate</strong></td>
</tr>
<tr>
<td><strong>Std Error</strong></td>
</tr>
<tr>
<td><strong>Chi-Square</strong></td>
</tr>
<tr>
<td><strong>Prob&gt;ChiSq</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 7.3 Logistic Fit Options</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Odds Ratios</strong></td>
</tr>
<tr>
<td><strong>Inverse Prediction</strong></td>
</tr>
<tr>
<td><strong>Logistic Plot</strong></td>
</tr>
<tr>
<td><strong>Plot Options</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>
Table 7.3 Logistic Fit Options  (Continued)

<table>
<thead>
<tr>
<th>ROC Curve</th>
<th>A Receiver Operating Characteristic curve is a plot of sensitivity by (1 – specificity) for each value of x. See “ROC Curves,” p. 229.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lift Curve</td>
<td>Produces a lift curve for the model. A lift curve shows the same information as a ROC curve, but in a way to dramatize the richness of the ordering at the beginning. The Y-axis shows the ratio of how rich that portion of the population is in the chosen response level compared to the rate of that response level as a whole. See the Modeling and Multivariate Methods book for details about lift curves.</td>
</tr>
<tr>
<td>Save Probability Formula</td>
<td>Creates new data table columns that contain formulas. See “Save Probability Formula,” p. 231.</td>
</tr>
<tr>
<td>Script</td>
<td>This menu contains options that are available to all platforms. They enable you to redo the analysis or save the JSL commands for the analysis to a window or a file. See “Script Menus,” p. 15 in the “Preliminaries” chapter.</td>
</tr>
</tbody>
</table>

ROC Curves

Suppose you have an x value that is a diagnostic measurement and you want to determine a threshold value of x that indicates the following:

- A condition exists if the x value is greater than the threshold.
- A condition does not exist if the x value is less than the threshold.

For example, you could measure a blood component level as a diagnostic test to predict a type of cancer. Now consider the diagnostic test as you vary the threshold and, thus, cause more or fewer false positives and false negatives. You then plot those rates. The ideal is to have a very narrow range of x criterion values that best divides true negatives and true positives. The Receiver Operating Characteristic (ROC) curve shows how rapidly this transition happens, with the goal being to have diagnostics that maximize the area under the curve.

Two standard definitions used in medicine are as follows:

- **Sensitivity**, the probability that a given x value (a test or measure) correctly predicts an existing condition. For a given x, the probability of incorrectly predicting the existence of a condition is 1 – sensitivity.
- **Specificity**, the probability that a test correctly predicts that a condition does not exist.

A ROC curve is a plot of sensitivity by (1 – specificity) for each value of x. The area under the ROC curve is a common index used to summarize the information contained in the curve.

When you do a simple logistic regression with a binary outcome, there is a platform option to request a ROC curve for that analysis. After selecting the ROC Curve option, a dialog box asks you to specify which level to use as “positive”. 

---

**Chapter 7**

**Performing Simple Logistic Regression**

**Logistic Platform Options**

**ROC Curves**

**Table 7.3 Logistic Fit Options  (Continued)**

<table>
<thead>
<tr>
<th>ROC Curve</th>
<th>A Receiver Operating Characteristic curve is a plot of sensitivity by (1 – specificity) for each value of x. See “ROC Curves,” p. 229.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lift Curve</td>
<td>Produces a lift curve for the model. A lift curve shows the same information as a ROC curve, but in a way to dramatize the richness of the ordering at the beginning. The Y-axis shows the ratio of how rich that portion of the population is in the chosen response level compared to the rate of that response level as a whole. See the Modeling and Multivariate Methods book for details about lift curves.</td>
</tr>
<tr>
<td>Save Probability Formula</td>
<td>Creates new data table columns that contain formulas. See “Save Probability Formula,” p. 231.</td>
</tr>
<tr>
<td>Script</td>
<td>This menu contains options that are available to all platforms. They enable you to redo the analysis or save the JSL commands for the analysis to a window or a file. See “Script Menus,” p. 15 in the “Preliminaries” chapter.</td>
</tr>
</tbody>
</table>
**Example of ROC Curves**

To demonstrate ROC curves, proceed as follows:

1. Open the Penicillin.jmp sample data table.
2. Select **Analyze > Fit Y by X**.
3. Select **Response** and click **Y, Response**.
4. Select **In(Dose)** and click **X, Factor**.
   
   Notice that JMP automatically fills in **Count** for **Freq**. **Count** was previously assigned the role of **Freq**.
5. Click **OK**.
6. From the red triangle menu, select **ROC Curve**.
7. Select **Cured** as the positive.
8. Click **OK**.

**Note**: This example shows a ROC Curve for a nominal response. For details about ordinal ROC curves, see the Recursive Partitioning chapter in the *Modeling and Multivariate Methods* book.

The results for the response by In(Dose) example are shown here. The ROC curve plots the probabilities described above, for predicting response. Note that in the ROC Table, the row with the highest Sens-(1-Spec) is marked with an asterisk.

**Figure 7.5 Examples of ROC Curve and Table**

![ROC Curve and Table](image-url)
If a test predicted perfectly, it would have a value above which the entire abnormal population would fall and below which all normal values would fall. It would be perfectly sensitive and then pass through the point (0,1) on the grid. The closer the ROC curve comes to this ideal point, the better its discriminating ability. A test with no predictive ability produces a curve that follows the diagonal of the grid (DeLong, et al. 1988).

The ROC curve is a graphical representation of the relationship between false-positive and true-positive rates. A standard way to evaluate the relationship is with the area under the curve, shown below the plot. In the plot, a yellow line is drawn at a 45 degree angle tangent to the ROC Curve. This marks a good cutoff point under the assumption that false negatives and false positives have similar costs.

Save Probability Formula

The Save Probability Formula option creates new data table columns. These data table columns save the following:

- formulas for linear combinations (typically called logits) of the $x$ factor
- prediction formulas for the response level probabilities
- a prediction formula that gives the most likely response

Inverse Prediction

Inverse prediction is the opposite of prediction. It is the prediction of $x$ values from given $y$ values. But in logistic regression, instead of a $y$ value, you have the probability attributed to one of the $Y$ levels. This feature only works when there are two response categories (a binary response). For example, in the rabbits study, you might want to know what dose of penicillin results in a 0.5 probability of curing a rabbit. In this case, the inverse prediction for 0.5 is called the $ED50$, the effective dose corresponding to a 50% survival rate.

Example of Inverse Prediction Using the Crosshair Tool

For inverse prediction, you predict the $x$ value that corresponds to a given probability of the lower response category. You use the crosshair tool to visually approximate an inverse prediction. See Figure 7.6, which shows the logistic plot that predicts $Response$ for given $ln(dose)$ values.

To see which value of $ln(dose)$ is equally likely either to cure or to be lethal, proceed as follows:

1. Open the Penicillin.jmp sample data table.
2. Select Analyze > Fit Y by X.
4. Select ln(Dose) and click X, Factor.
   Notice that JMP automatically fills in Count for Freq. Count was previously assigned the role of Freq.
5. Click OK.
6. Click on the crosshairs tool.
7. Place the horizontal crosshair line at about 0.5 on the vertical ($Response$) probability axis.
8. Move the cross-hair intersection to the prediction line, and read the $\ln(\text{dose})$ value that shows on the horizontal axis.

In this example, a rabbit with a $\ln(\text{dose})$ of approximately -0.9 is equally likely to be cured as it is to die.

![Figure 7.6 Example of Crosshair Tool on Logistic Plot](image)

**Example of Inverse Prediction Using the Inverse Prediction Option**

If your response has exactly two levels, the Inverse Prediction option enables you to request an exact inverse prediction. You are given the $x$ value corresponding to a given probability of the lower response category, as well as a confidence interval for that $x$ value.

To use the Inverse Prediction option, proceed as follows:

1. Open the Penicillin.jmp sample data table.
2. Select Analyze > Fit Y by X.
4. Select $\ln(\text{Dose})$ and click X, Factor.
   
   Notice that JMP automatically fills in Count for Freq. Count was previously assigned the role of Freq.
5. Click OK.
6. From the red triangle menu, select Inverse Prediction. See Figure 7.7.
7. Type 0.95 for the Confidence Level.
8. Select Two sided for the confidence interval.
9. Request the response probability of interest. Type 0.5 and 0.9 for this example, which indicates you are requesting the values for $\ln(\text{Dose})$ that correspond to a 0.5 and 0.9 probability of being cured.
10. Click OK.

The Inverse Prediction plot appears.
The estimates of the $x$ values and the confidence intervals are shown in the report as well as in the probability plot. For example, the value of $\ln(\text{Dose})$ that results in a 90% probability of being cured is estimated to be between -0.526 and 0.783.

The Fit Model platform also has an option that gives an inverse prediction with confidence limits. The *Modeling and Multivariate Methods* book gives more information about inverse prediction.

### Example of Ordinal Logistic Regression

This example uses the *AdverseR.jmp* sample data table to illustrate an ordinal logistic regression. Suppose you want to model the severity of an adverse event as a function of treatment duration value.

1. Open the *AdverseR.jmp* sample data table.
2. Right-click on the icon to the left of ADR SEVERITY and change the modeling type to ordinal.
Performing Simple Logistic Regression

Example of Ordinal Logistic Regression

3. Select **Analyze > Fit Y by X**.
4. Select **ADR SEVERITY** and click **Y, Response**.
5. Select **ADR DURATION** and click **X, Factor**.
6. Click **OK**.

![Figure 7.9 Example of Ordinal Logistic Report](image)

You interpret this report the same way as the nominal report. See “Logistic Report,” p. 223.

In the plot, markers for the data are drawn at their x-coordinate. When several data points appear at the same y position, the points are jittered. That is, small spaces appear between the data points so you can see each point more clearly.

Where there are many points, the curves are pushed apart. Where there are few to no points, the curves are close together. The data pushes the curves in that way because the criterion that is maximized is the product of the probabilities fitted by the model. The fit tries to avoid points attributed to have a small probability,
which are points crowded by the curves of fit. See the *Modeling and Multivariate Methods* book for more about computational details.

For details about the Whole Model Test report and the Parameter Estimates report, see “Logistic Report,” p. 223. In the Parameter Estimates report, an intercept parameter is estimated for every response level except the last, but there is only one slope parameter. The intercept parameters show the spacing of the response levels. They always increase monotonically.

### Additional Example of a Logistic Plot

This example uses the Car Physical Data.jmp sample data table to show an additional example of a logistic plot. Suppose you want to use weight to predict car size (Type) for 116 cars. Car size can be one of the following, from smallest to largest: Sporty, Small, Compact, Medium, or Large.

1. Open the Car Physical Data.jmp sample data table.
2. In the Columns panel, right-click on the icon to the left of Type, and select Ordinal.
3. Right-click on Type, and select Column Info.
4. From the Column Properties menu, select Value Ordering.
5. Move the data in the following top-down order: Sporty, Small, Compact, Medium, Large.
6. Click OK.
7. Select Analyze > Fit Y by X.
8. Select Type and click Y, Response.
9. Select Weight and click X, Factor.
10. Click OK.

   The report window appears.

---

**Figure 7.10** Example of Type by Weight Logistic Plot
Performing Simple Logistic Regression

In Figure 7.10, note the following observations:

- The first (bottom) curve represents the probability that a car at a given weight is Sporty.
- The second curve represents the probability that a car is Small or Sporty. Looking only at the distance between the first and second curves corresponds to the probability of being Small.
- As you might expect, heavier cars are more likely to be Large.
- Markers for the data are drawn at their $x$-coordinate, with the $y$ position jittered randomly within the range corresponding to the response category for that row.

If the $x$-variable has no effect on the response, then the fitted lines are horizontal and the probabilities are constant for each response across the continuous factor range. Figure 7.11 shows a logistic plot where Weight is not useful for predicting Type.

**Figure 7.11** Examples of Sample Data Table and Logistic Plot Showing No $y$ by $x$ Relationship

<table>
<thead>
<tr>
<th></th>
<th>Weight</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2500</td>
<td>Large</td>
</tr>
<tr>
<td>2</td>
<td>2500</td>
<td>Medium</td>
</tr>
<tr>
<td>3</td>
<td>2500</td>
<td>Compact</td>
</tr>
<tr>
<td>4</td>
<td>2500</td>
<td>Small</td>
</tr>
<tr>
<td>5</td>
<td>2500</td>
<td>Sporty</td>
</tr>
<tr>
<td>6</td>
<td>3500</td>
<td>Large</td>
</tr>
<tr>
<td>7</td>
<td>3500</td>
<td>Medium</td>
</tr>
<tr>
<td>8</td>
<td>3500</td>
<td>Compact</td>
</tr>
<tr>
<td>9</td>
<td>3500</td>
<td>Small</td>
</tr>
<tr>
<td>10</td>
<td>3500</td>
<td>Sporty</td>
</tr>
</tbody>
</table>

**Note:** To re-create the plots in Figure 7.11 and Figure 7.12, you must first create the data tables shown here, and then perform steps 7-10 at the beginning of this section.

If the response is completely predicted by the value of the factor, then the logistic curves are effectively vertical. The prediction of a response is near certain (the probability is almost 1) at each of the factor levels. Figure 7.12 shows a logistic plot where Weight almost perfectly predicts Type.

**Note:** In this case, the parameter estimates become very large and are marked *unstable* in the regression report.
Figure 7.12 Examples of Sample Data Table and Logistic Plot Showing an Almost Perfect y by x Relationship

<table>
<thead>
<tr>
<th>Weight</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2500</td>
</tr>
<tr>
<td>2</td>
<td>2600</td>
</tr>
<tr>
<td>3</td>
<td>3500</td>
</tr>
<tr>
<td>4</td>
<td>3600</td>
</tr>
<tr>
<td>5</td>
<td>4000</td>
</tr>
<tr>
<td>6</td>
<td>4600</td>
</tr>
<tr>
<td>7</td>
<td>4800</td>
</tr>
<tr>
<td>8</td>
<td>4800</td>
</tr>
<tr>
<td>9</td>
<td>4500</td>
</tr>
<tr>
<td>10</td>
<td>4500</td>
</tr>
</tbody>
</table>
Performing Simple Logistic Regression

Additional Example of a Logistic Plot
Comparing Paired Data
Using the Matched Pairs Platform

The Matched Pairs platform compares the means between two or more correlated variables and assesses the differences. For example, you might compare a blood pressure measurement taken on the same subject before a treatment and again after the treatment. A statistical method called the paired t-test takes the correlated responses into account.

The platform produces a graph of the paired differences by the paired means, and the paired t-test results for all three alternative hypotheses. Additional features provide for more than two matched responses and for a grouping column to test across samples, in a simple version of repeated measures analysis.

**Figure 8.1 Example of Matched Pairs Analysis**
Contents

Overview of the Matched Pairs Platform .................................................. 241
Example of Comparing Matched Pairs ....................................................... 241
Launch the Matched Pairs Platform .......................................................... 242
  Multiple Y Columns .............................................................................. 243
The Matched Pairs Report ...................................................................... 243
  Difference Plot and Report .................................................................. 244
  Across Groups ..................................................................................... 244
Matched Pairs Options .......................................................................... 245
Example of Comparing Matched Pairs Across Groups .............................. 246
Statistical Details ................................................................................. 247
  Graphics for Matched Pairs ............................................................... 247
  Correlation of Responses ................................................................. 249
  Comparison of Matched Pairs Analysis to Other t-Tests ...................... 249
Overview of the Matched Pairs Platform

The Matched Pairs platform compares row-by-row differences between two response columns using a paired \( t \)-test. Often, the two columns represent measurements on the same subject before and after some treatment. Alternatively, the measurements could represent data taken on the same subject with two different instruments.

If you have paired data arranged in two data table columns, then you are ready to use the Matched Pairs platform. However, if all of your measurements are in a single column, then perform one of the following tasks:

- Use the **Split** option in the **Tables** menu to split the column of measurements into two columns. Then you can use the Matched Pairs platform.
- For two response columns, create a third column that calculates the difference between the two responses. Then test that the mean of the difference column is zero with the Distribution platform.
- For the two responses stored in a single column, you can do a two-way analysis of variance. One factor (the ID variable) identifies the two responses and the other factor identifies the subject. Use the Fit Y by X Oneway platform with a blocking variable (the subject column), or use the Fit Model platform to do a two-way ANOVA. The test on the ID factor is equivalent to the paired \( t \)-test.

**Note:** If the data are paired, do not do a regular independent \( t \)-test. Do not stack the data into one column and use the Fit Y by X One-way ANOVA on the ID without specifying a block variable. To do this has the effect of ignoring the correlation between the responses. This causes the test to overestimate the effect if responses are negatively correlated, or to underestimate the effect if responses are positively correlated.

Example of Comparing Matched Pairs

This example uses the Therm.jmp sample data table. The data contains temperature measurements on 20 people. Temperature is measured using two types of thermometers: oral and tympanic (ear). You want to determine whether the two types of thermometers produce equal temperature readings. Note that the differences in temperature between the different people are not important. The matched pairs analysis is testing the differences between the thermometers.

1. Open the Therm.jmp sample data table.
2. Select **Analyze > Matched Pairs**.
3. Select Oral and Tympanic and click **Y, Paired Response**.
4. Click **OK**. The report window appears.
Comparing Paired Data

Example of Comparing Matched Pairs

The results show that, on average, the tympanic thermometer measures 1.12 degrees higher than the oral thermometer. The small p-value (Prob > |t|) indicates that this difference is statistically significant, and not due to chance.

Note that this matched pairs analysis does not indicate which thermometer is correct (if either), but only indicates that there is a difference between the thermometers.

Launch the Matched Pairs Platform

Launch the Matched Pairs platform by selecting Analyze > Matched Pairs.
Chapter 8
Comparing Paired Data

The Matched Pairs Report

For more details about launch windows, see “Launch Window Features,” p. 13 in the “Preliminaries” chapter.

**Multiple Y Columns**

You can have more than two responses. If the number of responses is odd, all possible pairs are analyzed. Table 8.2 shows an example for three responses.

**Table 8.2 Matrix for an Odd Number of Responses**

<table>
<thead>
<tr>
<th>Y1 by Y2</th>
<th>Y1 by Y3</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Y2 by Y3</td>
</tr>
</tbody>
</table>

If the number of responses is even, the Matched Pairs platform asks if you want to do all possible pairs. If you do not do all possible pairs, adjacent responses are analyzed as a pair. Table 8.3 shows the arrangement of analyses for four responses.

**Table 8.3 Matrix for an Even Number of Responses**

<table>
<thead>
<tr>
<th>Y1 by Y2</th>
<th>Y3 by Y4</th>
</tr>
</thead>
</table>

**The Matched Pairs Report**

Follow the instructions in “Example of Comparing Matched Pairs,” p. 241 to produce the report window shown in Figure 8.4.

The Matched Pairs report shows a Tukey mean-difference plot, summary statistics, and the results of the paired t-test. See “Difference Plot and Report,” p. 244. If you specified an **X, Grouping** variable, the report also includes the Across Groups report. See “Across Groups,” p. 244.
The Matched Pairs Report

Figure 8.4 Example of Matched Pairs Report

Note: The red triangle menu provides additional options that can add reports to the initial report window. See “Matched Pairs Options,” p. 245.

Difference Plot and Report

The Difference plot shows differences by means. In the Difference plot, note the following:

- The mean difference is shown as the horizontal line, with the 95% confidence interval above and below shown as dotted lines. If the confidence region includes zero, then the means are not significantly different at the 0.05 level. In this example the difference is significant.

- If you add a reference frame, the mean of pairs is shown by the vertical line. For details about a reference frame, see “Matched Pairs Options,” p. 245.

The Difference report shows the mean of each response, the difference of the means, and a confidence interval for the difference. The Difference report also shows the results of the paired t-test.

Across Groups

Note: The Across Groups report appears only if you have specified an X, Grouping variable.

To demonstrate the Across Groups report, proceed as follows:

1. Open the Dogs.jmp sample data table.
2. Select Analyze > Matched Pairs.
3. Select LogHist0 and LogHist1 and click Y, Paired Response.
4. Select drug and click X, Grouping.
5. Click OK. The Across Groups report appears within the report.

**Figure 8.5 Example of Across Groups Report**

<table>
<thead>
<tr>
<th>drug</th>
<th>Count</th>
<th>Mean Difference</th>
<th>Mean Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>morphine</td>
<td>8</td>
<td>0.6032</td>
<td>-2.284</td>
</tr>
<tr>
<td>trimeth</td>
<td>8</td>
<td>1.5217</td>
<td>-1.958</td>
</tr>
</tbody>
</table>

Test Across Groups

<table>
<thead>
<tr>
<th>F Ratio</th>
<th>Prob F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Difference</td>
<td>0.7566</td>
</tr>
<tr>
<td>Mean Mean</td>
<td>0.6803</td>
</tr>
</tbody>
</table>

The **Mean Difference** column shows the mean of the difference across rows in each group between the two paired columns. In other words, this is the within-subject by across-subject interaction, or split-plot by whole-plot interaction.

The **Mean Mean** column shows the mean of the mean across rows in each group across the two paired columns. In other words, this is the across-subject or whole-plot effect.

Within the **Test Across Groups** column, two F-tests determine whether the across-groups values are different:

- **Mean Difference** tests that the change across the pair of responses is different in different groups.
- **Mean Mean** tests that the average response for a subject is different in different groups.

The across-groups analysis corresponds to a simple repeated measures analysis. You can get the same test results using the **Manova** personality of the Fit Model platform.

See “Example of Comparing Matched Pairs Across Groups,” p. 246 for an example showing both the Matched Pairs Across Groups report and the corresponding MANOVA report.

### Matched Pairs Options

The following table describes the options in the Matched Pairs red triangle menu.

<table>
<thead>
<tr>
<th>Plot Dif by Mean</th>
<th>Shows or hides the plot of the paired differences by paired means. For a detailed description of this plot, see “Difference Plot and Report,” p. 244.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plot Dif by Row</td>
<td>Shows or hides the plot of paired differences by row number.</td>
</tr>
</tbody>
</table>
Example of Comparing Matched Pairs Across Groups

This example uses the Dogs.jmp sample data table. This example shows you how to produce both a Matched Pairs Across Groups report and the corresponding MANOVA report using Fit Model.

1. Open the Dogs.jmp sample data table.
2. Select Analyze > Matched Pairs.
3. Select LogHist0 and LogHist1 and click Y, Paired Response.
4. Select drug and click X, Grouping.
5. Click OK.

The report on the left in Figure 8.6 appears.

Now produce the Fit Model report using the same data table.

1. Select Analyze > Fit Model.
2. Select LogHist0 and LogHist1 and click Y.
3. Select drug and click Add.
4. Select the Manova personality.
5. Click **Run Model**.
6. In the Response Specification report, select **Repeated Measures** from the **Choose Response** menu.
7. Click **OK**.

**Figure 8.6 Examples of Matched Pairs Across Groups and Fit Model MANOVA with Repeated Measures**

The F Ratio for the Mean Difference in the Across Groups report corresponds to the F Ratio for Time*drug under the Within Subjects report. The F Ratio for the Mean Mean in the Across Groups report corresponds to the F Ratio for drug under Between Subjects in the Manova Fit report.

### Statistical Details

#### Graphics for Matched Pairs

The primary graph in the platform is a Tukey mean-difference (Cleveland 1994, p. 130), which plots the difference of the two responses on the y-axis against the mean of the two responses on the x-axis. This graph is the same as a scatterplot of the two original variables, but turned 45 degrees. A 45 degree rotation and rescaling turns the original coordinates into a difference and a mean.
Figure 8.7 Example of Transforming to Difference by Mean, Rotated by 45 Degrees

Before rotation, the axes represent $y_1$ and $y_2$.

\[
\begin{bmatrix}
1 & 1 \\
-1 & 1
\end{bmatrix}
\begin{bmatrix}
y_1 \\
y_2
\end{bmatrix}
= 
\begin{bmatrix}
y_2 + y_1 \\
y_2 - y_1
\end{bmatrix}
\]

After rotation, the axes represent a sum and difference.

region where $y_1 < y_2$.

line where $y_1 = y_2$.

region where $y_1 > y_2$.

mean difference and 95% CI

rescaled to mean of responses
Correlation of Responses

In most cases where the pair of measurements is taken from the same individual at different times, they are positively correlated. However, if they represent competing responses, the correlation can be negative.

Figure 8.8 shows how the positive correlation of the two responses becomes the small variance on the difference (the y-axis). If the correlation is negative, the ellipse is oriented in the other direction and the variance of the rotated graph is large on the y-axis.

Comparison of Matched Pairs Analysis to Other t-Tests

This example uses the Dogs.jmp sample data table to compare the Matched Pairs analysis (a paired t-test) result to the following:

- a one-sample t-test result, which produces the same correct results
- a two-sample independent t-test result, which produces incorrect results, because it ignores the correlation between the responses

Creating the Matched Pairs Analysis

1. Open the Dogs.jmp sample data table.
2. Select Analyze > Matched Pairs.
3. Select LogHist0 and LogHist1 and click Y, Paired Response.
4. Click OK.

The report on the left shows the paired t-test result. See Figure 8.9.

Comparing the Difference to Zero

Using the Dogs.jmp sample data table, proceed as follows:

1. Select Analyze > Distribution.
2. Select diff and click Y, Columns.
3. Click OK.
4. From the red triangle menu for \textit{diff}, select \textit{Test Mean}.
5. In the window that appears, type 0 for \textit{Specify Hypothesized Mean}.
6. Click \textit{OK}.

The report on the right shows the one-sample t-test result.

**Figure 8.9 Example of Compare Matched Pairs Analysis and One-Sample t-Test**

You can see that the Matched Pairs test result is equivalent to the one-sample t-test result on the differences of the responses. Now, compare the Matched Pairs test result to the two-sample independent t-test results.

**Creating the Two-Sample Independent t-Test**

Using the \textit{Dogs.jmp} sample data table, stack the \textit{LogHist0} and \textit{LogHist1} columns as follows:

1. Select \textit{Tables > Stack}.
2. Select \textit{LogHist0} and \textit{LogHist1} and click \textit{Stack Columns}.
3. Under \textit{Non-stacked columns}, select the \textit{Drop All} option.
4. Click \textit{OK}.

You now have a data table with two columns, \textit{Label} and \textit{Data}:

- The \textit{Data} column contains all of the values that were in the \textit{LogHist0} and \textit{LogHist1} columns in the original data table.
- The \textit{Label} column indicates in which of those two columns the row's data was originally.

Now, perform a one-way analysis and test the means, using the new stacked table you just created.

1. Select \textit{Analyze > Fit Y by X}.
2. Select \textit{Label} and click \textit{X, Factor}.
3. Select \textit{Data} and click \textit{Y, Response}.
4. Click \textit{OK}.
5. From the red triangle menu for Oneway Analysis, select \textit{Means/Anova/Pooled t}.

This produces the two-sample independent t-test report in Figure 8.10.
Compared to the report on the left in Figure 8.9, Figure 8.10 shows that the Matched Pairs test result is not equivalent to the two-sample independent t-test result. This is because the two-sample independent t-test does not account for the differences between the individual dogs. The Matched Pairs analysis adjusts for the differences between the dogs before comparing the two responses (LogHist0 and LogHist1). In other words, when you performed the two-sample independent t-test using Fit Y by X, the fact that the two responses are paired is lost.
Graph Builder provides a workspace where you can interactively create and modify graphs. Graph types include points, lines, bars, histograms, and more. You can use techniques such as trellising and overlay to examine relationships between several variables on the same graph.

