Statistical Discovery.™ From SAS.

Version 15

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

15.1

The correct bibliographic citation for this manual is as follows: SAS Institute Inc. 2020. *JMP® 15 Consumer Research*. Cary, NC: SAS Institute Inc.

JMP® 15 Consumer Research

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

September 2019

February 2020

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

https://www.jmp.com/getstarted

Contents

Consumer Research

1 Learn about JMP 9
Documentation and Additional Resources

Formatting Conventions 11
JMP Help 12
JMP Documentation Library 12
Additional Resources for Learning JMP 18
Tutorials 18
Sample Data Tables 18
Learn about Statistical and JSL Terms 19
Learn JMP Tips and Tricks 19
Tooltips 19
JMP User Community 20
Free Online Statistical Thinking Course 20
New User Welcome Kit 20
Statistics Knowledge Portal 20
JMP Training 20
JMP Books by Users 21
The JMP Starter Window 21
Technical Support 21

2 Introduction to Consumer Research 23
Overview of Customer and Behavioral Research Methods

3 Categorical Response Analysis 25
Analyze Survey and Other Counting Data

Example of the Categorical Platform 27
Launch the Categorical Platform 33
Response Roles 33
Columns Roles 38
Other Launch Window Options 38
The Categorical Report 40
Categorical Platform Options 41
Report Options 41

Statistical Testing Options 42
Additional Categorical Platform Options 44
Crosstab Table Options 47
Comparison Letters 48
Supercategories 49
Supercategories Options 50
Additional Examples of the Categorical Platform 51
Example of the Test for Response Homogeneity 51
Example of the Multiple Response Test 52
Example of Supercategories 54
Example of the Cell Chisq Test 57
Example of Compare Each Sample with Comparison Letters 58
Example of Compare Each Cell with Comparison Letters 59
Example of User-Specified Comparison with Comparison Letters 61
Example of Aligned Responses 62
Example of Conditional Association and Relative Risk 64
Example of Rater Agreement 66
Example of Repeated Measures 67
Example of the Multiple Responses 68
Example of Mean Score with Comparison Letters 73
Example of a Structured Report 75
Example of a Multiple Response with Nonresponse 76
Set Preferences 79
Statistical Details for the Categorical Platform 80
Rao-Scott Correction 80

4 Choice Models 81
Fit Models for Choice Experiments

Overview of the Choice Modeling Platform 83
Examples of the Choice Platform 85
One Table Format with No Choice 85
Multiple Table Format 88
Launch the Choice Platform 95
Launch Window for One Table, Stacked 96
Launch Window for Multiple Tables, Cross-Referenced 98
Choice Model Report 104
Effect Summary 104
Parameter Estimates 105
Likelihood Ratio Tests 106
Bayesian Parameter Estimates 106

Choice Platform Options 108
Willingness to Pay 111
Additional Examples 114
Example of Making Design Decisions 114
Example of Segmentation 123
Example of Logistic Regression Using the Choice Platform 127
Example of Logistic Regression for Matched Case-Control Studies 130
Example of Transforming Data to Two Analysis Tables 132
Example of Transforming Data to One Analysis Table 137
Statistical Details for the Choice Platform 139
Special Data Table Rules 139
Utility and Probabilities 140
Gradients 141

5 MaxDiff 143
Fit Models for MaxDiff Experiments
Overview of the MaxDiff Modeling Platform 145
Examples of the MaxDiff Platform 146
One Table Format 146
Multiple Table Format 149
Launch the MaxDiff Platform 152
Launch Window for One Table, Stacked 154
Launch Window for Multiple Tables, Cross-Referenced 156
MaxDiff Model Report 160
Effect Summary 160
MaxDiff Results 161
Parameter Estimates 162
Bayesian Parameter Estimates 164
Likelihood Ratio Tests 165
MaxDiff Platform Options 166
Comparisons Report 168
Save Bayes Chain 168

6 Uplift Models 171
Model the Incremental Impact of Actions on Consumer Behavior
Overview of the Uplift Platform 173
Example of the Uplift Platform 174
Launch the Uplift Platform 175
The Uplift Model Report 176
Uplift Model Graph 177
Uplift Report Options 179

