



**Version 14**

# **Consumer Research**

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

Marcel Proust

JMP, A Business Unit of SAS  
SAS Campus Drive  
Cary, NC 27513

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2018. *JMP® 14 Consumer Research*. Cary, NC: SAS Institute Inc.

## **JMP® 14 Consumer Research**

Copyright © 2018, SAS Institute Inc., Cary, NC, USA

ISBN 978-1-63526-497-5 (Hardcopy)

ISBN 978-1-63526-498-2 (EPUB)

ISBN 978-1-63526-499-9 (MOBI)

ISBN 978-1-63526-500-2 (Web PDF)

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.

SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513-2414.

March 2018

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 © 1998-2017 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.
- ZLIB Compression Library - Copyright © 1995-2005, Jean-Loup Gailly and Mark Adler.
- Made with Natural Earth. Free vector and raster map data @ [naturalearthdata.com](http://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.

- R software is Copyright © 1999-2012, R Foundation for Statistical Computing.
- MATLAB software is Copyright © 1984-2012, The MathWorks, Inc. Protected by U.S. and international patents. See [www.mathworks.com/patents](http://www.mathworks.com/patents). MATLAB and Simulink are registered trademarks of The MathWorks, Inc. See [www.mathworks.com/trademarks](http://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 license 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:

Copyright © 1985, 1986, 1992, 1993

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

## Consumer Research

---

<b>1</b>	<b>Learn about JMP</b>	<b>13</b>
	<b>Documentation and Additional Resources</b>	
	Formatting Conventions	15
	JMP Documentation	16
	JMP Documentation Library	16
	JMP Help	22
	Additional Resources for Learning JMP	22
	Tutorials	22
	Sample Data Tables	23
	Learn about Statistical and JSL Terms	23
	Learn JMP Tips and Tricks	23
	Tooltips	23
	JMP User Community	24
	JMP Books by Users	24
	The JMP Starter Window	24
	Technical Support	25
<b>2</b>	<b>Introduction to Consumer Research</b>	<b>27</b>
	<b>Overview of Customer and Behavioral Research Methods</b>	
<b>3</b>	<b>Categorical Response Analysis</b>	<b>29</b>
	<b>Analyze Survey and Other Counting Data</b>	
	Example of the Categorical Platform	31
	Launch the Categorical Platform	32
	Response Roles	32
	Columns Roles	34
	Other Launch Window Options	35
	The Categorical Report	36
	Categorical Platform Options	37
	Report Options	38
	Statistical Testing Options	39
	Additional Categorical Platform Options	41
	Responses Table Options	44

Comparison Letters .....	44
Supercategories .....	46
Supercategories Options .....	46
Set Preferences .....	47
Additional Examples of the Categorical Platform .....	48
Example of the Test for Response Homogeneity .....	48
Example of the Multiple Response Test .....	49
Example of the Cell Chisq Test .....	51
Example of Compare Each Sample with Comparison Letters .....	52
Example of Compare Each Cell with Comparison Letters .....	54
Example of User-Specified Comparison with Comparison Letters .....	56
Example of Conditional Association and Relative Risk .....	57
Example of Rater Agreement .....	59
Example of Repeated Measures .....	60
Examples of the Multiple Response Tab .....	61
Example of Mean Score with Comparison Letters .....	66
Example of a Structured Report .....	68
Statistical Details for the Categorical Platform .....	69
Rao-Scott Correction .....	69
<b>4 Choice Models .....</b>	<b>71</b>
<b>Fit Models for Choice Experiments</b>	
Overview of the Choice Modeling Platform .....	73
Examples of the Choice Platform .....	75
One Table Format with No Choice .....	75
Multiple Table Format .....	78
Launch the Choice Platform .....	84
Launch Window for One Table, Stacked .....	85
Launch Window for Multiple Tables, Cross-Referenced .....	86
Choice Model Report .....	91
Effect Summary .....	91
Parameter Estimates .....	92
Likelihood Ratio Tests .....	93
Bayesian Parameter Estimates .....	93
Choice Platform Options .....	95
Willingness to Pay .....	98
Additional Examples .....	101
Example of Making Design Decisions .....	101
Example of Segmentation .....	109
Example of Logistic Regression Using the Choice Platform .....	113

Example of Logistic Regression for Matched Case-Control Studies .....	116
Example of Transforming Data to Two Analysis Tables .....	118
Example of Transforming Data to One Analysis Table .....	123
Statistical Details for the Choice Platform .....	125
Special Data Table Rules .....	125
Utility and Probabilities .....	126
Gradients .....	127
<b>5 MaxDiff .....</b>	<b>129</b>
<b>Fit Models for MaxDiff Experiments</b>	
Overview of the MaxDiff Modeling Platform .....	131
Examples of the MaxDiff Platform .....	131
One Table Format .....	132
Multiple Table Format .....	134
Launch the MaxDiff Platform .....	138
Launch Window for One Table, Stacked .....	139
Launch Window for Multiple Tables, Cross-referenced .....	140
MaxDiff Model Report .....	144
Effect Summary .....	145
MaxDiff Results .....	146
Parameter Estimates .....	147
Bayesian Parameter Estimates .....	149
Likelihood Ratio Tests .....	150
MaxDiff Platform Options .....	151
Comparisons Report .....	153
Save Bayes Chain .....	153
<b>6 Uplift Models .....</b>	<b>155</b>
<b>Model the Incremental Impact of Actions on Consumer Behavior</b>	
Overview of the Uplift Platform .....	157
Example of the Uplift Platform .....	158
Launch the Uplift Platform .....	159
The Uplift Model Report .....	160
Uplift Model Graph .....	160
Uplift Report Options .....	163
<b>7 Multiple Factor Analysis .....</b>	<b>165</b>
<b>Analyze Agreement among Panelists</b>	
Overview of the Multiple Factor Analysis Platform .....	167
Example of Multiple Factor Analysis .....	167
Launch the Multiple Factor Analysis Platform .....	169

- Data Format ..... 171
- The Multiple Factor Analysis Report ..... 172
  - Summary Plots ..... 172
  - Consensus Map ..... 173
- Multiple Factor Analysis Platform Options ..... 173
- Statistical Details for the Multiple Factor Analysis Platform ..... 176

**A** **References** ..... 179

**Index** ..... 181

**Consumer Research**

# Chapter 1

## **Learn about JMP**

### **Documentation and Additional Resources**

---

This chapter includes details about JMP documentation, such as book conventions, descriptions of each JMP book, the Help system, and where to find other support.

**Contents**

Formatting Conventions ..... 15

JMP Documentation ..... 16

    JMP Documentation Library ..... 16

    JMP Help ..... 22

Additional Resources for Learning JMP ..... 22

    Tutorials ..... 22

    Sample Data Tables ..... 23

    Learn about Statistical and JSL Terms ..... 23

    Learn JMP Tips and Tricks ..... 23

    Tooltips ..... 23

    JMP User Community ..... 24

    JMP Books by Users ..... 24


    The JMP Starter Window ..... 24

Technical Support ..... 25

---

## Formatting Conventions

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
- Features that are for JMP Pro only are noted with the JMP Pro icon . For an overview of JMP Pro features, visit <https://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:  
<https://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.

Document Title	Document Purpose	Document Content
<i>Discovering JMP</i>	If you are not familiar with JMP, start here.	Introduces you to JMP and gets you started creating and analyzing data. Also learn how to share your results.
<i>Using JMP</i>	Learn about JMP data tables and how to perform basic operations.	Covers general JMP concepts and features that span across all of JMP, including importing data, modifying columns properties, sorting data, and connecting to SAS.



Document Title	Document Purpose	Document Content
<i>Basic Analysis</i>	Perform basic analysis using this document.	<p>Describes these Analyze menu platforms:</p> <ul style="list-style-type: none"> <li>• Distribution</li> <li>• Fit Y by X</li> <li>• Tabulate</li> <li>• Text Explorer</li> </ul> <p>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.</p>
<i>Essential Graphing</i>	Find the ideal graph for your data.	<p>Describes these Graph menu platforms:</p> <ul style="list-style-type: none"> <li>• Graph Builder</li> <li>• Scatterplot 3D</li> <li>• Contour Plot</li> <li>• Bubble Plot</li> <li>• Parallel Plot</li> <li>• Cell Plot</li> <li>• Scatterplot Matrix</li> <li>• Ternary Plot</li> <li>• Treemap</li> <li>• Chart</li> <li>• Overlay Plot</li> </ul> <p>The book also covers how to create background and custom maps.</p>
<i>Profilers</i>	Learn how to use interactive profiling tools, which enable you to view cross-sections of any response surface.	Covers all profilers listed in the Graph menu. Analyzing noise factors is included along with running simulations using random inputs.

Document Title	Document Purpose	Document Content
<i>Design of Experiments Guide</i>	Learn how to design experiments and determine appropriate sample sizes.	Covers all topics in the DOE menu and the Specialized DOE Models menu item in the Analyze > Specialized Modeling menu.
<i>Fitting Linear Models</i>	Learn about Fit Model platform and many of its personalities.	<div>Describes these personalities, all available within the Analyze menu Fit Model platform:</div> <ul style="list-style-type: none"><li>• Standard Least Squares</li><li>• Stepwise</li><li>• Generalized Regression</li><li>• Mixed Model</li><li>• MANOVA</li><li>• Loglinear Variance</li><li>• Nominal Logistic</li><li>• Ordinal Logistic</li><li>• Generalized Linear Model</li></ul>

Document Title	Document Purpose	Document Content
<i>Predictive and Specialized Modeling</i>	Learn about additional modeling techniques.	<p>Describes these Analyze &gt; Predictive Modeling menu platforms:</p> <ul style="list-style-type: none"> <li>• Modeling Utilities</li> <li>• Neural</li> <li>• Partition</li> <li>• Bootstrap Forest</li> <li>• Boosted Tree</li> <li>• K Nearest Neighbors</li> <li>• Naive Bayes</li> <li>• Model Comparison</li> <li>• Formula Depot</li> </ul> <p>Describes these Analyze &gt; Specialized Modeling menu platforms:</p> <ul style="list-style-type: none"> <li>• Fit Curve</li> <li>• Nonlinear</li> <li>• Functional Data Explorer</li> <li>• Gaussian Process</li> <li>• Time Series</li> <li>• Matched Pairs</li> </ul> <p>Describes these Analyze &gt; Screening menu platforms:</p> <ul style="list-style-type: none"> <li>• Response Screening</li> <li>• Process Screening</li> <li>• Predictor Screening</li> <li>• Association Analysis</li> <li>• Process History Explorer</li> </ul> <p>The platforms in the Analyze &gt; Specialized Modeling &gt; Specialized DOE Models menu are described in <i>Design of Experiments Guide</i>.</p>

Document Title	Document Purpose	Document Content
<i>Multivariate Methods</i>	Read about techniques for analyzing several variables simultaneously.	<p>Describes these Analyze &gt; Multivariate Methods menu platforms:</p> <ul style="list-style-type: none"> <li>• Multivariate</li> <li>• Principal Components</li> <li>• Discriminant</li> <li>• Partial Least Squares</li> <li>• Multiple Correspondence Analysis</li> <li>• Factor Analysis</li> <li>• Multidimensional Scaling</li> <li>• Item Analysis</li> </ul> <p>Describes these Analyze &gt; Clustering menu platforms:</p> <ul style="list-style-type: none"> <li>• Hierarchical Cluster</li> <li>• K Means Cluster</li> <li>• Normal Mixtures</li> <li>• Latent Class Analysis</li> <li>• Cluster Variables</li> </ul>
<i>Quality and Process Methods</i>	Read about tools for evaluating and improving processes.	<p>Describes these Analyze &gt; Quality and Process menu platforms:</p> <ul style="list-style-type: none"> <li>• Control Chart Builder and individual control charts</li> <li>• Measurement Systems Analysis</li> <li>• Variability / Attribute Gauge Charts</li> <li>• Process Capability</li> <li>• Pareto Plot</li> <li>• Diagram</li> <li>• Manage Spec Limits</li> </ul>

Document Title	Document Purpose	Document Content
<i>Reliability and Survival Methods</i>	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: <ul style="list-style-type: none"> <li>• Life Distribution</li> <li>• Fit Life by X</li> <li>• Cumulative Damage</li> <li>• Recurrence Analysis</li> <li>• Degradation and Destructive Degradation</li> <li>• Reliability Forecast</li> <li>• Reliability Growth</li> <li>• Reliability Block Diagram</li> <li>• Repairable Systems Simulation</li> <li>• Survival</li> <li>• Fit Parametric Survival</li> <li>• Fit Proportional Hazards</li> </ul>
<i>Consumer Research</i>	Learn about methods for studying consumer preferences and using that insight to create better products and services.	Describes these Analyze > Consumer Research menu platforms: <ul style="list-style-type: none"> <li>• Categorical</li> <li>• Choice</li> <li>• MaxDiff</li> <li>• Uplift</li> <li>• Multiple Factor Analysis</li> </ul>
<i>Scripting Guide</i>	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.
<i>JSL Syntax Reference</i>	Read about many JSL functions on functions and their arguments, and messages that you send to objects and display boxes.	Includes syntax, examples, and notes for JSL commands.


---

**Note:** The **Books** menu also contains two reference cards that can be printed: The *Menu Card* describes JMP menus, and the *Quick Reference* describes JMP keyboard shortcuts.

---

## 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 and view JMP Help on Windows using the **Help > Help Contents**, **Search Help**, and **Help Index** options. On Mac, select **Help > JMP Help**.
- Search the Help at <https://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”](#)
- [“Sample Data Tables”](#)
- [“Learn about Statistical and JSL Terms”](#)
- [“Learn JMP Tips and Tricks”](#)
- [“Tooltips”](#)
- [“JMP User Community”](#)
- [“JMP Books by Users”](#)
- [“The JMP Starter Window”](#)

## 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\14\Samples\Data

On Macintosh: \Library\Application Support\JMP\14\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.

## 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/>.

## JMP Books by Users

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

[https://www.jmp.com/en\\_us/software/books.html](https://www.jmp.com/en_us/software/books.html)

## 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 <https://www.jmp.com/support>, including the technical support phone number.



## Introduction to Consumer Research

### Overview of Customer and Behavioral Research Methods

---

JMP provides a full suite of tools for analyzing consumer and behavioral research data. You collect information about how customers use products or services, how satisfied they are with your offerings, and what new features they might desire. The resulting insights let you create better products and services, happier customers, and more revenue for your organization. Tools for analyzing these consumer research activities are located in the Consumer Research menu. Use the following platforms to analyze your data:

- The Categorical platform enables you to tabulate, plot, and compare categorical responses in your data, including multiple response data. You can use this platform to analyze data from surveys and other categorical response data, such as defect records and study participant demographics. Using the Categorical platform, you can analyze responses from data tables that are organized in many different ways. For more information, see [Chapter 3, “Categorical Response Analysis”](#).
- The Choice platform is designed for use in market research experiments, where the ultimate goal is to discover the preference structure of consumers. Then, this information is used to design products or services that have the attributes most desired by consumers. For more information, see [Chapter 4, “Choice Models”](#).
- The MaxDiff platform is an alternative to using standard preference scales to determine the relative importance of items being rated. A MaxDiff experiment forces respondents to report their most and least preferred options, thereby forcing respondents to rank options in terms of preference. For more information, see [Chapter 5, “MaxDiff”](#).
- The Uplift platform enables you to maximize the impact of your marketing budget by sending offers only to individuals who are likely to respond favorably, even when you have large data sets and many possible behavioral or demographic predictors. You can use uplift models to make such predictions. This method has been developed to help optimize marketing decisions, define personalized medicine protocols, or, more generally, to identify characteristics of individuals who are likely to respond to some action. For more information, see [Chapter 6, “Uplift Models”](#).
- The Multiple Factor Analysis platform enables you to analyze agreement among panelists in sensory data analysis. You can use MFA to analyze studies where items are measured on the same or different attributes by different instruments, individuals, or under different circumstances. For more information, see [Chapter 7, “Multiple Factor Analysis”](#).

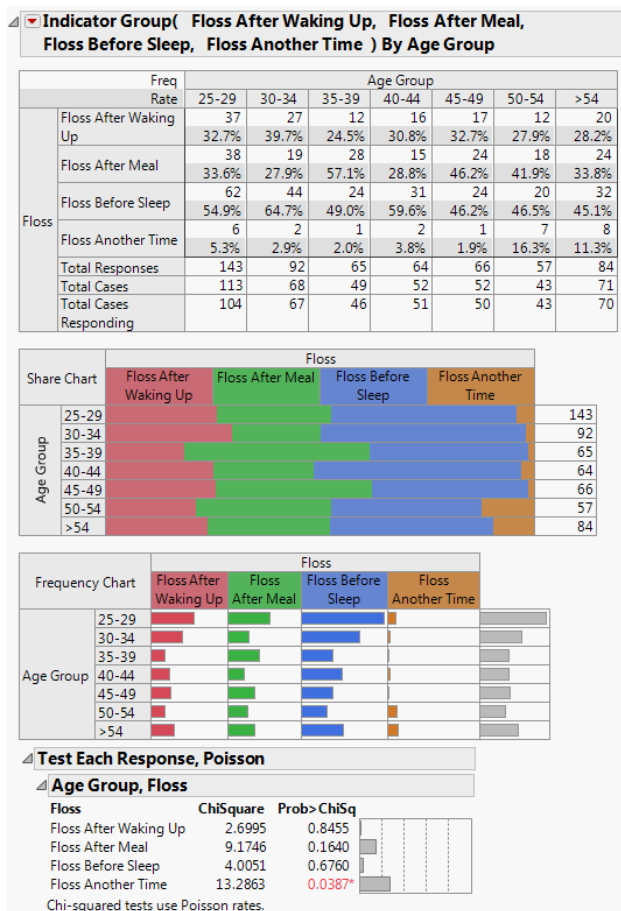


## Categorical Response Analysis

### Analyze Survey and Other Counting Data

The Categorical platform enables you to tabulate, plot, and compare categorical response data, including multiple response data. You can use this platform to analyze data from surveys and other categorical response data, such as defect records and study participant demographics. With the Categorical platform you can analyze responses from a rich variety of organizations of data. The Categorical launch window enables you to specify analyses as well as data formats.

**Figure 3.1** Categorical Analysis Example



**Contents**

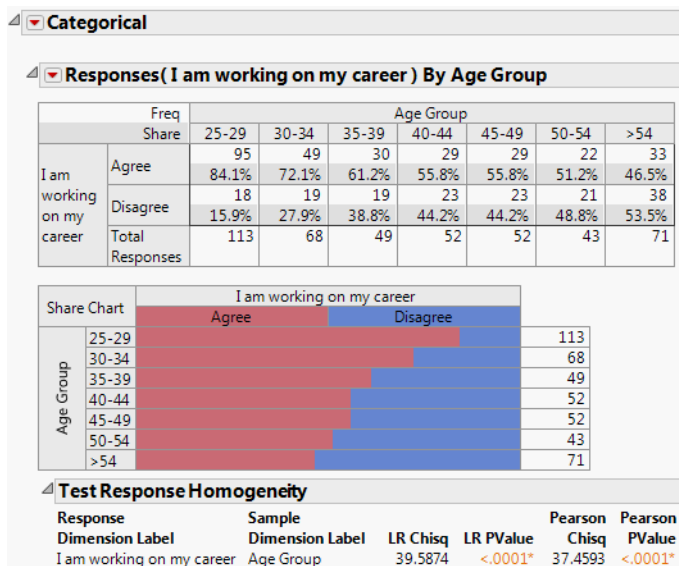
Example of the Categorical Platform .....	31
Launch the Categorical Platform .....	32
Response Roles .....	32
Columns Roles .....	34
Other Launch Window Options .....	35
The Categorical Report .....	36
Categorical Platform Options .....	37
Report Options .....	38
Statistical Testing Options .....	39
Additional Categorical Platform Options .....	41
Responses Table Options .....	44
Comparison Letters .....	44
Supercategories .....	46
Supercategories Options .....	46
Set Preferences .....	47
Additional Examples of the Categorical Platform .....	48
Example of the Test for Response Homogeneity .....	48
Example of the Multiple Response Test .....	49
Example of the Cell Chisq Test .....	51
Example of Compare Each Sample with Comparison Letters .....	52
Example of Compare Each Cell with Comparison Letters .....	54
Example of User-Specified Comparison with Comparison Letters .....	56
Example of Conditional Association and Relative Risk .....	57
Example of Rater Agreement .....	59
Example of Repeated Measures .....	60
Examples of the Multiple Response Tab .....	61
Example of Mean Score with Comparison Letters .....	66
Example of a Structured Report .....	68
Statistical Details for the Categorical Platform .....	69
Rao-Scott Correction .....	69

## Example of the Categorical Platform

This example uses the Consumer Preferences.jmp sample data table, which contains survey data on people's attitudes and opinions, and questions concerning oral hygiene. You can use the categorical platform to compare the response to a question by age groups.

1. Select **Help > Sample Data Library** and open Consumer Preferences.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select I am working on my career and click **Responses** on the Simple tab.
4. Select Age Group and click **X, Grouping Category**.
5. Click **OK**.
6. Click the Categorical red triangle and select **Crosstab Transposed**.
7. Click the Categorical red triangle and select **Test Response Homogeneity**.

**Figure 3.2** Question Responses by Age Group



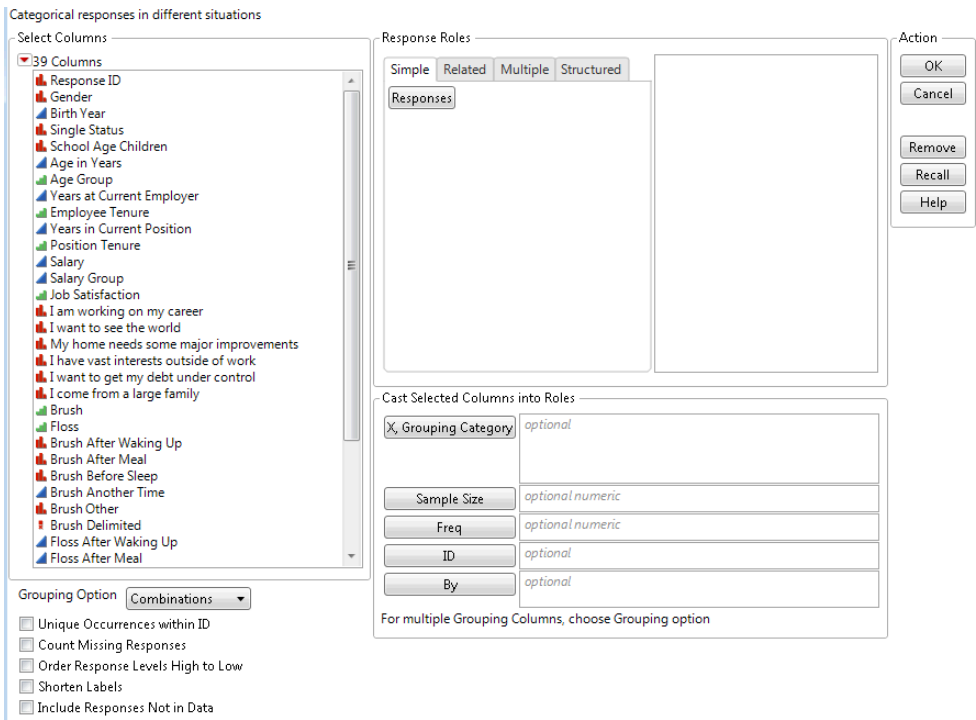
The question had two possible responses: Agree and Disagree. There are seven age groups of respondents with 43 to 113 respondents per group. The youngest age group (25 - 29) was the largest group surveyed with 113 responses. The highest positive response rate was 84.1% in the youngest age group. The share chart displays the share of respondents who agree or disagree with the question for each age group. As age increases, the portion of respondents who agree with the question decreases. Thus, the older respondents tend to be working on their careers less than the younger respondents.

**Tip:** Colors on the charts can be defined by using the Value Colors column property. See the Column Info Window chapter in *Using JMP*.

# Launch the Categorical Platform

Launch the Categorical platform by selecting **Analyze > Consumer Research > Categorical**.

**Figure 3.3** Categorical Platform Launch Window



## Response Roles

The launch window includes tabs for three specific types (or categories) of response roles (Simple, Related, and Multiple) and a Structured tab where you can create custom data summaries. The response role corresponds to the type of responses you want to analyze. Options on each tab correspond to how the responses are organized in your data table.



## Simple Tab

The Simple tab contains an option for the analysis of results that are contained in a single column.

**Responses** Adds one or more columns to your analysis. If multiple columns are selected, the categorical report contains a report for each individual column.

## Related Tab

The Related tab contains options for the analysis of a set of related columns.

**Aligned Responses** Summarizes data from multiple responses with the same response levels in a single report. This option is useful for survey data when you have many questions with the same set of responses. You can quickly summarize and compare response trends for all of the questions at once.

**Repeated Measures** Summarizes data from multiple columns where each column contains responses made at different time points. When an individual responds at multiple time points, the samples are called overlapping. When there are overlapping samples, the Kish correction is used. See Kish (1965, sec. 12.4).

**Rater Agreement** Summarizes data from multiple columns where each column is a rating for the same question, but by different individuals (raters).

## Multiple Tab

The Multiple tab contains options for the analysis of multiple responses recorded in one or more columns. A set of multiple responses could be from a survey where the response set allows for more than one choice (check all that apply questions). Another source of multiple responses is defect data where an item can have multiple defects. The options on the multiple tab are specific to how the data are organized in your data table.

**Multiple Response** Summarizes data from multiple columns where each column contains one response. The number of columns selected is the maximum number of responses for a single row. There can be many blanks in the columns.

**Multiple Response by ID** Summarizes data from a single column of responses with a second column containing an ID for the subject or part. This data structure can be thought of as a stacked format.

**Multiple Delimited** Summarizes data from a single column that contains multiple responses separated by a comma, semicolon, or tab.

**Indicator Group** Summarizes multiple responses that are stored in indicator columns. The data table has a column for each possible response, and each column is an indicator (for example, 0 and 1).

**Response Frequencies** Summarizes multiple responses that are stored in columns with frequency counts. This data format is the summarized version of the Indicator Group format.

**Free Text** Summarizes text data. The Free Text option launches a Text Explorer report inside the Categorical report window. See the Text Explorer chapter in the *Basic Analysis* book.

## Structured Tab

The Structured tab enables you to construct custom tables of summary statistics.

- Create side-by-side, crossed, and nested data summaries with the interactive table builder. The Structured table considers the innermost terms on the side of the table as responses and all other terms as grouping factors.
- Create multiple tables in a single the launch window.
- Use delimited multiple response columns when the column modeling type is set to Multiple Response. For more information about column modeling types, see [“The Column Info Window”](#) in the *Using JMP* book.

See [“Example of a Structured Report”](#) on page 68.

---

**Tip:** Use the Structured tab to test for response homogeneity with multiple response data.

---

## Columns Roles

The following roles are available:

**X, Grouping Category** Assigns a column as a grouping category. The responses are summarized for each group. If more than one grouping column is used, then by default the tabulation is nested (the Combinations Grouping Option). Use the Grouping Option to change the summarization.

**Sample Size** Assigns a column to define the number of individual units in the group to which that frequency is applicable, for multiple response roles with summarized data. For example, a Freq column might indicate 50 defects, where the sample size variable would reflect the defects for a batch of 100 units.

**Freq** Assigns a frequency variable to this role. This option is useful if your data are summarized.

- ID** Assigns a column that identifies the respondent. This option is required only when Multiple Response by ID is selected, and it is not used if entered for other response types.
- By** Produces a separate report for each level of the By variable. If more than one By variable is assigned, a separate report is produced for each possible combination of the levels of the By variables.

## Other Launch Window Options

Additional options are located in the lower left of the launch window. Alternatively, these options can be selected from the Categorical red triangle menu after you click OK in the launch window.

**Grouping Option** Defines how to use grouping variables in your analysis when more than one grouping column is specified.

**Combinations** Analyzes the response for combinations of the grouping variables. The first column in the grouping list is the outermost group.

**Each Individually** Analyzes the response for each grouping variable individually.

**Both** Provides reports for combinations of the grouping variables as well as for each grouping variable individually.

**Unique Occurrences within ID** Counts unique response levels within a participant. An ID variable must be specified.

**Count Missing Responses** Includes missing values as a category. Missing values can be either empty cells or a defined missing code in the Missing Value Codes column property. If a column contains only missing values, the missing values are counted regardless of this option.

---

**Note:** If this option is not selected, missing values are excluded from the analysis.

---

**Order Response Levels High to Low** Orders the responses from high to low. (The default ordering is low to high.) This option applies only to the response, not to grouping categories.

---

**Tip:** Use the Value Ordering column property to define a specific category ordering. See the Column Info Window chapter in *Using JMP*.

---

**Shorten Labels** Shortens value labels by removing prefixes and suffixes that are common to all labels.

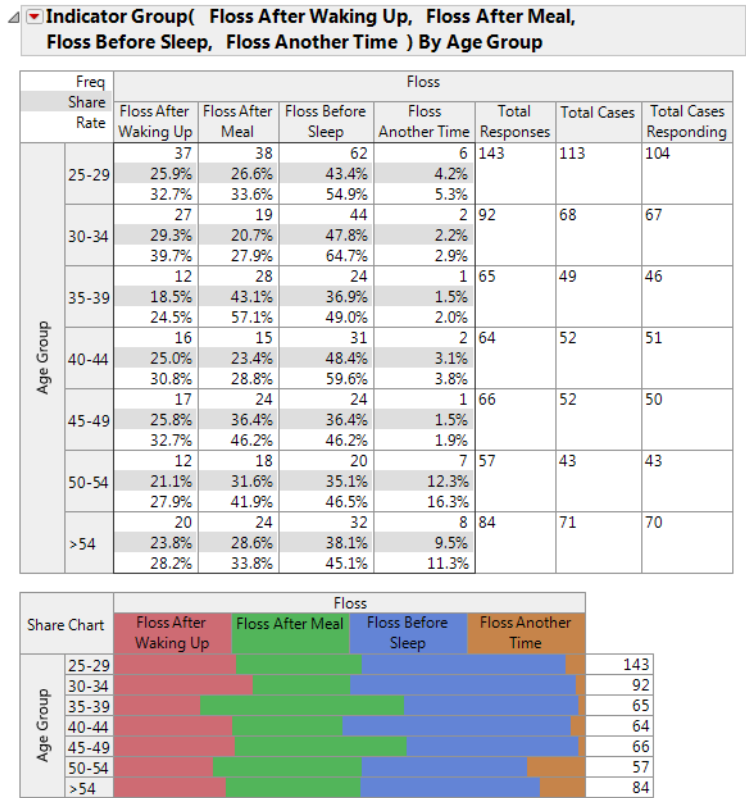
**Note:** This option applies only to value labels, not column names.

**Include Responses Not in Data** Includes categories with no responses in the report. The categories with no responses must be specified in the Value Labels column property. This option applies only to responses. Grouping categories include only categories with responses.

# The Categorical Report

The initial Categorical report shows a cross tabulation and a share chart for each set of responses selected. If you used the Structured tab, the initial report shows only a cross tabulation.

Figure 3.4 The Initial Categorical Report



The upper left corner of the table lists the items (Freq, Share, and Rate) that are included in each cell of the table. These items can be removed using the options in the red triangle menu.

- The Frequency count (Freq) is provided for each category with the total frequency (Total Responses) at the right of the table. When there are multiple responses, the summary columns at the right of the table also include the number of cases or the number of rows (Total Cases) and the number of responders (Total Cases Responding).
- The Share of Responses (Share) is determined by dividing each count by the total number of responses. The number represents the percent of the response among all the responses in the sample (frequency divided by response total). This is either a row percentage or a column percentage for transposed tables.
- The Rate (excluding missing values) is the frequency of response divided by the total cases responding.

In Figure 3.4, the number of responses and cases for each age group are displayed. Consider the first row of the table with the results for those in the 25 through 29 age group.

- In the first cell of the table there are 37 responses. This is the frequency of the respondents, aged 25 - 29, who floss after they wake up.
- There are 143 total responses for the 25 - 29 age group. Of the 143 responses, 25.9% (37/143) is the share who Floss After Waking Up. In other words, 25.9% of the times that 25 through 29-year-olds floss was after they wake up.
- Of the 113 total cases (respondents) in the 25 through 29 group, 32.7% (37/113) is the rate who responded that they floss after they wake up. In other words, 32.7% of 25 through 29-year-olds floss after they wake up.
- Because the number of responses (143) is greater than the number of cases responding (104), there are respondents in the 25 through 29 group who selected more than one response.
- Because the total number of cases (113) is greater than the total cases responding (104), there were nine responders in the 25 through 29 group who did not answer this question.

---

## Categorical Platform Options

The Categorical red triangle menu contains options that enable you to customize the report according to your needs. The options that are available in a particular report are determined by the response roles, use of grouping categories, and options selected in the launch window.

- [“Report Options”](#)
- [“Statistical Testing Options”](#)
- [“Additional Categorical Platform Options”](#)

- “Responses Table Options”

## Report Options

**Frequencies** Shows or hides the frequency in the Crosstab table. The frequency is the count of the responses in each category.

**Share of Responses (Share)** Shows or hides the share of responses in the Crosstab table. The share of responses is the percent of responses in each category.

**Rate Per Case (Rate)** (Available only for multiple responses.) Shows or hides rate per case in the Crosstab table. The rate per cases is the percent of responses in each category based on the total number of cases (regardless of if they were a respondent).

**Rate per Case Responding** (Available only for multiple responses.) Shows or hides rate per case responding in the Crosstab table. The rate per case responding is the percent of responses in each category based on the cases that responded.

**Share Chart** Shows or hides a divided bar chart. The bar length is proportional to the percentage of responses for each type. The column on the right shows the number of responses in each grouping category. If no grouping category is used, the column on the right shows the total number of responses.

---

**Tip:** You can change the colors in the share chart using the Value Colors column property. See the Column Info Window chapter in *Using JMP*.

---

**Frequency Chart** Shows or hides a Frequency Chart. The bars reflect the frequency of responses within each group. The scale is consistent across the chart. The gray bars at the far right are the total number of responses in each grouping category.

---

**Tip:** You can change the colors in the frequency chart using the Value Colors column property. See the Column Info Window chapter in *Using JMP*.

---

**Transposed Freq Chart** Shows or hides a transposed Frequency Chart. The bars reflect the frequency of responses within each group. The responses are the rows and the grouping levels are the columns in chart. The totals for each grouping level are represented by gray bars in the bottom row of the chart.

**Crosstab** Shows or hides the Crosstab table. The Crosstab table displays the response categories as column headings, and displays the grouping levels (when used) as row labels. The upper left cell of the table shows the labels for the items in each cell of the table (Freq, Share, and Rate). If the report contains a transposed Crosstab table, this option removes the transposed Crosstab table from the report.

**Crosstab Transposed** Shows or hides a transpose of the Crosstab table. The transposed Crosstab table displays the response categories as row labels, and displays the grouping levels (when used) as column headings. The upper left cell of the table shows the labels for the items in each cell of the table (Freq, Share, and Rate). If the report contains a Crosstab table, this option removes the Crosstab table from the report.

## Statistical Testing Options

The statistical testing options that are available depend on the response roles and the use of grouping variables in the analysis. Options include tests of response homogeneity, association, relative risk, and agreement.

**Test Multiple Response** (Available only for multiple response data with one or more grouping categories.) See [“Example of the Multiple Response Test”](#) on page 49. Contains the following tests for independence of responses across each grouping category:

**Count Test, Poisson** Shows or hides a test of independence of rates that uses Poisson regression. The frequency per unit is modeled by the sample categorical variable. The result is a likelihood ratio chi-square test of whether the rate of each individual response differs across grouping levels.

**Homogeneity Test, Binomial** Shows or hides the likelihood ratio chi-square test of independence for each individual response level. Each response category has a binomial distribution (selected or not selected).

**Test Response Homogeneity** (Available for a single response variable (simple or aligned) with one or more grouping categories. Available for multiple responses with one or more grouping within the Structured tab.) Shows or hides a report that contains tests for response homogeneity that depend on your approach:

- For independence of responses across grouping categories, likelihood ratio and Pearson chi-square tests are provided. See [“Example of the Test for Response Homogeneity”](#) on page 48.
- For analysis of multiple responses within the Structured tab, a Rao-Scott Chi-square test is provided.

**Cell Chisq** Shows or hides  $p$ -values for each cell in the table for a chi-square test of independence. A small  $p$ -value indicates a cell with an observed value that is larger or smaller than expected under the assumption that the rows are independent of the columns. The  $p$ -values are colored and shaded according to whether the count is larger or smaller than expected. See [“Example of the Cell Chisq Test”](#) on page 51.

**Compare Each Sample** (Available only for single responses with one or more grouping variables.) Shows or hides a report that contains pairwise likelihood ratio and Pearson

chi-square tests for independence of responses across levels of a grouping variable. See [“Example of Compare Each Sample with Comparison Letters”](#) on page 52.

**Compare Each Cell** (Available only for single and multiple responses with one or more grouping variables.) Shows or hides pairwise likelihood ratio chi-square, Pearson chi-square and Fisher’s exact tests for independence of each level of the response versus all other levels combined across levels of a grouping variable. See [“Example of Compare Each Cell with Comparison Letters”](#) on page 54.

**Relative Risk** (Available when the grouping variable has two levels and either the response has two levels or is a multiple response and the Unique occurrences within ID option has been selected.) Shows or hides the relative risks for a two-level grouping variable for each level of the response. See [“Example of Conditional Association and Relative Risk”](#) on page 57.

**Conditional Association** (Available only when the Unique occurrences within ID option has been selected.) Shows or hides the conditional probability of one response level given a second response level. See [“Example of Conditional Association and Relative Risk”](#) on page 57.

**Agreement Statistic** (Available only for Rater Agreement responses.) Shows or hides the Kappa coefficient of agreement and the Bowker test of symmetry. See [“Example of Rater Agreement”](#) on page 59.