**Figure 9.1 Examples Using Graph Builder**
## Contents

- Overview of Graph Builder ................................................................. 255
- Example Using Graph Builder ......................................................... 255
- Launch Graph Builder ....................................................................... 261
- The Graph Builder Window .............................................................. 262
- Graph Builder Options ..................................................................... 264
- Add Variables ..................................................................................... 268
- Change Variable Roles ...................................................................... 271
- Remove Variables ............................................................................ 272
- Add Multiple Variables to the X or Y Zone ...................................... 274
- Add Multiple Variables to Grouping Zones ...................................... 282
- Modify the Legend ............................................................................ 285
- Create Map Shapes ........................................................................... 286
  - Built-in Map Files .......................................................................... 287
  - Create Custom Map Files ............................................................... 288
- Additional Examples Using Graph Builder ......................................... 289
  - Examine Diamond Characteristics .................................................. 301
Overview of Graph Builder

You can interact with Graph Builder to create visualizations of your data. You start with a blank slate and drag and drop variables to place them where you want them. Instant feedback encourages exploration and discovery. Change your mind and move variables to new positions, or right-click to change your settings.

Graph Builder helps you see multi-dimensional relationships in your data with independent grouping variables for side-by-side or overlaid views. With many combinations to compare, you can create a trellis display of small graphs. Graph elements supported by Graph Builder include points, lines, bars, histograms, box plots, and contours.

The underlying philosophy of Graph Builder is to see your data. To that end, the default visualization elements impose no assumptions, such as normality. If there are not too many observations, you see all of them as marks on the graph. A smooth trend curve follows the data instead of an equation.

Once you see the data, you can draw conclusions directly, or decide where further analysis is needed to quantify relationships.

Example Using Graph Builder

You have data about nutrition information for candy bars. You want to find out which factors can best predict calorie levels. Working from a basic knowledge of food science, you believe that the fat content is a good place to start.

1. Open the Candy Bars.jmp sample data table.
2. Select Graph > Graph Builder.
3. Drag and drop Total fat g into the X zone.
4. Drag and drop Calories into the Y zone.
As you suspected, the candy bars with higher fat grams also have higher calories. But the relationship is not perfect. You can add other factors to try to increase the correlation. Next, determine whether cholesterol has an effect.

5. Drag and drop Cholesterol g into the **Group X** zone.
Eight levels of the variable make the graph difficult to read. Try putting Cholesterol g into the Wrap zone instead.

6. Click on Cholesterol g in the Group X zone and drag and drop it into the Wrap zone.

A scatterplot of Calories versus Total fat g is created for every level of Cholesterol g.
Figure 9.4 Example of Calories Versus Total fat g by Cholesterol g in the Wrap Zone

You can see that some of the cells have very little data; other cells have a lot of data. Among the cells that have a lot of data (cholesterol equals 0, 5, 10), there is still considerable variation in calories. So you decide to remove Cholesterol g.

7. Remove Cholesterol g by right-clicking on the Cholesterol g label in the Group X zone and selecting Remove.

Determine whether carbohydrates has any effect.

8. Drag and drop Carbohydrate g into the Wrap zone.
Carbohydrate g is a continuous variable with many values, so Graph Builder uses the percentiles to create five ranges of carbohydrate g levels. About the same number of points are displayed in each group. You can see that the relationship between calories and fat is relatively strong for each level of carbohydrate. It appears that carbohydrates adds additional predictive ability.

Now that you have determined that carbohydrates have a significant impact on calories, combine the five scatterplots into one scatterplot to directly compare the lines. You still want to identify the carbohydrate levels.

9. Drag and drop the Carbohydrate g label from the Group X zone to the Overlay zone.

The scatterplots combine into one, and the carbohydrate levels are individually colored.
Modify the legend title.

10. Right-click on the legend title (Carbohydrate g) and select **Legend Settings**.

11. Rename the Title to Carbohydrate grams.

12. Uncheck the marker items to remove them from the legend. Leave only the line items checked.

**Note**: For details about making changes to the legend, see “Modify the Legend,” p. 285.

13. Now that you are satisfied with this graph, click **Done**.
You now have a presentation-friendly graph that you can copy and paste outside of JMP.

To copy the entire graph:
14. Click the Selection Tool.
15. Click anywhere on the Graph Builder title bar.

The entire area is highlighted and ready to copy.

Launch Graph Builder

Launch Graph Builder by selecting Graph > Graph Builder.
The primary element in the Graph Builder window is the graph area. The graph area contains drop zones, and you can drag and drop variables from the Select Columns box into the zones. The following table describes the Graph Builder drop zones.

Table 9.1 Descriptions of Graph Builder Drop Zones

| X, Y | Drop variables here to assign them the X or Y role. Once a variable is placed here, no variable can be placed in Shape. |

For additional details, see the following sections:

- “Add Variables,” p. 268
- “Change Variable Roles,” p. 271
- “Remove Variables,” p. 272
- “Add Multiple Variables to the X or Y Zone,” p. 274
Table 9.1 Descriptions of Graph Builder Drop Zones (Continued)

<table>
<thead>
<tr>
<th>Drop Zones</th>
<th>Description</th>
</tr>
</thead>
</table>
| **Group X** | Subsets or partitions the data based on the variable or variables that you select. Displays the variable horizontally. Once a variable is placed here, no variable can be placed in **Wrap**.  
For additional details, see the following sections:  
- “Add Variables,” p. 268  
- “Change Variable Roles,” p. 271  
- “Remove Variables,” p. 272  
- “Add Multiple Variables to Grouping Zones,” p. 282 |
| **Group Y** | Subsets or partitions the data based on the variable or variables that you select. Displays the variable vertically.  
For additional details, see the following sections:  
- “Add Variables,” p. 268  
- “Change Variable Roles,” p. 271  
- “Remove Variables,” p. 272  
- “Add Multiple Variables to Grouping Zones,” p. 282 |
| **Shape** | Drop variables here to create map shapes. See “Create Map Shapes,” p. 286. This zone disappears if you have a variable in the X or Y zones. |
| **Wrap** | Subsets or partitions the data based on the variable or variables that you select. Wraps the data horizontally and vertically. Once a variable is placed here, no variable can be placed in **Group X**. |
| **Freq** | Drop a variable here to use it as a frequency or weight for graph elements that use statistics, such as mean or counts. |
| **Overlay** | Groups the Y variables by the selected variable, overlays the responses, and marks the levels with different colors. |
| **Color** | Drop variables here to color the graph:  
- If you are using a map, the map shapes are colored. See “Change Colors and Transparency,” p. 288.  
- If you are using a contour plot, colored contours appear.  
- If your graph contains points, they are colored. |
| **Legend** | Shows descriptions of graph elements. If you attempt to drop a variable here, the variable defaults to **Overlay**. For additional details, see “Modify the Legend,” p. 285. |
If you drop variables into the center area, JMP guesses which drop zone to put them into, based on whether the variables are continuous, ordinal or nominal.

- The X, Y, and Shape zones are positional, and influence the types of graph elements that are available.
- The Group X, Group Y, Wrap, and Overlay zones partition the data into subsets and lay out multiple graphs by either dividing the graph space or by overlaying the graphs.
- The Color and Freq zones modify certain graph elements.

Platform Buttons

There are three buttons on the Graph Builder window:

- **Undo** reverses the last change made to the window.
- **Start Over** returns the window to the default condition, removing all data and graph elements from the window, and all variables from the drop zones.
- **Done** hides the buttons and Select Columns box and removes all drop zone outlines. In this presentation-friendly format, you can copy the graph to other programs. To copy the graph, select **Edit > Copy**. To restore the window to the interactive mode, click **Show Control Panel** on the Graph Builder red triangle menu.

**Note:** For details about red triangle options or right-click menus, see “Graph Builder Options,” p. 264.

Graph Builder Options

The red triangle menu for Graph Builder contains these options:

- **Show Control Panel** shows or hides the platform buttons, the Select Columns box, and the drop zone borders.
- **Show Legend** shows or hides the legend.
- **Lock Scales** prevents axis scales and gradient legend scales from automatically adjusting in response to data or filtering changes.
- **Sampling** uses a random sample of the data to speed up graph drawing. If the sample size is zero, or greater than or equal to the number of rows in the data table, then sampling is turned off.
- **Script** contains options that are available to all platforms. They enable you to redo the analysis or save the JSL commands for the analysis to a window or a file. See “Script Menus,” p. 15 in the “Preliminaries” chapter.

Graph Builder Right-Click Menus

Graph Builder contains various right-click menus. See the following sections:

- “Right-Click Menu for a Graph,” p. 265
- “Additional Right-Click Menus,” p. 266
Right-Click Menu for a Graph

Right-clicking on a graph shows a menu of the available graph elements and other options. The following table describes all of the possible graph elements. The options that you see might vary, depending on whether the data is continuous or categorical.

Table 9.2 Descriptions of All Possible Graph Elements

<table>
<thead>
<tr>
<th>Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bar</td>
<td>Adds a bar chart to the graph.</td>
</tr>
<tr>
<td>Box plot</td>
<td>Adds box plots to the graph.</td>
</tr>
<tr>
<td>Contour</td>
<td>Adds density contours to the graph. See “Examine Diamond Characteristics,” p. 301.</td>
</tr>
<tr>
<td>Histogram</td>
<td>Adds a histogram to the graph.</td>
</tr>
<tr>
<td>Line</td>
<td>Adds a line plot to the graph.</td>
</tr>
<tr>
<td>Map Shapes</td>
<td>Adds map shapes to the graph. See “Create Map Shapes,” p. 286.</td>
</tr>
<tr>
<td>Mosaic</td>
<td>Adds a mosaic plot to the graph. Only appears for categorical variables.</td>
</tr>
<tr>
<td>Points</td>
<td>Adds points representing raw data to the graph.</td>
</tr>
<tr>
<td>Smoother</td>
<td>Adds a smoother line to the graph. See “About the Smoother,” p. 267.</td>
</tr>
</tbody>
</table>

Change and Add Elements in the Graph

In the right-click menu, each graph element currently in use on the graph is listed.

- You can change any existing element by selecting the Change to option.
- You can add any element by selecting the Add option.

Any changes that you make to a graph element apply to all graphs for that variable, across all grouping variables. The following table describes the right-click menu options for a graph and describes which graph elements each option is applicable to.

Table 9.3 Descriptions of Graph Right-Click Menu Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Graph Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bar Style</td>
<td>Bar</td>
<td>Changes the bar style.</td>
</tr>
<tr>
<td>Box Style</td>
<td>Box Plot</td>
<td>Changes the box plot style.</td>
</tr>
<tr>
<td>Change to</td>
<td>All graph elements</td>
<td>Provides options for changing the current graph element to another element. For example, if you have points showing, you can change them to appear as bars.</td>
</tr>
<tr>
<td>Error Bars</td>
<td>Line, Bar, and Points</td>
<td>Adds error bars to the graph. To see this option, the Summary Statistic must be set to Mean.</td>
</tr>
</tbody>
</table>
Table 9.3 Descriptions of Graph Right-Click Menu Options (Continued)

<table>
<thead>
<tr>
<th>Option</th>
<th>Graph Element</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal</td>
<td>Mosaic</td>
<td>Changes the primary direction of the mosaic plot to horizontal.</td>
</tr>
<tr>
<td>Jitter</td>
<td>Box Plot and Points</td>
<td>Turns jitter on or off.</td>
</tr>
<tr>
<td>Move Backward</td>
<td>All graph elements</td>
<td>If you have multiple graph elements, you can move them backward (or move them to the back).</td>
</tr>
<tr>
<td>Move Forward</td>
<td>All graph elements</td>
<td>If you have multiple graph elements, you can move them forward (or move them to the front).</td>
</tr>
<tr>
<td>Number of Levels</td>
<td>Contour</td>
<td>Specify the number of contour levels to display.</td>
</tr>
<tr>
<td>Outliers</td>
<td>Box Plot</td>
<td>Shows or hides outliers.</td>
</tr>
<tr>
<td>Remove</td>
<td>All graph elements</td>
<td>Removes the graph element.</td>
</tr>
<tr>
<td>Row order</td>
<td>Line</td>
<td>Connects line points in the order of their row numbers instead of the order of their X values.</td>
</tr>
<tr>
<td>Show Missing Shapes</td>
<td>Map Shapes</td>
<td>Shows or hides missing data from the map. This option is turned off by default.</td>
</tr>
<tr>
<td>Summary Statistic</td>
<td>Line, Bar, Points, and Map Shapes</td>
<td>Provides options for changing the statistic being plotted.</td>
</tr>
<tr>
<td>Vertical</td>
<td>Mosaic</td>
<td>Changes the primary direction of the mosaic plot to vertical.</td>
</tr>
<tr>
<td>X</td>
<td>All graph elements</td>
<td>This option only appears if you have multiple X variables. You can hide or show each one.</td>
</tr>
<tr>
<td>Y</td>
<td>All graph elements</td>
<td>This option only appears if you have multiple Y variables. You can hide or show each one.</td>
</tr>
</tbody>
</table>

**Note:** For a description of the Rows, Graph, Customize, and Edit menus, see the *Using JMP* book.

**Additional Right-Click Menus**

Depending on the area that you click on in Graph Builder, you see different right-click options.
Table 9.4 Descriptions of Additional Right-Click Menus

<table>
<thead>
<tr>
<th>Right-Click Area</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Y or X axis</td>
<td>The Remove and Swap commands appear. See “Use the Remove Command,” p. 272 or “Use the Swap Command,” p. 271.</td>
</tr>
<tr>
<td></td>
<td>For descriptions of the options below the line, see the Using JMP book.</td>
</tr>
<tr>
<td></td>
<td>For descriptions of the options below the line, see the Using JMP book.</td>
</tr>
<tr>
<td>A zone</td>
<td>These options might appear:</td>
</tr>
<tr>
<td></td>
<td>• <strong>Number of Levels.</strong> See “Separate Variables into Groups,” p. 270.</td>
</tr>
<tr>
<td></td>
<td>• <strong>Color</strong> changes the background color of the grouping zone.</td>
</tr>
<tr>
<td></td>
<td>• <strong>Remove.</strong> See “Use the Remove Command,” p. 272.</td>
</tr>
<tr>
<td></td>
<td>• <strong>X or Y Group Edge.</strong> See “Move Grouping Variable Labels,” p. 270.</td>
</tr>
<tr>
<td></td>
<td>• <strong>Swap.</strong> See “Use the Swap Command,” p. 271.</td>
</tr>
<tr>
<td>An item in the legend</td>
<td>For descriptions of these options, see the Using JMP book.</td>
</tr>
<tr>
<td>The empty space below</td>
<td>See “Graph Builder Options,” p. 264.</td>
</tr>
<tr>
<td>Select Columns</td>
<td>For descriptions of these options, see the Using JMP book.</td>
</tr>
<tr>
<td>The empty space below the</td>
<td>For descriptions of these options, see the Using JMP book.</td>
</tr>
<tr>
<td>legend or above or below</td>
<td></td>
</tr>
<tr>
<td>the graph</td>
<td></td>
</tr>
</tbody>
</table>

**About the Smoother**

The smoother is a cubic spline with a lambda of 0.05 and standardized X values. You can obtain the same spline in the Bivariate platform, by performing the following steps:

1. Select **Analyze > Fit Y by X**.
2. Select continuous variables for **Y, Response** and **X, Factor**.
3. Click **OK**.
4. From the red triangle menu for Bivariate Fit, select **Fit Spline > Other**.
5. Type in the smoothness value of 0.05.
Add Variables

To add a variable to a drop zone, click and drag the column name from the Select Columns box and drop it into the desired drop zone. Alternatively, you can highlight the column name in the Select Columns box and click in the desired drop zone. Both numeric and categorical variables can be added to all of the drop zones.

Example of Adding Variables

To show an example of adding variables to Graph Builder, make a scatterplot of weight versus height for males and females.

1. Open the Students jmp sample data table.
2. Select Graph > Graph Builder.
3. Click height and drag and drop it into the X zone.
A plot of the raw data appears. Random jitter is added in the Y direction. Since \textit{height} is a continuous variable and no other variable is in the X zone, dropping \textit{height} into the center of the chart would produce the same effect.

4. Click \textit{weight} and drag and drop it into the Y zone.

**Figure 9.10** Example of a Scatterplot of weight Versus height

Notice the following additional elements:

- A smoother (line) shows the general pattern of \textit{weight} as a function of \textit{height}.
- A legend describes the elements on the graph (in this case, the smoother).

5. Click \textit{sex} and drag and drop it into the \textbf{Group X} zone.
Side-by-side scatterplots (one for each level of sex) replace the initial scatterplot. You now see weight versus height for males and females.

**Move Grouping Variable Labels**

The grouping variable labels in Figure 9.11 can be relocated to another position on the graph. The Group X labels can be either on the top or the bottom of the graph. The Group Y labels can be either on the right or the left of the graph. For example, to move the sex labels from the top of the graph to the bottom of the graph, right-click on sex and select X Group Edge > Bottom. The variable labels move from the top to the bottom.

**Separate Variables into Groups**

When you add a categorical variable to the Group X or Group Y zone, a partition is created for each level of the variable. For example, if a variable has two levels, such as F and M in Figure 9.11, then the data is divided into two groups.

If you add a continuous variable to a grouping zone, Graph Builder uses quantiles of the data to divide the variable into five groups. Once the variable is added to the display, you can change the number of groups as follows:

1. Right-click on the grouping variable label and select Number of Levels.
2. Type in the number of levels that you want to display.
3. Click **OK**.

### Change Variable Roles

Use the following methods to change variable roles:

- “Use the Swap Command,” p. 271
- “Use the Clicking and Dragging Method to Change Variable Roles,” p. 272

#### Use the Swap Command

To demonstrate the **Swap** command, start from the graph in Figure 9.11. Swap the **Y** variable, **weight**, with the **X** variable, **height**. Proceed as follows:

1. Right-click on the **Y** variable, **weight**.
2. Select **Swap > height**.

![Graph Builder](image)

**Figure 9.12** Selecting the Swap Command

The scatterplots now show **weight** in the **X** role and **height** in the **Y** role. In this example, you could have right-clicked on **height** and selected **Swap > weight** to accomplish the same swap.
Use the Clicking and Dragging Method to Change Variable Roles

To demonstrate the clicking and dragging method for changing variable roles, start from the graph in Figure 9.11. Move the **Group X** variable, sex, to be the **Group Y** variable. Drag and drop sex to the **Group Y** zone.

**Figure 9.13** Example of Moving sex to the Group Y Zone

The variable **sex** now appears in the **Group Y** role and the scatterplots are stacked.

Remove Variables

There are two different methods you can use to remove variables, as follows:

- “Use the Remove Command,” p. 272
- “Use the Clicking and Dragging Method to Remove Variables,” p. 273

Use the Remove Command

To demonstrate the **Remove** command, start from the graph in Figure 9.11. Remove the X variable, height. Proceed as follows:

1. Right-click on the X variable, height.
2. Select **Remove**.
The *height* variable is removed from the graph. You are left with a plot of the *weight* variable for each level of *sex*.

**Use the Clicking and Dragging Method to Remove Variables**

To demonstrate the clicking and dragging method for removing variables, start from the graph in Figure 9.11. Click on the variable that you want to remove and drag it into a “dead” zone. Dead zones are indicated by the shaded areas in Figure 9.15.
Add Multiple Variables to the X or Y Zone

You can assign more than one variable to the same drop zone. By visualizing multiple Y variables across multiple X variables, you can discover multivariate relationships and interactions.

To add a variable to the left or right of the existing X variable, or to add a variable to the top or bottom of the existing Y variable, drag and drop the variable into the X or Y zone.

When you click and drag a variable, a blue shape appears, symbolizing where the variable is added. The following table illustrates what each shape looks like for each task.

Table 9.5 Descriptions of Tasks and Shapes

<table>
<thead>
<tr>
<th>Task</th>
<th>Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add an X variable to the left of another X variable</td>
<td>![Shape]</td>
</tr>
<tr>
<td>Add an X variable to the right of another X variable</td>
<td>![Shape]</td>
</tr>
</tbody>
</table>
Add Multiple Variables to the X or Y Zone

Example of Adding Multiple Variables to the X or Y Zone

Start from the graph in Figure 9.9. Add the weight variable to the left of the height variable. Click weight and drag and drop it into the X axis, to the left of height.

Table 9.5 Descriptions of Tasks and Shapes (Continued)

<table>
<thead>
<tr>
<th>Task</th>
<th>Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add a Y variable to the bottom of another Y variable</td>
<td><img src="image" alt="Diagram" /></td>
</tr>
<tr>
<td>Add a Y variable to the top of another Y variable</td>
<td><img src="image" alt="Diagram" /></td>
</tr>
</tbody>
</table>

Figure 9.16 Dragging and Dropping the weight Variable
A graph is created for both variables, with separate axes and scales. If grouping variables exist, a graph is created for both the existing variable and incoming variable for each combination of grouping variables.

**Merge Variables**

*Note:* This section assumes that you have variables existing in the X or Y zone where you want to add variables. If you have no variables existing in the X or Y zone, you can select multiple variables from the Select Columns list and drag them into the X or Y zone. This merges the selected variables.

Merging variables places both variables on the same axis and creates a graph for both variables. Merging is similar to adding variables, in that a graph element is added for both variables. But adding variables maintains separate axes and scales, and merging variables use the same axis and the same scale for the variables.

To demonstrate combining two continuous variables, start from the graph in Figure 9.9. Merge the weight variable with the height variable. Drag and drop weight to the center of the X zone, slightly above the axis. Before you drop the variable, a blue pentagon appears.

![Figure 9.17 Merging height and weight](image)

A graph element is added for weight that uses the same scale as height.
Add Multiple Variables to the X or Y Zone

If you merge a categorical variable with an existing continuous variable, then the categorical variable is transformed into integer values. For example, using the Students.jmp data, if sex is merged with height, the values of sex (F and M) are transformed into 0 and 1. The transformation allows the two variables to use the same axis and scale.

If you merge a numeric variable with an existing categorical variable, the result is ordered. See “Order Variables,” p. 277.

Order Variables

The levels of a character variable on a graph are ordered alphabetically by default. There are two ways you can modify this order:

- Use the Value Ordering column property. For details, see the Using JMP book.

Example of Ordering Variables Using a Second Variable

Only a nominal or ordinal variable can be ordered, and only by a numeric variable. The variable that does the ordering has to be numeric so that an average can be computed for each level of the categorical variable.

Note: If you try to order a numeric variable with another numeric variable, JMP defaults to a merge. See “Merge Variables,” p. 276.
For example, consider data about the sizes of cars. You want to see the different car sizes in ascending order.

1. Open the Cars.jmp sample data table.
2. Select **Graph > Graph Builder**.
3. Click **Size** and drag it into the **X** zone.

This variable represents the size of the vehicle. Eight levels are listed alphabetically on the X axis: compact (comp), heavy (hev), lightweight (lt), medium (med), mini, multi-purpose (mpv), pick-up truck (pu), and van. Since the levels are listed alphabetically, they are not ordered by size. Heavy comes before mini and lightweight. Another variable, **Wt** (weight) can act as a good substitute for size.

4. Click **Wt** and drag and drop it just above the **X** axis. Before you drop the variable, a blue pentagon appears.

**Figure 9.19 Example of Merging Wt and Size**

The levels of **Size** are now ordered according to the average **Wt** of all vehicles in that level, in ascending order. Notice that mini and lightweight are now ordered before heavy. The axis label is updated, signifying that an ordering variable is in use.

To verify that **Size** is actually ordered by **Wt**, click on **Wt** under **Select Columns** and drag and drop it into the **Y** zone. Figure 9.20 shows that the average **Wt** increases from the left to right.
To change the order from ascending to descending, proceed as follows:

1. Right-click in the X zone.
2. Deselect Ascending. See Figure 9.21.

The default ordering statistic is the mean. To use another statistic, select it from the Order Statistic menu shown in Figure 9.21. The available statistics are N, Mean, Min, Max, Sum, and % of Total.

To remove the ordering, right-click and select Remove Order.

Note: For details about the Axis and Edit menu options, see the Using JMP book.
Replace Variables

You can replace an existing variable with an incoming variable, maintaining a single graph element. If grouping variables exist, a single graph element is maintained for the incoming variable for each combination of grouping variables.

To demonstrate replacing variables, start from the graph in Figure 9.9. Replace height with weight. Drag and drop weight to the center portion directly in the X zone. Before you drop the variable, a blue hexagon appears.

*Figure 9.22* Example of Replacing height with weight

Once you drop the weight variable, the height variable is replaced, and a single graph element appears for weight.

Create a Second Y Axis

If you have two Y variables, you can move one of the variables to create a second Y axis. To demonstrate, proceed as follows:

1. Open the Students.jmp sample data table.
2. Select Graph > Graph Builder.
3. Click sex and drag and drop it into the X zone.
4. Click height and weight and drag and drop them into the Y zone.
5. Right-click on the height & weight label in the Y zone, and select **Move Right > weight**.
A new Y axis is created on the right of the graph for the weight variable.
Add Multiple Variables to Grouping Zones

You can add more than one variable to the Group X or Group Y zones. You place the incoming variable below or above the existing variable, depending on how you want your data to be ordered.

To add an additional variable to a grouping zone, simply drag and drop the variable into the drop zone. When you drag a variable, a blue shape appears, symbolizing where the variable is added. The following table illustrates what each shape looks like for each task.

Table 9.6 Descriptions of Tasks and Shapes

<table>
<thead>
<tr>
<th>Task</th>
<th>Shape</th>
</tr>
</thead>
<tbody>
<tr>
<td>Add the incoming variable below the existing variable</td>
<td><img src="image1" alt="Shape" /></td>
</tr>
<tr>
<td>Add the incoming variable above the existing variable</td>
<td><img src="image2" alt="Shape" /></td>
</tr>
<tr>
<td>Replace the existing variable with the incoming variable</td>
<td><img src="image3" alt="Shape" /></td>
</tr>
</tbody>
</table>

Example of Adding Multiple Variables to Grouping Zones

You have data on popcorn. You want to see the popcorn yields, grouped by popcorn type (gourmet or plain) and by batch size (small or large).

1. Open the Popcorn jmp sample data table.
2. Select Graph > Graph Builder.
3. Click popcorn and drag and drop it into Group X.
   This is the popcorn type.
4. Click yield and drag and drop it into the Y zone.
   You can see the yield by popcorn type.
To add the second variable, perform either step 5 or step 6, depending on where you want to place the incoming variable (above or below the existing variable).

5. To place the batch variable **above** the popcorn variable, drag and drop batch into the left side of the Group X zone. Before you drop the variable, a left-justified blue polygon appears.

6. To place the batch variable **below** the popcorn variable, drag and drop batch into the right side of the Group X zone. Before you drop the variable, a right-justified blue polygon appears.
Chapter 9

Add Multiple Variables to Grouping Zones

**Figure 9.27** Example of Adding batch Below popcorn

![Diagram](image1)

7. To see the data shape more clearly, right-click in the graph and select **Add > Box Plot**.

**Figure 9.28** Examples of Popcorn yield Grouped by popcorn and batch

![Diagram](image2)

Depending on where you place the incoming variable, you can look at the data in different ways. But you could come to the same conclusion: that batch size varies for gourmet popcorn, but batch size does not vary as much for plain popcorn.

**Replace Variables**

You can replace an existing variable with an incoming variable. To demonstrate replacing variables, start from the graph in Figure 9.25. Replace **popcorn** with **batch** in the **Group X** zone. Drag and drop **batch** into the center of the **Group X** zone. Before you drop the variable, a blue quadrilateral appears.
Once you drop the variable, groups are created for batch instead of popcorn.

**Order Grouping Variables**

Ordering the levels of grouping variables works the same as ordering the levels of variables in the X zone or Y zone. See "Order Variables," p. 277.

**Modify the Legend**

To modify a legend, right-click or double-click on the legend title. The following commands appear:

- **Legend Settings** opens a window where you can modify legend settings, such as the title and title position. See Figure 9.30.
- **Revert Legend** returns the legend to the default condition (if you have changed it).