7 Multiple Factor Analysis 181
Analyze Agreement among Panelists
Overview of the Multiple Factor Analysis Platform 183
Example of Multiple Factor Analysis 184
Launch the Multiple Factor Analysis Platform 187
Data Format 188
The Multiple Factor Analysis Report 190
Summary Plots 190
Consensus Map 191
Multiple Factor Analysis Platform Options 191
Statistical Details for the Multiple Factor Analysis Platform 194

A References 197

B Technology License Notices 199

Chapter 1

Learn about JMP

Documentation and Additional Resources

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

Contents

- Formatting Conventions . . . 11
- JMP Help . . . 12
- JMP Documentation Library . . . 12
- Additional Resources for Learning JMP . . . 18
  - Tutorials . . . 18
  - Sample Data Tables . . . 18
  - Learn about Statistical and JSL Terms . . . 19
  - Learn JMP Tips and Tricks . . . 19
  - Tooltips . . . 19
  - JMP User Community . . . 20
  - Free Online Statistical Thinking Course . . . 20
  - New User Welcome Kit . . . 20
  - Statistics Knowledge Portal . . . 20
  - JMP Training . . . 20
  - JMP Books by Users . . . 21
  - The JMP Starter Window . . . 21
- Technical Support . . . 21

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 (or sans-serif online) font.
- Code appears in `Lucida Sans Typewriter` (or monospace online) font.
- Code output appears in *`Lucida Sans Typewriter`* italic (or monospace italic online) font and is indented farther than the preceding code.
- **Helvetica bold** formatting (or bold sans-serif online) 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 Help

JMP Help in the Help menu enables you to search for information about JMP features, statistical methods, and the JMP Scripting Language (or *JSL*). You can open JMP Help in several ways:

- Search and view JMP Help on Windows by selecting the **Help > JMP Help**.
- On Windows, press the F1 key to open the Help system in the default browser.
- 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.

Note: The JMP Help is available for users with Internet connections. Users without an Internet connection can search all books in a PDF file by selecting **Help > JMP Documentation Library**. See “JMP Documentation Library” on page 12 for more information.

JMP Documentation Library

The Help system content is also available in one PDF file called *JMP Documentation Library*. Select **Help > JMP Documentation Library** to open the file. If you prefer searching individual PDF files of each document in the JMP library, download the files from https://www.jmp.com/documentation.

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

Document Title	Document Purpose	Document Content
Discovering JMP	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.
Using JMP	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
Basic Analysis	Perform basic analysis using this document.	Describes these Analyze menu platforms: • Distribution • Fit Y by X • Tabulate • Text Explorer Covers how to perform bivariate, one-way ANOVA, and contingency analyses through Analyze > Fit Y by X. How to approximate sampling distributions using bootstrapping and how to perform parametric resampling with the Simulate platform are also included.
Essential Graphing	Find the ideal graph for your data.	Describes these Graph menu platforms: • Graph Builder • Scatterplot 3D • Contour Plot • Bubble Plot • Parallel Plot • Cell Plot • Scatterplot Matrix • Ternary Plot • Treemap • Chart • Overlay Plot The book also covers how to create background and custom maps.
Profilers	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
Design of Experiments Guide	Learn how to design experiments and determine appropriate sample sizes.	Covers all topics in the DOE menu.
Fitting Linear Models	Learn about Fit Model platform and many of its personalities.	Describes these personalities, all available within the Analyze menu Fit Model platform: • Standard Least Squares • Stepwise • Generalized Regression • Mixed Model • MANOVA • Loglinear Variance • Nominal Logistic • Ordinal Logistic • Generalized Linear Model

Document Title	Document Purpose	Document Content
Predictive and Specialized Modeling	Learn about additional modeling techniques.	Describes these Analyze > Predictive Modeling menu platforms: • Neural • Partition • Bootstrap Forest • Boosted Tree • K Nearest Neighbors • Naive Bayes • Support Vector Machines • Model Comparison • Make Validation Column • Formula Depot Describes these Analyze > Specialized Modeling menu platforms: • Fit Curve • Nonlinear • Functional Data Explorer • Gaussian Process • Time Series • Matched Pairs Describes these Analyze > Screening menu platforms: • Modeling Utilities • Response Screening • Process Screening • Predictor Screening • Association Analysis • Process History Explorer