**Transition Report** (Available only for Repeated Measures responses.) Shows or hides transition counts and rates matrices for changes in responses across time. See [“Example of Repeated Measures”](#) on page 60.

**Test Options** Options available in this menu depend on your selected analysis.

**ChiSquare Test Choices** Specifies which chi-square tests of homogeneity are calculated for single responses. You can choose between Both LR and Pearson, LR Only, or Pearson Only, where LR refers to likelihood ratio.

**Show Warnings** Shows small sample size warnings for chi-square tests.

**Order by Significance** Reorders the reports so that the most significant reports are at the top.

**Hide Nonsignificant** Suppresses reports that are non-significant.



## Additional Categorical Platform Options

**Total Responses** Shows or hides the sum of the frequency counts for the response in Crosstab tables and share charts. The total is across each grouping category, when a grouping variable is used.

**Response Levels** Shows or hides the categories for the response column in Crosstab tables and share charts.

**Show Supercategories** (Available only when one or more supercategories is defined.) Shows or hides columns for supercategories in the Crosstab table and the Frequency Chart. For more information about supercategories, see [“Supercategories”](#) on page 46.

---

**Tip:** This option shows or hides the Supercategories. To hide the individual categories within the supercategory, use the Hide option in the Supercategories column property. Alternately, use the Response Levels option to hide all response levels so that only Supercategories remain unhidden.

---

**Total Cases** (Available only for multiple response columns.) Shows or hides a column in the Crosstab table that contains the number of cases (participants) in each group.

**Total Cases Responding** (Available only for multiple response columns.) Shows or hides a column in the Crosstab table that contains the number of cases (participants) who responded at least once. People who did not respond at all are not included. The total cases responding is less than or equal to the total cases.

**Mean Score** Shows or hides a column in the Crosstab table and share chart that contains the overall mean of the response or the mean for each grouping category. The mean is calculated based on a numerical value assigned to each response category.

- For numeric categories, the numeric value is the actual value.
- For non-numeric categories the value is the value assigned to the categories by the Value Scores column property.
- For categories without value scores, the value is based on a default assignment of 1 to the number of categories.

See [“Example of Mean Score with Comparison Letters”](#) on page 66.

**Mean Score Comparisons** Shows or hides the Compare Means column in the Crosstab table. This column compares the mean scores across grouping categories using the unpooled Satterthwaite  $t$  test for pairwise comparisons. See SAS Institute Inc. (2017). The results of the comparison are shown using letters. For more information about comparison letters, see [“Comparison Letters”](#) on page 44. For more information about specifying comparison groups, see [“Example of User-Specified Comparison with Comparison Letters”](#) on page 56.

**Std Dev Score** Shows or hides a column in the Crosstab table that contains the overall standard deviation of the response or the standard deviation of each grouping category.

**Order by Mean Score** (Appears only when there are more than one response and there are no grouping variables in the analyses.) Orders the response reports by the mean score.

**Save Tables** Saves specific portions of the reports to a new data table. Each option creates an individual data table for each report. The options available in this menu depend on your selected analysis.

---

**Note:** Supercategories are not included in the new tables.

---

**Save Frequencies** Saves the frequency counts from the Crosstab table to a new data table.

**Save Share of Responses** Saves the share of responses from the Crosstab table to a new data table.

**Save Contingency Table** Saves the complete Crosstab Table to a new data table.

**Save Rate Per Case** Saves the rates per case from the Crosstab table to a new data table.

**Save Transposed Frequencies** Saves the transposed frequency counts from the Crosstab table to a new data table.

**Save Transposed Share of Responses** Saves the transposed share of responses from the Crosstab to a new data table.

**Save Transposed Rate Per Case** Saves the transposed rates per case from the Crosstab table to a new data table.

**Save Test Rates** Saves the results of the Test Multiple Response option to a new data table.

**Save Test Homogeneity** Saves the results of the Test Response Homogeneity option to a new data table.

**Save Mean Scores** Saves the mean scores for each sample group to a new data table.

**Save tTests and pValues** Save *t* tests and *p*-values from the Mean Score Comparisons report to a new data table.

**Save Excel File** Creates a Microsoft Excel spreadsheet with the structure of the Crosstab format report. The option maps all of the tables to one sheet, with the response categories as rows, the sample levels as columns, sharing the headings for sample levels across multiple tables. When there are multiple elements in each table cell, you have the option to make them multiple or single cells in Microsoft Excel.

**Filter** Shows or hides the local data filter that enables you to filter the data used in a specific report. Sample levels that contain no responses are always hidden. To show the filtered headings in reports, select **Include Responses Not in Data** in the launch window. You can also select the **Set Preferences** red triangle menu and then select **Include Responses Not in Data**.

**Contents Summary** Shows or hides a Contents Summary report at the top of the Categorical report. The Contents Summary collects all of the tests and mean scores into a summary with links to the associated report.

**Show Columns Used in Report** Shows or hides Columns Used in Report information. This option affects only columns that have an SPSS or SAS Name or SPSS or SAS Label column property. When you import survey data from SAS or SPSS, the Name and Label column properties are automatically added to your JMP table. You can add a SAS or SPSS Name or Label column property using the Other column property. For example, if you use the SAS or SPSS Name column property to store a survey question, the column name can be a short name.

**Format Elements** Enables you to specify formats for Frequencies, Shares and Rates, and how zeros are displayed. By default, Frequencies are Fixed Dec format with 7 Width and 0 Decimals and Shares and Rates are Percent format with 6 Width and 1 Decimal.

**Arrange in Rows** Arranges the reports across the page instead of down the page. Enter the number of reports that you want to view across the window.

**Set Preferences** Enables you to set preferences for future launches of the Categorical platform in the current JMP session as well as in future JMP sessions. For more information, refer to [“Set Preferences”](#) on page 47.

**Category Options** Contains options (Grouping Option, Count Missing Responses, Order Response Levels High to Low, Shorten Labels, and Include Responses Not in Data) that are also presented on the launch window. If these options are selected here, the platform updates with the new setting. For more information about the Category Options, see [“Other Launch Window Options”](#) on page 35.

**Force Crosstab Shading** Forces shading on Crosstab reports even if the preference is set to no shading. If this option is not selected, the Crosstab reports are shaded according to the current setting of the Shade Alternate Table Rows preference.

**Relaunch Dialog** Enables you to return to the launch window and edit the specifications for an analysis.

See the JMP Reports chapter in the *Using JMP* book for more information about the following options:

**Local Data Filter** Shows or hides the local data filter that enables you to filter the data used in a specific report.

**Redo** Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

**Save Script** Contains options that enable you to save a script that reproduces the report to several destinations.

**Save By-Group Script** Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

## Responses Table Options

**Show Letters** Shows or hides the column letter IDs in the Crosstab table. These letters are used in many of the tests of homogeneity and are displayed automatically for those tests.

**Specify Comparison Groups** Enables you to specify specific comparison groups for tests of homogeneity. Use group comparison letters separated by a slash to represent each group. Separate multiple groups by commas. For example, to test A with E, B with D, and C with F, specify the groups as “A/E, B/D, C/F”. A Compare Each Cell report is provided for the defined comparison groups. See [“Example of User-Specified Comparison with Comparison Letters”](#) on page 56.

**Remove** Removes the report from the report window.

---

**Caution:** The Remove option cannot be undone.

---

---

## Comparison Letters

The Compare Each Cell, Compare Each Sample, and Mean Score Comparisons options use letters to identify sample levels. For more than 26 levels, numbers are appended to the letters. The letters are shown in the sample level headings of the Crosstab table when a comparison option is turned on.

**Figure 3.5** Crosstab Table with Comparison Letters

▼ **Responses( country ) By type**

		Freq	country			
		Share	American	European	Japanese	Compare
type	Family	A	74 47.7%	15 9.7%	66 42.6%	155 B
	Sporty	B	23 23.0%	21 21.0%	56 56.0%	100 c
	Work	C	18 37.5%	4 8.3%	26 54.2%	48 *

Default Comparison Groups: A/B/C

Shows letter of the category it is significantly different from at the higher share level

\* Base count warning 100 Uppercase Alpha Level 0.05

\*\* Base count minimum 30 Lowercase Alpha Level 0.1

▶ **Compare Each Sample**

If two sample levels are significantly different, the letter of the sample level with a smaller share of responses is placed into the comparison cell of the other level. An Uppercase letter indicates a stronger difference between levels than a lowercase letter. The default alpha level (significance level) for an uppercase letter is 0.05 and 0.10 for a lowercase letter. In Figure 3.5 the B indicates that there is a difference in the country of origin for Sporty and Family cars at the 0.05 significance level. The B is in the row for Family cars because the total responses for Family cars (155) is greater than the total for Sporty cars (100). The c in the Sporty row indicates that there is a difference at the 0.10 level between the country of origin when comparing Sporty to Work cars. The c is in the Sporty row because the total responses (100) is greater than the total responses (48) for Work cars.

Warnings for small counts are also included in the comparison cells. One asterisk indicates that the level has fewer than 100 responses and two asterisks indicate fewer than 30 responses. In Figure 3.5, notice that the row for Work has 48 total responses and is labeled with a single asterisk. You can change the alpha levels and warning counts in the Categorical platform preferences. For details about changing preferences, see [“Set Preferences”](#) on page 47.

---

**Tip:** If you want only one set of comparison letters in your report, set the lowercase letter alpha to 0 in the preferences.

---

See [“Example of Compare Each Sample with Comparison Letters”](#) on page 52, [“Example of Compare Each Cell with Comparison Letters”](#) on page 54, and [“Example of Mean Score with Comparison Letters”](#) on page 66.

---

## Supercategories

The term *supercategories* refers to the aggregation of response categories. For example, when using a five-point rating scale, you might want to know the percent of responses in the top two ratings (top two boxes). Such a grouping of responses can be defined using the Supercategories column property.

Supercategories add additional columns to Crosstab tables and Frequency Charts. Share Charts do not show supercategories, and supercategories are not applied to grouping columns.

To create a supercategory, follow these steps:

1. Select a column in your data table that contains categories that you would like to aggregate.
2. Select **Cols > Column Info**.
3. Click **Column Properties** and select **Supercategories**.
4. (Optional) To change the default name of the supercategory, enter a Supercategory Name.
5. Select one or more categories from the Column's Categories list.
6. Click **Add**.
7. (Optional) Select the supercategory and click the Supercategories red triangle menu for additional options.

## Supercategories Options

The following options are available in the Supercategories red triangle menu in the Column Properties window:

**Hide** Hides categories within a supercategory in the Crosstab table and frequency chart.

---

**Tip:** If you want the flexibility to show or hide the individual categories in your reports, then do not use the Hide option. Use the Response Level option in the Categorical red triangle menu.

---

**Net** (Available only for a multiple response column.) Prevents individual respondents from being counted twice when they appear in more than one supercategory.

**Add Mean** Includes mean statistics in the report.

**Add Std Dev** Includes standard deviation statistics in the report.

**Add All** Includes total responses in the report. By default, the Total Responses column is always included.

**Note:** Supercategories are supported for all response effects except Repeated Measures and Rater Agreement.

## Set Preferences

The Categorical red triangle menu has a Set Preferences option to enable you to specify settings and preferences.

**Figure 3.6** Set Preferences Window

Options are initialized to current state. Choose the option states, and check the Set check box to save option to Preferences as new default.

☒ Submit Platform Preferences  
☐ Create Platform Preference Script

Special Handling of Responses

☐ Set ☐ Count Missing Responses

☐ Set ☐ Order Response Levels High to Low

☐ Set ☐ Shorten Labels

☐ Set ☐ Include Responses Not in Data

☐ Set ☐ Unique Occurrences within ID

Which format to show tables

☐ Set ☐ Crosstab Transposed

Which elements to show in reports

☐ Set ☒ Frequencies

☐ Set ☒ Share Of Responses

☐ Set ☒ Rate Per Case

☐ Set ☐ Mean Score

☐ Set ☐ Mean Score Comparisons

☐ Set ☐ Std Dev Score

Which Graphs to show

☐ Set ☐ Share Chart

☐ Set ☒ Frequency Chart

☐ Set ☐ Transposed Freq Chart

Which detail reports

These only appear in specific contexts even if default is on

☐ Set ☐ Test Response Homogeneity

☐ Set ☐ Compare Each Pair

☐ Set ☐ Compare Each Cell

☐ Set ☐ Cell Chisq

☐ Set ☐ Show Warnings

☐ Set ☐ ChiSquare Test Choices Both LR and Pearson ▼

Options for Compare Each Cell

☐ Set Uppercase Alpha Level 0.05

☐ Set Lowercase Alpha Level 0.1

☐ Set Base count minimum 30

☐ Set Base count warning 100

☐ Set Compare Each Cell Test Fisher's Exact Test ▼

Options for Structured Marginals

☐ Set ☒ Total Responses

☐ Set ☒ Response Levels

☐ Set ☒ Show Supercategories

☐ Set ☒ Total Cases

☐ Set ☒ Total Cases Responding

OK Cancel

Select the **Set** box for the options that you want to set. Select the option box if you want the option to appear by default, or deselect the option box if you do not want the option to appear by default. To submit the changes that you make to the platform preferences, select the **Submit Platform Preferences** box. To save the change that you make as a preference script, select the **Create Platform Preference Script** box. When the Categorical platform is launched, the

preferences associated with the current preference set are used to create the Categorical report.

---

**Note:** Running the script submits the preferences to the platform preferences. You can use the platform preference script to share a preference set among multiple users, or to save the settings for specific projects.

---

---

## Additional Examples of the Categorical Platform

- [“Example of the Test for Response Homogeneity”](#)
- [“Example of the Multiple Response Test”](#)
- [“Example of the Cell Chisq Test”](#)
- [“Example of Compare Each Sample with Comparison Letters”](#)
- [“Example of Compare Each Cell with Comparison Letters”](#)
- [“Example of User-Specified Comparison with Comparison Letters”](#)
- [“Example of Conditional Association and Relative Risk”](#)
- [“Example of Rater Agreement”](#)
- [“Example of Repeated Measures”](#)
- [“Examples of the Multiple Response Tab”](#)
- [“Example of Mean Score with Comparison Letters”](#)
- [“Example of a Structured Report”](#)

### Example of the Test for Response Homogeneity

This example uses the Car Poll.jmp sample data table, which contains data collected from car polls. The data include demographics about the individuals polled and information about the cars that they own. You want to explore the relationship between marital status and origin of car. You also want to test for the homogeneity of the responses. That is, you want to test to see whether the distribution of the origin of cars is the same for married and single respondents.

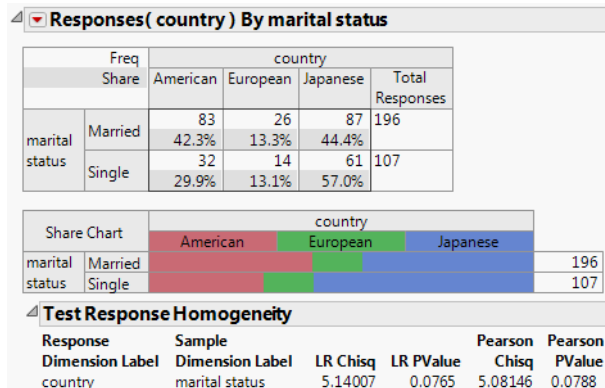
There are two versions of this test, the Pearson and the Likelihood Ratio, both provide chi-square statistics and  $p$ -values.

1. Select **Help > Sample Data Library** and open Car Poll.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select country and click **Responses** on the Simple tab.
4. Select marital status and click **X, Grouping Category**.



5. Click **OK**.
6. Click the Categorical red triangle and select **Test Response Homogeneity**.

**Figure 3.7** Test Response Homogeneity



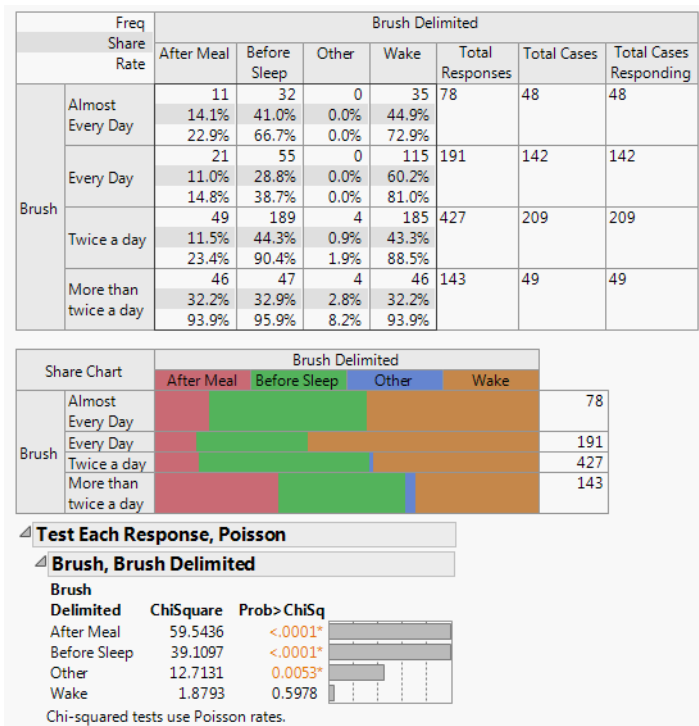
The Share Chart indicates that the married group is evenly split between ownership of American and Japanese cars. In the single group, Japanese cars are the most frequently owned. The test for response homogeneity has a significance of about 0.08. Therefore, the difference in response probabilities across marital status is not statistically significant at an alpha level of 0.05.

## Example of the Multiple Response Test

This example uses the Consumer Preferences.jmp sample data table, which contains survey data on people's attitudes and opinions, as well as questions concerning oral hygiene. You can use the Test Multiple Response option to test if the response rates for each brushing time (Brush Delimited) is the same across groups (Brush). The groups are defined by the frequency that responders brush their teeth.

1. Select **Help > Sample Data Library** and open Consumer Preferences.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select Brush Delimited and click **Multiple Delimited** on the Multiple tab.
4. Select Brush and click **X, Grouping Category**.
5. Click **OK**.
6. Click the Categorical red triangle and select **Test Multiple Response > Count Test, Poisson**.

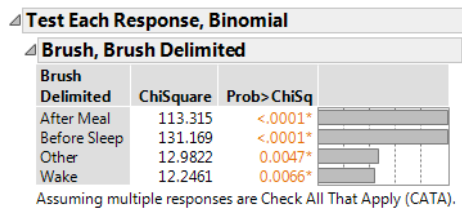
Figure 3.8 Test Multiple Response, Poisson



The  $p$ -values show that the response rates for After Meal, Before Sleep, and Other are significantly different across brushing groups. Wake is not significantly different across brushing groups. From the Crosstab table, you can see that most people brush their teeth when they wake up regardless of how frequently they brush their teeth.

7. Click the Categorical red triangle menu and select **Test Multiple Response > Homogeneity Test, Binomial**.

Figure 3.9 Test Multiple Response, Binomial



The Homogeneity Test, Binomial option always produces a larger test statistic (and therefore a smaller  $p$ -value) than the Count Test, Poisson option. The binomial distribution compares not only the rate at which the response occurred (the number of people who

reported that they brush upon waking) but also the rate at which the response did not occur (the number of people who did not report that they brush upon waking).

In this example, the proportion of responders for each response (After Meal, Before Sleep, Wake, and Other) differ across the age groups. The  $p$ -value for each response is less than 0.05.

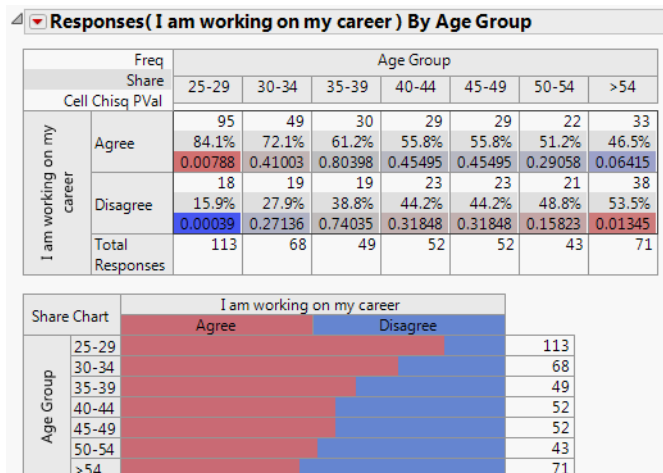
**Note:** JMP detects a multiple response column by the Multiple Response modeling type or the Multiple Response column property.

## Example of the Cell Chisq Test

This example uses the Consumer Preferences.jmp sample data table, which contains survey data on people's attitudes and opinions, as well as questions concerning oral hygiene. You explore the distribution of responses to the statement "I am working on my career" across age groups.

1. Select **Help > Sample Data Library** and open Consumer Preferences.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select I am working on my career and click **Responses** on the Simple tab.
4. Select Age Group and click **X, Grouping Category**.
5. Click **OK**.
6. Click the Categorical red triangle and select **Crosstab Transposed**.
7. Click the Categorical red triangle and select **Cell Chisq**.

**Figure 3.10** Cell Chisq



Small  $p$ -values indicate that there is a significant difference between the observed cell count and the expected cell count. The  $p$ -values are colored by significance level from dark red for cells with significantly higher counts than expected to dark blue for cells with significantly lower counts than expected. The expected cell count is based on the observed row and column totals.

For example, the expected number of responses in the 25 through 29-year-old group who agree is  $(287 \cdot 113) / 448 = 72.4$ ; the observed value was 95. This observed value, with a  $p$ -value of 0.00788, is significantly larger than the expected value. The number of responses in the 25 through 29 group who agree with “I am working on my career” is higher than expected were the response to this question was independent of age.

## Example of Compare Each Sample with Comparison Letters

This example uses the Consumer Preferences.jmp sample data table, which contains survey data on people’s attitudes and opinions, as well as questions concerning oral hygiene. You explore the distribution of responses to the statement “I am working on my career” between each age group.

1. Select **Help > Sample Data Library** and open Consumer Preferences.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select I am working on my career and click **Responses** on the Simple tab.
4. Select Age Group and click **X, Grouping Category**.
5. Click **OK**.
6. Click the Categorical red triangle and select **Compare Each Sample**.

Figure 3.11 Compare Each Sample

Responses (I am working on my career) By Age Group

Freq Share		I am working on my career		Total Responses	Compare
		Agree	Disagree		
Age Group	25-29 A	95 84.1%	18 15.9%	113	b, C, D, E, F, G
	30-34 B	49 72.1%	19 27.9%	68	d, e, F*
	35-39 C	30 61.2%	19 38.8%	49	*
	40-44 D	29 55.8%	23 44.2%	52	*
	45-49 E	29 55.8%	23 44.2%	52	*
	50-54 F	22 51.2%	21 48.8%	43	*
	>54 G	33 46.5%	38 53.5%	71	B*

Default Comparison Groups: A/B/C/D/E/F/G

Shows letter of the category it is significantly different from at the higher share level

\* Base count warning 100 Uppercase Alpha Level 0.05  
\*\* Base count minimum 30 Lowercase Alpha Level 0.1

Compare Each Sample

Letter comparisons use Pearson Chisq		LR PValues		Pearson PValues	
Age Group, I am working on my career	LR Chi-square p-value on pairs			Pearson Chi-square p-value on pairs	
A	1.0000	0.0552	0.0020	0.0001	0.0001
B	0.0552	1.0000	0.2183	0.0641	0.0261
C	0.0020	0.2183	1.0000	0.5781	0.3312
D	0.0001	0.0641	0.5781	1.0000	0.6540
E	0.0001	0.0641	0.5781	1.0000	0.6540
F	<.0001	0.0261	0.3312	0.6540	1.0000
G	<.0001	0.0020	0.1108	0.3083	0.6276

The Crosstab table summarizes the statement “I am working on my career” across age groups. The cells of the table contain the frequency (count) and share (percent) of those who agree or disagree with the statement for each age group. In addition, the Crosstab includes comparison letters. Each group is labeled with a letter in a column to the right of the group label. The comparison column uses the letters to interpret the outcome of the statistical test of independence between groups.

The Compare Each Sample outline provides  $p$ -values from the pairwise Pearson and Chi-square likelihood ratio chi-square tests. The  $p$ -values are reported in symmetric matrices labeled by the comparison letters.

For this example we make the following observations:

- The comparison column for the 25 - 29 group contains all letters b - g. Thus, the 25 - 29 group has significantly different response rates to the statement “I am working on my career” as compared to all other groups. Because the letter b is lowercase, the difference between the 25 - 29 group and the 30 - 34 group is significant at the 0.10 level. All other letters are uppercase indicating differences significant at the 0.05 level.
- The >54 group, denoted by letter G, is significantly different from the 30 - 34 group, denoted by B. The letter for the comparison is in the cell for group G because group G has a higher share of responders (71 versus 68) than group B.
- The single asterisks in the comparison cells are small sample size warnings. A single asterisk indicates that a group has more than 30 but fewer than 100 responses.

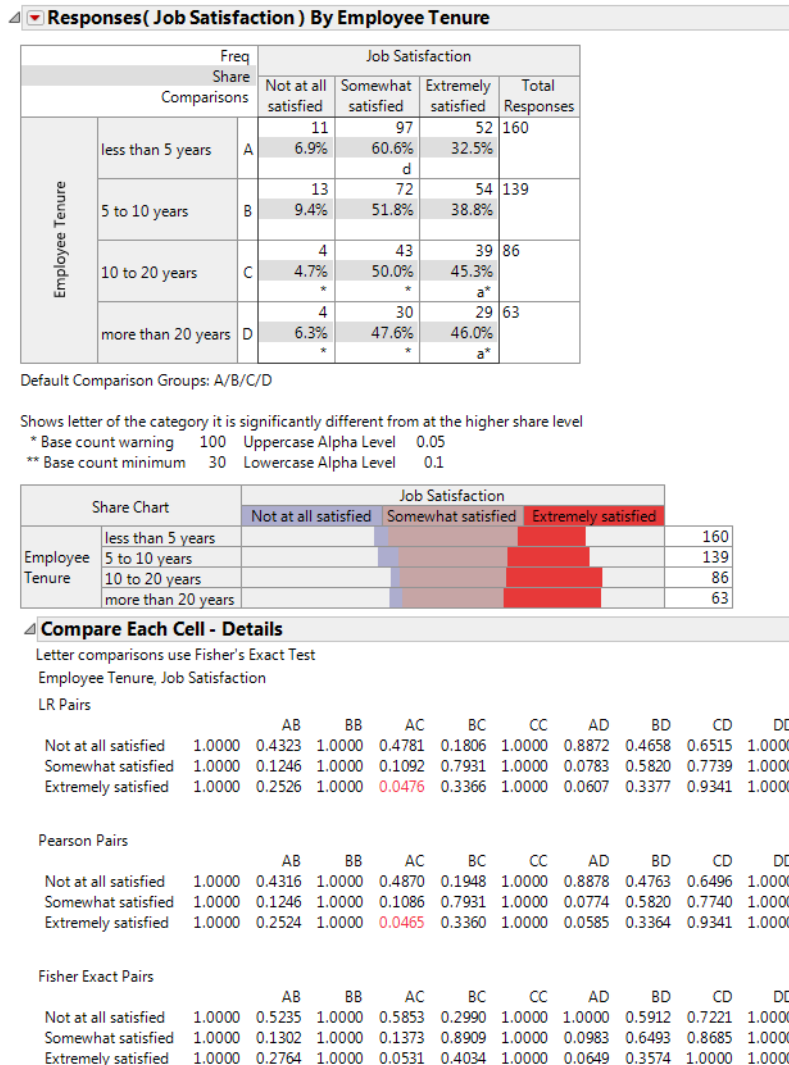
- A double asterisk, not observed in this example, would indicate a group size of fewer than 30.

## Example of Compare Each Cell with Comparison Letters

This example uses the Consumer Preferences.jmp sample data table, which contains survey data on people's attitudes and opinions, as well as questions concerning oral hygiene. You explore the distribution of the responses to the statement "I am working on my career" between employee tenure groups.

1. Select **Help > Sample Data Library** and open Consumer Preferences.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select Job Satisfaction and on the Simple tab click **Responses**.
4. Select Employee Tenure and click **X, Grouping Category**.
5. Click **OK**.
6. Click the Categorical red triangle menu and select **Compare Each Cell**.

Figure 3.12 Compare Each Cell



The  $p$ -values for pairwise Pearson Chi-square, likelihood ratio chi-square, and Fisher's exact tests for independence are provided in tables. The tables are labeled by comparison letters. The comparison letters are shown in the Crosstab table to the right of the group labels. Response rates that differ by groups are indicated with a comparison letter in the Crosstab table cells.

Employees with fewer than 5 years of tenure are somewhat satisfied at a greater rate than those with 20 years of tenure. This finding is noted by the letter d in the Somewhat satisfied cell in the first row of the Crosstab table. In addition, these same employees are Extremely satisfied at a lower rate than group D as noted by the letter a in the Extremely satisfied cell of

the last row of the Crosstab table. The letters are placed in the cell with the higher share of responses.

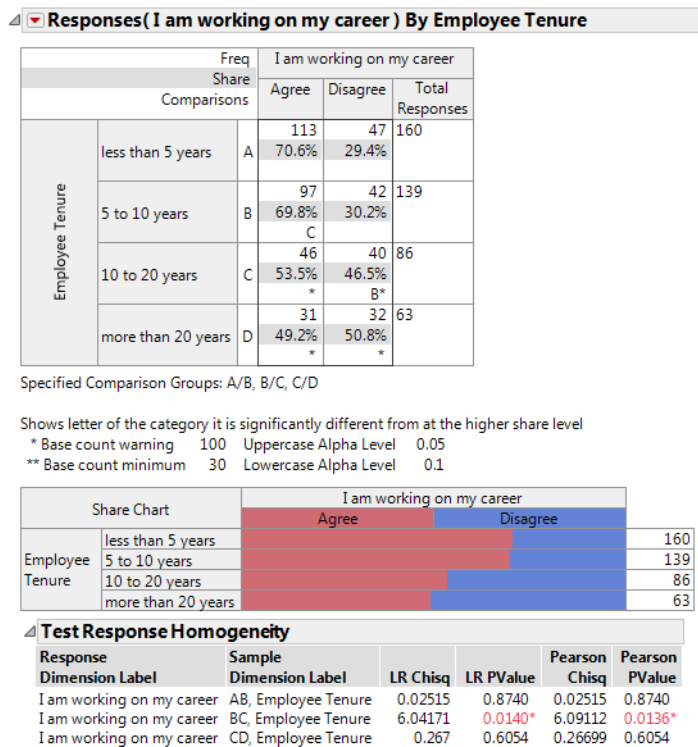
## Example of User-Specified Comparison with Comparison Letters

This example uses the Consumer Preferences.jmp sample data table, which contains survey data on people's attitudes and opinions, as well as questions concerning oral hygiene. You define specific comparison groups across which to compare the responses to the statement "I am working on my career".

1. Select **Help > Sample Data Library** and open Consumer Preferences.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select I am working on my career and on the Simple tab click **Responses**.
4. Select Employee Tenure and click **X, Grouping Category**.
5. Click **OK**.
6. Click the Responses red triangle menu and select **Show Letters**.
7. Click the Responses red triangle menu and select **Specify Comparison Groups**.
8. Enter A/B, B/C, C/D.
9. Click the Categorical red triangle menu and select **Test Response Homogeneity**.



**Figure 3.13** Specify Comparison Example



The test of response homogeneity compares Group A to B, Group B to C and Group C to D. Group B (5 to 10 years) agrees with the statement “I am working on my career” more often than those in group C (10 to 20 years). This difference in agreement rates is statistically significant as the Pearson *p*-value is 0.0136.

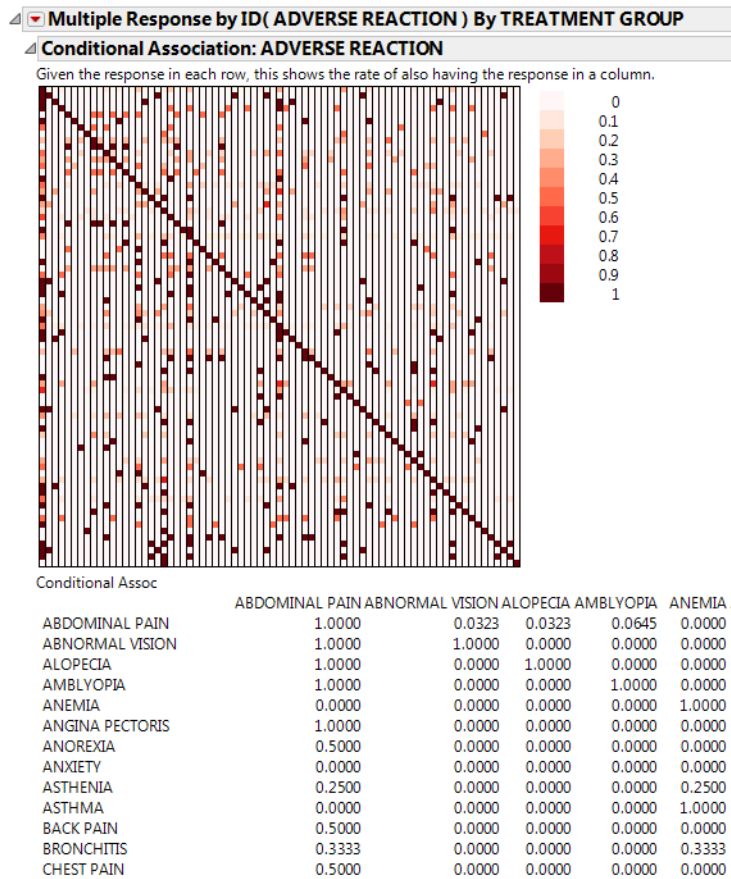
## Example of Conditional Association and Relative Risk

This example uses the AdverseR.jmp sample data table, which contains adverse reactions from a clinical trial. Use this data to explore the conditional association of adverse events and then the relative risk of the events in the treatment group as compared to the control.

1. Select **Help > Sample Data Library** and open AdverseR.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select ADVERSE REACTION and on the Multiple tab click **Multiple Response by ID**.
4. Select TREATMENT GROUP and click **X, Grouping Category**.
5. Select PATIENT ID and click **ID**.
6. Select **Unique Occurrences within ID** and click **OK**.

7. Click the Categorical red triangle menu and select **Conditional Association**.

Figure 3.14 Conditional Association Report (Partial Report)

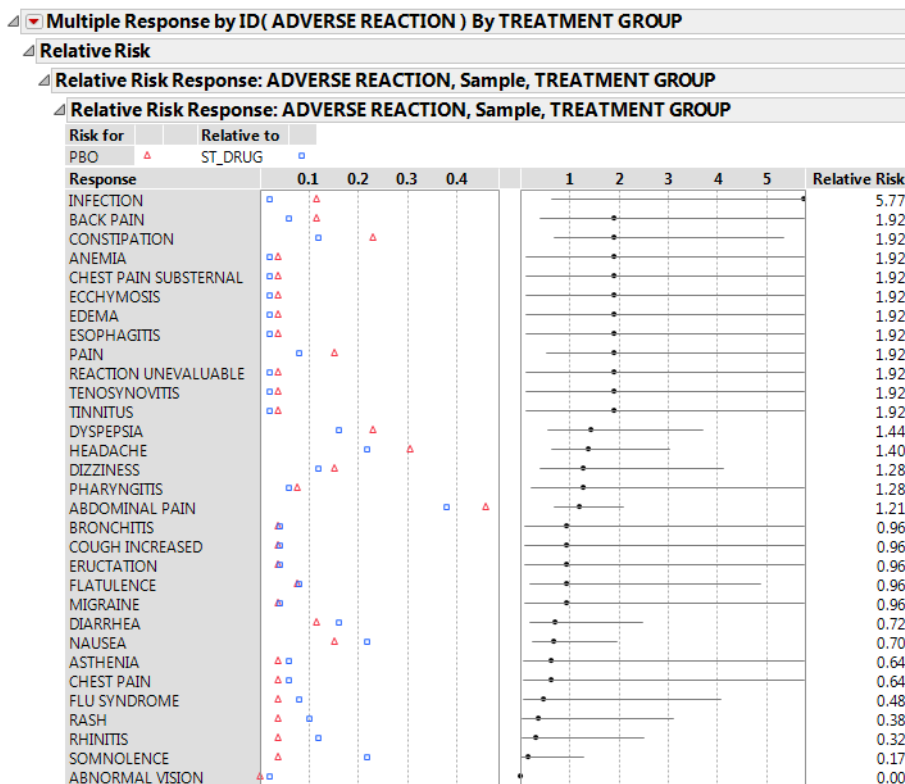


The conditional association matrix provides the conditional probability of one adverse reaction given the presence of another reaction. The probabilities are across all groups. The probability of abnormal vision given that a patient has abdominal pain is 0.0323.

**Tip:** Place your pointer on the heat map for conditional probabilities.

- 8. Click the Categorical red triangle menu and select **Relative Risk**.
- 9. Select PBO in the window and click **OK**.
- 10. Right-click the Relative Risk Report in the window and select **Sort by Column**.
- 11. Select Relative Risk and click **OK**.

**Figure 3.15** Relative Risk Report (Partial Report)



The Relative Risk option computes relative risks for different responses as the ratio of the risk for each level of the grouping variable. The default relative risk report lists the response name, the risk (rate) for each level of the grouping variable, a plot of the relative risk with 95% confidence intervals, and the relative risk estimate. Here you can compare the relative risk of the adverse reactions by treatment group. The relative risk of an infection is 5.7 times greater for PBO relative to ST\_DRUG. However, the confidence interval is very wide and includes a relative risk of 1.0. A relative risk of 1.0 occurs when the risk is equal for each level of the grouping variable.

Right-click and select **Columns > Lower 95%** and **Columns > Upper 95%** to add 95% confidence intervals on the relative risk estimates to the report table.