**Note:** For a description of the Edit menu options, see the Using JMP book.

*Figure 9.30 The Legend Settings Window*
Create Map Shapes

Use Graph Builder to create maps using the Shape zone. If a column contains the names of geographical regions (such as countries, states, provinces, counties), you can assign the column to the Shape zone to create a map of the regions.

Example of Creating Map Shapes

This example uses the Crime.jmp sample data table, which contains data on crime rates for each US state.

1. Open the Crime.jmp sample data table.
2. Select Graph > Graph Builder.
3. Drag and drop State into the Shape zone.
4. Drag and drop murder into the Color zone.

<table>
<thead>
<tr>
<th>Title</th>
<th>The name of the legend.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Check boxes</td>
<td>Shows or hides items in the legend.</td>
</tr>
<tr>
<td>Up and down arrows</td>
<td>Changes the order of items in the legend.</td>
</tr>
<tr>
<td>Title Position</td>
<td>Places the legend title on top or to the left of the items in the legend.</td>
</tr>
<tr>
<td>Item Direction</td>
<td>Displays the legend horizontally or vertically.</td>
</tr>
<tr>
<td>Item Wrap</td>
<td>Sets the legend to be $n$ items tall (if vertical) or $n$ items wide (if horizontal).</td>
</tr>
<tr>
<td>Preview</td>
<td>Shows your changes to the legend.</td>
</tr>
<tr>
<td>OK</td>
<td>Commits your changes to the legend.</td>
</tr>
<tr>
<td>Cancel</td>
<td>Cancels your changes to the legend.</td>
</tr>
<tr>
<td>Help</td>
<td>Opens the online Help.</td>
</tr>
</tbody>
</table>

Table 9.7 Descriptions of the Legend Settings Window Options
Figure 9.31 Example of murder by state

Note the following:

- The latitude and longitude appear on the Y and X axes.
- The legend shows the colors that correspond to the murder rates. Since murder is a continuous variable, the colors are on a gradient.

**Built-in Map Files**

When you use the Shape zone, JMP searches through its built-in map files for matching names. Built-in map files include the following:

- world countries
- states and counties in the United States
- first-level divisions for Canada, China, the United Kingdom, France, Italy, Japan, and Germany

By default, JMP installs map files in the following directory:

- **Windows**: C:\Program Files\SAS\JMP\9\Maps
- **Macintosh**: /Library/Application Support/JMP/9/Maps

Each map consists of two JMP data files with a common prefix:

- -Name file that contains the unique names for the different regions
- -XY file that contains the latitude and longitude coordinates of the boundaries
Create Map Shapes

The two files are implicitly linked by a Shape ID column. For example:

- The World-Country-Name.jmp file contains the exact names and abbreviations for different countries throughout the world.
- The World-Country-XY.jmp file contains the latitude and longitude numbers for each country, by Shape ID.

**Note:** If JMP does not recognize the names in your data, check the built-in map files for the spelling that JMP recognizes. For example, United Kingdom is a country name, but not Great Britain.

Create Custom Map Files

You can create your own map files by following the same pattern as the built-in files. In order for JMP to automatically find your files, place them in the following directory:

- **Windows XP:** `<username>\Local Settings\Application Data\SAS\JMP\Maps`
- **Windows Vista:** `<username>\AppData\Local\SAS\JMP\Maps`
- **Macintosh:** `<username>/Library/Application Support/JMP/Maps`

Or, you can link the map files to your data files explicitly with the Map Role column property. Each set of map files that you create must contain a -Name file and a -XY file. The first column in both files must be the ascending, numeric Shape ID variable. The -Name file can contain any other columns. Columns that are marked with the Map Role/Shape Name Definition are searched for shape identification and must contain unique values.

If you import an ESRI SHP file, it is opened in this format.

For explicitly linked shape data, use the Map Role/Shape Name Use column property and provide the location and column of the map file data to use.

The *Using JMP* book includes more information about how to structure the map data tables and your analysis data table. Steps for adding the Map Role column property are also included.

Change Colors and Transparency

To change colors and transparency for a map, right-click on the color bar in the legend. The right-click options vary, depending on whether the Color variable is continuous or categorical (nominal or ordinal).

However, for both types of variables, you can change the transparency.

Continuous variables use a color gradient. To change the color theme, proceed as follows:

1. Right-click on the color bar and select **Gradient**.
2. In the Gradient Settings window, select a different Color Theme.

Categorical (nominal or ordinal) variables use a singular coloring system, where each level of the variable is colored differently.
To change the color of one of the variable levels, proceed as follows:
1. Right-click on the color of the variable level that you want to change and select **Fill Color**.
2. Select the new color.

**Note:** For more details about color options, see the *Using JMP* book.

---

**Additional Examples Using Graph Builder**

This section contains the following additional examples that use Graph Builder:

- “Measure Global Oil Consumption and Production,” p. 289
- “Analyze Popcorn Yield,” p. 296
- “Examine Diamond Characteristics,” p. 301

**Measure Global Oil Consumption and Production**

You have data about oil consumption and production, measured in barrels per day, for selected countries. You want to find out which countries consume the most oil, and which countries produce the most oil.

1. Open the *Oil Use.jmp* sample data table.
2. Select **Graph > Graph Builder**.
3. Drag and drop **Country** into the Y zone. You can resize the Graph Builder window and the graph if necessary.

**Note:** Notice that the default ordering for **Country** is ascending alphabetical (starting point is at the bottom). You can change the sorting order within a data table by using either the Value Ordering or Row Order Levels commands. For details, see the *Using JMP* book.
4. Drag and drop Production and Consumption to the X zone.
Markers appear on the graph for both variables, along with a legend that identifies the different colors.
Because the default graph element is set to points, it is difficult to read and learn from the graph. Change the points to bars to make the graph easier to interpret.

5. To change the points to bars, right-click on the graph and select Points > Change to > Bar.
Experiment with the presentation of the bar graph.

6. Change the default side-by-side bars to stacked bars by right-clicking on the graph and selecting **Bar > Bar Style > Stacked**.
7. Go back to the Oil Use.jmp sample data table.
8. Double-click on an empty column heading to create a new column.
9. Double-click on the new column title, and rename it Negative Consumption. Click OK.
10. With the new column highlighted, select Cols > Formula.
11. Click Consumption.
12. Click +/-.
13. Click OK.

The new Negative Consumption column contains the negative consumption values. Use this column to see the consumption bar going in the opposite direction from the production bar. Launch Graph Builder and create the graph again.

14. Select Graph > Graph Builder.
15. Drag and drop Country into the Y zone. You can resize the Graph Builder window and the graph if necessary.
16. Drag and drop Production and Negative Consumption to the X zone.
17. Right-click on the graph and select **Points > Change to > Bar**.
18. Right-click on the graph and select **Bar > Bar Style > Stacked**.

**Figure 9.36** Example of Stacked Bars for Production and Negative Consumption

You can clearly see consumption on the left in red, and production on the right in blue. Sort the countries by their oil consumption. You can use the **Consumption** column as an ordering variable.

19. Drag and drop **Consumption** into the graph, to the right of the **Y** zone.

Before you drop the variable, a blue polygon appears.
The Country variables are now ordered by Consumption.
You can clearly see the countries that consume the most oil at the top of the graph. You can also see each country's oil production.

**Analyze Popcorn Yield**

You have data about popcorn. The data includes popcorn type (plain or gourmet), how much oil was used, the batch size (small or large), the yield, and the trial number. You want to determine how these factors affect the popcorn yield: popcorn type, oil, and batch size.

**Note:** These data are artificial, but inspired from an experiment reported in Box, Hunter, and Hunter (1978).

1. Open the Popcorn.jmp sample data table.
2. Run the attached Fit Model script attached to the data table by clicking on the red triangle next to **Fit Model** and select **Run Script**.
   A three-way ANOVA model is fit to the data.
3. Click on the disclosure button next to **Effect Tests** to open the report.

The popcorn*batch interaction has a small p-value (0.0026). From this, you conclude that there is a significant interaction between popcorn and batch.
4. Save the model predictions to the data table. From the red triangle menu next to Response yield, select **Save Columns > Prediction Formula**. See Figure 9.40.

Notice that a new column is added to the data table, **Pred Formula yield**.

5. Save the prediction intervals to the data table. From the red triangle menu next to Response yield, select **Save Columns > Mean Confidence Interval**. See Figure 9.40.

Notice that two new columns are added to the data table: **Lower 95% Mean yield** and **Upper 95% Mean yield**.

Figure 9.40 The Save Columns Menu

Now you can use Graph Builder to visualize the interaction between popcorn and batch.

6. Select **Graph > Graph Builder**.

7. Drag and drop these columns into the Y zone:
   - Pred Formula yield
   - Lower 95% Mean yield
   - Upper 95% Mean yield

8. Drag and drop popcorn into the X zone.

9. Drag and drop batch into the Group X zone.

10. Drag and drop oil amt into the Group Y zone.

11. Because the Jitter option is turned on, there are two data points per combination of factors, and the prediction at a given combination is the same. Turn off the jitter by right-clicking in the graph and deselecting **Points > Jitter**.
Format the graph to see interval bars for **Lower 95% Mean yield** and **Upper 95% Mean yield**, and to see points for **Pred Formula yield**.

12. Right-click on the graph and select **Add > Bar**.
   
   This adds bars for all three responses.

13. Change the bar style to interval by right-clicking on the graph and selecting **Bar > Bar Style > Interval**.
The interval bar style currently spans from Lower 95% Mean yield to Pred Formula yield, but you want it to span up to Upper 95% Mean yield.

14. Remove the bar element for Pred Formula yield by right-clicking on the graph and deselecting Bar > Y > Pred Formula yield.

Now the confidence interval spans from the lower to upper value.
Figure 9.43 Example of Correct Interval Span

Remove the point graph element for Lower 95% Mean yield and Upper 95% Mean yield.

15. Right-click on the graph and select Points > Y, and deselect the Lower 95% Mean yield and Upper 95% Mean yield individually.

To make the points easier to see, increase the size of the points.

16. Right-click on the graph, and select the XXL option under Graph > Marker Size. Do this for each quadrant of the graph.

**Note:** You can also change the graph title and labels for the X and Y zones. Click on the label and type in the new text.
Additional Examples Using Graph Builder

From Figure 9.44, you can see the following relationships:

- For large batches, there is no difference between plain and gourmet popcorn.
- For small batches, the gourmet popcorn has a higher yield than the plain popcorn.
- For the oil amount, the relationship is the same whether the oil amount is little or lots, so there is no three-way interaction.

Examine Diamond Characteristics

You have data about diamonds, including their carat weight and price. Examine the relationship between carat weight and price.

1. Open the Diamonds Data.jmp sample data table.
2. Select Graph > Graph Builder.
3. Drag and drop Price into the Y zone.
4. Drag and drop Carat Weight into the X zone.
5. Right-click on the graph and select Add > Contour.
Note: The 3 screenshots in this section are new and need to be added to the catalog.

Because there are numerous points, a contour plot appears by default. A smoother also appears. To see how many points there are, temporarily change the contour plot to points.

6. Right-click on the graph and select Contour > Change to > Points.
You can see that the points are difficult to interpret. Some points are likely overlapping, making the density unclear. Change the points back to a contour plot.

7. Right-click on the plot and select **Points > Change to > Contour.**
The darker the density, the more data. In Figure 9.47, notice the following:

- Most people in the sample purchased diamonds with a carat weight of about 0.5, 0.75, and 1.0.
- Most people who purchased diamonds with a carat weight of 0.5 paid about $850-$2,000 dollars.
- Most people who purchased diamonds with a carat weight of 0.75 paid about $1,600-$2,700 dollars.
- Most people who purchased diamonds with a carat weight of 1.0 paid about $3,800-$4,800 dollars.
Chapter 10
Creating Summary Charts
Using the Chart Platform

The Chart platform on the Graph menu charts continuous variables versus categorical variables. The continuous variables are summarized for each categorical level. Chart supports several chart types and arrangements. Chart is similar to the Tables > Summary command and is useful for making graphical representations of summary statistics. Here are some types of charts available from the Chart platform:

- bar and stacked bar
- pie
- needle
- line
- point
- range

If you want to make a plot of individual data points (rather than summaries of data points), we recommend using the Overlay Plot platform. See the “Creating Overlay Plots” chapter for details about overlay plots.

Figure 10.1 Examples of Charts

[Images of different types of charts as described in the text]
Contents

Example of the Chart Platform ........................................... 307
Launch the Chart Platform .................................................. 309
  Plot Statistics for Y Variables ...................................... 312
  Use Categorical Variables ........................................... 313
  Use Grouping Variables .............................................. 314
  Adding Error Bars ..................................................... 316
The Chart Report ............................................................. 317
  Legends ........................................................................ 317
  Ordering ...................................................................... 318
  Coloring Bars in a Chart .............................................. 318
Chart Platform Options ....................................................... 319
  General Platform Options ............................................ 319
  Y Options .................................................................. 321
Examples of Charts ............................................................. 321
  Plot a Single Statistic .................................................. 322
  Plot Multiple Statistics .............................................. 322
  Plot Counts of Variable Levels ................................... 323
  Plot Multiple Statistics with Two X Variables ............... 325
  Create a Stacked Bar Chart .......................................... 326
  Create a Pie Chart ..................................................... 327
  Create a Range Chart .................................................. 329
  Create a Chart with Ranges and Lines for Statistics ....... 330
Example of the Chart Platform

Here is a simple example that shows how to plot the mean height of students based on their age group.

1. Open the Students.jmp sample data table.
2. Select Graph > Chart.
3. Select height and click Statistics.
4. Select Mean from the menu of statistics.
5. Select age and click Categories, X, Levels.

Figure 10.2 The Completed Chart Launch Window

6. Click OK.
This bar chart shows the following:

- The mean height of the students in this class increases with age.
- The largest increase occurs at the earliest ages.
- The mean changes very little for the older students.

You might expect the change of the mean height to be different for males and females.

7. From the red triangle menu for Chart, select **Script > Relaunch Analysis**.

   The Chart launch window appears, already filled out for you. If you clicked **OK** now, you would see a duplicate of the chart you already have.

8. Open the Additional Roles outline.

9. Select **sex** and click **Grouping**.

10. Click **OK**.
These two bar charts confirm your assumption. The mean of height for girls in the class rises early and then remains stable. The mean of height for boys rises more dramatically overall, and also continues to increase at later ages.

Launch the Chart Platform

Launch the Chart platform by selecting **Graph > Chart**.
Creating Summary Charts

Example of the Chart Platform

Figure 10.5 The Chart Launch Window

![Chart Launch Window Diagram]

The Chart launch window contains many controls that are described in "Launch Window Features," p. 13 in the “Preliminaries” chapter. Controls that are unique to this window are described in Table 10.1.

In the Chart launch window, you can assign the following:

- Up to two X variables, which appear on the x-axis in the same order that you assign them in the launch window.
- As many Y variables (statistics) as you want. If the data is already summarized, select Data as the statistics option.

Table 10.1 Description of the Chart Launch Window

<table>
<thead>
<tr>
<th>Cast Selected Columns Into Roles</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Statistics</strong></td>
</tr>
<tr>
<td>Use this menu to select the statistic to chart for each Y variable.</td>
</tr>
<tr>
<td>See “Plot Statistics for Y Variables,” p. 312.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><strong>Categories, X, Levels</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Select up to two X variables whose levels are categories on the x-axis. The Chart platform produces a bar for each level or combination of levels of the X variables. If you do not specify an X variable, the chart has a bar for each row in the data table.</td>
</tr>
<tr>
<td>See “Use Categorical Variables,” p. 313.</td>
</tr>
</tbody>
</table>
**Example of the Chart Platform**

### Additional Roles

<table>
<thead>
<tr>
<th>Role</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grouping</strong></td>
<td>For one or more Grouping variables, independent results are produced for each level or combination of levels of the grouping variables. These results appear in the same report window, but in separate plots.</td>
</tr>
<tr>
<td></td>
<td>See “Use Grouping Variables,” p. 314.</td>
</tr>
<tr>
<td><strong>Weight</strong></td>
<td>Assigns a variable to give the observations different weights.</td>
</tr>
<tr>
<td><strong>Freq</strong></td>
<td>Assigns a frequency variable. This is useful if you have summarized data.</td>
</tr>
<tr>
<td><strong>By</strong></td>
<td>By variables cause plots to be created in separate outline nodes.</td>
</tr>
</tbody>
</table>

### Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Overlay</strong></td>
<td>If this option is selected, all Y variables are plotted in one graph. If this option is not selected, each Y variable is plotted in its own graph. This option is selected by default where available.</td>
</tr>
<tr>
<td></td>
<td>This option is available for all chart types except Pie Chart.</td>
</tr>
<tr>
<td><strong>Chart Orientation</strong></td>
<td>Select Vertical for a vertical chart or Horizontal for a horizontal chart. Vertical is the default option.</td>
</tr>
<tr>
<td></td>
<td>This option is available for all chart types except Pie Chart.</td>
</tr>
<tr>
<td><strong>Chart Type</strong></td>
<td>Select the type of chart that you want. Available chart types are Bar Chart, Line Chart, Pie Chart, Needle Chart, and Point Chart. Selecting a chart controls which of the other options are available. You can always change these options after the chart appears. See “Y Options,” p. 321.</td>
</tr>
<tr>
<td><strong>Show Points</strong></td>
<td>Shows the points in the plot. This option is selected by default where available.</td>
</tr>
<tr>
<td></td>
<td>This option is available for all chart types except Bar Chart and Pie Chart.</td>
</tr>
<tr>
<td><strong>Connect Points</strong></td>
<td>Connects the points in the plot. Show Points does not have to be selected to connect points. This option is selected by default where available.</td>
</tr>
<tr>
<td></td>
<td>This option is available only for Line Chart.</td>
</tr>
<tr>
<td><strong>Add Error Bars to Mean</strong></td>
<td>Adds error bars when the Mean statistic is selected for at least one Y variable and at least one X variable is assigned. This option is not selected by default.</td>
</tr>
<tr>
<td></td>
<td>This option is available only for Line Chart, and additional options are added to the Chart launch window. See “Adding Error Bars,” p. 316.</td>
</tr>
</tbody>
</table>
Plot Statistics for Y Variables

You can plot the raw data for Y variables, or you can plot as many statistics as you want on the y-axis. The Statistics menu in the Chart launch window lists the available statistics. To specify the y-axis, highlight one or more numeric columns in the Select Columns list and select from the list of statistics. If all the statistics requested are counting statistics (for example, N) for the same column, that column is used as the category variable.

The available statistics in the Chart platform are described in the following table. They are the same as those computed by statistical platforms in the Analyze menu and the Summary command in the Tables menu.

Table 10.2 Descriptions of Y Variables Statistics

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>The value of each row in the data table when there is no categorical variable. If there is a categorical variable, <strong>Data</strong> produces a point plot within the variable's levels.</td>
</tr>
<tr>
<td>N</td>
<td>The number of nonmissing values. Also used to compute statistics when there is no column assigned as a weight variable. The Chart platform shows <strong>N</strong> for each level of a categorical variable.</td>
</tr>
<tr>
<td>% of Total</td>
<td>The percentage of the total number of rows represented by each level of the <strong>Categories, X, Levels</strong> variable. If summary statistics are requested on a continuous variable, then the % of Total equals the proportion of the sum represented by each level of the <strong>Categories, X, Levels</strong> variable.</td>
</tr>
<tr>
<td>N Missing</td>
<td>The number of missing values in each level of a categorical variable.</td>
</tr>
<tr>
<td>Min</td>
<td>The least value, excluding missing values, in the level of a categorical variable.</td>
</tr>
<tr>
<td>Max</td>
<td>The greatest value in the level of a categorical variable.</td>
</tr>
<tr>
<td>Sum Wgt</td>
<td>The sum of all values in a column assigned as Weight. Also used instead of <strong>N</strong> to compute other statistics. <strong>Chart</strong> shows the sum of the weight variable for each level of a categorical variable.</td>
</tr>
<tr>
<td>Sum</td>
<td>The sum of all values in each level of a categorical variable.</td>
</tr>
<tr>
<td>Mean</td>
<td>The arithmetic average of a column's values. The mean is the sum of nonmissing values divided by the number of nonmissing values.</td>
</tr>
</tbody>
</table>
Use Categorical Variables

You can assign zero, one, or two X variables whose levels are categories on the x-axis. The Chart platform produces a bar (or a needle, or a pie slice, and so on) for each level or combination of levels of the X variables. If you do not specify any X variable, the chart has a bar for each row in the data table.

The following table shows what type of chart to expect based on the number of X and Y variables.

### Table 10.3 Chart Results by X and Y Variables

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>Type of Chart</th>
</tr>
</thead>
<tbody>
<tr>
<td>none</td>
<td>one or more</td>
<td>If you do not specify a variable for categories, most statistics produce a bar (or a needle, or a pie slice, and so on) for each observation in the data table. This is useful when your data is already summarized. In that case, you usually specify Data as the statistic to plot. Each bar reflects the value of the Y variable. See &quot;Plot a Single Statistic,&quot; p. 322, for an example.</td>
</tr>
<tr>
<td>one or two</td>
<td>none</td>
<td>Plots the counts for each level of the X variable. For two X variables, the counts for each level of both X variables are included (or overlaid) in a single chart. See “Plot Counts of Variable Levels,” p. 323, for an example.</td>
</tr>
</tbody>
</table>
Creating Summary Charts
Example of the Chart Platform

Table 10.3 Chart Results by X and Y Variables (Continued)

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>Type of Chart</th>
</tr>
</thead>
<tbody>
<tr>
<td>one or two</td>
<td>one or more</td>
<td>Plots the selected statistics for each level of the X variable. For two X variables, the selected statistics for each level of the X variables are included (or overlaid) in a single chart. See “Plot Multiple Statistics,” p. 322, and “Plot Multiple Statistics with Two X Variables,” p. 325, for examples.</td>
</tr>
</tbody>
</table>

Use Grouping Variables

If you specify one grouping variable, the result is a separate chart for each level of the grouping variable. All charts are under the same outline title. If you used the same variable as a By variable instead, the same separate charts are produced, but each chart is under its own outline title.

If you specify two or more grouping variables, the result is a matrix of charts. Each chart shows a combination of one level from each of the grouping variables.

If there are multiple statistics, the Overlay option is checked by default, and the Y variables (statistics) are plotted on the same chart for each level of the grouping variable. However, the levels of the grouping variable cannot be overlaid into the same plot frame. For example, if the levels of your grouping variable are Male and Female, the Overlay option cannot be used to combine the two Male and Female graphs into one graph. To see that type of result, use Categories, X, Levels instead of Grouping variables.

Example of Two Grouping Variables

1. Open the Car Poll.jmp sample data table.
2. Select the Chart command from the Graph menu.
3. Select age and click Statistics.
4. Select Mean from the drop-down list of statistics.
5. Select sex and click Categories, X, Levels.
6. Open the Additional Roles outline.
7. Select marital status and country and click Grouping.
8. Click OK.
Example of Two Grouping Variables and Two Category Variables

If you use multiple grouping and category variables, the multiple group labels appear around the borders of the charts, and the multiple X variables cause divisions within charts.

1. Open the Car Poll.jmp sample data table.
2. Select the Chart command from the Graph menu.
3. Select age and click Statistics.
4. Select Mean from the drop-down list of statistics.
5. Select sex and type and click Categories, X, Levels.
6. Open the Additional Roles outline.
7. Select marital status and country and click Grouping.
8. Click OK.
Adding Error Bars

Error bars are available when the **Mean** statistic is selected for at least one Y variable, and at least one X variable is assigned. Error Bars are not available for pie charts. Selecting **Add Error Bars to Mean** causes additional options to appear in the Chart launch window.

After the option is checked, select a type of error bar from the menu that appears. Some of the types of error bar have an additional numeric field. The following table describes the different types of error bars that are available.

**Table 10.4** Types of Error Bars

<table>
<thead>
<tr>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Range</strong></td>
<td>Creates error bars based on the range of the data.</td>
</tr>
<tr>
<td><strong>Standard Error</strong></td>
<td>Creates error bars based on the standard error of the mean. You can specify the number of standard errors.</td>
</tr>
<tr>
<td><strong>Standard Deviation</strong></td>
<td>Creates error bars based on the standard deviation of the data. You can specify the number of standard deviations.</td>
</tr>
</tbody>
</table>
The Chart Report

Follow the instructions in “Example of the Chart Platform,” p. 307 to produce the report shown in Figure 10.8.

Figure 10.8 shows a standard bar chart. Charts can be bar charts, pie charts, line charts, needle charts, point charts, and range charts. For information about additional options for the report, see “Chart Platform Options,” p. 319.

### Legends

Legends are shown as needed. If your chart uses different colors or markers to show levels of one or two X variables, a legend below the chart shows them. If your chart uses different colors or markers to show more than one statistic, a legend to the right of the chart shows them.
Ordering

By default, the Chart platform orders the bars using one of the common orders supported by JMP (months, days of the week, and so on). However, if the grouping column has a **Row Order Levels** column property, the levels are ordered in that order. If the grouping column has a **Value Ordering** column property, it uses that order. If both **Row Order Levels** and **Value Ordering** properties are defined, the **Value Ordering** property has precedence. With neither property in effect, bars are drawn in alphanumeric order.

Coloring Bars in a Chart

There are a few ways to color bars after the chart has been created.

**Manually Set the Color of All Bars**

1. Ensure that no bars are selected.
2. From the red triangle menu for Chart, select **Level Options > Colors**.
3. Select a color from the color palette that appears.

**Set the Color of a Single Bar**

1. Select a bar in the chart
2. From the red triangle menu for Chart, select **Level Options > Colors**.
3. Select a color from the color palette that appears.
Chart Platform Options

The basic Chart report is shown in Figure 10.8 “The Initial Chart Report Window,” p. 317.

The Chart platform has plotting options on the red triangle menu on the Chart title bar. When you select one of these options at the platform level, it affects all plots in the report if no legend levels are highlighted. If one or more plot legend levels are highlighted, the options affect only those levels. There is also a single-plot options menu for each Y variable, which appears when you highlight a Y variable legend beneath the plot and right-click.

The individual plot options are the same as those in the Y Options submenu at the platform level.

General Platform Options

When you select one of these options at the platform level, it affects all plots in the report if no legend levels are highlighted. If one or more plot legend levels are highlighted, the options affect only those plots.

