Discovering JMP

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

Marcel Proust

**Discovering JMP 13®**

Copyright © 2016, SAS Institute Inc., Cary, NC, USA

ISBN 978-1-62960-570-8 (EPUB)

All rights reserved. Produced in the United States of America.

**For a hard-copy book**: No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

**For a web download or e-book**: Your use of this publication shall be governed by the terms established by the vendor at the time you acquire this publication.

The scanning, uploading, and distribution of this book via the Internet or any other means without the permission of the publisher is illegal and punishable by law. Please purchase only authorized electronic editions and do not participate in or encourage electronic piracy of copyrighted materials. Your support of others’ rights is appreciated.

**U.S. Government License Rights; Restricted Rights**: The Software and its documentation is commercial computer software developed at private expense and is provided with RESTRICTED RIGHTS to the United States Government. Use, duplication or disclosure of the Software by the United States Government is subject to the license terms of this Agreement pursuant to, as applicable, FAR 12.212, DFAR 227.7202-1(a), DFAR 227.7202-3(a) and DFAR 227.7202-4 and, to the extent required under U.S. federal law, the minimum restricted rights as set out in FAR 52.227-19 (DEC 2007). If FAR 52.227-19 is applicable, this provision serves as notice under clause (c) thereof and no other notice is required to be affixed to the Software or documentation. The Government’s rights in Software and documentation shall be only those set forth in this Agreement.


September 2016

SAS® and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.
SAS software may be provided with certain third-party software, including but not limited to open-source software, which is licensed under its applicable third-party software license agreement. For license information about third-party software distributed with SAS software, refer to http://support.sas.com/thirdpartylicenses.

**Technology License Notices**

- **Scintilla** - Copyright © 1998-2014 by Neil Hodgson <neilh@scintilla.org>. All Rights Reserved.

  Permission to use, copy, modify, and distribute this software and its documentation for any purpose and without fee is hereby granted, provided that the above copyright notice appear in all copies and that both that copyright notice and this permission notice appear in supporting documentation.

  **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.**

- **Telerik RadControls**: Copyright © 2002-2012, Telerik. Usage of the included Telerik RadControls outside of JMP is not permitted.


- **Made with Natural Earth. Free vector and raster map data @ naturalearthdata.com.**

- **Packages** - Copyright © 2009-2010, Stéphane Sudre (s.sudre.free.fr). All rights reserved.

  Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

  Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

  Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

  Neither the name of the WhiteBox nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

  **THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED**
WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE
DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR
ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES
(INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS
OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY
THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING
NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE,
EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

• iODBC software - Copyright © 1995-2006, OpenLink Software Inc and Ke Jin
  (www.iodbc.org). All rights reserved.
  Redistribution and use in source and binary forms, with or without modification, are
  permitted provided that the following conditions are met:
  – Redistributions of source code must retain the above copyright notice, this list of
    conditions and the following disclaimer.
  – Redistributions in binary form must reproduce the above copyright notice, this list
    of conditions and the following disclaimer in the documentation and/or other
    materials provided with the distribution.
  – Neither the name of OpenLink Software Inc. nor the names of its contributors may
    be used to endorse or promote products derived from this software without specific
    prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS “AS IS” AND
ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED
WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE
DISCLAIMED. IN NO EVENT SHALL OPENLINK OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT,
INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING,
BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE,
DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF
LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR
OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF
THE POSSIBILITY OF SUCH DAMAGE.

• bzip2, the associated library “libbzip2“, and all documentation, are Copyright ©
  1996-2010, Julian R Seward. All rights reserved.
  Redistribution and use in source and binary forms, with or without modification, are
  permitted provided that the following conditions are met:
Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

The origin of this software must not be misrepresented; you must not claim that you wrote the original software. If you use this software in a product, an acknowledgment in the product documentation would be appreciated but is not required.

Altered source versions must be plainly marked as such, and must not be misrepresented as being the original software.

The name of the author may not be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE AUTHOR “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE AUTHOR BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See www.mathworks.com/patents. MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See www.mathworks.com/trademarks for a list of additional trademarks. Other product or brand names may be trademarks or registered trademarks of their respective holders.
- libopc is Copyright © 2011, Florian Reuter. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

- Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

- Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and / or other materials provided with the distribution.
Neither the name of Florian Reuter nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

This software is provided by the copyright holders and contributors “as is” and any express or implied warranties, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose are disclaimed. In no event shall the copyright owner or contributors be liable for any direct, indirect, incidental, special, exemplary, or consequential damages (including, but not limited to, procurement of substitute goods or services; loss of use, data, or profits; or business interruption) however caused and on any theory of liability, whether in contract, strict liability, or tort (including negligence or otherwise) arising in any way out of the use of this software, even if advised of the possibility of such damage.

- libxml2 - Except where otherwise noted in the source code (e.g. the files hash.c, list.c and the trio files, which are covered by a similar licence but with different Copyright notices) all the files are:

  Copyright © 1998 - 2003 Daniel Veillard. All Rights Reserved.

  Permission is hereby granted, free of charge, to any person obtaining a copy of this software and associated documentation files (the “Software”), to deal in the Software without restriction, including without limitation the rights to use, copy, modify, merge, publish, distribute, sublicense, and/or sell copies of the Software, and to permit persons to whom the Software is furnished to do so, subject to the following conditions:

  The above copyright notice and this permission notice shall be included in all copies or substantial portions of the Software.

  The software is provided “as is”, without warranty of any kind, express or implied, including but not limited to the warranties of merchantability, fitness for a particular purpose and noninfringement. In no event shall Daniel Veillard be liable for any claim, damages or other liability, whether in an action of contract, tort or otherwise, arising from, out of or in connection with the software or the use or other dealings in the software.

  Except as contained in this notice, the name of Daniel Veillard shall not be used in advertising or otherwise to promote the sale, use or other dealings in this Software without prior written authorization from him.

- Regarding the decompression algorithm used for UNIX files:

The Regents of the University of California. All rights reserved.

THIS SOFTWARE IS PROVIDED BY THE REGENTS AND CONTRIBUTORS “AS IS” AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE REGENTS OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE.

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the University nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

• Snowball - Copyright © 2001, Dr Martin Porter, Copyright © 2002, Richard Boulton. All rights reserved.

Redistribution and use in source and binary forms, with or without modification, are permitted provided that the following conditions are met:

1. Redistributions of source code must retain the above copyright notice, this list of conditions and the following disclaimer.

2. Redistributions in binary form must reproduce the above copyright notice, this list of conditions and the following disclaimer in the documentation and/or other materials provided with the distribution.

3. Neither the name of the copyright holder nor the names of its contributors may be used to endorse or promote products derived from this software without specific prior written permission.

THIS SOFTWARE IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE
DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT HOLDER OR
CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL,
SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT
LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF
USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED
AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT
LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN
ANY WAY OUT OF THE USE OF THIS SOFTWARE, EVEN IF ADVISED OF THE
POSSIBILITY OF SUCH DAMAGE.

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 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
• details on registering for JMP training
• 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/
**Contents**

**Discovering JMP**

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Learn about JMP</td>
<td>13</td>
</tr>
<tr>
<td></td>
<td>About This Book</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Gallery of JMP Graphs</td>
<td>15</td>
</tr>
<tr>
<td></td>
<td>Formatting Conventions</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>JMP Documentation</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>JMP Documentation Library</td>
<td>37</td>
</tr>
<tr>
<td></td>
<td>JMP Help</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Additional Resources for Learning JMP</td>
<td>43</td>
</tr>
<tr>
<td></td>
<td>Tutorials</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Sample Data Tables</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Learn about Statistical and JSL Terms</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Learn JMP Tips and Tricks</td>
<td>44</td>
</tr>
<tr>
<td></td>
<td>Tooltips</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>JMP User Community</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>JMPer Cable</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>JMP Books by Users</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>The JMP Starter Window</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>Technical Support</td>
<td>46</td>
</tr>
<tr>
<td>2</td>
<td>Introducing JMP</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>Basic Concepts</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>Concepts That You Should Know</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>How Do I Get Started?</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>Starting JMP</td>
<td>49</td>
</tr>
<tr>
<td></td>
<td>Using Sample Data</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>Understanding Data Tables</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>Understanding the JMP Workflow</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>Step 1: Launching a Platform and Viewing Results</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>Step 2: Removing the Box Plot</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Step 3: Requesting Additional Output</td>
<td>56</td>
</tr>
<tr>
<td></td>
<td>Step 4: Interacting with Platform Results</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>How is JMP Different from Excel?</td>
<td>58</td>
</tr>
</tbody>
</table>
## 3 Work with Your Data

### Prepare Your Data for Graphing and Analyzing

- Get Your Data into JMP .......................................................... 61
  - Copy and Paste Data .......................................................... 62
  - Import Data ........................................................................... 62
  - Enter Data ............................................................................. 65
  - Transfer Data from Excel ....................................................... 67
- Work with Data Tables ............................................................. 68
  - Edit Data .............................................................................. 68
  - Select, Deselect, and Find Values ......................................... 70
  - View or Change Column Information .................................... 73
  - Calculate Values with Formulas ........................................... 75
  - Filter Data ............................................................................. 77
- Manage Data ............................................................................. 78
  - View Summary Statistics ....................................................... 79
  - Create Subsets ..................................................................... 83
  - Join Data Tables .................................................................... 85
  - Sort Tables ............................................................................ 87

## 4 Visualize Your Data

### Common Graphs

- Analyze Single Variables ....................................................... 90
  - Histograms .......................................................................... 90
  - Bar Charts .......................................................................... 92
- Compare Multiple Variables .................................................... 94
  - Scatterplots ......................................................................... 95
  - Scatterplot Matrix ............................................................... 99
  - Side-by-Side Box Plots .......................................................... 102
  - Overlay Plots ....................................................................... 105
  - Variability Chart ................................................................ 108
  - Graph Builder ..................................................................... 111
  - Bubble Plots ....................................................................... 117

## 5 Analyze Your Data

### Distributions, Relationships, and Models

- About This Chapter ................................................................. 124
- The Importance of Graphing Your Data .................................. 124
- Understand Modeling Types .................................................... 126
  - Example of Viewing Modeling Type Results ....................... 127
  - Change the Modeling Type .................................................... 128
- Analyze Distributions .............................................................. 130
  - Distributions of Continuous Variables ................................. 130
  - Distributions of Categorical Variables ................................. 133
Analyze Relationships ................................................................. 136
Use Regression with One Predictor ............................................. 136
Compare Averages for One Variable ............................................. 141
Compare Proportions ................................................................. 144
Compare Averages for Multiple Variables ................................. 146
Use Regression with Multiple Predictors ..................................... 150

6 The Big Picture
Exploring Data in Multiple Platforms ........................................ 157
Before You Begin ........................................................................ 158
Explore Data in Multiple Platforms ............................................. 158
Analyze Distributions ................................................................. 158
Analyze Patterns and Relationships ............................................. 162
Analyze Similar Values ............................................................... 166

7 Save and Share Your Work
Save and Re-create Results ............................................................ 171
Save Platform Results in Journals ............................................... 172
Example of Creating a Journal ..................................................... 172
Add Analyses to a Journal .......................................................... 173
Create Projects .......................................................................... 173
Save and Run Scripts .................................................................. 174
Example of Saving and Running a Script ...................................... 175
About Scripts and JSL ................................................................. 176
Save Reports as Interactive HTML .............................................. 176
Interactive HTML Contains Data ............................................... 177
Example of Creating Interactive HTML ................................. 177
Save a Report as a PowerPoint Presentation ................................... 178
Create Adobe Flash Versions of the Profiler, Bubble Plot, or Distribution Platforms ........................................ 179
Example of Saving an Adobe Flash Version of a Bubble Plot .... 179
Create Dashboards ..................................................................... 181
Example of Combining Windows ............................................... 181
Example of Creating a Dashboard with Two Reports ................... 183
Save the Dashboard as an Add-In ............................................... 184

8 Special Features
Automatic Analysis Updates and SAS Integration ......................... 187
Automatically Update Analyses and Graphs ................................ 188
Example of Using Automatic Recalc ............................................ 188
Change Preferences .................................................................... 191
Example of Changing Preferences .............................................. 192
Integrate JMP and SAS ............................................................... 194
Example of Creating SAS Code .................................................. 195
Example of Submitting SAS Code ............................................... 195
Discovering JMP provides a general introduction to the JMP software. This guide assumes that you have no knowledge of JMP. Whether you are an analyst, researcher, student, professor, or statistician, this guide gives you a general overview of JMP’s user interface and features.

This guide introduces you to the following information:

- Starting JMP
- The structure of a JMP window
- Preparing and manipulating data
- Using interactive graphs to learn from your data
- Performing simple analyses to augment graphs
- Customizing JMP and special features

This guide contains six chapters. Each chapter contains examples that reinforce the concepts presented in the chapter. All of the statistical concepts are at an introductory level. The sample data used in this book are included with the software. Here is a description of each chapter:

- **Chapter 2, “Introducing JMP”** provides an overview of the JMP application. You learn how content is organized and how to navigate the software.
- **Chapter 3, “Work with Your Data”** describes how to import data from a variety of sources, and prepare it for analysis. There is also an overview of data manipulation tools.
- **Chapter 4, “Visualize Your Data”** describes graphs and charts that you can use to visualize and understand your data. The examples range from simple analyses involving a single variable, to multiple-variable graphs that enable you to see relationships between many variables.
- **Chapter 5, “Analyze Your Data”** describes many commonly used analysis techniques. These techniques range from simple techniques that do not require the use of statistical methods, to advanced techniques, where knowledge of statistics is useful.
- **Chapter 6, “The Big Picture”** shows you how to analyze distributions, patterns, and similar values in several platforms.
- **Chapter 7, “Save and Share Your Work”** describes sharing your work with non JMP users in PowerPoint presentations, interactive HTML, and Adobe Flash. Saving analyses as scripts and saving work in journals and projects for JMP users are also covered.
• Chapter 8, “Special Features” describes how to automatically update graphs and analyses as data changes, how to use preferences to customize your reports, and how JMP interacts with SAS.

After reviewing this guide, you will be comfortable navigating and working with your data in JMP.

JMP is available for both Windows and Macintosh operating systems. However, the material in this guide is based on a Windows operating system.
Here are pictures of many of the graphs that you can create with JMP. Each picture is labeled with the platform used to create it. For more information about the platforms and these and other graphs, see the documentation on the Help > Books menu.

**Histogram**
Analyze > Distribution

**Bivariate**
Analyze > Fit Y by X
Discovering JMP

Oneway
Analyze > Fit Y by X

Mosaic Plot
Analyze > Fit Y by X

Logistic
Analyze > Fit Y by X

Oneway t Test
Analyze > Fit Y by X
Matched Pairs
Analyze > Specialized Modeling > Matched Pairs

LS Means Plot
Analyze > Fit Model

Partition
Analyze > Predictive Modeling > Partition

Leverage Plot
Analyze > Fit Model

MANOVA
Analyze > Fit Model

Neural Diagram
Analyze > Predictive Modeling > Neural
Actual by Predicted
Analyze > Predictive Modeling > Model Comparison

Nonlinear Fit
Analyze > Specialized Modeling > Nonlinear

Surface Profiler
Analyze > Specialized Modeling > Gaussian Process

Time Series
Analyze > Specialized Modeling > Time Series
Screening
Analyze > Specialized Modeling > Specialized DOE Models > Fit Two Level Screening

FDR pValue Plot
Analyze > Screening > Response Screening

Scatterplot Matrix
Analyze > Multivariate Methods > Multivariate

Dendrogram
Analyze > Clustering > Hierarchical Cluster
Self Organizing Map
Analyze > Clustering > K Means Cluster

Principal Components
Analyze > Multivariate Methods > Principal Components

Canonical Plot
Analyze > Multivariate Methods > Discriminant

Loadings Plot
Analyze > Multivariate Methods > Partial Least Squares
Discovering JMP

XBar and R Charts
Analyze > Quality and Process > Control Chart Builder

Average Chart
Analyze > Quality and Process > Measurement Systems Analysis

Variability Chart
Analyze > Quality and Process > Variability/Attribute Chart

Goal Plot
Analyze > Quality and Process > Process Capability
Individual Measurement Chart
Moving Range Chart
Analyze > Quality and Process > Control Chart > IR

XBar Chart
Analyze > Quality and Process > Control Chart > XBar

Pareto Plot
Analyze > Quality and Process > Pareto Plot

Ishikawa Chart
Fishbone Chart
Analyze > Quality and Process > Diagram
Compare Distributions
Analyze > Reliability and Survival > Life Distribution

Nonparametric Overlay
Analyze > Reliability and Survival > Fit Life by X

Scatterplot
Analyze > Reliability and Survival > Fit Life by X

MCF Plot
Analyze > Reliability and Survival > Recurrence Analysis
Overlay
Analyze > Reliability and Survival > Degradation

Prediction Interval
Analyze > Reliability and Survival > Destructive Degradation

Existing and Future Risk Set
Analyze > Reliability and Survival > Reliability Growth

Forecast
Analyze > Reliability and Survival > Reliability Forecast
Reliability Block Diagram
Analyze > Reliability and Survival > Reliability Block Diagram

Failure Plot
Analyze > Reliability and Survival > Survival

Survival Quantiles
Analyze > Reliability and Survival > Fit Parametric Survival

Baseline Survival
Analyze > Reliability and Survival > Fit Proportional Hazards
Mixture Profiler
Analyze > Consumer Research > Categorical

Factor Loading Plot
Analyze > Consumer Research > Factor Analysis

Prediction Profile
Analyze > Consumer Research > Choice

Characteristic Curves
Analyze > Consumer Research > Item Analysis
Multiple Correspondence Analysis
Analyze > Consumer Research > Multiple Correspondence Analysis

Uplift Model
Analyze > Consumer Research > Uplift

Dual Plot
Analyze > Consumer Research > Item Analysis

Line Graphs
Graph > Graph Builder
Box Plots
Graph > Graph Builder

Pie Chart
Graph > Chart

Stacked Bar Chart
Graph > Chart

Needle and Line Chart
Graph > Overlay Plot
Dot and Line Chart
Graph > Overlay Plot

Three Dimensional Scatterplot
Graph > Scatterplot 3D

Three Dimensional Scatterplot
Graph > Scatterplot 3D

Contour Plot
Graph > Contour Plot
Bubble Plot
Graph > Bubble Plot

Parallel Plot
Graph > Parallel Plot

Cell Plot
Graph > Cell Plot

Treemap
Graph > Treemap
Discovering JMP

Scatterplot Matrix
Graph > Scatterplot Matrix

Ternary Plot
Graph > Ternary Plot

Contour Profiler
Graph > Contour Profiler

Prediction Profiler
Graph > Profiler
Discovering JMP

Surface Plot
Graph > Surface Plot

Multidimensional Scaling
Analyze > Consumer Research > Multidimensional Scaling

Mixture Profiler
Graph > Mixture Profiler

MaxDiff
Analyze > Consumer Research > MaxDiff
Stress Patterns Plot
Analyze > Reliability and Survival > Cumulative Damage

Repairable Systems Simulation
Analyze > Reliability and Survival > Repairable Systems Simulation

Latent Class Analysis
Analyze > Clustering > Latent Class Analysis

Predictor Screening
Analyze > Screening > Predictor Screening
Process Screening
Analyze > Screening > Process Screening

Text Explorer
Analyze > Text Explorer
This chapter includes the following information:

- book conventions
- JMP documentation
- JMP Help
- additional resources, such as the following:
  - other JMP documentation
  - tutorials
  - indexes
  - Web resources
  - technical support options
The following conventions help you relate written material to information that you see on your screen:

- Sample data table names, column names, pathnames, filenames, file extensions, and folders appear in **Helvetica** font.
- Code appears in **Lucida Sans Typewriter** font.
- Code output appears in **Lucida Sans Typewriter italic** font and is indented farther than the preceding code.
- **Helvetica bold** formatting indicates items that you select to complete a task:
  - buttons
  - check boxes
  - commands
  - list names that are selectable
  - menus
  - options
  - tab names
  - text boxes
- The following items appear in italics:
  - words or phrases that are important or have definitions specific to JMP
  - book titles
  - variables
  - script output
- Features that are for JMP Pro only are noted with the JMP Pro icon 📀. For an overview of JMP Pro features, visit [http://www.jmp.com/software/pro/](http://www.jmp.com/software/pro/).

**Note:** Special information and limitations appear within a Note.

**Tip:** Helpful information appears within a Tip.
JMP Documentation

JMP offers documentation in various formats, from print books and Portable Document Format (PDF) to electronic books (e-books).

- Open the PDF versions from the Help > Books menu.
- All books are also combined into one PDF file, called JMP Documentation Library, for convenient searching. Open the JMP Documentation Library PDF file from the Help > Books menu.
- You can also purchase printed documentation and e-books on the SAS website: http://www.sas.com/store/search.ep?keyWords=JMP

JMP Documentation Library

The following table describes the purpose and content of each book in the JMP library.