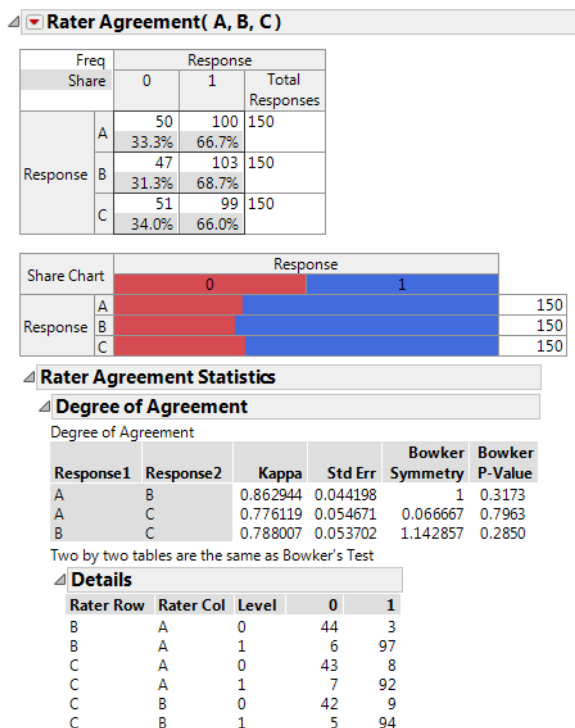
## Example of Rater Agreement

This example uses the Attribute Gauge.jmp sample data, which has the ratings (0/1) from three operators rating 50 parts 3 times.

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

2. Select **Analyze > Consumer Research > Categorical**.
3. Select A, B, and C.
4. On the Related tab, click **Rater Agreement**.
5. Click **OK**.

**Figure 3.16** Rater Agreement Report



The rater agreement is strong as shown by the Kappa statistics. The Kappa statistic can take on a value between 0 (no agreement) to 1.0 (perfect agreement). The details section provides 2x2 tables for each pair of raters. The Bowker test of symmetry tests the null hypothesis that cell proportions are symmetric for all pairs of cells ( $p_{ij} = p_{ji}$  for all  $i, j$ ). Here  $p$ -values for the Bowker test are all greater than 0.05, which supports symmetry between raters.









## Example of Repeated Measures

This example uses the Presidential Elections.jmp sample data table, which contains United States presidential election results for each state from 1980 through 2012. Use this data to explore Repeated Measures where we consider the election results the repeated measures.

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

2. Select **Analyze > Consumer Research > Categorical**.
3. Select 1980 Winner through 2012 Winner.
4. On the Related tab, click **Repeated Measures**.
5. Click **OK**.
6. Near the bottom of the report window, click on the gray Transition Report disclosure icon to open the Transition Report.

**Figure 3.17** Repeated Measures Transition Report

Transition Report									
All From	to	Transition Counts			Transition Rates				
		Democrat	Republican		Democrat	Republican		1984 Winner	
All 1980 Winner	1984 Winner	Democrat	1	5	Democrat	1.0000	0.1020		
		Republican	0	44	Republican	0.0000	0.8980		
All 1984 Winner	1988 Winner	Democrat	1	0	Democrat	0.1000	0.0000		
		Republican	9	40	Republican	0.9000	1.0000		
All 1988 Winner	1992 Winner	Democrat	10	0	Democrat	0.3125	0.0000		
		Republican	22	18	Republican	0.6875	1.0000		
All 1992 Winner	1996 Winner	Democrat	29	3	Democrat	0.9355	0.1579		
		Republican	2	16	Republican	0.0645	0.8421		
All 1996 Winner	2000 Winner	Democrat	20	11	Democrat	1.0000	0.3667		
		Republican	0	19	Republican	0.0000	0.6333		
All 2000 Winner	2004 Winner	Democrat	18	2	Democrat	0.9474	0.0645		
		Republican	1	29	Republican	0.0526	0.9355		
All 2004 Winner	2008 Winner	Democrat	19	0	Democrat	0.6786	0.0000		
		Republican	9	22	Republican	0.3214	1.0000		
All 2008 Winner	2012 Winner	Democrat	26	2	Democrat	1.0000	0.0833		
		Republican	0	22	Republican	0.0000	0.9167		

Between 1980 and 1984, there were 5 Democratic states that transitioned to Republican states. In 1980, they voted democratic but voted Republican in 1984. Between 2008 and 2012, there were two Democratic states that transitioned to Republican. All other states voted the same way in both the 2008 and 2012 elections.

## Examples of the Multiple Response Tab

The following examples use sample data tables that contain the same information organized in different data table layouts. The data come from testing a fabrication line on three different occasions under two different conditions. Each set of operating conditions (or batch) yielded 50 units for inspection. Inspectors recorded seven types of defects. Each unit could have zero, one, or more than one defect. A unit could have more than one defect of the same kind.

Multiple Response

The Failure3MultipleField.jmp sample data table has a row for each unit and multiple columns for defects, where defects are entered one per column. In this example, there are three columns for defects. Thus, any one unit had at most three defects.

1.

Select **Help > Sample Data Library** and open **Quality Control/Failure3MultipleField.jmp**.
2.

Select **Analyze > Consumer Research > Categorical**.
3.

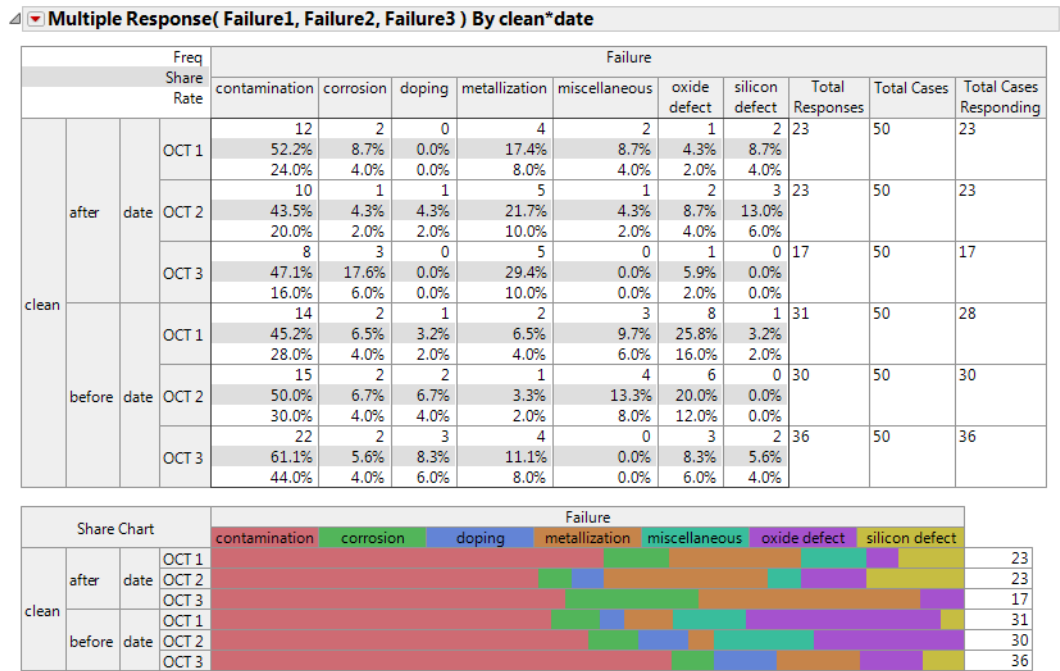
Select **Failure1**, **Failure2**, and **Failure3**.
4.

On the **Multiple** tab, click **Multiple Response**.
5.

Select **clean** and **date** and click **X, Grouping Category**.
6.

Click **OK**.

Figure 3.18 Multiple Response Report

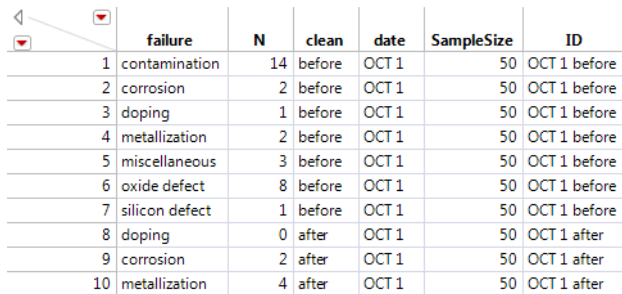


The Crosstab table has a row for each batch and a column for each defect type. The frequency, share, and rate of each defect within each batch are shown in the table cells. For example, for the batch after cleaning on OCT 1, there were 12 contamination defects representing 12/23 or 52.2% of defects for that batch. The 12 contamination defects were from 50 units. Therefore, the rate per unit was 24%.

## Multiple Response by ID

The Failure3ID.jmp sample data table has a row for each defect type within each batch, a column for the number of each defect type, and an ID column for each batch.

**Figure 3.19** Failure3ID Data Table (Partial Table)



	failure	N	clean	date	SampleSize	ID
1	contamination	14	before	OCT 1	50	OCT 1 before
2	corrosion	2	before	OCT 1	50	OCT 1 before
3	doping	1	before	OCT 1	50	OCT 1 before
4	metallization	2	before	OCT 1	50	OCT 1 before
5	miscellaneous	3	before	OCT 1	50	OCT 1 before
6	oxide defect	8	before	OCT 1	50	OCT 1 before
7	silicon defect	1	before	OCT 1	50	OCT 1 before
8	doping	0	after	OCT 1	50	OCT 1 after
9	corrosion	2	after	OCT 1	50	OCT 1 after
10	metallization	4	after	OCT 1	50	OCT 1 after

1. Select **Help > Sample Data Library** and open Quality Control/Failure3ID.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select failure and click Multiple Response by ID on the Multiple tab.
4. Select clean and date and click **X, Grouping Category**.
5. Select SampleSize and click **Sample Size**.
6. Select N and click **Freq**.
7. Select ID and click **ID**.
8. Click **OK**.

The resulting report is the same as the report shown in Figure 3.18.

## Multiple Delimited

The Failures3Delimited.jmp sample data table has a row for each unit with a single column in which the defects are recorded, delimited by a comma. Note in the partial data table, shown in Figure 3.20, that some units did not have any observed defects, so the failureS column is empty.

**Figure 3.20** Failure3Delimited.jmp Data Table (Partial Table)

	failureS	clean	date	ID	ID Label
1		before	OCT 1	1	OCT 1 before
2	oxide defect	before	OCT 1	1	OCT 1 before
3	contamination,oxide defect	before	OCT 1	1	OCT 1 before
4		before	OCT 1	1	OCT 1 before
5	contamination	before	OCT 1	1	OCT 1 before
6	oxide defect	before	OCT 1	1	OCT 1 before
7	contamination	before	OCT 1	1	OCT 1 before
8		before	OCT 1	1	OCT 1 before
9		before	OCT 1	1	OCT 1 before
10	metallization,contamination	before	OCT 1	1	OCT 1 before

1. Select **Help > Sample Data Library** and open Quality Control/ Failures3Delimited.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select failureS and click **Multiple Delimited** on the Multiple tab.
4. Select clean and date and click **X, Grouping Category**.
5. Click **OK**.

When you click **OK**, you get the report in Figure 3.18

**Note:** If more than one delimited column is specified, separate analyses are produced for each column.

## Indicator Group

The Failures3Indicators.jmp sample data table has a row for each unit and indicator columns for each defect type. The data entry in each defect columns is a 0 if that defect was not observed and a 1 if the defect was observed for the unit.

**Figure 3.21** Faliure3Indicators.jmp Data Table (Partial Table)

	clean	date	ID	ID Label	contamination	corrosion	doping	metallization	miscellaneous	oxide defect	silicon defect
1	before	OCT 1	1	OCT 1 before	0	0	0	0	0	0	0
2	before	OCT 1	1	OCT 1 before	0	0	0	0	0	1	0
3	before	OCT 1	1	OCT 1 before	1	0	0	0	0	1	0
4	before	OCT 1	1	OCT 1 before	0	0	0	0	0	0	0
5	before	OCT 1	1	OCT 1 before	1	0	0	0	0	0	0
6	before	OCT 1	1	OCT 1 before	0	0	0	0	0	1	0

1. Select **Help > Sample Data Library** and open Quality Control/Failures3Indicators.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select contamination, corrosion, doping, metallization, miscellaneous, oxide defect, and silicon defect and click **Indicator Group** on the Multiple tab.
4. Select clean and date and click **X, Grouping Category**.



5. Click **OK**.

When you click **OK**, you get the report in Figure 3.18.

## Response Frequencies

The Failure3Freq.jmp sample data table has a row for each batch, columns for each defect type, and a column for the batch size. The data entries in the defect columns are the frequency of occurrence of the defect in the batch.

**Figure 3.22** Failure3Freq.jmp Data Table

	clean	date	contamination	corrosion	doping	metallization	miscellaneous	oxide defect	silicon defect	SampleSize
1	after	OCT 1	12	2	0	4	2	1	2	50
2	after	OCT 2	10	1	1	5	1	2	3	50
3	after	OCT 3	8	3	0	5	0	1	0	50
4	before	OCT 1	14	2	1	2	3	8	1	50
5	before	OCT 2	15	2	2	1	4	6	0	50
6	before	OCT 3	22	2	3	4	0	3	2	50

1. Select **Help > Sample Data Library** and open Quality Control/Failure3Freq.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select the frequency variables (contamination, corrosion, doping, metallization, miscellaneous, oxide defect, silicon defect).
4. On the Multiple tab, click **Response Frequencies**.
5. Select clean and date and click **X, Grouping Category**.
6. Select Sample Size and click **Sample Size**.
7. Click **OK**.

**Figure 3.23** Defect Rate Output

Sample Size: SampleSize

☒
**Response Frequencies( contamination, corrosion, doping, metallization, miscellaneous, oxide defect, silicon defect ) By clean\*date**

			Freq Share Rate	Response							Total Responses	Total Cases	Total Cases Responding
				contamination	corrosion	doping	metallization	miscellaneous	oxide defect	silicon defect			
clean	after	date	OCT 1	12	2	0	4	2	1	2	23	50	50
				52.2%	8.7%	0.0%	17.4%	8.7%	4.3%	8.7%			
				24.0%	4.0%	0.0%	8.0%	4.0%	2.0%	4.0%			
		OCT 2	10	1	1	5	1	2	3	23	50	50	
			43.5%	4.3%	4.3%	21.7%	4.3%	8.7%	13.0%				
			20.0%	2.0%	2.0%	10.0%	2.0%	4.0%	6.0%				
	OCT 3	8	3	0	5	0	1	0	17	50	50		
		47.1%	17.6%	0.0%	29.4%	0.0%	5.9%	0.0%					
		16.0%	6.0%	0.0%	10.0%	0.0%	2.0%	0.0%					
	before	date	OCT 1	14	2	1	2	3	8	1	31	50	50
				45.2%	6.5%	3.2%	6.5%	9.7%	25.8%	3.2%			
				28.0%	4.0%	2.0%	4.0%	6.0%	16.0%	2.0%			
OCT 2		15	2	2	1	4	6	0	30	50	50		
		50.0%	6.7%	6.7%	3.3%	13.3%	20.0%	0.0%					
		30.0%	4.0%	4.0%	2.0%	8.0%	12.0%	0.0%					
OCT 3	22	2	3	4	0	3	2	36	50	50			
	61.1%	5.6%	8.3%	11.1%	0.0%	8.3%	5.6%						
	44.0%	4.0%	6.0%	8.0%	0.0%	6.0%	4.0%						

Share Chart			Response							
			contamination	corrosion	doping	metallization	miscellaneous	oxide defect	silicon defect	
clean	after	date	OCT 1							23
			OCT 2							23
			OCT 3							17
	before	date	OCT 1							31
			OCT 2							30
			OCT 3							36

The resulting output is the same as that in Figure 3.18 with the exception of the Total Cases Responding column in the Crosstab table. Here, the defect data was summarized. From the summarized table, there is no record of the number of units with zero defects. Thus, the Total Cases Responding is the full batch size of 50 for each batch.

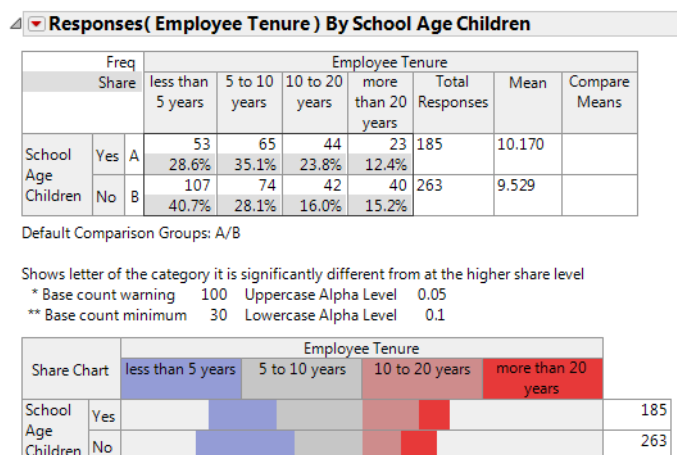
## Example of Mean Score with Comparison Letters

This example uses the Consumer Preferences.jmp sample data table to explore the relationship between employee tenure and having school age children. The Employee Tenure column is a numeric column with values 1, 2, 3, and 4. These values have been assigned Value Labels using the Value Labels column property. To evaluate a mean employee tenure using the mean score option in the categorical platform, you need to assign Value Scores to the column values. For more information about column properties, refer to The Column Info Window chapter in the *Using JMP* book.

1. Select **Help > Sample Data Library** and open Consumer Preferences.jmp.
2. In the data table, right-click the Employee Tenure column heading and select **Column Properties > Value Scores**.
3. Enter 1 for **Value** and 3 for **Score** and click **Add**.

4. Enter 2 for **Value** and 7.5 for **Score** and click **Add**.
5. Enter 3 for **Value** and 15 for **Score** and click **Add**.
6. Enter 4 for **Value** and 25 for **Score** and click **Add**.
7. Click **OK**.
8. Select **Analyze > Consumer Research > Categorical**.
9. Select **Employee Tenure** and click **Responses** on the Simple tab.
10. Select **School Age Children** and click **X, Grouping Category**.
11. Click **OK**.
12. Click the Categorical red triangle and select **Mean Score**.
13. Click the Categorical red triangle and select **Mean Score Comparison**.

**Figure 3.24** Categorical Report with Mean Scores



The mean employee tenure for those with school age children is 10.17 and 9.53 for those without school age children. The means are not statistically different as the Compare Means column in the Crosstab table is empty. If there was a difference, a letter would indicate the difference.

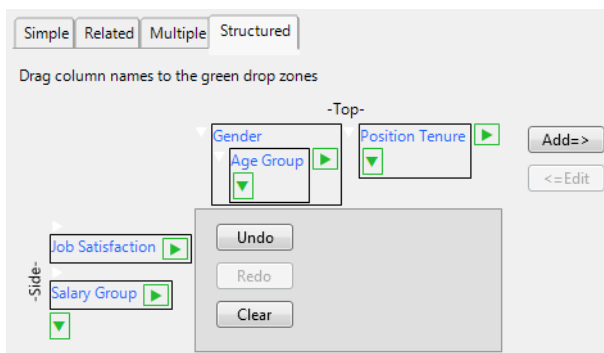
**Tip:** Be aware of how your data is recorded when using the mean score option. If your data is recorded as coded numeric data with value labels, the mean value calculations are based on the numeric data. If the numeric data does not have meaning, use **Value Scores** to assign meaningful data values to the response levels.

## Example of a Structured Report

This example uses the Consumer Preferences.jmp sample data table to compare job satisfaction and salary against gender by age group and position tenure to explore using the Structured tab.

1. Select **Help > Sample Data Library** and open Consumer Preferences.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select the Structured tab.
4. Drag Gender to the green drop zone at the **Top** of the table on the Structured tab.
5. Drag Age Group to the green drop zone just below Gender.
6. Drag Position Tenure to the green drop zone at the **Top** of the table next to Gender.
7. Drag Job Satisfaction to the green drop zone at the **Side** of the table.
8. Drag Salary Group to the green drop zone at the **Side** of the table under Job Satisfaction.

**Figure 3.25** Structured Tab Report Setup



9. Click **Add=>**.
10. Click **OK**.
11. Click the Categorical red triangle and select **Test Response Homogeneity**.

Figure 3.26 Structured Tab Report Example

Job Satisfaction + Salary Group By Gender*Age Group + Position Tenure																						
		Freq Share	Gender														Position Tenure					
			M							F												
			25-29	30-34	35-39	40-44	45-49	50-54	>54	25-29	30-34	35-39	40-44	45-49	50-54	>54	less than 5 years	5 to 10 years	10 to 20 years	more than 20 years		
Job Satisfaction	Not at all satisfied	3 5.5%	3 10.3%	1 3.3%	3 8.8%	3 7.9%	2 8.3%	2 4.1%	3 5.2%	1 2.6%	0 0.0%	5 27.8%	2 14.3%	1 5.3%	3 13.6%	15 7.2%	11 8.2%	6 6.8%	0 0.0%			
	Somewhat satisfied	32 58.2%	14 48.3%	15 50.0%	19 55.9%	19 50.0%	10 41.7%	26 53.1%	34 58.6%	30 76.9%	12 63.2%	6 33.3%	5 35.7%	11 57.9%	9 40.9%	122 58.7%	70 52.2%	40 45.5%	10 55.6%			
	Extremely satisfied	20 36.4%	12 41.4%	14 46.7%	12 35.3%	16 42.1%	12 50.0%	21 42.9%	8 36.2%	7 20.5%	7 36.8%	7 38.9%	7 50.0%	7 36.8%	10 45.5%	51 34.1%	52 39.6%	42 47.7%	8 44.4%			
	Total Responses	55	29	30	34	38	24	49	58	39	19	18	14	19	22	208	134	88	18			
	less than 40000	20 36.4%	6 20.7%	4 13.3%	8 23.5%	6 15.8%	5 20.8%	10 20.4%	35 60.3%	19 48.7%	6 31.6%	6 33.3%	2 14.3%	7 36.8%	4 18.2%	80 38.5%	40 29.9%	15 17.0%	3 16.7%			
Salary Group	40000 to 60000	16 29.1%	14 48.3%	11 36.7%	12 35.3%	14 36.8%	7 29.2%	13 26.5%	15 25.9%	10 25.6%	5 42.1%	5 33.3%	3 50.0%	3 31.6%	4 54.5%	32 32.2%	25 28.4%	19 42.0%	5 50.0%			
	60000 to 80000	8 14.5%	5 17.2%	7 23.3%	7 20.6%	8 21.1%	8 33.3%	10 20.4%	5 8.6%	5 12.8%	3 15.8%	3 16.7%	4 21.4%	5 21.1%	22 22.7%	15 15.4%	17 18.7%	10 21.6%	5 27.8%			
	80000 to 120000	9 16.4%	3 10.3%	6 20.0%	4 11.8%	7 18.4%	3 12.5%	7 14.3%	1 1.7%	2 5.1%	1 5.3%	1 11.1%	1 7.1%	1 5.3%	0 0.0%	19 9.1%	17 12.7%	10 11.4%	1 5.6%			
	greater than 120000	2 3.6%	1 3.4%	2 6.7%	3 8.8%	3 7.9%	1 4.2%	9 18.4%	2 3.4%	3 7.7%	1 5.3%	1 5.6%	1 7.1%	1 5.3%	4 4.5%	10 4.8%	14 10.4%	7 8.0%	0 0.0%			
	Total Responses	55	29	30	34	38	24	49	58	39	19	18	14	19	22	208	134	88	18			

Test Response Homogeneity					
Response Dimension Label	Sample Dimension Label	LR Chisq	LR PValue	Pearson Chisq	Pearson PValue
Job Satisfaction	Gender = M, Age Group	4.65556	0.9685	4.59049	0.9703
Job Satisfaction	Gender = F, Age Group	23.8406	0.0214*	24.9917	0.0149*
Salary Group	Gender = M, Age Group	22.0707	0.5750	23.8074	0.4727
Salary Group	Gender = F, Age Group	27.6209	0.2764	26.3992	0.3332
Job Satisfaction	Position Tenure	8.00461	0.2378	6.76025	0.3436
Salary Group	Position Tenure	25.8336	0.0113*	24.0752	0.0199*

The structured tab report contains the table that you specified in the structured tab. The tests for response homogeneity are for each combination of grouping variables. We see that, for males, there is no difference in job satisfaction across age groups (Pearson  $p$ -value = 0.9703). For women, there is a difference in job satisfaction across age groups (Pearson  $p$ -value = 0.0149). The middle aged women tend to be the least satisfied with their jobs. Share and frequency charts can be added to your report to visualize your results.

## Statistical Details for the Categorical Platform

### Rao-Scott Correction

The Rao-Scott correction is applied to the test of response homogeneity for multiple responses. See Lavassani et al.(2009). The test of response homogeneity for multiple responses is available only for the structured tab.

In the case of a multiple response, you can have overlapping samples, meaning a single participant can provide more than one response. The Pearson chi-square test is not appropriate for multiple responses, because the multiple responses violate the Pearson chi-square test assumption of independence. In addition, expected values calculated using the marginal totals are influenced by the multiple responses because the totals are larger than if multiple responses were not allowed.

Rao-Scott chi-square statistic is defined as follows:

$$\chi_C^2 = \frac{\chi^2}{\bar{\delta}}$$

where

$\chi^2$  is the standard Pearson Chi-squared statistic and  $\bar{\delta}$  is the correction factor defined by

$$\bar{\delta} = 1 - \frac{m_{++}}{n_+ C}$$

where

$m_{++}$  is the total count of the multiple responses

$n_+$  is the total number of participants and

$C$  is the number of response levels (number of columns in the Crosstab table).

The degrees of freedom is  $(R-1)C$  or the number of rows minus 1 times the number of columns.

# Chapter 4

## Choice Models

### Fit Models for Choice Experiments

Use the Choice platform to analyze the results of choice experiments conducted in the course of market research. Choice experiments are used to help discover which product or service attributes your potential customers prefer. You can use this information to design products or services that have the attributes that your customers most desire.

The Choice platform enables you to do the following:

- Use information about subject (customer) traits as well as product attributes.
- Analyze choice experiments where respondents were allowed to select “none of these”.
- Integrate data from one, two, or three sources.
- Use the integrated profiler to understand, visualize, and optimize the response (utility) surface.
- Obtain subject-level scores for segmenting or clustering your data.
- **JMP PRO** Estimate subject-specific coefficients using a Bayesian approach.
- Use bias-corrected maximum likelihood estimators (Firth 1993).

**Figure 4.1** Choice Platform Utility Profiler



**Contents**

Overview of the Choice Modeling Platform .....	73
Examples of the Choice Platform .....	75
One Table Format with No Choice .....	75
Multiple Table Format .....	78
Launch the Choice Platform .....	84
Launch Window for One Table, Stacked .....	85
Launch Window for Multiple Tables, Cross-Referenced .....	86
Choice Model Report .....	91
Effect Summary .....	91
Parameter Estimates .....	92
Likelihood Ratio Tests .....	93
Bayesian Parameter Estimates .....	93
Choice Platform Options .....	95
Willingness to Pay .....	98
Additional Examples .....	101
Example of Making Design Decisions .....	101
Example of Segmentation .....	109
Example of Logistic Regression Using the Choice Platform .....	113
Example of Logistic Regression for Matched Case-Control Studies .....	116
Example of Transforming Data to Two Analysis Tables .....	118
Example of Transforming Data to One Analysis Table .....	123
Statistical Details for the Choice Platform .....	125
Special Data Table Rules .....	125
Utility and Probabilities .....	126
Gradients .....	127



---

# Overview of the Choice Modeling Platform

Choice modeling, pioneered by McFadden (1974), is a powerful analytic method used to estimate the probability of individuals making a particular choice from presented alternatives. Choice modeling is also called conjoint choice modeling, discrete choice analysis, and conditional logistic regression.

A choice experiment studies customer preferences for a set of product or process (in the case of a service) attributes. Respondents are presented sets of product attributes, called *profiles*. Each respondent is shown a small set of profiles, called a *choice set*, and asked to select the preference that he or she most prefers. Each respondent is usually presented with several choice sets. Use the Choice platform to analyze the results of a choice experiment.

---

**Note:** You can design your choice experiment using the Choice Design platform. See the Discrete Choice Design chapter in *Design of Experiments Guide*.

---

Because customers vary in how they value attributes, many market researchers view market segmentation as an important step in analyzing choice experiments. Otherwise, you risk designing a product or process that pleases the “average” customer, who does not actually exist, and ignoring the preferences of market segments that *do* exist.

For background on choice modeling, see Louviere et al. (2015), Train (2009), and Rossi et al. (2005).

## The Choice Platform

The Choice Modeling platform uses a form of conditional logistic regression to estimate the probability that a configuration is preferred. Unlike simple logistic regression, choice modeling uses a linear model to model choices based on response attributes and not solely upon subject characteristics. In choice modeling, a respondent might choose between two cars that are described by a combination of ten attributes, such as price, passenger load, number of cup holders, color, GPS device, gas mileage, anti-theft system, removable-seats, number of safety features, and insurance cost.

The Choice platform allows respondents to *not* make a choice from among a set of profiles. The *no choice* option is treated as a product with a single attribute (“Select none of these”) that respondents are allowed to select. The parameter estimate for the No Choice attribute can then be interpreted in many ways, depending on the assumptions of the model. The Choice platform also enables you to obtain subject-level information, which can be useful in segmenting preference patterns.

You can obtain bias-corrected maximum likelihood estimators as described by Firth (1993). This method has been shown to produce better estimates and tests than MLEs without bias correction. In addition, bias-corrected MLEs improve separation problems that tend to occur

in logistic-type models. Refer to Heinze and Schemper (2002) for a discussion of the separation problem in logistic regression.

---

**Note:** The Choice platform is not appropriate to use for fitting models that involve ranking, scoring, or nested hierarchical choices. You can use PROC MDC in SAS/ETS for these analyses.

---

### Choice Designs in Developing Products and Services

Although customer satisfaction surveys can disclose what is wrong with a product or service, they fail to identify consumer preferences with regard to specific product attributes. When engineers design a product, they routinely make hundreds or thousands of small design decisions. If customer testing is feasible and research participants (subjects) are available, you can use choice experiments to guide some design decisions.

Decreases in survey deployment, modeling, and prototyping costs facilitate the customer evaluation of many attributes and alternatives as a product is designed. Choice modeling can be used in Six Sigma programs to improve consumer products, or, more generally, to make the products that people want. Choice experiments obtain data on customer preferences, and choice modeling analysis reveals such preferences.

### Segmentation

Market researchers sometimes want to analyze the preference structure for each subject separately in order to see whether there are groups of subjects that behave differently. However, there are usually not enough data to do this with ordinary estimates. If there are sufficient data, you can specify the subject identifier as a “By groups” in the Response Data or you could introduce a subject identifier as a subject-side model term. This approach, however, is costly if the number of subjects is large.

If there are not sufficient data to specify “By groups,” you can segment in JMP by clustering subjects using the Save Gradients by Subject option. The option creates a new data table containing the average Hessian-scaled gradient on each parameter for each subject. For an example, see “[Example of Segmentation](#)” on page 109. For details about the gradient values, see “[Gradients](#)” on page 127.

**JMP PRO** In JMP Pro, you can request that the Choice platform use a Hierarchical Bayes approach in order to facilitate market segmentation. Bayesian modeling provides subject-specific estimates of model parameters (also called part-worths) in choice models that can be analyzed through hierarchical clustering or some other type of cluster analysis to reveal market segments.

---

## Examples of the Choice Platform

In a study of pizza preferences, each respondent is presented with four choice sets, each containing two profiles. The Choice platform can analyze data that is in a one table format or a multiple data format. In the multiple table format, information about responses, choice sets, and subjects is saved in different data tables. In the one table format, that information is contained in a single data table.

- [“One Table Format with No Choice”](#) on page 75 shows how to analyze a subset of the available data in a one table format.
- [“Multiple Table Format”](#) on page 78 shows how to bring together information from different tables into one Choice analysis

### One Table Format with No Choice

In this example, some respondents do not express a preference for either profile. The respondent makes “no choice”. When a respondent does not express a preference, the respondent’s choice indicator is entered as missing.

1. Select **Help > Sample Data Library** and open **Pizza Combined No Choice.jmp**.  
Choice sets are defined by the combination of Subject and Trial. Notice that there are missing values in the Indicator column for some choice sets.
2. Select **Analyze > Consumer Research > Choice**.  
The One Table, Stacked data format is the default.
3. Click **Select Data Table**.
4. Select **Pizza Combined No Choice** and click **OK**.
5. Complete the launch window as follows:
  - Select Indicator and click **Response Indicator**.
  - Select Subject and click **Subject ID**.
  - Select Trial and click **Choice Set ID**.
  - Select Crust, Cheese, and Topping and click **Add** in the Construct Profile Effects panel.
  - Select Gender and click **Add** in the Construct Subject Effects (Optional) panel.

**Figure 4.2** Completed Launch Window

Data Format: One Table, Stacked

Select Data Table: Pizza Combined No Choice

Select Columns:

- Gender
- Subject
- Trial
- Profile Name
- Indicator
- Crust
- Cheese
- Topping

Pick Role Variables:

Response Indicator: Indicator

Subject ID: Subject

Choice Set ID: Trial

Grouping: optional

Construct Profile Effects:

Add

Cross

Nest

Macros

Degree: 2

Transform

Construct Subject Effects (Optional):

Add

Cross

Nest

Macros

Degree: 2

Transform

Run Model

Help

Remove

☒ Firth Bias-Adjusted Estimates

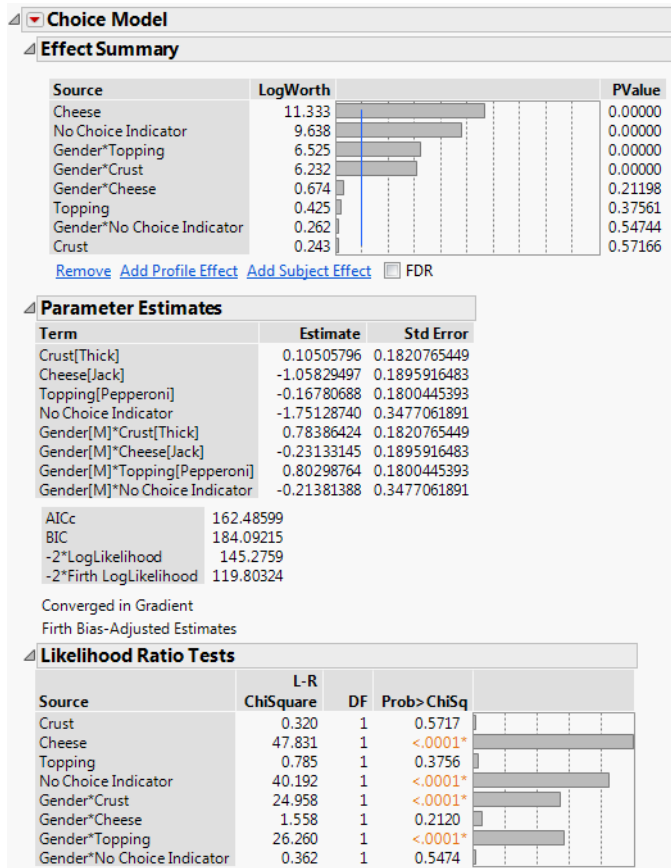
☐ Hierarchical Bayes

Number of Bayesian Iterations: 5000

☒ Respondent is allowed to select "None" or "No Choice"

6. Check the box next to **Respondent is allowed to select "None" or "No Choice"**.
7. Click **Run Model**.

**Figure 4.3** Report Showing No Choice as an Effect



The Effect Summary report shows the effects in order of significance. Cheese is the most significant effect, followed by the No Choice Indicator, which is treated as a model effect. The subject effect interactions Gender\*Topping and Gender\*Crust are also significant, indicating that preferences for Topping and Crust depend on Gender market segments.

To get some insight on the nature of the No Choice responses, select and view those choice sets that resulted in No Choice.

- In the data table, right-click in a cell in the Indicator column where the response is missing and select **Select Matching Cells**.
- In the Rows panel, right-click **Selected** and select **Data View**.

**Figure 4.4** Choice Sets with No Choice Responses

		Gender	Subject	Trial	Profile Name	Indicator	Crust	Cheese	Topping
	1	F	2	4	TrimPepperjack	• Thin	Jack	Pepperoni	
	2	F	2	4	TrimOri	• Thin	Mozzarella	Pepperoni	
	3	M	7	2	Trimella	• Thin	Mozzarella	None	
	4	M	7	2	TrimJack	• Thin	Jack	None	
	5	M	7	3	Trimella	• Thin	Mozzarella	None	
	6	M	7	3	TrimJack	• Thin	Jack	None	
	7	F	8	2	ThickElla	• Thick	Mozzarella	None	
	8	F	8	2	ThickJack	• Thick	Jack	None	
	9	M	11	4	ThickOri	• Thick	Mozzarella	Pepperoni	
	10	M	11	4	ThickJackoni	• Thick	Jack	Pepperoni	
	11	F	14	2	TrimOri	• Thin	Mozzarella	Pepperoni	
	12	F	14	2	TrimPepperjack	• Thin	Jack	Pepperoni	
	13	F	18	3	ThickJack	• Thick	Jack	None	
	14	F	18	3	ThickElla	• Thick	Mozzarella	None	
	15	F	24	3	ThickJack	• Thick	Jack	None	
	16	F	24	3	TrimOri	• Thin	Mozzarella	Pepperoni	
	17	M	29	1	TrimPepperjack	• Thin	Jack	Pepperoni	
	18	M	29	1	ThickJackoni	• Thick	Jack	Pepperoni	

In the table in Figure 4.4, consider the profiles in the first seven choice sets, which are defined by the Subject and Trial combinations in rows 1 to 14. The only difference within each choice set is the Cheese. There is an indication that some respondents might not be able to detect the difference in cheeses. However, the analysis takes the No Choice Indicator into account and concludes that, despite this behavior, Cheese is significant.