Document Title	Document Purpose	Document Content
Multivariate Methods	Read about techniques for analyzing several variables simultaneously.	Describes these Analyze > Multivariate Methods menu platforms: • Multivariate • Principal Components • Discriminant • Partial Least Squares • Multiple Correspondence Analysis • Structural Equation Models • Factor Analysis • Multidimensional Scaling • Item Analysis Describes these Analyze > Clustering menu platforms: • Hierarchical Cluster • K Means Cluster • Normal Mixtures • Latent Class Analysis • Cluster Variables
Quality and Process Methods	Read about tools for evaluating and improving processes.	Describes these Analyze > Quality and Process menu platforms: • Control Chart Builder and individual control charts • Measurement Systems Analysis • Variability / Attribute Gauge Charts • Process Capability • Model Driven Multivariate Control Chart • Pareto Plot • Diagram • Manage Spec Limits

Document Title	Document Purpose	Document Content
Reliability and Survival Methods	Learn to evaluate and improve reliability in a product or system and analyze survival data for people and products.	Describes these Analyze > Reliability and Survival menu platforms: • Life Distribution • Fit Life by X • Cumulative Damage • Recurrence Analysis • Degradation • Destructive Degradation • Reliability Forecast • Reliability Growth • Reliability Block Diagram • Repairable Systems Simulation • Survival • Fit Parametric Survival • Fit Proportional Hazards
Consumer Research	Learn about methods for studying consumer preferences and using that insight to create better products and services.	Describes these Analyze > Consumer Research menu platforms: • Categorical • Choice • MaxDiff • Uplift • Multiple Factor Analysis
Scripting Guide	Learn about taking advantage of the powerful JMP Scripting Language (JSL).	Covers a variety of topics, such as writing and debugging scripts, manipulating data tables, constructing display boxes, and creating JMP applications.
JSL Syntax Reference	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.

Additional Resources for Learning JMP

In addition to reading 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"
- "Free Online Statistical Thinking Course"
- "New User Welcome Kit"
- "Statistics Knowledge Portal"
- "JMP Training"
- "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, 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\15\Samples\Data

On macOS: \Library\Application Support\JMP\15\Samples\Data

In JMP Pro, sample data is installed in the JMPPRO (rather than JMP) 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 https://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 and get help on the commands.

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 (or *hover labels*) 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 macOS.

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.

Free Online Statistical Thinking Course

Learn practical statistical skills in this free online course on topics such as exploratory data analysis, quality methods, and correlation and regression. The course consists of short videos, demonstrations, exercises, and more. Visit https://www.jmp.com/statisticalthinking.

New User Welcome Kit

The New User Welcome Kit is designed to help you quickly get comfortable with the basics of JMP. You'll complete its thirty short demo videos and activities, build your confidence in using the software, and connect with the largest online community of JMP users in the world. Visit https://www.jmp.com/welcome.

Statistics Knowledge Portal

The Statistics Knowledge Portal combines concise statistical explanations with illuminating examples and graphics to help visitors establish a firm foundation upon which to build statistical skills. Visit https://www.jmp.com/skp.

JMP Training

SAS offers training on a variety of topics led by a seasoned team of JMP experts. Public courses, live web courses, and on-site courses are available. You might also choose the online e-learning subscription to learn at your convenience. Visit https://www.jmp.com/training.

JMP Books by Users

Additional books about using JMP that are written by JMP users are available on the JMP website. Visit https://www.jmp.com/books.

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 macOS) **> 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 macOS, 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.

Chapter 2

Introduction to Consumer Research

Overview of Customer and Behavioral Research Methods

Consumer Research 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. 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. 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. 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. It can do this 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. 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. See Chapter 7, “Multiple Factor Analysis”.