Table 10.5 Descriptions of Chart Platform Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overlay</td>
<td>Displays a single overlaid chart when you have more than one Y (statistics) variable. Each statistic can be assigned a different type of chart (such as line and bar) and overlaid in a single plot. Overlay is selected by default. The axis notation only shows for the last chart displayed if the charts are not overlaid. When Overlay is not selected, the platform shows duplicate axis notation for each chart.</td>
</tr>
<tr>
<td>Vertical, Horizontal</td>
<td>Changes horizontal charts to vertical charts (Vertical), or vertical charts to horizontal charts (Horizontal). Affects all charts in the report. Pie charts are converted to bar chats.</td>
</tr>
<tr>
<td>Pie Chart</td>
<td>Changes a horizontal or vertical chart type to a pie chart.</td>
</tr>
<tr>
<td>Range Chart</td>
<td>Displays a range chart. You can change any chart that includes at least two statistics in a single plot into a range chart. See “Create a Range Chart,” p. 329, for an example of a range chart.</td>
</tr>
<tr>
<td>Option</td>
<td>Description</td>
</tr>
<tr>
<td>------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Add Error Bars to Mean</td>
<td>Adds error bars to charts based on means. A window opens prompting you to select the type of error bar. If error bars already exist on a chart, you can change the error bar type. See “Adding Error Bars,” p. 316 for a description of error bar types.</td>
</tr>
<tr>
<td>Stack Bars</td>
<td>Stacks the bars from levels of a subgroup end-to-end. To use this option, you need two <strong>Categories, X, Levels</strong> variables and a statistic. See “Create a Stacked Bar Chart,” p. 326, for an example of stacking bars.</td>
</tr>
<tr>
<td>Y Options</td>
<td>Contains the options described in “Y Options,” p. 321. To apply these options to a single Y variable, highlight that variable in the legend first.</td>
</tr>
<tr>
<td>Level Options</td>
<td>Selects colors and markers. If no levels (bars, points, or pie slices) are selected, the color or marker that you select is applied to all levels. If you select one or more levels, the color or marker that you select is applied only to the selected levels.</td>
</tr>
<tr>
<td>Label Options</td>
<td>Attaches labels to your plots. In the <strong>Label Options</strong> menu, the first two options (<strong>Show Labels</strong> and <strong>Remove Labels</strong>) turn labels on and off. The last three options (<strong>Label by Value, Label by Percent of Total Values, Label By Row</strong>) specify what label should appear. Only one label can be shown at a time. Label options are also available by right-clicking in the chart.</td>
</tr>
<tr>
<td>Thick Connecting Line</td>
<td>Toggles the connecting line in a line chart to be thick or thin.</td>
</tr>
<tr>
<td>Show Y Legend</td>
<td>Shows the Y legend of the plot. This option is on by default for overlaid charts.</td>
</tr>
<tr>
<td>Show Level Legend</td>
<td>Shows the level legend of the plot. This option is on by default when the <strong>Show Separate Axes</strong> option is selected.</td>
</tr>
<tr>
<td>Show Separate Axes</td>
<td>Duplicates the axis notation for each chart when there are multiple charts. By default, the axis notation only shows for the last chart displayed if the charts are not overlaid. This option is not available for grouped charts.</td>
</tr>
<tr>
<td>Ungroup Charts</td>
<td>Moves level identifiers from the right side of the charts to beneath the charts for individual charts when a grouping variable is specified.</td>
</tr>
<tr>
<td>Script</td>
<td>This menu contains commands that are available to all platforms. They enable you to redo the analysis or save the JSL commands for the analysis to a window or a file. See “Script Menus,” p. 15 in the “Preliminaries” chapter.</td>
</tr>
</tbody>
</table>
Chapter 10

Creating Summary Charts

Examples of Charts

Y Options

The following section describes the Y Options submenu. These commands apply to all Y variables, unless you have a legend level highlighted, then they apply to only the highlighted Y variable.

Click on the legend within a plot to highlight a Y. If you right-click on a highlighted legend level, the commands to modify that Y appear. The commands then affect only the highlighted Y.

Table 10.6 Descriptions of Y Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bar Chart</td>
<td>Displays a bar for each level of the chart variables. The default chart is a bar chart.</td>
</tr>
<tr>
<td>Line Chart</td>
<td>Replaces a bar chart with a line chart and connects each point with a straight line. Select the Show Points option to show or hide the points. Line Chart is also available as a platform option, which then applies to all charts at once.</td>
</tr>
<tr>
<td>Needle Chart</td>
<td>Replaces each bar with a line drawn from the axis to the plotted value. Needle Chart is also available as a platform option, which then applies to all charts at once.</td>
</tr>
<tr>
<td>Point Chart</td>
<td>Shows only the plotted points, without connecting them.</td>
</tr>
<tr>
<td>Show Points</td>
<td>Toggles the point markers on a line or needle chart on or off.</td>
</tr>
<tr>
<td>Connect Points</td>
<td>Toggles the line connecting points on or off.</td>
</tr>
<tr>
<td>Show Error Bars</td>
<td>Toggles the error bars on plots of means. Note that this option is available only for plots that involve means of variables.</td>
</tr>
<tr>
<td>Overlay Color</td>
<td>Assigns a color to statistics (y-axis) to identify them in overlaid charts.</td>
</tr>
<tr>
<td>Overlay Marker</td>
<td>Assigns a marker to statistics, to identify them in overlaid charts.</td>
</tr>
<tr>
<td>Pen Style</td>
<td>Selects a line style for connecting lines.</td>
</tr>
<tr>
<td>Label Format</td>
<td>Specifies the format, field width, and number of decimals for labels. Thousands separators can be turned on or off. Enter the values in the window that appears.</td>
</tr>
</tbody>
</table>

Examples of Charts

Some of the examples that follow show charts that results from different combinations of X variables and Y statistics.

- “Plot a Single Statistic,” p. 322
- “Plot Counts of Variable Levels,” p. 323
- “Plot Multiple Statistics,” p. 322
- “Plot Multiple Statistics with Two X Variables,” p. 325

The other examples show how to produce different types of charts.
Plot a Single Statistic

If you do not specify a variable for categories, most statistics produce a bar for each observation in the data table. This is useful when your data is already summarized. In that case, you usually specify Data as the statistic to plot. Each bar reflects the value of the Y variable.

1. Open the Trial1.jmp sample data table.

This data table contains data from a popcorn experiment. Each row is an experiment, and the yield column is the amount of popped corn that resulted.

2. Select Graph > Chart.

3. Select yield and click Statistics.

4. Select Data from the menu of statistics.

5. Click OK.

The bar chart in Figure 10.10 shows a bar for each experiment (each row) in the data table.

![Figure 10.10 Example of a Chart with One Statistic and No Categories](image)

Plot Multiple Statistics

Chart more than one statistic in a single chart to compare them. If you do not assign an X variable, the chart displays a bar or point for each row in the data table.

1. Open the Financial.jmp sample data table.

2. Select Graph > Chart.
3. Select Type and click Categories, X, Levels.
4. Select Sales($M) and Assets($Mil.) and click Statistics.
5. Select Mean from the menu of statistics.
6. Click OK.

The bar chart in Figure 10.11 compares the mean of sales and assets for each type of company.

**Figure 10.11 Example of a Chart with Two Statistics and One Category**

---

**Plot Counts of Variable Levels**

If you assign one or two X variables without specifying any Y variables, JMP produces bar charts that show counts for each level of the X variables.

**One X Variable**

1. Open the Companies.jmp sample data table.
2. Select Graph > Chart.
3. Select Size Co and click Categories, X, Levels.
4. Click OK.

The bar chart in Figure 10.12 shows a bar for each level of the Size Co variable.
Two X Variables

If you specify two X variables (with no statistics variables), JMP divides the data into groups based on the levels in the two X variables and plots the number of members in each group.

1. Open the Companies.jmp sample data table.
2. Select Graph > Chart.
3. Select Type and click Categories, X, Levels.
4. Select Size Co and click Categories, X, Levels.

The order that you assign these variables is important. The levels of the second variable are nested within the levels of the first variable.

5. Click OK.

The bar chart on the left in Figure 10.13 shows the levels for the size of computer companies and of pharmaceutical companies. The bar chart on the right shows the results if you first selected Size Co and then Type as category variables.
Plot Multiple Statistics with Two X Variables

When you assign two category variables, the result is a chart that shows the statistics for each level of the second category variable. This variable is nested within each level of the first category variable. The chart shows each Y side by side, using a common axis.

1. Open the Companies.jmp sample data table.
2. Select Graph > Chart.
3. Select Type and click Categories, X, Levels.
4. Select Size Co and click Categories, X, Levels.
5. Select Sales ($M) and Profits ($M) and click Statistics.
6. Select Mean from the menu of statistics.
7. Click OK.

The chart on the left in Figure 10.14 shows the result.

8. To see a separate chart for each statistic, click to deselect Overlay from the red triangle menu for Chart.

The chart on the right in Figure 10.14 shows the result.
Create a Stacked Bar Chart

When you have two X levels and a single Y variable, stack the bars by selecting the Stack Bars command from the platform menu.

1. Open the Companies.jmp sample data table.
2. Select Graph > Chart.
3. Select Type and click Categories, X, Levels.
4. Select Size Co and click Categories, X, Levels.
5. Select Sales (SM) and click Statistics.
6. Select Mean from the menu of statistics.
7. Click OK.
8. Select Stack Bars from the red triangle menu for Chart.
Chapter 10
Creating Summary Charts
Examples of Charts

Figure 10.15 Example of a Stacked Bar Chart

Create a Pie Chart

You can create a pie chart either in the Chart launch window or in the report window after you create another type of chart.

Create a Pie Chart in the Chart Launch Window
1. Open the Companies.jmp sample data table.
2. Select Graph > Chart.
3. Select Size Co and click Categories, X, Levels.
4. Select Pie Chart from the menu of chart types.
5. Click OK.

Change a Bar Chart to a Pie Chart
1. Open the Companies.jmp sample data table.
2. Select Graph > Chart.
3. Select Size Co and click Categories, X, Levels.
4. Click OK.
5. Select Pie Chart from the red triangle menu for chart.
   You can also right-click in the chart and select Chart Options > Pie Chart.
Both of the above steps produce the same pie chart shown in Figure 10.16.

**Figure 10.16 Example of a Pie Chart**

The pie chart shows that many more companies are small then are big or medium.

**Show the Values or Percentages for Each Category**

1. Starting with the pie chart in Figure 10.16, select **Label Options > Show Labels** from the red triangle menu for Chart.
   
   You can also right-click in the chart and select **Label > Show Labels**.

   **Label By Value** is the default setting, so the pie chart now shows the number of companies that are big, medium, and small. See the chart on the left in Figure 10.16.

2. To show percentages instead, select **Label Options > Label by Percent of Total Values** from the red triangle menu for Chart.
   
   You can also right-click in the chart and select **Label > Label by Percent of Total Values**.

   The chart on the right shows the percentage of each company size.
Create a Range Chart

A range chart shows the range between two values. In this example, use a range chart to compare the profits and sales of differently sized companies.

1. Open the Companies.jmp sample data table.
2. Select Graph > Chart.
3. Select Size Co and click Categories, X, Levels.
4. Select Profits ($M) and Sales ($M) and click Statistics.
5. Select Mean from the menu of statistics.
6. Click OK.
7. Select Range Chart from the red triangle menu for Chart.
The chart shows a much larger difference between profits and sales than that for medium and small companies.

Create a Chart with Ranges and Lines for Statistics

The Stock Prices.jmp sample data table contains data for the dates and values of a stock over time. The variable YearWeek is a computed column representing the year and week in a single variable. Use a range chart to show the high, low, and average close values for each stock. For those weeks where data exists for multiple days, the average of the values is plotted.

1. Open the Stock Prices.jmp sample data table.
2. Select Graph > Chart.
3. Select YearWeek and click Categories, X, Levels.
4. Select High and click Statistics.
5. Select Max from the menu of statistics.
6. Select Low and click Statistics.
7. Select Min from the menu of statistics.
8. Select Close and click Statistics.
9. Select Mean from the menu of statistics.
10. Click OK.
11. Select Range Chart from the red triangle menu for Chart.
12. In the legend, right-click Mean(Close) and select Connect Points.
Figure 10.19 Example of a Combined Range and Line Chart

The range for each date shows the highest and lowest values the stock reached during that week. The line shows the stock's average closing price for that week.
The **Overlay Plot** command in the **Graph** menu produces plots of a single X column and one or more numeric Ys and does not accept non-numeric values for the y-axis. Curves can also be shown as separate plots for each Y with a common x-axis. Plots can be modified with range and needle options, color, log axes, and grid lines. Curves with two different scales can be overlaid on the same plot with the addition of a right axis.

**Figure 11.1 Examples of Overlay Plot Graphs**
Contents

Example of an Overlay Plot ................................................................. 335
Launch the Overlay Plot Platform ...................................................... 336
The Overlay Plot Report ................................................................. 337
Overlay Plot Options ................................................................. 338
  General Platform Options ...................................................... 338
  Y Options .............................................................................. 341
Additional Examples of Overlay Plots ............................................. 342
  Function Plots ........................................................................ 342
  Plotting Two or More Variables with a Second Y-axis ................ 343
  Grouping Variables .................................................................. 345
Example of an Overlay Plot

This example shows you how to plot two variables on a single $y$-axis.

1. Open the Spring.jmp sample data table.
   The table has a row for each day in the month of April. The column named April is the numeric day of the month, and the remaining columns are various weather statistics.

2. Select the Overlay Plot command from the Graph menu.

3. Select Humid1:PM and Humid4:PM and click Y.
   These two columns are the humidity measurements at 1:00 p.m. and 4:00 p.m.

4. Select April and click X.

5. Click OK.
   The plot shown in Figure 11.2 appears. Initially, this platform overlays all specified Y columns. The legend below the plot shows individual markers and colors that identify each Y column.

To help you quickly differentiate between the Ys, select Y Options > Connect Points from the Overlay Plot red triangle menu. Adjacent points are connected for each Y variable, as shown in Figure 11.3.
Launch the Overlay Plot Platform

Launch Overlay Plot by selecting **Graph > Overlay Plot**.

The Overlay Plot Launch window contains many controls that are described in “Launch Window Features,” p. 13 in the “Preliminaries” chapter. Controls that are unique to this window are described in Table 11.2.
In the Overlay Plot Launch window, you assign the following:

- one X variable of any modeling type
- as many numeric Y variables as you want

**Table 11.1 Description of the Overlay Plot Launch Window**

<table>
<thead>
<tr>
<th>Cast Selected Columns Into Roles</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>X, Y</td>
<td>You can graph many numeric Y variables against a single X variable.</td>
</tr>
<tr>
<td>Left Scale/Right Scale</td>
<td>The columns assigned to the Y role have a left- or right-pointing arrow to the left of the column name. This arrow designates on which vertical axis (on the left or right of the plot) the variable appears. Change the designation by highlighting the column in the Y list and clicking the <strong>Left Scale/Right Scale</strong> button.</td>
</tr>
<tr>
<td>Grouping</td>
<td>This option produces a matrix of graphs for each Grouping variable.</td>
</tr>
<tr>
<td>By</td>
<td>This option produces a separate graph for each level of the By variable. If two By variables are assigned, a separate graph for each possible combination of the levels of both By variables is produced.</td>
</tr>
</tbody>
</table>

**Options**

| Sort X                           | This option causes the points to be connected in order of ascending X values. Otherwise, the points are connected in row order. This option is selected by default. |
| X Log Scale                      | This option applies a log scale to the x-axis. |
| Left Y Log Scale                 | This option applies a log scale to the left y-axis. It is available only if one or more Y variables are left-scaled. (See **Left Scale/Right Scale** in this table.) |
| Right Y Log Scale                | This option applies a log scale to the right y-axis. It is available only if one or more Y variables are right-scaled. (See **Left Scale/Right Scale** in this table.) |

---

**The Overlay Plot Report**

Follow the instructions in “Example of an Overlay Plot,” p. 335 to produce the plot shown in Figure 11.5. Initially, this platform overlays all specified Y columns. The legend below the plot shows individual markers and colors that identify each Y column. For information about additional options for the report, see “Overlay Plot Options,” p. 338.
Creating Overlay Plots

Overlay Plot Options

Figure 11.5 The Overlay Plot Report

Overlay Plot Options

The basic Overlay Plot report is shown in Figure 11.5.

The Overlay Plot platform has plotting options accessed from the red triangle menu on the Overlay Plot title bar. When you select one of these options at the platform level, it affects all plots in the report if no legend levels are highlighted. If one or more plot legend levels are highlighted, the options affect only those plots. There is also a single-plot options menu for each Y variable, which appears when you highlight a Y variable legend beneath the plot and right-click.

The individual plot options are the same as those in the Y Options submenu at the platform level.

General Platform Options

When you select one of these options at the platform level, it affects all plots in the report if no legend levels are highlighted. If one or more plot legend levels are highlighted, the options affect only those plots.
Table 11.2 Descriptions of Overlay Plot Platform Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overlay Plots</td>
<td>Contains options for overlaying:</td>
</tr>
<tr>
<td>Overlay Y's</td>
<td>Overlays all variables assigned to the Y role on one plot. This option is on</td>
</tr>
<tr>
<td></td>
<td>by default and unavailable if only one Y variable is assigned.</td>
</tr>
<tr>
<td>Overlay Groups</td>
<td>Overlays groups and produces a legend. This option is off by default and</td>
</tr>
<tr>
<td></td>
<td>unavailable if no grouping variables are assigned. See “Overlay Groups,” p.339.</td>
</tr>
<tr>
<td>No Overlay</td>
<td>Turns off overlaying for both Ys and groups. Creates a separate plot for</td>
</tr>
<tr>
<td></td>
<td>each Y and each group. This option is off by default unless only one Y</td>
</tr>
<tr>
<td></td>
<td>variable is assigned and no grouping variables are assigned. In this case,</td>
</tr>
<tr>
<td></td>
<td>no overlaying options are available.</td>
</tr>
<tr>
<td>Separate Axes</td>
<td>Assigns each plot its own set of xy-axes. If Separate Axes is off, the</td>
</tr>
<tr>
<td></td>
<td>vertical axis is shared across the same row of plots and the horizontal</td>
</tr>
<tr>
<td></td>
<td>axis is shared on the same column of plots. The default setting is off,</td>
</tr>
<tr>
<td>Uniform Y Scale</td>
<td>Uses the same Y scale for all grouped plots. The default setting is off.</td>
</tr>
<tr>
<td>Connect Thru Missing</td>
<td>Connects adjacent points in the plot, regardless of missing values. The</td>
</tr>
<tr>
<td></td>
<td>default setting is off.</td>
</tr>
<tr>
<td>Range Plot</td>
<td>Connects the lowest and highest points at each x value with a line with</td>
</tr>
<tr>
<td></td>
<td>bars at each end. The Needle and Range Plot options are mutually exclusive.</td>
</tr>
<tr>
<td>Y Options</td>
<td>Contains options for the Y variables. See “Y Options,” p. 341.</td>
</tr>
<tr>
<td>Ungroup Plots</td>
<td>Creates a separate chart for each level of a grouping variable.</td>
</tr>
<tr>
<td>Arrange Plots</td>
<td>Enables you to specify the number of plots in each row.</td>
</tr>
<tr>
<td>Script</td>
<td>This menu contains options that are available to all platforms. They enable</td>
</tr>
<tr>
<td></td>
<td>you to redo the analysis or save the JSL commands for the analysis to a</td>
</tr>
<tr>
<td></td>
<td>window or a file. See “Script Menus,” p. 15 in the “Preliminaries” chapter.</td>
</tr>
</tbody>
</table>

**Overlay Groups**

Figure 11.6 shows the effect that the **Overlay Groups** option has on an overlay plot with one Y variable, one X variable, and a grouping variable. The grouping variable has two levels. The plot on the left has **Overlay Groups** turned off, so a separate graph is produced for the two levels of the grouping variable. The plot on the right has **Overlay Groups** turned on, so there is a single graph that uses colors and markers to show the two levels of the grouping variable. A legend describing the levels is added under the graph.
Creating Overlay Plots

Chapter 11

Overlay Plot Options

Figure 11.6 Overlay Groups: Off (left) and On (right)

Separate Axes

Figure 11.7 shows the effect that the **Separate Axes** option has on an overlay plot with two Y variables and one X variable. The **Overlay Y’s** option is turned off, so a separate plot is produced for each Y variable. The plot on the left has **Separate Axes** turned off, so the two graphs share a single x-axis. The plot on the right has **Separate Axes** turned on, so both graphs have their own x-axis.
Creating Overlay Plots

Overlay Plot Options

Figure 11.7 Separate Axes: Off (left) and On (right)

Y Options

Each Y variable is labeled in a legend beneath the plot. The Y options are available from the Y Options menu from the red triangle menu for Overlay Plot. You can also access the Y Options menu by right-clicking on any Y variable in the legend.

Note: If no Y variables are selected, any Y options that you select affect all Y variables. If one or more of the Y variables are selected, any Y options that you select affect only those you have selected.

Selecting and Deselecting Y Variables in the Legend

Hold the SHIFT key and click to select multiple contiguous legend levels.

Hold the CONTROL key and click to select multiple discontiguous legend levels.

Hold the CONTROL key and click a selected legend level to deselect it.
Creating Overlay Plots

Chapter 11

Additional Examples of Overlay Plots

The following sections show several examples of different overlay plots.

Function Plots

Overlay Plot normally assumes you want a function plot when the Y column contains a formula. However, formulas that contain random number functions are more frequently used with simulations, where function plotting is not often wanted. Therefore, the Function Plot option is off by default when a random number function is present, but on for all other functions.

To see an example of a function plot

1. Open the sample data file Density Compare.jmp.
2. From the Graph menu, select Overlay Plot.
3. Assign gamma1, gamma3, and gamma5 as the Y variables.

Table 11.3 Descriptions of Y Options

<table>
<thead>
<tr>
<th>Description</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show Points</td>
<td>A toggle that either shows or hides points in the graph.</td>
</tr>
<tr>
<td>Connect Points</td>
<td>A toggle that either connects the points with lines or turns the connecting lines off. You can use Connect Points without showing points.</td>
</tr>
<tr>
<td>Needle</td>
<td>Draws a vertical line from each point to the x-axis.</td>
</tr>
<tr>
<td>Step</td>
<td>Draws a horizontal line from each point to the x value of the following point, and then a vertical line to that point. You can use Step without showing points.</td>
</tr>
</tbody>
</table>

Note: The Connect Points, Needle, and Step options are mutually exclusive.

Function Plot

Plots a formula (stored in the Y column) as a smooth curve. To use this function, store a formula in a column that is a function of a single X column. Assign the formula to the Y role. For an example, see “Function Plots,” p. 342.

Connect Color

Displays the JMP color palette for assigning colors to connecting lines.

Overlay Marker

Displays the JMP marker palette for assigning markers to plotted points.

Overlay Marker Color

Assigns a color to all points of the selected variable.

Line Style

Enables the choice of dashed, dotted, or other line styles.

Line Width

Enables the choice of line widths.
4. Assign X\text{gamma} as the X variable.
5. Click OK.
6. Turn off the Show Points option by selecting Y Options > Show Points from the red triangle menu.

**Figure 11.8 Function Plot**

---

**Plotting Two or More Variables with a Second Y-axis**

A second y-axis is useful for plotting data with different scales on the same plot, such as a stock's closing price and its daily volume, or temperature and pressure. For example, consider plotting the selling price of an inexpensive stock against the Dow Jones Industrial Average.

1. Open the Stock Prices.jmp sample data table.
2. From the Graph menu, select Overlay Plot.
3. Assign High, Low, Close, and Volume as Y variables.
4. Select Volume in the Y list and click Left Scale/Right Scale.
   This action assigns Volume to the right axis, leaving the others on the left axis. The arrows to the left of the Y variables show you which axis each variable is assigned.
5. Assign Date as the X variable.
6. Click OK.
7. From the red triangle menu for Overlay Plot, select Y Options > Connect Points.
Creating Overlay Plots
Additional Examples of Overlay Plots

Figure 11.9 Dual Axis Overlay Plot

The variables High, Low, and Close are the stock prices of the same stock and thus are on the same scale. Volume is a different scale entirely, representing the trading volume of the entire Dow Jones Industrial Average.

To see why this matters, perform the same steps above without clicking the Left Scale/Right Scale button for Volume. Compare the resulting graph in Figure 11.10 to Figure 11.9.

Figure 11.10 Single Axis Overlay Plot
Grouping Variables

The Overlay Plot platform allows the production of several plots in one window through the use of grouping variables. With one grouping variable, a stacked vector of plots appears, with one plot for each level of the grouping variable. Two grouping variables result in a matrix of plots.

1. Open the students.jmp sample data table.
2. From the Graph menu, select Overlay Plot.
3. Assign weight as the Y variable and height as the X variable.
4. Assign age and sex as grouping variables.
5. Click OK.

A portion of the resulting plot is shown in Figure 11.11.

Figure 11.11 Grouped Plots Without Separate Axes

Select the Separate Axes option from the red triangle menu to produce plots that do not share axes. Compare Figure 11.12 to Figure 11.12.
Figure 11.12  Grouping Variables
Chapter 12

Creating Three-Dimensional Scatterplots
Using the Scatterplot 3D Platform

The Scatterplot 3D platform shows the values of numeric columns in the associated data table in a rotatable, three-dimensional view. Up to three columns that you select from the associated data table are displayed at one time. See Figure 12.1.

To help visualize variation in higher dimensions, the 3D scatterplot can show a biplot representation of the points and variables when you request principal components. The most prominent directions of data are displayed on the 3D scatterplot report.

**Figure 12.1 Example of a 3D Scatterplot**
Example of a 3D Scatterplot

This example uses the Iris.jmp sample data table, which includes measurements of sepal length, sepal width, petal length, and petal width for three species of iris.

1. Open the Iris.jmp sample data table.
2. Select Graph > Scatterplot 3D.
4. Click OK.

Now you can spin the 3D scatterplot to see the relationships between the variables. In this example, the data points are formatted in blue, red, and green. You might want to spin the scatterplot to see more clearly the relationships between the red and green points.
Launch the Scatterplot 3D Platform

Launch the Scatterplot 3D platform by selecting Graph > Scatterplot 3D.

**Figure 12.3** The Scatterplot 3D Launch Window

**Table 12.1** Description of the Scatterplot 3D Launch Window

| Y, Columns | Select the variables to plot on the 3D scatterplot. The order in which you select the variables determines where the data points appear on the axes:  
|            | • The first variable appears on the x axis.  
|            | • The second variable appears on the y axis.  
|            | • The third variable appears on the z axis.  
|            | You can assign the remaining variables interactively through the drop-down menus below the scatterplot. |
| Weight     | Identifies the data table column whose variables assign weight (such as importance or influence) to the data. When you specify a Weight variable, JMP draws the points as balls. The balls are scaled so that their volume represents the weight value. You click and drag the Circle Size slider below the scatterplot to resize the balls. |
| Freq       | Identifies the data table column whose variables assign a frequency to each row. This option is useful when a frequency is assigned to each row in summarized data. |
| By         | Produces a separate 3D scatterplot for each By variable value. When two By variables are assigned, a separate graph is produced for each combination of both By variables. |
The Scatterplot 3D Report

To produce the 3D scatterplot shown in Figure 12.4, follow the instructions in “Example of a 3D Scatterplot,” p. 349.

The Scatterplot 3D report shows a three-dimensional spinnable view of your data. See Figure 12.4. In the launch window, you select two or three variables and then create the report. The variables are displayed on the 3D scatterplot’s x, y, and z axes.

Figure 12.4 Example of Information Displayed on the Scatterplot 3D Report

After you create a 3D scatterplot, you can add features such as displaying ellipses around specific data points, showing separate principal components, rotating components, connecting points, and more. See “Scatterplot 3D Platform Options,” p. 355 for details.
Creating Three-Dimensional Scatterplots

The Scatterplot 3D Report

You can also assign colors and symbols (or markers) to data points either on the 3D scatterplot itself or in the associated data table. See “Assign Colors and Markers to Data Points,” p. 354 and “Assign Colors and Markers in the Data Table,” p. 354.

Spin the 3D Scatterplot

You spin the 3D scatterplot report in four ways:

- Click and drag an empty area on the 3D scatterplot. The 3D scatterplot spins in the direction you dragged the mouse.

  **Note:** Click and drag on an empty area on the 3D scatterplot, not on an axis or data point. Dragging the axis rescales the axis. Dragging a data point only selects the point.

- Slide the mouse wheel. The 3D scatterplot spins up and down only.
- Hold down an arrow key. (Before using an arrow on the number keypad, verify that NUM LOCK is turned off.)
- Hold down ESC. The 3D scatterplot spins left and right only.

In each case, the 3D scatterplot spins as long as you hold down the mouse button, arrow key, or ESC key. The spinning also continues as you slide the mouse wheel.

You can also spin the 3D scatterplot continuously as follows:

- Click and drag: Hold down SHIFT, click and drag an empty area on the plot, and then release SHIFT. The faster you drag the mouse, the faster the 3D scatterplot spins.
- Mouse wheel: Hold down SHIFT, slide the wheel, and then release the wheel. The 3D scatterplot spins up and down only.
- Arrow keys: Hold down SHIFT, press the arrow key, then release SHIFT.
- ESC key: Hold down SHIFT and press ESC. The 3D scatterplot spins up and down only.