To see how to further analyze data of this type, see [“Find Optimal Profiles”](#) on page 82.

## Multiple Table Format

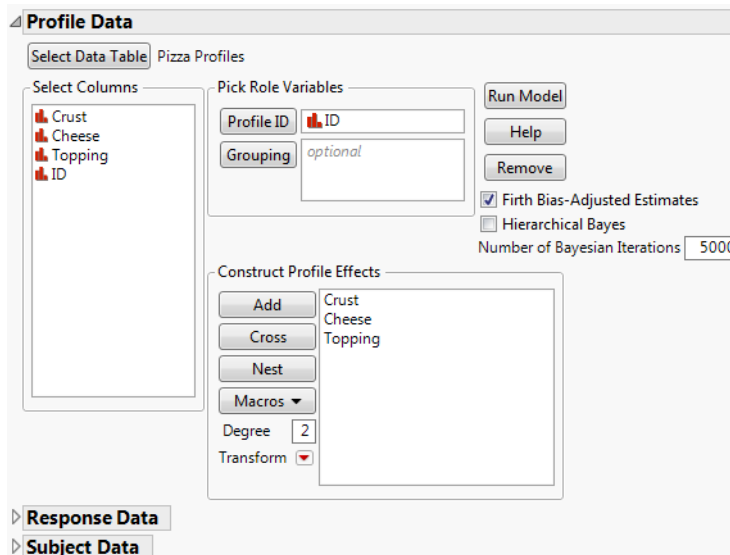
In this example, you examine pizza choices where three attributes, with two levels each, are presented to the respondents. The study was designed such that the respondents had to make a choice. The analysis uses three data tables: *Pizza Profiles.jmp*, *Pizza Responses.jmp*, and *Pizza Subjects.jmp*. Although you can always arrange your data into a single table, a multi-table approach can be more convenient than a one table analysis when you have additional profile and subject variables that you want to include in your analysis.

1. Select **Help > Sample Data Library** and open *Pizza Profiles.jmp*, *Pizza Responses.jmp*, and *Pizza Subjects.jmp*.
  - The profile data table, *Pizza Profiles.jmp*, lists all the pizza choice combinations that you want to present to the subjects. Each choice combination is given an ID.
  - The responses data table, *Pizza Responses.jmp*, contains the design and results. For the experiment, each subject is given four choice sets, where each choice set consists of two choice profiles (Choice1 and Choice2). The subject selects a preference (Choice) for each choice set. For information about how to construct a choice design, see the Discrete Choice Designs chapter in the *Design of Experiments Guide*. Notice that each value in the

Choice column is an ID value in the Profile data table that contains the attribute information.

- The subjects data table, Pizza Subjects.jmp, includes a Subject ID column and a single characteristic of the subject, Gender. Each value of Subject in the Pizza Subjects.jmp data table corresponds to values in the Subject column in the Pizza Responses.jmp data table.
2. Select **Analyze > Consumer Research > Choice** to open the launch window.
- 
- Note:** This can be done from any of the three open data tables.
- 
3. From the Data Format menu, select **Multiple Tables, Cross-Referenced**.  
There are three separate sections, one for each of the data sources.
  4. Click **Select Data Table** under Profile Data.  
A Profile Data Table window appears, which prompts you to specify the data table for the profile data.
  5. Select Pizza Profiles.jmp and click **OK**.
  6. Select ID and click **Profile ID**.
  7. Select Crust, Cheese, and Topping and click **Add**.

**Figure 4.5** Profile Data

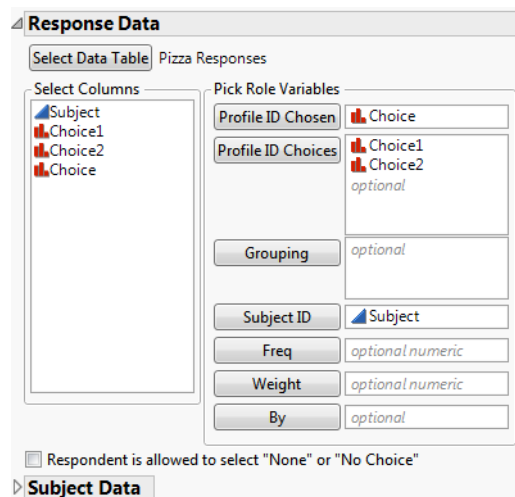


8. Click the disclosure icon next to Response Data to open the outline and click **Select Data Table**.
9. Select Pizza Responses.jmp and click **OK**.

10. Do the following:

- Select Choice and click **Profile ID Chosen**.
- Select Choice1 and Choice2 and click **Profile ID Choices**.
- Select Subject and select **Subject ID**.

**Figure 4.6** Response Data Window



Choice1 and Choice2 are the profiles presented to a subject in each of four choice sets. The Choice column contains the chosen preference between Choice1 and Choice2.

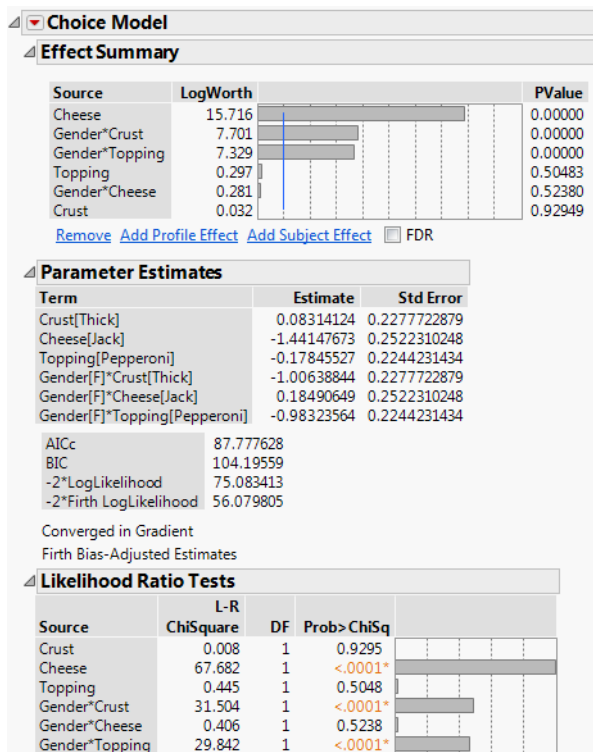
11. Click the disclosure icon next to Subject Data to open the outline and click **Select Data Table**.
12. Select Pizza Subjects.jmp and click **OK**.
13. Select Subject and click **Subject ID**.
14. Select Gender and click **Add**.



Figure 4.7 Subject Data Window

15. Click **Run Model**.

Figure 4.8 Choice Model Results



Six effects are entered into the model. The effects Crust, Cheese, and Topping are product attributes. The interaction effects, Gender\*Crust, Gender\*Cheese, and Gender\*Topping are

subject-effect interactions with the attributes. These interaction effects enable you to construct products that meet market-segment preferences.

**Note:** For Choice models, subject effects cannot be entered as main effects. They appear only as interaction terms.

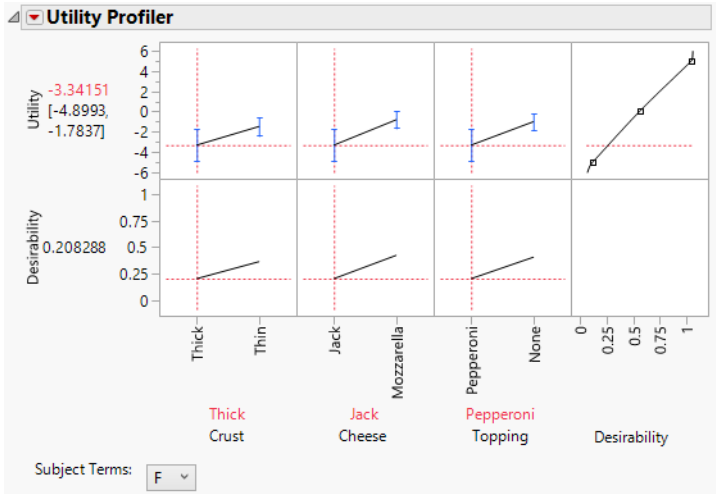
The Effect Summary and Likelihood Ratio Tests reports show strong interactions between Gender and Crust and between Gender and Topping. Notice that the main effects of Crust and Topping are not significant. If you had not included subject-level effects, you might have overlooked important information relative to market segmentation.

Find Optimal Profiles

Next, you use the Utility Profiler to explore your results and find optimal settings for the attributes.

- 1. Click the Choice Model red triangle and select **Utility Profiler**.  
The Subject Terms menu beneath the profiler indicates that it is showing results for females.
- 2. Click the red triangle next to Utility Profiler and select **Optimization and Desirability > Desirability Functions**.

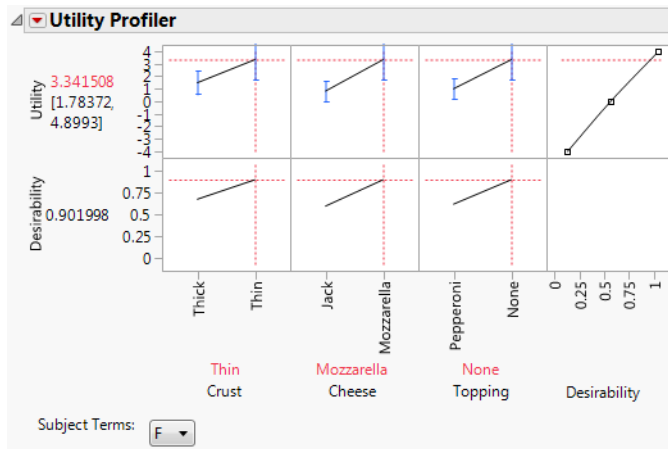
Figure 4.9 Utility Profiler with Desirability Function



A desirability function that maximizes utility is added to the profiler. See the Profiler chapter in the *Profilers* book.

- 3. Click the red triangle next to Utility Profiler and select **Optimization and Desirability > Maximize Desirability**.

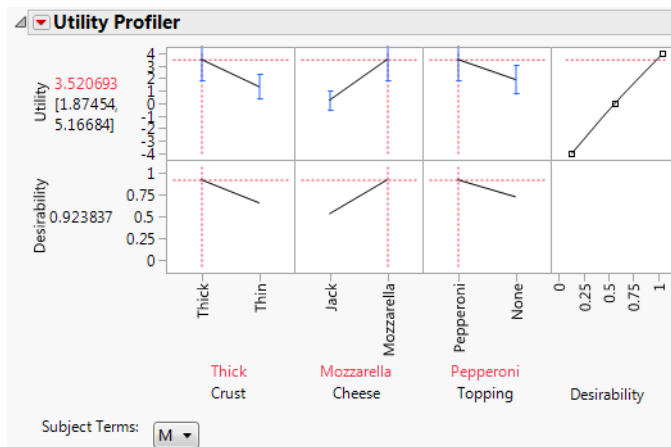
**Figure 4.10** Utility Profiler with Optimal Settings for Females



The optimal settings for females are a thin crust, Mozzarella cheese, and no topping.

- From the Subject Terms menu, select M.

**Figure 4.11** Utility Profiler with Male Level Factor Setting



The optimal settings for males are a thick crust, Mozzarella cheese, and a Pepperoni topping.

In this example, understanding the preferences of gender-defined market segments enables you to provide two pizza choices that appeal to two segments of customers.

---

## Launch the Choice Platform

Launch the Choice platform by selecting **Analyze > Consumer Research > Choice**.

Your data for the Choice platform can be combined in a single data table or it can reside in two or three separate data tables. When the Choice window opens, the first menu item asks you to specify the Data Format.

### One Table, Stacked

For this format, the data are combined into a single data table. There is a row for every profile presented to a subject and an indicator of whether that profile was selected.

For an example of data in the one-table format, see [“One Table Format with No Choice”](#) on page 75. For details, see [“Launch Window for One Table, Stacked”](#) on page 85.

### Multiple Tables, Cross-referenced

For this format, the data are stored in two or three separate tables: a Profile Data and Response Data table are required and a Subject Data table is optional. The Choice Launch Window contains three sections, each corresponding to a different data table. You can expand or collapse each section of the launch window, as needed.

For an example of data in the multiple-tables format, see [“Multiple Table Format”](#) on page 78. For details, see [“Launch Window for Multiple Tables, Cross-Referenced”](#) on page 86.

## Launch Window for One Table, Stacked

**Figure 4.12** Launch Window for One Table, Stacked Data Format

**Select Data Table** Select or open the data table that contains the combined data. Select Other to open a file that is not already open.

**Response Indicator** A column that contains values that indicate the preferred choice. A 1 indicates the preferred profile and a 0 indicates the other profiles. If respondents are given an option to select no preference, enter missing values for choice sets where no preference is indicated. See [“Respondent is allowed to select “None” or “No Choice””](#) on page 86.

**Subject ID** An identifier for the study participant.

**Choice Set ID** An identifier for the choice set presented to the subject for a given preference determination.

**Grouping** A column which, when used with the Choice Set ID column, uniquely designates each choice set. For example, if a choice set has Choice Set ID = 1 for Survey = A, and another choice set has Choice Set ID = 1 for Survey = B, then Survey should be used as a Grouping column.

**Construct Profile Effects** Add effects constructed from the attributes in the profiles.

For information about the Construct Profile Effects panel, see the Construct Model Effects section in the Model Specification chapter of the *Fitting Linear Models* book.

**Construct Subject Effects (Optional)** Add effects constructed from subject-related factors.

For information about the Construct Subject Effects panel, see the Construct Model Effects section in the Model Specification chapter of the *Fitting Linear Models* book.

**Firth Bias-adjusted Estimates** Computes bias-corrected MLEs that produce better estimates and tests than MLEs without bias correction. These estimates also improve separation problems that tend to occur in logistic-type models. Refer to Heinze and Schemper (2002) for a discussion of the separation problem in logistic regression.

**JMP PRO Hierarchical Bayes** Uses a Bayesian approach to estimate subject-specific parameters. See “Bayesian Parameter Estimates” on page 93.

**JMP PRO Number of Bayesian Iterations** (Applicable only if Hierarchical Bayes is selected.) The total number of iterations of the adaptive Bayes algorithm used to estimate subject-specific parameters. This number includes a burn-in period of iterations that are discarded. The number of burn-in iterations is equal to half of the Number of Bayesian Iterations specified on the launch window.

**Respondent is allowed to select “None” or “No Choice”** Enters a No Choice Indicator into the model for response rows containing missing values. For the One Table, Stacked data format, the No Choice rows must contain (numeric) missing values in the Response Indicator column. The option appears at the bottom of the launch window.

## Launch Window for Multiple Tables, Cross-Referenced

Figure 4.13 Launch Window for Multiple Tables, Cross-Referenced Data Format

Data Format: Multiple Tables, Cross-Referenced

**Profile Data**

Select Data Table: Pizza Profiles

Select Columns:

- Crust
- Cheese
- Topping
- ID

Pick Role Variables:

Profile ID: required

Grouping: optional

Run Model

Help

Remove

☒ Firth Bias-Adjusted Estimates

☐ Hierarchical Bayes

Number of Bayesian Iterations: 5000

Construct Profile Effects:

Add

Cross

Nest

Macros

Degree: 2

Transform: ☒

Response Data

Subject Data

Figure 4.13 shows the launch window for Multiple Tables, using Pizza Profiles.jmp as the Profile table.

In the case of Multiple Tables, Cross-referenced, the launch window has three sections:

- “Profile Data” on page 87
- “Response Data” on page 88
- “Subject Data” on page 90

## Profile Data

The profile data table describes the attributes associated with each choice. Each attribute defines a column in the data table. There is a row for each profile. A column in the table contains a unique identifier for each profile. Figure 4.14 shows the Pizza Profiles.jmp data table and a completed Profile Data panel.

**Figure 4.14** Profile Data Table and Completed Profile Data Outline

	Crust	Cheese	Topping	ID
1	Thick	Mozzarella	Pepperoni	ThickOni
2	Thick	Mozzarella	None	ThickElla
3	Thick	Jack	Pepperoni	ThickJackoni
4	Thick	Jack	None	ThickJack
5	Thin	Mozzarella	Pepperoni	TrimOni
6	Thin	Mozzarella	None	Trimella
7	Thin	Jack	Pepperoni	TrimPepperjack
8	Thin	Jack	None	TrimJack

**Profile Data**

Select Data Table: Pizza Profiles

Select Columns: Crust, Cheese, Topping, ID

Pick Role Variables: Profile ID: ID, Grouping: optional

Run Model, Help, Remove

☒ Firth Bias-Adjusted Estimates  
☐ Hierarchical Bayes  
 Number of Bayesian Iterations: 5000

Construct Profile Effects: Add, Cross, Nest, Macros, Degree: 2, Transform: None

Response Data, Subject Data

**Select Data Table** Select or open the data table that contains the profile data. Select Other to open a file that is not already open.

**Profile ID** Identifier for each row of attribute combinations (profile). If the **Profile ID** column does not uniquely identify each row in the profile data table, you need to add **Grouping** columns. Add **Grouping** columns until the combination of **Grouping** and **Profile ID** columns uniquely identify the row, or profile.

**Grouping** A column which, when used with the Profile ID column, uniquely designates each choice set. For example, if Profile ID = 1 for Survey = A, and a different Profile ID = 1 for Survey = B, then Survey would be used as a **Grouping** column.

**Construct Profile Effects** Add effects constructed from the attributes in the profiles.

For information about the Construct Profile Effects panel, see the Construct Model Effects section in the Model Specification chapter of the *Fitting Linear Models* book.

**Firth Bias-adjusted Estimates** Computes bias-corrected MLEs that produce better estimates and tests than MLEs without bias correction. These estimates also improve separation problems that tend to occur in logistic-type models. Refer to Heinze and Schemper (2002) for a discussion of the separation problem in logistic regression.



**Hierarchical Bayes** Uses a Bayesian approach to estimate subject-specific parameters. See “[Bayesian Parameter Estimates](#)” on page 93.



**Number of Bayesian Iterations** (Applicable only if Hierarchical Bayes is selected.) The total number of iterations of the adaptive Bayes algorithm used to estimate subject effects. This number includes a burn-in period of iterations that are discarded. The number of burn-in iterations is equal to half of the Number of Bayesian Iterations specified on the launch window.

## Response Data

The response data table includes a subject identifier column, columns that list the profile identifiers for the profiles in each choice set, and a column containing the preferred profile identifier. There is a row for each subject and choice set. Grouping variables can be used to distinguish choice sets when the data contain more than one group of choice sets. Figure 4.15 shows the Pizza Responses.jmp data table and a completed Response Data panel.

Grouping variables can be used to align choice indices when more than one group is contained within the data.



**Figure 4.15** Response Data Table and Completed Responses Data Outline

	Subject	Choice1	Choice2	Choice
1	1	ThickJack	TrimPepperjack	TrimPepperjack
2	1	TrimPepperjack	ThickElla	ThickElla
3	1	TrimOni	Trimella	TrimOni
4	1	ThickElla	ThickJack	ThickElla
5	2	Trimella	ThickJackoni	Trimella
6	2	TrimJack	ThickElla	ThickElla
7	2	Trimella	TrimPepperjack	Trimella
8	2	TrimPepperjack	TrimOni	TrimOni
9	3	TrimOni	ThickJackoni	TrimOni
10	3	TrimPepperjack	ThickElla	ThickElla
11	3	ThickJackoni	TrimPepperjack	ThickJackoni
12	3	ThickOni	Trimella	ThickOni
13	4	ThickElla	ThickOni	ThickElla
14	4	TrimPepperjack	ThickJack	ThickJack

**Response Data**

Select Data Table: Pizza Responses

Select Columns

- Subject
- Choice1
- Choice2
- Choice

Pick Role Variables

Profile ID Chosen: Choice

Profile ID Choices: Choice1, Choice2 (optional)

Grouping: (optional)

Subject ID: Subject

Freq: (optional numeric)

Weight: (optional numeric)

By: (optional)

☐ Respondent is allowed to select "None" or "No Choice"

**Subject Data**

**Select Data Table** Select or open the data table that contains the response data. Select Other to open a file that is not already open.

**Profile ID Chosen** The Profile ID from the Profile data table that represents the subject's selected profile.

**Grouping** A column which, when used with the Profile ID Chosen column, uniquely designates each choice set.

**Profile ID Choices** The Profile IDs of the set of possible profiles.

**Subject ID** An identifier for the study participant.

**Freq** A column containing frequencies. If  $n$  is the value of the Freq variable for a given row, then that row is used in computations  $n$  times. If it is less than 1 or missing, then JMP does not use it to calculate any analyses.

**Weight** A column containing a weight for each observation in the data table. The weight is included in analyses only when its value is greater than zero.

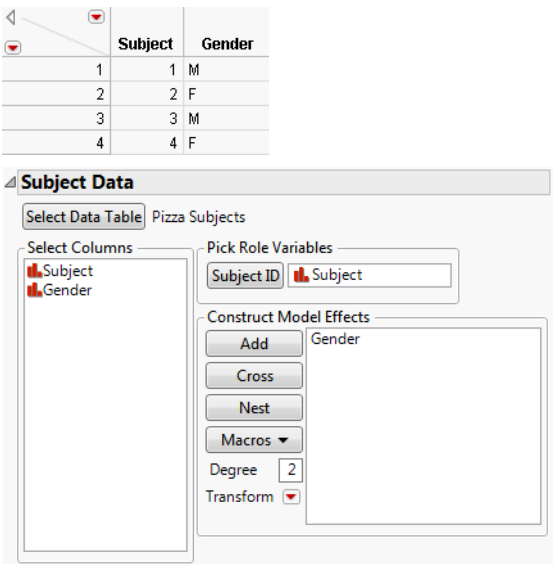
**By** A column whose levels define separate analyses. For each level of the specified column, the corresponding rows are analyzed as a separate analysis on a separate table. The results are presented in separate reports. If more than one By variable is assigned, a separate analysis is produced for each possible combination of the levels of the By variables.

**Respondent is allowed to select “None” or “No Choice”** Enters a No Choice Indicator into the model for response rows containing missing values. For the Multiple Tables, Cross-Referenced data format, the No Choice rows must contain (categorical) missing values in the Profile ID Chosen column in the Response Data table. The option appears at the bottom of the Response Data panel.

Subject Data

The subject data table is optional and depends on whether you want to model subject effects. The table contains a column with the subject identifier used in the response table, and columns for attributes or characteristics of the subjects. You can put subject data in the response data table, but you should specify the subject effects in the Subject Data outline. Figure 4.16 shows the Pizza Subjects.jmp data table and a completed Subject Data panel.

Figure 4.16 Subject Data Table and Completed Subject Data Outline



**Select Data Table** Select or open the data table that contains the subject data. Select Other to open a file that is not already open.

**Subject ID** Unique identifier for the subject.

**Construct Model Effects** Add effects constructed from columns in the subject data table.

For information about the Construct Model Effects panel, see the Construct Model Effects section in the Model Specification chapter of the *Fitting Linear Models* book.

---

## Choice Model Report

- [“Effect Summary”](#)
- [“Parameter Estimates”](#)
- [“Likelihood Ratio Tests”](#)
- [“Bayesian Parameter Estimates”](#)

### Effect Summary

The Effect Summary report appears if your model contains more than one effect and if it can be calculated quickly. (If the report does not appear, select Likelihood Ratio Tests from the red triangle menu to make both reports appear.) It lists the effects estimated by the model and gives a plot of the LogWorth (or FDR LogWorth) values for these effects. The report also provides controls that enable you to add or remove effects from the model. The model fit report updates automatically based on the changes made in the Effects Summary report. For details, see the Effect Summary Report section in the Standard Least Squares Report and Options chapter in the *Fitting Linear Models* book.

The Effect Summary report does not appear when Bayesian Subject Effects is checked in the launch window. This is because likelihood ratio tests are not conducted in this case.

### Effect Summary Table Columns

The Effect Summary table contains the following columns:

**Source** Lists the model effects, sorted by ascending  $p$ -values.

**LogWorth** Shows the LogWorth for each model effect, defined as  $-\log_{10}(p\text{-value})$ . This transformation adjusts  $p$ -values to provide an appropriate scale for graphing. A value that exceeds 2 is significant at the 0.01 level (because  $-\log_{10}(0.01) = 2$ ).

**FDR LogWorth** Shows the False Discovery Rate LogWorth for each model effect, defined as  $-\log_{10}(\text{FDR PValue})$ . This is the best statistic for plotting and assessing significance. Select the **FDR** check box to replace the LogWorth column with the **FDR LogWorth** column.

**Bar Chart** Shows a bar chart of the LogWorth (or FDR LogWorth) values. The graph has dashed vertical lines at integer values and a blue reference line at 2.

**PValue** Shows the  $p$ -value for each model effect. This is the  $p$ -value corresponding to the significance test displayed in the Likelihood Ratio Tests report.

**FDR PValue** Shows the False Discovery Rate  $p$ -value for each model effect calculated using the Benjamini-Hochberg technique. This technique adjusts the  $p$ -values to control the false discovery rate for multiple tests. Select the **FDR** check box to replace the **PValue** column with the **FDR PValue** column.

For details about the FDR correction, see Benjamini and Hochberg (1995). For details about the false discovery rate, see the Response Screening chapter in the *Predictive and Specialized Modeling* book or Westfall et al. (2011).

## Effect Summary Table Options

The options below the summary table enable you to add and remove effects:

**Remove** Removes the selected effects from the model. To remove one or more effects, select the rows corresponding to the effects and click the Remove button.

**Add Profile Effect** Opens a panel that contains a list of all columns in the data table for the OneTable, Stacked data format, and for the columns in the Profile Data table for the Multiple Tables, Cross-Referenced data format. Select columns that you want to add to the model, and then click Add below the column selection list to add the columns to the model. Click Close to close the panel.

**Add Subject Effect** Opens a panel that contains a list of all columns in the data table for the OneTable, Stacked data format, and for the columns in the Subject Data table for the Multiple Tables, Cross-Referenced data format. Select columns that you want to add to the model, and then click Add below the column selection list to add the columns to the model. Click Close to close the panel.

## Parameter Estimates

The Parameter Estimates report gives estimates and standard errors of the coefficients of utility associated with the effects listed in the Term column. The coefficients associated with attributes are sometimes referred to as *part-worths*. When the Firth Bias-Adjusted Estimates option is selected in the launch window, the parameter estimates are based on the Firth bias-corrected maximum likelihood estimators. These estimates considered to be more accurate than MLEs without bias correction. For details about utility, see “[Utility and Probabilities](#)” on page 126.

### Comparison Criteria

The following fit statistics are shown as part of the report and can be used to compare models: AICc (corrected Akaike's Information Criterion), BIC (Bayesian Information Criterion),  $-2 \times \text{LogLikelihood}$ , and  $-2 \times \text{Firth LogLikelihood}$ . For details and formulas, see the section Likelihood, AICc, and BIC in the Statistical Details appendix of the *Fitting Linear Models* book.

The  $-2 \times \text{Firth LogLikelihood}$  fit statistic is included in the report when the Firth Bias-Adjusted Estimates option is selected in the launch window. Note that this option is checked by default. The decision to use or not use the Firth Bias-Adjusted Estimates does not affect the AICc score or the  $-2 \times \text{LogLikelihood}$  results.

---

**Note:** For each of these statistics, a smaller value indicates a better fit.

---

## Likelihood Ratio Tests

The Likelihood Ratio Test report appears by default if the model is fit in less than five seconds. If the report does not appear, you can select the Likelihood Ratio Tests option from the Choice Model red triangle menu. The report gives the following:

**Source** Lists the effects in the model.

**L-R ChiSquare** The value of the likelihood ratio ChiSquare statistic for a test of the corresponding effect.

**DF** The degrees of freedom for the ChiSquare test.

**Prob>ChiSq** The  $p$ -value for the ChiSquare test.

**Bar Graph** Shows a bar chart of the L-R ChiSquare values.

## Bayesian Parameter Estimates

(Appears only if Hierarchical Bayes is selected on the launch window.) The Bayesian Parameter Estimates report gives results for model effects. The estimates are based on a Hierarchical Bayes fit that integrates the subject-level covariates into the likelihood function and estimates their effects on the parameters directly. The subject-level covariates are estimated using a Bayesian procedure combined with the Metropolis-Hastings algorithm. See Train (2001). Posterior means and variances are calculated for each model effect. The algorithm also provides subject-specific estimates of the model effect parameters. See [“Save Subject Estimates”](#) on page 97.

During the estimation process, each individual is assigned his or her own vector of parameter estimates, essentially treating the estimates as random effects and covariates. The vector of

coefficients for an individual is assumed to come from a multivariate normal distribution with arbitrary mean and covariance matrix. The likelihood function for the utility parameters for a given subject is based on a multinomial logit model for each subject’s preference within a choice set, given the attributes in the choice set. The prior distribution for a given subject’s vector of coefficients is normal with mean equal to zero and a diagonal covariance matrix with the same variance for each subject. The covariance matrix is assumed to come from an inverse Wishart distribution with a scale matrix that is diagonal with equal diagonal entries.

For each subject, a number of burn-in iterations at the beginning of the chain is discarded. By default, this number is equal to half of the Number of Bayesian Iterations specified on the launch window.

Figure 4.17 Bayesian Parameter Estimates Report

Choice Model			
Bayesian Parameter Estimates			
Term	Posterior Mean	Posterior Std Dev	Subject Std Dev
Crust[Thick]	0.27529781	0.724891965	2.815958945
Cheese[Jack]	-7.01110292	4.697169974	2.593184268
Topping[Pepperoni]	-1.06702410	0.941602303	2.260232580
Total Iterations	5000		
Burn-In Iterations	2500		
Number of Respondents	32		
Avg Log Likelihood After Burn-In	-14.19342		

**Term** The model term.

**Posterior Mean** The parameter estimate for the term’s coefficient. For each iteration after the burn-in period, the mean of the subject-specific coefficient estimates is computed. The Posterior Mean is the average of these means.

**Tip:** Select the red-triangle option Save Bayes Chain to see the individual estimates for each iteration.

**Posterior Std Dev** The standard deviation of the means of the subject-specific estimates over the iterations after burn-in.

**Subject Std Dev** The standard deviation of the subject-specific estimates.

**Tip:** Select the red-triangle option Save Subject Estimates to see the individual estimates.

**Total Iterations** The total number of iterations performed, including the burn-in period.

**Burn-In Iterations** The number of burn-in iterations. This number is equal to half of the Number of Bayesian Iterations specified on the launch window.

**Number of Respondents** The number of subjects.

**Avg Log Likelihood After Burn-In** The average of the log-likelihood function, computed on values obtained after the burn-in period.

---

## Choice Platform Options

The Choice Modeling platform has many available options. To access these options, select the Choice Model red triangle menu.

---

**Note:** When you use Hierarchical Bayes, the subject-level estimates are based on Monte Carlo sampling. For this reason, results obtained for the options below vary from run to run.

---

**Likelihood Ratio Tests** See [“Likelihood Ratio Tests”](#) on page 93.

**JMP PRO Show MLE Parameter Estimates** (Available only if Hierarchical Bayes is selected on the launch window.) Shows non-Firth maximum likelihood estimates and standard errors for the coefficients of model terms. These estimates are used as starting values for the Hierarchical Bayes algorithm.

**Joint Factor Tests** (Not available if Hierarchical Bayes is selected on the launch window.) Tests each factor in the model by constructing a likelihood ratio test for all the effects involving that factor. For more information about Joint Factor Tests, see the Standard Least Squares Report and Options chapter in the *Fitting Linear Models* book.

**Confidence Intervals** (Available for Firth Bias-Adjusted Estimates) Shows a confidence interval for each parameter in the Parameter Estimates report.

**Confidence Limits** (Available for Hierarchical Bayes) Shows confidence limits for each parameter in the Bayesian Parameter Estimates report. The limits are constructed based on the 2.5 and 97.5 quantiles of the posterior distribution.

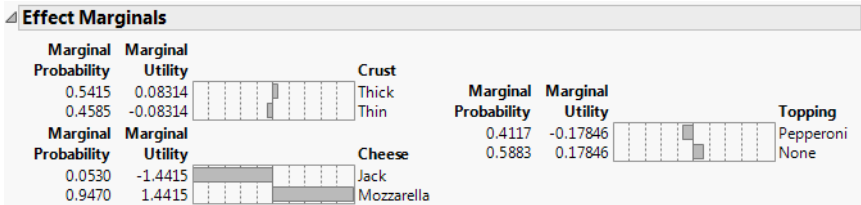
**Correlation of Estimates** If Hierarchical Bayes was not selected, shows the correlations between the maximum likelihood parameter estimates.

If you selected Hierarchical Bayes, shows the correlation matrix for the posterior means of the parameter estimates. The correlations are calculated from the iterations after burn-in. The posterior means from each iteration after burn-in are treated as if they are columns in a data table. The Correlation of Estimates table is obtained by calculating the correlation matrix for these columns.

**Effect Marginals** Shows marginal probabilities and marginal utilities for each main effect in the model. The marginal probability is the probability that an individual selects attribute A over B with all other attributes at their mean or default levels.

In Figure 4.18, the marginal probability of any subject choosing a pizza with mozzarella cheese, thick crust and pepperoni, over that same pizza with Monterey Jack cheese instead of mozzarella, is 0.9470.

Figure 4.18 Example of Marginal Effects



**Utility Profiler** Shows the predicted utility for different factor settings. The utility is the value predicted by the linear model. See [“Find Optimal Profiles”](#) on page 82 for an example of the Utility Profiler. For details about utility, see [“Utility and Probabilities”](#) on page 126. For details about the Utility Profiler options, see the Prediction Profiler Options section in the Profiler chapter of the *Profilers* book.

**Probability Profiler** Enables you to compare choice probabilities among a number of potential products. This predicted probability is defined as

$$(\exp(U))/(\exp(U) + \exp(U_b))$$

where  $U$  is the utility for the current settings and  $U_b$  is the utility for the baseline settings. This implies that the probability for the baseline settings is 0.5. For details, see [“Utility and Probabilities”](#) on page 126.

See [“Comparisons to Baseline”](#) on page 105 for an example of using the Probability Profiler. For details about the Probability Profiler options, see the Prediction Profiler Options section in the Profiler chapter of the *Profilers* book.

**Multiple Choice Profiler** Provides the number of probability profilers that you specify. This enables you to set each profiler to the settings of a given profile so that you can compare the probabilities of choosing each profile relative to the others. See [“Multiple Choice Comparisons”](#) on page 107 for an example of using the Multiple Choice Profiler. For details about the Multiple Choice Profiler options, see the Prediction Profiler Options section in the Profiler chapter of the *Profilers* book.

**Comparisons** Performs comparisons between specific alternative choice profiles. Enables you to select the factors and the values that you want to compare. You can compare specific configurations, including comparing all settings on the left or right by selecting



the **Any** check boxes. If you have subject effects, you can select the levels of the subject effects to compare. Using Any does not compare all combinations across features, but rather all combinations of comparisons, one feature at a time, using the left settings as the settings for the other factors.

**Figure 4.19** Utility Comparisons Window

**Willingness to Pay** Requires that your data table contains a continuous price column. Calculates how much a price must change allowing for the new feature settings to produce the same predicted outcome. The result is calculated using the Baseline settings for each background setting.

**Save Utility Formula** When the analysis is on multiple data tables, creates a new data table that contains a formula column for utility. The new data table contains a row for each subject and profile combination, and columns for the profiles and the subject effects. When the analysis is on one data table, a new Utility Formula column is added.

**Save Gradients by Subject** Constructs a new table that has a row for each subject containing the average (Hessian-scaled-gradient) steps for the likelihood function on each parameter. This corresponds to using a Lagrangian multiplier test for separating that subject from the remaining subjects. These values can later be clustered, using the built-in-script, to indicate unique market segments represented in the data. For more details, see [“Gradients”](#) on page 127. For an example, see [“Example of Segmentation”](#) on page 109.

**JMP PRO Save Subject Estimates** (Available only if Hierarchical Bayes is selected.) Creates a table where each row contains the subject-specific parameter estimates for each effect. The distribution of subject-specific parameter effects for each effect is centered at the estimate for the term given in the Bayesian Parameter Estimates report. The Subject Acceptance Rate gives the rate of acceptance for draws of new parameter estimates during the Metropolis-Hastings step. Generally, an acceptance rate of 0.20 is considered to be good. See [“Bayesian Parameter Estimates”](#) on page 93.

**JMP PRO Save Bayes Chain** (Available only if Hierarchical Bayes is selected.) Creates a table that gives information about the chain of iterations used in computing subject-specific Bayesian estimates. See [“Save Bayes Chain”](#) on page 100.

**Model Dialog** Shows the Choice launch window, which can be used to modify and re-fit the model. You can specify new data sets, new IDs, and new model effects.

See the JMP Reports chapter in the *Using JMP* book for more information about the following options:

**Redo** Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

**Save Script** Contains options that enable you to save a script that reproduces the report to several destinations.

**Save By-Group Script** Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

## Willingness to Pay

The term *willingness to pay* refers to the price that a customer is willing to pay for new features, calculated to match a customer’s utility for baseline features. For example, suppose that a customer is willing to pay \$1,000 for a computer with a 40 GB hard drive. Willingness to Pay for an 80 GB hard drive is calculated by setting the Hard drive feature to 80 GB and then solving for the price that delivers the same utility as the \$1000 40 GB hard drive.

### Willingness to Pay Launch Window Options

When you select the Willingness to Pay option, the Willingness to Pay launch window is shown. The launch window in Figure 4.20 is obtained by selecting the Willingness to Pay option in the report that results from running the **Choice** data table script in *Laptop Profile.jmp*.

**Factor** The variables from the analysis. These can be product features or subject-specific attributes.

**Baseline** The baseline setting for each factor. If the factor is categorical, select the baseline value from a list. If the factor is numeric, enter the baseline value.

**Role** The type of factor. You can choose from the following list:

**Feature Factor** A product or service feature from the experiment that you want to price.

**Price Factor** A price factor in the experiment. The price factor must be continuous, and there can be only one specified price factor for each Willingness to Pay analysis.