<table>
<thead>
<tr>
<th>Document Title</th>
<th>Document Purpose</th>
<th>Document Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discovering JMP</td>
<td>If you are not familiar with JMP, start here.</td>
<td>Introduces you to JMP and gets you started creating and analyzing data.</td>
</tr>
<tr>
<td>Using JMP</td>
<td>Learn about JMP data tables and how to perform basic operations.</td>
<td>Covers general JMP concepts and features that span across all of JMP, including importing data, modifying columns properties, sorting data, and connecting to SAS.</td>
</tr>
<tr>
<td>Basic Analysis</td>
<td>Perform basic analysis using this document.</td>
<td>Describes these Analyze menu 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>• Tabulate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Text Explorer</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Covers how to perform bivariate, one-way ANOVA, and contingency analyses through Analyze &gt; Fit Y by X. How to approximate sampling distributions using bootstrapping and how to perform parametric resampling with the Simulate platform are also included.</td>
</tr>
<tr>
<td>Document Title</td>
<td>Document Purpose</td>
<td>Document Content</td>
</tr>
<tr>
<td>------------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>---------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><em>Essential Graphing</em></td>
<td>Find the ideal graph for your data.</td>
<td>Describes these Graph menu platforms:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Graph Builder</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Overlay Plot</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Scatterplot 3D</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Contour Plot</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Bubble Plot</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Parallel Plot</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Cell Plot</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Treemap</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Scatterplot Matrix</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Ternary Plot</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Chart</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The book also covers how to create background and custom maps.</td>
</tr>
<tr>
<td><em>Profilers</em></td>
<td>Learn how to use interactive profiling tools, which enable you to view cross-sections of any response surface.</td>
<td>Covers all profilers listed in the Graph menu. Analyzing noise factors is included along with running simulations using random inputs.</td>
</tr>
<tr>
<td><em>Design of Experiments Guide</em></td>
<td>Learn how to design experiments and determine appropriate sample sizes.</td>
<td>Covers all topics in the DOE menu and the Specialized DOE Models menu item in the Analyze &gt; Specialized Modeling menu.</td>
</tr>
<tr>
<td>Document Title</td>
<td>Document Purpose</td>
<td>Document Content</td>
</tr>
<tr>
<td>--------------------------</td>
<td>-------------------------------------------------------</td>
<td>----------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><em>Fitting Linear Models</em></td>
<td>Learn about Fit Model platform and many of its personalities.</td>
<td>Describes these personalities, all available within the Analyze menu Fit Model platform:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Standard Least Squares</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Stepwise</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Generalized Regression</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Mixed Model</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• MANOVA</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Loglinear Variance</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Nominal Logistic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Ordinal Logistic</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Generalized Linear Model</td>
</tr>
</tbody>
</table>


<table>
<thead>
<tr>
<th>Document Title</th>
<th>Document Purpose</th>
<th>Document Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Predictive and Specialized</td>
<td>Learn about additional modeling techniques.</td>
<td>Describes these Analyze &gt; Predictive Modeling menu platforms:</td>
</tr>
<tr>
<td>Modeling</td>
<td></td>
<td>• Modeling Utilities</td>
</tr>
<tr>
<td>Specialized</td>
<td></td>
<td>• Neural</td>
</tr>
<tr>
<td>Modeling</td>
<td></td>
<td>• Partition</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Bootstrap Forest</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Boosted Tree</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• K Nearest Neighbors</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Naive Bayes</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Model Comparison</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Formula Depot</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Describes these Analyze &gt; Specialized Modeling menu platforms:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Fit Curve</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Nonlinear</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Gaussian Process</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Time Series</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Matched Pairs</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Describes these Analyze &gt; Screening menu platforms:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Response Screening</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Process Screening</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Predictor Screening</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Association Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>The platforms in the Analyze &gt; Specialized Modeling &gt; Specialized DOE Models menu are described in Design of Experiments Guide.</td>
</tr>
<tr>
<td><strong>Document Title</strong></td>
<td><strong>Document Purpose</strong></td>
<td><strong>Document Content</strong></td>
</tr>
<tr>
<td>-------------------------</td>
<td>--------------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Multivariate Methods</strong></td>
<td>Read about techniques for analyzing several variables simultaneously.</td>
<td>Describes these Analyze &gt; Multivariate Methods menu platforms:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Multivariate</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Principal Components</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Discriminant</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Partial Least Squares</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Describes these Analyze &gt; Clustering menu platforms:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Hierarchical Cluster</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• K Means Cluster</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Normal Mixtures</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Latent Class Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Cluster Variables</td>
</tr>
<tr>
<td><strong>Quality and Process Methods</strong></td>
<td>Read about tools for evaluating and improving processes.</td>
<td>Describes these Analyze &gt; Quality and Process menu platforms:</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Control Chart Builder and individual control charts</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Measurement Systems Analysis</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Variability / Attribute Gauge Charts</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Process Capability</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Pareto Plot</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• Diagram</td>
</tr>
<tr>
<td>Document Title</td>
<td>Document Purpose</td>
<td>Document Content</td>
</tr>
<tr>
<td>----------------------</td>
<td>----------------------------------------------------------------------------------</td>
<td>-------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
</tbody>
</table>
| **Reliability and Survival Methods** | Learn to evaluate and improve reliability in a product or system and analyze survival data for people and products. | Describes these Analyze > Reliability and Survival menu platforms:  
  - Life Distribution  
  - Fit Life by X  
  - Cumulative Damage  
  - Recurrence Analysis  
  - Degradation and Destructive Degradation  
  - Reliability Forecast  
  - Reliability Growth  
  - Reliability Block Diagram  
  - Repairable Systems Simulation  
  - Survival  
  - Fit Parametric Survival  
  - Fit Proportional Hazards |
| **Consumer Research** | Learn about methods for studying consumer preferences and using that insight to create better products and services. | Describes these Analyze > Consumer Research menu platforms:  
  - Categorical  
  - Multiple Correspondence Analysis  
  - Multidimensional Scaling  
  - Factor Analysis  
  - Choice  
  - MaxDiff  
  - Uplift  
  - Item Analysis |
| **Scripting Guide**   | Learn about taking advantage of the powerful JMP Scripting Language (JSL).       | Covers a variety of topics, such as writing and debugging scripts, manipulating data tables, constructing display boxes, and creating JMP applications. |
JMP Help

JMP Help is an abbreviated version of the documentation library that provides targeted information. You can open JMP Help in several ways:

- On Windows, press the F1 key to open the Help system window.
- Get help on a specific part of a data table or report window. Select the Help tool from the Tools menu and then click anywhere in a data table or report window to see the Help for that area.
- Within a JMP window, click the Help button.
- Search the Help at http://jmp.com/support/help/ (English only).

Additional Resources for Learning JMP

In addition to JMP documentation and JMP Help, you can also learn about JMP using the following resources:

- Tutorials (see “Tutorials” on page 44)
- Sample data (see “Sample Data Tables” on page 44)
- Indexes (see “Learn about Statistical and JSL Terms” on page 44)
- Tip of the Day (see “Learn JMP Tips and Tricks” on page 44)
- Web resources (see “JMP User Community” on page 45)
- JMPer Cable technical publication (see “JMPer Cable” on page 45)
- Books about JMP (see “JMP Books by Users” on page 46)
- JMP Starter (see “The JMP Starter Window” on page 46)
Learn about JMP

Additional Resources for Learning JMP

- Teaching Resources (see “Sample Data Tables” on page 44)

Tutorials

You can access JMP tutorials by selecting Help > Tutorials. The first item on the Tutorials menu is 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 designing an experiment and comparing a sample mean to a constant.

Sample Data Tables

All of the examples in the JMP documentation suite use sample data. Select Help > Sample Data Library to open the sample data directory.

To view an alphabetized list of sample data tables or view sample data within categories, select Help > Sample Data.

Sample data tables are installed in the following directory:

- On Windows: C:\Program Files\SAS\JMP\13\Samples\Data
- On Macintosh: \Library\Application Support\JMP\13\Samples\Data

In JMP Pro, sample data is installed in the JMPPRO (rather than JMP) directory. In JMP Shrinkwrap, sample data is installed in the JMPSW directory.

To view examples using sample data, select Help > Sample Data and navigate to the Teaching Resources section. To learn more about the teaching resources, visit http://jmp.com/tools.

Learn about Statistical and JSL Terms

The Help menu contains the following indexes:

Statistics Index Provides definitions of statistical terms.

Scripting Index Lets you search for information about JSL functions, objects, and display boxes. You can also edit and run sample scripts from the Scripting Index.

Learn JMP Tips and Tricks

When you first start JMP, you see the Tip of the Day window. This window provides tips for using JMP.
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 for details.

### Tooltips

JMP provides descriptive tooltips when you place your cursor over items, such as the following:

- Menu or toolbar options
- Labels in graphs
- Text results in the report window (move your cursor in a circle to reveal)
- Files or windows in the Home Window
- Code in the Script Editor

**Tip:** On Windows, you can hide tooltips in the JMP Preferences. Select **File > Preferences > General** and then deselect **Show menu tips**. This option is not available on Macintosh.

### JMP User Community

The JMP User Community provides a range of options to help you learn more about JMP and connect with other JMP users. The learning library of one-page guides, tutorials, and demos is a good place to start. And you can continue your education by registering for a variety of JMP training courses.

Other resources include a discussion forum, sample data and script file exchange, webcasts, and social networking groups.

To access JMP resources on the website, select **Help > JMP User Community** or visit [https://community.jmp.com/](https://community.jmp.com/).

### JMPer Cable

The JMPer Cable is a yearly technical publication targeted to users of JMP. The JMPer Cable is available on the JMP website:

[http://www.jmp.com/about/newsletters/jmpercable/](http://www.jmp.com/about/newsletters/jmpercable/)
JMP Books by Users

Additional books about using JMP that are written by JMP users are available on the JMP website:


The JMP Starter Window

The JMP Starter window is a good place to begin if you are not familiar with JMP or data analysis. Options are categorized and described, and you launch them by clicking a button. The JMP Starter window covers many of the options found in the Analyze, Graph, Tables, and File menus. The window also lists JMP Pro features and platforms.

- To open the JMP Starter window, select View (Window on the Macintosh) > JMP Starter.
- To display the JMP Starter automatically when you open JMP on Windows, select File > Preferences > General, and then select JMP Starter from the Initial JMP Window list. On Macintosh, select JMP > Preferences > Initial JMP Starter Window.

Technical Support

JMP technical support is provided by statisticians and engineers educated in SAS and JMP, many of whom have graduate degrees in statistics or other technical disciplines.

Many technical support options are provided at http://www.jmp.com/support, including the technical support phone number.
JMP (pronounced *jump*) is a powerful and interactive data visualization and statistical analysis tool. Use JMP to learn more about your data by performing analyses and interacting with the data using data tables, graphs, charts, and reports.

JMP enables researchers to perform a wide range of statistical analyses and modeling. JMP is equally useful to the business analyst who wants to quickly uncover trends and patterns in data. With JMP, you do not have to be an expert in statistics to get information from your data.

For example, you can use JMP to do the following:

- Create interactive graphs and charts to explore your data and discover relationships.
- Discover patterns of variation across many variables at once.
- Explore and summarize large amounts of data.
- Develop powerful statistical models to predict the future.

**Figure 2.1 Examples of JMP Reports**
Concepts That You Should Know

Before you begin using JMP, you should be familiar with these concepts:

- Enter, view, edit, and manipulate data using JMP data tables.
- Select a platform from the Analyze, Graph, or DOE menus. Platforms contain interactive windows that you 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.

How Do I Get Started?

The general workflow in JMP is simple:

1. Get your data into JMP.
2. Select a platform and complete its launch window.
3. Explore your results and discover where your data takes you.

This workflow is described in more detail in “Understanding the JMP Workflow” on page 53.

Typically, you start your work in JMP by using graphs to visualize individual variables and relationships among your variables. Graphs make it easy to see this information, and to see the deeper questions to ask. Then you use analysis platforms to dig deeper into your problems and find solutions.

- The “Work with Your Data” chapter on page 61 shows you how to get data into JMP.
- The “Visualize Your Data” chapter on page 89 shows you how to use some of the useful graphs JMP provides to look more closely at your data.
- The “Analyze Your Data” chapter on page 123 shows you how to use some of the analysis platforms.
- The “The Big Picture” chapter on page 157 shows you how to analyze distributions, patterns, and similar values in several platforms.

Each chapter teaches through examples. The following sections in this chapter describe data tables and general concepts for working in JMP.
Starting JMP

Start JMP in two ways:

- Double-click the JMP icon, normally found on your desktop. This starts JMP, but does not open any existing JMP files.
- Double-click an existing JMP file. This starts JMP and opens the file.

The initial view of JMP includes the Tip of the Day window and the Home Window on Windows; on Macintosh, the Tip of the Day and JMP Starter, and Home Window initially appear.

The JMP Starter window classifies actions and platforms using categories.

**Figure 2.2 The JMP Starter**

On the left is a list of categories. Click a category to see the features and the commands related to that category. The JMP Starter also lists JMP Pro features and platforms.

The Home Window helps you organize and access files in JMP.
To open the Home Window on Windows, select View > Home Window. On Macintosh, select Window > JMP Home. The Home Window includes links to the following:

- the data tables and report windows that are currently open
- files that you have opened recently

For more details about the Home Window, see the Get Started chapter in the Using JMP book. Almost all JMP windows contain a menu bar and a toolbar. You can find most JMP features in three ways:

- using the menu bar
- using the toolbar buttons
- using the buttons on the JMP Starter window

**About the Menu Bar and Toolbars**

The menus and toolbars are hidden in many windows. To see them, place your mouse pointer over the blue bar under the window’s title bar. The menus in the JMP Starter window, the Home Window, and all data tables are always visible.
Using Sample Data

The examples in this book and the other JMP books use sample data tables. The default location on Windows for the sample data is:

- C:/Program Files/SAS/JMP/13/Samples/Data
- C:/Program Files/SAS/JMPPro/13/Samples/Data
- C:/Program Files/SAS/JMPSW/13/Samples/Data

The Sample Data Index groups the data tables by category. Click a disclosure button to see a list of data tables for that category, and then click a link to open a data table.

Macintosh sample data is installed in /Library/Application Support/JMP/13/Samples/Data.

Opening a JMP Sample Data Table

1. From the Help menu, select Sample Data.
2. Open the Data Tables used in Discovering JMP list by clicking on the disclosure button next to it.
3. Click the name of the data table to use it in the examples in this book.

Sample Import Data

Use files from other applications to learn how to import data into JMP.

The default location on Windows for the sample import data is:

- C:/Program Files/SAS/JMP/13/Samples/Import Data
- C:/Program Files/SAS/JMPPro/13/Samples/Import Data
- C:/Program Files/SAS/JMPSW/13/Samples/Import Data

Understanding Data Tables

A data table is a collection of data organized in rows and columns. It is similar to a Microsoft® Excel® spreadsheet, but with some important differences that are discussed in “How is JMP Different from Excel?” on page 58. A data table might also contain other information like notes, variables, and scripts. These supplementary items are discussed in later chapters.

Open the VA Lung Cancer data table to see the data table described here.
A data table contains the following parts:

**Data grid**  The data grid contains the data arranged in rows and columns. Generally, each row in the data grid is an observation, and the columns (also called variables) give information about the observations. In Figure 2.4, each row corresponds to a test subject, and there are twelve columns of information. Although all twelve columns cannot be shown in the data grid, the Columns panel lists them all. The information given about each test subject includes the time, cell type, treatment, and more. Each column has a header, or name. That name is not part of the table’s total count of rows.

**Table panel**  The table panel can contain table variables or table scripts. In Figure 2.4, there is one saved script called Model that can automatically re-create an analysis. This table also has a variable named Notes that contains information about the data. Table variables and table scripts are discussed in a later chapter.

**Columns panel**  The columns panel shows the total number of columns, whether any columns are selected, and a list of all the columns by name. The numbers in parentheses (12/0) show that there are twelve columns, and that no columns are selected. An icon to the left of each column name shows that column’s modeling type. Modeling types are described in
“Understand Modeling Types” on page 126 in the “Analyze Your Data” chapter. Icons to the right show any attributes assigned to the column. See “View or Change Column Information” on page 73 in the “Work with Your Data” chapter for more information about these icons.

**Rows panel** The rows panel shows the number of rows in the data table, and how many rows are selected, excluded, hidden, or labeled. In Figure 2.4, there are 137 rows in the data table.

**Thumbnail links to report windows** This area shows thumbnails of all reports based on the data table. Place your mouse pointer over a thumbnail to see a larger preview of the report window. Double-click a thumbnail to bring the report window to the front.

Interacting with the data grid, which includes adding rows and columns, entering data, and editing data, is discussed in the “Work with Your Data” chapter on page 61. If you open multiple data tables, each one appears in a separate window.

---

**Understanding the JMP Workflow**

Once your data is in a data table, you can create graphs or plots, and perform analyses. All features are located in platforms, which are found primarily on the Analyze or Graph menus. They are called platforms because they do not just produce simple static results. Platform results appear in report windows, are highly interactive, and are linked to the data table and to each other.

The platforms under the Analyze and Graph menus provide a variety of analytical features and data exploration tools.

The general steps to produce a graph or analysis are as follows:

1. Open a data table.
2. Select a platform from the Graph or Analyze menu.
3. Complete the platform launch window to set up your analysis.
4. Click **OK** to create the report window that contains your graphs and statistical analyses.
5. Customize your report by using report options.
6. Save, export, and share your results with others.

Later chapters discuss these concepts in greater detail.

The following example shows you how to perform a simple analysis and customize it in four steps. This example uses the Companies.jmp file sample data table to show a basic analysis of the variable Profits ($M).
**Step 1: Launching a Platform and Viewing Results**

1. Select **Help > Sample Data Library** and open **Companies.jmp**.
2. Select **Analyze > Distribution** to open the Distribution launch window.
3. Select **Profits ($M)** in the Select Columns box and click the **Y, Columns** button.
   - The variable **Profits ($M)** appears in the **Y, Columns** role. See Figure 2.5 for the completed window.
   - Another way to assign variables is to click and drag columns from the Select Columns box to any of the roles boxes.

**Figure 2.5 Assign Profits ($M)**

4. Click **OK**.
   - The Distribution report window appears.
The report window contains basic plots or graphs and preliminary analysis reports. The results appear in an outline format, and you can show or hide any report by clicking on the disclosure button.

Red triangle menus contain options and commands to request additional graphs and analyses at any time.

- On Windows, place your mouse pointer over the blue bar at the top of the window to see the menu bar and the toolbars.
- On Windows, click the data table button in the lower right corner to view the data table that was used to create this report. On Macintosh, click the Show Data Table button in the upper right corner of the report window.
- On Windows, click the JMP Home Window button in the lower right corner to view the Home Window. On Macintosh, select Window > JMP Home.
Step 2: Removing the Box Plot

Continue using the Distribution report that you created earlier.

1. Click the red triangle next to Profits ($M) to see a menu of report options.
2. Deselect Outlier Box Plot to turn the option off.

   The outlier box plot is removed from the report window.

Figure 2.7 Removing the Outlier Box Plot

Step 3: Requesting Additional Output

Continue to use the same report window.

1. From the red triangle menu next to Profits ($M), select Test Mean.
   The Test Mean window appears.
2. Enter 500 in the Specify Hypothesized Mean box.
3. Click OK.
   The test for the mean is added to the report window.
Chapter 2
Discovering JMP

Introducing JMP
Understanding the JMP Workflow

Figure 2.8 Test for the Mean

Step 4: Interacting with Platform Results

All platforms produce results that are interactive. For example:

- Reports can be shown or hidden.
- Additional graphs and statistical details can be added or removed to suit your purposes.
- Platform results are connected to the data table and to each other.

For example, to close the Quantiles report, click the disclosure button next to Quantiles.

Figure 2.9 Close the Quantiles Report

Platform results are connected to the data table. The histogram in Figure 2.10 shows that a group of companies makes a much higher profit that the others. To quickly identify that group, click on the histogram bar for them. The corresponding rows in the data table are selected.
How is JMP Different from Excel?

There are a number of important differences between JMP and Excel or other spreadsheet applications.

Table 2.1 JMP and Excel Differences

<table>
<thead>
<tr>
<th>Formulas</th>
<th>Excel</th>
<th>JMP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Formulas are applied</td>
<td>to individual cells.</td>
<td>only to entire columns. “Calculate Values with Formulas” on page 75 in the “Work with Your Data” chapter describes how to use formulas.</td>
</tr>
</tbody>
</table>

| Column Names             | Excel | |
|--------------------------|-------| Column names are part of the grid. Numbered rows and labeled columns extend past the data. Numeric and character data reside in the same column. |
Discovering JMP

How is JMP Different from Excel?