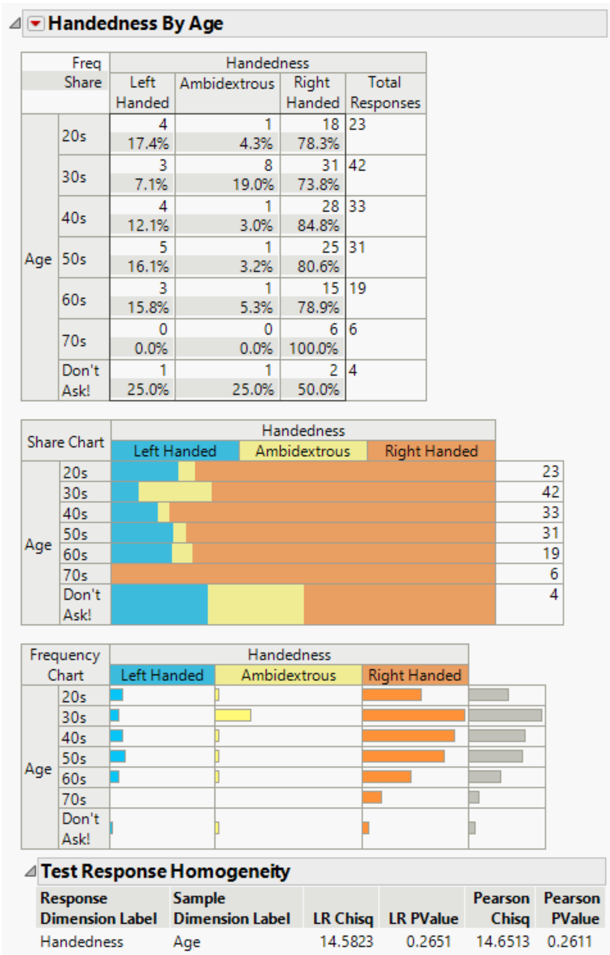
Chapter 3

Categorical Response Analysis

Analyze Survey and Other Counting Data

The Categorical platform enables you to tabulate, chart, and compare categorical response data, including multiple response data. You can analyze data from surveys and other categorical response data, such as defect records and study participant demographics. Many different data types and formats are supported.

Figure 3.1 Categorical Analysis Example



Handedness By Age

Freq / Share		Left Handed	Ambidextrous	Right Handed	Total Responses
Age	20s	4 17.4%	1 4.3%	18 78.3%	23
	30s	3 7.1%	8 19.0%	31 73.8%	42
	40s	4 12.1%	1 3.0%	28 84.8%	33
	50s	5 16.1%	1 3.2%	25 80.6%	31
	60s	3 15.8%	1 5.3%	15 78.9%	19
	70s	0 0.0%	0 0.0%	6 100.0%	6
	Don't Ask!	1 25.0%	1 25.0%	2 50.0%	4

Test Response Homogeneity

Response Dimension Label	Sample Dimension Label	LR Chisq	LR PValue	Pearson Chisq	Pearson PValue
Handedness	Age	14.5823	0.2651	14.6513	0.2611

Contents

Example of the Categorical Platform 27
Launch the Categorical Platform 33
  Response Roles 33
  Columns Roles 38
  Other Launch Window Options 38
The Categorical Report 40
Categorical Platform Options 41
  Report Options 41
  Statistical Testing Options 42
  Additional Categorical Platform Options 44
  Crosstab Table Options 47
Comparison Letters 48
Supercategories 49
  Supercategories Options 50
Additional Examples of the Categorical Platform 51
  Example of the Test for Response Homogeneity 51
  Example of the Multiple Response Test 52
  Example of Supercategories 54
  Example of the Cell Chisq Test 57
  Example of Compare Each Sample with Comparison Letters 58
  Example of Compare Each Cell with Comparison Letters 59
  Example of User-Specified Comparison with Comparison Letters 61
  Example of Aligned Responses 62
  Example of Conditional Association and Relative Risk 64
  Example of Rater Agreement 66
  Example of Repeated Measures 67
  Example of the Multiple Responses 68
  Example of Mean Score with Comparison Letters 73
  Example of a Structured Report 75
  Example of a Multiple Response with Nonresponse 76
Set Preferences 79
Statistical Details for the Categorical Platform 80
  Rao-Scott Correction 80

Example of the Categorical Platform

This example uses the Color Preference Survey.jmp sample data table, which contains survey data on color preferences. You can use the Categorical platform to summarize results from different question types. You are interested in the following types of questions:

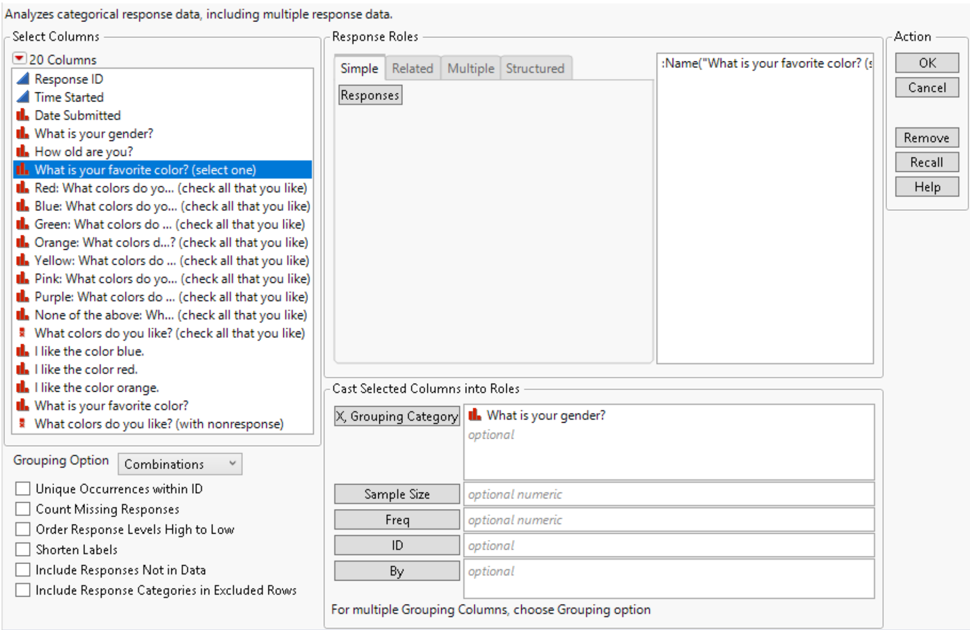
- A single response question
- Three aligned ranking questions
- A multiple response question
- A single response question segmented by a secondary question

You will set up four analyses (one for each type of question) in the Categorical launch window. Alternatively, you can use one launch tab at a time.

1. Select **Help > Sample Data Library** and open Color Preference Survey.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select What is your favorite color? (select one) and click **Responses**.
4. Select What is your gender? and click **X, Grouping Category**.

Note: If you want to run the single analysis of favorite color by gender, click **OK** now.

Figure 3.2 Completed Simple Tab in the Categorical Launch Window

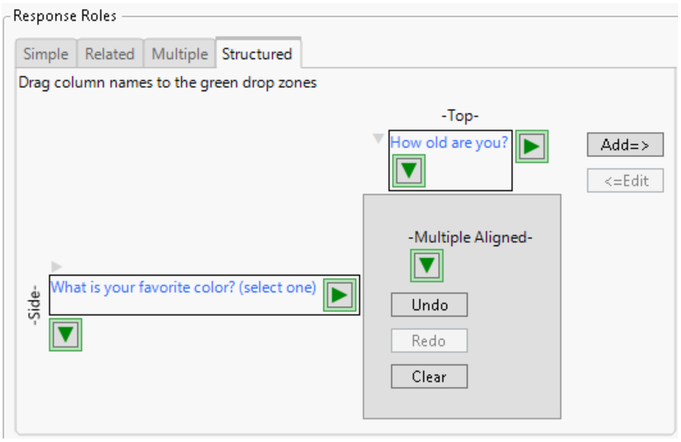


5. Select the **Related** tab.
6. Select I like the color blue. through I like the color orange. and click **Aligned Responses**.

   This enters three questions with rating scales to be analyzed together. The gender grouping category applies to this analysis.
7. Select the **Multiple** tab.
8. Select Red: What colors do you like? (check all that you like) through None of the above: What colors do you like? (check all that you like) and click **Multiple Response**.

   This enters the multiple response question "What colors do you like?" where each response is in an individual column for analysis. The gender grouping category applies to this analysis.
9. Select the **Structured** tab.
10. Drag What is your favorite color? (select one) to the Side green arrow and drag How old are you? to the Top green arrow.