**Background Constant** A factor that you want to hold constant in the Willingness to Pay calculation. Generally, these are subject-specific variables.

**Background Variable** A factor that you want to hold constant, at each of its levels, in the Willingness to Pay calculation. Generally, these are subject-level factors. Specifying a subject factor as a Background Variable rather than a Background Constant provides Willingness to Pay estimates for all levels of the variable.

**Include baseline settings in report table** Adds the baseline settings with a price change of zero to the Willingness to Pay report.

**Tip:** If you make an output table, use this option to display all the baseline settings as well as the attribute settings.

**Output data table also** Creates a data table containing the Willingness to Pay report.

**Figure 4.20** Willingness to Pay Launch Window

Factor	Baseline	Role
Hard Disk	40 GB	Feature Factor
Speed	1.5 GHz	Feature Factor
Battery Life	4 hours	Feature Factor
Price	1000	Price Factor

Enter baseline factor values and role.  
Exactly one role must be a continuous pricing factor.

☐ Include baseline settings in report table

☐ Output data table also

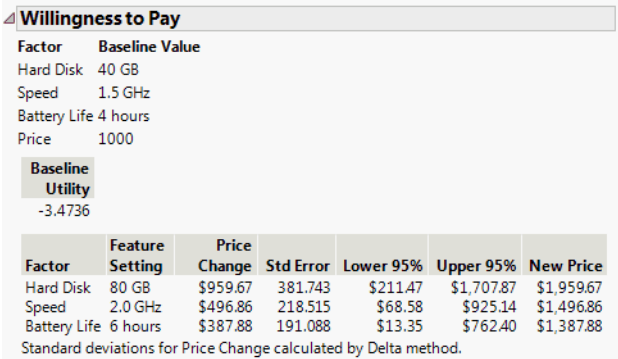
OK Cancel Help

Once you complete your first Willingness to Pay calculation, the platform remembers the baseline values and assigned roles that you selected. This enables you to do multiple Willingness to Pay comparisons without having to re-enter the baseline information. If there is no factor called Price, but there is a continuous factor used in the analysis, the continuous factor is automatically assigned as the Price factor in the Willingness to Pay window. Common cost variables that are not prices in the traditional sense include factors such as travel time or distance.

Willingness to Pay Report

The Willingness to Pay report displays the baseline value for each factor, as well as baseline utility values. For each factor, the report shows the feature setting, estimated price change, and new price. If there are no interaction or second-order effects, standard errors and confidence intervals are also shown. These are calculated using the delta method.

Figure 4.21 Willingness to Pay Report



JMP PRO Save Bayes Chain

You can use the Bayes Chain data table to determine whether your estimates have stabilized. The table that is created has a number of rows equal to the Number of Bayesian Iterations (specified on the launch window) plus one. The first row, Iteration 1, gives the starting values. The following rows show the results of the iterations, in order. The columns are arranged as follows:

- Iteration** Gives the iteration number, where the first row shows starting values.
- Log Likelihood** The log-likelihood of the model for that iteration. You can plot the Log Likelihood against Iteration to view behavior over the burn-in and tuning periods.
- Adaptive Sigma for <model effect>** Gives the estimate of the square root of the diagonal entries of the inverse Wishart distribution scale matrix for the corresponding effect.
- Acceptance for <model effect>** Gives the sampling acceptance rate for the corresponding effect.
- Mean of <model effect>** Gives the estimated mean for the corresponding effect.
- Variance of <model effect>** Gives the estimated variance for the corresponding effect.

---

## Additional Examples

- [“Example of Making Design Decisions”](#)
- [“Example of Segmentation”](#)
- [“Example of Logistic Regression Using the Choice Platform”](#)
- [“Example of Logistic Regression for Matched Case-Control Studies”](#)
- [“Example of Transforming Data to Two Analysis Tables”](#)
- [“Example of Transforming Data to One Analysis Table”](#)

### Example of Making Design Decisions

You can use the Choice Modeling platform to determine the relative importance of product attributes. Even if the attributes of a particular product that are important to the consumer are known, information about preference trade-offs with regard to these attributes might be unknown. By gaining such information, a market researcher or product designer is able to incorporate product features that represent the optimal trade-off from the perspective of the consumer. This example illustrates the advantages of this approach to product design.

It is already known that four attributes are important for laptop design: hard-disk size, processor speed, battery life, and selling price. The data gathered for this study are used to determine which of four laptop attributes (Hard Disk, Speed, Battery Life, and Price) are most important. It also assesses whether there are Gender or Job effects associated with these attributes.

This example has the following sections:

- [“Complete the Launch Window”](#) on page 101
- [“Analyze the Model”](#) on page 103
- [“Comparisons to Baseline”](#) on page 105
- [“Multiple Choice Comparisons”](#) on page 107

### Complete the Launch Window

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

---

**Note:** If you prefer not to follow the manual steps in this section, click the green triangle next to the script **Choice with Gender** to run the model, and go to [“Analyze the Model”](#) on page 103.

---

2. Click the green triangle next to the **Open Profile and Subject Tables** script.  
The script opens the **Laptop Profile.jmp** and **Laptop Subjects.jmp** data tables.

3. Select **Analyze > Consumer Research > Choice**.

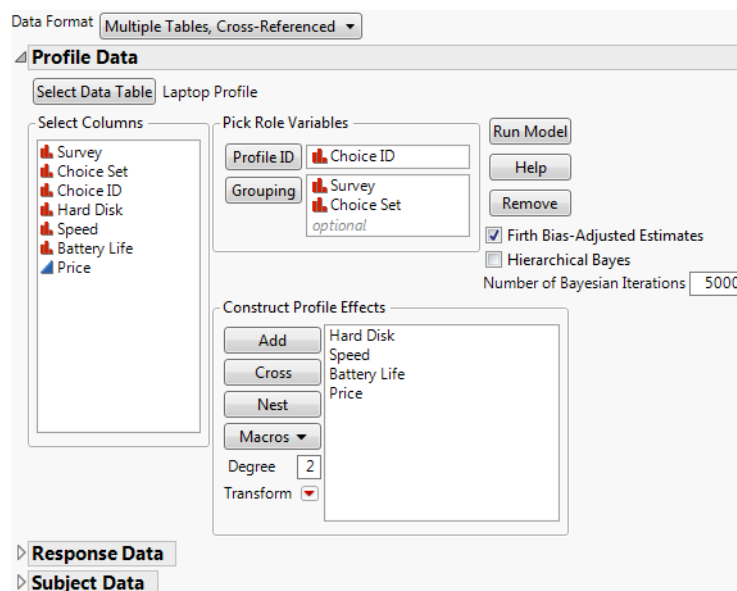
---

**Note:** This can be done from any of the three open data tables.

---

4. From the Data Format list, select **Multiple Tables, Cross-Referenced**.
5. Click **Select Data Table** under Profile Data and select Laptop Profile.jmp. Select Choice ID and click Profile ID.
6. Select Hard Disk, Speed, Battery Life, and Price and click **Add**.
7. Select Survey and Choice Set and click **Grouping**.

**Figure 4.22** Profile Data Window for Laptop Study



8. Open the **Response Data** outline.
9. From the **Select Data Table** list, select Laptop Runs.jmp.
10. Complete the Response Data table as follows:
  - Select Response and click **Profile ID Chosen**.
  - Select Choice1 and Choice2 and click **Profile ID Choices**.
  - Select Survey and Choice Set and click **Grouping**
  - Select Person and click **Subject ID**. The Response Data window is shown in Figure 4.23.

**Figure 4.23** Response Data Window for Laptop Study

**Response Data**

Select Data Table: Laptop Runs

Select Columns:

- Survey
- Choice Set
- Choice1
- Choice2
- Response
- Person

Pick Role Variables:

Profile ID Chosen: Response

Profile ID Choices: Choice1, Choice2 (optional)

Grouping: Survey, Choice Set (optional)

Subject ID: Person

Freq: optional numeric

Weight: optional numeric

By: optional

☐ Respondent is allowed to select "None" or "No Choice"

**Subject Data**

11. Open the **Subject Data** outline.
12. From the **Select Data Table** list, select Laptop Subjects.jmp.
13. Select Person and click **Subject ID**.
14. Select Gender click **Add**.

The Subject Data window is shown in Figure 4.24.

**Figure 4.24** Subject Data Window for Laptop Study

**Subject Data**

Select Data Table: Laptop Subjects

Select Columns:

- Person
- Gender
- Job

Pick Role Variables:

Subject ID: Person

Construct Model Effects:

Add: Gender

Cross

Nest

Macros

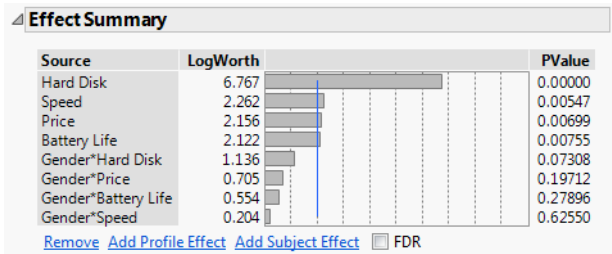
Degree: 2

Transform: ☒

## Analyze the Model

1. Click **Run Model**.

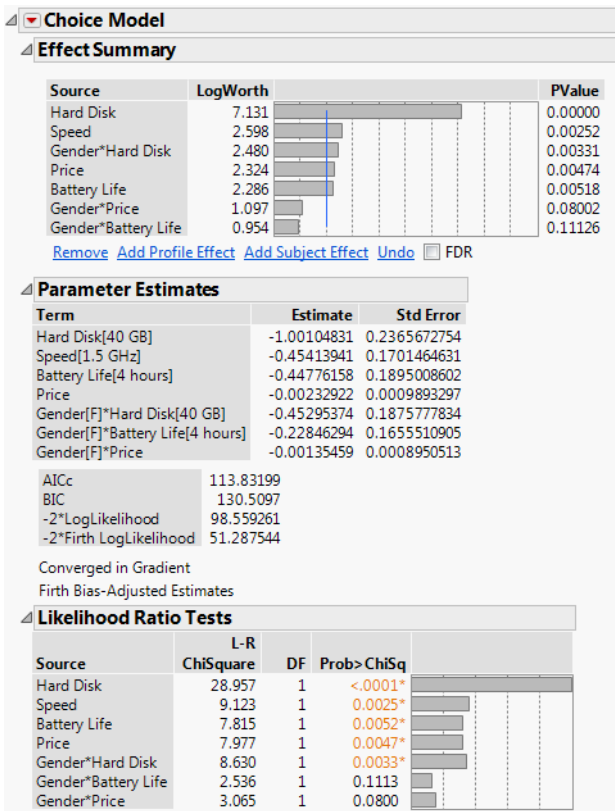
Figure 4.25 Laptop Effect Summary



The Effect Summary report shows that Hard Disk is the most significant effect. You can reduce the model by removing terms with a  $p$ -value greater than 0.15. This process should be done one term at a time. Here, Gender\*Speed is the least significant effect, with a  $p$ -value of 0.625.

2. In the Effect Summary report, select Gender\*Speed and click **Remove**.

Figure 4.26 Laptop Results

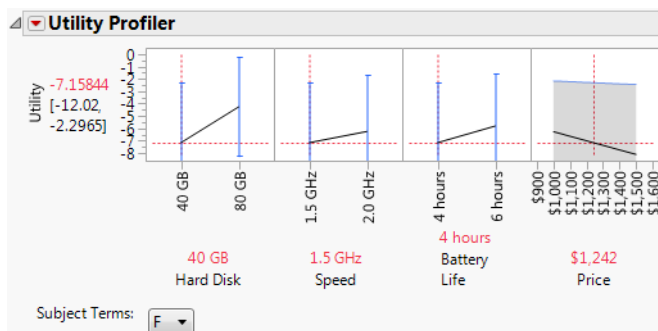




Once Gender\*Speed is removed from the model, all effects have a  $p$ -value of 0.15 or less. Therefore, you use this as your final model.

- Click the Choice Model red triangle and select **Utility Profiler**.

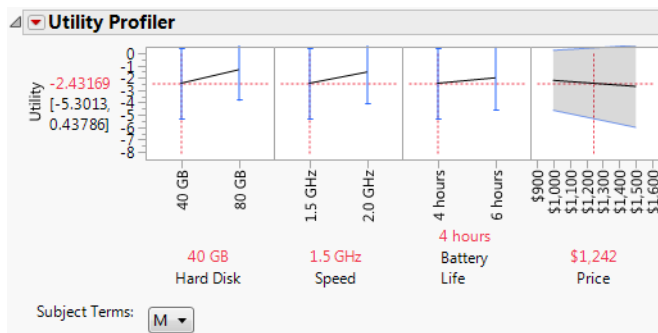
**Figure 4.27** Laptop Profiler Results for Females



**Tip:** If your utility profiler does not look like Figure 4.27, click the red triangle next to Utility Profiler and select **Appearance > Adapt Y Axis**.

- From the list next to Subject Terms, select **M**.

**Figure 4.28** Laptop Profiler Results for Males in Development



The interaction effect between Gender and Hard Disk is highly significant, with a  $p$ -value of 0.0033. See [Figure 4.26](#) on page 104. In the Utility Profilers, check the slope for Hard Disk for both levels of Gender. You see that the slope is steeper for females than for males.

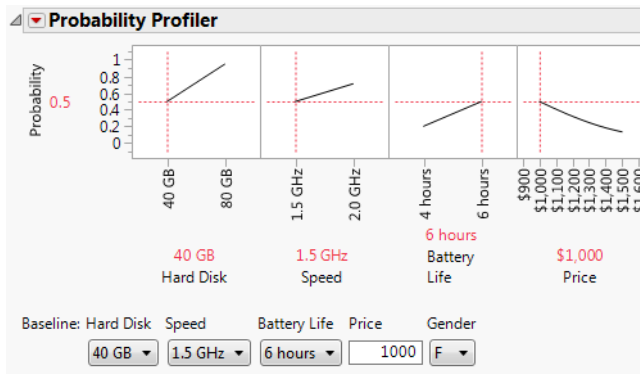
## Comparisons to Baseline

Suppose you are developing a new product. You want to explore the likelihood that a customer selects the new product over the old product, or over a competitor's product. Use the Probability Profiler to compare profiles to a baseline profile.

In this example, your company is currently producing laptops with 40 GB hard drives, 1.5 GHz processors, and 6-hour battery life, that cost \$1,000. You are looking for a way to make your product more desirable by changing as few factors as possible. You set the current product configuration as the baseline. JMP adjusts the probabilities so that the probability of preference for the baseline configuration is 0.5. Then you compare the probabilities of other configurations to the baseline probability.

1. Do one of the following:
  - Follow the steps in “[Complete the Launch Window](#)” on page 101. Then complete step 1 and step 2 in “[Analyze the Model](#)” on page 103.
  - In the Laptop Runs. jmp sample data table, click the green triangle next to the **Choice Reduced Model** script.
2. Click the Choice Model red triangle and select **Probability Profiler**.  
Note that the Probability Profiler is for Gender = F. You can change this later.
3. Using the menus and text box below the profiler, in the Baseline area, specify the Baseline settings as 40 GB, 1.5 GHz, 6 hours, and 1000.
4. Now set these as the values in the Probability Profiler. To set the Price at \$1000, click \$1242 above Price under the rightmost profiler cell, and type 1000. Then click outside the text box.

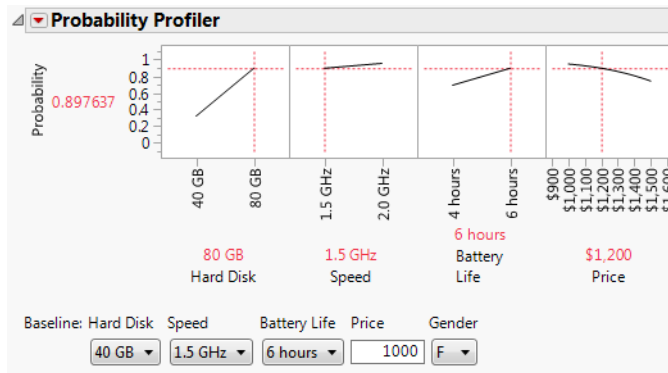
**Figure 4.29** Probability Profiler with Text Entry Area for Price



This configuration has probability 0.5.

5. In the Probability Profiler, move the slider for HardDisk to 80 GB.  
Notice that, with this change, the probability is relatively insensitive to increases in Price.
6. Click the \$1000 label above the Price cell in the profiler, type **\$1,200**, and click outside the text box.

**Figure 4.30** Laptop Probability Profiler Results with Baseline Effects



An increase in Hard Disk size from 40 GB to 80 GB and an increase in price to \$1200 coincides with an increased probability of preference, from 0.50 to 0.90 for females. Change the Gender effect in the Baseline to **M**. The probability of preference is 0.71.

## Multiple Choice Comparisons

Use the Multiple Choice Profiler to compare product profiles.

- You currently produce a low-end laptop with a small hard drive, a slow processor, and low battery life. You charge \$1000.
- Company A produces a product with a fast processor speed and high battery life at a reasonable price of \$1200.
- Company B makes the biggest hard drives with the fastest speed, but at a high price of \$1500 and low battery life.

You want to gain market share by increasing only one area of performance, and price.

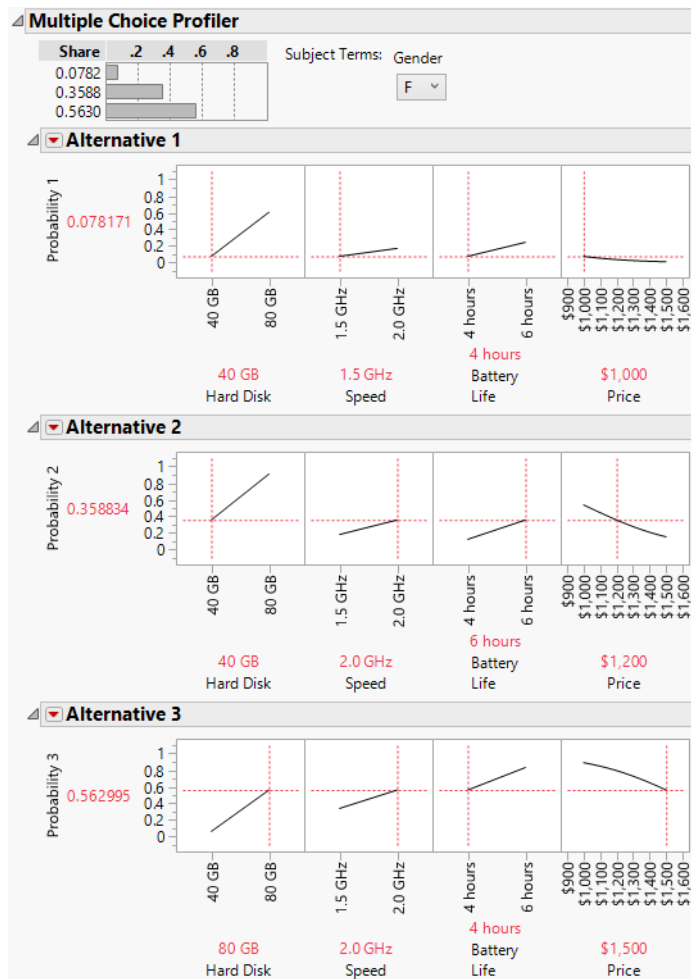
1. Do one of the following:
  - Follow the steps in [“Complete the Launch Window”](#) on page 101. Then complete step 1 and step 2 in [“Analyze the Model”](#) on page 103.
  - In the Laptop Runs.jmp sample data table, click the green triangle next to the **Choice Reduced Model** script.
2. Click the Choice Model red triangle and select **Multiple Choice Profiler**.  
A window appears, asking for the number of alternative choices to profile. Accept the default number of 3.
3. Click **OK**.

Three Alternative profilers appear. Notice that the profilers are set for Gender = F.

Each factor in each profiler is set to its default values. Alternative 1 indicates the product that you want to develop. Alternative 2 indicates Company A's product. Alternative 3 indicates Company B's product.

4. For Alternative 1, set Hard Disk to 40 GB, Speed to 1.5 GHz, Battery Life to 4 hours, and Price to \$1,000.
5. For Alternative 2, set Hard Disk to 40 GB, Speed to 2.0 GHz, Battery Life to 6 hours, and Price to \$1,200.
6. For Alternative 3, set Hard Disk to 80 GB, Speed to 2.0 GHz, Battery Life to 4 hours, and Price to \$1,500.

**Figure 4.31** Multiple Choice Profiler for Females

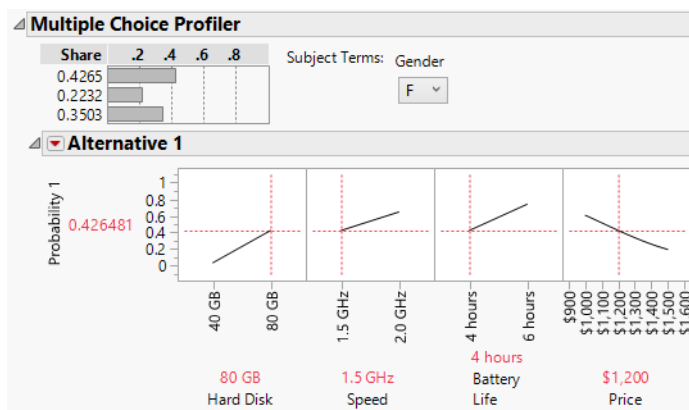


You can see that Company B has the greatest Share of 0.5630. It is obvious that with your company's settings, very few females buy your product.

You want to increase your market share by upgrading your company's laptop in one of the performance areas while increasing price. The slope of the line in Alternative 1's Hard Disk profile suggests increasing hard disk space increases market share the most.

7. For Alternative 1, set Hard Disk to **80 GB** and Price to **\$1,200**.

**Figure 4.32** Multiple Choice Profiler with Improved Laptop



By increasing hard disk space, you can increase the price of your laptop and expect a market share among females of about 43%. This share exceeds that of Company B's high-performance laptop and is much better than the market share with the initial low-end settings seen in Figure 4.31.

Explore the settings that increase your market share for males. If you increase both Hard Disk size and Speed, you can capture a 44% market share among males.

## Example of Segmentation

In this example, you attempt to identify market segments for pizza preferences.

To see how to complete the launch window for this example, see step 1 to step 15 in the example ["Multiple Table Format"](#) on page 78. Otherwise, follow the instructions below.

### Define Clusters

1. Select **Help > Sample Data Library** and open **Pizza Responses.jmp**.
2. Click the green triangle next to the **Choice** script.
3. Click the Choice Model red triangle and select **Save Gradients by Subject**.

A data table appears with gradient forces saved for each main effect and subject interaction.

Figure 4.33 Gradients by Subject for Pizza Data, Partial View

Untitled

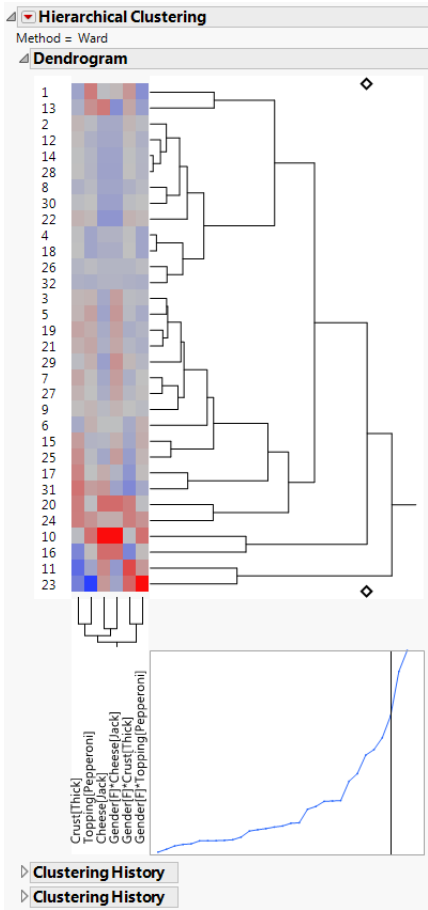
Hierarchical Cluster

Merge Data Back

	Subject	Crust[Thick]	Cheese[Jack]	Topping[Pepperoni]	Gender[F]*Crust [Thick]	Gender[F]*Cheese [Jack]	Gender[F]*Topping [Pepperoni]
1	1	-0.00959	-0.00168	0.014876	0.009585	0.001685	-0.01488
2	2	0.002373	-0.00758	-0.00239	0.002373	-0.00758	-0.00239
3	3	0.002129	-0.0079	0.003031	-0.00213	0.007899	-0.00303
4	4	-0.00106	-0.00485	-0.00901	-0.00106	-0.00485	-0.00901
5	5	0.002828	-0.00945	0.00725	-0.00283	0.009453	-0.00725
6	6	-0.0073	-0.00089	0.003761	-0.0073	-0.00089	0.003761

4. Click the green triangle next to the **Hierarchical Cluster** script.

Figure 4.34 Dendrogram of Subject Clusters for Pizza Data



The script runs a hierarchical cluster analysis on all columns in the gradient table, except for Subject. Click on either diamond to see that the rows have been placed into three clusters.

- Click the red triangle next to Hierarchical Clustering and select **Save Clusters**.

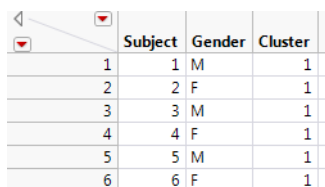
A new column called Cluster is added to the data table containing the gradients. Each subject has been assigned a Cluster value that is associated with other subjects having similar gradient forces. See the Hierarchical Cluster platform chapter in the *Multivariate Methods* book for a discussion of other Hierarchical Clustering options.

You can delete the gradient columns because they were used only to obtain the clusters.

- Select all columns except Subject and Cluster. Right-click on the selected columns and select **Delete Columns**.
- Click the green triangle next to the **Merge Data Back** script (Figure 4.33).

The cluster information is merged into the Subject data table. The columns in the Subject data table are now Subject, Gender, and Cluster, as shown in Figure 4.35.

**Figure 4.35** Subject Data with Cluster Column

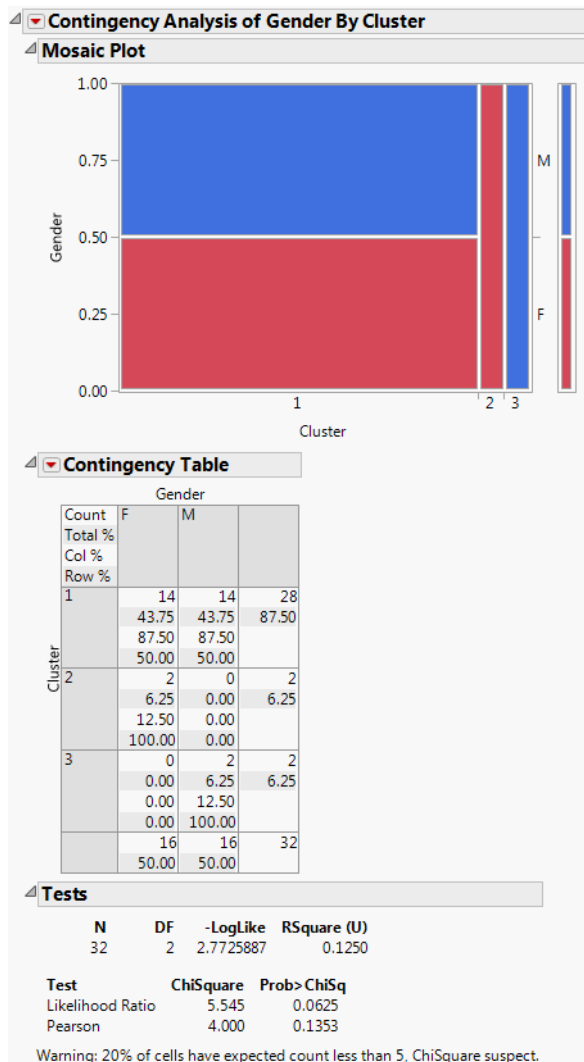


	Subject	Gender	Cluster
1	1	M	1
2	2	F	1
3	3	M	1
4	4	F	1
5	5	M	1
6	6	F	1

This table can now be used for further analysis.

### Explore the Clusters

- Click the icon to the left of the Cluster variable in the columns panel and select **Nominal**.
- Select **Analyze > Fit Y by X**.
- Select Gender and click **Y, Response**.
- Select Cluster and click **X, Factor**.
- Click **OK**.

**Figure 4.36** Contingency Analysis of Gender by Cluster

You see the following:

- Cluster 1 is evenly divided between males and females
- Cluster 2 consists of only females
- Cluster 3 consists of only males

If desired, you could now refit and analyze the model with the addition of the Cluster variable.

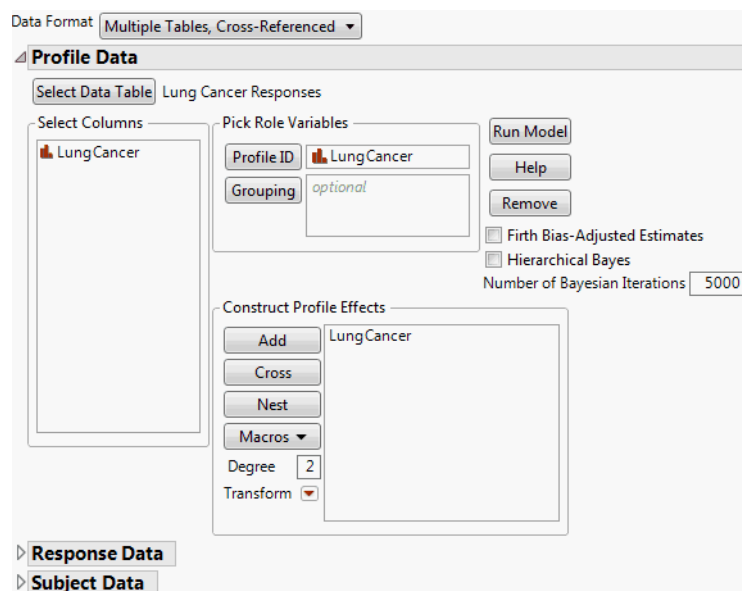


## Example of Logistic Regression Using the Choice Platform

### Use the Choice Platform

1. Select **Help > Sample Data Library** and open Lung Cancer Responses.jmp and Lung Cancer Choice.jmp.  
Notice Lung Cancer Responses.jmp has only one column (Lung Cancer) with two rows (Cancer and NoCancer).
2. Select **Analyze > Consumer Research > Choice**
3. Select **Multiple Tables, Cross-Referenced** from the list next to Data Format.
4. Click **Select Data Table**, select Lung Cancer Responses.jmp, and click **OK**.
5. Select Lung Cancer and click **Profile ID**.
6. Select Lung Cancer and click **Add**.
7. Uncheck the Firth Bias-Adjusted Estimates box.

**Figure 4.37** Completed Profile Data Panel



8. Open the **Response Data** outline.
9. Click **Select Data Table**, select Lung Cancer Choice.jmp, and click **OK**.
10. Do the following:
  - Select Lung Cancer and click **Profile ID Chosen**.
  - Select Choice1 and Choice2 and click **Profile ID Choices**.

- Select Count and click **Freq.**

Figure 4.38 Completed Response Data Panel

The screenshot shows the 'Response Data' panel for a table named 'Lung Cancer Choice'. On the left, under 'Select Columns', a list of variables includes Smoker, Lung Cancer, Count, Choice1, and Choice2. On the right, under 'Pick Role Variables', several roles are assigned: 'Profile ID Chosen' is Lung Cancer; 'Profile ID Choices' includes Choice1 and Choice2 (marked as optional); 'Grouping' is optional; 'Subject ID' is optional; 'Freq' is Count; 'Weight' is optional numeric; and 'By' is optional. At the bottom, there is a checkbox labeled 'Respondent is allowed to select "None" or "No Choice"' which is currently unchecked.

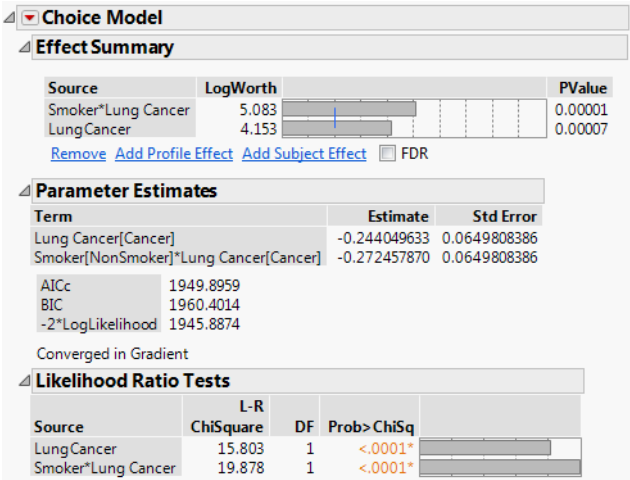
11. Open the **Subject Data** outline.
12. Click **Select Data Table**, select Lung Cancer Choice.jmp, and click **OK**.
13. Select Smoker and click **Add**.

Figure 4.39 Completed Subject Data Panel

The screenshot shows the 'Subject Data' panel for the same 'Lung Cancer Choice' table. The 'Select Columns' list on the left remains the same. In the 'Pick Role Variables' section, 'Subject ID' is set to optional. The 'Construct Model Effects' section on the right has 'Add' and 'Cross' buttons visible, with 'Smoker' listed in the adjacent box. Below these, there are buttons for 'Nest', 'Macros' (with a dropdown arrow), 'Degree' set to 2, and a 'Transform' dropdown menu.

14. Click **Run Model**.

Figure 4.40 Choice Modeling Logistic Regression Results



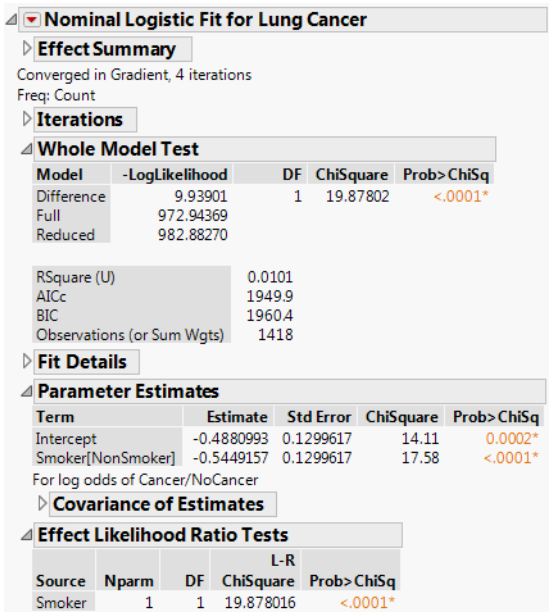
Use the Fit Model Platform

1. Select **Help > Sample Data Library** and open Lung Cancer.jmp.
2. Select **Analyze > Fit Model**.

Because the data table contains a model script, the Model Specification window is automatically completed. The **Nominal Logistic** personality is selected.

3. Click **Run**.

Figure 4.41 Fit Model Nominal Logistic Regression Results



Notice that the likelihood ratio chi-square test for Smoker\*Lung Cancer in the Choice model matches the likelihood ratio chi-square test for Smoker in the Logistic model. The reports shown in Figure 4.40 and Figure 4.41 support the conclusion that smoking has a strong effect on developing lung cancer. See the Logistic Regression chapter in the *Fitting Linear Models* book for more details.

### Example of Logistic Regression for Matched Case-Control Studies

This section provides an example using the Choice platform to perform logistic regression on the results of a study of endometrial cancer with 63 matched pairs. The data are from the Los Angeles Study of the Endometrial Cancer Data reported in Breslow and Day (1980). The goal of the case-control analysis was to determine the relative risk for gallbladder disease, controlling for the effect of hypertension. The Outcome of 1 indicates the presence of endometrial cancer, and 0 indicates the control. Gallbladder and Hypertension data indicators are also 0 or 1.

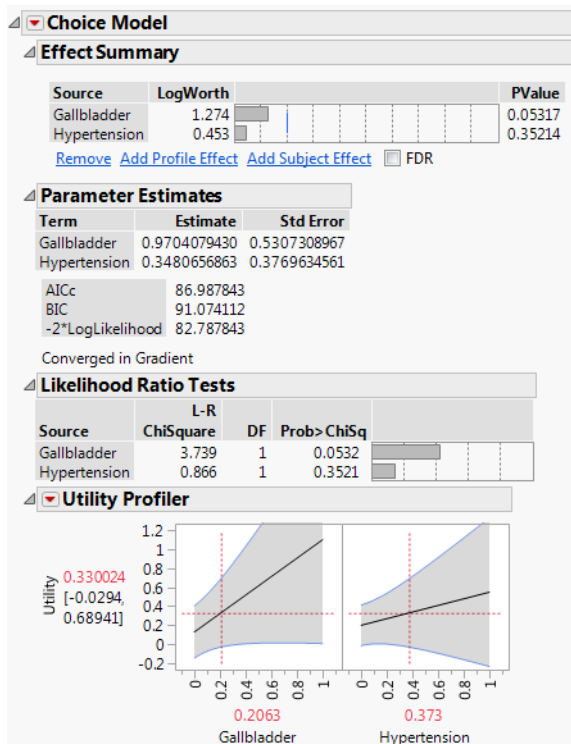
For details about performing logistic regression using the Choice platform, see “Logistic Regression” on page 126.

1. Select **Help > Sample Data Library** and open Endometrial Cancer.jmp.
2. Select **Analyze > Consumer Research > Choice**.
3. Check that the Data Format selected is **One-Table, Stacked**.

4. Click the **Select Data Table** button.
5. Select Endometrial Cancer as the profile data table. Click **OK**.
6. Select Outcome and click **Response Indicator**.
7. Select Pair and click **Grouping**.
8. Select Gallbladder and Hypertension and click **Add** in the Construct Profile Effects window.
9. Deselect the **Firth Bias-Adjusted Estimates** check box.
10. Click **Run Model**.
11. Click the Choice Model red triangle and select **Utility Profiler**.