JMP Column names are not part of the grid. There are no rows and columns beyond the existing data. The grid is only as big as the data. A column is either numeric or character. If a column contains both character and numeric data, the entire column’s data type is character, and the numbers are treated as character data.

“Understand Modeling Types” on page 126 in the “Analyze Your Data” chapter describes how data type influences platform results.

### Tables and Worksheets

<table>
<thead>
<tr>
<th>Excel</th>
<th>A single spreadsheet contains several tables, or worksheets.</th>
</tr>
</thead>
<tbody>
<tr>
<td>JMP</td>
<td>JMP does not have the concept of worksheets. Each data table is a separate .jmp file and appears in a separate window.</td>
</tr>
</tbody>
</table>

### Data Grid

<table>
<thead>
<tr>
<th>Excel</th>
<th>Data can be located anywhere in the data grid.</th>
</tr>
</thead>
<tbody>
<tr>
<td>JMP</td>
<td>Data always begins in row 1 and column 1.</td>
</tr>
</tbody>
</table>

### Analysis and Graph Reports

<table>
<thead>
<tr>
<th>Excel</th>
<th>All data, analyses, and graphs are placed inside the data grid.</th>
</tr>
</thead>
<tbody>
<tr>
<td>JMP</td>
<td>Results appear in a separate window.</td>
</tr>
</tbody>
</table>

---

Table 2.1 JMP and Excel Differences  (Continued)
Before graphing or analyzing your data, the data has to be in a data table and in the proper format. This chapter shows some basic data management tasks, including the following:

- Creating new data tables
- Opening existing data tables
- Importing data from other applications into JMP
- Managing your data

**Figure 3.1 Example of a Data Table**

<table>
<thead>
<tr>
<th>Companies</th>
<th>Type</th>
<th>Sales (SM)</th>
<th>Profits (SM)</th>
<th># Employ</th>
<th>profit/emp</th>
<th>Assets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Computer</td>
<td>855.1</td>
<td>31.0</td>
<td>7523</td>
<td>4120.70</td>
<td>615.2</td>
</tr>
<tr>
<td>2</td>
<td>Pharmaceutical</td>
<td>5453.5</td>
<td>859.8</td>
<td>40929</td>
<td>21007.11</td>
<td>4851.6</td>
</tr>
<tr>
<td>3</td>
<td>Computer</td>
<td>2153.7</td>
<td>153.9</td>
<td>8200</td>
<td>18685.84</td>
<td>2233.7</td>
</tr>
<tr>
<td>4</td>
<td>Pharmaceutical</td>
<td>6747.0</td>
<td>1102.2</td>
<td>56816</td>
<td>21690.02</td>
<td>5881.5</td>
</tr>
<tr>
<td>5</td>
<td>Computer</td>
<td>5284.0</td>
<td>454.0</td>
<td>12086</td>
<td>37620.15</td>
<td>2743.9</td>
</tr>
<tr>
<td>6</td>
<td>Pharmaceutical</td>
<td>9422.0</td>
<td>747.0</td>
<td>54100</td>
<td>138077.86</td>
<td>8497.0</td>
</tr>
<tr>
<td>7</td>
<td>Computer</td>
<td>2878.1</td>
<td>333.3</td>
<td>9500</td>
<td>35084.21</td>
<td>2000.4</td>
</tr>
<tr>
<td>8</td>
<td>Computer</td>
<td>709.3</td>
<td>41.4</td>
<td>5000</td>
<td>8280.00</td>
<td>488.1</td>
</tr>
<tr>
<td>9</td>
<td>Computer</td>
<td>2952.1</td>
<td>-80.4</td>
<td>1800.0</td>
<td>-37800.0</td>
<td>1800.7</td>
</tr>
<tr>
<td>10</td>
<td>Computer</td>
<td>784.7</td>
<td>89.0</td>
<td>4708</td>
<td>18903.99</td>
<td>955.8</td>
</tr>
<tr>
<td>11</td>
<td>Computer</td>
<td>1324.3</td>
<td>-119.7</td>
<td>1374.0</td>
<td>-8711.79</td>
<td>1040.2</td>
</tr>
<tr>
<td>12</td>
<td>Pharmaceutical</td>
<td>4175.0</td>
<td>939.5</td>
<td>2820.0</td>
<td>33315.00</td>
<td>5848.0</td>
</tr>
<tr>
<td>13</td>
<td>Computer</td>
<td>11899.0</td>
<td>829.0</td>
<td>65000</td>
<td>8726.32</td>
<td>10075.0</td>
</tr>
<tr>
<td>14</td>
<td>Computer</td>
<td>873.6</td>
<td>79.5</td>
<td>8200</td>
<td>9695.12</td>
<td>808.0</td>
</tr>
<tr>
<td>15</td>
<td>Pharmaceutical</td>
<td>9844.0</td>
<td>1022.0</td>
<td>63100</td>
<td>13020.46</td>
<td>7919.0</td>
</tr>
<tr>
<td>16</td>
<td>Pharmaceutical</td>
<td>969.2</td>
<td>227.4</td>
<td>3418</td>
<td>65530.13</td>
<td>784.0</td>
</tr>
<tr>
<td>17</td>
<td>Pharmaceutical</td>
<td>6698.4</td>
<td>1485.4</td>
<td>34400</td>
<td>43470.93</td>
<td>6756.7</td>
</tr>
<tr>
<td>18</td>
<td>Computer</td>
<td>5956.0</td>
<td>412.0</td>
<td>56000</td>
<td>7357.14</td>
<td>4500.0</td>
</tr>
<tr>
<td>19</td>
<td>Pharmaceutical</td>
<td>5803.7</td>
<td>681.1</td>
<td>42100</td>
<td>16178.15</td>
<td>8324.8</td>
</tr>
<tr>
<td>20</td>
<td>Computer</td>
<td>2959.3</td>
<td>252.8</td>
<td>31404</td>
<td>8049.93</td>
<td>5611.1</td>
</tr>
<tr>
<td>21</td>
<td>Pharmaceutical</td>
<td>1193.3</td>
<td>85.5</td>
<td>8527</td>
<td>10144.25</td>
<td>1791.7</td>
</tr>
</tbody>
</table>
Get Your Data into JMP

- To copy and paste data from another application, see “Copy and Paste Data” on page 62.
- To import data from another application, see “Import Data” on page 62.
- To enter data directly into a data table, see “Enter Data” on page 65
- To open a data table, double-click on the file, or use the File > Open command.

You can also import data into JMP from a database. For more information, see the Import Your Data chapter in the Using JMP book.

This chapter uses sample data tables and sample import data that is installed with JMP. To find these files, see “Using Sample Data” on page 51 in the “Introducing JMP” chapter.

Copy and Paste Data

You can move data into JMP by copying and pasting from another application, such as Excel or a text file.

1. Open the VA Lung Cancer.xls file in Excel. This file is located in the Sample Import Data folder.
2. Select all of the rows and columns, including the column names. There are 12 columns and 138 rows.
3. Copy the selected data.
4. In JMP, select File > New > Data Table to create an empty table.
5. Select Edit > Paste with Column Names to paste the data and column headings.
   If the data that you are pasting into JMP does not have column names, then you can use Edit > Paste.

Import Data

You can move data into JMP by importing data from another application, such as Excel, SAS, or text files. The basic steps to import data are as follows:

1. Select File > Open.
2. Navigate to your file’s location.
3. If your file is not listed in the Open Data File window, select the correct file type from the Files of type menu.
4. Click Open.
Example of Importing a Microsoft Excel File

1. Select File > Open.
2. Navigate to the Samples/Import Data folder.
3. Select Team Results.xls.
   Note the rows and columns on which the data begin. The spreadsheet also contains two worksheets. In this example, you import the Ungrouped Team Results worksheet.
4. Click Open.
   The spreadsheet opens in the Excel Import Wizard, where a preview of the data appears along with import options.
   Text from the first row of the spreadsheet are column headings. However, you want text in row 3 of the spreadsheet to be converted to column headings.
5. Next to Column headers start on row, type 3, and press Enter. The column headings are updated in the data preview. The value for the first row of data is updated to 4.
6. Save the settings only for this worksheet:
   – Deselect Use for all worksheets in the lower left corner of the window.
   – Select Ungrouped Team Results in the upper right corner of the window.
7. Click Import to convert the spreadsheet as you specified.

When you import Excel files, JMP predicts whether columns headings exist, and if the column names are on row one. The copy and paste method is recommended for the following situations:
- If the column names are located in a row other than row one
- If the file does not include column names and the data does not start in row one
- If the file contains column names and the data does not start in row two

See “Copy and Paste Data” on page 62 and the Import Your Data chapter in the Using JMP book for more information about importing Excel files.

Example of Importing a Text File

One way to import a text file is to let JMP assume the data’s format and place the data in a data table. This method uses settings that you can specify in Preferences. See the JMP Preferences chapter in the Using JMP book for information about setting text import preferences.

Another way to import a text file is to use a Text Preview window to see what your data table will look like after importing, and make adjustments. The following example shows you how to use Text Import Preview window.

1. Select File > Open.
2. Navigate to the Samples/Import Data folder.
4. At the bottom of the Open window, select Data with Preview.
5. Click Open.

**Figure 3.2 Initial Preview Window**

This text file has a title on the first line, column names on the third line, and the data starts on line four. If you opened this directly in JMP, the Animals Data line would be the first column name, and all the column names and data afterward would be out of sync. The Preview window lets you adjust the settings before you open the file, and see how your adjustments affect the final data table.

6. Enter 3 in the **File contains column names on line** field.
7. Enter 4 in the **Data starts on line** field.
8. Click **Next**.
   
   In the second window, you can exclude columns from the import and change the data modeling of the columns. For this example, use the default settings.
9. Click **Import**.

The new data table has columns named species, subject, miles, and season. The species and season columns are character data. The subject and miles columns are continuous numeric data.
Enter Data

You can enter data directly in a data table. The following example shows you how to enter data that was collected over several months into a data table.

Scenario

Table 3.1 shows the data from a study that investigated a new blood pressure medication. Each individual’s blood pressure was measured over a six-month period. Two doses (300mg and 450mg) of the medication were used, along with a control and placebo group. The data shows the average blood pressure for each group.

Table 3.1 Blood Pressure Data

<table>
<thead>
<tr>
<th>Month</th>
<th>Control</th>
<th>Placebo</th>
<th>300mg</th>
<th>450mg</th>
</tr>
</thead>
<tbody>
<tr>
<td>March</td>
<td>165</td>
<td>163</td>
<td>166</td>
<td>168</td>
</tr>
<tr>
<td>April</td>
<td>162</td>
<td>159</td>
<td>165</td>
<td>163</td>
</tr>
<tr>
<td>May</td>
<td>164</td>
<td>158</td>
<td>161</td>
<td>153</td>
</tr>
<tr>
<td>June</td>
<td>162</td>
<td>161</td>
<td>158</td>
<td>151</td>
</tr>
<tr>
<td>July</td>
<td>166</td>
<td>158</td>
<td>160</td>
<td>148</td>
</tr>
<tr>
<td>August</td>
<td>163</td>
<td>158</td>
<td>157</td>
<td>150</td>
</tr>
</tbody>
</table>

Enter Data in a New Data Table

1. Select File > New > Data Table to create an empty data table.
   A new data table has one column and no rows.
2. Select the column name and change the name to Month. See Figure 3.3

   **Note:** To rename a column, you can also double-click the column name or select the column and press Enter.
3. Select **Rows > Add Rows**.
   The Add Rows window appears.
4. Since you want to add six rows, type 6.
5. Click **OK**. Six empty rows are added to the data table.
6. Enter the **Month** information by clicking in a cell and typing.

**Figure 3.4** Month Column Completed

In the columns panel, look at the modeling type icon to the left of the column name. It has changed to reflect that **Month** is now nominal (previously it was continuous). Compare the modeling type shown for Column 1 in Figure 3.3 and for **Month** in Figure 3.4. This difference is important and is discussed in “**View or Change Column Information**” on page 73.

7. Double-click in the space on the right side of the **Month** column to add the **Control** column.
8. Change the name to **Control**.
9. Enter the **Control** data as shown in Table 3.1. Your data table now consists of six rows and two columns.
10. Continue adding columns and entering data as shown in Table 3.1 to create the final data table with six rows and five columns.

**Change the Data Table Name**

1. Double-click on the data table name (Untitled) in the Table Panel.
2. Type the new name (Blood Pressure).

**Figure 3.5** Changing the Data Table Name

Double-click here. Type the new name.
Transfer Data from Excel

You can use the JMP Add In for Excel to transfer a spreadsheet from Excel to JMP:

- a data table
- Graph Builder
- Distribution platform
- Fit Y by X platform
- Fit Model platform
- Time Series platform
- Control Chart platform

Set JMP Add In Preferences in Excel

To configure JMP Add In Preferences:

1. In Excel, select JMP > Preferences.
   The JMP Preferences window appears.

2. Accept the default Data Table Name (File name_Worksheet name) or type a name.
3. Select to Use the first rows as column names if the first row in the worksheet contains column headers.
4. If you selected to use the first rows a column headers, type the number of rows used.
5. Select to Transfer Hidden Rows if the worksheet contains hidden rows to be included in the JMP data table.
6. Select to Transfer Hidden Column if the worksheet contains hidden columns to be included in the JMP data table.
7. Click OK to save your preferences.
Transfer to JMP

To transfer an Excel worksheet to JMP:

1. Open the Excel file.
2. Select the worksheet to transfer.
3. Select **JMP** and then select the JMP destination:
   - Data Table
   - Graph Builder
   - Distribution platform
   - Fit Y by X platform
   - Fit Model platform
   - Time Series platform
   - Control Chart platform

The Excel worksheet is opened as a data table in JMP and the selected platform’s launch window appears.

---

Work with Data Tables

This section contains the following information:

- “Edit Data” on page 68
- “Select, Deselect, and Find Values” on page 70
- “View or Change Column Information” on page 73
- “Calculate Values with Formulas” on page 75
- “Filter Data” on page 77

**Tip:** Consider setting the Autosave Time Out value in the General preferences to automatically save open data tables at the specified number of minutes. This autosave value also applies to journals, scripts, projects, and reports.

---

Edit Data

You can enter or change data, either a few cells at a time or for an entire column. This section contains the following information:

- “Change Values” on page 69
- “Recode Values” on page 69
Change Values

To change a value, select a cell and type the change. You can also double-click a cell to edit it.

**Note:** Double-clicking in a cell is not the same as selecting a cell. A single click selects a cell. You can select more than one cell at the same time, and you can perform certain actions on selected cells. Double-clicking only lets you edit a cell. For more information about selecting rows, columns, and cells, see “Select, Deselect, and Find Values” on page 70.

Recode Values

Use the recoding tool to change all of the values in a column at once. For example, suppose that you are interested in comparing the sales of computer and pharmaceutical companies. Your current company labels are Computer and Pharmaceutical. You want to change them to Technical and Drug. Going through all 32 rows of data and changing all the values would be tedious, inefficient, and error-prone, especially if you had many more rows of data. Recode is a better option.

1. Select Help > Sample Data Library and open Companies.jmp.
2. Select the Type column by clicking once on the column heading.
5. Click Done and select the In Place option from the list.

![Recode Window](image_url)

All cells are updated automatically to the new values.

Create Patterned Data

Use the Fill options to populate a column with patterned data. The Fill options are especially useful if your data table is large, and typing in the values for each row would be cumbersome.
Example of Filling a Column with the Pattern

1. Add a new column.
2. Enter 1 in the first cell, 2 in the second cell, and 3 in the third cell.
3. Select the three cells, and right-click anywhere in the selected cells to see a menu.
4. Select Fill > Repeat sequence to end of table.

The rest of the column is filled with the sequence (1, 2, 3, 1, 2, 3, ...).

To continue a pattern instead of repeating it (1, 2, 3, 4, 5, 6, ...), select Continue sequence to end of table. This command can also be used to generate patterns like (1, 1, 2, 2, 2, 3, 3, 3, ...).

The Fill options can recognize simple arithmetic and geometric sequences. For character data, the Fill options only repeat the values.

Select, Deselect, and Find Values

You can select rows, columns, or cells within a data table. For example, to create a subset of an existing data table, you must first select the parts of the table that you want to subset. Also, selecting rows can make data points stand out on a graph. Select rows and columns manually by clicking, or select rows that meet certain search criteria. This section contains the following information:

- “Select and Deselect Rows” on page 70
- “Select and Deselect Columns” on page 71
- “Select and Deselect Cells” on page 72
- “Search for Values” on page 72

Select and Deselect Rows

Table 3.2  Selecting and Deselecting Rows

<table>
<thead>
<tr>
<th>Task</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select rows one at a time</td>
<td>Click on the row number.</td>
</tr>
<tr>
<td>Select multiple adjacent rows</td>
<td>Click and drag on the row numbers.</td>
</tr>
<tr>
<td></td>
<td>or</td>
</tr>
<tr>
<td></td>
<td>Select the beginning row, and then hold down the Shift key and click the last row number.</td>
</tr>
<tr>
<td>Select multiple non-adjacent rows</td>
<td>Select the first row, and then hold down the Ctrl key and click the other row numbers.</td>
</tr>
<tr>
<td>Deselect rows one at a time</td>
<td>Hold down the Ctrl key and click the row numbers.</td>
</tr>
</tbody>
</table>
Table 3.2 Selecting and Deselecting Rows (Continued)

<table>
<thead>
<tr>
<th>Task</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deselect all rows</td>
<td>Click in the lower-triangular space in the top left corner of the table. See Figure 3.8.</td>
</tr>
</tbody>
</table>

Figure 3.8 Deselecting Rows
To deselect all rows at once, click here.

Table 3.3 Selecting and Deselecting Columns

<table>
<thead>
<tr>
<th>Task</th>
<th>Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Select columns one at a time</td>
<td>Click the column heading.</td>
</tr>
</tbody>
</table>
| Select multiple adjacent columns | Click and drag across the column headings.  
|                               | or                                                                                         |
|                               | Select the beginning column, and then hold down the Shift key and click the last header.    |
| Select multiple non-adjacent columns | Select the first column, and then hold down the Ctrl key and click the other column headings. |
| Deselect columns one at a time | Hold down the Ctrl key and click the column heading.                                         |
| Deselect all columns         | Click in the upper-triangular space in the top left corner of the table. See Figure 3.9.    |
In a data table that has thousands or tens of thousands of rows, it can be difficult to locate a particular cell by scrolling through the table. If you are looking for specific information, use the Search feature to find it. If data is matches the search criteria, the cell is selected and the data grid scrolls to show it in the window. For example, the Companies.jmp data table contains information about a company that has total sales of $11,899. Use the Search feature to find that cell.

Example of Searching for a Value

1. Select Edit > Search > Find to launch the Search window.
2. In the Find what box, enter 11899.
3. Click Find. JMP finds the first cell that has 11,899 in it, and selects it.
If multiple cells meet the search criteria, click **Find** again to find the next cell that matches the search term.

You can also search for multiple rows at once, with each row matching some criteria.

**Example of Select All Rows That Correspond to Medium-Sized Companies**

1. Select **Rows > Row Selection > Select Where** to open the **Select rows** window.
2. In the column list box on the left, select **Size Co**.
3. In the text box on the right, enter medium.
4. Click **OK**.

**Figure 3.10** Select Rows Window

JMP selects all of the rows that have **Size Co** equal to medium. There are seven.

**View or Change Column Information**

Information about a column is not limited to the data in the column. Data type, modeling type, format, and formulas can also be set.

To view or change column characteristics, double-click the column heading. Or, right-click the column heading and select **Column Info**. The Column Info window appears.
Figure 3.11 Column Info Window

Column Name  Enter or change the column name. No two columns can have the same column name.

Data Type    Select one of the following data types:
- Numeric specifies the column values as numbers.
- Character specifies the column values as non-numeric, such as letters or symbols.
- Row State specifies the column values as row states. This is an advanced topic. See The Column Info Window chapter in the Using JMP book.

Modeling Type Modeling types define how values are used in analyses. Select one of the following modeling types:
- Continuous values are numeric only.
- Ordinal values are either numeric or character, and are ordered categories.
- Nominal values are either numeric or character, but not ordered.

Format       Select a format for numeric values. This option is not available for character data. Here are a few of the most common formats:
- Best lets JMP choose the best display format.
- Fixed Dec specifies the number of decimal places that appear.
- Date specifies the syntax for date values.
- Time specifies the syntax for time values.
- Currency specifies the type of currency and decimal points that are used for currency values.
Column Properties  Set special column properties such as formulas, notes, and value orders. See The Column Info Window chapter in the Using JMP book.

Lock  Lock a column, so that the values in the column cannot be changed.

Calculate Values with Formulas

Use the Formula Editor to create columns that contain calculated values.

Scenario

The sample data table On-Time Arrivals.jmp reflects the percent of on-time arrivals for several airlines. The data was collected for March, June, and August of 1999.