In addition to automatically spinning the plot, you can oscillate the plot. Hold down SHIFT and CTRL and then click and drag the plot. The plot shakes up and down or left to right, depending on the direction in which you dragged the plot.

To stop the spinning or oscillating, click on the plot or press ESC.

Change Variables on the Axes

The variables on each axis are determined by the order in which you select the variables in the launch window. For example, the first variable that you select is displayed on the x axis. The second variable is displayed on the y axis.

After you create a 3D scatterplot, you can change the variable assigned to an axis, plot a different set of variables, or sequence through all combinations of the variables.

1. To change the variable on a specific axis, select the axis control drop-down menu and select a different variable.
2. To add a third variable, select the blank axis control drop-down menu, select **Other**, select the variable, and then click **OK**.

3. To sequence through combinations of all variables, click the **Next Axis Set** button until the variables that you want to plot are displayed.

### Adjust the Axes

You can manually move or rescale the axis coordinates by clicking and dragging the axis. This option shows a different set of coordinates on the 3D scatterplot. It also lets you change the space displayed between the coordinates (or the coordinate scaling).

You can also specify axis properties by double-clicking the axis and modifying settings in the specifications window.

**To move the coordinates on the axis**

1. Place your cursor over the middle of the axis.
2. Click and drag the axis.

**To modify coordinate scaling**

1. Place your cursor over the end of the axis.
2. Click and drag the axis.

**To rescale an axis precisely**

1. Place your cursor over the middle of the axis (the axis, not the label).
2. Double-click the axis.

---

**Table 12.2** Descriptions of Axis Options

<table>
<thead>
<tr>
<th>Scale</th>
<th>Changes the scale of the axes. See the <em>Using JMP</em> book.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Format</td>
<td>Specifies how to display numbers on the 3D scatterplot. See the <em>Using JMP</em> book.</td>
</tr>
</tbody>
</table>
Assign Colors and Markers to Data Points

Each point in the 3D scatterplot corresponds to a row in the associated data table. To highlight points on the 3D scatterplot, you assign colors and markers to the points. The colors and markers are then displayed on the 3D scatterplot and in the data table.

When you click a point, the following items are selected:

- the point in the 3D scatterplot
- the corresponding row in the associated data table
- the point in any other opened 3D scatterplots, if applicable

To select one point, click the point.

To select several points, double-click the 3D scatterplot and drag the cursor over the points. A box is displayed to indicate which points are selected.

To deselect points, double-click the 3D scatterplot.

To assign a color or marker to selected data points, proceed as follows:

1. To assign a color to the selected point, select Rows > Colors and then select the color.
2. To assign a marker to the selected point, select Rows > Markers and then select the marker.

Assign Colors and Markers in the Data Table

You can assign colors and markers to rows in the data table. The colors and markers appear next to the row number in the data table and on the 3D scatterplot. This option distinguishes points for each variable, and you can save the settings in the data table; assigning colors and markers to specific data points (as described in “Assign Colors and Markers to Data Points,” p. 354) only highlights them for the current scatterplot.

See Using JMP for details about assigning colors and markers in the data table. For details about changing the size, quality, or transparency of markers, see “Scatterplot 3D Settings,” p. 365.
## Scatterplot 3D Platform Options

The red triangle menu next to Scatterplot 3D contains options to customize the display and to compute, rotate, and save principal or rotated components.

### Table 12.3 Description of Scatterplot 3D Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Show Points</strong></td>
<td>Shows or hides the data points on the graph.</td>
</tr>
<tr>
<td><strong>Show Controls</strong></td>
<td>Shows or hides the source and axis controls displayed beneath the 3D scatterplot. See Figure 12.4.</td>
</tr>
<tr>
<td><strong>Drop Lines</strong></td>
<td>Draws lines from each point to the plane created by the ( x ) and ( z ) variables that you selected on the launch window.</td>
</tr>
<tr>
<td><strong>Connect Points</strong></td>
<td>Connects the points with a line. Points can be connected on the data as a whole or in groups. You can also group data by a specific variable.</td>
</tr>
<tr>
<td><strong>Normal Contour Ellipsoids</strong></td>
<td>Draws one or more normal contour ellipsoids, that is, three-dimensional ellipses that encompass a specified portion of points. You specify whether you want an ellipsoid for all of the data or for each group. You can also control the size and transparency of the ellipsoids. For details, see &quot;Normal Contour Ellipsoids,&quot; p. 357.</td>
</tr>
<tr>
<td><strong>Ellipsoid Coverage</strong></td>
<td>Changes the size of normal contour ellipsoids. Type a value between 0 and 1, where the greater the value creates a bigger the ellipsoid. The actual values “0” and “1” produce no ellipsoid, so you will see a warning if you try to use those values. This option only appears after you add a normal contour ellipsoid to the 3D scatterplot.</td>
</tr>
<tr>
<td><strong>Ellipsoid Transparency</strong></td>
<td>Changes the surface of normal contour ellipsoids. The greater the value, the more opaque the ellipsoid. This option only appears after you add a normal contour ellipsoid to the 3D scatterplot.</td>
</tr>
<tr>
<td><strong>Nonpar Density Contour</strong></td>
<td>Draws nonparametric density contours, which approximately encompass a specified proportion of the points. You specify whether you want a density contour for all of the data or for each group. For details, see &quot;Nonparametric Density Contours,&quot; p. 359.</td>
</tr>
<tr>
<td><strong>Drop Line Thickness</strong></td>
<td>Changes the width of drop lines. This option only appears after you add drop lines to the 3D scatterplot.</td>
</tr>
<tr>
<td><strong>Principal Components</strong></td>
<td>Calculates principal components on all variables. This changes the axes of the plot to have principal component scores. Biplot rays are displayed by default. You can remove them by selecting <strong>Biplot Rays</strong> from the red triangle menu. For details about principal components, see the <em>Modeling and Multivariate Methods</em> book.</td>
</tr>
</tbody>
</table>
### Table 12.3 Description of Scatterplot 3D Options (Continued)

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Std Prin Components</strong></td>
<td>Calculates principal components (as with the Principal Components option) but scales the principal component scores to have unit variance. If this option is not selected, the scores have variance equal to the corresponding eigenvalue. With standardized principal components, the correlation between the variables and the principal component scores is equal to the values in the eigenvector. This helps you quickly assess the relative importance of the variables. For details, see the <em>Modeling and Multivariate Methods</em> book.</td>
</tr>
<tr>
<td></td>
<td><strong>Note:</strong> Select this option if you want GH' rather than JK' biplots. GH' biplots try to preserve relationships between variables; JK' biplots try to preserve relationships between observations. The interpoint distance shown by GH' biplots is less meaningful, but the angles of the GH' biplot rays measure correlations better.</td>
</tr>
<tr>
<td><strong>Rotated Components</strong></td>
<td>Specifies the number of factors that you want to rotate and the rotation method. You rotate components to better align the directions of the factors with the original variables so that the factors might be more interpretable. For details, see the <em>Modeling and Multivariate Methods</em> book.</td>
</tr>
<tr>
<td><strong>Biplot Rays</strong></td>
<td>Shows or hides biplot rays that correspond to the principal components. You must have already selected Principal Components, Std Prin Components, or Rotated Components for this option to appear.</td>
</tr>
<tr>
<td><strong>Remove Prin Comp</strong></td>
<td>Removes principal components, standardized principal components, and rotated components from the scatterplot 3D report. The 3D scatterplot reverts to its original display before principal components were selected. This option, however, does not remove any saved principal components from the data table. This option only appears after you add principal, standard, or rotated components to the 3D scatterplot.</td>
</tr>
<tr>
<td><strong>Save Prin Components</strong></td>
<td>Saves the specified number of current principal component scores as new columns in the current data table. These columns also include the formulas used for the principal components. For n variables in the components list, n principal component columns are created and named Prin1, Prin2, ... Prin_n. This option only appears after you add principal, standard, or rotated components to the 3D scatterplot.</td>
</tr>
</tbody>
</table>
Scatterplot 3D Platform Options

Normal Contour Ellipsoids

A normal contour ellipsoid is a 3-dimensional ellipse that encompasses a specified portion of points. The ellipsoid is computed from a contour of the multivariate normal distribution fit to the points. The ellipsoid is a function of the means, standard deviations, and correlations of variables on the plot. See the *Modeling and Multivariate Methods* book for details about multivariate normal distributions.

When you add an ellipsoid, two formatting options are available:

- **Coverage** changes the portion of data points covered by the ellipsoid. The larger the value, the bigger the ellipsoid.
- **Transparency** changes the surface of the ellipsoid from transparent to opaque. The larger the value, the more opaque the ellipsoid.

The coverage and transparency options also appear in the red triangle menu after you add the ellipsoid.

When you add normal contour ellipsoids to a 3D scatterplot, you specify whether you want an ellipsoid for all of the data or for a specific group of data. The ellipsoid for each set of grouped data is color-coded to differentiate one group from another.

You display and remove normal contour ellipsoids by selecting and deselecting **Normal Contour Ellipsoids** from the red triangle menu.

The examples in this section use the *Iris.jmp* sample data table, which includes measurements of sepal length, sepal width, petal length, and petal width for three species of iris.

**Example of an Ungrouped Normal Contour Ellipsoid**

This example shows how to add a normal contour ellipsoid over 75% of the data points. The ellipsoid is 25% transparent.

1. Open the *Iris.jmp* sample data table.
2. Select **Graph > Scatterplot 3D**.
3. Select **Sepal length, Sepal width, Petal length, and Petal width** and click **Y, Columns**.
4. Click OK.
5. From the red triangle menu, select **Normal Contour Ellipsoids**. Notice that **Ungrouped** is already selected.
6. Type .75 next to Coverage.
7. Type .25 next to Transparency.
8. Click OK.

**Figure 12.6** Example of an Ungrouped Normal Contour Ellipsoid

---

**Example of Grouped Normal Contour Ellipsoids**

This example shows how to group measurements by species and to format each group with a normal contour ellipsoid. The ellipsoids cover 75% of the data points and are 50% transparent. The contours are color-coded based on species.

1. Open the iris.jmp sample data table.
2. Select **Graph > Scatterplot 3D**.
3. Select **Sepal length, Sepal width, Petal length, and Petal width** and click **Y, Columns**.
4. Click **OK**.

5. From the red triangle menu, select **Normal Contour Ellipsoids**.

6. Select **Grouped by Column**.

7. Select **Species**.

8. Type .75 next to **Coverage**.

9. Type .5 next to **Transparency**.

10. Click **OK**.

---

**Figure 12.7** Example of Grouped Normal Contour Ellipsoids

---

**Nonparametric Density Contours**

The nonparametric density contour shows contours that approximately encompass a specified proportion of the points. You add nonparametric density contours to see patterns in point density when the scatterplot is darkened by thousands of points.

This feature is particularly valuable when you have many points on a 3D scatterplot; the contours can be so dark that you cannot see the structure. In this situation, you remove the points so that only the contours are displayed. See “**Example of Optimizing a Dense Nonparametric Density Contour**,” p. 364 for details.
When you add nonparametric density contours to a 3D scatterplot, you specify whether you want a contour for all of the data or for a specific group of data. The contour for each set of grouped data is color-coded to differentiate one group from another.

You display and remove nonparametric density contours by selecting and deselecting Nonpar Density Contours from the red triangle menu.

The examples in this section use the Iris.jmp sample data table, which includes measurements of sepal length, sepal width, petal length, and petal width for three species of iris.

**Example of an Ungrouped Nonparametric Density Contour**

This example shows how to add a nonparametric density contour over all data points.

1. Open the Iris.jmp sample data table.
2. Select Graph > Scatterplot 3D.
4. Click OK.
5. From the red triangle menu, select Nonpar Density Contour. Notice that Ungrouped is already selected.
6. Click OK.
The Density Contour Controls options are displayed below the 3D scatterplot. These options let you select additional contours and change each contour’s formatting.
Creating Three-Dimensional Scatterplots

Scatterplot 3D Platform Options

By default, one contour is selected and formatted in blue with a 10% density and 25% transparency. This means that about 10% of the data is outside the contour surface and 90% is inside. You select the second and third contour levels (and modify settings, if desired) to display them on the 3D scatterplot.

Table 12.4 Description of the Density Contour Controls Window

| Contour Quantile | Controls which contours are shown and lets you customize the contour formatting. 
|---|---|
| Density Level | Represents the volume and density of the points. As the contours go from smaller to larger values, the contours cover less volume but more dense areas. A .9 contour represents the 10% densest part of the total, where the points are closest together. Click and drag the slider below “Contour Quantile,” or enter a value next to the slider. 
| Transparency | Changes the surface of density contours. The greater the value, the more opaque the contour. Enter a value in the box. 
| Color | Changes the color of the contour. Click the colored box and select a different color. (This option only appears for ungrouped density contours.) 

Changes to these settings take affect immediately.

| Resolution | Changes the resolution of the contours. A higher resolution results in a less granular drawing of the contours but takes more time to display. 

| Column Bandwidth | Changes the smoothness of the fitted density. A higher bandwidth results in a smoother fitted density. Type a new bandwidth for each variable, or click and drag the sliders. Click Apply to display your changes.

Example of a Grouped Nonparametric Density Contour with Modified Transparency and Density Levels

This example shows how to group data points and format each nonparametric density contour.

1. Open the iris jmp sample data table.
2. Select Graph > Scatterplot 3D.
4. Click OK.
5. From the right triangle menu, select Nonpar Density Contour.
7. Select Species and click OK. A different colored contour is displayed for each of the three species.
8. Type .25 in the first Color Quantile box. 25% of the data points appear outside the contour surfaces, which results in smaller contours.
9. Type .5 in the first Transparency box. The contours are 50% opaque.
10. Select the second check box. A second set of contours appears to further illustrate the density of the data points.

**Figure 12.11** Adding a Second Nonparametric Density Contour

**Figure 12.10** Changing the Nonparametric Density Contour Transparency and Density
Creating Three-Dimensional Scatterplots

Scatterplot 3D Platform Options

You can now format the second levels of contours and turn on the third level of contours.

The options for formatting the grouped and ungrouped nonparametric density contours are similar. The only difference is that you cannot change the color of each grouped nonparametric density contour. See Table 12.4 for options.

**Example of Optimizing a Dense Nonparametric Density Contour**

When you have many points on a 3D scatterplot, the contours can be so dark that you cannot see the structure. In this situation, you remove the points so that only the contours are displayed.

To remove points from a 3D scatterplot, select **Show Points** from the red triangle menu. You can further optimize the contours by changing their size, color, and transparency. See Table 12.4 for details.

---

**Figure 12.12** Example of Optimizing a Dense Nonparametric Density Contour

---

**Context Menu**

Right-click the 3D Scatterplot to see the context menu.

**Table 12.5** Descriptions of Context Menu Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reset</td>
<td>Returns the orientation of the scatterplot to its original state.</td>
</tr>
<tr>
<td><strong>Settings</strong></td>
<td>Provides options to change the appearance of the 3D scatterplot. See “Scatterplot 3D Settings,” p. 365.</td>
</tr>
<tr>
<td>Hide Lights Border</td>
<td>Shows and hides a border that displays the lights. The lights highlight different portions of the 3D scatterplot. Right-click a light to turn it on or off and to change the color.</td>
</tr>
<tr>
<td>Wall Color</td>
<td>Changes the color of the 3D scatterplot.</td>
</tr>
</tbody>
</table>
Chapter 12  
Creating Three-Dimensional Scatterplots  
Scatterplot 3D Platform Options

365

Scatterplot 3D Platform Options

Scatterplot 3D Settings

To customize properties such as the marker size, text size, and grid lines, right-click the 3D scatterplot and select Settings. The Settings window appears. As you modify the settings, a preview appears on the 3D scatterplot.

Figure 12.13 The Scatterplot 3D Settings Window

Note the following:
- Move the sliders left to decrease the selected property or to the right to increase the selected property.
- To move the Settings window around the scatterplot, click and drag the top portion of the window.

Table 12.5 Descriptions of Context Menu Options (Continued)

<table>
<thead>
<tr>
<th>Background Color</th>
<th>Changes the color surrounding the 3D scatterplot.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rows</td>
<td>You can color, mark, exclude, hide, and label points that correspond to rows in the associated data table. You must select the points before selecting this option. See the Using JMP book.</td>
</tr>
<tr>
<td>Use Hardware Acceleration</td>
<td>Turns hardware acceleration on or off for machines that support acceleration. This option might display the scatterplot faster. If not, try updating your graphics drivers.</td>
</tr>
<tr>
<td>Show ArcBall</td>
<td>Shows and hides a globe around the 3D scatterplot. This option helps you visualize the rotation of the scatterplot. Select whether you want the ArcBall to appear always, only when you drag the scatterplot, or never.</td>
</tr>
</tbody>
</table>

Table 12.6 Descriptions of Scatterplot 3D Settings Window Options

<table>
<thead>
<tr>
<th>Reset</th>
<th>Done</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resets the default settings.</td>
<td>Closes the window.</td>
</tr>
</tbody>
</table>
### Table 12.6 Descriptions of Scatterplot 3D Settings Window Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Walls</td>
<td>Adds or removes the 3D scatterplot walls. Without walls, the background color of the 3D scatterplot is displayed.</td>
</tr>
<tr>
<td>Grids</td>
<td>Shows or hides the coordinate lines.</td>
</tr>
<tr>
<td>Axes</td>
<td>Shows or hides the variable names that appear above each axis.</td>
</tr>
<tr>
<td>Box</td>
<td>Shows or hides the box. Without the box, the 3D scatterplot is displayed as an open plot.</td>
</tr>
<tr>
<td>Zoom</td>
<td>Enlarges or shrinks the 3D scatterplot.</td>
</tr>
</tbody>
</table>
| Orthographic | Changes the view of the scatterplot from 3-dimensional to an orthographic projection. In the orthographic view, the walls of the scatterplot do not converge to a vanishing point. This means that you can compare near and far distances and see the structure between data points.  
**Note:** If you turn off orthographic view and completely decrease the perspective, the walls of the scatterplot do not converge. This is the same effect that you get when you turn on orthographic view. |
| Perspective | Increases or decreases the perspective. Large values create a view that is unnaturally large and visually disorienting. In this case, you need to resize the scatterplot window to show the entire plot. |
| Marker Size | Increases or decreases the size of the data point markers. |
| Marker Quality | Increases and decreases the data marker quality. For example, when you increase the marker quality, some markers have an opaque center. Other symbol markers are formatted in bold. Increase the zoom to see these changes in quality. |
| Marker Transparency | Increases or decreases the transparency of the data markers. |
| Text Size  | Increases or decreases the text size. |
| Line Width | Changes the width of the coordinate and axes lines. |
Chapter 13

Creating Contour Plots

Using the Contour Plot Platform

The Contour Plot command in the Graph menu constructs contours of a response in a rectangular coordinate system. A contour plot shows a three-dimensional surface in two dimensions. Contours delineate changes in the third dimension.

Here are some of the options available with the Contour platform:

- specify the number of contour levels
- choose to plot contour lines or filled contours
- show or hide data points
- label contours with response values
- define and use a custom coloring scheme

Figure 13.1 Examples of Contour Plots
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example of a Contour Plot</td>
<td>369</td>
</tr>
<tr>
<td>Launch the Contour Plot Platform</td>
<td>369</td>
</tr>
<tr>
<td>The Contour Plot Report</td>
<td>371</td>
</tr>
<tr>
<td>Contour Plot Options</td>
<td>371</td>
</tr>
<tr>
<td>Fill Areas</td>
<td>372</td>
</tr>
<tr>
<td>Contour Specification</td>
<td>373</td>
</tr>
<tr>
<td>Contour Plot Save Options</td>
<td>375</td>
</tr>
<tr>
<td>Use Formulas for Specifying Contours</td>
<td>375</td>
</tr>
</tbody>
</table>
Example of a Contour Plot

To create a contour plot, you need two variables for the x- and y-axes and at least one more variable for contours. You can also use several y-variables. This example uses the Little Pond.jmp sample data table. X and Y are coordinates of a pond. Z is the depth.

1. Open the Little Pond.jmp sample data table.
2. Select Graph > Contour Plot.
3. Select the X and Y coordinates and click X.
4. Select the depth, Z, and click Y.

Note: In graphing platforms, the X and Y roles refer to the axes of the graph and not to the analytical concepts of factors and responses.

5. Click OK.

Figure 13.2 Example of a Contour Plot with Legend

The x- and y-axes are coordinates and the contour lines are defined by the depth variable. This contour plot is essentially a map of a pond showing depth.

Launch the Contour Plot Platform

By default, the contour levels used in the plot are values computed from the data. You can specify your own number of levels and level increments in the Launch window before you create the plot. You can also do so in the red triangle menu for Contour Plot after you create the plot. You can use a column formula to compute the contour variable values.
Creating Contour Plots

Example of a Contour Plot

Figure 13.3 The Contour Plot Launch Window

Table 13.1 Description of the Contour Plot Launch Window

<table>
<thead>
<tr>
<th>Role</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y</td>
<td>Columns assigned to the Y role are used as variables to determine the contours of the plot. You must specify at least one, and you can specify more than one. You can also assign a column with a formula to this role. If you do so, the formula should be a function of exactly two variables. Those variables should be the x variables entered in the Launch window.</td>
</tr>
<tr>
<td>X</td>
<td>Columns assigned to the X role are used as the variables for the x- and y-axes. You must specify exactly two columns for X.</td>
</tr>
<tr>
<td>By</td>
<td>This option produces a separate graph for each level of the By variable. If two By variables are assigned, a separate graph for each possible combination of the levels of both By variables is produced.</td>
</tr>
<tr>
<td>Options</td>
<td></td>
</tr>
<tr>
<td>Contour Values</td>
<td>Specify your own number of levels and level increments. See “Contour Specification,” p. 373.</td>
</tr>
<tr>
<td>Fill Areas</td>
<td>Fill the areas between contour lines using the contour line colors.</td>
</tr>
</tbody>
</table>
The Contour Plot Report

Follow the instructions in “Example of a Contour Plot,” p. 369 to produce the plot shown in Figure 13.4. The legend for the plot shows individual markers and colors for the $Y$ variable. For information about additional options for the report, see “Contour Plot Options,” p. 371.

Contour Plot Options

Using the options in the red triangle menu next to Contour Plot, you can tailor the appearance of your contour plot and save information about its construction.
Creating Contour Plots

Contour Plot Options

Table 13.2 Descriptions of Contour Plot Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show Data Points</td>
<td>Shows or hides ((x, y)) points. The points are hidden by default.</td>
</tr>
<tr>
<td>Show Missing Data</td>
<td>Shows or hides points with missing (y) values. Only available if Show Data Points is selected.</td>
</tr>
<tr>
<td>Points</td>
<td></td>
</tr>
<tr>
<td>Show Contours</td>
<td>Shows or hides the contour lines or fills. The contour lines are shown by default.</td>
</tr>
<tr>
<td>Show Boundary</td>
<td>Shows or hides the boundary of the total contour area. The boundary is shown by default.</td>
</tr>
<tr>
<td>Fill Areas</td>
<td>Fills the areas between the contours with a solid color. It is the same option that is available in the Launch window. If you leave it deselected in the Launch window, you can see the line contours before filling the areas.</td>
</tr>
<tr>
<td>Label Contours</td>
<td>Shows or hides the label ((z)-value) of the contour lines.</td>
</tr>
<tr>
<td>Color Theme</td>
<td>Select another color theme for the contours.</td>
</tr>
<tr>
<td>Reverse Colors</td>
<td>Reverses the order of the colors assigned to the contour levels.</td>
</tr>
<tr>
<td>Change Contours</td>
<td>Set your own number of levels and level increments.</td>
</tr>
<tr>
<td>Save</td>
<td>This menu has options to save information about contours, triangulation, and grid coordinates to data tables.</td>
</tr>
<tr>
<td></td>
<td>See “Contour Plot Save Options,” p. 375.</td>
</tr>
<tr>
<td>Script</td>
<td>This menu contains commands that are available to all platforms. They enable you to redo the analysis or save the JSL commands for the analysis to a window or a file. See “Script Menus,” p. 15 in the “Preliminaries” chapter.</td>
</tr>
</tbody>
</table>

Fill Areas

If you select Fill Areas, the areas between contour lines are filled with the contour line colors. This option is available in the Launch window and in the red triangle menu for Contour Plot. Figure 13.5 shows a plot with contour lines on the left and a plot with the contour areas filled on the right.
Areas are filled from low to high values. An additional color is added in the filled contour plot for the level above the last, and highest, contour line.

**Contour Specification**

If you do not select options in the Launch window, the default plot spaces the contour levels equally within the range of the $Y$ variable. The default colors are assigned from low- to high-level values by cycling from top to bottom through the middle column of the color palette. You can see the colors palette with the Colors command in the Rows menu or by right-clicking on an item in the Contour Plot legend.

You can specify contour levels either in the Launch window (the Specify button) or in the report window from the red triangle menu for Contour Plot (the Specify Contours option).
Specify

This option is both in the Launch window and on the red triangle menu for Contour Plot (the Specify Contours option).

Selecting this option displays the Contour Specification window. See Figure 13.7. Using this window, you can do the following:

- change the number of contours
- specify minimum and maximum values to define the range of the response to be used in the plot
- change the increment between contour values

You supply any three of the four values, and the remaining value is computed for you. Click on the check box to deselect one of the numbers and automatically select the remaining check box.

Figure 13.7 The Contour Specification Window

Colors are automatically assigned and are determined by the number of levels in the plot. After the plot appears, you can right-click (press CONTROL and click on the Macintosh) on any contour in the plot legend and choose from the JMP color palette to change that contour color.

Retrieve

This option is both in the Launch window (the Retrieve button) and on the red triangle menu for Contour Plot (the Retrieve Contours option).

Note: Neither the button nor the menu option are active unless there is an open data table in addition to the table that has the contour plotting values. When you click Retrieve or select Retrieve Contours, a window with a list of open data tables appears.

Using this option, you can retrieve the following from an open JMP data table:

- the number of contours
- an exact value for each level
- a color for each level

From the list of open data tables, select the data table that contains the contour levels.
For level value specification, the Contour Plot platform looks for a numeric column with the same name as the response column that you specified in the Launch window. The number of rows in the data table defines the number of levels.

If there is a row state column with color information, those colors are used for the contour levels. Otherwise, the default platform colors are used.

**Revert Contours**

This option appears only on the red triangle menu for Contour Plot.

If you have specified your own contours, selecting this option reverts your Contour Plot back to the default contours.

**Contour Plot Save Options**

This menu has options to save information about contours, triangulation, and grid coordinates to data tables.

- **Save Contours** creates a new JMP data table with columns for the following:
  - the \( x \)- and \( y \)-coordinate values generated by the Contour platform for each contour
  - the response computed for each coordinate set
  - the curve number for each coordinate set

The number of observations in this table depends on the number of contours you specified. You can use the coordinates and response values to look at the data with other JMP platforms. For example, you can use the Scatterplot 3D platform to get a three-dimensional view of the pond.

- **Generate Grid** displays a window that prompts you for the grid size that you want. When you click **OK**, the Contour platform creates a new JMP data table with the following:
  - the number of grid coordinates you requested
  - the contour values for the grid points computed from a linear interpolation

- **Save Triangulation** creates a new JMP data table that lists coordinates of each triangle used to construct the contours.

**Use Formulas for Specifying Contours**

Most often you construct a contour plot for a table of recorded response values such as the Little Pond data table. In that case, in the launch window, **Use Table Data** is checked and the **Specify Grid** button is unavailable. However, if a column has a formula and you specify that column as the response (\( Y \)), the **Specify Grid** button becomes active.