The report is shown in Figure 4.42.

**Figure 4.42** Logistic Regression on Endometrial Cancer Data



Likelihood Ratio tests are given for each factor. Note that Gallbladder is nearly significant at the 0.05 level ( $p$ -value = 0.0532). Use the Utility Profiler to visualize the impact of the factors on the response.

## Example of Transforming Data to Two Analysis Tables

Consider the data from Daganzo, found in Daganzo Trip.jmp. This data set contains the travel time for three transportation alternatives and the preferred transportation alternative for each subject.

### Add Choice Mode and Subjects

1. Select **Help > Sample Data Library** and open the Daganzo Trip.jmp data table.

Figure 4.43 Partial Daganzo Trip Table

	Subway	Bus	Car	Choice
1	16.481	16.196	23.89	2
2	15.123	11.373	14.182	2
3	19.469	8.822	20.819	2
4	18.847	15.649	21.28	2
5	12.578	10.671	18.335	2

Each Choice number listed must first be converted to one of the travel mode names. This transformation is easily done by using the **Choose** function in the formula editor, as follows.

2. Select **Cols > New Columns**.
3. Specify the Column Name as Choice Mode and the modeling type as **Nominal**.
4. Click the **Column Properties** and select **Formula**.
5. Click **Conditional** in the functions list, select **Choose**, and press the comma key twice to obtain additional arguments for the function.
6. Click Choice for the Choose expression (expr), and double click each clause entry box to enter “Subway”, “Bus”, and “Car” (with the quotation marks) as shown in Figure 4.44.

Figure 4.44 Choose Function for Choice Mode Column of Daganzo Data

Choose(Choice)	1	=	"Subway"
	2	=	"Bus"
	else	=	"Car"

7. Click **OK** in the Formula Editor window.
8. Click **OK** in the New Column window.

The new Choice Mode column appears in the data table. Because each row contains a choice made by each subject, another column containing a sequence of numbers should be created to identify the subjects.

9. Select **Cols > New Columns**.

10. Specify the Column Name as Subject.
11. Click **Missing/Empty** next to Initialize Data and select **Sequence Data**.
12. Click **OK**.

**Figure 4.45** Partial Daganzo Trip Data with New Choice Mode and Subject Columns

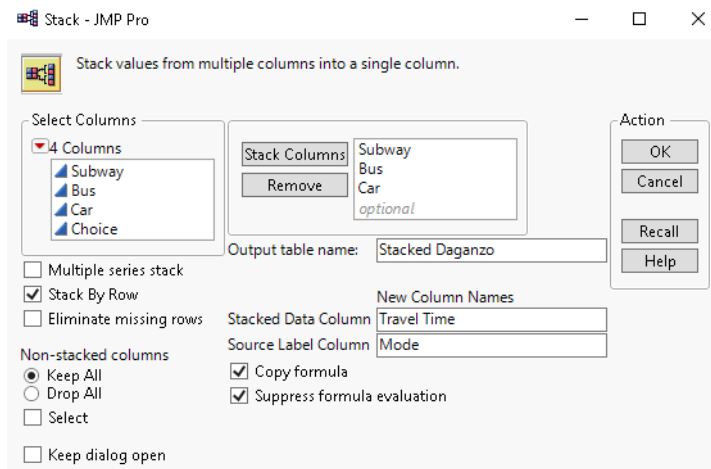
	Subway	Bus	Car	Choice	Choice Mode	Subject
1	16.481	16.196	23.89	2	Bus	1
2	15.123	11.373	14.182	2	Bus	2
3	19.469	8.822	20.819	2	Bus	3
4	18.847	15.649	21.28	2	Bus	4
5	12.578	10.671	18.335	2	Bus	5
6	11.513	20.582	27.838	1	Subway	6

### Stack the Data

In order to construct the Profile data, each alternative needs to be expressed in a separate row.

1. Select **Tables > Stack**.
2. Select Subway, Bus, and Car and click **Stack Columns**.
3. For the Output table name, type Stacked Daganzo. Type Travel Time for the Stacked Data Column and Mode for the Source Label Column.

**Figure 4.46** Stack Dialog for Daganzo Data



4. Click **OK**.

**Figure 4.47** Partial Stacked Daganzo Table

	Choice	Choice Mode	Subject	Mode	Travel Time
1	2	Bus	1	Subway	16.481
2	2	Bus	1	Bus	16.196
3	2	Bus	1	Car	23.89
4	2	Bus	2	Subway	15.123
5	2	Bus	2	Bus	11.373
6	2	Bus	2	Car	14.182
7	2	Bus	3	Subway	19.469

**Make the Profile Data Table**

For the Profile Data Table, you need the Subject, Mode, and Travel Time columns.

1. Select the Subject, Mode, and Travel Time columns and select **Tables > Subset**.
2. Select **All Rows** and **Selected Columns** and click **OK**.

A partial data table is shown in Figure 4.48. Note the default table name is Subset of Stacked Daganzo.

**Figure 4.48** Partial Subset Table of Stacked Daganzo Data

	Subject	Mode	Travel Time
1	1	Subway	16.481
2	1	Bus	16.196
3	1	Car	23.89
4	2	Subway	15.123
5	2	Bus	11.373
6	2	Car	14.182
7	3	Subway	19.469

**Make the Response Data Table**

For the Response Data Table, you need the Subject and Choice Mode columns, but you also need a column for each possible choice.

3. From the Daganzo Trip.jmp data, select the Subject and Choice Mode columns.
4. Select **Tables > Subset**.
5. Select **All Rows** and **Selected Columns** and click **OK**.  
Note that the default table name is Subset of Daganzo Trip.
6. Select **Cols > New Columns**.
7. For the Column prefix, type Choice.
8. Select **Data Type>Character**.
9. Enter 3 for the Number of columns to add.
10. Click **OK**.

The columns Choice 1, Choice 2, and Choice 3 have been added.



11. Type "Bus" (without quotation marks) in the first row of Choice 1. Right-click the cell and select **Fill > Fill to end of table**.
12. Type "Subway" (without quotation marks) in the first row of Choice 2. Right-click the cell and select **Fill > Fill to end of table**.
13. Type "Car" (without quotation marks) in the first row of Choice 3. Right-click the cell and select **Fill > Fill to end of table**.

The resulting table is shown in Figure 4.49.

**Figure 4.49** Partial Subset Table of Daganzo Data with Choice Set

	Choice Mode	Subject	Choice 1	Choice 2	Choice 3
1	Bus	1	Bus	Subway	Car
2	Bus	2	Bus	Subway	Car
3	Bus	3	Bus	Subway	Car
4	Bus	4	Bus	Subway	Car
5	Bus	5	Bus	Subway	Car
6	Subway	6	Bus	Subway	Car
7	Subway	7	Bus	Subway	Car

### Fit the Model

Now that you have separated the original Daganzo Trip.jmp table into two separate tables, you can run the Choice Platform.

1. Select **Analyze > Consumer Research > Choice**.
2. From the Data Format list, select **Multiple Tables, Cross-Referenced**.
3. Specify the model, as shown in Figure 4.50.

**Figure 4.50** Choice Dialog Box for Subset of Daganzo Data

Data Format: Multiple Tables, Cross-Referenced

### Profile Data

Select Data Table: Stacked Daganzo

Select Columns:

- Choice
- Choice Mode
- Subject
- Mode
- Travel Time

Pick Role Variables:

Profile ID: Mode

Grouping: Subject (optional)

Run Model

Help

Remove

☒ Firth Bias-Adjusted Estimates

☐ Hierarchical Bayes

Number of Bayesian Iterations: 5000

Construct Profile Effects:

Add: Travel Time

Cross

Nest

Macros

Degree: 2

Transform:

### Response Data

Select Data Table: Subset of Daganzo Trip

Select Columns:

- Choice Mode
- Subject
- Choice 1
- Choice 2
- Choice 3

Pick Role Variables:

Profile ID Chosen: Choice Mode

Profile ID Choices: Choice 1, Choice 2, Choice 3 (optional)

Grouping: Subject (optional)

Subject ID: optional

Freq: optional numeric

Weight: optional numeric

By: optional

☐ Respondent is allowed to select "None" or "No Choice"

4. Click **Run Model**.

The resulting parameter estimate now expresses the utility coefficient for Travel Time and is shown in Figure 4.51.

**Figure 4.51** Parameter Estimate for Travel Time of Daganzo Data

Choice Model			
Parameter Estimates			
Term	Estimate	Std Error	
Travel Time	-0.341768586	0.0745222259	
AICc	68.766653		
BIC	70.595342		
-2*LogLikelihood	66.683319		
-2*Firth LogLikelihood	61.490004		
Converged in Gradient			
Firth Bias-Adjusted Estimates			
Likelihood Ratio Tests			
Source	ChiSquare	DF	Prob>ChiSq
Travel Time	43.178	1	<.0001*

The negative coefficient implies that increased travel time has a negative effect on consumer utility or satisfaction. The likelihood ratio test result indicates that the Choice model with the effect of Travel Time is significant.

## Example of Transforming Data to One Analysis Table

Rather than creating two or three tables, it can be more practical to transform the data so that only one table is used. For the one-table format, the subject effect is added as in the previous example. A response indicator column is added instead of using three different columns for the choice sets (Choice 1, Choice 2, Choice 3). The transformation for the one-table scenario includes the following steps.

1. Create or open Stacked Daganzo.jmp from the “[Stack the Data](#)” steps shown in “[Example of Transforming Data to Two Analysis Tables](#)” on page 118.
2. Select **Cols > New Columns**.
3. Type Response as the Column Name.
4. Click **Column Properties** and select **Formula**.
5. Select **Conditional** in the functions list and then select **If**.
6. Select the column Choice Mode for the expression (expr).
7. Enter “=” and select Mode.
8. Type 1 for the **Then Clause** and 0 for the **Else Clause**.
9. Click **OK** in the Formula Editor window. Click **OK** in the New Column window.

The completed formula should look like Figure 4.52.

Figure 4.52 Formula for Response Indicator for Stacked Daganzo Data

If 

Choice Mode == Mode

= 1

else

= 0

10. Select the Subject, Travel Time, and Response columns and then select **Tables > Subset**.
11. Select **All Rows** and **Selected Columns** and click **OK**.

A partial listing of the new data table is shown in Figure 4.53.

Figure 4.53 Partial Table of Stacked Daganzo Data Subset

	Subject	Travel Time	Response
1	1	16.481	0
2	1	16.196	1
3	1	23.89	0
4	2	15.123	0
5	2	11.373	1
6	2	14.182	0
7	3	19.469	0

12. Select **Analyze > Consumer Research > Choice** to open the launch window and specify the model as shown in Figure 4.54.

Figure 4.54 Choice Dialog Box for Subset of Stacked Daganzo Data for One-Table Analysis

Data Format

One Table, Stacked

Select Data Table

Subset of Stacked Daganzo

Select Columns

Subject

Travel Time

Response

Pick Role Variables

Response Indicator

Response

Subject ID

required

Choice Set ID

required

Grouping

Subject

optional

Construct Profile Effects

Add

Cross

Nest

Macros

Degree

2

Transform

Travel Time

Construct Subject Effects (Optional)

Add

Cross

Nest

Macros

Degree

2

Transform

Run Model

Help

Remove

☒ Firth Bias-Adjusted Estimates

☐ Hierarchical Bayes

Number of Bayesian Iterations

5000

☐ Respondent is allowed to select "None" or "No Choice"

13. Click **Run Model**.

**Figure 4.55** Parameter Estimate for Travel Time of Daganzo Data from One-Table Analysis

Choice Model			
Parameter Estimates			
Term	Estimate	Std Error	
Travel Time	-0.341768586	0.0745222259	
AICc	68.766653		
BIC	70.595342		
-2*LogLikelihood	66.683319		
-2*Firth LogLikelihood	61.490004		
Converged in Gradient			
Firth Bias-Adjusted Estimates			
Likelihood Ratio Tests			
Source	ChiSquare	DF	Prob>ChiSq
Travel Time	43.178	1	<.0001*

Notice that the result is identical to that obtained for the two-table model, shown earlier in Figure 4.51.

This chapter illustrates the use of the Choice Modeling platform with simple examples. This platform can also be used for more complex models, such as those involving more complicated transformations and interaction terms.

## Statistical Details for the Choice Platform

- “Special Data Table Rules”
- “Utility and Probabilities”
- “Gradients”

### Special Data Table Rules

#### Default Choice Set

If in every trial, you can choose any of the response profiles, you can omit the **Profile ID Choices** selection under **Pick Role Variables** in the Response Data section of the Choice launch window. The Choice Model platform then assumes that all choice profiles are available on each run.

#### Subject Data with Response Data

If you have subject data in the Response data table, select this table as the **Select Data Table** under the Subject Data. In this case, a **Subject ID** column does not need to be specified. In fact, it is not used. It is generally assumed that the subject data repeats consistently in multiple runs for each subject.

## Logistic Regression

Ordinary logistic regression can be performed with the Choice Modeling platform.

---

**Note:** The Fit Y by X and Fit Model platforms are more convenient to use than the Choice Modeling platform for logistic regression modeling. This section is used only to demonstrate that the Choice Modeling platform can be used for logistic regression, if desired.

---

If your data are already in the choice-model format, you might want to use the steps given below for logistic regression analysis. However, three steps are needed:

- Create a trivial Profile data table with a row for each response level.
- Put the explanatory variables into the Response data.
- Specify the Response data table, again, for the Subject data table.

For examples of conducting Logistic Regression using the Choice Platform, see [“Example of Logistic Regression Using the Choice Platform”](#) on page 113 and [“Example of Logistic Regression for Matched Case-Control Studies”](#) on page 116.

## Utility and Probabilities

Parameter estimates from the choice model identify consumer *utility*, or marginal utilities in the case of a linear utility function. Utility is the level of satisfaction consumers receive from products with specific attributes and is determined from the parameter estimates in the model.

The choice statistical model is expressed as follows:

Let  $X[k]$  represent a subject attribute design row, with intercept

Let  $Z[j]$  represent a choice attribute design row, without intercept

Then, the probability of a given choice for the  $k$ 'th subject to the  $j$ 'th choice of  $m$  choices is:

$$P_{i,jk} = \frac{\exp(\beta'(X[k] \otimes Z[j]))}{\sum_{l=1}^m \exp(\beta'(X[k] \otimes Z[l]))}$$

where:

- $\otimes$  is the Kronecker rowwise product
- the numerator calculates for the  $j$ 'th alternative actually chosen
- the denominator sums over the  $m$  choices presented to the subject for that trial

## Gradients

The gradient values that you obtain when you select the Save Gradients by Subject option are the subject-aggregated Newton-Raphson steps from the optimization used to produce the estimates. At the estimates, the total gradient is zero, and  $\Delta = H^{-1}g = 0$ , where  $g$  is the total gradient of the log-likelihood evaluated at the MLE, and  $H^{-1}$  is the inverse Hessian function or the inverse of the negative of the second partial derivative of the log-likelihood.

But, the disaggregation of  $\Delta$  results in the following:

$$\Delta = \sum_{ij} \Delta_{ij} = \sum H^{-1} g_{ij} = 0,$$

Here  $i$  is the subject index,  $j$  is the choice response index for each subject,  $\Delta_{ij}$  are the partial Newton-Raphson steps for each run, and  $g_{ij}$  is the gradient of the log-likelihood by run.

The mean gradient step for each subject is then calculated as follows:

$$\bar{\Delta}_i = \sum_j \frac{\Delta_{ij}}{n_i},$$

where  $n_i$  is the number of runs per subject. The  $\bar{\Delta}_i$  are related to the force that subject  $i$  is applying to the parameters. If groups of subjects have truly different preference structures, these forces are strong, and they can be used to cluster the subjects. The  $\bar{\Delta}_i$  are the gradient forces that are saved. You can then cluster these values using the Clustering platform.





Use MaxDiff (maximum difference scaling) as an alternative to standard preference scales to determine the relative importance of items being rated. MaxDiff forces respondents to report their most and least preferred options. This often results in rankings that are more definitive than rankings obtained using standard preference scales.

The MaxDiff platform enables you to do the following:

- Use information about respondent (subject) traits as well as product attributes.
- Integrate data from one, two, or three sources.
- Obtain subject-level scores for segmenting or clustering your data.
- **JMP PRO** Estimate subject-specific coefficients using a Bayesian approach.
- Use bias-corrected maximum likelihood estimators (Firth 1993).

Figure 5.1 MaxDiff All Comparisons Report

All Levels Comparison Report										
Difference (Row-Column) Standard Error of Difference Wald p-Value	Gyro	Sour Cream and Onion	Truffle Fries	Biscuits and Gravy	Reuben	Ketchup	All Dressed	Dill Pickle	Barbecue	Southern Barbecue
Gyro	0	-1.1747 0.33327 0.00062	-1.3082 0.35477 0.00035	-1.2688 0.34456 0.00036	-0.6992 0.34758 0.04668	-0.8229 0.32802 0.01356	-0.9467 0.33571 0.00569	-1.027 0.35251 0.00432	-2.4417 0.41499 4.32e-6	-1.9535 0.38939 2.01e-6
Sour Cream and Onion	1.17469 0.33327 0.00062	0	-0.1335 0.31379 0.67137	-0.0941 0.30716 0.75983	0.47548 0.30932 0.12709	0.35175 0.31915 0.27279	0.22804 0.28158 0.41975	0.14766 0.30707 0.63156	-1.267 0.37654 0.00105	-0.7788 0.34896 0.02764
Truffle Fries	1.30817 0.35477 0.00035	0.13348 0.31379 0.67137	0	0.03935 0.31345 0.90032	0.60896 0.31817 0.05821	0.48523 0.31157 0.12223	0.36152 0.31136 0.24809	0.28114 0.30871 0.36444	-1.1335 0.39213 0.00462	-0.6453 0.34904 0.06715
Biscuits and Gravy	1.26882 0.34456 0.00036	0.09413 0.30716 0.75983	-0.0393 0.31345 0.90032	0	0.56961 0.30639 0.06567	0.44588 0.29752 0.1368	0.32217 0.29974 0.28477	0.24179 0.28249 0.39388	-1.1729 0.36267 0.00161	-0.6847 0.34038 0.0467
Reuben	0.69921 0.34758 0.04668	-0.4755 0.30932 0.12709	-0.609 0.31817 0.05821	-0.5696 0.30639 0.06567	0	-0.1237 0.30893 0.68956	-0.2474 0.30148 0.41355	-0.3278 0.30328 0.28208	-1.7425 0.37117 7.69e-6	-1.2543 0.35324 0.00056
Ketchup	0.82294 0.32802 0.01356	-0.3517 0.31915 0.27279	-0.4852 0.31157 0.12223	-0.4459 0.29752 0.1368	0.12373 0.30893 0.68956	0	-0.1237 0.29728 0.67812	-0.2041 0.27938 0.46661	-1.6188 0.36911 2.64e-5	-1.1305 0.34808 0.00154
All Dressed	0.94665 0.33571 0.00569	-0.228 0.28158 0.41975	-0.3615 0.31136 0.24809	-0.3222 0.29974 0.28477	0.24744 0.30148 0.41355	0.12371 0.29728 0.67812	0	-0.0804 0.29447 0.78538	-1.4951 0.36838 9.22e-5	-1.0068 0.3418 0.00393
Dill Pickle	1.02703 0.35251 0.00432	-0.1477 0.30707 0.63156	-0.2811 0.30871 0.36444	-0.2418 0.28249 0.39388	0.32782 0.30328 0.28208	0.20409 0.27938 0.46661	0.08038 0.29447 0.78538	0	-1.4147 0.3539 0.00012	-0.9264 0.35532 0.01038
Barbecue	2.44172 0.41499 4.32e-6	1.26703 0.37654 0.00105	1.13355 0.39213 0.00462	1.1729 0.36267 0.00161	1.7425 0.37117 7.69e-6	1.61877 0.36911 2.64e-5	1.49507 0.36838 9.22e-5	1.41468 0.3539 0.00612	0	0.48824 0.39164 0.21514
Southern Barbecue	1.95348 0.38939 2.01e-6	0.77879 0.34896 0.02764	0.6453 0.34904 0.06715	0.68465 0.34038 0.0467	1.25426 0.35324 0.00056	1.13053 0.34808 0.00154	1.00683 0.3418 0.00393	0.92644 0.35532 0.01038	-0.4882 0.39164 0.21514	0

## Contents

Overview of the MaxDiff Modeling Platform.....	131
Examples of the MaxDiff Platform.....	131
One Table Format .....	132
Multiple Table Format .....	134
Launch the MaxDiff Platform .....	138
Launch Window for One Table, Stacked .....	139
Launch Window for Multiple Tables, Cross-referenced .....	140
MaxDiff Model Report.....	144
Effect Summary .....	145
MaxDiff Results.....	146
Parameter Estimates .....	147
Bayesian Parameter Estimates.....	149
Likelihood Ratio Tests .....	150
MaxDiff Platform Options .....	151
Comparisons Report.....	153
Save Bayes Chain.....	153

---

## Overview of the MaxDiff Modeling Platform

MaxDiff, also known as *best-worst scaling* (BWS), is a choice-based measurement method. Rather than asking a respondent to report one favorite choice among several alternative profiles, MaxDiff asks a respondent to report both a *best* and a *worst* choice. The MaxDiff approach can provide more information about preferences than an approach where a respondent reports only a favorite choice. For background on MaxDiff studies, see Louviere et al. (2015). For background on choice modeling, see Louviere et al. (2015), Train (2009), and Rossi et al. (2005).

MaxDiff analysis uses the framework of random utility theory. A choice is assumed to have an underlying value, or *utility*, to respondents. The MaxDiff platform estimates these utilities. The MaxDiff platform also estimates the probabilities that a choice is preferred over other choices. This is done using conditional logistic regression. See McFadden (1974).

---

**Note:** One-factor MaxDiff studies can be designed using the MaxDiff Design platform. See the MaxDiff Design chapter in the *Design of Experiments Guide*.

---

### Segmentation and Bayesian Subject-Level Effects

Market researchers sometimes want to analyze the preference structure for each subject separately in order to see whether there are groups of subjects that behave differently. If there are sufficient data, you can specify “By groups” in the Response Data or you could introduce a Subject identifier as a subject-side model term. This approach, however, is costly if the number of subjects is large. Other segmentation techniques discussed in the literature include Bayesian and mixture methods.

If there are not sufficient data to specify “By groups,” you can segment in JMP by clustering subjects using response data and the **Save Gradients by Subject** option. The option creates a new data table containing the average Hessian-scaled gradient on each parameter for each subject. For an example, see “[Example of Segmentation](#)” on page 109 in the “Choice Models” chapter. For details about the gradient values, see “[Gradients](#)” on page 127 in the “Choice Models” chapter.

 MaxDiff also provides a Hierarchical Bayesian approach to estimating subject-level effects. This approach can be useful in market segmentation.

---

## Examples of the MaxDiff Platform

Thirty respondents participated in a MaxDiff study to compare seven flavors of potato chips. Each choice set consisted of three profiles (potato chip flavors). For each choice set, a

respondent's favorite choice was recorded as 1 and his or her least favorite choice was recorded as -1. Intermediate choices were recorded as 0.

The MaxDiff platform can analyze data that is presented in a one-table format or in a multiple-table format. In the multiple table format, information about responses, choice sets, and subjects is saved in different data tables. In the one-table format, that information is contained in a single data table.

- “[One Table Format](#)” on page 132 shows how to analyze a subset of the available data in a one-table format. Note that you could add additional profile and subject data to the single table for a more complete analysis.
- “[Multiple Table Format](#)” on page 134 shows how to bring together information from different tables into one MaxDiff analysis.

## One Table Format

1. Select **Help > Sample Data Library** and open **Potato Chip Combined.jmp**.
2. Select **Analyze > Consumer Research > MaxDiff**.  
Note that the default Data Format is set to One Table, Stacked.
3. Click **Select Data Table**.
4. Select **Potato Chip Combined.jmp** and click **OK**.
5. Assign roles to columns as follows. The completed launch dialog is shown in Figure 5.2.
  - Select **Response** and click **Response Indicator**.
  - Select **Respondent** and click **Subject ID**.
  - Select **Choice Set ID** and click **Choice Set ID**.
  - Select **ProfileID** and click **Add** in the Construct Profile Effects panel.

**Figure 5.2** Completed MaxDiff Launch Window

Data Format: One Table, Stacked

Select Data Table: Potato Chip Combined

Select Columns:

- Respondent
- Survey ID
- Choice Set ID
- Profile ID
- Response

Pick Role Variables:

Response Indicator: Response

Subject ID: Respondent

Choice Set ID: Choice Set ID

Grouping: optional

Construct Profile Effects:

Add: Profile ID

Cross

Nest

Macros

Degree: 2

Transform

Construct Subject Effects (Optional):

Add

Cross

Nest

Macros

Degree: 2

Transform

Run Model

Help

Remove

☒ Firth Bias-Adjusted Estimates

☐ Hierarchical Bayes

Number of Bayesian Iterations: 5000

Best: 1

Worst: -1

Note that the setting for Worst choice changed to -1 when you specified the Response column as the Response Indicator variable.

- Click **Run Model**.

**Figure 5.3** MaxDiff Report for Potato Chip Combined.jmp

MaxDiff Model

MaxDiff Results

Marginal Utility	Marginal Probability	Profile ID
1.2774	0.2895	Barbecue
0.7892	0.1777	Southern Barbecue
0.1439	0.0932	Truffle Fries
0.1046	0.0896	Biscuits and Gravy
0.0104	0.0815	Sour Cream and Onion
-0.137	0.0703	Dill Pickle
-0.218	0.0649	All Dressed
-0.341	0.0574	Ketchup
-0.465	0.0507	Reuben
-1.164	0.0252	Gyro

Parameter Estimates

Likelihood Ratio Tests

Source	ChiSquare	DF	Prob>ChiSq
Profile ID	78.317	9	<.0001*

The report indicates that Profile ID is significant, indicating that preferences for the various chip types differ significantly. The highest Marginal Utility is for Barbecue chips. The estimated probability that Barbecue chips are preferred to other chip types is 0.2895.

- Click the red triangle next to MaxDiff Model and select **All Levels Comparison Report**.

**Figure 5.4** All Comparisons Report

All Levels Comparison Report										
Difference (Row-Column) Standard Error of Difference Wald p-Value	Gyro	Sour Cream and Onion	Truffle Fries	Biscuits and Gravy	Reuben	Ketchup	All Dressed	Dill Pickle	Barbecue	Southern Barbecue
Gyro	0	-1.1747	-1.3082	-1.2688	-0.6992	-0.8229	-0.9467	-1.027	-2.4417	-1.9535
	0	0.33327	0.35477	0.34456	0.34758	0.32802	0.33571	0.35251	0.41499	0.38939
		0.00062	0.00035	0.00036	0.04668	0.01356	0.00569	0.00432	4.32e-8	2.01e-6
Sour Cream and Onion	1.17469	0	-0.1335	-0.0941	0.47548	0.35175	0.22804	0.14766	-1.267	-0.7788
	0.33327	0	0.31379	0.30716	0.30932	0.31915	0.28158	0.30707	0.37654	0.34896
	0.00062		0.67137	0.75983	0.12709	0.27279	0.41975	0.63156	0.00105	0.02764
Truffle Fries	1.30817	0.13348	0	0.03935	0.60896	0.48523	0.36152	0.28114	-1.1335	-0.6453
	0.35477	0.31379	0	0.31345	0.31817	0.31157	0.31136	0.30871	0.39213	0.34904
	0.00035	0.67137		0.90032	0.05821	0.12223	0.24809	0.36444	0.00462	0.06715
Biscuits and Gravy	1.26882	0.09413	-0.0393	0	0.56961	0.44588	0.32217	0.24179	-1.1729	-0.6847
	0.34456	0.30716	0.31345	0	0.30639	0.29752	0.29974	0.28249	0.36267	0.34038
	0.00036	0.75983	0.90032		0.06567	0.1368	0.28477	0.39388	0.00161	0.0467
Reuben	0.69921	-0.4755	-0.609	-0.5696	0	-0.1237	-0.2474	-0.3278	-1.7425	-1.2543
	0.34758	0.30932	0.31817	0.30639	0	0.30893	0.30148	0.30328	0.37117	0.35324
	0.04668	0.12709	0.05821	0.06567		0.68956	0.41355	0.28208	7.69e-6	0.00056
Ketchup	0.82294	-0.3517	-0.4852	-0.4459	0.12373	0	-0.1237	-0.2041	-1.6188	-1.1305
	0.32802	0.31915	0.31157	0.29752	0.30893	0	0.29728	0.27938	0.36911	0.34808
	0.01356	0.27279	0.12223	0.1368	0.68956		0.67812	0.46661	2.64e-5	0.00154
All Dressed	0.94665	-0.228	-0.3615	-0.3222	0.24744	0.12371	0	-0.0804	-1.4951	-1.0068
	0.33571	0.28158	0.31136	0.29974	0.30148	0.29728	0	0.29447	0.36838	0.3418
	0.00569	0.41975	0.24809	0.28477	0.41355	0.67812		0.78538	9.22e-5	0.00393
Dill Pickle	1.02703	-0.1477	-0.2811	-0.2418	0.32782	0.20409	0.08038	0	-1.4147	-0.9264
	0.35251	0.30707	0.30871	0.28249	0.30328	0.27938	0.29447	0	0.3539	0.35532
	0.00432	0.63156	0.36444	0.39388	0.28208	0.46661	0.78538		0.00012	0.01038
Barbecue	2.44172	1.26703	1.13355	1.1729	1.7425	1.61877	1.49507	1.41468	0	0.48824
	0.41499	0.37654	0.39213	0.36267	0.37117	0.36911	0.36838	0.3539	0	0.39164
	4.32e-8	0.00105	0.00462	0.00161	7.69e-6	2.64e-5	9.22e-5	0.00012		0.21514
Southern Barbecue	1.95348	0.77879	0.6453	0.68465	1.25426	1.13053	1.00683	0.92644	-0.4882	0
	0.38939	0.34896	0.34904	0.34038	0.35324	0.34808	0.3418	0.35532	0.39164	0
	2.01e-6	0.02764	0.06715	0.0467	0.00056	0.00154	0.00393	0.01038	0.21514	

Each comparison is the difference in estimated utilities between the chip type labeling the row and the chip type labeling the column. Small  $p$ -values are colored with an intense blue or red color, depending on the sign of the difference. For example, based on the blue colors across the Gyro row, you can see that Gyro chips have significantly lower utility than all other chip types. Barbecue chips have higher utility than all other chip types, though they do not differ significantly from Southern Barbecue chips.

**Note:** Because the All Comparisons Report  $p$ -values are not corrected for multiple comparisons, use them as a guide.

## Multiple Table Format

This version of the potato chip study uses three data tables: Potato Chip Profiles.jmp, Potato Chip Responses.jmp, and Potato Chip Subjects.jmp. Although you can always arrange your data into a single table, a multi-table approach can be more convenient than a one-table

analysis when you have additional profile and subject variables that you want to include in your analysis.

## Complete the Launch Window

1. Select **Help > Sample Data Library** and open the Potato Chip Responses.jmp sample data table.

---

**Note:** If you prefer not to follow the steps for completing the launch window, click the green triangle next to the **MaxDiff for Flavor** script. Then proceed to [“Explore the Model”](#) on page 137.

---

2. Click the green triangle next to the **Open Profile and Subject Tables** script.
  - The profile data table, Potato Chip Profiles.jmp, lists all the potato chip types in the study (Flavor) along with information on the country of origin (Product Of). Each choice has a Profile ID.
  - The subjects data table, Potato Chip Subjects.jmp, lists the respondents. It also gives additional information about each respondent: Citizenship and Gender.
  - The responses data table, Potato Chip Responses.jmp, lists the respondents. For each respondent, the Survey ID and Choice Set ID for each set of profiles is listed, along with the Profile ID values for each choice set. The table also contains response data in the Best Profile and Worst Profile columns.
3. From any of the three data tables, select **Analyze > Consumer Research > MaxDiff**.
4. From the Data Format list, select Multiple Tables, Cross-referenced.

There are three separate outlines, one for each of the data sources.
5. Click **Select Data Table** under Profile Data.

A Profile Data Table window appears, which prompts you to specify the data table for the profile data.
6. Select Potato Chip Profiles.jmp and click **OK**.

The columns from this table appear in the **Select Columns**.
7. Select Profile ID from the Select Columns list and click **Profile ID** under **Pick Role Variables**.
8. Select Flavor and click **Add** under **Construct Model Effects**.

Note that Product Of is another profile effect that you could add to the effects list.

**Figure 5.5** Complete Profile Data Outline

Data Format: Multiple Tables, Cross-Referenced

**Profile Data**

Select Data Table: Potato Chip Profiles

Select Columns:

- Profile ID
- Flavor
- Product Of

Pick Role Variables:

Profile ID: Profile ID

Grouping: optional

Run Model

Help

Remove

☒ Firth Bias-Adjusted Estimates

☐ Hierarchical Bayes

Number of Bayesian Iterations: 5000

Construct Profile Effects:

Add

Cross

Nest

Macros

Degree: 2

Transform: optional

Flavor

9. Open the Response Data outline. Click **Select Data Table**.
10. Select Potato Chip Responses.jmp and click **OK**.
11. Assign roles to columns as follows. The completed launch dialog is shown in Figure 5.6.
  - Select Best Profile and click **Best Choice**.
  - Select Worst Profile and click **Worst Choice**.
  - Select Choice 1, Choice 2, and Choice 3 and click **Profile ID Choices**.
  - Select Respondent and click **Subject ID**.

**Figure 5.6** Completed Response Data Outline

Data Format: Multiple Tables, Cross-Referenced

**Response Data**

Select Data Table: Potato Chip Responses

Select Columns:

- Respondent
- Survey ID
- Choice Set ID
- Choice 1
- Choice 2
- Choice 3
- Best Profile
- Worst Profile

Pick Role Variables:

Best Choice: Best Profile

Worst Choice: Worst Profile

Profile ID Choices: Choice 1, Choice 2, Choice 3

Grouping: optional

Subject ID: Respondent

Freq: optional numeric

Weight: optional numeric

By: optional

Run Model

Help

Remove

☒ Firth Bias-Adjusted Estimates

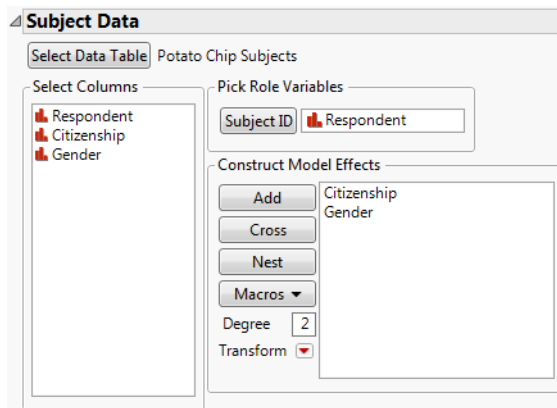
☐ Hierarchical Bayes

Number of Bayesian Iterations: 5000



12. Open the Subject Data outline. Click **Select Data Table**.
13. Select Potato Chip Subjects.jmp and click **OK**.
14. Select Respondent and click **Subject ID**.
15. Select Citizenship and Gender and click **Add** under **Construct Model Effects**.

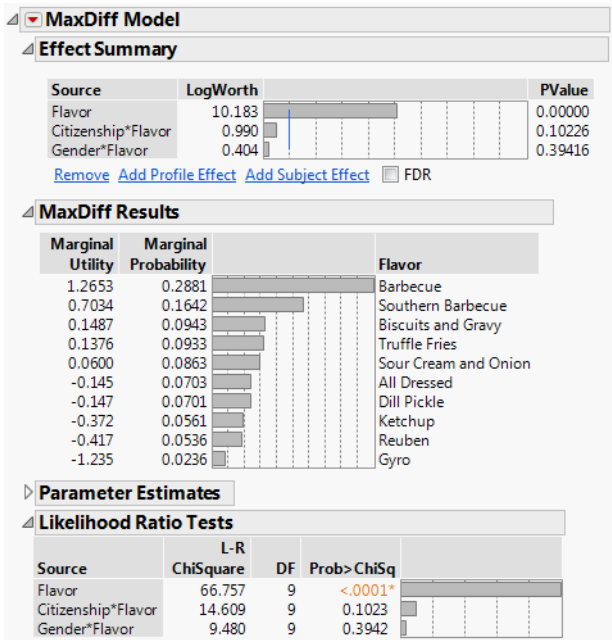
**Figure 5.7** Completed Subject Data Outline



## Explore the Model

1. Click **Run Model**.

Figure 5.8 MaxDiff Model Report



The Effect Summary report in Figure 5.8 shows the terms in the model and gives *p*-values for their significance. Notice that Flavor is a profile effect, and that each of Citizenship\*Flavor and Gender\*Flavor is an interaction of a subject and a profile effect.

The Likelihood Ratio Tests report indicates that Flavor is significant.

# Launch the MaxDiff Platform

Launch the MaxDiff platform by selecting **Analyze > Consumer Research > MaxDiff**.

Your data for the MaxDiff platform can be combined in a single data table or it can reside in two or three separate data tables. When the Choice window opens, specify whether you are using one or several data tables by selecting from the Data Format list.

## One Table, Stacked

For this format, the data are combined into a single data table with a row for every profile presented to a subject and an indicator for the best and worst choices in that profile.

For an example of data in the one-table format, see “[One Table Format](#)” on page 132. For details, see “[Launch Window for One Table, Stacked](#)” on page 139.

## Multiple Tables, Cross-referenced

Your data can reside in two or three separate tables: a Profile Data and Response Data table are required, and a Subject Data table is optional. The MaxDiff Launch Window contains three sections, each corresponding to a different data table. You can expand or collapse each section of the launch window, as needed.

For an example of data in the multiple-tables format, see [“Multiple Table Format”](#) on page 134. For details, see [“Launch Window for Multiple Tables, Cross-referenced”](#) on page 140.