Create the Formula

Suppose that you want to create a new column containing the average on-time percentage for each airline.

1.  Add a new column.
2.  Right-click the column heading of the new column and select Formula. The Formula Editor window appears.
Create the formula for the average on-time percentage of each airline:

3. From the Columns list, select March 1999.
4. Click the $+$ button on the keypad.
5. Select June 1999, followed by another $+$ sign.

Figure 3.13  Sum of the Months

\[ \text{March 1999} + \text{June 1999} + \text{August 1999} \]

Notice that only August 1999 is selected (has the red box around it).

7. Click on the box surrounding the entire formula.

Figure 3.14  Entire Formula Selected

\[ \text{March 1999} + \text{June 1999} + \text{August 1999} \]

8. Click the $+$ button.
9. Type a 3 in the denominator box, and then click outside of the formula in any of the white space.

**Figure 3.15 Completed Formula**

\[
\frac{[\text{March 1999} + \text{June 1999} + \text{August 1999}]}{3}
\]

10. Click OK

The new column contains the averages.

The Formula Editor has many built-in arithmetic and statistical functions. For example, another way to calculate the average on-time arrival percentage is to use the Mean function in the Statistical functions list. For details about all of the Formula Editor functions, see the Formula Editor chapter in the *Using JMP* book.

**Filter Data**

Use the **Data Filter** to interactively select complex subsets of data, hide these subsets in plots, or exclude them from analyses. For example, look at profit per employee for computer and pharmaceutical companies.

1. Select **Help > Sample Data Library** and open **Companies.jmp**.
2. Select **Analyze > Distribution**.
3. Select **profit/emp** and click **Y, Columns**.
4. Click OK.
5. From the red triangle menu for **profit/emp**, select **Display Options > Horizontal Layout**.

**Figure 3.16 Distribution of profit/emp**

6. Turn on Automatic Recalc by selecting **Redo > Automatic Recalc** from the red triangle menu for **Distributions**.
Work with Your Data

Chapter 3
Manage Data

When this option is on, every change that you make (for example, hiding or excluding points) causes your report window to automatically update itself.

7. In the data table, select **Rows > Data Filter**.
8. Select **Type** and click **Add**.
9. Make sure that Select and Include are selected.
10. To filter out the Pharmaceutical companies from the Distribution results, and include only the Computer companies, click the **Computer** box in the Data Filter window.

   The distribution results update to only include Computer companies.

Figure 3.17 Filter for Computer Companies

Conversely, to change the Distribution results to include only the Pharmaceutical companies, click the **Pharmaceutical** button on the Data Filter window.

Manage Data

The commands on the **Tables** menu (and **Tabulate** on the **Analyze** menu) summarize and manipulate data tables into the format that you need for graphing and analyzing. This section describes five of these commands:

**Summary**  Creates a table that contains summary statistics that describe your data.
**Tabulate**  Provides a drag and drop workspace to create summary statistics.
**Subset**  Creates a table that contains a subset of your data.
**Join**  Joins the data from two data tables into one new data table.
Sort  Sorts your data by one or more columns.

For complete details about these and the other Tables menu commands, see the Reshape Data chapter in the Using JMP book.

**View Summary Statistics**

Summary statistics, such as sums and means, can instantly provide useful information about your data. For example, if you look at the annual profit of each company out of thirty-two companies, it’s difficult to compare the profits of small, medium, and large companies. A summary shows that information immediately.

Create summary tables by using either the Summary or Tabulate commands. The Summary command creates a new data table. As with any data table, you can perform analyses and create graphs from the summary table. The Tabulate command creates a report window with a table of summary data. You can also create a table from the Tabulate report.

**Summary**

A summary table contains statistics for each level of a grouping variable. For example, look at the financial data for computer and pharmaceutical companies. Suppose that you want to calculate the mean of sales and the mean of profits, for each combination of company type and size.

1. Select Help > Sample Data Library and open Companies.jmp.
2. Select Tables > Summary.
3. Select Type and Size Co and click Group.
4. Select Sales ($M) and Profits ($M) and click Statistics > Mean.
5. Click OK.

JMP calculates the mean of Sales ($M) and the mean of Profit ($M) for each combination of Type and Size Co.

The summary table contains the following:

- There are columns for each grouping variable (in this example, Type, and Size Co).
- The N Rows column shows the number of rows from the original table that correspond to each combination of grouping variables. For example, the original data table contains 14 rows corresponding to small computer companies.
There is a column for each summary statistic requested. In this example, there is a column for the mean of Sales ($M) and a column for the mean of Profits ($M).

The summary table is linked to the source table. Selecting a row in the summary table also selects the corresponding rows in the source table.

**Tabulate**

Use the Tabulate command to drag columns into a workspace, creating summary statistics for each combination of grouping variables. This example shows you how to use Tabulate to create the same summary information that you just created using Summary.

1. Select Help > Sample Data Library and open Companies.jmp.
2. Select Analyze > Tabulate.

**Figure 3.20 Tabulate Workspace**

3. Select both Type and Size Co.
4. Drag and drop them into the Drop zone for rows.
5. Right-click a heading and select Nest Grouping Columns.
   The initial tabulation shows the number of rows per group.

6. Select both Sales ($M) and Profits ($M), and drag and drop them over the N in the table.

The tabulation now shows the sum of Sales ($M) and the sum of Profits ($M) per group.
7. The final step is to change the sums to means. Right-click **Sum** (either of them) and select **Statistics > Mean**.

**Figure 3.25** Final Tabulation

![Figure 3.25](image)

The means are the same as those obtained using the Summary command. Compare Figure 3.25 to Figure 3.19.

**Create Subsets**

If you want to look closely at only part of your data table, you can create a subset. For example, suppose that you have already compared the sales and profits of big, medium, and small computer and pharmaceutical companies. Now you want to look at the sales and profits of only the medium-sized companies.

Creating a subset is a two-step process. First select the target data, and then extract the data into a new table.

**Subset with the Subset Command**

1. Select **Help > Sample Data Library** and open **Companies.jmp**.

**Selecting the Rows and Columns That You Want to Subset**

2. Select **Rows > Row Selection > Select Where**.
3. Select **Size Co** in the column list box on the left.
4. Enter medium in the text enter box.
5. Click **OK**.
6. Hold down the Ctrl key and select the **Type**, **Sales ($M)**, and **Profits ($M)** columns.

**Creating the Subset Table**

7. Select **Tables > Subset** to launch the Subset window.
8. Select **Selected columns** to subset only the columns that you selected. You can also customize your subset table further by selecting additional options.

9. Click **OK**.

The resulting subset data table has seven rows and three columns. For complete details about the Subset command, see the Reshape Data chapter in the *Using JMP* book.

**Subset with the Distribution Platform**

Another way to create subsets uses the connection between platform results and data tables.

**Example of Creating a Subset Using the Distribution Command**

1. Select **Help > Sample Data Library** and open *Companies.jmp*.
2. Select **Analyze > Distribution**.
3. Select **Type** and click **Y, Columns**.
4. Click **OK**.
5. Double-click on the histogram bar that represents Computer to create a subset table of the Computer companies.

**Caution:** This method creates a *linked* subset table. This means if you make any changes to the data in the subset table, the corresponding value changes in the source table.
Join Data Tables

Use the Join option to combine information from multiple data tables into a single data table. For example, suppose that you have a data table containing results from an experiment on popcorn yields. In another data table, you have the results of a second experiment on popcorn yields. To compare the two experiments or to analyze the trials using both sets of results, you need to have the data in the same table. Also, the experimental data was not entered into the data tables in the same order. One of the columns has a different name, and the second experiment is incomplete. This means that you cannot copy and paste from one table into another.

Example of Joining Two Data Tables
1. Select Help > Sample Data Library and open Trial1.jmp and Little.jmp.
2. Click on Trial1.jmp to make it the active data table.
3. Select Tables > Join.
4. In the Join ‘Trial1’ With box, select Little.
5. From the Matching Specification menu, select By Matching Columns.
6. In the Source Columns boxes, select popcorn in both boxes, and then click Match.
   - In the same way, match batch and oil amt to oil in both boxes.
   - Your matching columns do not have to have the same name.
7. Select Include non-matches for both tables.
   - Since one experiment is partial, you want to include all rows, including any with missing data.
8. To avoid duplicate columns, select the Select columns for joined table option.
9. From Trial1, select all four columns and click Select.
10. From Little, select only yield and click Select.
12. Click **OK**.

---

**Figure 3.27 Completed Join Window**

---

**Figure 3.28 Joined Table**
Sort Tables

Use the Sort command to sort a data table by one or more columns in the data table. For example, look at financial data for computer and pharmaceutical companies. Suppose that you want to sort the data table by Type, then by Profits ($M). Also, you want Profits ($M) to be in descending order within each Type.

1. Select Help > Sample Data Library and open Companies.jmp.
2. Select Tables > Sort.
3. Select Type and click By to assign Type as a sorting variable.
4. Select Profits ($M) and click By.

At this point, both variables are set to be sorted in ascending order. See the ascending icon next to the variables in Figure 3.29.

**Figure 3.29** Sort Ascending Icon

5. To change Profits ($M) to sort in descending order, select Profits ($M) and click the descending button.

**Figure 3.30** Change Profits to Descending

The icon next to Profits ($M) changes to descending.
6. Select the **Replace Table** check box.

   When selected, the **Replace Table** option tells JMP to sort the original data table instead of creating a new table with the sorted values. This option is not available if there are any open report windows created from the original data table. Sorting a data table with open report windows might change how some of the data is displayed in the report window, especially in graphs.

7. Click **OK**.

   The data table is now sorted by type alphabetically, and by descending profit totals within type.
Visualizing your data is an important first step. The graphs described in this chapter help you discover important details about your data. For example, histograms show you the shape and range of your data, and help you find unusual data points.

This chapter presents several of the most common graphs and plots that enable you to visualize and explore data in JMP. This chapter is an introduction to some of JMP’s graphical tools and platforms. Use JMP to visualize the distribution of single variables, or the relationships among multiple variables.

Figure 4.1 Visualizing Data with JMP
Analyse Single Variables

Single-variable graphs, or *univariate* graphs, let you look closely at one variable at a time. When you begin to look at your data, it’s important to learn about each variable before looking at how the variables interact with each other. Univariate graphs let you visualize each variable individually.

This section covers two graphs that show the distribution of a single variable:

- “Histograms” on page 90, for continuous variables
- “Bar Charts” on page 92, for categorical variables

Use the Distribution platform to create both of these graphs. Distribution produces a graphical description and descriptive statistics for each variable.

**Histograms**

The histogram is one of the most useful graphical tools for understanding the distribution of a continuous variable. Use a histogram to find the following in your data:

- the average value and variation
- extreme values

**Figure 4.2** Example of a Histogram

**Scenario**

This example uses the Companies.jmp data table, which contains data on profits for a group of companies.

A financial analyst wants to explore the following questions:

- Generally, how much profit does each company earn?
- What is the average profit?
Are there any companies that earn either extremely high or extremely low profits compared to the other companies?

To answer these questions, use a histogram of Profits ($M).

**Create the Histogram**

1. Select **Help > Sample Data Library** and open **Companies.jmp**.
2. Select **Analyze > Distribution**.
3. Select Profits ($M) and click Y, Columns.

**Figure 4.3 Distribution Window for Profits ($M)**

4. Click OK.

**Figure 4.4 Histogram of Profits ($M)**

**Interpret the Histogram**

The histogram provides these answers:

- Most companies’ profits are between $-1000 and $1500.

  All the bars except for one are located in this range. Also, more companies’ profits range from $0 to $500 than any other range. The bar representing that range is much longer than the others.
• The average profit is a little less than $500.

  The middle of the diamond in the box plot indicates the mean value. In this case, the mean is slightly lower than the $500 mark.

• One company has significantly higher profits than the others, and might be an outlier. An outlier is a data point that is separated from the general pattern of the other data points.

  This outlier is represented by a single, very short bar at the top of the histogram. The bar is small and represents a small group (in this case, a single company), and it is widely separated from the rest of the histogram bars.

In addition to the histogram, this report includes the following:

• The box plot, which is another graphical summary of the data. For detailed information about the box plot, see the Graph Builder chapter in the Essential Graphing book.

• Quantiles and Summary Statistics reports. These reports are discussed in “Analyze Distributions” on page 130 in the “Analyze Your Data” chapter.

Interact with the Histogram

Data tables and reports are all connected in JMP. Click on a histogram bar to select the corresponding rows in the data table.

Bar Charts

Use a bar chart to visualize the distribution of a categorical variable. A bar chart looks similar to a histogram, since they both have bars that correspond to the levels of a variable. A bar chart shows a bar for every level of the variable, whereas the histogram shows a range of values for the variable.

Figure 4.5 Example of a Bar Chart

Scenario

This example uses the Companies.jmp data table, which contains data on the size and type of a group of companies.
A financial analyst wants to explore the following questions:

- What is the most common type of company?
- What is the most common size for a company?

To answer these questions, use bar charts of Type and Size Co.

**Create the Bar Chart**

1. Select **Help > Sample Data Library** and open **Companies.jmp**.
2. Select **Analyze > Distribution**.
3. Select Type and Size Co and click **Y, Columns**.
4. Click **OK**.

**Figure 4.6** Bar Charts of Type and Size Co

**Interpret the Bar Charts**

The bar charts provide these answers:

- There are more computer companies than pharmaceutical companies. The bar that represents computer companies is larger than the bar that represents pharmaceutical companies.
- The most common company size is small. The bar that represents small companies is larger than the bars that represent medium and big companies.
The additional summary output gives detailed frequencies. This report is discussed in “Distributions of Categorical Variables” on page 133 in the “Analyze Your Data” chapter.

Interact with the Bar Charts

As is the case with histograms, click on individual bars to highlight rows of the data table. If more than one graph is created, clicking on a bar in one bar chart highlights the corresponding bar or bars in the other bar chart.

For example, suppose that you want to see the distribution of company size for the pharmaceutical companies. Click the Pharmaceutical bar in the Type bar chart, and the pharmaceutical companies are highlighted on the Size Co bar chart. Figure 4.7 shows that although most companies in this data table are small, most of the pharmaceutical companies are medium or big.

Also, the corresponding rows in the data table are selected.

Figure 4.7 Clicking Bars
Click on this bar to select the corresponding data in the other chart.

Compare Multiple Variables

Use multiple-variable graphs to visualize the relationships and patterns between two or more variables. This section covers the following graphs:

Table 4.1 Multiple-Variable Graphs

| “Scatterplots” on page 95 | Use scatterplots to compare two continuous variables. |
Chapter 4
Discovering JMP
Visualize Your Data
Compare Multiple Variables

Table 4.1 Multiple-Variable Graphs (Continued)

<table>
<thead>
<tr>
<th>Method</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Scatterplot Matrix” on page 99</td>
<td>Use scatterplot matrices to compare several pairs of continuous variables.</td>
</tr>
<tr>
<td>“Side-by-Side Box Plots” on page 102</td>
<td>Use side-by-side box plots to compare one continuous and one categorical variable.</td>
</tr>
<tr>
<td>“Overlay Plots” on page 105</td>
<td>Use overlay plots to compare one or more variables on the Y-axis to another variable on the X-axis. Overlay plots are especially useful if the X variable is a time variable, because you can compare how two or more variables change across time.</td>
</tr>
<tr>
<td>“Variability Chart” on page 108</td>
<td>Use variability charts to compare one continuous Y variable to one or more categorical X variables. Variability charts show differences in means and variability across several categorical X variables.</td>
</tr>
<tr>
<td>“Graph Builder” on page 111</td>
<td>Use Graph Builder to create and change graphs interactively.</td>
</tr>
<tr>
<td>“Bubble Plots” on page 117</td>
<td>Bubble plots are specialized scatterplots that use color and bubble sizes to represent up to five variables at once. If one of your variables is a time variable, you can animate the plot to see your other variables change through time.</td>
</tr>
</tbody>
</table>

Scatterplots

The scatterplot is the simplest of all the multiple-variable graphs. Use scatterplots to determine the relationship between two continuous variables and to discover whether two continuous variables are correlated. Correlation indicates how closely two variables are related. When you have two variables that are highly correlated, one might influence the other. Or, both might be influenced by other variables in a similar way.
Scenario

This example uses the Companies.jmp data table, which contains sales figures and the number of employees of a group of companies. A financial analyst wants to explore the following questions:

- What is the relationship between sales and the number of employees?
- Does the amount of sales increase with the number of employees?
- Can you predict average sales from the number of employees?

To answer these questions, use a scatterplot of Sales ($M) versus # Employ.

Create the Scatterplot

1. Select Help > Sample Data Library and open Companies.jmp.
2. Select Analyze > Fit Y by X.
3. Select Sales ($M) and Y, Response.
4. Select # Employ and X, Factor.
Figure 4.9  Fit Y by X Window

5. Click OK.

Figure 4.10  Scatterplot of Sales ($M) versus # Employ

Interpret the Scatterplot

One company has a large number of employees and high sales, represented by the single point at the top right of the plot. The distance between this data point and all the rest makes it difficult to visualize the relationship between the rest of the companies. Remove the point from the plot and re-create the plot by following these steps:

1. Click on the point to select it.
2. Select Rows > Hide and Exclude. The data point is hidden and no longer included in calculations.
Note: The difference between hiding and excluding is important. Hiding a point removes it from any graphs but statistical calculations continue to use the point. Excluding a point removes it from any statistical calculations but does not remove it from graphs. When you both hide and exclude a point, you remove it from all calculations and from all graphs.

3. To re-create the plot without the outlier, select Redo > Redo Analysis from the red triangle menu for Bivariate. You can close the original report window.

Figure 4.11 Scatterplot with the Outlier Removed

The updated scatterplot provides these answers:

- There is a relationship between the sales and the number of employees.
  The data points have a discernible pattern. They are not scattered randomly throughout the graph. You could draw a diagonal line that would be near most of the data points.

- Sales do increase with the number of employees, and the relationship is linear.
  If you drew that diagonal line, it would slope from bottom left to top right. This slope shows that as the number of employees increases (left to right on the bottom axis), sales also increases (bottom to top on the left axis). A straight line would be near most of the data points, indicating a linear relationship. If you would have to curve your line to be near the data points, there would still be a relationship (because of the pattern of the points). However, that relationship would not be linear.

- You can predict average sales from the number of employees.
  The scatterplot shows that sales generally increase as the number of employees does. You could predict the sales for a company if you knew only the number of employees of that company. Your prediction would be on that imaginary line. It would not be exact, but it would approximate the real sales.
Interact with the Scatterplot

As with other JMP graphics, the scatterplot is interactive. Place your mouse pointer over the point in the bottom right corner with the mouse to reveal the row number and the x and y values.

Figure 4.12 Place Your Mouse Pointer Over a Point

Click on a point to highlight the corresponding row in the data table. Select multiple points by doing one of the following:

- Click and drag with the mouse around the points. This selects points in a rectangular area.
- Select the lasso tool, and then click and drag around multiple points. The lasso tool selects an irregularly shaped area.

Scatterplot Matrix

A scatterplot matrix is a collection of scatterplots organized into a grid (or matrix). Each scatterplot shows the relationship between a pair of variables.
**Figure 4.13** Example of a Scatterplot Matrix

**Scenario**

This example uses the Solubility.jmp data table, which contains data for solubility measurements for 72 different solutes.

A lab technician wants to explore the following questions:

- Is there a relationship between any pair of chemicals? (There are six possible pairs.)
- Which pair has the strongest relationship?

To answer these questions, use a scatterplot matrix of the four solvents.

**Create the Scatterplot Matrix**

1. Select **Help > Sample Data Library** and open Solubility.jmp.
2. Select **Graph > Scatterplot Matrix**.
3. Select Ether, Chloroform, Benzene, and Hexane, and click **Y, Columns**.
Chapter 4
Visualize Your Data

Discovering JMP
Compare Multiple Variables

Figure 4.14 Scatterplot Matrix Window

4. Click OK.

Figure 4.15 Scatterplot Matrix

Interpret the Scatterplot Matrix

The scatterplot matrix provides these answers:

- All six pairs of variables are positively correlated. As one variable increases, the other variable increases too.
- The strongest relationship appears to be between Benzene and Chloroform. The data points in the scatterplot for Benzene and Chloroform are the most tightly clustered along an imaginary line.
Interact with the Scatterplot Matrix

If you select a point in one scatterplot, it is selected in all the other scatterplots. For example, if you select a point in the Benzene versus Chloroform scatterplot, the same point is selected in the other five plots.

Figure 4.16 Selected Points

Side-by-Side Box Plots

Side-by-side box plots show the following:

- the relationship between one continuous variable and one categorical variable
- differences in the continuous variable across levels of the categorical variable
**Scenario**

This example uses the Analgesics.jmp data table, which contains data on pain measurements taken on patients using three different drugs.

A researcher wants to explore the following questions:

- Are there differences in the average amount of pain control among the drugs?
- Does the *variability* in the pain control given by each drug differ? A drug with high variability would not be as reliable as a drug with low variability.