When you click **Specify Grid**, the window shown in Figure 13.8 appears.
You can complete the Specify Grid window and define the contour grid in any way, regardless of the rows in the existing data table. This feature is also available with table templates that have one or more columns defined by formulas but no rows.
A bubble plot is a scatter plot that represents its points as circles, or bubbles. Bubble plots can be dynamic (animated over time) or static (fixed bubbles that do not move). Use bubble plots to:

- dynamically animate bubbles using a time variable, to see patterns and movement across time
- use size and color to clearly distinguish between different variables
- aggregate data (rows) into a single bubble, simplifying the bubble plot

Because you can see up to five dimensions at once (x position, y position, size, color, and time), bubble plots can produce dramatic visualizations and readily show patterns and trends.

**Note:** Dynamic bubble plots were pioneered by Hans Rosling, Professor of International Health, Karolinska Institutes, and the people involved in the Gapminder.org project.
### Contents

- Example of a Dynamic Bubble Plot .................................................. 379
- Launch the Bubble Plot Platform ....................................................... 380
- Interact with the Bubble Plot .......................................................... 382
  - Control Animation for Dynamic Bubble Plots ................................. 383
  - Specify a Time or ID Variable ......................................................... 384
  - Select Bubbles .............................................................................. 388
  - Use the Brush Tool ....................................................................... 389
- Bubble Plot Platform Options ........................................................... 389
- Show Roles ....................................................................................... 390
- Additional Examples ......................................................................... 392
  - Example of a Static Bubble Plot ........................................................ 392
  - Example of a Bubble Plot with a Categorical Y Variable ....................... 396
Chapter 14
Creating Bubble Plots

Example of a Dynamic Bubble Plot

This example uses the PopAgeGroup.jmp sample data table, which contains population data for countries and regions around the world. Examine the relationship between the proportion of younger and older people in the sample populations.

1. Open the PopAgeGroup.jmp sample data table.
2. Select Graph > Bubble Plot.
   The launch window appears.

3. Select Portion60+ and click Y.
   The portion of the population that are 60 years or older becomes the y coordinate.
4. Select Portion 0-19 and click X.
   The portion of the population that are 0-19 years becomes the x coordinate.
5. Select Country and click ID.
   All the rows for each country are aggregated into a single bubble.
6. Select Year and click Time.
   The bubble plot shows a unique plot for each year's data.
7. Select Pop and click Sizes.
   The sizes of the bubbles reflect the overall population values.
8. Select Region and click Coloring.
   The bubble for each region is colored differently.
9. Click OK.
   The report window appears.
10. Click Go to see the animated, dynamic report. Alternatively, you can click Step to move forward by one year.

11. (Optional) To view a legend that identifies each color with its region, select Legend from the red triangle menu.

As time progresses, you can see that the portion of the population that is 0-19 years decreases, and the portion of the population that is 60 years or more increases.

**Launch the Bubble Plot Platform**

Launch the Bubble Plot platform by selecting Graph > Bubble Plot.
### Figure 14.4 The Bubble Plot Launch Window

![Image of Bubble Plot Launch Window]

### Table 14.1 Description of the Bubble Plot Launch Window

<table>
<thead>
<tr>
<th>Column</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Y, X</strong></td>
<td>The Y and X columns become the $y$ and $x$ coordinates of the bubbles in the plot. These values can be continuous or categorical (nominal or ordinal).</td>
</tr>
<tr>
<td><strong>ID</strong></td>
<td>(Optional) ID variables identify rows that should be aggregated and shown as a single bubble. The default coordinates of each bubble are the averaged $x$ and $y$ values, and the default size of each bubble is the sum of the sizes of all aggregated members. Specifying a second ID variable provides a hierarchy of categories, but the bubbles are not split by the second category until they are selected and split interactively. If you specify a second ID variable, <strong>Split</strong> and <strong>Combine</strong> buttons appear in the report window. For example, you might specify a country as the first ID variable, resulting in a separate aggregated bubble for each country. A second ID variable, perhaps designating regions within each country, would further split each country when the interactive Split button under the graph is pressed.</td>
</tr>
</tbody>
</table>

For details about the Split and Combine buttons, see “Control Animation for Dynamic Bubble Plots,” p. 383.

For more details about ID variables, see “Specify a Time or ID Variable,” p. 384.
Creating Bubble Plots
Interact with the Bubble Plot

Table 14.1 Description of the Bubble Plot Launch Window (Continued)

| Time       | Maintains separate coordinates, sizes, and colors for each unique time period. The bubble plot shows these values for a single time period. For example, if the Time column contains years, the bubble plot is updated to show data by each year.
|            | Click and drag to move the time label on the plot.
|            | When data is missing within a time period, the value is linearly interpolated. If data is missing for the first or last time period, the value is not estimated, but left as missing.

| Sizes      | Controls the size of the bubbles. The area of the bubbles is proportional to the Size value. There is a minimum bubble size, to keep bubbles visible, even if the size value is zero. If Size is left blank, the default bubble size is proportional to the number of rows in that combination of Time and ID.

| Coloring   | Colors the bubbles according to the selected variable. If the selected variable is categorical (nominal or ordinal), each category is colored distinctly. If the selected variable is continuous, a gradient of colors is used.

| By         | Place a column here to produce a separate bubble plot for each level of the variable.

Interact with the Bubble Plot

Use the Bubble Plot platform in one of two modes:

- **Static mode**, where the bubbles are fixed and do not animate over time (no Time variable is specified). See “Example of a Static Bubble Plot,” p. 392.
- **Dynamic mode**, where the bubbles are animated over time (a Time variable is specified). See “Example of a Dynamic Bubble Plot,” p. 379.

You interact with both static and dynamic bubble plots in different ways.
Control Animation for Dynamic Bubble Plots

Use sliders and buttons to control the animation of dynamic bubble plots.

Figure 14.5 Animation Controls

<table>
<thead>
<tr>
<th>Slider or Button</th>
<th>Description</th>
<th>Visibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;Time variable&gt;</td>
<td>Controls which time values appear in the bubble plot. You manually drag the slider to see a progression of time. Click and drag on the time variable in the bubble plot to move its position.</td>
<td>Only appears if you have specified a variable for Time.</td>
</tr>
<tr>
<td>Speed</td>
<td>Adjusts the speed of the animation.</td>
<td>Only appears if you have specified a variable for Time.</td>
</tr>
<tr>
<td>Circle Size</td>
<td>Adjusts the size of the bubbles. The bubbles maintain their relative size, but their absolute size can be adjusted.</td>
<td>Appears on all bubble plots.</td>
</tr>
<tr>
<td>Step</td>
<td>Adjusts the time value by one unit and shows the next time value.</td>
<td>Only appears if you have specified a variable for Time.</td>
</tr>
<tr>
<td>Prev</td>
<td>Adjusts the time value by one unit and shows the previous time value.</td>
<td>Only appears if you have specified a variable for Time.</td>
</tr>
<tr>
<td>Go</td>
<td>Animates the bubble plot. Moves through all of the time values in order, and loops back to the beginning when the last time period is reached.</td>
<td>Only appears if you have specified a variable for Time.</td>
</tr>
<tr>
<td>Stop</td>
<td>Stops the bubble plot animation.</td>
<td>Only appears if you have specified a variable for Time.</td>
</tr>
<tr>
<td>Split</td>
<td>Separates the bubble represented by the first, larger ID variable into its smaller constituent parts, which are defined by the second, smaller ID column. Select the bubble and click Split. See Figure 14.9.</td>
<td>Only appears if you have specified two ID variables.</td>
</tr>
</tbody>
</table>
Specify a Time or ID Variable

- For dynamic plots, you might specify only a Time variable. See “Specify Only a Time Variable,” p. 384.
- For static plots, you might specify only ID variables. See “Specify Only ID Variables and Splitting a Bubble,” p. 386.

Specify Only a Time Variable

For dynamic bubble plots, you might specify only a Time variable and no ID variable. The resulting bubble plot contains a single moving bubble that tracks the series as the Time value changes.

1. Open the PopAgeGroup.jmp sample data table.
2. Select Graph > Bubble Plot.
3. Select Portion60+ and click Y.
4. Select Portion 0-19 and click X.
5. Select Year and click Time.
6. Click OK.

The initial report window appears.
7. Click on the bubble to select it.
   All rows in the data table are also highlighted.

8. From the red triangle menu, select **Trail Bubbles** and **Trail Lines**.

9. Click **Go**.
   The bubble plot animates, showing a trail for the single bubble.
For static bubble plots, you might specify one or two ID variables and no Time variable. The resulting bubble plot contains a bubble at each ID value. Note that although this bubble plot is static, you can perform splitting on bubbles.

1. Open the PopAgeGroup.jmp sample data table.
2. Select Graph > Bubble Plot.
3. Select Portion60+ and click Y.
4. Select Portion 0-19 and click X.
5. Select Region and Country and click ID.
6. Select Region and click Coloring.
7. Click OK.

The initial report window appears.
8. From the red triangle menu, select **Legend**.

Split the bubble representing the region of North America into countries.

9. Click on the bubble representing North America (hover over a bubble to see its label, or use the legend to find the color of North America.)

10. Click **Split**.

You see that the North America bubble has split into three bubbles, representing the countries within the region of North America (the United States of America, Canada, and Mexico).
Creating Bubble Plots

Chapter 14

Interact with the Bubble Plot

Figure 14.9 Splitting the North America Bubble

Select Bubbles

Click on a bubble to select it. Note the following:

- Visually, selected bubbles become darker or brighter, and non-selected bubbles are more transparent.
- If the bubble was not filled initially, selection fills it.
- If no bubbles are selected, all of the bubbles are semi-transparent.

When you select a bubble, all of the rows in the data table that correspond to the selected bubble are highlighted. Note the following:

- If the bubble is an aggregate based on an ID column, all of the rows for that ID are highlighted. Otherwise, the one row represented by that bubble is highlighted.
- If you specify an ID and a Time variable, selecting a bubble highlights all of the rows for that ID, across all of the Time levels.

If you select a row from the data table, it is selected in the associated bubble plot. Note the following:

- If you have not specified a Time variable, selecting one row from the data table highlights the corresponding bubble in the plot.
- If you have specified a Time variable, selecting one row from the data table highlights the corresponding bubble for only that time period in the dynamic bubble plot.
Use the Brush Tool

Use the brush tool to temporarily select bubbles and obtain more information about the selected bubbles. For dynamic bubble plots, turn on the Trail Bubbles or the Trail Lines option to show the selected bubbles during animation. When you select bubbles with the brush tool, the corresponding rows are highlighted in the associated data table.

Note: For a more granular examination of the highlighted rows, use the Tables > Subset command or the Row Editor. See the Using JMP book.

Bubble Plot Platform Options

The following table describes the options in the Bubble Plot red triangle menu.

<table>
<thead>
<tr>
<th>Table 14.3 Platform Options</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Filled</strong></td>
</tr>
<tr>
<td><strong>Trail Bubbles</strong></td>
</tr>
<tr>
<td><strong>Trail Lines</strong></td>
</tr>
<tr>
<td><strong>All Labels</strong></td>
</tr>
<tr>
<td><strong>No Labels</strong></td>
</tr>
<tr>
<td><strong>Legend</strong></td>
</tr>
<tr>
<td><strong>Selectable Across Gaps</strong></td>
</tr>
<tr>
<td><strong>Show Roles</strong></td>
</tr>
</tbody>
</table>
Creating Bubble Plots

Chapter 14

Bubble Plot Platform Options

Show Roles

Using the Show Roles option in the red triangle menu, you can make changes to your existing variables without having to relaunch the platform and start your analysis over.

Follow the instructions in “Example of a Dynamic Bubble Plot,” p. 379 to produce the report window shown in Figure 14.10.

Table 14.3 Platform Options (Continued)

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Split All</td>
<td>Splits all bubbles into their constituent parts. Unlike the Split button, the bubbles do not have to be selected. Only appears if you have specified two ID variables.</td>
</tr>
<tr>
<td>Combine All</td>
<td>Combines all constituent bubbles within a group into their larger bubble. Unlike the Combine button, the bubbles do not have to be selected. Only appears if you have specified two ID variables.</td>
</tr>
<tr>
<td>Aggregation Options</td>
<td>Alters how the X, Y, and Sizes roles are computed. By default, the values are calculated using means for X and Y, and sums for Sizes.</td>
</tr>
<tr>
<td></td>
<td>- Selecting X as Sum or Y as Sum computes the X and Y values using sums.</td>
</tr>
<tr>
<td></td>
<td>- Deselecting Size as Sum computes Size values using means.</td>
</tr>
<tr>
<td>Save for Adobe Flash platform (.SWF)</td>
<td>Saves the bubble plot as .SWF files that are Adobe Flash player compatible. You can use these files in presentations and in web pages. An HTML page is also saved that shows you the correct code for using the resulting .SWF file. For more information about this option, go to <a href="http://www.jmp.com/support/swfhelp/en">http://www.jmp.com/support/swfhelp/en</a> or <a href="http://www.jmp.com/support/notes/34/780.html">http://www.jmp.com/support/notes/34/780.html</a>.</td>
</tr>
<tr>
<td>Script</td>
<td>This menu contains options that are available to all platforms. They enable you to redo the analysis or save the JSL commands for the analysis to a window or a file. See “Script Menus,” p. 15 in the “Preliminaries” chapter.</td>
</tr>
</tbody>
</table>
**Bubble Plot Platform Options**

**Figure 14.10** Example of Bubble Plot with Show Roles Selected

![Bubble Plot Example](image)

**Change the Variable Assigned to a Role**

To change the variable assigned to a role, click on a blue underlined role name. For example, in Figure 14.10, to change the Coloring variable from Region to Country:

1. Click on the Coloring link.
   
   The Select column for Coloring window appears.

2. Click on Country.

3. Click **OK**.
   
   Country now replaces Region as the Coloring variable in the bubble plot.

**Remove a Variable**

To remove an existing variable from the bubble plot, make sure that nothing is selected in the Select column for *Role* window, and click **OK**. For example, in Figure 14.10, to remove the Sizes variable (Pop), proceed as follows:

1. Click on the Sizes link.
   
   The Select column for Sizes window appears.

2. Ensure that nothing is selected. If a variable is selected, deselect it by holding down the CTRL key (COMMAND key on the Macintosh) and clicking on the variable.
3. Click OK.
   The Sizes role now appears with an empty box.

**Note:** The X and Y variables can only be changed and cannot be removed.

**Add a Variable**

Once you have removed an existing variable from the bubble plot, there are two ways to add a new variable:

- Click on the blue underlined role name. See “Change the Variable Assigned to a Role,” p. 391.
- In the data table, click on the variable in the column panel, and drag it into the empty role box.

**Additional Examples**

The following examples further illustrate the features of the Bubble Plot platform.

**Example of a Static Bubble Plot**

This example uses the SATByYear.jmp sample data table, which contains SAT verbal and math test scores for a selection of the US population in 2004.

1. Open the SATByYear.jmp sample data table.
2. Select Graph > Bubble Plot.
3. Select SAT Verbal and click Y.
4. Select SAT Math and click X.
5. Select State and click ID.
7. Click OK.
   The report window appears.
From Figure 14.11, you draw the following conclusions:

- Higher verbal scores appear to be associated with higher math scores, since the two track very closely in the bubble plot. This signifies a correlation between verbal and math scores.

- The larger bubbles represent the US states that have a high percentage of individuals taking the SAT test in 2004. These larger bubbles are all grouped together in the lower left of the graph. This shows that when a state has a high percentage of individuals taking the test, both the math and verbal scores are low.

Instead of grouping the bubbles primarily by state, group the bubbles primarily by region as follows:

1. From the red triangle menu, select Show Roles.
2. Click on the ID link.
3. Select Region, and click OK.
   Region is now the primary ID variable.
4. Click on the ID2 link.
5. Select State, and click OK.
   State is now the secondary ID variable.
6. Click on the bubble that represents the Southwest region (hover over a bubble or click on it to see its label).

7. Click **Split**.

   Now the bubbles are split by the secondary ID variable, which is **State**. You now see each state within the Southwest region.
From Figure 14.13, you see that there is significant variation between the scores from the Southwest states.

8. Click **Combine** to combine the southwest states again.

9. To do a comparison, click on the New England bubble (hover over a bubble or click on it to see its label).

10. Click **Split**.
Figure 14.14 Example of New England Region Split by State

You see that the New England states do not have as much variation as the Southwest states.

Example of a Bubble Plot with a Categorical Y Variable

All of the examples shown so far use continuous Y variables. If you use a categorical (nominal or ordinal) Y variable, the bubble plot appears differently.

This example uses the blsPriceData.jmp sample data table, which shows the price of commodities over several years. Because the value of the US dollar changes over time, a column named Price/Price2000 shows the ratio of a commodity’s price at any given time to the price in the year 2000.

1. Open the blsPriceData.jmp sample data table.
2. Select Graph > Bubble Plot.
3. Select Series and click Y.
4. Select Price/Price2000 and click X.
5. Select date and click Time.
6. Click OK.

The report window appears.

This produces a bubble plot that, when animated by clicking Go, shows the price bubbles moving side to side according to their price ratio.
For easier readability, add grid lines as follows:

7. Double-click on the categorical axis.
8. In the Y Axis Specification window, select the box under Gridline next to Major.
9. Click OK.

To animate the bubble plot, click Go. The price bubbles move side to side, according to their price ratio.
Using parallel and cell plots, you can visualize each cell in a data table.

- Parallel plots draw connected line segments that represent each row in a data table. Parallel plots were initially developed by Inselberg (1985) and later popularized by Wegman (1990).

- Cell plots are direct representations of a data table, since they draw a rectangular array of cells where each cell corresponds to a data table entry. Cell plots were popularized by genomics applications to browse large numbers of values for gene expression levels.

Figure 15.1 Example of a Parallel Plot and a Cell Plot
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example of a Parallel Plot</td>
<td>401</td>
</tr>
<tr>
<td>Launch the Parallel Plot Platform</td>
<td>402</td>
</tr>
<tr>
<td>The Parallel Plot Report</td>
<td>403</td>
</tr>
<tr>
<td>Interpreting Parallel Plots</td>
<td>404</td>
</tr>
<tr>
<td>Parallel Plot Platform Options</td>
<td>405</td>
</tr>
<tr>
<td>Additional Examples of Parallel Plots</td>
<td>406</td>
</tr>
<tr>
<td>Examine Iris Measurements</td>
<td>406</td>
</tr>
<tr>
<td>Examine Student Measurements</td>
<td>407</td>
</tr>
<tr>
<td>Example of a Cell Plot</td>
<td>409</td>
</tr>
<tr>
<td>Launch the Cell Plot Platform</td>
<td>410</td>
</tr>
<tr>
<td>The Cell Plot Report</td>
<td>410</td>
</tr>
<tr>
<td>Cell Plot Platform Options</td>
<td>411</td>
</tr>
<tr>
<td>Context Menu for Cell Plots</td>
<td>412</td>
</tr>
<tr>
<td>Additional Example of a Cell Plot</td>
<td>413</td>
</tr>
</tbody>
</table>
Example of a Parallel Plot

This example uses the Dogs.jmp sample data table, which contains histamine level measurements for 16 dogs that were given two different drugs. The histamine levels were taken at zero, one, three, and five minutes. Examine the variation in the histamine levels for each drug.

1. Open the Dogs.jmp sample data table.

To see the differences by drug, color the parallel plot lines by drug:

2. Select Rows > Color or Mark by Column.
3. Select drug and click OK.

Morphine appears red and trimeth appears blue.

Create the parallel plot:

4. Select Graph > Parallel Plot.
5. Select hist0, hist1, hist3, and hist5 and click Y, Response.
6. Click OK.

The report window appears.

Figure 15.2 Parallel Plot of Histamine Variables

Each connected line segment represents a single observation. Click on a line segment to see which observation (or row) it corresponds to in the data table.

For further exploration, isolate the trimeth values:

7. Select Rows > Data Filter.
8. Select drug and click Add.
9. Select trimeth.

Only the trimeth values are highlighted in the parallel plot.
Creating Cell-Based Plots

Chapter 15

Example of a Parallel Plot

Figure 15.3 Trimeth Values Highlighted

From Figure 15.3, you observe the following about the histamine levels for dogs given trimeth:

- For most of the dogs, the histamine levels had a sharp drop at one minute.
- For four of the dogs, the histamine levels remained high, or rose higher. You might investigate this finding further, to determine why the histamine levels were different for these dogs.

Launch the Parallel Plot Platform

Launch the Parallel Plot platform by selecting Graph > Parallel Plot.

Figure 15.4 The Parallel Plot Launch Window
Table 15.1 Description of the Parallel Plot Launch Window

<table>
<thead>
<tr>
<th>Y, Response</th>
<th>Variables appear on the horizontal axis of the parallel plot. These values are plotted and connected in the parallel plot.</th>
</tr>
</thead>
<tbody>
<tr>
<td>X, Grouping</td>
<td>Produces a separate parallel plot for each level of the variable.</td>
</tr>
<tr>
<td>By</td>
<td>Identifies a column that creates a report consisting of separate analyses for each level of the specified variable.</td>
</tr>
<tr>
<td>Scale Uniformly</td>
<td>Represents all variables on the same scale, adding a y-axis to the plot. Without this option, each variable is on a different scale. To allow for proper comparisons, select this option if your variables are measured on the same scale.</td>
</tr>
<tr>
<td>Center at zero</td>
<td>Centers the parallel plot (not the variables) at zero.</td>
</tr>
</tbody>
</table>

The Parallel Plot Report

To produce the plot shown in Figure 15.5, follow the instructions in “Example of a Parallel Plot,” p. 401.

Figure 15.5 The Parallel Plot Report

A parallel plot is one of the few types of coordinate plots that show any number of variables in one plot. However, the relationships between variables might be evident only in the following circumstances:

- when the variables are side by side
- if you assign a color to the variables to track groups
- if you select lines to track groups

Note: If the columns in a parallel plot use the Spec Limits column property, the specification limits appear as red lines.
Interpreting Parallel Plots

To help you interpret parallel plots, compare the parallel plot with a scatterplot. In each of the following figures, the parallel plot appears on the left, and the scatterplot appears on the right.

**Strong Positive Correlation**

The following relationship shows a strong positive correlation. Notice the coherence of the lines in the parallel plot.

![Figure 15.6 Strong Positive Correlation](image)

**Strong Negative Correlation**

A strong negative correlation, by contrast, shows a narrow neck in the parallel plot.

![Figure 15.7 Strong Negative Correlation](image)

**Collinear Groups**

Now, consider a case that encompasses both situations: two groups, both strongly collinear. One has a positive slope, the other has a negative slope. In Figure 15.8, the positively sloped group is highlighted.
Single Outlier

Finally, consider the case of a single outlier. The parallel plot shows a general coherence among the lines, with a noticeable exception.

Parallel Plot Platform Options

The following table describes the options within the red triangle menu for Parallel Plot.

**Table 15.2 Descriptions of Parallel Plot Red Triangle Options**

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show Reversing Checkboxes</td>
<td>Reverses the scale for one or more variables.</td>
</tr>
<tr>
<td>Script</td>
<td>This menu contains options that are available to all platforms. They enable you to redo the analysis or save the JSL commands for the analysis to a window or a file. See “Script Menus,” p. 15 in the “Preliminaries” chapter.</td>
</tr>
</tbody>
</table>
Additional Examples of Parallel Plots

The following examples further illustrate using the Parallel Plots platform.

Examine Iris Measurements

The following example uses the Fisher’s Iris data set (Mardia, Kent, and Bibby 1979). The Iris jmp sample data table contains measurements of the sepal length and width and petal length and width in centimeters for three species of Iris flowers: setosa, versicolor, and virginica. To find characteristics that differentiate the three species, examine these measurements.

Examine Three Species in One Parallel Plot

1. Open the Iris jmp sample data table.
2. Select Graph > Parallel Plot.
4. Select the Scale Uniformly check box.
5. Click OK.

The report window appears.

Figure 15.10 Three Species in One Parallel Plot
In this parallel plot, the three species are all represented in the same plot. The colors correspond to the three species, as follows:

- Blue corresponds to virginica.
- Green corresponds to versicolor.
- Red corresponds to setosa.

From Figure 15.10, you observe the following:

- For sepal width, the setosa values appear to be higher than the virginica and versicolor values.
- For petal width, the setosa values appear to be lower than the virginica and versicolor values.

**Examine Three Species in Different Parallel Plots**

1. From the Iris.jmp sample data table, select Graph > Parallel Plot.
2. Select Sepal length, Sepal width, Petal length, and Petal width and click Y, Response.
3. Select Species and click X, Grouping.
4. Click OK.

The report window appears.

![Figure 15.11 Three Species in Different Parallel Plots](image)

Each species is represented in a separate parallel plot.

**Examine Student Measurements**

The following example uses the Big Class.jmp sample data table, which contains data on age, sex, height, and weight for 40 students. Examine the relationships between different variables.

1. Open the Big Class.jmp sample data table.
2. Select Graph > Parallel Plot.
4. Select age and click X, Grouping.
5. Select sex and click By.
6. Select the Scale Uniformly check box.
7. Click OK.

**Figure 15.12 Height and Weight by Sex, Grouped by Age**

From Figure 15.12, you observe the following:

- Among the 13 year old females, one female's weight is lower than the other females in her age group. If you click on the line representing the lower weight, the respective individual (Susan) is highlighted in the data table.

- Among the 14 year old females, one female's weight is higher than the other females in her age group. If you click on the line representing the higher weight, the respective individual (Leslie) is highlighted in the data table.
Example of a Cell Plot

This example uses the Dogs.jmp sample data table, which contains histamine level measurements for 16 dogs that were given two different drugs. The histamine levels were taken at zero, one, three, and five minutes. Examine the variation in the histamine levels for each drug.

1. Open the Dogs.jmp sample data table.
2. Select Graph > Cell Plot.
3. Select drug, hist0, hist1, hist3, and hist5 and click Y, Response.
4. Click OK.

The report window appears.

From Figure 15.13, notice the following:

- There are two types of drugs, represented by two distinct colors.
- Histamine levels are assigned colors from a gradient of blue to red.
- Any missing values are delineated by an X.
- The third row from the bottom is selected, and black lines appear next to the cells.
Launch the Cell Plot Platform

Launch the Cell Plot platform by selecting Graph > Cell Plot.

Figure 15.14 The Cell Plot Launch Window

Table 15.3 Description of the Cell Plot Launch Window

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Y, Response</td>
<td>Variables appear on the horizontal axis of the cell plot. Each cell represents a value.</td>
</tr>
<tr>
<td>X, Grouping</td>
<td>Produces a separate cell plot for each level of the variable.</td>
</tr>
<tr>
<td>Label</td>
<td>Labels each row by the specified variable. See “Additional Example of a Cell Plot,” p. 413.</td>
</tr>
<tr>
<td>By</td>
<td>Identifies a column that creates a report consisting of separate analyses for each level of the variable.</td>
</tr>
<tr>
<td>Scale Uniformly</td>
<td>Represents all variables on the same scale. Without this option, each variable is on a different scale.</td>
</tr>
<tr>
<td>Center at zero</td>
<td>Centers the cell plot at zero.</td>
</tr>
</tbody>
</table>

The Cell Plot Report

To produce the plot shown in Figure 15.15, follow the instructions in “Example of a Cell Plot,” p. 409.
Cell plots are direct representations of a data table, drawn as a rectangular array of cells with each cell corresponding to a data table entry. Colors are assigned to each cell based on the range and type of values found in the column.

- Nominal variables use a distinct color for each level. You can customize nominal and ordinal colors using the Value Colors property of data columns, available through the Column Info command.
- Continuous variables are assigned a gradient of colors to show the smooth range of values in the variable.
- Ordinal variables are scaled like continuous variables in order.
- When some outliers are present, the scale uses all but the extreme categories for the 90% middle of the distribution, so that the outliers do not overly influence the scale.

The cell plot appears with a one-to-one correspondence of a colored cell representing each data table entry.