## Launch Window for One Table, Stacked

Figure 5.9 shows the one-table launch window populated using Potato Chip Combined.jmp.

**Figure 5.9** Launch Window for One Table, Stacked Data Format

**Select Data Table** Select or open the data table that contains the combined data. Select Other to open a file that is not already open.

**Response Indicator** A column containing the preference data. Use two of the values 1, -1, and 0 for the Best and Worst choices, and the third value for profiles that are not Best or Worst. Unless you specify a different coding using the menus next to Best and Worst in the lower left portion of the window, a 1 will indicate the Best choice and a -1 the Worst choice.

**Subject ID** An identifier for the study participant.

**Choice Set ID** An identifier for the set of profiles presented to the subject for a given preference determination.

**Grouping** A column which, when used with the Choice Set ID, uniquely designates each choice set. For example, if a choice set has Choice Set ID = 1 for Survey = A, and another choice set has Choice Set ID = 1 for Survey = B, then Survey should be used as a **Grouping** column.

**Construct Profile Effects** Add effects constructed from the attributes for the profiles.

For information about the Construct Profile Effects panel, see the Construct Model Effects section in the Model Specification chapter of the *Fitting Linear Models* book.

**Construct Subject Effects (Optional)** Add effects constructed from subject-related factors.

For information about the Construct Subject Effects panel, see the Construct Model Effects section in the Model Specification chapter of the *Fitting Linear Models* book.

**Firth Bias-adjusted Estimates** Computes bias-corrected MLEs that produce better estimates and tests than MLEs without bias correction. These estimates also improve separation problems that tend to occur in logistic-type models. Refer to Heinze and Schemper (2002) for a discussion of the separation problem in logistic regression.



**Hierarchical Bayes** Uses a Bayesian approach to estimate subject-specific parameters. See “Bayesian Parameter Estimates” on page 149.



**Number of Bayesian Iterations** (Applicable only if Hierarchical Bayes is selected.) The total number of iterations of the adaptive Bayes algorithm used to estimate subject effects. This number includes a burn-in period of iterations that are discarded. The number of burn-in iterations is equal to half of the Number of Bayesian Iterations specified on the launch window.

## Launch Window for Multiple Tables, Cross-referenced

Figure 5.10 shows the multiple-table launch window, with the Profile Data outline populated using Potato Chip Profile.jmp.

**Figure 5.10** Launch Window for Multiple Tables, Cross-Referenced Data Format

**Profile Data**

Select Data Table: Potato Chip Profiles

Select Columns:

- Profile ID
- Flavor
- Product Of

Pick Role Variables:

Profile ID: Profile ID

Grouping: optional

Run Model: Run Model, Help, Remove

☒ Firth Bias-Adjusted Estimates

☐ Hierarchical Bayes

Number of Bayesian Iterations: 5000

Construct Profile Effects:

Add, Cross, Nest, Macros, Degree: 2, Transform

Flavor

Response Data

Subject Data

In the case of Multiple Tables, Cross-referenced, the launch window has three sections:

- “Profile Data” on page 141
- “Response Data” on page 142
- “Subject Data” on page 143

## Profile Data

The profile data table describes the attributes associated with each choice. Each choice can comprise many different attributes, and each attribute is listed as a column in the data table. There is a row for each possible choice, and each possible choice contains a unique ID.

**Select Data Table** Select or open the data table that contains the profile data. Select Other to open a file that is not already open.

**Profile ID** Identifier for each row of choice combinations. If the **Profile ID** column does not uniquely identify each row in the profile data table, you need to add **Grouping** columns. Add **Grouping** columns until the combination of **Grouping** and **Profile ID** columns uniquely identifies the row, or profile.

**Grouping** A column which, when used with the Choice Set ID column, uniquely designates each choice set. For example, if **Profile ID** = 1 for Survey = A, and a different **Profile ID** = 1 for Survey = B, then Survey would be used as a **Grouping** column.

**Construct Profile Effects** Add effects constructed from the attributes in the profiles.

For information about the Construct Profile Effects panel, see the Construct Model Effects section in the Model Specification chapter of the *Fitting Linear Models* book.

**Firth Bias-adjusted Estimates** Computes bias-corrected MLEs that produce better estimates and tests than MLEs without bias correction. These estimates also improve separation problems that tend to occur in logistic-type models. Refer to Heinze and Schemper (2002) for a discussion of the separation problem in logistic regression.

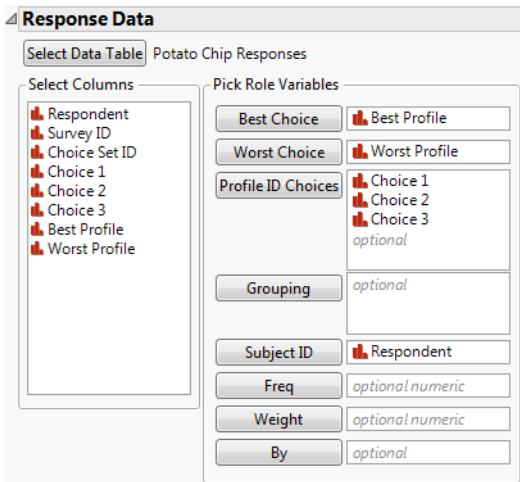
**JMP PRO Hierarchical Bayes** Uses a Bayesian approach to estimate subject-specific parameters. See “Bayesian Parameter Estimates” on page 149.

**JMP PRO Number of Bayesian Iterations** (Applicable only if Hierarchical Bayes is selected.) The total number of iterations of the adaptive Bayes algorithm used to estimate subject effects. This number includes a burn-in period of iterations that are discarded. The number of burn-in iterations is equal to half of the Number of Bayesian Iterations specified on the launch window.

Response Data

Figure 5.11 shows the Response Data outline populated using Potato Chip Responses.jmp.

Figure 5.11 Response Data Outline



The response data table contains the study results. It gives the choice set IDs for each trial as well as the profiles selected as best and worst by the subject. The Response data are linked to the Profile data through the choice set columns and the choice response column. Grouping variables can be used to align choice indices when more than one group is contained within the data.

**Select Data Table** Select or open the data table that contains the profile data. Select Other to open a file that is not already open.

**Best Choice** The Response table column containing the Profile ID of the profile that the subject designated as Best.

**Worst Choice** The Response table column containing the Profile ID of the profile that the subject designated as Worst.

**Profile ID Choices** The columns that contain the **Profile IDs** of the set of possible choices for each choice set.

**Grouping** A column which, when used with the Profile ID Chosen column, uniquely designates each choice set.

**Subject ID** A unique identifier for the study participant.

**Freq** A column containing frequencies. If  $n$  is the value of the Freq variable for a given row, then that row is used in computations  $n$  times. If it is less than 1 or missing, then JMP does not use it to calculate any analyses.

**Weight** A column containing a weight for each observation in the data table. The weight is included in analyses only when its value is greater than zero.

**By** A column whose levels define separate analyses. For each level of the specified column, the corresponding rows are analyzed as a separate analysis on a separate table. The results are presented in separate reports. If more than one By variable is assigned, a separate analysis is produced for each possible combination of the levels of the By variables.

## Subject Data

Figure 5.12 shows the Subject Data outline populated using Potato Chip Subjects.jmp.

**Figure 5.12** Subject Data Outline

---

**Note:** A subject data table is optional, depending on whether subject effects are to be modeled.

---

The subject data table contains the Subject ID and one or more columns of attributes or characteristics for each subject. The subject data table contains the same number of rows as subjects and has an identifier column that matches a similar column in the Response data table.

---

**Note:** You can include subject data in the response data table, but you need to specify subject effects in the Subject Data outline.

---

**Select Data Table** Select or open the data table that contains the subject data. Select Other to open a file that is not already open.

**Subject ID** Unique identifier for the subject.

**Construct Model Effects** Add effects constructed from columns in the subject data table.

For information about the Construct Model Effects panel, see the Construct Model Effects section in the Model Specification chapter of the *Fitting Linear Models* book.

---

## MaxDiff Model Report

The MaxDiff Model window shows some of the following reports by default, depending on your selections in the launch window.

- “Effect Summary”
- “MaxDiff Results”



- “Parameter Estimates”
- “Bayesian Parameter Estimates”
- “Likelihood Ratio Tests”

For descriptions of the platform options, see “MaxDiff Platform Options” on page 151.

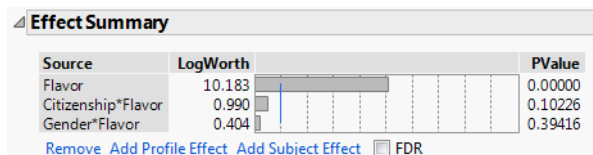
## Effect Summary

The Effect Summary report appears if your model contains more than one effect. It lists the effects estimated by the model and gives a plot of the LogWorth (or FDR LogWorth) values for these effects. The report also provides controls that enable you to add or remove effects from the model. The model fit report updates automatically based on the changes made in the Effects Summary report. For details, see the Effect Summary Report section in the Standard Least Squares Report and Options chapter in the *Fitting Linear Models* book.

The Effect Summary report does not appear when Bayesian Subject Effects is checked in the launch window. This is because likelihood ratio tests are not conducted in this case.

Figure 5.13 shows the Effect Summary report obtained by running the script **MaxDiff for Flavor** in *Potato Chip Responses.jmp*.

**Figure 5.13** Effect Summary Report



### Effect Summary Table Columns

The Effect Summary table contains the following columns:

**Source** Lists the model effects, sorted by ascending  $p$ -values.

**LogWorth** Shows the LogWorth for each model effect, defined as  $-\log_{10}(p\text{-value})$ . This transformation adjusts  $p$ -values to provide an appropriate scale for graphing. A value that exceeds 2 is significant at the 0.01 level (because  $-\log_{10}(0.01) = 2$ ).

**FDR LogWorth** Shows the False Discovery Rate LogWorth for each model effect, defined as  $-\log_{10}(\text{FDR PValue})$ . This is the best statistic for plotting and assessing significance. Select the **FDR** check box to replace the LogWorth column with the **FDR LogWorth** column.

**Bar Graph** Shows a bar graph of the LogWorth (or FDR LogWorth) values. The graph has dashed vertical lines at integer values and a blue reference line at 2.

**PValue** Shows the  $p$ -value for each model effect. This is the  $p$ -value corresponding to the significance test displayed in the Likelihood Ratio Tests report.

**FDR PValue** Shows the False Discovery Rate  $p$ -value for each model effect calculated using the Benjamini-Hochberg technique. This technique adjusts the  $p$ -values to control the false discovery rate for multiple tests. Select the **FDR** check box to replace the **PValue** column with the **FDR PValue** column.

For details about the FDR correction, see Benjamini and Hochberg (1995). For details about the false discovery rate, see the Response Screening chapter in the *Predictive and Specialized Modeling* book or Westfall et al. (2011).

## Effect Summary Table Options

The options below the summary table enable you to add and remove effects:

**Remove** Removes the selected effects from the model. To remove one or more effects, select the rows corresponding to the effects and click the Remove button.

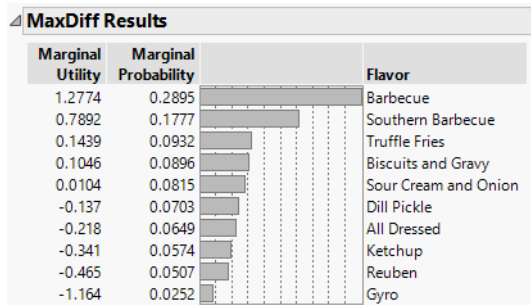
**Add Profile Effect** Opens a panel that contains a list of all columns in the data table for the OneTable, Stacked data format, and for the columns in the Profile Data table for the Multiple Tables, Cross-Referenced data format. Select columns that you want to add to the model, and then click Add below the column selection list to add the columns to the model. Click Close to close the panel.

**Add Subject Effect** Opens a panel that contains a list of all columns in the data table for the OneTable, Stacked data format, and for the columns in the Subject Data table for the Multiple Tables, Cross-Referenced data format. Select columns that you want to add to the model, and then click Add below the column selection list to add the columns to the model. Click Close to close the panel.

## MaxDiff Results

Figure 5.14 shows the MaxDiff Results report obtained by running the script **MaxDiff with No Subject Effects** in Potato Chip Responses.jmp.

**Figure 5.14** MaxDiff Results Report



For each Profile effect specified in the launch window, the following are displayed:

**Marginal Utility** An indicator of the perceived value of the corresponding level of the effect. Larger values suggest that the feature is of greater value.

**Marginal Probability** The estimated probability that a subject expresses a preference for the corresponding level of the effect over all other levels. For each effect, the marginal probabilities sum to one.

**Bar Graph** Shows a bar graph of the marginal probabilities.

**Effect Column** Gives the name of the effect and a list of its levels. The levels define the features to which the Marginal Utility and Marginal Probability estimates apply.

## Parameter Estimates

This report gives details about parameter estimates, fit criteria, and the fitting algorithm.

Figure 5.15 shows the Parameter Estimates report obtained by running the script **MaxDiff for Flavor** in Potato Chip Responses.jmp.

Figure 5.15 Parameter Estimates Report

Parameter Estimates		
Term	Estimate	Std Error
Flavor[All Dressed]	-0.15616234	0.2227718641
Flavor[Barbecue]	1.22210326	0.2978972284
Flavor[Biscuits and Gravy]	0.16398758	0.2249076601
Flavor[Dill Pickle]	-0.17356828	0.2151668801
Flavor[Gyro]	-1.11927509	0.2827016652
Flavor[Ketchup]	-0.47308393	0.2319900823
Flavor[Reuben]	-0.50927309	0.2294979936
Flavor[Sour Cream and Onion]	0.21115573	0.2450361368
Flavor[Southern Barbecue]	0.70149945	0.2695322771
Citizenship[Canadian]*Flavor[All Dressed]	-0.04368106	0.2239442590
Citizenship[Canadian]*Flavor[Barbecue]	-0.16180196	0.2978036763
Citizenship[Canadian]*Flavor[Biscuits and Gravy]	0.05734172	0.2233312021
Citizenship[Canadian]*Flavor[Dill Pickle]	-0.09824391	0.2188896544
Citizenship[Canadian]*Flavor[Gyro]	0.43257276	0.2907051874
Citizenship[Canadian]*Flavor[Ketchup]	-0.38035261	0.2349720398
Citizenship[Canadian]*Flavor[Reuben]	-0.34677939	0.2342643217
Citizenship[Canadian]*Flavor[Sour Cream and Onion]	0.56678250	0.2355493683
Citizenship[Canadian]*Flavor[Southern Barbecue]	-0.00712532	0.2720518372
Gender[Female]*Flavor[All Dressed]	-0.26955535	0.2106170508
Gender[Female]*Flavor[Barbecue]	0.39881410	0.2961368976
Gender[Female]*Flavor[Biscuits and Gravy]	0.09184462	0.2245258406
Gender[Female]*Flavor[Dill Pickle]	-0.11461679	0.2131955199
Gender[Female]*Flavor[Gyro]	-0.40538827	0.2950786212
Gender[Female]*Flavor[Ketchup]	-0.07317652	0.2135446256
Gender[Female]*Flavor[Reuben]	0.20780054	0.2227994637
Gender[Female]*Flavor[Sour Cream and Onion]	0.27353428	0.2337547848
Gender[Female]*Flavor[Southern Barbecue]	0.07570279	0.2627935035
AICc	397.44423	
BIC	456.27172	
-2*LogLikelihood	327.00945	
-2*Firth LogLikelihood	244.39836	
Converged in Gradient		
Firth Bias-Adjusted Estimates		

**Term** Lists the terms in the model.

**Estimate** An estimate of the parameter associated with the corresponding term. In discrete choice experiments, parameter estimates are sometimes referred to as *part-worths*. Each part-worth is the coefficient of utility associated with the given term. By default, these estimates are based on the Firth bias-corrected maximum likelihood estimators and therefore are considered to be more accurate than MLEs without bias correction.

**Std Error** An estimate of the standard deviation of the parameter estimate.

### Comparison Criteria

The following fit statistics are shown as part of the report and can be used to compare models: AICc (corrected Akaike's Information Criterion), BIC (Bayesian Information Criterion),  $-2 \times \text{LogLikelihood}$ , and  $-2 \times \text{Firth LogLikelihood}$ . See the Statistical Details appendix in the *Fitting Linear Models* book for details on the first three of these measures.

The  $-2 \times \text{Firth LogLikelihood}$  value is included in the report only when the Firth Bias-adjusted Estimates check box is checked in the launch window. This option is checked by default.

For each of these statistics, a smaller value indicates a better fit.

## JMP PRO Bayesian Parameter Estimates

(Appears only if Hierarchical Bayes is selected on the launch window.) The Bayesian Parameter Estimates report gives results for model effects. The estimates are based on a Hierarchical Bayes fit that integrates the subject-level covariates into the likelihood function and estimates their effects on the parameters directly. The subject-level covariates are estimated using a version of the algorithm described in Train (2001), which incorporates Adaptive Bayes and Metropolis-Hastings approaches. Posterior means and variances are calculated for each model effect. The algorithm also provides subject-specific estimates of the model effect parameters. See “[Save Subject Estimates](#)” on page 152.

During the estimation process, each individual is assigned his or her own vector of parameter estimates, essentially treating the estimates as random effects and covariates. The vector of coefficients for an individual is assumed to come from a multivariate normal distribution with arbitrary mean and covariance matrix. The likelihood function for the utility parameters for a given subject is based on a multinomial logit model for each subject’s preference within a choice set, given the attributes in the choice set. The prior distribution for a given subject’s vector of coefficients is normal with mean equal to zero and a diagonal covariance matrix with the same variance for each subject. The covariance matrix is assumed to come from an inverse Wishart distribution with a scale matrix that is diagonal with equal diagonal entries.

For each subject, a number of burn-in iterations at the beginning of the chain is discarded. By default, this number is equal to half of the Number of Bayesian Iterations specified on the launch window.

**Figure 5.16** Bayesian Parameter Estimates Report

Bayesian Parameter Estimates			
Term	Posterior Mean	Posterior Std Dev	Subject Std Dev
Product Of[Canada]	-0.341776602	0.1656807233	0.2008820509
Citizenship[Canadian]*Product Of[Canada]	-0.021887377	0.1665704936	0.1752016641
Gender[Female]*Product Of[Canada]	-0.312611454	0.1699325105	0.1747635982
Total Iterations	5000		
Burn-In Iterations	2500		
Number of Respondents	30		
Avg Log Likelihood After Burn-In	-197.0257		

**Term** The model term.

**Posterior Mean** The parameter estimate for the term’s coefficient. For each iteration after the burn-in period, the mean of the subject-specific coefficient estimates is computed. The Posterior Mean is the average of these means.

---

**Tip:** Select the red-triangle option Save Bayes Chain to see the individual estimates for each iteration.

---

**Posterior Std Dev** The standard deviation of the means of the subject-specific estimates over the iterations after burn-in.

**Subject Std Dev** The standard deviation of the subject-specific estimates around the posterior mean.

**Tip:** Select the red-triangle option Save Subject Estimates to see the individual estimates.

**Total Iterations** The total number of iterations performed, including the burn-in period.

**Burn-In Iterations** The number of burn-in iterations, which are discarded. This number is equal to half of the Number of Bayesian Iterations specified on the launch window.

**Number of Respondents** The number of subjects

**Avg Log Likelihood After Burn-In** The average of the log-likelihood function, computed on values obtained after the burn-in period.

Likelihood Ratio Tests

Figure 5.17 shows the Likelihood Ratio Tests report obtained by running the script **MaxDiff for Flavor** in **Potato Chip Responses.jmp**.

Figure 5.17 Likelihood Ratio Tests

Likelihood Ratio Tests				
Source	L-R		Prob>ChiSq	
	ChiSquare	DF		
Flavor	66.757	9	<.0001*	
Citizenship*Flavor	14.609	9	0.1023	
Gender*Flavor	9.480	9	0.3942	

**Source** Lists the effects in the model.

**L-R ChiSquare** The value of the likelihood ratio ChiSquare statistic for a test of the corresponding effect.

**DF** The degrees of freedom for the ChiSquare test.

**Prob>ChiSq** The *p*-value for the ChiSquare test.

**Bar Graph** Shows a bar graph of the L-R ChiSquare values.

---

## MaxDiff Platform Options

**JMP PRO Show MLE Parameter Estimates** (Available only if Hierarchical Bayes is selected on the launch window.) Shows non-Firth maximum likelihood estimates and standard errors for the coefficients of model terms. These estimates are used as starting values for the Hierarchical Bayes algorithm.

**Joint Factor Tests** (Not available if Hierarchical Bayes is selected on the launch window.) Tests each factor in the model by constructing a likelihood ratio test for all the effects involving that factor. For more information about Joint Factor Tests, see the Standard Least Squares Report and Options chapter in the *Fitting Linear Models* book.

**Confidence Intervals** If Hierarchical Bayes was not selected, shows a confidence interval for each parameter in the Parameter Estimates report.

If you selected Hierarchical Bayes, the confidence intervals appear in the Bayesian Parameter Estimates report. The intervals are constructed assuming a normal distribution and are based on the Posterior Mean and Posterior Std Dev.

**Correlation of Estimates** If Hierarchical Bayes was not selected, shows the correlations between the maximum likelihood parameter estimates.

If you selected Hierarchical Bayes, shows the correlation matrix for the posterior means of the parameter estimates. The correlations are calculated from the iterations after burn-in. The posterior means from each iteration after burn-in are treated as if they are columns in a data table. The Correlation of Estimates table is obtained by calculating the correlation matrix for these columns.

**Comparisons** Performs comparisons between specific alternative choice profiles. Enables you to select factor values and the values that you want to compare. You can compare specific configurations, including comparing all settings on the left or right by selecting the **Any** check boxes. Using **Any** does not compare all combinations across features, but rather all combinations of comparisons, one feature at a time, using the left settings as the settings for the other factors. See [“Comparisons Report”](#) on page 153.

**All Levels Comparison Report** Shows the All Levels Comparison Report, which contains a table with information on all pairwise comparisons of profiles. If you are modeling subject effects, you must specify a combination of subject effects and the table is specific to that combination of subject effects. Each cell of the table shows the difference in utilities for the row level and column level, the standard error of the difference, and a Wald  $p$ -value for a test of no difference.

---

**Caution:** The  $p$ -values are not corrected for multiple comparisons. Use these results as a guide.

---

The Wald  $p$ -values are colored. A saturated blue (respectively, red) color indicates that the Difference (Row - Column) is negative (respectively positive). The intensity of the red and blue coloring indicates the degree of significance, with a highly saturated red or blue meaning that the difference is highly significant.

**Save Utility Formula** When the analysis is on multiple data tables, creates a new data table that contains a formula column for utility. The new data table contains a row for each subject and profile combination, and columns for the profiles and the subject effects. When the analysis is on one data table, a new Utility Formula column is added.

**Save Gradients by Subject** Constructs a new table that has a row for each subject containing the average (Hessian-scaled-gradient) steps for the likelihood function on each parameter. This corresponds to using a Lagrangian multiplier test for separating that subject from the remaining subjects. These values can later be clustered, using the built-in-script, to indicate unique market segments represented in the data. For more details, see [“Example of Segmentation”](#) on page 109 in the “Choice Models” chapter.

**JMP PRO Save Subject Estimates** (Available only if Hierarchical Bayes is selected.) Creates a table where each row contains the subject-specific parameter estimates for each effect. The distribution of subject-specific parameter effects for each effect is centered at the estimate for the term given in the Bayesian Parameter Estimates report. The Subject Acceptance Rate gives the rate of acceptance for draws of new parameter estimates during the Metropolis-Hastings step. Generally, an acceptance rate of 0.20 is considered to be good. See [“Bayesian Parameter Estimates”](#) on page 149.

**JMP PRO Save Bayes Chain** (Available only if Hierarchical Bayes is selected.) Creates a table that gives information on the chain of iterations used in computing subject-specific Bayesian estimates. See [“Save Bayes Chain”](#) on page 153.

**Model Dialog** Shows the MaxDiff launch window that resulted in the current analysis, which can be used to modify and re-fit the model. You can specify new data sets, new IDs, and new model effects.

See the JMP Reports chapter in the *Using JMP* book for more information about the following options:

**Redo** Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

**Save Script** Contains options that enable you to save a script that reproduces the report to several destinations.



**Save By-Group Script** Contains options that enable you to save a script that reproduces the platform report for all levels of a By variable to several destinations. Available only when a By variable is specified in the launch window.

## Comparisons Report

The Comparisons report is shown when you specify pairwise comparisons. It contains the following columns:

**Factor** Shows the levels of the subject factors that you specified.

**Compared 1** Shows the factor and levels for the profile variables in the first component of the comparison.

**Compared 2** Shows the factor and levels for the profile variables in the second component of the comparison.

**Utility 1** Shows the estimated utility of the first component for the subjects specified in the Factor column.

**Utility 2** Shows the estimated utility of the second component for the subjects specified in the Factor column.

**Probability 1** Shows the predicted probability that the first component is preferred to the second for the subjects specified in the Factor column.

**Probability 2** Shows the predicted probability that the second component is preferred to the first for the subjects specified in the Factor column.

**Odds Ratio 1** Probability 1 divided by Probability 2.

**Odds Ratio 2** Probability 2 divided by Probability 1.

**Comparison Difference** Utility 1 minus Utility 2.

**Standard Deviation** The sample standard error of the estimated Comparison Difference.

## Save Bayes Chain

You can use the Bayes Chain data table to determine if your estimates have stabilized. The table that is created has a number of rows equal to the Number of Bayesian Iterations (specified on the launch window) plus one. The first row, Iteration 1, gives the starting values. The following rows show the results of the iterations, in order. The columns are arranged as follows:

**Iteration** Gives the iteration number, where the first row shows starting values.

**Log Likelihood** The log-likelihood of the model for that iteration. You can plot the Log Likelihood against Iteration to view behavior over the burn-in and tuning periods.

**Adaptive Sigma for <model effect>** Gives the estimate of the square root of the diagonal entries of the inverse Wishart distribution scale matrix for the corresponding effect.

**Acceptance for <model effect>** Gives the sampling acceptance rate for the corresponding effect.

**Mean of <model effect>** Gives the estimated mean for the corresponding effect.

**Variance of <model effect>** Gives the estimated variance for the corresponding effect.

# Chapter 6

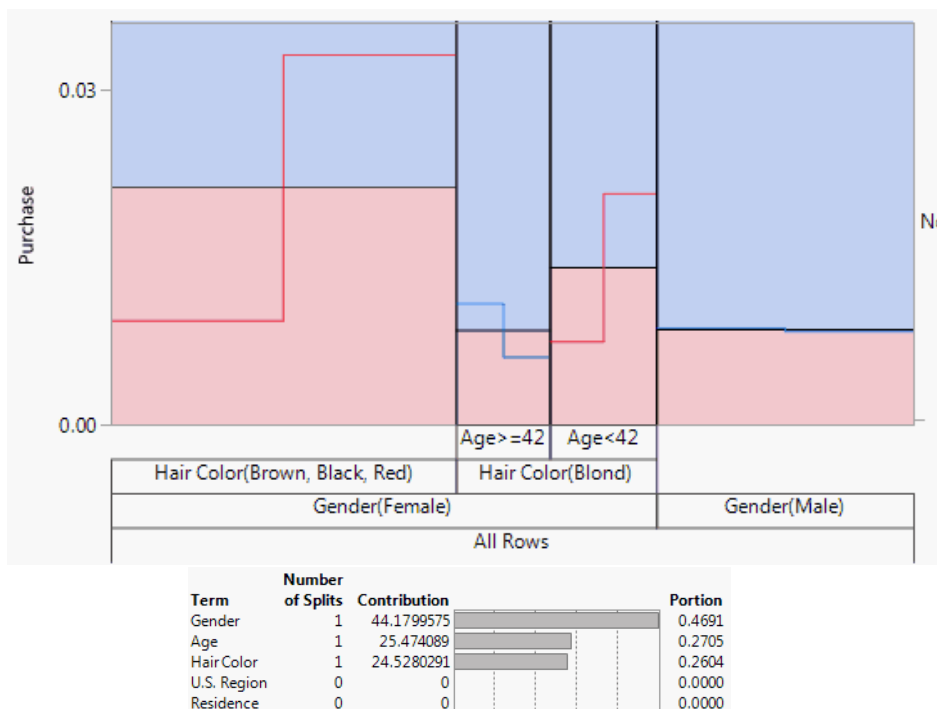
## JMP<sup>PRO</sup> Uplift Models

### Model the Incremental Impact of Actions on Consumer Behavior

**JMP<sup>PRO</sup>** Many features in this platform are available only in JMP Pro and noted with this icon.

Use uplift modeling to optimize marketing decisions, to define personalized medicine protocols, or, more generally, to identify characteristics of individuals who are likely to respond to an intervention. Also known as incremental modeling, true lift modeling, or net modeling, uplift modeling differs from traditional modeling techniques in that it finds the interactions between a treatment and other variables. It directs focus to individuals who are likely to react positively to an action or treatment.

**Figure 6.1** Example of Uplift for a Hair Product Marketing Campaign



**Contents**

Overview of the Uplift Platform ..... 157

Example of the Uplift Platform ..... 158

Launch the Uplift Platform ..... 159

The Uplift Model Report ..... 160

    Uplift Model Graph ..... 160

    Uplift Report Options..... 163



## Overview of the Uplift Platform

Use the Uplift platform to model the incremental impact of an action, or *treatment*, on individuals. An uplift model helps identify groups of individuals who are most likely to respond to the action. Identification of these groups leads to efficient and targeted decisions that optimize resource allocation and impact on the individual. See Radcliffe and Surry (2011).

The Uplift platform fits partition models. While traditional partition models find splits to optimize a prediction, uplift models find splits to maximize a treatment difference.

The uplift partition model accounts for the fact that some individuals receive the treatment, while others do not. It does this by fitting a linear model to each possible (binary) split. A continuous response is modeled as a linear function of the split, the treatment, and the interaction of the split and treatment. A categorical response is expressed as a logistic function of the split, the treatment, and the interaction of the split and treatment. In both cases, the interaction term measures the difference in uplift between the groups of individuals in the two splits.

The criterion used by the Uplift platform in defining splits is the significance of the test for interaction over all possible splits. However, predictor selection based solely on  $p$ -values introduces bias favoring predictors with many levels. For this reason, JMP adjusts  $p$ -values to account for the number of levels. (See the paper “Monte Carlo Calibration of Distributions of Partition Statistics” on the JMP website.) The splits in the Uplift platform are determined by maximizing the adjusted  $p$ -values for  $t$  tests of the interaction effects. The logworth for each adjusted  $p$ -value, namely  $-\log_{10}(\text{adj } p\text{-value})$ , is reported.



## Example of the Uplift Platform

The Hair Care Product.jmp sample data table results from a marketing campaign designed to increase purchases of a hair coloring product targeting both genders. For purposes of designing the study and tracking purchases, 126,184 “club card” members of a major beauty supply chain were identified. Approximately half of these members were randomly selected and sent a promotional offer for the product. Purchases of the product over a subsequent three-month period by all club card members were tracked.

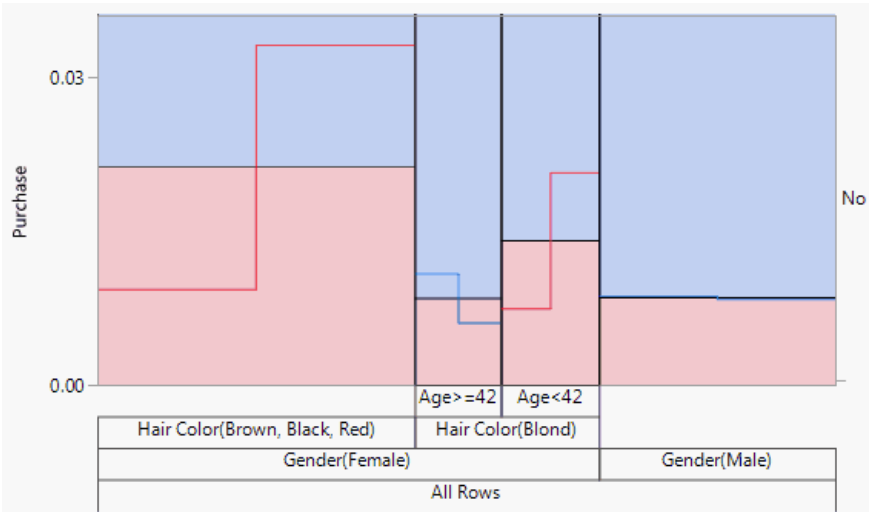
The data table shows a Promotion column, indicating whether the member received promotional material. The column Purchase indicates whether the member purchased the product over the test period. For each member, the following information was assembled: Gender, Age, Hair Color (natural), U.S. Region, and Residence (whether the member is located in an urban area). Also shown is a Validation column consisting of about 33% of the subjects.

For a categorical response, the Uplift platform interprets the first level in its value ordering as the response of interest. This is why the column Purchase has the Value Ordering column property. This property ensures that “Yes” responses are first in the ordering.

1. Select **Help > Sample Data Library** and open Hair Care Product.jmp.
2. Select **Analyze > Consumer Research > Uplift**.
3. From the Select Columns list:
  - Select Promotion and click **Treatment**.
  - Select Purchase and click **Y, Response**.
  - Select Gender, Age, Hair Color, U.S. Region, and Residence, and click **X, Factor**.
  - Select Validation and click **Validation**.
4. Click **OK**.
5. Below the Graph in the report that appears, click **Go**.

Based on the validation set, the optimal Number of Splits is determined to be three. The Graph is shown in Figure 6.2. Note that the vertical scale has been modified in order to show the detail.

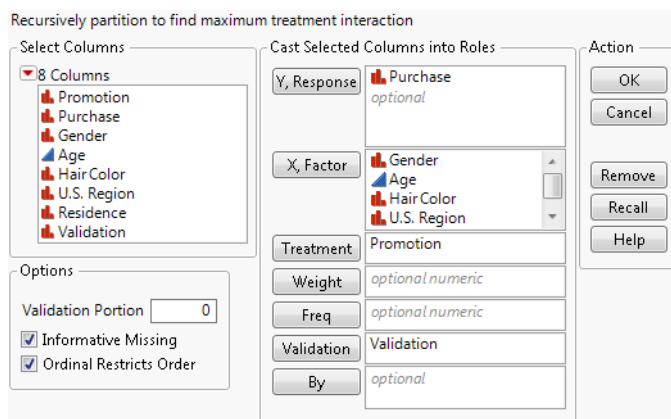
**Figure 6.2** Graph after Three Splits



The graph indicates that uplift in purchases occurs for females with black, red, or brown hair and for younger females (Age < 42) with blond hair. For older blond-haired women (Age ≥ 42) and males, the promotion has a negative effect.

## **JMP PRO** Launch the Uplift Platform

To launch the Uplift platform, select **Analyze > Consumer Research > Uplift**. Figure 6.3 shows a launch window for the Hair Care Product.jmp sample data table. The columns that you enter for Y, Response, and X, Factor can be continuous or categorical. In typical usage, the Treatment column is categorical, and often has only two levels. If your Treatment column contains more than two levels, the first level is treated as Treatment1 and the remaining levels are combined in Treatment2.

**Figure 6.3** Launch Window for Uplift

You can specify your own Validation column, or designate a random portion of your data to be selected as a Validation Portion. If you click the Validation button with no columns selected in the Select Columns list, you can add a validation column to your data table. For more information about the Make Validation Column utility, see the Modeling Utilities chapter in the *Predictive and Specialized Modeling* book.

The following options are also available:

**Informative Missing** If selected, enables missing value categorization for categorical predictors and informative treatment of missing values for continuous predictors.

**Ordinal Restricts Order** If selected, restricts consideration of splits to those that preserve the ordering.

---

## **JMP PRO** The Uplift Model Report

- “Uplift Model Graph”
- “Uplift Report Options”

## **JMP PRO** Uplift Model Graph

The graph represents the response on the vertical axis. The horizontal axis corresponds to observations, arranged by nodes. For each node, a black horizontal line shows the mean response. Within each split, there is a subsplit for treatment shown by a red or blue line. These lines indicate the mean responses for each of the two treatment groups within the split. The value ordering of the treatment column determines the placement order of these lines. As



nodes are split, the graph updates to show the splits beneath the horizontal axis. Vertical lines divide the splits.

Beneath the graph are the control buttons: **Split**, **Prune**, and **Go**. The Go button only appears if there is a validation set. Also shown is the name of the Treatment column and its two levels, called Treatment1 and Treatment2. If more than two levels are specified for the Treatment column, all but the first level are treated as a single level and combined into Treatment2.

To the right of the Treatment column information is a report showing summary values relating to prediction. (Keep in mind that prediction is not the objective in uplift modeling.) The report updates as splitting occurs. If a validation set is used, values are shown for both the training and the validation sets.

**RSquare** The RSquare for the regression model associated with the tree. Note that the regression model includes interactions with the treatment column. An RSquare closer to 1 indicates a better fit to the data than does an RSquare closer to 0.

---

**Note:** A low RSquare value suggests that there may be variables not in the model that account for the unexplained variation. However, if your data are subject to a large amount of inherent variation, even a useful uplift model may have a low RSquare value.

---

**RMSE** The root mean square error (RMSE) for the regression model associated with the tree. RMSE is only given for continuous responses. For more details, see the *Fitting Linear Models* book.

**N** The number of observations.

**Number of Splits** The number of times splitting has occurred.

**AICc** The Corrected Akaike Information Criterion (AICc), computed using the associated regression model. AICc is only given for continuous responses. For more details, see the Statistical Details appendix in the *Fitting Linear Models* book.