To answer these questions, use a side-by-side box plot for the pain levels and the drug categories.

**Create the Side-by-Side Box Plots**

1. Select Help > Sample Data Library and open Analgesics.jmp.
2. Select Analyze > Fit Y by X.
4. Select drug and click X, Factor.
Figure 4.18 Fit Y by X Window

5. Click OK.
6. From the red triangle menu, select Display Options > Box Plots.

Figure 4.19 Side-by-Side Box Plots

Interpret the Side-by-Side Box Plots

Box plots are designed according to the following principles:

- The line through the box represents the median.
- The middle half of the data is within the box.
- The majority of the data falls between the ends of the whiskers.
- A data point outside the whiskers might be an outlier.

The box plots in Figure 4.19 show these answers:

- There is evidence to believe that patients on drug A feel less pain, since the box plot for drug A is lower on the pain scale than the others.
- Drug B appears to have higher variability than Drugs A and C, since the box plot is taller.
There is one point for drug C that is a lot lower than the other points for drug C. Place your mouse pointer over it with your mouse to see that it is row 26 of the data table. That point looks like it is more similar to the data in drug group A or B. The information in row 26 deserves investigation. There might have been a typographical error when the data was recorded.

**Overlay Plots**

Like scatterplots, overlay plots show the relationship between two or more variables. However, if one of the variables is a time variable, an overlay plot shows trends across time better than scatterplots do.

**Figure 4.20** Example of an Overlay Plot

![Overlay Plot](image)

**Note:** To plot data over time, you can also use Graph Builder, bubble plots, control charts, and variability charts. For complete details about Graph Builder and bubble plots, see the Graph Builder chapter in the *Essential Graphing* book. Refer to the Control Chart Builder chapter and the Variability Gauge Charts chapter in the *Quality and Process Methods* book for information about control charts and variability charts.

**Scenario**

This example uses the Stock Prices.jmp data table, which contains data on the price of a stock over a three-month period.

A potential investor wants to explore the following questions:

- Has the stock’s closing price changed over the past three months?
  
  To answer this question, use an overlay plot of the stock’s closing price over time.

- How do the stock’s high and low prices relate to each other?
To answer this question, use another overlay plot of the stock’s high and low prices over time.

Create the first overlay plot to answer the first question, and then create a second overlay plot to answer the second question.

Create the Overlay Plot of the Stock’s Price over Time

1. Select Help > Sample Data Library and open Stock Prices.jmp.
2. Select Graph > Overlay Plot.
3. Select Close and click Y.
4. Select Date and click X.

Figure 4.21 Overlay Plot Window

5. Click OK.

Figure 4.22 Overlay Plot of the Closing Price over Time
Interpret and Interact with the Overlay Plot

The overlay plot shows that the closing stock price has been decreasing over the last several months. To see the trend more clearly, connect the points and add grid lines.

1. From the red triangle menu, select Connect Thru Missing.
2. Double-click the Y axis.
3. Select the Major Grid Lines check box.
4. Click OK.

Figure 4.23 Connected Points and Grid Lines

The potential investor can see that although the stock price has gone up and down over the past three months, the overall trend has been downward.

Create the Overlay Plot of the Stock’s High and Low Prices

Use an overlay plot to plot more than one Y variable. For example, suppose that you want to see both the high and the low prices on the same plot.

1. Follow the steps in “Create the Overlay Plot of the Stock’s Price over Time” on page 106, this time assigning both High and Low to the Y role.
2. Connect the points and add grid lines as shown in “Interpret and Interact with the Overlay Plot” on page 107.
**Figure 4.24 Two Y Variables**

The legend at the bottom of the plot shows the colors and markers used for the High and Low variables in the graph. The overlay plot shows that the High price and Low price track each other very closely.

**Answer the Questions**

Both of the overlay plots answer the two questions asked at the beginning of this example.

- The first plot shows that the price of this stock has not remained the same, but has been decreasing.
- The second plot shows that the high and low prices of this stock are not very different from each other. The stock price does not vary wildly on any given day.

**Variability Chart**

In the graphs described so far, you specified only a single X variable. Use a variability chart to specify multiple X variables and see differences in means and variability across all of your variables at once.
**Scenario**

This example uses the Popcorn.jmp data table with data from a popcorn maker. The yield (the volume of popcorn for a given measure of kernels) was measured for each combination of popcorn style, batch size, and amount of oil used.

The popcorn maker wants to explore the following question:

- Which combination of factors results in the highest popcorn yield?

To answer this question, use a variability chart of the yield versus the style, batch size, and oil amount.

**Create the Variability Chart**

1. Select Help > Sample Data Library and open Popcorn.jmp.
2. Select Analyze > Quality and Process > Variability/Attribute Gauge Chart.
4. Select popcorn and click X, Grouping.
5. Select batch and click X, Grouping.
6. Select oil amt and click X, Grouping.

**Note:** The order in which you assign the variables to the X, Grouping role is important, because the order in this window determines their nesting order in the variability chart.
7. Click OK.

The top chart is the variability chart, showing the yield broken down by each combination of the three variables. The bottom chart shows the standard deviation for each combination of the three variables. Since the bottom chart does not show the yield, hide it.

8. Deselect Std Dev Chart on the red triangle menu.

**Figure 4.27** Results Window

Interpret the Variability Chart

The variability chart for yield indicates that small, gourmet batches produce the highest yield.
To be more specific, the popcorn maker might ask this additional question: Is the yield high because those batches are small, or because those batches are gourmet?

The variability chart shows the following:

- The yield from small, plain batches is low.
- The yield from large, gourmet batches is low.

Given this information, the popcorn maker can conclude that only the combination of small and gourmet at the same time results in batches with high yield. It would have been impossible to reach this conclusion with a chart that only allowed a single variable.

**Graph Builder**

Use Graph Builder to interactively create and modify graphs. So far, all of the graphs have been created by launching a platform and specifying variables. To create a different type of graph, you must launch a different platform. In Graph Builder, you can change the variables and change the graphs at any time.

Use Graph Builder to accomplish the following tasks:

- Change variables by dragging and dropping them in and out of the graph.
- Create a different type of graph with a few mouse clicks.
- Partition the graph horizontally or vertically.
Figure 4.28 Example of a Graph That Was Created with Graph Builder

Note: Only some of the Graph Builder features are covered here. For complete details, see the Graph Builder chapter in the Essential Graphing book.

Scenario

This example uses the Profit by Product.jmp data table, which contains profit data for multiple product lines.

A business analyst wants to explore the following question:

• How is the profitability different between product lines?

To answer this question, use a line plot that displays revenue, product cost, and profit data across different product lines.

Create the Graph

1. Select Help > Sample Data Library and open Profit by Product.jmp.
2. Select Graph > Graph Builder.
3. Click Quarter and then drag and drop it onto the X zone to assign Quarter as the X variable.

4. Click Revenue, Product Cost, and Profit, and drag and drop them onto the Y zone to assign all three variables as Y variables.

   The X and Y zones are now axes.

**Note:** You can also click on variables and then click a zone to assign them. However, after a zone becomes an axis, drag and drop additional variables onto the axis rather than clicking on the variables and axis.
Figure 4.30  After Adding Y and X Variables

Based on the variables that you are using, Graph Builder shows side-by-side box plots.

5. To change the box plots to a line plot, click the Line icon.
6. To create a separate chart for each product, click Product Line, and drag and drop it into the Wrap zone.

A separate line plot is created for each product.
Interpret the Graph

Figure 4.32 shows revenue, cost, and profit broken down by product line. The business analyst was interested in seeing the difference in profitability between product lines. The line plots in Figure 4.32 can provide some answers, as follows:

- Credit products, deposit products, and revolving credit products produce more revenue than fee-based products, third-party products, and other products.
- However, the profits of all the product lines are similar.

The data table also includes data on sales channels. The business analyst wants to see how revenue, product cost, and profit differ between different sales channels.

1. To remove Product Line from the graph, click the title of the graph (Product Line) and drag and drop it into any empty area within Graph Builder.

2. To add Channel as the wrap variable, click Channel and drag and drop it into the Wrap zone.
Figure 4.33 Line Plots Showing Sales Channels

Figure 4.33 provides this answer: revenue and product cost for ATMs are the highest and are growing the most quickly.

**Bubble Plots**

A bubble plot is a scatterplot that represents its points as bubbles. You can change the size and color of the bubbles, and even animate them over time. With the ability to represent up to five dimensions (x position, y position, size, color, and time), a bubble plot can produce dramatic visualizations and make data exploration easy.
Figure 4.34  Example of a Bubble Plot

Scenario

This example uses the PopAgeGroup.jmp data table, which contains population statistics for 116 countries or territories between the years 1950 to 2004. Total population numbers are broken out by age group, and not every country has data for every year.

A sociologist wants to explore the following question:

- Is the age of the population of the world changing?

To answer this question, look at the relationship between the oldest (more than 59) and the youngest (younger than 20) portions of the population. Use a bubble plot to determine how this relationship changes over time.

Create the Bubble Plot

1. Select Help > Sample Data Library and open PopAgeGroup.jmp.
2. Select Graph > Bubble Plot.
3. Select Portion60+ and click Y.
   This corresponds to the Y variable on the bubble plot.
4. Select Portion 0-19 and click X.
This corresponds to the X variable on the bubble plot.

5. Select Country and click ID.

Each unique level of the ID variable is represented by a bubble on the plot.

6. Select Year and click Time.

This controls the time indexing when the bubble plot is animated.

7. Select Pop and click Sizes.

This controls the size of the bubbles.

8. Select Region and click Coloring.

Each level of the Coloring variable is assigned a unique color. So in this example, all the bubbles for countries located in the same region have the same color. The bubble colors that appear in Figure 4.36 are the JMP default colors.

Figure 4.35  Bubble Plot Window

9. Click OK.
**Figure 4.36** Initial Bubble Plot

Interpret the Bubble Plot

Because the time variable (in this case, year) starts in 1950, the initial bubble plot shows the data for 1950. Animate the bubble plot to cycle through all the years by clicking the play/pause button. Each successive bubble plot shows the data for that year. The data for each year determines the following:

- The X and Y coordinates
- The bubble’s sizes
- The bubble’s coloring
- Bubble aggregation

**Note:** For detailed information about how the bubble plot aggregates information across multiple rows, see the Bubble Plots chapter in the *Essential Graphing* book.

The bubble plot for 1950 shows that if a country’s proportion of people younger than 20 is high, then the proportion of people more than 59 is low.
Click the play/pause button to animate the bubble plot through the range of years. As time progresses, the Portion 0-19 decreases and the Portion 60+ increases.

- plays the animation, turns to a pause button after you click it.
- pauses the animation.
- manually controls the animation back one unit of time.
- manually controls the animation forward one unit of time.

**Year** is used to change the time index manually.

**Speed** controls the speed of the animation.

**Bubble Size** controls the absolute sizes of the bubbles, while maintaining the relative sizes.

The sociologist wanted to know how the age of the world’s population is changing. The bubble plot indicates that the population of the world is getting older.

**Interact with the Bubble Plot**

Click to select a bubble to see the trend for that bubble over time. For example, in the 1950 plot, the large bubble in the middle is Japan.

**To See the Pattern of Population Changes in Japan through the Years**

1. Click in the middle of the Japan bubble to select it.
2. From the red triangle menu, select **Trail Bubbles > Selected**.
3. Click the play button.

As the animation progresses through time, the Japan bubble leaves a trail of bubbles that illustrates its history.
**Figure 4.37** Japan’s History of Population Shifts

Focusing on the Japan bubble, you can see the following over time:

- The proportion of the population 19 years old or less decreased.
- The proportion of the population 60 years old or more increased.
Analyzing your data helps you make informed decisions. Data analysis often involves these actions:

- Examining distributions
- Discovering relationships
- Hypothesis testing
- Building models

**Figure 5.1 Analysis Examples**
About This Chapter

Before you analyze your data, review the following information:

- “The Importance of Graphing Your Data” on page 124
- “Understand Modeling Types” on page 126

The rest of this chapter shows you how to use some basic analytical methods in JMP:

- “Analyze Distributions” on page 130
- “Analyze Relationships” on page 136

For a description of advanced modeling and analysis techniques, refer to the following JMP books:

- Fitting Linear Models
- Multivariate Methods
- Predictive and Specialized Modeling
- Consumer Research
- Reliability and Survival Methods
- Quality and Process Methods

The Importance of Graphing Your Data

Graphing, or visualizing, your data is important to any data analysis, and should always occur before the use of statistical tests or model building. To illustrate why data visualization should be an early step in your data analysis process, consider the following example:


   This data consists of four pairs of X and Y variables.

2. In the Table panel, click the green triangle next to the The Quartet script.

   The script creates a simple linear regression on each pair of variables using Fit Y by X. The Show Points option is turned off, so that none of the data can be seen on the scatterplots. Figure 5.2 shows the model fit and other summary information for each regression.
Notice that all four models and the RSquare values are nearly identical. The fitted model in each case is essentially $Y = 3 + 0.5X$, and the RSquare value in each case is essentially 0.66. If your data analysis took into account only the above summary information, you would likely conclude that the relationship between $X$ and $Y$ is the same in each case. However, at this point, you have not visualized your data. Your conclusion might be wrong.

**To Visualize the Data, Add the Points to All Four Scatterplots**

1. Hold down the Ctrl key.
2. From the red triangle menu for any one of the Bivariate Fits, select **Show Points**.
The scatterplots show that the relationship between $X$ and $Y$ is not the same for the four pairs, although the lines describing the relationships are the same:

- Plot 1 represents a linear relationship.
- Plot 2 represents a non-linear relationship.
- Plot 3 represents a linear relationship, except for one outlier.
- Plot 4 has all the data at $x = 8$, except for one point.

This example illustrates that conclusions that are based on statistics alone can be inadequate. A visual exploration of the data should be an early part of any data analysis.

**Understand Modeling Types**

In JMP, data can be of different types. JMP refers to this as the modeling type of the data. Table 5.1 describes the three modeling types in JMP.
Different modeling types produce different results in JMP. To see an example of the differences, follow these steps:

1. Select Help > Sample Data Library and open Linnerud.jmp.
2. Select Analyze > Distribution.
3. Select Age and Weight and click Y, Columns.
4. Click OK.

### Table 5.1 Modeling Types

<table>
<thead>
<tr>
<th>Modeling Type and Description</th>
<th>Examples</th>
<th>Specific Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous</td>
<td>Height</td>
<td>The time to complete a test might be 2 hours, or 2.13 hours.</td>
</tr>
<tr>
<td></td>
<td>Temperature</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Time</td>
<td></td>
</tr>
<tr>
<td>Ordinal</td>
<td>Month (1,2,...,12)</td>
<td>The month of the year can be 2 (February) or 3 (March), but not 2.13. February comes before March.</td>
</tr>
<tr>
<td></td>
<td>Letter grade (A, B,...F)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Size (small, medium, large)</td>
<td></td>
</tr>
<tr>
<td>Nominal</td>
<td>Gender (M or F)</td>
<td>The gender can be M or F, with no order. Gender categories can also be represented by a number (M=1 and F=2).</td>
</tr>
<tr>
<td></td>
<td>Color</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Test result (pass or fail)</td>
<td></td>
</tr>
</tbody>
</table>

**Example of Viewing Modeling Type Results**

Different modeling types produce different results in JMP. To see an example of the differences, follow these steps:

1. Select Help > Sample Data Library and open Linnerud.jmp.
2. Select Analyze > Distribution.
3. Select Age and Weight and click Y, Columns.
4. Click OK.
Although *Age* and *Weight* are both numeric variables, they are not treated the same. Table 5.2 compares the differences between the results for *weight* and *age*.

### Table 5.2 Results for weight and age

<table>
<thead>
<tr>
<th>Variable</th>
<th>Modeling Type</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Weight</td>
<td>Continuous</td>
<td>Histogram, Quantiles, and Summary Statistics</td>
</tr>
<tr>
<td>Age</td>
<td>Ordinal</td>
<td>Bar chart and Frequencies</td>
</tr>
</tbody>
</table>

### Change the Modeling Type

To treat a variable differently, change the modeling type. For example, in Figure 5.4, the modeling type for *Age* is ordinal. Remember that for an ordinal variable, JMP calculates frequency counts. Suppose that you wanted to find the average age instead of frequency counts. Change the modeling type to continuous, which shows the mean age.
1. Double-click the Age column heading. The Column Info window appears.
2. Change the Modeling Type to Continuous.

**Figure 5.5 Column Info Window**

3. Click OK.

4. Repeat the steps in the example (see “Example of Viewing Modeling Type Results” on page 127) to create the distribution. Figure 5.6 shows the distribution results when Age is ordinal and continuous.
When age is ordinal, you can see the frequency counts for each age. For example, age 48 appears two times. When age is continuous, you can find the mean age, which is nearly 48 (47.677)

**Analyze Distributions**

To analyze a single variable, you can examine the distribution of the variable, using the Distribution platform. Report content for each variable varies, depending on whether the variable is categorical (nominal or ordinal) or continuous.

**Note:** For complete details about the Distribution platform, see the Distributions chapter in the Basic Analysis book.

**Distributions of Continuous Variables**

Analyzing a continuous variable might include questions such as the following:
Analyze Your Data

Chapter 5
Discovering JMP

Analyze Distributions

131

Does the shape of the data match any known distributions?
Are there any outliers in the data?
What is the average of the data?
Is the average statistically different from a target or historical value?
How spread out are the data? In other words, what is the standard deviation?
What are the minimum and maximum values?

You can answer these and other questions with graphs, summary statistics, and simple statistical tests.

Scenario

This example uses the Car Physical Data.jmp data table, which contains information about 116 different car models.

A planning specialist has been asked by a railroad company to determine the possible issues involved in transporting cars by train. Using the data, the planning specialist wants to explore the following questions:

• What is the average car weight?
• How spread out are the cars’ weights (standard deviation)?
• What are the minimum and maximum weights of cars?
• Are there any outliers in the data?

Use a histogram of weight to answer these questions.

Create the Histogram

1. Select Help > Sample Data Library and open Car Physical Data.jmp.
2. Select Analyze > Distribution.
4. Click OK.
5. To rotate the report window, select Display Options > Horizontal Layout from the red triangle menu next to Weight.
The report window contains three sections:

- A histogram and a box plot to visualize the data.
- A Quantiles report that shows the percentiles of the distribution.
- A Summary Statistics report that shows the mean, standard deviation, and other statistics.

**Interpret the Distribution Results**

Using the results presented in Figure 5.7, the planning specialist can answer the questions.

**What is the average car weight?** The Histogram shows a weight of around 3,000 lbs.

**How spread out are the weights (standard deviation)?** The Summary Statistics show a weight of around 2,958 lbs. The Summary Statistics show a standard deviation of around 536 lbs.

**What are the minimum and maximum weights?** The Histogram shows a minimum of around 1,500 lbs. and a maximum of around 4,500 lbs. The Quantiles show a minimum of around 1,695 lbs. and a maximum of around 4,285 lbs.

**Are there any outliers?** No.

The default report window in Figure 5.7 provides a minimal set of graphs and statistics. Additional graphs and statistics are available on the red triangle menu.

**Draw Conclusions**

Based on other research, the railroad company has determined that an average weight of 3000 pounds is the most efficient to transport. Now, the planning specialist needs to find out whether the average car weight in the general population of cars that they might transport is 3000 pounds. Use a t-test to draw inferences about the broader population based on this sample of the population.
Test Conclusions

1. From the red triangle menu for Weight, select Test Mean.
2. In the window that appears, type 3000 in the Specify Hypothesized Mean box.
3. Click OK.

Figure 5.8 Test Mean Results

Interpret the t-Test

The primary result of a t-test is the p-value. In this example, the p-value is 0.396 and the analyst is using a significance level of 0.05. Since 0.396 is greater than 0.05, you cannot conclude that the average weight of car models in the broader population is significantly different from 3000 pounds. Had the p-value been lower than the significance level, the planning specialist would have concluded that the average car weight in the broader population is significantly different from 3000 pounds.

Distributions of Categorical Variables

Analyzing a categorical (ordinal or nominal) variable might include questions such as the following:

- How many levels does the variable have?
- How many data points does each level have?
- Is the data uniformly distributed?
- What proportions of the total do each level represent?

Scenario

See the scenario in “Distributions of Continuous Variables” on page 130.
Now that the railroad company has determined that the average weight of the cars is not significantly different from the target weight, there are more questions to address.

The planning specialist wants to answer these questions for the railroad company:

- What are the types of cars?
- What are the countries of origin?

To answer these questions, look at the distribution for Type and Country.

**Create the Distribution**

1. Select Help > Sample Data Library and open Car Physical Data.jmp.
2. Select Analyze > Distribution.
3. Select Country and Type and click Y, Columns.
4. Click OK.