For details about the options in the red triangle menu, see “Cell Plot Platform Options,” p. 411.

**Cell Plot Platform Options**

The following table describes the options within the red triangle menu for Cell Plot.
Table 15.4 Descriptions of Cell Plot Red Triangle Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Legend</td>
<td>Shows or hides a legend.</td>
</tr>
<tr>
<td>Arrange Plots</td>
<td>Specifies how many plots to put on the same row before starting the next row of plots. This option is available only if you specify an X, Grouping variable.</td>
</tr>
<tr>
<td>Script</td>
<td>This menu contains options that are available to all platforms. They enable you to redo the analysis or save the JSL commands for the analysis to a window or a file. See “Script Menus,” p. 15 in the “Preliminaries” chapter.</td>
</tr>
</tbody>
</table>

Context Menu for Cell Plots

The following table describes the options that appear when you right-click on a cell plot.

Note: For details about the context options that appear when you right-click on labels, see the Using JMP book.

Table 15.5 Descriptions of Cell Plot Context Menu Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Graph Type</td>
<td>Determines the appearance of the graph. See “Graph Type,” p. 413.</td>
</tr>
<tr>
<td>Color Theme</td>
<td>Shows a list of color themes that affect continuous variables in color maps. The default color theme is Blue to Gray to Red (corresponding to small values to middle values to large values). Use White to Black to create a gray-scale plot. Note: To see custom colors, you must first create them. Go to File &gt; Preferences &gt; Reports and click on the Customize Themes button.</td>
</tr>
<tr>
<td>Sort Ascending</td>
<td>Sorts the rows of the plot from lowest to highest by the values of a column. To sort, right-click in the plot under a column and select Sort Ascending. The entire plot is rearranged to accommodate the sorting. See “Additional Example of a Cell Plot,” p. 413.</td>
</tr>
<tr>
<td>Sort Descending</td>
<td>Sorts the rows of the plot from highest to lowest by the values of a column. To sort, right-click in the plot under a column and select Sort Descending. The entire plot is rearranged to accommodate the sorting.</td>
</tr>
<tr>
<td>No Separator Lines</td>
<td>Draws or removes lines separating the columns.</td>
</tr>
</tbody>
</table>
Graph Type

Use the Graph Type option to change the appearance of the graph.

**Figure 15.16  Graph Types**

![Graph Types Diagram]

---

**Additional Example of a Cell Plot**

This example uses the SAT.jmp sample data table, which contains SAT test scores (divided into verbal and mathematics portions) for all 50 United States.

1. Open the SAT.jmp sample data table.
2. Select Graph > Cell Plot.
3. Select all of the Verbal scores for all of the years, and click Y, Response.
4. Select all of the Math scores for all of the years, and click Y, Response.
5. Select State and click Label.
6. Click OK.
   The report window appears.
7. Right-click on the plot under 2004 Verbal (the top left cell) and select Sort Ascending.
   This sorts the cell plot by the verbal scores for 2004.
From Figure 15.17, you notice the following:

- Hawaii has the lowest verbal scores for 2004, and South Dakota has the highest verbal scores for 2004.
- There is a contrast between Hawaii’s math and verbal scores. Hawaii has average math scores (represented by gray color values) but low verbal scores (represented by blue color values). Hawaii appears to be an outlier, since it has a strikingly different pattern for its math scores and its verbal scores.
- There is very little contrast between North Dakota’s math and verbal scores. North Dakota’s math and verbal scores are generally high (represented by red color values).

For a description of color themes, see “Context Menu for Cell Plots,” p. 412.
Chapter 16

Creating Tree Maps
Using the Tree Map Platform

Tree maps are useful for observing patterns among groups that have many levels. Tree maps are like bar charts that have been folded over in two dimensions so that there is no unused space. Rather than drawing a single bar for each measurement, a tree map can show the magnitude of a measurement by varying the size or color of a rectangular area. Use tree maps when your data contains many categories, to visualize many groups.

Tree maps were named and popularized by Ben Schneiderman, who has an extensive Web site about the idea (http://www.cs.umd.edu/hcil/treemap-history/index.shtml)

Figure 16.1 Example of a Tree Map
Contents

Example of Tree Maps ................................................................. 417
Launch the Tree Map Platform .................................................. 419
   Categories ................................................................. 420
   Sizes ................................................................. 421
   Ordering ............................................................. 422
   Coloring ......................................................... 424
Tree Map Report Window ...................................................... 426
Tree Map Platform Options .................................................. 427
   Context Menu ..................................................... 428
Additional Tree Map Examples .............................................. 428
   Examine Pollution Levels ................................................ 428
   Examine Causes of Failure ........................................... 430
   Examine Patterns in Car Safety ...................................... 431
Example of Tree Maps

Tree maps can be useful in cases where histograms or bar charts are ineffective. This example uses the Cities.jmp sample data table, which contains meteorological and demographic statistics for 52 cities. Compare the bar chart to the tree map.

Create a bar chart representing ozone levels in each of the 52 cities:

1. Open the Cities.jmp sample data table.
2. Select Graph > Chart.
   The launch window appears.
3. Select OZONE and click Statistics.
4. Select Mean.
5. Select city and click Categories, X, Levels.
6. Click OK.
   The report window appears.

![Figure 16.2 Ozone Levels in a Bar Chart](image)

Although it is easy to see that there is a single large measurement, each bar looks similar. Subtle distinctions are difficult to see.

Create a tree map representing ozone levels in each of the 52 cities:

1. Return to the Cities.jmp sample data table.
2. Select Graph > Tree Map.
   The launch window appears.
3. Select POP (population) and click Sizes.
4. Select city and click **Categories**.
5. Select OZONE and click **Coloring**.
6. Click **OK**.

The report window appears.

**Figure 16.3** Ozone Levels in a Tree Map

Compare the bar chart to the tree map. Because the tree map folds the data over two dimensions (size and color), each city's data looks more distinctive than it did in the bar chart.

Note the following about this tree map:

- The magnitude of the ozone level for each city is represented by color.
- Each rectangle is colored based on a continuous color spectrum, with bright blue on the lowest end and bright red on the highest end. In this tree map, Des Moines has the lowest ozone levels, and Los Angeles has the highest ozone levels.
- Ozone levels somewhere in the middle decrease the intensity of the color, so pale blue, and pink indicate levels that are closer to the mean ozone level.
- Cities colored black have missing ozone values.

Since no size or ordering options have been specified, the area of each rectangle is proportional to the number of rows for that city, and the tree is split alphabetically.
Launch the Tree Map Platform

Launch the Tree Map platform by selecting **Graph > Tree Map**.

**Figure 16.4** The Tree Map Launch Window

<table>
<thead>
<tr>
<th>Select Columns</th>
<th>Cart</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>by</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Y</td>
<td></td>
<td></td>
</tr>
<tr>
<td>State</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Region</td>
<td></td>
<td></td>
</tr>
<tr>
<td>pop-rn</td>
<td></td>
<td></td>
</tr>
<tr>
<td>POP</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Max IM. f Jan</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OZONE</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SO2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NO</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PM10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lead</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Table 16.1** Description of the Tree Map Launch Window

<table>
<thead>
<tr>
<th></th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sizes</td>
<td>Determines the size of the rectangles based on the levels of the specified variable. See “Sizes,” p. 421.</td>
</tr>
<tr>
<td>Categories</td>
<td>(Required) Specifies the category that comprises the tree map. See “Categories,” p. 420.</td>
</tr>
<tr>
<td>Ordering</td>
<td>Changes the ordering from alphabetical (where values progress from the top left to the lower right) to order by the specified variable. You can specify more than one ordering variable. See “Ordering,” p. 422.</td>
</tr>
</tbody>
</table>
| Coloring | Colors the rectangles corresponding to the levels of the specified variable.  
  • If the variable is continuous, the colors are based on a continuous color spectrum.  
  • If the variable is categorical, the default colors are selected in order from JMP’s color palette. See “Coloring,” p. 424. |
| By     | Identifies a column that creates a report consisting of separate tree maps for each level of the variable. |
Categories

The only required variable role for the Tree Map platform is Categories. If you specify only a Categories variable and no other variables, the rectangles in the tree map have these attributes:

- They are colored from a rotating color palette.
- They are arranged alphabetically.
- They are sized by the number of occurrences in each group.

For example, using the Cities.jmp sample data table, specify city as the Categories variable. See Figure 16.5.

**Figure 16.5 City as Categories Variable**

Most of the rectangles are equal in size, because there is only one data point per city. The city of Portland has two data points, so its rectangle is twice as big.

**Specifying Two Categories Variables**

If you specify two Categories variables, the tree map is grouped by the first variable, and sorts within groups by the second variable. For example, using the Cities.jmp sample data table, specify Region and city (in that order) as the Categories variables.
Creating Tree Maps

Example of Tree Maps

Figure 16.6 Tree Map with Two Categories Variables

All of the states are grouped by their region and region labels appear on the tree map. Each region is ordered alphabetically by city.

Sizes

If you want the size of the rectangles to correspond to the levels of a variable, specify a Sizes variable. The rectangle size is proportional to the sum of the Sizes variable across all of the rows corresponding to a category. If you do not specify a Sizes variable, the rectangle size is proportional to the number of rows for each category.

For example, specify a Sizes variable as follows:

1. Open the Cities.jmp sample data table.
2. Select Graph > Tree Map.
3. Select POP (population) and click Sizes.
4. Select city and click Categories.
5. Click OK.

The report window appears.
Creating Tree Maps

Example of Tree Maps

Figure 16.7 POP as Sizes Variable

- Cities with large populations are represented by large rectangles, such as New York and Los Angeles.
- Cities with smaller populations are represented by smaller rectangles, such as Cheyenne and Dubuque.

Some rectangles are too small to show their labels (for example, Cheyenne and Dubuque). Click on a small rectangle to select the corresponding row in the data table, where you can see the label.

Ordering

By default, the rectangles in a tree map appear in alphabetical order, with values progressing from the top left to the lower right. To change this ordering, specify an Ordering variable.

If you specify a single Ordering variable, the rectangles are clustered, with the high levels or large values together, and the low levels or small values together. For example, using the Cities.jmp sample data table, specify an Ordering variable as follows:

1. Open the Cities.jmp sample data table.
2. Select Graph > Tree Map.
3. Select city and click Categories.
4. Select POP (population) and click Ordering.
5. Click OK.

The report window appears.
Example of Tree Maps

If you specify two Ordering variables, the tree map arranges the rectangles horizontally by the first ordering variable, and vertically by the second ordering variable. This approach can be useful for geographic data. For example, in the Cities.jmp sample data table, the X and Y columns correspond to the geographic location of the cities. Specify the X and Y columns as Ordering variables:

1. Open the Cities.jmp sample data table.
2. Select Graph > Tree Map.
3. Select city and click Categories.
4. Select X and click Ordering.
   The X variable corresponds to the western and eastern US states.
5. Select Y and click Ordering.
   The Y variable corresponds to the northern and southern US states.
6. Click OK.
   The report window appears.
The western and eastern US states are arranged horizontally, and the northern and southern US states are arranged vertically.

**Coloring**

If you specify a Coloring variable, the colors of the rectangles correspond to the levels of the variable.

- If the variable is continuous, the colors are based on a continuous color spectrum. The Blue to Gray to Red color theme is applied by default. The color of each category is based on the average value of all the rows. Blue represents the lowest values, and red represents the highest values. The color is most intense at the extremes of the variable, and paler colors correspond to levels that are close to the mean. For example, see “Example of a Continuous Coloring Variable,” p. 425.

- If the variable is categorical, the JMP Default color theme is applied by default. For example, see “Example of a Categorical Coloring Variable,” p. 425.

**Note:** If you have used the Value Colors column property to color a column, that property determines the colors of the categories.

- If you do not specify a Coloring variable, colors are chosen from a rotating color palette.

For more details about colors, see the Color Theme and Change Color Column options in “Tree Map Platform Options,” p. 427.
Example of a Continuous Coloring Variable

For example, using the Cities.jmp sample data table, specify a continuous Coloring variable as follows:

1. Open the Cities.jmp sample data table.
2. Select Graph > Tree Map.
3. Select city and click Categories.
4. Select OZONE and click Coloring.
5. Click OK.

The report window appears.

Note that the size of the rectangles is still based on the number of occurrences of the Categories variable, but the colors are mapped to ozone values. The high ozone value for Los Angeles clearly stands out. Missing values appear as black rectangles.

Example of a Categorical Coloring Variable

For example, using the Cities.jmp sample data table, specify a categorical Coloring variable as follows:

1. Open the Cities.jmp sample data table.
2. Select Graph > Tree Map.
3. Select city and click Categories.
4. Select Region and click Coloring.
5. Click OK.

The report window appears.
All of the cities belonging to the same region are colored the same color. The colors are chosen from JMP's color palette.

**Tree Map Report Window**

To produce the plot shown in Figure 16.12, follow the instructions in “Example of Tree Maps,” p. 417.
Tree map rectangles can have the following attributes:

- Categories add labels to the rectangles. You can specify one or two categories. You can show or hide labels using the Show Labels option. See “Context Menu,” p. 428.

- Rectangle size is determined by one of the following:
  - The Sizes variable, if you specify one.
  - If you do not specify a Sizes variable, size is determined by the size of the category.

- Rectangle color is determined by one of the following:
  - If you specify a continuous Coloring variable, colors are based on a continuous color spectrum. The Blue to Gray to Red color theme is applied by default. The color of each category is based on the average value of all the rows. Blue represents the lowest values, and red represents the highest values. The color is most intense at the extremes of the variable, and paler colors correspond to levels that are close to the mean.
  - If you specify a categorical (nominal or ordinal) Coloring variable, the JMP Default color theme is applied by default.
  - If you do not specify a Coloring variable, colors are chosen from a rotating color palette.
  - If you have used the Value Colors column property to color a column, that property determines the colors of the categories.

- The order of the rectangles is determined by one of the following:
  - The Ordering variable, if you specify one.
  - If you do not specify an Ordering variable, the order is alphabetical by default, with values progressing from the top left to the lower right.

For information about the options in the red triangle menu, see “Tree Map Platform Options,” p. 427.

Tree Map Platform Options

The following table describes the options within the red triangle menu next to Tree Map.

Table 16.2 Descriptions of Red Triangle Options

<table>
<thead>
<tr>
<th>Change Color Column</th>
<th>Change the column that is currently used to color the rectangles.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Color Theme</strong></td>
<td>Change the colors representing the high, middle, and low values of the color column. This option is available only if you have specified a Coloring variable.</td>
</tr>
<tr>
<td><strong>Color Range</strong></td>
<td>Specify the range that you want applied to the color gradient. The default low value is the column minimum, and the default high value is the column maximum. This option is available only if you have specified a continuous column as the Coloring variable.</td>
</tr>
<tr>
<td><strong>Legend</strong></td>
<td>Shows or hides a legend that defines the coloring used on the tree map. This option is available only if you have specified a Coloring variable.</td>
</tr>
</tbody>
</table>
Creating Tree Maps

Additional Tree Map Examples

Table 16.2 Descriptions of Red Triangle Options (Continued)

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Script</td>
<td>This menu contains options that are available to all platforms. They enable you to redo the analysis or save the JSL commands for the analysis to a window or a file. See “Script Menus,” p. 15 in the “Preliminaries” chapter.</td>
</tr>
</tbody>
</table>

Context Menu

The following table describes the options available by right-clicking on the tree map.

Table 16.3 Descriptions of Right-Click Menu Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Suppress Box Frames</td>
<td>Suppresses the black lines outlining each box.</td>
</tr>
<tr>
<td>Ignore Group Hierarchy</td>
<td>Flattens the hierarchy and sorts by the Ordering columns without using grouping, except to define cells.</td>
</tr>
<tr>
<td>Show Labels</td>
<td>Shows or hides the Categories labels. If you have specified two Categories, the secondary labels are hidden or shown. If you hide the Categories labels, hover over a rectangle to show the primary or secondary category label.</td>
</tr>
<tr>
<td>Show Group Labels</td>
<td>Shows or hides the group labels. This option is available only if more than one variable is assigned to the Categories role.</td>
</tr>
<tr>
<td>Group Label Background</td>
<td>Adjust the transparency of the group label. This option is available only if more than one variable is assigned to the Categories role.</td>
</tr>
</tbody>
</table>

Additional Tree Map Examples

The following examples further illustrate using the Tree Map platform.

Examine Pollution Levels

Using the Cities.jmp sample data table, examine the distribution of different pollution measurements (ozone and lead) across selected cities in the United States.

First, examine ozone levels in the selected cities.

1. Open the Cities.jmp sample data table.
2. Select Graph > Tree Map.
3. Select POP (population) and click Sizes.
4. Select city and click Categories.
5. Select X and Y and click Ordering.
6. Select OZONE and click Coloring.
7. Click **OK**.

   The report window appears.

**Figure 16.13** OZONE Levels for Selected Cities

From Figure 16.13, you observe the following:

- Los Angeles has a high level of ozone. It is also a western state, and that is reflected by its positioning in the tree map.
- Chicago and Houston have slightly elevated ozone levels.
- New York and Washington have slightly lower-than-average ozone levels.

Next, examine lead levels in the selected cities. Perform the same steps as before, but substitute **Lead** with **OZONE** for the **Coloring** variable.
From Figure 16.14, you observe the following:

- Cleveland and Seattle have high levels of lead. Cleveland is near the middle of the country, and Seattle is in the northwest. These locations are reflected by their positioning in the tree map.
- Interestingly, most other cities have rather low lead levels.
- Raleigh and Phoenix have missing values for lead measurements.

**Examine Causes of Failure**

This example uses the Failure3.jmp sample data table, which contains the common causes of failure during the fabrication of integrated circuits. Examine the causes of failure and when it occurs.

1. Open the Failure3.jmp sample data table, located in the Quality Control folder.
2. Select Graph > Tree Map.
3. Select N and click Sizes.
4. Select failure and clean and click Categories.
5. Select clean and click Coloring.
6. Click OK.

The report window appears.
From Figure 16.15, you observe the following:

- Contamination is the biggest cause of failure.
- Contamination occurs more often before the circuits were cleaned, rather than after they were cleaned.

**Examine Patterns in Car Safety**

This example uses the Cars.jmp sample data table, which contains impact measurements of crash-test dummies in automobile safety tests. Compare these measurements for different automobile makes and models during the years 1990 and 1991.

1. Open the Cars.jmp sample data table.
2. Select Rows > Data Filter.
3. Select Year.
4. Click And.
5. Select 90 and 91.

The rows corresponding to 1990 and 1991 are highlighted in the data table.

6. Select Tables > Subset.
7. Ensure that Selected Rows is selected and click OK.

A new data table (Subset of Cars) appears that contains only the data corresponding to the years 1990 and 1991.
Now, using the Subset of Cars data table, create the tree map.

1. Select **Graph > Tree Map.**
2. Select **Wt** (weight) and click **Sizes.**
3. Select **Make** and **Model** and click **Categories.**
4. Select **L Leg** and click **Coloring.**
   
   *L Leg represents a measurement of injuries resulting from the deceleration speed of the left leg, where more deceleration causes more injury.*
5. Click **OK.**
   
   The report window appears.

---

**Figure 16.16 Left Leg Deceleration Injuries**

From Figure 16.16, you can see that the Club Wagon and S10 Pickup 4x4 have the largest number of left leg deceleration injuries.

You can examine other safety measurements without re-launching the Tree Map platform, as follows:

1. From the red triangle menu, select **Change Color Column.**
2. Select **Head IC.**
3. Click **OK.**
   
   The tree map updates to reflect head injuries instead of left leg injuries.
From Figure 16.17, you notice the following:

- Although the S10 Pickup 4x4 had a high number of left leg deceleration injuries, it has a lower number of head injuries.
- The Club Wagon still has a high number of head injuries, in addition to the high number of left leg deceleration injuries.
- The Trooper II 4x4 had a low number of left leg deceleration injuries (see Figure 16.16), but it has a high number of head injuries.
Creating Tree Maps
Additional Tree Map Examples
Using the Scatterplot Matrix platform, you can assess the relationships between multiple variables simultaneously. A scatterplot matrix is an ordered collection of bivariate graphs. For further analysis, you can customize the scatterplots with density ellipses for all of your data, or for only groups of your data.

**Figure 17.1 Example of a Scatterplot Matrix**
Contents

Example of a Scatterplot Matrix .............................................................. 437
Launch the Scatterplot Matrix Platform ................................................. 437
  Change the Matrix Format ................................................................. 439
The Scatterplot Matrix Report ............................................................ 440
Scatterplot Matrix Options ................................................................. 441
Example Using a Grouping Variable ..................................................... 442
  Create a Grouping Variable ............................................................... 444
Example of a Scatterplot Matrix

This example shows you how to create a scatterplot matrix.

1. Open the Students.jmp sample data table.
2. Select Graph > Scatterplot Matrix.
3. Select age, sex, height, and weight and click Y, Columns.
4. Click OK.

![Figure 17.2 Example of a Scatterplot Matrix](image)

Launch the Scatterplot Matrix Platform

Launch the Scatterplot Matrix platform by selecting Graph > Scatterplot Matrix.
Creating Scatterplot Matrices
Example of a Scatterplot Matrix

Figure 17.3 The Scatterplot Matrix Launch Window

Table 17.1 Description of the Scatterplot Matrix Launch Window

| Y, Columns and X | • If you assign variables to the Y, Columns role only, they appear on both the horizontal and vertical axes.  
|                  | • If you assign variables to both the Y, Columns and X role, then the Y, Columns variables appear on the vertical axis. The X variables appear on the horizontal axis. This approach enables you to produce rectangular matrices, or matrices that have different, yet overlapping, sets of variables forming the axes of the matrix. |
| Group            | If you assign a variable to the Group role, you can add shaded density ellipses for each level of the Group variable. See “Example Using a Grouping Variable,” p. 442. |
| By               | This option produces a separate scatterplot matrix for each level of the By variable. If two By variables are assigned, a separate graph for each possible combination of the levels of both By variables is produced. |
| Matrix Format    | The Matrix Format can be one of three arrangements: Upper Triangular, Lower Triangular, or Square. See “Change the Matrix Format,” p. 439. |
Chapter 17

Creating Scatterplot Matrices

Example of a Scatterplot Matrix

439

Change the Matrix Format

The Matrix Format can be one of three arrangements: Upper Triangular, Lower Triangular, or Square.

Figure 17.4 Examples of Matrix Formats

Lower Triangular

Upper Triangular

Square
Creating Scatterplot Matrices

The Scatterplot Matrix Report

Follow the instructions in “Example of a Scatterplot Matrix,” p. 437 to produce the plot shown in Figure 17.5.

The Scatterplot Matrix report shows an ordered grouping of bivariate graphs. In each graph, you can examine the relationships between each pair of variables.

**Figure 17.5  Example of a Scatterplot Matrix Report**

![Scatterplot Matrix Report](image)

**Note:** For information about additional options for the report, see “Scatterplot Matrix Options,” p. 441.

In this example, you can see that the graph for weight versus height is different from the graph for sex versus age. If you turn off jitter by clicking on the red triangle menu and selecting Points Jittered, the difference becomes even more pronounced.
Chapter 17

Creating Scatterplot Matrices

Figure 17.6 Example of a Scatterplot Matrix with No Jitter

The weight versus height graph shows continuous data, and the sex versus age graph shows categorical data.

Scatterplot Matrix Options

The following table describes the options within the red triangle menu next to Scatterplot Matrix.

Table 17.2 Descriptions of the Scatterplot Matrix Options

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show Points</td>
<td>Shows or hides the points in the scatterplots.</td>
</tr>
<tr>
<td>Points Jittered</td>
<td>Turns the jittering of the points in the scatterplot on or off.</td>
</tr>
<tr>
<td>Fit Line</td>
<td>Fits a simple regression line and its mean confidence interval to the scatterplots.</td>
</tr>
<tr>
<td>Density Ellipses</td>
<td>Shows or hides the outline of the density ellipses. See “Example Using a Grouping Variable,” p. 442.</td>
</tr>
<tr>
<td>Shaded Ellipses</td>
<td>Colors the area within each ellipse. See “Example Using a Grouping Variable,” p. 442.</td>
</tr>
<tr>
<td>Ellipses Coverage</td>
<td>Enables you to choose an $\alpha$-level for the ellipses to cover.</td>
</tr>
<tr>
<td>Ellipses Transparency</td>
<td>Enables you to choose the transparency of the shaded ellipses, where 0 is completely transparent and 1 is completely opaque.</td>
</tr>
</tbody>
</table>
Example Using a Grouping Variable

This example shows you how to create a scatterplot matrix using a grouping variable.

1. Open the *Iris.jmp* sample data table.
2. Select *Graph > Scatterplot Matrix*.
3. Select *Sepal length*, *Sepal width*, *Petal length*, and *Petal width* and click *Y, Columns*.
4. Select *Species* and click *Group*.
5. Click *OK*.

---

**Table 17.2** Descriptions of the Scatterplot Matrix Options

<table>
<thead>
<tr>
<th><strong>Nonpar Density</strong></th>
<th>Shows or hides the nonparametric density, which represents the areas where the data points are the most dense. The nonparametric density estimation is helpful when you have a lot of points and the density of the points is difficult to see. There are two quantile density contours. One contour includes 50% of the points, and the other contour includes 100% of the points.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Group By</strong></td>
<td>In the Group By window, you can perform the following actions:</td>
</tr>
<tr>
<td></td>
<td>• If you did not select a <strong>Group</strong> variable in the launch window, you can add one now.</td>
</tr>
<tr>
<td></td>
<td>• If you did select a <strong>Group</strong> variable in the launch window:</td>
</tr>
<tr>
<td></td>
<td>– you can remove the existing <strong>Group</strong> variable.</td>
</tr>
<tr>
<td></td>
<td>– you can add additional <strong>Group</strong> variables.</td>
</tr>
<tr>
<td></td>
<td>See “Example Using a Grouping Variable,” p. 442.</td>
</tr>
<tr>
<td><strong>Script</strong></td>
<td>This menu contains options that are available to all platforms. They enable you to redo the analysis or save the JSL commands for the analysis to a window or a file. See “Script Menus,” p. i5 in the “Preliminaries” chapter.</td>
</tr>
</tbody>
</table>
To make the groupings stand out, proceed as follows:

6. From the red triangle menu, select **Density Ellipses**.
7. From the red triangle menu, select **Shaded Ellipses**.
Create a Grouping Variable

If your data does not already have a grouping variable, you can create one using the Cluster platform. Using the Iris.jmp data, assume that the Species column does not exist. You know that the data comes from three species of Iris flowers, so you want to create three clusters within a group.

Proceed as follows:

1. Using the Iris.jmp sample data table, select Analyze > Multivariate Methods > Cluster.
2. Select Sepal length, Sepal width, Petal length, and Petal width and click Y, Columns.
3. Click OK.
4. From the red triangle menu, select Number of Clusters.
5. Type 3 to represent the three different Iris species.
6. Click OK.
7. From the red triangle menu, select Save Clusters.
8. Close the Hierarchical Cluster report window, and go back to the Iris.jmp data table.
   You can see that a Cluster column has been added to the Iris.jmp data table.
9. Perform the Scatterplot Matrix analysis. Follow the instructions in the section “Example Using a Grouping Variable,” p. 442, but use Cluster as the grouping variable.

Figure 17.9 Example of a Scatterplot Matrix Using a Cluster Variable
The **Ternary Plot** command in the **Graph** menu produces a three-axis plot.

Ternary plots are a way of displaying the distribution and variability of three-part compositional data. (For example, the proportion of sand, silt, and clay in soil or the proportion of three chemical agents in a trial drug.) You can use data expressed in proportions or use absolute measures.

The ternary display is a triangle with sides scaled from 0 to 1. Each side represents one of the three components. A point is plotted so that a line drawn perpendicular from the point to each leg of the triangle intersect at the component values of the point.