## **Uplift Decision Tree**

The decision tree shows the splits used to model uplift. See Figure 6.4 for an example using the Hair Care Product.jmp sample data table. Each node contains the following information:

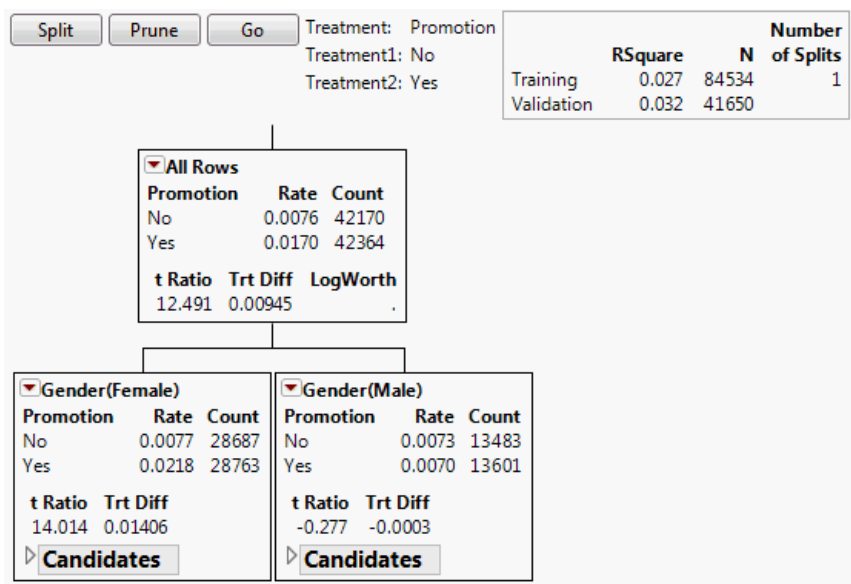
**Treatment** The name of the treatment column is shown, with its two levels.

**Rate** Only appears for two-level categorical responses. For each treatment level, the proportion of subjects in this node who responded.

**Mean** Only appears for continuous responses. For each treatment level, the mean response for subjects in this node.

- Count** The number of subjects in this node in the specified treatment level.
- t Ratio** The  $t$  ratio for the test for a difference in response across the levels of Treatment for subjects in this node. If the response is categorical, it is treated as continuous (values 0 and 1) for this test.
- Trt Diff** The difference in response means across the levels of Treatment. This is the uplift, assuming that:
- The first level in the treatment column’s value ordering represents the treatment.
  - The response is defined so that larger values reflect greater impact.
- LogWorth** The value of the logworth for the subsequent split based on the given node.

Figure 6.4 Nodes for First Split



Candidates Report

Each node also contains a Candidates report. This report gives the following information:

- Term** The model term.
- LogWorth** The maximum logworth over all possible splits for the given term. The logworth corresponding to a split is  $-\log_{10}$  of the adjusted  $p$ -value.
- F Ratio** When the response is continuous, this is the F Ratio associated with the interaction term in a linear regression model. The regression model specifies the response as a linear

function of the treatment, the binary split, and their interaction. When the response is categorical, this is the ChiSquare value for the interaction term in a nominal logistic model.

**Gamma** When the response is continuous, this is the coefficient of the interaction term in the linear regression model used in computing the  $F$  ratio. When the response is categorical, this is an estimate of the interaction constructed from Firth-adjusted log-odds ratios.

**Cut Point** If the term is continuous, this is the point that defines the split. If the term is categorical, this describes the first (left) node.

## **JMP<sup>®</sup> PRO** Uplift Report Options

With the exception of the options described below, all of the red triangle options for the Uplift report are described in the documentation for the Partition platform. For details about these options, see the Partition Models chapter in the *Predictive and Specialized Modeling* book.

### **JMP<sup>®</sup> PRO** Minimum Size Split

This option presents a window where you enter a number or a fractional portion of the total sample size to define the minimum size split allowed. To specify a number, enter a value greater than or equal to 1. To specify a fraction of the sample size, enter a value less than 1. The default value for the Uplift platform is set to 25 or the floor of the number of rows divided by 2,000, whichever value is greater.

### **JMP<sup>®</sup> PRO** Column Uplift Contributions

This table and plot address a column's contribution to the uplift tree structure. A column's contribution is computed as the sum of the  $F$  Ratio values associated with its splits. Recall that these values measure the significance of the treatment-by-split interaction term in the linear regression model.

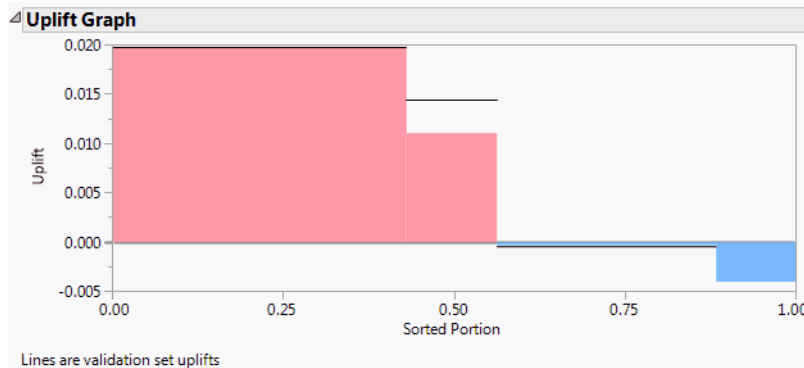
### **JMP<sup>®</sup> PRO** Uplift Graph

Consider the observations in the training set. Define uplift for an observation as the difference between the predicted probabilities or means across the levels of Treatment for the observation's terminal node. These uplift values are sorted in descending order. On its vertical axis, the Uplift Graph shows the uplift values. On its horizontal axis, the graph shows the proportion of observations with each uplift value.

See Figure 6.5 for an example of an Uplift Graph for the Hair Care Product.jmp sample data table after three splits. Note that, for two groups of subjects (males and non-blond women in the Age  $\geq 42$  group), the promotion has a negative effect.

The horizontal lines shown on the Uplift Graph delineate the graph for the validation set. Specifically, the decision tree is evaluated for the validation set and the Uplift Graph is constructed from the estimated uplifts.

**Figure 6.5** Uplift Graph



## JMP<sup>®</sup> PRO Save Columns

**Save Difference** Saves the estimated difference in mean responses across levels of Treatment for the observation's node. This is the estimated uplift.

**Save Difference Formula** Saves the formula for the Difference, or uplift.

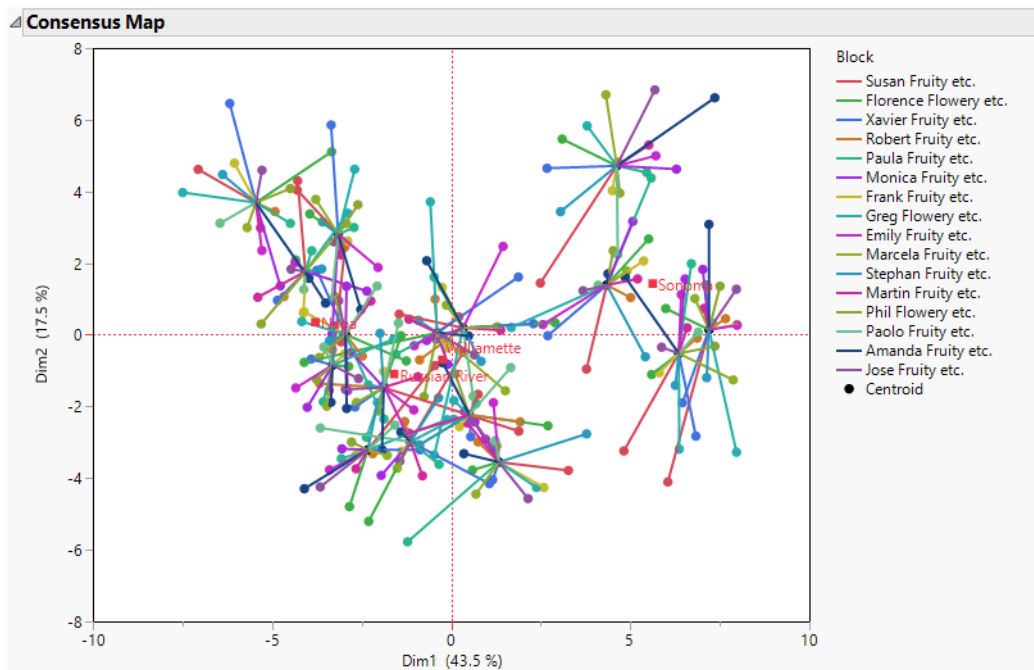
**Publish Difference Formula** Creates the difference formula and saves it as a formula column script in the Formula Depot platform. If a Formula Depot report is not open, this option creates a Formula Depot report. See the Formula Depot chapter in the *Predictive and Specialized Modeling* book.

## Multiple Factor Analysis

### Analyze Agreement among Panelists

Multiple factor analysis (MFA) is an analytical method closely related to principal components analysis (PCA). MFA uses eigenvalue decomposition to transform multiple measurements on the same items into orthogonal principal components. These components can help you understand how the items are similar and how they are different. MFA uses multiple table or consensus PCA techniques.

**Figure 7.1** Consensus Map in Multiple Factor Analysis



Contents

Overview of the Multiple Factor Analysis Platform ..... 167

Example of Multiple Factor Analysis..... 167

Launch the Multiple Factor Analysis Platform..... 169

    Data Format ..... 171

The Multiple Factor Analysis Report..... 172

    Summary Plots..... 172

    Consensus Map ..... 173

Multiple Factor Analysis Platform Options ..... 173

Statistical Details for the Multiple Factor Analysis Platform ..... 176

---

## Overview of the Multiple Factor Analysis Platform

Multiple factor analysis (MFA) is an analytical method that is closely related to principal components analysis (PCA). However, MFA differs from PCA in that it combines measurements from more than one table. Such tables are sometimes called sub-tables or sub-matrices. Each sub-table has the same number of rows, which represent the items or products being tested. In JMP, sub-tables are represented as groups of columns in a single data table. Each column group is called a block. Note the following about blocks:

- The number of columns in a block can vary. For example, in sensory analysis, a block represents a panelist. Some panelists might rate fewer attributes of a product than other panelists.
- Each block of columns can represent different measurements entirely. MFA scales each block to enable global analysis of all measurements.

The primary goal of MFA is to find groupings of products (rows in a data table) that are similar. A secondary goal is to identify outlier panelists. An outlier panelist results are so different from the rest of the group that they change the study results. Supplementary variables can be used investigate why items group together.

You can use MFA to analyze studies where items are measured on the same or different attributes by different instruments, individuals, or under different circumstances. MFA is frequently used in sensory analysis to account for different measurements among panelists. Traditional sensory analysis can entail hours of up-front training to ensure that panelists' measurements are consistent with each other. For example, consider a juice product with sensory measurements described as "fruity", "sweet", and "refreshing". In traditional sensory analysis, each panelist would have to be trained and tested to make sure reporting on distinct sensory measurements was consistent across panelists. MFA enables the researcher to perform a PCA-like analysis with untrained panelists.

When you use MFA, the same items are measured each time and the measurements can be arranged into internally consistent groups or blocks. For sensory analysis, the rows are the items measured, and the columns are the sensory aspects recorded by each panelist (there is a block for each panelist). Missing observations are replaced by the column mean.

For more information about multiple factor analysis, see Abdi et al. (2013).

---

## Example of Multiple Factor Analysis

This example uses data from a simulated sensory panel study of wine characteristics. Participants rated 16 wines on a number of characteristics from 1 (no intensity) to 10

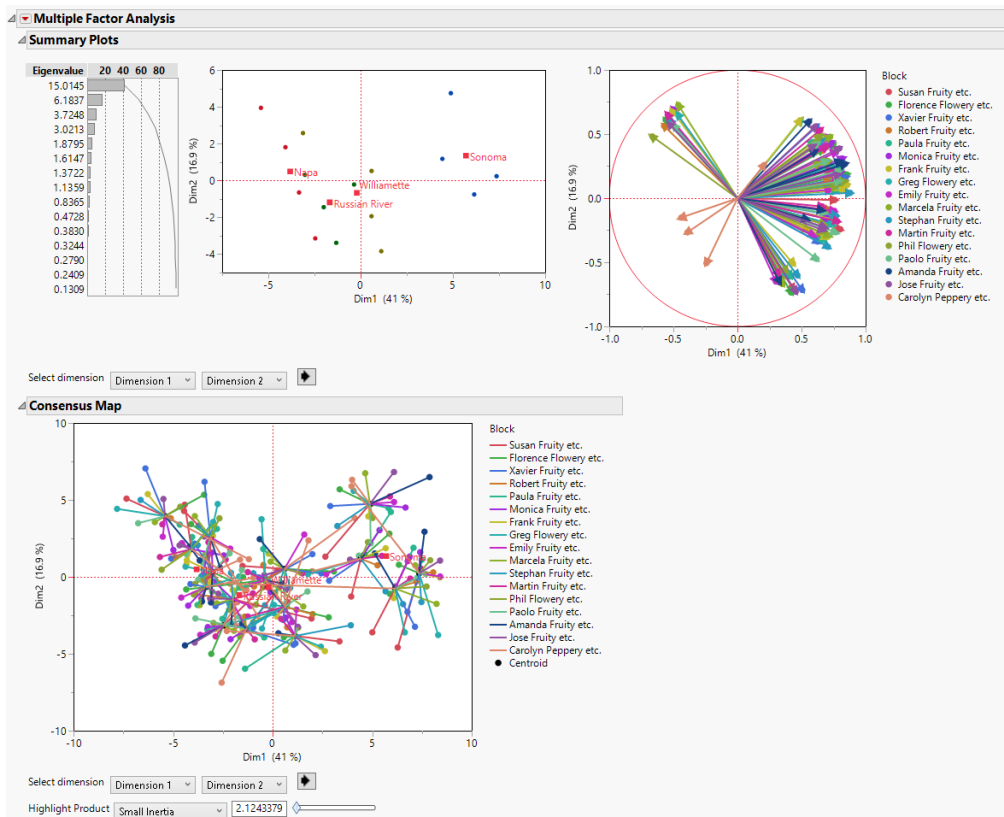
(prominent intensity). You want to better understand how the 16 wines are similar or different.

1. Select **Help > Sample Data Library** and open Wine Sensory Data.jmp.
2. Select **Analyze > Consumer Research > Multiple Factor Analysis**.
3. Select Vineyard and click **Product ID**.
4. Select Region and click **Z, Supplementary**.
5. Select all of the column groups from Carolyn Peppery etc. to Jose Fruity etc. and click **Add Block**.

**Note:** The columns in this data table are grouped into one block for each panelist. For ungrouped data, select the columns for a block, click **Add Block**, and repeat for each block.

6. Click **Run Model**.

**Figure 7.2** Initial Multiple Factor Analysis Report

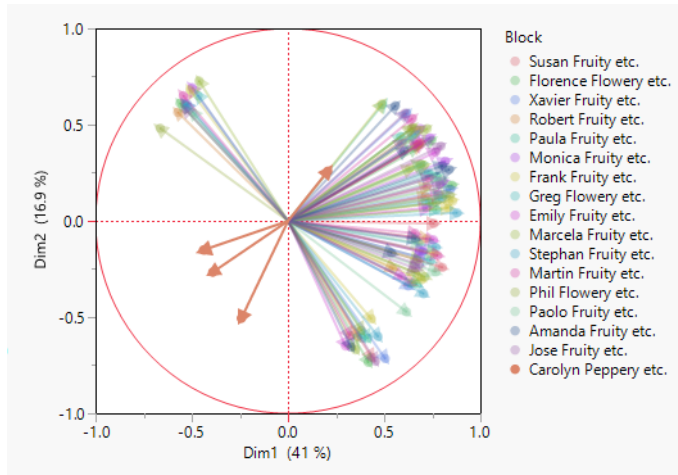


Notice the following in the Summary Plots:



- In the plot of the factor scores in the first two dimensions, the wines tend to cluster together according to their regions.
  - In the loading plot, the rays in the lower left quadrant correspond to Carolyn. They indicate a difference between Carolyn and the other raters.
7. In the legend next to the loading plot, click **Carolyn Peppery etc.** to highlight her results.

**Figure 7.3** Loading Plot with Results for Carolyn Highlighted

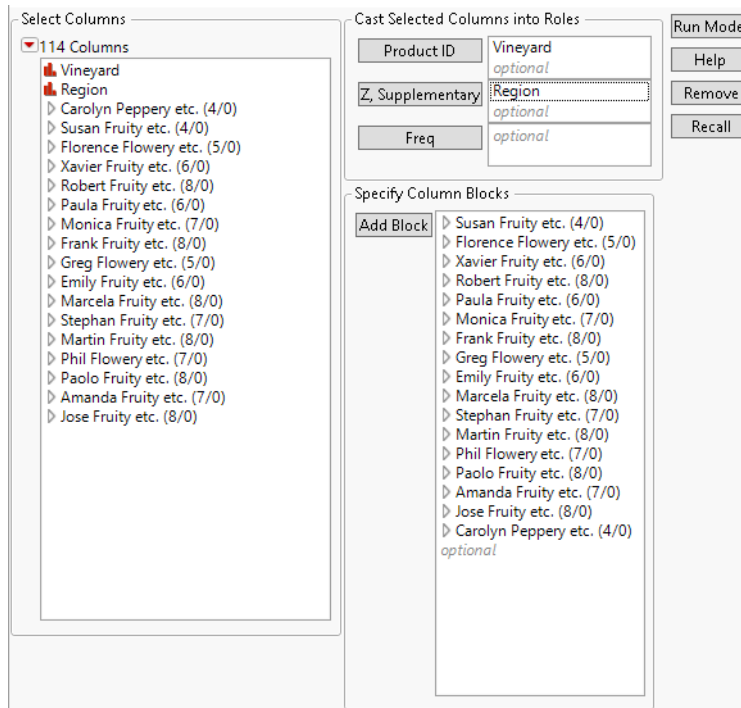


Carolyn's results differ from the other panelists. You might want to re-run the analysis without her results. See [Figure 7.6](#) on page 172 for the results of the analysis without Carolyn.

---

## Launch the Multiple Factor Analysis Platform

Launch the Multiple Factor Analysis Platform by selecting **Analyze > Consumer Research > Multiple Factor Analysis**.

**Figure 7.4** The Multiple Factor Analysis Launch Window


**Product ID** Specifies columns of items or products to be analyzed.

**Z, Supplementary** Specifies the columns to be used as supplementary variables. These variables are those with which you are interested in identifying associations, but they are not included in the calculations.

**Freq** Identifies one column whose numeric values assign a frequency to each row in the analysis.

**Add Block** Performs one of the following actions:

- Adds individual columns as a single block.
- Adds a column group as a block.
- Adds individual columns to a selected block.

---

**Tip:** If you group the columns into blocks before running the platform, you can select multiple column groups and cast them as blocks in a single action. Otherwise, you must select each group of columns for each block and click Add Block, one block at a time. Double-click on a block name to change it.

---

The Multiple Factor Analysis platform uses a data table that contains column groups, or sub-tables. Each column group, referred to as a block, can have a different number of columns. Each block of columns can represent different measurements. The columns or blocks do not have to be in JMP column groups. However, the platform is easier to launch when the columns are grouped into blocks in the data table.

The data table rows represent the items that are being measured. Observations for each item must be in a single row. For example, Figure 7.5 shows a table that is measuring attributes of 16 wines from different vineyards. The column panel shows the column groups, or sub-tables, for each panelist.

**Figure 7.5** Partial View of a Data Table for Multiple Factor Analysis

The screenshot displays the SAS Enterprise Guide interface. The top menu bar includes File, Edit, Tables, Rows, Cols, DOE, Analyze, Graph, Tools, View, Window, and Help. Below the menu bar is a toolbar with various icons for file operations, data manipulation, and analysis. The main workspace shows a data table with the following structure:

	Vineyard	Region	Carolyn Peppery	Carolyn Tannic	Carolyn Aromatic
1	Rabbit Stone	Napa	5	8	8
2	Orchid Lake	Napa	6	4	3
3	Flying Cat	Napa	9	4	9
4	Zack's Zin Reserve	Napa	9	4	3
5	Golden Acres Estate	Russian River	1	10	5
6	Star Trail Reserve	Russian River	8	4	9
7	Dancer's Leap	Russian River	6	8	4
8	Regency Park	Russian River	10	10	1
9	Sandy Beach	Sonoma	8	4	5
10	Quick Creek Canyon	Sonoma	5	8	8
11	Deer Ridge	Sonoma	2	1	5
12	Oak Arbor	Sonoma	3	3	9
13	Beechtree Mountain	Williamette	1	6	8
14	Pinetree Vinyards	Williamette	6	7	2
15	Sunny Hill	Williamette	5	3	3
16	Green Estate	Williamette	10	7	4

The left pane shows the project tree with the following structure:

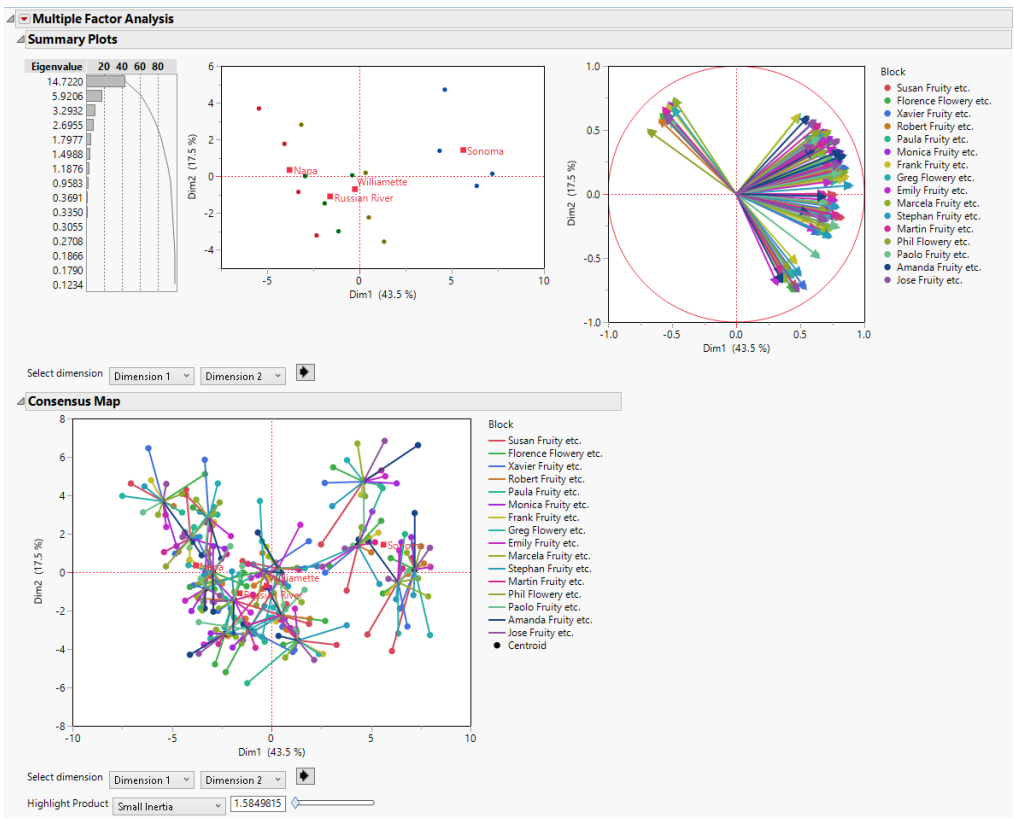
- Columns (114/0)
  - Vineyard \*
  - Region
  - Carolyn Peppery etc. (4/0)
  - Susan Fruity etc. (4/0)
    - Susan Fruity
    - Susan Flowery
    - Susan Spicy
    - Susan Crispness
  - Florence Flowery etc. (5/0)
  - Xavier Fruity etc. (6/0)
    - Xavier Fruity
    - Xavier Spicy
    - Xavier Crispness
    - Xavier Alcohol
    - Xavier Savory
    - Xavier Lightness
  - Robert Fruity etc. (8/0)
  - Paula Fruity etc. (6/0)
  - Monica Fruity etc. (7/0)

**Note:** Missing observations are replaced by the column mean. When missing observations result in no variation for a column, the column is excluded from the analysis. Missing rules are applied to all variables, including supplementary variables.

# The Multiple Factor Analysis Report

The initial Multiple Factor Analysis report shows a table of eigenvalues, summary plots, and a consensus map.

Figure 7.6 Multiple Factor Analysis Report



## Summary Plots

The Summary Plot report has the following three sections:

- The first section shows the eigenvalues of the consensus PCA with a plot of the cumulative percent of variance explained by each component. A consensus PCA is used to obtain a common representation of the blocks of data. Consensus PCA refers to the principal component solution of the weighted sub-tables and is used to obtain a common representation of the blocks of data.

- The middle section is a plot of factor scores with a marker for each row (item). If supplementary variables are used, there is a labeled marker for each level of the variables. Items that cluster together in this plot are considered to be similar.
- The third section is a loading plot of factor loadings for each block.

---

**Tip:** In the loading plot legend, click on a block to highlight it in the plot.

---

**Select dimension** Controls the dimensions plotted on the score and loading plots. The first control selects the horizontal dimension, and the second control selects the vertical dimension.

## Consensus Map

The Consensus Map report contains a plot that overlays the individual panelist responses with the average response among all panelists for each item. You might use this map to investigate response consistency among panelists. For example, if a given panelist's points fall consistently farther from the average of the other panelists, then that panelist might be a candidate to exclude from the analysis.

**Select dimension** Controls the dimensions plotted on the consensus map. The first control selects the horizontal dimension, and the second control selects the vertical dimension.

**Highlight Product** Controls the transparency of the items according to their inertia score. Small inertia indicates items that panelists have good agreement on. Large inertia indicates items that panelists do not agree on.

**Small Inertia** Highlights items with inertia less than or equal to the value in the text box. To adjust the cutoff for the inertia value, use the slider or enter a value in the text box.

**Large Inertia** Highlights items with inertia greater than or equal to the value in the text box. To adjust the cutoff for the inertia value, use the slider or enter a value in the text box.

**Min and Max Inertia** Highlights the items with the smallest and largest inertia. The text box and slider have no impact on the results when this option is selected.

---

## Multiple Factor Analysis Platform Options

The Multiple Factor Analysis red triangle menu includes the following options.

**Block Weights** Shows or hides the first eigenvalue for each block as well as the weight of that eigenvalue. The weight is the inverse of the square root of the first eigenvalue.

**Eigenvalues** Shows or hides a table of eigenvalues that correspond to the consensus principal components, in order, from largest to smallest.

**Eigenvectors** Shows or hides a table of the eigenvectors for each of the consensus principal components, in order, from left to right. Using these coefficients to form a linear combination of the original variables produces the consensus principal component variables. Following the standard convention, eigenvectors have norm 1.

**Variable Loadings** Shows or hides the loadings for each column. The degree of transparency for the table values indicates the distance of the absolute loading value from zero. Absolute loading values that are closer to zero are more transparent than absolute loading values that are farther from zero.

**Variable Partial Contributions** Shows or hides a table that contains the partial contributions of variables. The partial contributions enable you to see the percentage of variance that each variable contributes to each consensus principal component.

**Variable Squared Cosines** Shows or hides a table that contains the squared cosines of variables. The sum of the squared cosine values across consensus principal components is equal to 1 for each variable. The squared cosines enable you to see how well the variables are represented by the consensus principal components.

---

**Tip:** For the variable loadings, variable partial contributions, and variable squared cosines, values near zero indicate that the variable has little effect on the consensus principal component. Values far from zero indicate a strong effect. The degree of transparency for the table values highlights these effects.

---

**Summary Plots** Shows or hides the summary plots. The summary plots include the plot of the eigenvalues or score plot and the loading plot.

**Consensus Map** Shows or hides the consensus map. See [“Consensus Map”](#) on page 173.

**Biplot** Shows or hides a plot that is an overlay of the score and loadings plots. Use the controls to select any two dimensions for the plot.

**Partial Axes Plot** Shows or hides a partial axes plot. This plot displays correlations between PCA scores from separate block analyses and the consensus principal component. Use the controls to select any two dimensions for the plot. Click on a block in the legend to highlight that block in the plot.

### Display Options

**Arrow Lines** Enables you to show or hide arrows on the loading plot, and the partial axes plot. Arrows are shown if the number of variables is 1000 or fewer. If there are more than 1000 variables, the arrows are off by default.

**Show Labels** Shows or hides column name labels on all points in the loading plot, bi-plot, and the partial axes plot.

**RV Correlations** Shows or hides a matrix of squared correlation coefficients between blocks.

**Lg Coefficients** Shows or hides a matrix of similarity measures between blocks.

**Block Partial and Consensus Correlations** Shows or hides a matrix of correlation coefficients between block partial scores and consensus principal component scores. The matrix is rectangular because only correlations between concordant dimensions are displayed.

**Block Partial Contributions** Shows or hides the sum of the variable contributions within the block.

**Block Partial Inertias** Shows or hides the block contribution multiplied by the eigenvalue for the principal component and then divided by 100.

**Block Squared Cosines** Shows or hides the block inertia squared and divided by the sum of squares used to calculate the eigenvalues. The values have a range between 0 and 1, and the sum of the Block Squared Cosines for a single block across all principal components is 1. The Block Squared Cosines can be considered as the percentage of the block variance explained by each principal component.

**Save Individual Scores** Saves the item consensus principal components to new columns in the data table. If one or more categorical supplementary variables are used, this option also saves individual scores for each level of the supplementary variables to a new data table.

**Save Individual Squared Cosines** Saves the item squared cosines to new columns in the data table. If one or more categorical supplementary variables are used, this option also saves categorical supplementary variable squared cosines to a new data table.

**Save Individual Partial Contributions** Saves the item partial contributions to new columns in the data table. If one or more categorical supplementary variables are used, this option also saves categorical supplementary variable squared cosines to a new data table.

**Save Block Partial Scores** Saves the block partial scores to a new data table.

**Save Partial Axes Coordinates** Saves the partial axes coordinates to a new data table.

See the JMP Reports chapter in the *Using JMP* book for more information about the following options:

**Local Data Filter** Shows or hides the local data filter that enables you to filter the data used in a specific report.

**Redo** Contains options that enable you to repeat or relaunch the analysis. In platforms that support the feature, the Automatic Recalc option immediately reflects the changes that you make to the data table in the corresponding report window.

**Save Script** Contains options that enable you to save a script that reproduces the report to several destinations.

---

## Statistical Details for the Multiple Factor Analysis Platform

Multiple factor analysis combines information from sub-tables into a set of orthogonal columns that describe the items in the rows of the table. The basic procedure is as follows:

- Perform PCA on each sub-table.
- Record the first eigenvalue of each sub-table to create a matrix of weights.
- Concatenate the sub-tables side-by-side, center and normalize the matrix.
- Perform a generalized PCA on the concatenated table via the singular value decomposition. Generalized PCA is used to constrain the solution using the sub-table weights.

This results in three matrices of generalized right and left singular vectors and singular values. These are then used to derive component scores, eigenvalues, and component loadings for the consensus across sub-tables. These three matrices are the result of decomposing the many columns from the original measurements into a few interpretable dimensions that explain the similarities and differences between the objects being measured.

### Calculations

For MFA, a singular value decomposition of the  $\mathbf{X}$  matrix can be defined as follows:

$$\mathbf{X} = \mathbf{P}\mathbf{\Delta}\mathbf{Q}^T \quad \text{with the constraint} \quad \mathbf{P}^T\mathbf{M}\mathbf{P} = \mathbf{Q}^T\mathbf{A}\mathbf{Q} = \mathbf{I}$$

The matrices use are as follows:

$\mathbf{X}$  is an  $n \times p$  centered and normalized matrix of sub-tables. In consumer research there are  $n$  products and  $p$  panelists' ratings.

$\mathbf{Q}$  is a  $p \times q$  matrix of right singular vectors, which are weighted by the MFA singular values to obtain the loadings on  $q$  principal components.

$\mathbf{\Delta}$  is a  $q \times q$  diagonal matrix of singular values from the generalized PCA. As with PCA, the magnitude of the squared singular values, or eigenvalues, represent the importance of each principal component in the combined analysis.



**P** is an  $n \times q$  matrix of right singular vectors, which are weighted by the MFA singular values to obtain the  $q$  principal components of the compromise.

**M** is the  $n \times n$  diagonal matrix of mass weights.

**A** is the  $p \times p$  diagonal matrix of block or panelist weights.

For more information about multiple factor analysis, see Abdi et al. ([2013](#)).

### **Mass Weight**

JMP calculations use  $N - 1$  for mass weight calculations. These calculations affect individual and block partial scores.



# Appendix **A**

## References

---

- Abdi, H., Williams, L. J., and Valentin, D. (2013). "Multiple factor analysis: principal component analysis for multitable and multiblock data sets." *WIREs Comp Stat*, 5:149-179. doi: 10.1002/wics.1246.
- Benjamini, Y., and Hochberg, Y. (1995). "Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing." *Journal of the Royal Statistical Society, Series B* 57:289-300.
- Breslow, N. E., and Day, N. E. (1980). *The Analysis of Case-Control Studies*. Statistical Methods in Cancer Research, IARC Scientific Publications, vol. 1, no. 32. Lyon: International Agency for Research on Cancer. <https://www.iarc.fr/en/publications/pdfs-online/stat/sp32/SP32.pdf>. Retrieved 3/9/2017.
- Firth, D. (1993). "Bias Reduction of Maximum Likelihood Estimates." *Biometrika* 80:27-38.
- Heinze, G., and Schemper, M. (2002). "A Solution to the Problem of Separation in Logistic Regression." *Statistics in Medicine* 21:2409-2419.
- Kish, L. (1965). *Survey Sampling*. New York: John Wiley & Sons.
- Lavassani, K. M., Movahedi, B., and Kumar, V. (2009). "Developments in Analysis of Multiple Response Survey Data in Categorical Data Analysis: The Case of Enterprise System Implementation in Large North American Firms." *Journal of Applied Quantitative Methods* 4:45-53.
- Louviere, J. J., Flynn, T. N., and Marley, A. A. (2015). *Best-Worst Scaling: Theory, Methods and Applications*. Cambridge: Cambridge University Press.
- McFadden, D. (1974). "Conditional Logit Analysis of Qualitative Choice Behavior." in *Frontiers in Econometrics* edited by P. Zarembka, 105-142. New York: Academic Press. See <http://eml.berkeley.edu/reprints/mcfadden/zarembka.pdf>. Retrieved 4/25/2016.
- Radcliffe, N. J., and Surry, P. D. (2011). *Real-World Uplift Modeling with Significance-Based Uplift Trees*. Portrait Technical Report TR-2011-1, Stochastic Solutions. <http://www.stochasticsolutions.com/pdf/sig-based-up-trees.pdf>. Retrieved 1/13/2017.
- Rossi, P. E., Allenby, G. M., and McCulloch, R. (2005). *Bayesian Statistics and Marketing*. Chichester, UK: John Wiley & Sons.
- SAS Institute Inc. (2017). "The TTEST Procedure." In *SAS/STAT 14.3 User's Guide*. Cary NC: SAS Institute Inc. [http://go.documentation.sas.com/?docsetId=statug&docsetTarget=statug\\_ttest\\_details14.htm&docsetVersion=14.3&locale=en](http://go.documentation.sas.com/?docsetId=statug&docsetTarget=statug_ttest_details14.htm&docsetVersion=14.3&locale=en). Retrieved 12/22/2017

- Train, K. E. (2001). "A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit." Department of Economics, University of California, Berkley.  
<https://eml.berkeley.edu/~train/compare.pdf>. Retrieved 1/13/2017.
- Train, K. E. (2009). *Discrete Choice Methods and Simulation*. 2nd ed. Cambridge University Press.
- Westfall, P. H., Tobias, R. D., and Wolfinger, R. D. (2011). *Multiple Comparisons and Multiple Tests Using SAS*. 2nd ed. Cary, NC: SAS Institute Inc.

# Index

## Consumer Research

---

### A

All Levels Comparison Report, MaxDiff [151](#)

### B

best-worst scaling [131](#)

### C

Categorical Platform [29](#)

Free Text [34](#)

Launch Window [32](#)

Report [36](#)

Structured [34](#)

Supercategories [48](#)

Choice modeling [71–125](#)

Choice platform

Effect Marginals [96](#)

Joint Factor Tests [95](#)

No Choice option [86, 90](#)

Profile data [87](#)

Save Gradients by Subject [74, 97, 131](#)

Save Utility Formula [97](#)

segmentation [74, 131](#)

Subject data [90](#)

Willingness to Pay [97](#)

Comparisons with Letters [41, 48](#)

Comparisons, MaxDiff [151](#)

Confidence Intervals, MaxDiff [151](#)

Correlation of Estimates, MaxDiff [151](#)

### E

Effect Marginals, Choice platform [96](#)

### J

Joint Factor Tests

MaxDiff [151](#)

Joint Factor Tests, Choice platform [95](#)

### L

Likelihood Ratio Tests [95](#)

### M

MaxDiff

All Levels Comparison Report [151](#)

best-worst scaling [131](#)

Correlation of Estimates [151](#)

pizza choice example [131–133](#)

Profile Data [141](#)

Response Data [142](#)

Save Gradients by Subject [152](#)

Save Utility Formula [152](#)

MaxDiff, Joint Factor Tests [151](#)

### N

No Choice option, Choice platform [86, 90](#)

### P

pizza choice example [109–112](#)

pizza choice example, MaxDiff [131–133](#)

Profile data, Choice platform [87](#)

Profile Data, MaxDiff [141](#)

### R

Response Data, MaxDiff [142](#)

### S

Save Gradients by Subject, Choice platform [74, 97, 131](#)

Save Gradients by Subject, MaxDiff [152](#)

Save Utility Formula, Choice platform [97](#)

Save Utility Formula, MaxDiff [152](#)  
Subject data, Choice platform [90](#)

## T

Test Each Response [39](#)  
Test Response Homogeneity [39](#)

## U

Uplift Model [155](#)  
    platform [159](#)  
    report options [163](#)

## W-Z

Willingness to Pay, Choice platform [97](#)