**Figure 5.9 Distribution for Country and Type**

**Interpret the Distribution Results**

The report window includes a bar chart and a Frequencies report for Country and Type. The bar chart is a graphical representation of the frequency information provided in the Frequencies report. The Frequencies report contains the following:

- Categories of data. For example, Japan is a category of Country, and Sporty is a category of Type.
• Total counts for each category.
• Proportion of the total each category represents.

For example, there are 22 compact cars, or about 19% of the 116 observations.

**Interact with the Distribution Results**

Selecting a bar in one chart also selects the corresponding data in the other chart. For example, select the Japan bar in the Country bar chart to see that a large number of Japanese cars are sporty.

**Figure 5.10  Japanese Cars**

Select the Other category to see that a majority of these cars are small or compact, and almost none are large.

**Figure 5.11  Other Cars**
Analyze Relationships

Scatterplots and other such graphs can help you visualize relationships between variables. Once you have visualized relationships, the next step is to analyze those relationships so that you can describe them numerically. That numerical description of the relationship between variables is called a model. Even more importantly, a model also predicts the average value of one variable (Y) from the value of another variable (X). The X variable is also called a predictor. Generally, this model is called a regression model.

With JMP, the Fit Y by X platform and the Fit Model platform creates regression models.

Note: Only the basic platforms and options are covered here. For complete details and explanations of all platform options, see Basic Analysis, Essential Graphing, and the books listed in “About This Chapter” on page 124.

Table 5.3 shows the four primary types of relationships.

<table>
<thead>
<tr>
<th>X</th>
<th>Y</th>
<th>Section</th>
</tr>
</thead>
<tbody>
<tr>
<td>Continuous</td>
<td>Continuous</td>
<td>• “Use Regression with One Predictor” on page 136</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “Use Regression with Multiple Predictors” on page 150</td>
</tr>
<tr>
<td>Categorical</td>
<td>Continuous</td>
<td>• “Compare Averages for One Variable” on page 141</td>
</tr>
<tr>
<td></td>
<td></td>
<td>• “Compare Averages for Multiple Variables” on page 146</td>
</tr>
<tr>
<td>Categorical</td>
<td>Categorical</td>
<td>“Compare Proportions” on page 144</td>
</tr>
<tr>
<td>Continuous</td>
<td>Categorical</td>
<td>Logistic regression is an advanced topic. See the Logistic Analysis chapter in the Basic Analysis book.</td>
</tr>
</tbody>
</table>

Use Regression with One Predictor

Scenario

This example uses the Companies.jmp data table, which contains financial data for 32 companies from the pharmaceutical and computer industries.
Intuitively, it makes sense that companies with more employees can generate more sales revenue than companies with fewer employees. A data analyst wants to predict the overall sales revenue for each company based on the number of employees.

To accomplish this task, do the following:

- “Discover the Relationship” on page 137
- “Fit the Regression Model” on page 137
- “Predict Average Sales” on page 139

**Discover the Relationship**

First, create a scatterplot to see the relationship between the number of employees and the amount of sales revenue. This scatterplot was created in “Create the Scatterplot” on page 96 in the “Visualize Your Data” chapter. After hiding and excluding one outlier (a company with significantly more employees and higher sales), the plot in Figure 5.12 shows the result.

![Figure 5.12 Scatterplot of Sales ($M) versus # Employ](image)

This scatterplot provides a clearer picture of the relationship between sales and the number of employees. As expected, the more employees a company has, the higher sales that it can generate. This visually confirms the data analyst’s guess, but it does not predict sales for a given number of employees.

**Fit the Regression Model**

To predict the sales revenue from the number of employees, fit a regression model. From the red triangle for Bivariate Fit, select Fit Line. A regression line is added to the scatterplot and reports are added to the report window.
Within the reports, look at the following results:

- the p-value of <.0001
- the RSquare value of 0.618

From these results, the data analyst can conclude the following:

- The p-value is less than the significance level of 0.05. Therefore, including the number of employees in the prediction model significantly improves the ability to predict average sales.
- Since the RSquare value in this example is large, this confirms that a prediction model based on the number of employees can predict sales revenue. The RSquare value shows the strength of a relationship between variables, also called the correlation. A correlation of 0 indicates no relationship between the variables, and a correlation of 1 indicates a perfect linear relationship.
Predict Average Sales

Use the regression model to predict the average sales a company might expect if they have a certain number of employees. The prediction equation for the model is included in the report:

\[
\text{Average sales} = 1059.68 + 0.092 \times \text{employees}
\]

For example, in a company with 70,000 employees, the equation is as follows:

\[
7499.68 = 1059.68 + 0.092 \times 70000
\]

In the lower right area of the current scatterplot, there is an outlier that does not follow the general pattern of the other companies. The data analyst wants to know whether the prediction model changes when this outlier is excluded.

Exclude the Outlier

1. Click on the outlier.
2. Select \textit{Rows > Exclude/Unexclude}.
3. Fit this model by selecting \textit{Fit Line} from the red triangle menu for \textit{Bivariate Fit}.

The following are added to the report window (see Figure 5.14):

- a new regression line
- a new Linear Fit report, which includes:
  - a new prediction equation
  - a new RSquare value
Interpret the Results

Using the results in Figure 5.14, the data analyst can make the following conclusions:

- The outlier was pulling down the regression line for the larger companies, and pulling the line up for the smaller companies.
- The new model fits the data better, since the new RSquare value (0.88) is closer to 1 than the first RSquare value (0.618).
**Draw Conclusions**

Using the new prediction equation, the predicted average sales for a company with 70,000 employees can be calculated as follows:

\[ \$8961.37 = 631.37 + 0.119 \times 70,000 \]

The prediction for the first model was $7499.68, so this model predicts a higher sales total by $1461.69.

The second model, after removing the outlier, describes and predicts sales totals based on the number of employees better than the first model. The data analyst now has a good model to use.

**Compare Averages for One Variable**

If you have a continuous Y variable, and a categorical X variable, you can compare averages across the levels of the X variable.

**Scenario**

This example uses the Companies.jmp data table, which contains financial data for 32 companies from the pharmaceutical and computer industries.

A financial analyst wants to explore the following question:

- How do the profits of computer companies compare to the profits of pharmaceutical companies?

To answer this question, fit Profits ($M) by Type.

**Discover the Relationship**

1. Select Help > Sample Data Library and open Companies.jmp.
2. If you still have the Companies.jmp sample data table open, you might have rows that are excluded or hidden. To return the rows to the default state (all rows included and none hidden), select Rows > Clear Row States.
3. Select Analyze > Fit Y by X.
4. Select Profits ($M) and click Y, Response.
5. Select Type and click X, Factor.
6. Click OK.
Figure 5.15  Profits by Company Type

There is an outlier in the Computer Type. The outlier is stretching the scale of the plot and making it difficult to compare the profits. Exclude and hide the outlier:

1. Click on the outlier.
2. Select **Rows > Exclude/Unexclude**. The data point is no longer included in calculations.
3. Select **Rows > Hide/Unhide**. The data point is hidden from all graphs.
4. To re-create the plot without the outlier, select **Redo > Redo Analysis** from the red triangle menu for Oneway Analysis. You can close the original Scatterplot window.

Figure 5.16  Updated Plot

Removing the outlier gives the financial analyst a clearer picture of the data.

5. To continue analyzing the relationship, select these options from the red triangle menu for Oneway Analysis:
   - **Display Options > Mean Lines**. This adds mean lines to the scatterplot.
– **Means and Std Dev.** This displays a report that provides averages and standard deviations.

**Figure 5.17 Mean Lines and Report**

![Oneway Analysis of Profits ($M) By Type](image)

**Means and Std Deviations**

<table>
<thead>
<tr>
<th>Level</th>
<th>Number</th>
<th>Mean</th>
<th>Std Dev</th>
<th>Std Err</th>
<th>Mean</th>
<th>Lower 95%</th>
<th>Upper 95%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Computer</td>
<td>19</td>
<td>55.753</td>
<td>357.009</td>
<td>82.11</td>
<td>-116.8</td>
<td>228.25</td>
<td></td>
</tr>
<tr>
<td>Pharmaceutical</td>
<td>12</td>
<td>690.075</td>
<td>429.988</td>
<td>124.13</td>
<td>416.9</td>
<td>963.28</td>
<td></td>
</tr>
</tbody>
</table>

**Interpret the Results**

The financial analyst wanted to know how the profits of computer companies compared to the profits of pharmaceutical companies. The updated scatterplot shows that pharmaceutical companies have higher average profits than computer companies. In the report, if you subtract one mean value from the other, the difference in profit is about $635 million. The plot also shows that some of the computer companies have negative profits and all of the pharmaceutical companies have positive profits.

**Perform the T-test**

The financial analyst has looked at only a sample of companies (the companies in the data table). The financial analyst now wants to examine these questions:

- Does a difference exist in the broader population, or is the difference of $635 million due to chance?
- If there is a difference, what is it?

To answer these questions, perform a two-sample t-test. A t-test lets you use data from a sample to make inferences about the larger population.

To perform the t-test, select **Means/Anova/Pooled t** from the red triangle for Oneway Analysis.
Figure 5.18  t Test Results

The p-value of 0.0001 is less than the significance level of 0.05, which indicates statistical significance. Therefore, the financial analyst can conclude that the difference in average profits for the sample data is not due to chance alone. This means that in the larger population, the average profits for pharmaceutical companies are different from the average profits for computer companies.

**Draw Conclusions**

Use the confidence interval limits to determine how much difference exists in the profits of both types of companies. Look at the Upper CL Dif and Lower CL Dif values in Figure 5.18. The financial analyst concludes that the average profit of pharmaceutical companies is between $343 million and $926 million higher than the average profit of computer companies.

**Compare Proportions**

If you have categorical X and Y variables, you can compare the proportions of the levels within the Y variable to the levels within the X variable.

**Scenario**

This example continues to use the Companies.jmp data table. In “Compare Averages for One Variable” on page 141, a financial analyst determined that pharmaceutical companies have higher profits on average than do computer companies.

The financial analyst wants to know whether the size of a company affects profits more for one type of company than the other? However, before examining this question, the financial analyst needs to know whether the populations of computer and pharmaceutical companies consist of the same proportions of small, medium, and big companies.

**Discover the Relationship**

1. Select Help > Sample Data Library and open Companies.jmp.
2. If you still have the Companies.jmp data file open from the previous example, you might have rows that are excluded or hidden. To return the rows to the default state (all rows included and none hidden), select Rows > Clear Row States.
3. Select Analyze > Fit Y by X.
4. Select Size Co and click **Y, Response**.
5. Select Type and click **X, Factor**.
6. Click **OK**.

**Figure 5.19** Company Size by Company Type

The Contingency Table contains information that is not applicable for this example. From the red triangle menu for **Contingency Table** deselect **Total %** and **Col %** to remove that information. Figure 5.20 shows the updated table.
Interpret the Results

The statistics in the Contingency Table are graphically represented in the Mosaic Plot. Together, the Mosaic Plot and the Contingency Table compare the percentages of small, medium, and big companies between the two industries. For example, the Mosaic Plot shows that the computer industry has a higher percentage of small companies compared to the pharmaceutical industry. The Contingency Table shows the exact statistics: 70% of computer companies are small, and about 17% of pharmaceutical companies are small.

Interpret the Test

The financial analyst has looked at only a sample of companies (the companies in the data table). The financial analyst needs to know whether the percentages differ in the broader populations of all computer and pharmaceutical companies.

To answer this question, use the p-value from the Pearson test in the Tests report. See Figure 5.19. Since the p-value of 0.011 is less than the significance level of 0.05, the financial analyst concludes the following:

- The differences in the sample data are not due to chance alone.
- The percentages differ in the broader population.

Now the financial analyst knows that the proportions of small, medium, and big companies are different, and can answer the question: Does the size of company affect profits more for one type of company than the other?

Compare Averages for Multiple Variables

The section “Compare Averages for One Variable” on page 141, compared averages across the levels of a categorical variable. To compare averages across the levels of two or more variables at once, use the Analysis of Variance technique (or ANOVA).
Scenario

The financial analyst can answer the question that we started to work through in the Comparing Proportions section, which is: Does the size of the company have a larger effect on the company’s profits, based on type (pharmaceutical or computer)?

To answer this question, compare the company profits by these two variables:

- Type (pharmaceutical or computer)
- Size (small, medium, big)

Discover the Relationship

To visualize the differences in profit for all of the combinations of type and size, use a graph:

1. Select Help > Sample Data Library and open Companies.jmp.
2. Select Graph > Graph Builder. The Graph Builder window appears.
3. Click Profits ($M) and drag and drop it into the Y zone.
4. Click Size Co and drag and drop it into the X zone.
5. Click Type and drag and drop it into the Group X zone.

Figure 5.21 Graph of Company Profits

The graph shows that one big computer company has very large profits. That outlier is stretching the scale of the graph, making it difficult to compare the other data points.
6. Select the outlier, then right-click and select **Rows > Row Exclude**. The point is removed, and the scale of the graph automatically updates.

7. Click on the Bar icon. Comparing mean profits is easier with bar charts than with points.

**Figure 5.22** Graph with Outlier Removed

![Figure 5.22 Graph with Outlier Removed](image)

The updated graph shows that pharmaceutical companies have higher average profits. The graph also shows that profits differ between company sizes for only the pharmaceutical companies. When the effect of one variable (company size) changes for different levels of another variable (company type), this is called an *interaction*.

**Quantify the Relationship**

Because this data is only a sample, the financial analyst needs to determine the following:

- if the differences are limited to this sample and due to chance
- if the same patterns exist in the broader population

1. Return to the **Companies.jmp** sample data table that has the data point excluded. See “Discover the Relationship” on page 147.

2. Select **Analyze > Fit Model**.
3. Select Profits ($M) and click Y.
4. Select both Type and Size Co.
5. Click the Macros button and select Full Factorial.
6. From the Emphasis menu, select Effect Screening.
7. Select the Keep dialog open option.

**Figure 5.23** Completed Fit Model Window

8. Click Run. The report window shows the model results.

To decide whether the differences in profits are real, or due to chance, examine the **Effect Tests** report.

**Note:** For complete details about all of the **Fit Model** results, see the Model Specification chapter in the *Fitting Linear Models* book.

**View Effect Tests**

The Effect Tests report (see Figure 5.24) shows the results of the statistical tests. There is a test for each of the effects included in the model on the Fit Model window: Type, Size Co, and Type*Size Co.
First, look at the test for the interaction in the model: the Type*Size Co effect. Figure 5.22 showed that the pharmaceutical companies appeared to have different profits between company sizes. However, the effect test indicates that there is no interaction between type and size as it relates to profit. The p-value of 0.218 is large (greater than the significance level of 0.05). Therefore, remove that effect from the model, and re-run the model.

1. Return to the Fit Model window.
2. In the Construct Model Effects box, select the Type*Size Co effect and click Remove.
3. Click Run.

The p-value for the Size Co effect is large, indicating that there are no differences based on size in the broader population. The p-value for the Type effect is small, indicating that the differences that you saw in the data between computer and pharmaceutical companies is not due to chance.

**Draw Conclusions**

The financial analyst wanted to know whether the size of the company has a larger effect on the company’s profits, based on type (pharmaceutical or computer). The financial analyst can now answer the question as follows:

- There is a real difference in profits between computer and pharmaceutical companies in the broader population.
- There is no correlation between the company’s size and type and its profits.

**Use Regression with Multiple Predictors**

The section “Use Regression with One Predictor” on page 136 showed you how to build simple regression models consisting of one predictor variable and one response variable.
Multiple regression predicts the average response variable using two or more predictor variables.

Scenario

This example uses the Candy Bars.jmp data table, which contains nutrition information for candy bars.

A dietitian wants to predict calories using the following information:
• Total fat
• Carbohydrates
• Protein

Use multiple regression to predict the average response variable using these three predictor variables.

Discover the Relationship

To visualize the relationship between calories and total fat, carbohydrates, and protein, create a scatterplot matrix:

1. Select Help > Sample Data Library and open Candy Bars.jmp.
2. Select Graph > Scatterplot Matrix.
4. Select Total fat g, Carbohydrate g, and Protein g, and click X.
5. Click OK.

Figure 5.26 Scatterplot Matrix Results
The scatterplot matrix shows that there is a positive correlation between calories and all three variables. The correlation between calories and total fat is the strongest. Now that the dietitian knows that there is a relationship, the dietitian can build a multiple regression model to predict average calories.

**Build the Multiple Regression Model**

Continue to use the Candy Bars.jmp sample data table.

1. Select **Analyze > Fit Model**.
2. Select Calories and click **Y**.
3. Select Total Fat g, Carbohydrate g, and Protein g and click **Add**.
4. Next to Emphasis, select **Effect Screening**.

**Figure 5.27  Fit Model Window**

5. Click **Run**.

The report window shows the model results. To interpret the model results, focus on these areas:

- “**View the Actual by Predicted Plot**” on page 153
- “**Interpret the Parameter Estimates**” on page 153
- “**Use the Prediction Profiler**” on page 154

**Note:** For complete details about all of the model results, see the Model Specification chapter in the *Fitting Linear Models* book.
View the Actual by Predicted Plot

The Actual by Predicted Plot shows the actual calories versus the predicted calories. As the predicted values come closer to the actual values, the points on the scatterplot fall closer around the red line. See Figure 5.28. Because the points are all very close to the line, you can see that the model predicts calories based on the chosen factors well.

Figure 5.28 Actual by Predicted Plot

Another measure of model accuracy is the RSq value (which appears below the plot in Figure 5.28). The RSq value measures the percentage of variability in calories, as explained by the model. A value closer to 1 means a model is predicting well. In this example, the RSq value is 0.99.

Interpret the Parameter Estimates

The Parameter Estimates report shows the following information:

- The model coefficients
- P-values for each parameter

Figure 5.29 Parameter Estimates Report

In this example, the p-values are all very small (<.0001). This indicates that all three effects (fat, carbohydrate, and protein) contribute significantly when predicting calories.
You can use the model coefficients to predict the value of calories for particular values of fat, carbohydrate, and protein. For example, suppose that you want to predict the average calories for any candy bar that has these characteristics:

- Fat = 11 g
- Carbohydrate = 43 g
- Protein = 2 g

Using these values, you can calculate the predicted average calories as follows:

\[
277.92 = -5.9643 + 8.99\times11 + 4.0975\times43 + 4.4013\times2
\]

The characteristics in this example are the same as the Milky Way candy bar (on row 59 of the data table). The actual calories for the Milky Way are 280, showing that the model predicts well.

**Use the Prediction Profiler**

Use the Prediction Profiler to see how changes in the factors affect the predicted values. The profile lines show the magnitude of change in calories as the factor changes. The line for Total fat g is the steepest, meaning that changes in total fat have the largest effect on calories.

**Figure 5.30** Prediction Profiler

Click and drag the vertical line for each factor to see how the predicted value changes. You can also click the current factor values and change them. For example, click on the factor values and type the values for the Milky Way candy bar (row 59).
Draw Conclusions

The dietitian now has a good model to predict calories of a candy bar based on its total fat, carbohydrates, and protein.
JMP provides a host of statistical discovery platforms to help you explore different aspects of your data. You might start with a simple look at individual variables in histograms and then progress to multivariate and cluster analyses to get a deeper look. Each step of the way, you learn more about your data.

This chapter steps through an analysis of the Cereal.jmp sample data table that is installed with JMP. You learn how to explore the data in the Distribution, Multivariate, and Hierarchical Clustering platforms.

**Figure 6.1 Multiple Analyses in JMP**
Before You Begin

One of the powerful features in JMP is its linked analyses. The graphs and reports that you create are linked to each other through the data table. As shown in Figure 6.1, data that are selected in the data table are also selected in the three report windows. As you work through the examples in this chapter, keep the JMP windows open to see these interactions yourself.

Explore Data in Multiple Platforms

Which cereals are part of a healthy diet? The Cereal.jmp sample data (real data gathered from boxes of popular cereals) provides statistics on fiber content, calories, and other nutritional information. To identify the healthiest cereals, you step through interpreting histograms and descriptive statistics, correlations and outlier detection, scatterplots, and cluster analysis.

Analyze Distributions

The Distribution platform illustrates the distribution of a single variable (univariate analysis) using histograms, additional graphs, and reports. The word univariate simply means involving one variable instead of two (bivariate) or many (multivariate). However, you can examine the distribution of several individual variables within a single report. The report content for each variable changes depending on whether the variable is categorical (nominal or ordinal) or continuous.

- For categorical variables, the initial graph is a histogram. The histogram shows a bar for each level of the ordinal or nominal variable. The reports show counts and proportions.
- For 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 reports show selected quantiles and summary statistics.

Once you know how your data are distributed, you can plan the appropriate type of analysis going forward.

Note: For details about the Distribution platform, see the Distributions chapter in the Basic Analysis book.