![Figure 18.1 Examples of Ternary Plots](image)
## Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example of a Ternary Plot</td>
<td>447</td>
</tr>
<tr>
<td>Launch the Ternary Plot Platform</td>
<td>449</td>
</tr>
<tr>
<td>The Ternary Plot Report</td>
<td>450</td>
</tr>
<tr>
<td>Mixtures and Constraints</td>
<td>450</td>
</tr>
<tr>
<td>Ternary Plot Platform Options</td>
<td>451</td>
</tr>
<tr>
<td>Example of Using a Contour Function</td>
<td>452</td>
</tr>
</tbody>
</table>
Example of a Ternary Plot

This example uses the Pogo Jumps.jmp sample data table. The data, adapted from Aitchison (1986), show measurements for pogo jumps of seven finalists in the 1985 Hong Kong Pogo-Jump Championship. A single pogo jump is the total jump distance in three consecutive bounces, referred to as yat, yee, and sam.

1. Open the Pogo Jumps.jmp sample data table.
2. From the Graph menu, select Ternary Plot.
3. Select Yat, Yee, and Sam and click X, Plotting.
4. Click OK.

Figure 18.2 Example of a Ternary Plot

Use the crosshairs tool to determine exact coordinates of points within the plot.
To get a better idea of how the three bounces contribute to total distance, assign each contestant’s points a different color.

1. Right-click on the plot and select **Row Legend**.
2. Select **Finalist** in the column list box.
3. Select **Standard** from the **Markers** menu.
4. Click **OK**.
5. To make the markers easier to see, right-click on the plot and select **Marker Size > 3, Large**.

---

**Figure 18.3 Using the Crosshairs Tool**

---

**Figure 18.4 Pogo Data Colored by Finalist**
Note that most of the finalists are consistent in the composition of total distance. However, two finalists, Jao and Ko, both have one jump that is not consistent with their other jumps. For example, for three of Jao’s jumps the Yat composed about 50% of the total distance. But for the other jump the Yat composed only 30% of the total distance. That jump is not consistent with the others. A similar observation can be made about Ko’s jumps.

**Launch the Ternary Plot Platform**

Launch Ternary Plot by selecting *Graph > Ternary Plot*. The Ternary Plot Launch window contains many controls that are described in “Launch Window Features,” p. 13 in the “Preliminaries” chapter. Controls that are unique to this window are described in Table 18.1.

**Figure 18.5** The Ternary Plot Launch Window

<table>
<thead>
<tr>
<th>Table 18.1 Description of the Ternary Plot Launch Window</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>X, Plotting</strong></td>
</tr>
<tr>
<td><strong>Contour Formula</strong></td>
</tr>
<tr>
<td><strong>By</strong></td>
</tr>
</tbody>
</table>
The Ternary Plot Report

Follow the instructions in “Example of a Ternary Plot,” p. 447 to produce the plot shown in Figure 18.6. Each of the three sides of a ternary plot represents a proportion of 0%, with the point of the triangle opposite that base representing a proportion of 100%. As a proportion increases in any one sample, the point representing that sample moves from the base to the opposite point of the triangle.

Figure 18.6 The Ternary Plot Report

Mixtures and Constraints

Ternary Plot uses the Mixture column property to shade the portion of the graph that is out of bounds. The only constraints that the Ternary plot recognizes are the mixture sum and the mixture bounds. The Ternary plot does not recognize a general linear constraint like the Mixture Profiler does. For information about setting the Mixture column property in the Column Info window, see the Using JMP book.

An Example Using Mixture Constraints

1. Open the Plasticizer.jmp sample data table.
   The p1, p2, and p3 columns all have Mixture Column Properties defined.
2. From the Graph menu, select Ternary Plot.
3. Select p1, p2, and p3 and click X, Plotting.
4. Click OK.
For more information about mixtures, see the section on the Mixture Profiler in the Profilers chapter in the *Modeling and Multivariate Methods* book.

**Ternary Plot Platform Options**

The red triangle menu next to Ternary Plot contains options to modify the plot and add a 3D graph.

**Table 18.2 Descriptions of Ternary Plot Options**

<table>
<thead>
<tr>
<th>Option</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ref Labels</td>
<td>Shows or hides the tick marks and tick labels on the three axes. The default plot shows the labels.</td>
</tr>
<tr>
<td>Ref Lines</td>
<td>Shows or hides the triangular grid lines. The default plot shows the grid lines.</td>
</tr>
<tr>
<td>Contour Fill</td>
<td>Allows filling of contours if a contour formula is specified in the plot. You can select <strong>Lines Only</strong>, <strong>Fill Above</strong>, or <strong>Fill Below</strong>. The default platform shows lines only.</td>
</tr>
<tr>
<td>3D Graph</td>
<td>Available only when four or more X variables are assigned. To see this option on the menu, press the SHIFT key and click on the red triangle menu. Adds a Ternary 3D graph to the report window showing your data points in a three-dimensional scatterplot.</td>
</tr>
<tr>
<td>Show Points</td>
<td>Shows or hides the plotted points. The default plot shows the points.</td>
</tr>
<tr>
<td>Script</td>
<td>This menu contains commands that are available to all platforms. They enable you to redo the analysis or save the JSL commands for the analysis to a window or a file. See “Script Menus,” p. 15 in the “Preliminaries” chapter.</td>
</tr>
</tbody>
</table>
Example of Using a Contour Function

The data in Fish Patty.jmp is adapted from Cornell (1990) and comes from an experiment to optimize the texture of fish patties. The columns Mullet, Sheepshead, and Croaker represent what proportion of the patty came from those fish types. The column Temperature represents the oven temperature used to bake the patties. The column Rating is the response and is a measure of texture acceptability, where higher is better. A response surface model was fit to the data and the prediction formula was stored in the column Predicted Rating. (For more information, see the section on Mixture Profilers in the Modeling and Multivariate Methods book.)

1. Open the Fish Patty.jmp sample data table.
2. From the Graph menu, select Ternary Plot.
3. Select Mullet, Sheepshead, and Croaker and click X, Plotting.
4. Select Predicted Rating and click Contour Formula.
5. Click OK.
6. From the red triangle menu, select Contour Fill > Fill Above.

The manufacturer wants the rating to be at least 5. You can drag the slider for Temperature and see the contours for the Predicted Rating change. Each point represents a mixture of the three fish. Any given mixture of fish types receives different ratings according to the temperature at which the patties are baked.

In this example, the red shaded area shows the mixture of fish that results in a rating of 5 to 5.5. Any purple areas show the mixture of fish that results in a rating of 5.5 and above. At 400 degrees, a mixture of mostly sheepshead and mullet with very little croaker results in a rating of 5 and above.


Hsu, J. (1989), “Multiple Comparison Procedures” ASA Short Course notes, Columbus OH: Ohio State University.


Linnerud: see Rawlings (1988).


Myers, R.H. (1989), *Classical and Modern Regression with Applications*, Boston: PWS-KENT.


References


Umetrics (1995), *Multivariate Analysis (3-day course)*, Winchester, MA.


Western Electric Company (1956), *Statistical Quality Control Handbook*, currently referred to as the *AT&T Statistical Quality Control Handbook*.


**Index**

**Basic Analysis and Graphing**

**Numerics**

<table>
<thead>
<tr>
<th>Numerical Option</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>3D Correspondence Analysis option</td>
<td>201–202</td>
</tr>
<tr>
<td>3D Graph option</td>
<td>451</td>
</tr>
<tr>
<td>3D scatterplots, see Scatterplot 3D platform</td>
<td></td>
</tr>
<tr>
<td>5% Contours option</td>
<td>121</td>
</tr>
<tr>
<td>3D Correspondence Analysis option</td>
<td></td>
</tr>
<tr>
<td>3D Graph option</td>
<td></td>
</tr>
<tr>
<td>3D scatterplots, see Scatterplot 3D platform</td>
<td></td>
</tr>
<tr>
<td>5% Contours option</td>
<td></td>
</tr>
</tbody>
</table>

**A**

<table>
<thead>
<tr>
<th>Analysis Option</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Across Groups report</td>
<td>244</td>
</tr>
<tr>
<td>Add Error Bars to Mean option</td>
<td>311, 316–317, 320</td>
</tr>
<tr>
<td>Add option</td>
<td>265</td>
</tr>
<tr>
<td>Aggregation options</td>
<td>390</td>
</tr>
<tr>
<td>Agreement Statistic option</td>
<td>190, 207–209, 216</td>
</tr>
<tr>
<td>All fits</td>
<td>67–69</td>
</tr>
<tr>
<td>All Graphs option</td>
<td>134</td>
</tr>
<tr>
<td>All Labels option</td>
<td>389</td>
</tr>
<tr>
<td>All Pairs, Tukey HSD test</td>
<td>151, 155–157</td>
</tr>
<tr>
<td>Analysis of Means charts</td>
<td>148–149</td>
</tr>
<tr>
<td>for Proportions option</td>
<td>190, 197–198</td>
</tr>
<tr>
<td>options</td>
<td>149–150</td>
</tr>
<tr>
<td>Analysis of Means Methods</td>
<td>132, 146–150</td>
</tr>
<tr>
<td>Analysis of Variance</td>
<td></td>
</tr>
<tr>
<td>See also Oneway</td>
<td></td>
</tr>
<tr>
<td>report</td>
<td>100–101, 142–143</td>
</tr>
<tr>
<td>ANOM</td>
<td>147</td>
</tr>
<tr>
<td>See also Analysis of Means</td>
<td></td>
</tr>
<tr>
<td>for Variances</td>
<td>147–148</td>
</tr>
<tr>
<td>for Variances with Levene (ADM)</td>
<td>147</td>
</tr>
<tr>
<td>with Transformed Ranks</td>
<td>147</td>
</tr>
<tr>
<td>Arrange Plots option</td>
<td>339, 412</td>
</tr>
<tr>
<td>automatic recalc option</td>
<td>17</td>
</tr>
<tr>
<td>Axes option</td>
<td>366</td>
</tr>
<tr>
<td>Axis Options window</td>
<td>353</td>
</tr>
</tbody>
</table>

**B**

| Background Color option                                | 365     |
| Bar Chart option                                       | 321     |
| Bar option                                             | 265     |
| Bar Style option                                       | 265     |
| Bartlett test                                          | 167     |
| Beta Binomial fit                                      | 70–71   |
| Beta fit                                               | 64      |
| Binomial fit                                           | 69      |
| Biplot Rays option                                     | 356     |
| Bivariate Normal Ellipse menu                          | 118–121 |
| Bivariate platform                                     | 87      |
| example                                               | 89      |
| launching                                             | 89      |
| options                                               | 91–93, 118–122 |
| report window                                          | 90      |
| Block button                                           | 86      |
| Block Means report                                     | 144     |
| Box option                                             | 366     |
| Box plots                                              | 134, 265 |
| Box Style option                                       | 265     |
| Brown-Forsythe test                                    | 167     |
| brush tool                                             | 389     |
| Bubble Plot platform                                   | 377     |
| animating dynamic                                      | 383     |
| brush tool                                             | 389     |
| By variable                                            | 382     |
| categorical Y variable                                 | 396–397 |
| dynamic                                               | 377, 379–380, 382 |
| example                                               | 379–380, 392–397 |
| launching                                             | 380–382 |
| options                                               | 389–392 |
| selecting bubbles                                      | 388     |
| splitting a bubble                                     | 386–388 |
| static                                                | 377, 382, 392–396 |

**C**

| Capability Analysis option                            | 57–60, 79–82 |
| Categories variable                                    | 419–421     |
| Categories, X, Levels button                          | 310, 313–314 |
| CDF plot                                               | 47–48, 133, 177–178 |
| Cell Labeling option                                   | 192     |
| Cell Plot platform                                     | 399     |
| examples                                              | 409, 413–414 |
| launching                                             | 410     |
options 411–413
  report window 410–411
Center at zero option 403, 410
Change Contours options 372–375
Change to option 265
Chart Orientation options 311
Chart platform 305
  coloring bars 318–319
  examples 307–309, 321–331
  launching 309–317
  legends 317
  options 319–321
  Ordering 318
  report window 317–319
Chart Type options 311
Circle Size slider 383
Cochran Armitage Trend Test 191, 214–215
Cochran Mantel Haenszel test 190, 206–207
Color by Density Quantile option 121
Color option 267
Color Theme option 193, 372, 412
Color zone 263
Coloring variable 382, 419, 424–426
Column Bandwidth option 362
Combine All option 390
Combine button 384
Compare Densities option 178
Compare Means options 132, 150–162
  comparison circles 135, 151–153, 182–183
Composition of Densities option 178
Confid Curves Fit option 119
Confid Curves Indiv option 119
Confid Quantile option 162
Confid Shaded Fit option 120
Confid Shaded Indiv option 120
Confidence Interval options 39, 51
Confidence Limits option 75
Connect Color option 342
Connect Means option 135
Connect Points option 311, 321, 342, 355
Connect Thru Missing option 339
Connecting Letters Report 162
Contingency platform 185
  example 187–188
  launching 188
  options 189–191
  report window 188
Contingency Table 190, 193–195
Continuous Fit options 61–69
Contour Fill option 121, 451
Contour Formula button 449
Contour Lines option 121
Contour option 265
Contour Plot platform 367
  example 369
  launching 369–371
  options 371–376
  report window 371
  using formulas 375–376
Contour Quantile option 362
Contour Values buttons 370
Correlation report 113
Correspondence Analysis option 190, 199–205
Count Axis option 35, 41
Covariance of Estimates report 228

D
data tables 3, 8
Densities options 133, 178–179
Density Axis option 35, 42
Density Contour Controls options 361–362
Density Curve option 71
Density Ellipse option 92, 112–114, 441
density estimation fits 93
detailed Comparisons Report 162
details report 202
diagnostic Plot option 71, 73–75
difference Matrix option 162
disclosure buttons 3
discrete Fit options 69–71
display Options 134–135, 150, 198
  for categorical variables 34
  for continuous variables 39–40
displayBox Scripting index 7
distribution platform 19
categorical variables in 21
  continuous variables in 21
  example 21–23
  launching 23
  options 34–53
  report window 24–26
documentation
  basic Analysis and Graphing book 4
  Design of Experiments book 6
  discovering JMP book 4
  Modeling and Multivariate Methods book 5
  Quality and Reliability Methods book 6
  Scripting Guide book 6
  using JMP book 4
Index

Drop Line options 355
Dunn All Pairs for Joint Ranks test 164
Dunn With Control for Joint Ranks test 164
dynamic bubble plots 377, 382

E
Each Pair, Student’s t test 150, 153–155
Ellipses Coverage option 441
Ellipses Transparency option 441
Ellipsoid Coverage option 355
Ellipsoid Transparency option 355
Equal Variances option 109
Equivalence Test option 133, 172–173
Error Bars option 265
Exact Test 191, 215
See also Fisher’s Exact Test
Exponential fit 63
Extreme Value fit 63

F
Fill Areas option 370, 372–373
Filled option 389
Fisher’s Exact Test 197
Fit Each Value
command 92, 108–109
menu 118–120
report 108
Fit Line command 92, 95–103, 122, 441
Fit Mean
command 92–94
menu 118–120
report 94
Fit Orthogonal command 92, 109–112, 123
Fit Polynomial command 92, 95–103
Fit Special command 92, 103–106
Fit Spline command 92, 106–108, 123
Fit X to Y option 110
Fit Y by X platform 83
Fix Parameters option 71
font conventions 8
Freq zone 263
Frequencies
option 34
report 29–30
Function Plot option 342

G
Gamma fit 63
Gamma Poisson fit 69–70
GLog fit 66–67
Go button 383
Goodness of Fit
tests 75–76
Gradient Between Ends option 193
Gradient Between Selected Points option 193
Grand Mean option 135
Graph Builder 253, 255
adding multiple variables 274–285
adding variables 268–271
buttons 264
changing the legend 285
examples 255–261, 289–304
launching 261
maps 286, 289
moving grouping labels 270
moving variables 272
options 264–268
removing variables 272–274
right-click menus 264–268
zones 262–264
Graph Builder Customize 266
Graph in Variance Scale option 150
Graph Type option 412
Grids option 366
Group button 438, 442
Group By option 91, 116–118, 442
Group X zone 263
Group Y zone 263
Grouping button 311, 314, 337

H
help
buttons 6
menu 3
tool 6
Hide Lights Border option 364
Histogram Options 41
histograms 26–29
borders option 91, 116
color option 35, 41
creating subsets 26
highlighting data 26–27
option in Graph Builder 265
option in Oneway 135
Index

options for categorical variables 35
options for continuous variables 40–42
rescaling 26
resizing 26–27
selecting data 28
specifying data 26
Horizontal Layout option 34, 40
Horizontal option 266, 319

I
ID variable 381, 386–388
Inverse Prediction option 228
Iterations report 225

J
jitter 234, 266
JMP
  Menu Card 6
  Quick Reference Card 6
tutorials 7
user community online 8
JMP.com 8
Johnson fits 65–66
JSL Functions index 8

K
Kernel Control option 121
Kolmogorov Smirnov Test 163
Kruskal-Wallis test, see Wilcoxon Test

L
Label Contours option 372
Label Format option 321
Label Options for charts 320
Lack of Fit report 97–99
launch windows 3, 11
Left Scale/Right Scale button 337
Left Y Log Scale option 337
Legend 263, 389, 412
Level Options 320
Levene test 167
Lift Curve option 229
Line Chart option 321
Line Color option 119
Line of Fit option 265
Line option 265
Line Style option 119, 342
Line Width option 119, 342
Line Width slider 366
Linear Fit
  menu 118–120
  report 96–103
Lock Scales option 264
Logistic platform 219
  See also logistic regression
  Covariance of Estimates report 228
  examples 221–223, 233–237
  Iterations report 225
  launching 223
  logistic plot 225, 228
  options 228–233
  Parameter Estimates report 227
  report window 223–228
  Whole Model Test report 225–227
logistic regression 219, 221
LogNormal fit 62
LSD Threshold Matrix option 162

M
Macros option 193
Mann-Whitney test, see Wilcoxon Test
Map Shapes option 265
maps
  create custom 288
Marker Quality slider 366
Marker Size slider 366
Marker Transparency slider 366
Matched Pairs platform 239, 241
  Across Groups report 244
  examples 241–242, 246–247
  launching 242
  multiple Y columns 243
  options 245
  report window 243–245
  statistical details 247
Tukey mean-difference plot 243–244
Matching Column option 134, 180–181
Matching Dotted Lines option 135
Matching Lines option 135, 180
Matrix Format option 438–439
Mean CI Lines option 135
Mean Diamonds option 135, 144–145
Mean Error Bars option 135, 145–146
Mean Lines option 135, 145–146
Mean of Means option 135
Means and Std Dev option 132
Means for Oneway Anova report 143–144
Means/Anova option 132, 138–146
Means/Anova/Pooled t option 132, 138–146
Measures of Association option 191, 212–214
Median Reference Line option 75
Median Test 163
Mesh Plot option 121
Mixture column property 450
Model Clustering option 121
modeling types 9–10
Moments
  option 40
  report 32–33
More Moments option 40
Mosaic in Graph Builder 265
Mosaic Plot
  in Contingency 189, 191–193
  option in Distribution 35–36
Move Backward option 266
Move Forward option 266

N
Needle Chart option 321
Needle option 342
No Labels option 389
No Overlay option 339
No Separator Lines option 412
Nonpar Density command 92, 114–115, 442
Nonpar Density Contour option 355
Nonparametric
  options 132
  tests 162–167
Nonparametric Bivariate Density report 115
Nonparametric Density Contours option 359–364
Nonparametric Multiple Comparisons tests 164
Normal Contour Ellipsoids option 355, 357–359
Normal fit 62
Normal Mixtures fits 64
Normal Quantile Plot 42–43, 133, 176–177
Number of Levels option 266–267, 270

O
O’Brien test 167
Object Scripting index 7
Odds Ratios option 190, 228
One-way platform 125, 127
  example 127–128
  launching 129
  options 130–135
  report window 129
Ordered Differences Report 162
Ordering variable 419, 422–424
Orthogonal Fit Ratio menu 118–120
Orthogonal Regression report 112
Orthographic option 366
Outlier Box Plot 43–45
Outliers option 266
Overlay Color option 321
Overlay Groups option 339
Overlay Marker Color option 342
Overlay Marker option 321, 342
Overlay option in Chart 311, 319
Overlay Plot platform 333
  example 335–336, 342–346
  launching 336–337
  options 338–342
  report window 337
Overlay Y’s option 339
Overlay zone 263

P
paired t-test 239, 241
Parallel Plot platform 399
  examples 401–402, 406–408
  launching 402
  options 405
  report window 403–405
Parameter Estimates report 101–103, 227
Pen Style option 321
Percent for quantiles option 312
Perspective slider 366
Pie Chart option 319
platforms 9–12
  common features 13
  description of 3
  design 9
  launch windows 3, 13
  modeling types 9
  multiple 8
  report windows 3
Plot Actual by Quantile option 176
Plot Dif by Mean option 245
Plot Dif by Row option 245
Plot Options menu in Logistic 228
Plot Quantile by Actual option 176
Plot Residuals option 120
Point Chart option 321
Points Jittered option 135, 440–441
Points option 134, 265
Points Spread option 135
Poisson fit 69
Polynomial Fit
  report 96–103
Polynomial Fit Degree menu 118–120
Power option 133, 173–176, 183–184
Prediction Interval option 54–55, 78
Prev button 383
Principal Components option 355
Prob Axis option 35, 41
Proportion of Densities option 178
Q
Quantile Box Plot 45–46
Quantile Density Contours menu 118–122
Quantiles
  option 40, 71, 77, 131, 136–137
  report 31
question mark (?) tool, see help tool
R
Range Chart option 319
Range Plot option 339
red triangle menus 3
Ref Labels option 451
Ref Lines option 451
Reference Frame option 246
regression fits for Bivariate 93
Relative Risk option 190, 209–210
Remove command 266–267
Remove Fit option 72, 120
Remove Prin Comp option 356
Report option 119
report windows 3, 11
Reset option 364–365
Resolution option 362
Reverse Colors option 372
Revert to Old Colors option 193
Right Y Log Scale option 337
ROC Curve option 229
Rotate option 75
Rotated Components option 356
Row order option 266
row profile 199
Rows option 365
S
Sampling option 264
Save Coefficients option 121
Save Colors to Column option 193
Save commands in Distribution 39, 52–53
Save Density Formula command 72
Save Density Grid option 122
Save Density Quantile option 121
Save Fitted Quantiles command 72
Save for Adobe Flash platform (.SWF) option 34, 390
Save options
  in Contour Plot 372, 375
  in Oneway 134
Save Predicteds option 120
Save Prediction Formula option 121
Save Prin Components option 356
Save Probability Formula option 229
Save Residuals option 120
Save Rotated Components option 357
Save Spec Limits command 72
Save Value Ordering option 201
Saved Transformed command 72
Scale Uniformly option 403, 410
Scatterplot 3D platform 347
  adjusting axes 353
  assigning markers 354
  changing variables 352
  coloring points 354
  example 349
  launching 350
  options 355, 364–366
  report window 351–354
  Settings window 364–365
  spinning the plot 352
Scatterplot Matrix platform 435
  examples 437, 442–444
  launching 437
  options 441
  report window 440–441
Script All By-Groups menu 16
Script menu 15–17
Select Colors for Values window 192
Select Points by Density option 121
Select Points Inside option 121
Select Points Outside option 121
Selectable Across Gaps option 389
selecting
  analyses 11
  bubbles 388
Separate Axes option 339–340
Separate Bars option 35
Set Alpha Level option
  for Bivariate 120
  for Contingency 190, 198
  for Matched Pairs 246
  for One-way 133, 149
Set Bin Width option 41
Set Colors option 192
Set Spec Limits for K Sigma option 72
Shaded Contour option 121
Shaded Ellipses option 441
Shadowgram option 41
Shape zone 263
shapefiles 288
Show ArcBall option 365
Show Boundary option 372
Show Center Line option 150
Show Contours option 372
Show Controls option 355
Show Counts option 35, 42
Show Data Points option 372
Show Decision Limit Shading option 150
Show Decision Limits option 150
Show Error Bars option 321
Show Legend option 264
Show Level Legend option 320
Show Missing Data Points option 372
Show Missing Shapes option 266
Show Needles option 150
Show Percents option 35, 42
Show Points option
  for Bivariate 91
  for Chart 311, 321
  for Overlay Plot 342
  for Scatterplot 3D 355
  for Scatterplot Matrix 441
  for Ternary Plot 451
Show Reversing Checkboxes option 405
Show Roles option 389–392
Show Separate Axes option 320
Show Summary Report option 149, 198
Show Y Legend option 320
Sign Test 246
Sizes variable 382, 419, 421–422
Smooth Curve fit 65
Smooother option 265, 267
Smoothing Spline Fit
  menu 118–121
  report 107
Sort Ascending option 412
Sort Descending option 412
Sort X option 337
Spec Limits option 72, 76–77
Specified Variance Ratio option 110
Specify Grid button 371, 375
Specify Transformation or Constraint window 103–105
Speed slider 383
Split All option 390
Split button 383, 387
Stack Bars option 320
Stack option 34
static bubble plots 377, 382, 392–396
Statistics button 310, 312–313
Statistics index 7
Std Dev Lines option 135, 145–146
Std Error Bars option 35, 41
Std Prin Components option 356
Steel With Control test 164
Steel-Dwass All Pairs test 164
Stem and Leaf plot 46–47
Step button 383
Step option 342
Stop button 383
Student’s t test 150, 153–155
Subset option 26
Summary of Fit report 96–97, 139
Summary Statistic option 266
Swap command 267, 271
Switch Response Level for Proportion option 198

T

t test option 132
  report 140–142
Ternary Plot platform 445
  examples 447–449, 452
  launching 449
  options 451
  report window 450–451
Test Mean option 48–49
Test Probabilities option 36–38
Test Std Dev option 50–51
Tests
   option in Contingency 190
   report 195–196
Text Size slider 366
Thick Connecting Line option 320
Time variable 382–386
Tip of the Day window 8
Tolerance Interval option 55–57, 78–79
Trail Bubbles option 389
Trail Lines option 389
Transformed Fit menu 118–120
   report 106
Tree Map platform 415
   By variable 419
   Categories variable 419–421
   Coloring variable 419, 424–426
   examples 417–418, 428–433
   launching 419–426
   options 427–428
   Ordering variable 419, 422–424
   report window 426–427
   Sizes variable 419, 421–422
Tukey mean-difference plot 243–244
   tutorials 7
Two Sample Test for Proportions 191, 211–212

U
Unequal Variances 132, 167–171
Ungroup Charts option 320
Ungroup Plots option 339
Uniform Scaling option 34
Uniform Y Scale option 339
Univariate Variances, Prin Comp option 109
Use Hardware Acceleration option 365
Use Table Data option 371

V
van der Waerden Test 161
Vertical option 35, 41, 266, 319

W–Z
Wall Color option 364
Walls option 366
Weibull fits 62
Welch’s test 169, 171
Whole Model Test report 225–227
Wilcoxon Each Pair test 164
Wilcoxon Signed Rank test 246
Wilcoxon Test 162
With Best, Hsu MCB test 151, 157–158
With Control, Dunnett’s test 151, 159–160
Wrap zone 263
X Axis proportional option 135, 144–145
X Group Edge option 267
X Log Scale option 337
X option in Graph Builder 266
X zone 262
X, Continuous Regressor button 86
X, Grouping button 86
X, Grouping Category button 86
X, Plotting button 449
X, Regressor button 86
Y Group Edge option 267
Y option in Graph Builder 266
Y Options
   in Chart 320–321
   in Overlay Plot 341–342
Y zone 262
Y, Categorical Response button 86
Y, Response Category button 86
Zoom slider 366