Scenario

You want to view the nutritional values of cereals so that you can eat more healthily. Analyzing distributions of cereal data reveals answers to the following questions:

- Which cereals contain the highest amount of fiber?
- What is the average, minimum, and maximum number of calories?
- What is the average amount of fat?
- Which cereal contains the most fat?
- Are there any outliers in the data?

Create the Distributions

1. Select Help > Sample Data Library and open Cereal.jmp.
2. Select Analyze > Distribution.
3. Press Ctrl and click Manufacturer, Calories, Fat, and Fiber.
4. Click Y, Columns and then click OK.

Figure 6.2 Distributions for Manufacturer, Calories, Fat, and Fiber

In the Fiber distributions, notice the following:
- Fiber One and All-Bran with Extra Fiber contain the most fiber as shown in the Fiber box plot. These cereals are outliers in terms of fiber content.

To display the name of the cereal next to a data point in graphs, the row that contains Fiber One in Cereal.jmp is labeled. To see the entire label, drag the right-most vertical border to the right. Place your cursor over the unlabeled data point to see “All Bran with Extra Fiber”. 
In the Fat distributions, notice the following:

- Place your cursor over the top data point (the x marker) in the Fat box plot to see that 100% Nat. Bran Oats & Honey is the highest in fat.
- In the Fat Quantiles report, the median amount of fat is 1 gram.

In the Calories Quantiles report, notice the following:

- The maximum number of calories is 250.
- The minimum number of calories is 50.

5. In the Manufacturer histogram, click on the bar for Nabisco.

Figure 6.3 Distributions for Nabisco Cereals

The Calories, Fat, and Fiber distributions for Nabisco cereals are highlighted in the other histograms. You can view the Calories, Fat, and Fiber distributions for the Nabisco cereals relative to the Calories, Fat, and Fiber distributions for the overall data. For example, the Fat distribution of Nabisco cereals seems to be lower than the Fat distribution for the overall data.

6. Click below the last Fiber bar to deselect all bars.

7. Press Shift and click all histogram bars in the Fiber histogram with a value above 8.
The highest-fiber cereals are highlighted in the Calories and Fat histograms. Because the histograms are linked, note that some of the high-fiber cereals are also low in fat.

8. Press Ctrl and Shift and deselect the two Calories histogram bars that are at or near 200. High calorie cereals are eliminated from the histograms.

**Figure 6.4** High-Fiber Cereals

**Figure 6.5** High-Fiber and Low-Calorie Cereals

**Tip:** Leave the Distributions report open. You will use it later in a cluster analysis. See “Analyze Similar Values” on page 166.

**Interpret the Results**

Looking at the results, you can answer the following questions:

**Which cereals are highest in fiber?**  The Fiber box plot shows that All-Bran with Extra Fiber and Fiber One have the highest amount of fiber. These two cereals are outliers.
What is the average, minimum, and maximum number of calories? The Calories histogram shows that the number of calories range from 50 to 275. The Calories Quantiles show that the number of calories range from 50 to 250, and the median number of calories is 120. The distribution is not uniform.

What is the average amount of fat? The Fat Quantiles report shows that the median amount of fat is 1 gram.

Which cereal contains the most fat? The Fat box plot shows that 100% Nat. Bran Oats & Honey is the highest in fat. This cereal is an outlier.

Draw Conclusions

To increase the amount of fiber in your diet, you decide to try All-Bran with Extra Fiber and Fiber One. These cereals also happen to have a low number of calories and fat. Most cereals do not greatly increase the amount of fat in your diet, but you plan to avoid the high fat 100% Nat. Bran Oats & Honey. And although most cereals are relatively low in fat, they are not necessarily low in calories.

Analyze Patterns and Relationships

Now that you have identified which cereals to eat or avoid, you want to see how the cereal variables relate to each other. The Multivariate platform enables you to observe patterns and relationships between variables. From the Multivariate report, you can do the following:

• summarize the strength of the linear relationships between each pair of response variables using the Correlations table
• identify dependencies, outliers, and clusters using the Scatterplot Matrix
• use other techniques to examine multiple variables, such as partial, inverse, and pairwise correlations, covariance matrices, and principal components

Note: For details about the Multivariate platform, see the Correlations and Multivariate Techniques chapter in the Multivariate Methods book.

Scenario

You want to see the relationships between variables such as fat and calories. Analyzing the cereal data in the Multivariate platform reveals answers to the following questions:

• Which pairs of variables are highly correlated?
• Which pairs of variables are not correlated?
Create the Multivariate Report

1. In the Cereal.jmp data table, click the bottom triangle at the top of the Columns panel to deselect the rows (Figure 6.6).

Figure 6.6 Deselecting Rows

2. Select Analyze > Multivariate Methods > Multivariate.

3. Select Calories through Potassium, click Y, Columns, and then click OK.

The Multivariate report appears. The report contains the Correlations report and Scatterplot Matrix by default. The Correlations report is a matrix of correlation coefficients that summarizes the strength of the linear relationships between each pair of response (Y) variables. The dark numbers indicate a higher degree of correlation.

Figure 6.7 Correlations Report

<table>
<thead>
<tr>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Calories</td>
</tr>
<tr>
<td>Calories</td>
</tr>
<tr>
<td>Protein</td>
</tr>
<tr>
<td>Fat</td>
</tr>
<tr>
<td>Sodium</td>
</tr>
<tr>
<td>Fiber</td>
</tr>
<tr>
<td>Complex Carbs</td>
</tr>
<tr>
<td>Tot Carbs</td>
</tr>
<tr>
<td>Sugars</td>
</tr>
<tr>
<td>Calories fr Fat</td>
</tr>
<tr>
<td>Potassium</td>
</tr>
</tbody>
</table>

Note the following:

- In the Calories column, the number of calories is highly correlated with all variables except for sodium and fiber.
- In the Fiber column, fiber and potassium appear to be highly correlated.
- In the Sodium column, sodium is not highly correlated with the other variables.

The density ellipses in the Scatterplot Matrix further illustrates relationships between variables. Figure 6.8 shows a portion of the plot.
By default, a 95% bivariate normal density ellipse is in each scatterplot. Assuming that each pair of variables has a bivariate normal distribution, this ellipse encloses approximately 95% of the points. If the ellipse is fairly round and is not diagonally oriented, the variables are uncorrelated. If the ellipse is narrow and diagonally oriented, the variables are correlated.

Note the following:

– The ellipses are fairly round in the Sodium row. This shape indicates that Sodium is uncorrelated with other variables.

– The blue x markers, which represent Nat. Bran Oats & Honey, Cracklin’ Oat Bran, and Banana Nut Crunch, appear outside the ellipses in the Fat row. This placement indicates that the data is an outlier (due to the amount of fat in the cereal).

You will further explore a scatterplot matrix later.

4. Select **Pairwise Correlations** from the Multivariate red triangle menu to display the Pairwise Correlations report.

**Figure 6.9** Portion of the Pairwise Correlations Report
The Pairwise Correlations report lists the Pearson product-moment correlations for each pair of Y variables. The report also shows significance probabilities and compares the correlations in a bar chart.

5. To quickly see which pairs are highly correlated, right-click in the report and select the Sort by Column, Signif Prob, Ascending checkbox, and then click OK.

The most highly correlated pairs appear at the top of the report. The small p-values for the pairs indicate evidence of correlation. The most significant correlation is between Tot Carbo (total carbohydrates) and Calories.

**Figure 6.10** Small p-values for Pairs

| Variable   | by Variable  | Correlation | Count | Lower 95% | Upper 95% | Signif Prob | -.8 | -.6 | -.4 | -.2 | 0   | .2  | .4  | .6  | .8  |
|------------|--------------|-------------|-------|-----------|-----------|-------------|-----|-----|-----|-----|-----|-----|-----|-----|
| Tot Carbo  | Calories     | 0.9075      | 76    | 0.8576    | 0.9406    | .001*       |     |     |     |     |     |     |     |     |
| Calories   | Fat          | 0.9013      | 76    | 0.8483    | 0.9395    | .001*       |     |     |     |     |     |     |     |     |
| Potassium  | Fiber        | 0.8328      | 76    | 0.7747    | 0.8898    | .001*       |     |     |     |     |     |     |     |     |
| Tot Carbo  | Complex Carbs| 0.7773      | 76    | 0.6691    | 0.8332    | .001*       |     |     |     |     |     |     |     |     |
| Protein    | Calories     | 0.7041      | 76    | 0.5690    | 0.8022    | .001*       |     |     |     |     |     |     |     |     |
| Tot Carbo  | Protein      | 0.6937      | 76    | 0.5552    | 0.7949    | .001*       |     |     |     |     |     |     |     |     |
| Potassium  | Protein      | 0.6782      | 76    | 0.5344    | 0.7838    | .001*       |     |     |     |     |     |     |     |     |
| Calories   | Fat          | 0.6709      | 76    | 0.5248    | 0.7786    | .001*       |     |     |     |     |     |     |     |     |
| Complex Carbs| Calories   | 0.6588      | 76    | 0.5221    | 0.7771    | .001*       |     |     |     |     |     |     |     |     |
| Complex Carbs| Protein   | 0.6486      | 76    | 0.4956    | 0.7625    | .001*       |     |     |     |     |     |     |     |     |
| Fat        | Calories     | 0.6480      | 76    | 0.4922    | 0.7807    | .001*       |     |     |     |     |     |     |     |     |
| Fiber      | Protein      | 0.5770      | 76    | 0.3668    | 0.6877    | .001*       |     |     |     |     |     |     |     |     |
| Potassium  | Tot Carbo    | 0.5375      | 76    | 0.3551    | 0.6805    | .001*       |     |     |     |     |     |     |     |     |
| Sugars     | Calories     | 0.5060      | 76    | 0.3167    | 0.6566    | .001*       |     |     |     |     |     |     |     |     |
| Calories   | Tot Carbo    | 0.4435      | 76    | 0.2659    | 0.6238    | .001*       |     |     |     |     |     |     |     |     |
| Potassium  | Calories     | 0.4431      | 76    | 0.2442    | 0.6094    | .001*       |     |     |     |     |     |     |     |     |
| Calories   | Fat          | 0.4369      | 76    | 0.2345    | 0.6029    | .001*       |     |     |     |     |     |     |     |     |

**Interpret the Results**

Looking at the results, you can answer the following questions:

**Which pairs of variables are highly correlated?** The Correlations report and Scatterplot Matrix show that the number of calories is highly correlated with all variables except for sodium and fiber. The Pairwise Correlations report shows that Tot Carbo (total carbohydrates) and Calories is the most correlated pair of variables.

**Which pairs of variables are not correlated?** The Correlations report and Scatterplot Matrix show that Sodium is not correlated with the other variables.

**Draw Conclusions**

You confirm the previous decision to avoid the high fat 100% Nat. Bran Oats & Honey. Trying All-Bran with Extra Fiber and Fiber One was also a smart decision. These two high-fiber cereals have the added benefit of contributing a lower number of calories, fat, and sugars and a higher amount of potassium. You also decide to avoid high-carbohydrate cereals because they likely contain a large number of calories.
Chapter 6
Explore Data in Multiple Platforms

Analyze Similar Values

Clustering is a multivariate technique that groups observations together that share similar values across a number of variables. Hierarchical clustering combines rows in a hierarchical sequence that is portrayed as a tree. Cereals with certain characteristics, such as high-fiber cereals, are grouped in clusters so that you can view similarities among cereals.

Note: For details about hierarchical clustering, see the Hierarchical Cluster chapter in the Multivariate Methods book.

Scenario

You want to know which cereals are similar to each other and which ones are dissimilar. Analyzing clusters of cereal data reveals answers to the following questions:

- Which cluster of cereals provides little nutritional value?
- Which cluster of cereals is high in vitamins and minerals and contains a low amount of sugar and fat?
- Which cluster of cereals contains high fiber and low calories?

Create the Hierarchical Cluster Graph

1. With Cereal.jmp displayed, select Analyze > Clustering > Hierarchical Cluster.
2. Select Calories through Enriched, click Y, Columns, and then click OK.

The Hierarchical Clustering report appears. Figure 6.11 shows a portion of the report. The clusters are colored according to the data table row states.

Figure 6.11 Portion of the Hierarchical Clustering Report

3. Select Color Clusters from the Hierarchical Clustering red triangle menu.
The clusters are colored according to their relationships in the dendrogram.

**Figure 6.12 Colored Clusters**

The cereals have similar characteristics within each cluster. For example, judging by the names of the cereals in cluster one, you guess that the cereals are high in fiber.

Notice how All-Bran with Extra Fiber and Fiber One are grouped in cluster one. These cereals are more similar to each other than the other two cereals in the cluster.

**Figure 6.13 Similar Cereals in Cluster One**

4. To select cluster one, click on the red horizontal line on the right.

The four cereals are highlighted in red.
5. To see the similar characteristics in the cluster, select **Cluster Summary** from the red triangle menu.

   The Cluster Summary graph at the bottom of the report shows the mean value of each variable across each cluster. For example, the cereals in this cluster contain more fiber and potassium than cereals in other clusters.

6. Select Scatterplot Matrix from the red triangle menu.

   This option is an alternative to creating a scatterplot matrix in the Multivariate platform. Note the Fiber plot in the Potassium row. The selected cereals are located on the right side of the plot between 8 and 13 grams. This location indicates that the cereals in cluster one are high in fiber and potassium.
Figure 6.16 Cluster One Characteristics

Interpret the Results

Clicking through the clusters and looking at the Cluster Summary report, you can see the following characteristics:

- Cluster one cereals, such as Fiber One and All-Bran, contain high fiber and potassium and low calories.
- Cluster two, which contains many favorite children’s cereals, is high in sugar and low in fiber, complex carbohydrates, and protein.
- Cluster three cereals (Puffed Rice and Puffed Wheat) are low in calories but provide little nutritional value.
- Cluster four cereals, such as Total Corn Flakes and Multi-Grain Cheerios, provide 100% of your daily requirement of vitamins and minerals. They are low in fat, fiber, and sugar.
- Cluster five cereals are high in protein and fat and low in sodium. The cluster consists of cereals such as Banana Nut Crunch and Quaker Oatmeal.
- Cluster six cereals are low in fat and high in sodium and carbohydrates. Traditional cereals such as Wheaties and Grape-Nuts are in this cluster.
- Cluster seven cereals are high in calories and low in fiber. Many cereals that include dried fruit are in this cluster (Mueslix Healthy Choice, Low Fat Granola w Raisins, Oatmeal Raisin Crisp, Raisin Nut Bran, and Just Right Fruit & Nut).
- Cluster eight cereals are low in sodium and sugar, and high in complex carbohydrates, protein, and potassium. Shredded Wheat and Mini-Wheat cereals are in this cluster.

By looking at the joins in the dendrogram, you can see which cereals in each cluster are most similar.

- In cluster one, Fiber One is similar in nutritional value to All-Bran with Extra Fiber. 100% Bran and All-Bran are also similar. Each pair of similar cereals are made by different companies, so the cereals are competing against each other.
• In cluster two, Frosted Flakes and Honey Frosted Wheaties are similar even though one is a corn flake and the other is a wheat flake. Lucky Charms and Frosted Cheerios are similar. Cap’n’Crunch and Trix are also similar.

**Draw Conclusions**

Based on your desire to eat more fiber and fewer calories, you decide to try the cereals in cluster one. You will avoid cereals in cluster three, which consists of puffed wheat and puffed rice and have little nutritional value. And you will try cereals in the highly nutritious cluster four.
Once you have generated results from your data, JMP provides you with multiple ways to share your work with others. Here are some of the ways that you can share your work:

- Saving platform results as journals or projects
- Saving results, data tables, and other files in projects
- Saving scripts to reproduce results in data tables
- Creating Adobe Flash (SWF) versions of platform results
- Saving results as Interactive HTML (.htm, html)
- Saving results as a PowerPoint presentation (.pptx)
- Sharing results in a dashboard

Figure 7.1  Example of an Adobe Flash (SWF) Bubble Plot
Save Platform Results in Journals

Save platform reports for future viewing by creating a journal of the report window. The journal is a copy of the report window. You can edit or append additional reports to an existing journal. The journal is not connected to the data table. A journal is an easy way to save the results from several report windows in a single report window that you can share with others.

Example of Creating a Journal

1. Select Help > Sample Data Library and open Companies.jmp.
2. Select Analyze > Distribution.
3. Select both Type and Size Co and click Y, Columns.
4. Click OK.
5. From the red triangle menu for Type, select Histogram Options > Show Counts.
6. From the red triangle menu for Size Co, select Mosaic Plot.
7. Select Edit > Journal to journal these results. The results are duplicated in a journal window.

Figure 7.2 Journal of Distribution Results
The results in the journal are not connected to the data table. In the Type bar chart, if you click the Computer bar, no rows are selected in the data table.

Since the journal is a copy of your results, most of the red triangle menus do not exist. A journal does have a red triangle menu for each new report that you add to the journal. This menu has two options:

**Rerun in new window**  If you have the original data table that was used to create the original report, this option runs the analysis again. The result is a new report window.

**Edit Script**  This option opens a script window that contains a JSL script to re-create the analysis. JSL is a more advanced topic that is covered in the *Scripting Guide* and *JSL Syntax Reference*.

### Add Analyses to a Journal

If you perform another analysis, you can add the results of the analysis to the existing journal.

1. With a journal open, select **Analyze > Distribution**.
2. Select **profit/emp** and click **Y, Columns**.
3. Click **OK**.
4. Select **Edit > Journal**. The results are appended to the bottom of the journal.

### Create Projects

Save multiple JMP file types (such as data tables, reports, journals, and scripts) in a single file by creating a project. The project file contains all the information needed to re-open all of the included files.

#### Example of Creating a Project

**Create a Report**

1. Select **Help > Sample Data Library** and open Companies.jmp.
2. Select **Analyze > Distribution**.
3. Select **Profits ($M)** and **profit/emp**, and click **Y, Columns**.
4. Click **OK**.

**Create a Project and Add the Data Table and the Report to It**

1. To start a new project, select **File > New > Project**. A window appears that shows the untitled project.
2. Right-click the project (Untitled) and select **Rename**, and then enter a new name (Finance).
3. To add the Distribution results to the project, right-click on the project name and select **Add Window**.
4. In the Add Windows to Project window, select the Distribution results.
5. Click **OK**. The Distribution results are added to the project.
6. To add the `Companies.jmp` data table to the project, repeat step 3 and select the `Companies` data table from the window.
7. Click **OK**. The data table is added to the project.

**Figure 7.4** The Final Project

You can double-click on the links in the project to open the data table and re-create the Distribution results.

---

**Save and Run Scripts**

Most platform options in JMP are scriptable, meaning that most actions that you perform can be saved as a JMP Scripting Language (JSL) script. You can use a script to reproduce your actions or results at any time.
Example of Saving and Running a Script

Create a Report

1. Select Help > Sample Data Library and open Companies.jmp.
2. Select Analyze > Distribution.
3. Select Type and profit/emp and click Y, Columns.
4. Click OK.
5. From the red triangle menu for Type, select these options:
   - Histogram Options > Show Counts
   - Confidence Interval > 0.95
6. From the red triangle menu for profit/emp, select these options:
   - Outlier Box Plot, to remove the outlier box plot
   - CDF Plot
7. From the red triangle menu for Distributions, select Stack.

Save the Script to the Data Table and Run It

1. To save this analysis, select Save Script > To Data Table from the red triangle menu for Distributions. The Distribution script appears in the Table panel.

Figure 7.5 Distribution Script

2. Close the Distribution report window.
3. To re-create the analysis, click the green triangle next to the Distribution script.
Figure 7.6 Running the Distribution Script

Tip: Right-click the table script to view more options.

About Scripts and JSL

The script that you saved in this section contains JMP Scripting Language (JSL) commands. JSL is a more advanced topic that is covered in the Scripting Guide and JSL Syntax Reference.

Save Reports as Interactive HTML

Interactive HTML enables JMP users to share reports that contain dynamic graphs so that even non JMP users can explore the data. The JMP report is saved as a web page in HTML 5 format, which you can email to users or publish on a website. Users then explore the data as they would in JMP.

Interactive HTML provides a subset of features from JMP:
- Explore interactive graph features, such as selecting histogram bars and viewing data values.
- View data by brushing.
- Show or hide report sections.
• Place your cursor over the report for tooltips.
• Increase the marker size.

Figure 7.7 Brushing Data in Interactive HTML

Many changes that you make to the graphs, such as ordered variables, horizontal histograms, background colors, and colored data points, are saved in the web page. Graphs and tables that are closed when you save the content remain closed on the web page until the user opens them.

Interactive HTML Contains Data

When you save reports as interactive HTML in JMP, your data are embedded in the HTML. The content is unencrypted, because web browsers cannot read encrypted data. To avoid sharing sensitive data, save your results as a non-interactive web page. (Select File > Save As > HTML File on Windows, or File > Export > HTML on Macintosh.)

Example of Creating Interactive HTML

Create a Report
1. Select Help > Sample Data Library and open Big Class.jmp.
2. Select Analyze > Distribution.
4. Click OK.
Save and Share Your Work

Chapter 7
Discovering JMP

Save as Interactive HTML

1. On Windows, select File > Save As and select Interactive HTML with Data from the Save as type list. On Macintosh, select File > Export > Interactive HTML with Data.

2. Name and save the file (or export on Macintosh).
   The output appears in your default browser.

Figure 7.8 Interactive HTML Output

For information about exploring interactive HTML output, visit http://www.jmp.com/support/help/InteractiveHTML/13/ShareJMPReports.shtml.

Save a Report as a PowerPoint Presentation

Create a presentation by saving JMP results as a Microsoft PowerPoint presentation (.pptx). Rearrange JMP content and edit text in PowerPoint after saving as a .pptx file. Sections of a JMP report are exported into PowerPoint differently.

• Report headings are exported as editable text boxes.
• Graphs are exported as images. Certain graphical elements, such as legends, are exported as separate images. Images resize to fit the slide in PowerPoint.
Use the selection tool to select the sections that you want to save in your presentation. Delete unwanted content once after you open the file in PowerPoint.

**Note:** On Windows, PowerPoint 2007 is the minimum version required to open .pptx files created in JMP. On Macintosh, at least PowerPoint 2011 is required.

1. In JMP, create the report.
2. (Windows) Select **File > Save As** and select **PowerPoint Presentation** from the Save as type list.
3. (Macintosh) Select **File > Export > PowerPoint Presentation** and click **Next**.
4. Select a graphic file format from the list.
   On Windows, EMF is the default format. On Macintosh, PDF is the default format.
5. Name and save the file (or export on Macintosh).
   The file opens in Microsoft PowerPoint because **Open presentation after save** is selected by default.

**Note:** The native EMF graphics produced on Windows are not supported on Macintosh. The native PDF graphics produced on Macintosh are not supported on Windows. For cross-platform compatibility, change the default graphics file format by selecting **File > Preferences > General**. Then, change the **Image Format for PowerPoint** to either PNG or JPEG.

---

**Create Adobe Flash Versions of the Profiler, Bubble Plot, or Distribution Platforms**

To share the interactive results of a profiler, bubble plot, or distribution outside of JMP, you can export an SWF file and use the Adobe Flash Player to view it. You can import the SWF file into presentations and applications. You can also save the results as an HTML page with the SWF output embedded.

**Example of Saving an Adobe Flash Version of a Bubble Plot**

**Create a Bubble Plot in JMP**

1. Select **Help > Sample Data Library** and open PopAgeGroup.jmp.
2. Select **Graph > Bubble Plot**.
3. Select Portion60+ and click **Y**.
4. Select Portion 0-19 and click **X**.
5. Select Country and click **ID**.
6. Select Year and click **Time**.
7. Select Pop and click **Sizes**.
8. Select Region and click **Coloring**.
9. Click **OK**.

**Figure 7.9** Initial Bubble Plot

Save the Bubble Plot as an SWF File

1. Select **Save for Adobe Flash Platform (SWF)** from the red triangle menu.
2. In the Save As SWF window, select the location to which you want to save the file.
3. Click **Save**. The Adobe Flash version of the Bubble Plot is saved as HTML and appears in a web browser.
Create Dashboards

A dashboard is a visual tool that lets you run and present reports on a regular basis. You can show reports, data filters, selection filters, data tables, and graphics on a dashboard. The content shown on the dashboard is updated when you open the dashboard.

Example of Combining Windows

You can quickly create dashboards by merging several open windows in JMP. Combining windows provides options to view a summary of statistics and include a selection filter.

1. Select Help > Sample Data Library and open Birth Death.jmp.
2. Run the Distribution and Bivariate table scripts.
3. Select **Window > Combine Windows**.
   The Combine Windows window appears.

   **Tip:** On Windows, you can also select Combine Windows from the Arrange Menu option in the lower right corner of JMP windows.

4. Select **Summary Report View** to display the graphs and omit the statistical reports
5. In the Combine column, select **Birth Death - Bivariate of death by birth** and **Birth Death - Distribution**.
6. In the Filter By column, select **Birth Death - Distribution**.

**Figure 7.11 Combine Windows Options**

![Combine Windows Options](image)

7. Click **OK**.
   The two reports are combined into one window. Notice the filter icon 🕒 at the top of the Distribution report. When you select a bar in one of the histograms, the corresponding data in the Bivariate graph are selected.
Example of Creating a Dashboard with Two Reports

Suppose that you created two reports and want to run the reports again the next day against an updated set of data. This example shows how to create a dashboard from the reports in Dashboard Builder.

2. Run the table scripts named “Distribution: Profitability by Lead Studio and Genre” and “Graph Builder: World and Domestic Gross by Genre”.
3. From any window, select File > New > Dashboard.
   Templates for common layouts appear.
4. Select the 2x1 Dashboard template.
   A box with room for two reports appears on the workspace.
5. In the Reports list, double-click the report thumbnails to put them on the dashboard.
6. Select Preview Mode from the Dashboard Builder red triangle menu.
   A preview of the dashboard appears. Notice that the graphs are linked to each other and the data table. They also have the same red triangle options as the Distribution and Graph Builder platforms.
Save and Share Your Work
Create Dashboards

Chapter 7
Discovering JMP

7. Click Close Preview.

Figure 7.13 Dashboard with Two Reports

For more information about creating dashboards, see the Extend JMP chapter in the Using JMP book.

Save the Dashboard as an Add-In

A JMP add-in is a JSL script that you can run anytime from the JMP Add-Ins menu. Saving a dashboard as an add-in enables you to share a dashboard with another JMP user.

After creating the dashboard as shown in “Example of Creating a Dashboard with Two Reports” on page 183, follow these steps to save it as an add-in.

1. From the Dashboard Builder red triangle menu, select **Save Script > To Add-In**.
   Add-In Builder opens.
2. Next to Add-In name, type Hollywood Movies Dashboard (the add-in filename).
3. Click the **Menu Items** tab and type Hollywood Movies Dashboard next to **Menu item name** (the name of the add-in menu item).
4. Click **Save** and save the add-in to your desktop.
   The add-in is saved and installed in your Add-Ins menu.
5. Click **Close** on the Add-In Builder window.

6. From the JMP main menu, select **Add-Ins** and select **Hollywood Movies Dashboard**.

   The Graph Builder and Distribution reports are generated from Hollywood Movies.jmp.

For more information about Add-In Builder, see the Creating Applications chapter in the *Scripting Guide*. 
Using some of the special features in JMP, you can do the following:

- Update analyses or graphs automatically
- Customize platform results
- Integrate with SAS to use advanced analytical features

**Figure 8.1** Examples of Special Features

```plaintext
DATA Candy_Bars; INPUT Calories Total_fat_g Carbohydrate_g Protein_g; Lines;
310 20 28 6
230 12 27 4
200 12 24 3
170 6 21 2
200 2.5 43 1
200 16 26 3
180 1.5 42 2
190 11 21 2
230 12 28 3
; RUN;

PROC GLM DATA=Candy_Bars ALPHA=0.05;
MODEL Calories = Total_fat_g Carbohydrate_g Protein_g;
RUN;
```
Automatically Update Analyses and Graphs

When you make a change to a data table, you can use the Automatic Recalc feature to automatically update analyses and graphs that are associated with the data table. For example, if you exclude, include, or delete values in the data table, that change is instantly reflected in the associated analyses or graphs. Note the following information:

- Some platforms do not support Automatic Recalc. For more information, see the JMP Reports chapter in the Using JMP book.
- For the supported platforms in the Analyze menu, Automatic Recalc is turned off by default. However, for the supported platforms in the Quality and Process menu, Automatic Recalc is turned on by default, except for the Variability/Attribute Gauge Chart, Capability, and Control Chart.
- For the supported platforms in the Graph menu, Automatic Recalc is turned on by default.

Example of Using Automatic Recalc

This example uses the Companies.jmp sample data table, which contains financial data for 32 companies from the pharmaceutical and computer industries.

1. Select Help > Sample Data Library and open Companies.jmp.
2. Select Analyze > Fit Y by X.
3. Select Sales ($M) and click Y, Response.
4. Select # Employ and click X, Factor.
5. Click OK.

Figure 8.2 Initial Scatterplot

The initial scatterplot shows that one company has significantly more employees and sales than the other companies. You decide that this company is an outlier, and you want to exclude
that point. Before you exclude the point, turn on Automatic Recalc so that your scatterplot is updated automatically when you make the change.

6. Turn on Automatic Recalc by selecting **Redo > Automatic Recalc** from the red triangle menu.

7. Click on the outlier to select it.

8. Select **Rows > Exclude/Unexclude**. The point is excluded from the analysis and the scatterplot is automatically updated.

**Figure 8.3** Updated Scatterplot

If you fit a regression line to the data, the point in the lower right corner is an outlier, and influences the slope of the line. If you then exclude the outlier with Automatic Recalc turned on, you can see the slope of the line change.

9. Fit a regression line by selecting **Fit Line** from the red triangle menu. Figure 8.4 shows the regression line and analysis results added to the report window.
10. Click on the outlier to select it.

11. Select Rows > Exclude/Unexclude. The regression line and analysis results are automatically updated, reflecting the exclusion of the point.

**Tip:** When you exclude a point, the analyses are recalculated without the data point, but the data point is not hidden in the scatterplot. To also hide the point in the scatterplot, select the point, and then select Rows > Hide and Exclude.
Change Preferences

You can change preferences in JMP using the Preferences window. To open the Preferences window, select **File > Preferences**.
On the left side of the Preferences window is a list of Preference groups. On the right side of the window are all of the preferences that you can change for the selected category.

**Example of Changing Preferences**

Every platform report window has options that you can turn on or off. However, your changes to these options are not remembered the next time you use the platform. If you want JMP to remember your changes every time you use the platform, change those options in the Preferences window.

This example shows how to set the Distribution platform so that an Outlier Box Plot is not added to the initial report.

**Create a Distribution Using the Default Preference Setting**

1. Select Help > Sample Data Library and open Companies.jmp.
2. Select Analyze > Distribution.
3. Select Profits ($M) and click Y, Columns.
4. Click OK.

**Figure 8.7** Distribution Report Window

The histogram is vertical, and the graphs includes an outlier box plot. To change the histogram to horizontal and remove the outlier box, select the appropriate options from the red triangle menu for Profits ($M). However, if you want those preferences to be in effect every time you use the platform, then change them in the Preferences window.

**Change the Preference for the Outlier Box Plot and Run Distribution Again**

1. Select File > Preferences.
2. Select Platforms from the preference group.
3. Select Distribution from the Platforms list.
4. Select the Horizontal Layout option to turn it on.
5. Deselect the Outlier Box Plot option to turn it off.
6. Click **OK**.
7. Repeat the Distribution analysis. See “Create a Distribution Using the Default Preference Setting” on page 192.

The histogram is now horizontal and the outlier box plot does not appear. These preferences remain the same until you change them.

For details about all of the preferences, see the JMP Preferences chapter in the *Using JMP* book.

### Integrate JMP and SAS

**Note:** You must have access to SAS, either on your local machine or on a server, to use SAS through JMP.

Using JMP, you can interact with SAS as follows:

- Write or create SAS code in JMP.
- Submit SAS code and view the results in JMP.
- Connect to a SAS Metadata Server or a SAS Server on a remote machine.
• Connect to SAS on your local machine.
• Open and browse SAS data sets.
• Retrieve and view data sets generated by SAS.

For complete details about integrating JMP and SAS, see the Import Your Data chapter in the Using JMP.

Example of Creating SAS Code

This example uses the Candy Bars.jmp sample data table, which contains nutrition data for candy bars.

1. Select Help > Sample Data Library and open Candy Bars.jmp.
2. Select Analyze > Fit Model.
3. Select Calories and click Y.
4. Select Total fat g, Carbohydrate g, and Protein g, and click Add.
5. From the red triangle menu for Model Specification, select Create SAS Job.

Figure 8.9 shows the SAS code. (Not all of the data is shown.)

Figure 8.9 SAS Code

```
DATA Candy_Bars; INPUT Calories Total_fat g Carbohydrate g Protein g; Lines;
310 20 28 6
230 12 27 4
220 12 24 3
170 8 21 3
260 2.5 43 1
260 16 28 5
150 1.5 42 2
190 11 21 2
230 12 28 3
;
RUN;
```

```
PROC GLM DATA=Candy_Bars ALPHA=0.05;
MODEL Calories = Total_fat g Carbohydrate g Protein g;
RUN;
```

Example of Submitting SAS Code

1. Select Help > Sample Data Library and open Candy Bars.jmp.
2. Select Analyze > Fit Model.
3. Select Calories and click Y.
4. Select Total fat g, Carbohydrate g, and Protein g, and click Add.
5. From the red triangle menu for Model Specification, select Submit to SAS.
6. In the Connect to SAS Server window (see Figure 8.10), choose a method to connect to SAS (if you are not already connected). For this example, select Connect to SAS on this machine.
Figure 8.10 Connect to SAS Server

7. Click OK.

JMP connects to SAS. SAS runs the model and sends the results back to JMP. The results can appear as SAS output, HTML, RTF, PDF, or JMP report format (you can choose the format using JMP Preferences). Figure 8.11 shows the results formatted as a JMP report. For details, see the Import Your Data chapter in the Using JMP book.

Figure 8.11 SAS Results Formatted as a JMP Report
Index
Discovering JMP

A
Actual by Predicted plot 153
Add-In Builder 184
Adobe Flash
  creating bubble plots 179
  creating Profiler results 179
analyses
  adding to journals 173
  analyzing variable relationships 136
  automatically updating 188–191
analyzing data 123
Anscombe, F.J. 124
automatic recalc 188–191
averages, comparing
  multiple variables 146
  single variable 141

B
bar charts 92
  creating 93
  example 92–94
  interacting with 94
  interpreting 93
box plots 102
  changing to line plots 114
  creating 103
  Distribution platform and 132
  example 103
  interpreting 104
  outlier box plots 56
bubble plots 117
  creating 118
  creating Adobe Flash version 179
  example 118
  interacting with 121
  interpreting 120
  options 121
trail bubbles 121

categorical variables
  comparing averages 141
  comparing proportions 144
  example 133–135
  relationships 136

cells, selecting or deselecting 72
character data type 74
charts
  bar charts 92–94
  creating separate charts 115
  variability charts 108
clustering 166
code, SAS
  creating 195
  submitting 195
Column Info window 73–74
column names 74
  changing 74
  entering 74
  unique 74
Column Properties menu 75
columns
  adding 66
  locking 75
  naming 74
  selecting or deselecting 71
columns panel 52
combining reports in dashboard 181
contingency tables 145
continuous modeling type 74, 127
continuous variables 127
  comparing averages 141
  example of analyzing 131
  fitting lines 137
modeling types 127
regression with one predictor 136
relationships 136
copying and pasting data 62
correlation 95

D

dashboard, creating 181
data
analyzing 123
changing
modeling types 128
multiple values 69
single value 69
copying and pasting 62
creating patterned 69
deselecting
cells 72
columns 71
rows 70
excluding data points 97
filtering 77
finding means 83
graphing 124
importing 62–63
including data points 97
moving into JMP 62
searching for values 72
selecting
cells 72
columns 71
data points 99
rows 70
unexcluding 97
data filter 77
data grid 52
data points
excluding 97
including 97
selecting 99
data tables 48
See also modeling types
adding
columns 66
rows 66
to projects 174

changing name of 66
Column Properties menu 75
columns panel 52
creating 65
data grid 52
data types 74
example 61
formatting numeric values 74
joining 85
opening 62
opening sample 51
rows panel 53
sorting by columns 87
subsetting 83
table panel 52
typing data into 65
data types 74
disclosure buttons 48, 55
Distribution platform 130
box plot and 132
effect, example, categorical variables 133–135
example, continuous variables 131
histograms and 132
Linear Fit report 139
Quantiles report 132
rotating the report window 131
Summary Statistics report 132
t-tests 132

E

Effect Tests report 149
example, analyses 123
excluding data 97

F

Fill options 69
filtering data 77
Fit Model platform
Actual by Predicted plot 153
analyzing variable relationships 136
Effect Tests report 149
launching 152
Parameter Estimates report 153
Prediction Profiler 154
Fit Y by X platform
analyzing variable relationships 136
contingency table 145
launching 141
Means and Std Dev report 143
Mosaic Plot 146
Tests report 146
t-tests 143
fitting lines 137
formula editor 75

G
Graph Builder 111
changing box plots to line plots 114
creating graphs 112
creating separate charts 115
example 112
interpreting graphs 116
launching 147
graphing data 124
graphs
automatically updating 188–191
creating 112
multiple variable 94
See also Graph Builder
single variable 90
grid lines, adding to overlay plots 107

H
hierarchical clustering 166
histograms 90
creating 91
Distribution platform and 132
example 90
interacting with 92, 135
interpreting 91
Home window 49
HTML 5, save output as 176

I
icons, modeling type 66
importing data 62
interaction 148
interactive HTML, save report as 176

J
JMP 47
copying and pasting data into 62
creating SAS code in 195
getting started 48–58
importing data 62
integrating with SAS 194
starting JMP 49
submitting SAS code in 195
versus Microsoft Excel 58
JMP scripting language 176
joining data tables 85
journals
adding analyses 173
creating 172
saving platform results 172
JSL (JMP scripting language) 176

L
lasso tool 99
launch windows 48
line plots, changing box plots to 114
linear fit report 139
lines
adding mean lines 142
fitting 137
locking columns 75

M
mean lines 142
mean, testing 133
Means and Std Dev report 143
Microsoft Excel
importing data 63
versus JMP 58
modeling types
See also categorical variables
changing 128
continuous 74, 127
example 127
icons 66
nominal 74, 127
ordinal 74, 127
Mosaic Plot 146
multiple regression 150, 152
multivariate analysis 162

N
names
  columns 74
  data tables 66
nominal modeling type 74
  See also categorical variables
numeric data type 74
numeric values, formatting 74

O
ordinal modeling type 74, 127
  See also categorical variables
outlier box plots 56
outliers 92, 104
  example 137, 139
  excluding 139
  implications of 142, 147
  removing 97
overlay plots 105–106
  adding grid lines 107
  connecting points 107
  example 105
  interacting with 107
  interpreting 107
  legends 108
  multiple variables and 107

P
Parameter Estimates report 153
patterned data, creating 69
platforms 48, 53
  examples 54–58
  launch windows 48
  launching 54
  report windows and 48, 55, 57
  saving results 172
  viewing results 54
plots
  See specific plot type
PowerPoint, saving report as 178
Prediction Profiler 154

predictor variables 136
preferences, changing 191
Profiler platform, creating Adobe Flash
  results 179
projects
  adding
    data tables 174
    results 174
  creating 173
proportions, comparing 144

Q
Quantiles report 132

R
recoding data 69
red triangle menus 48, 55
regression
  multiple 150, 152
  one predictor 136
  two predictors 150
report windows 55
  closing or opening 57
  description 48
  disclosure buttons 48, 55
  interacting with 57
  platform options 48
  red triangle menus 48, 55
  reports in 48
  rotating 131
reports
  Effect Tests 149
  Linear Fit 139
  Means and Std Dev 143
  Parameter Estimates 153
  Quantiles 132
  Summary Statistics 132
  Tests 146
results, saving platform 172
rotating the report window 131
row state data type 74
rows
  adding 66
  clear row states 144
  excluding or unexcluding 97
returning to default state 144
selecting or deselecting 70
rows panel 53

tests
Distribution platform 132
Fit Y by X platform 143
two-sample 143

V
values
changing multiple 69
changing single 69
formatting numeric 74
searching for 72

variability 103
variability charts 108
creating 109
example of 109
grouping order 109
hiding 110
interpreting 110

variables
analyzing relationships 136
categorical 133, 136, 141, 144
comparing averages for multiple 146
comparing averages for single 141
continuous 127, 130, 136–137, 141
graphs with multiple 94
graphs with single 90
modeling types 127
overlay plots with multiple 107
predictor 136
types of relationships 136

W-Z
windows
JMP Starter window 49
Tip of the Day window 49

T
table panel 52
testing means 133
Tests report 146
Tip of the Day window 49
trail bubbles 121

S
sample data tables 51
SAS
creating code 195
integrating with JMP 194
submitting code 195

saving results in journal 172
scatterplot matrices 99
creating 100
every example 100
interacting with 102
interpreting 101
scatterplot matrix 164
Scatterplot Matrix platform
example 151
launching 151

scatterplots 95–96
example 96
interacting with 99
interpreting 97
scripts
creating 174
JSL and 176
running 174

searching for values 72
side-by-side box plots
See box plots
sorting data tables by columns 87
standard deviations 143
starting JMP 49
statistics, summary 79
subsetting data tables 83
summary statistics 79
Summary Statistics report 132

Index
Discovering JMP