Figure 3.3 Structured Tab in Categorical Launch Window

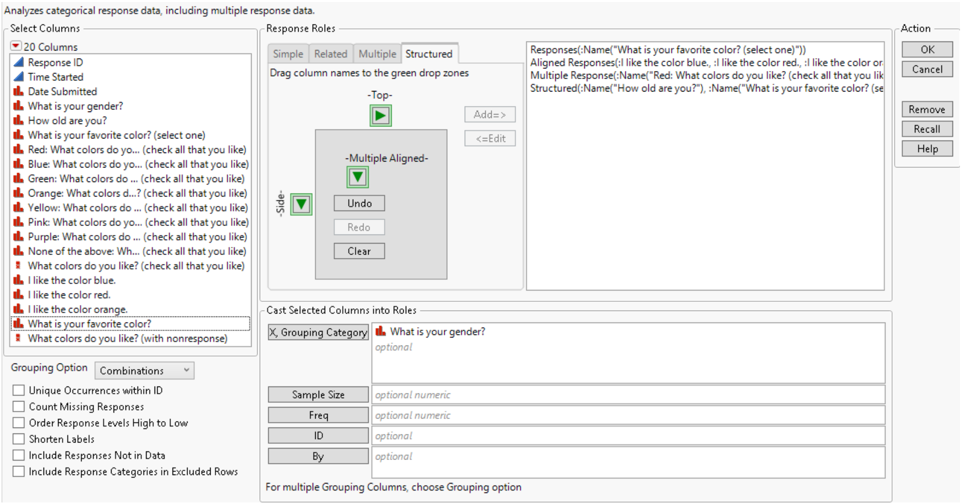


11. Click **Add**.

    The What is your gender? column specified in the X. Grouping Category role does not apply to the structured analysis.

Tip: Click on the green arrows in the structured tab for a column list, where you can click to add a column.

Figure 3.4 Completed Launch Window for Full Example



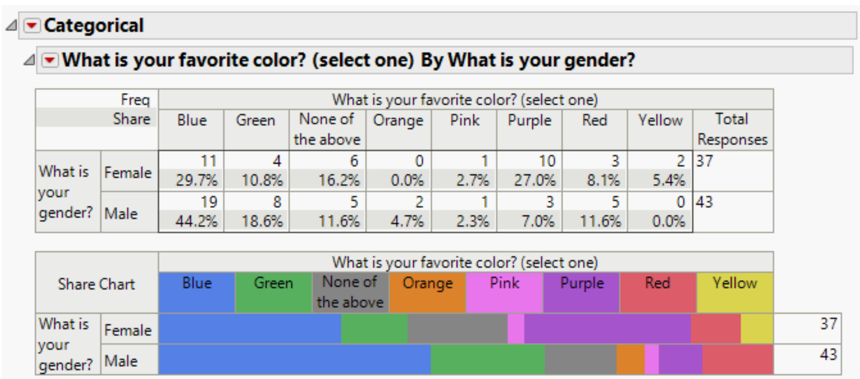
12. Click **OK**.

The results for each analysis are stacked vertically in a single report window.

Simple Tab Report

The first section shows the results for the simple response survey question, “What is your favorite color?”, grouped by gender. Each respondent was asked to select one color from the list: Red, Blue, Green, Orange, Yellow, Pink, Purple, or None of the above.

Figure 3.5 Simple Response: Favorite Color by Gender



Freq / Share		What is your favorite color? (select one)								
		Blue	Green	None of the above	Orange	Pink	Purple	Red	Yellow	Total Responses
What is your gender?	Female	11 29.7%	4 10.8%	6 16.2%	0 0.0%	1 2.7%	10 27.0%	3 8.1%	2 5.4%	37
	Male	19 44.2%	8 18.6%	5 11.6%	2 4.7%	1 2.3%	3 7.0%	5 11.6%	0 0.0%	43

The analysis shows that blue is the favorite color for both genders. For males, the frequency of the 43 male respondents selecting blue is 19. This corresponds to a share of 44.2% or 19/43. For females, 11 of 37 female respondents (frequency) or 29.7% (share) selected blue.

Tip: You can define the colors in the charts using the Value Colors column property. See The Column Info Window chapter in *Using JMP*.

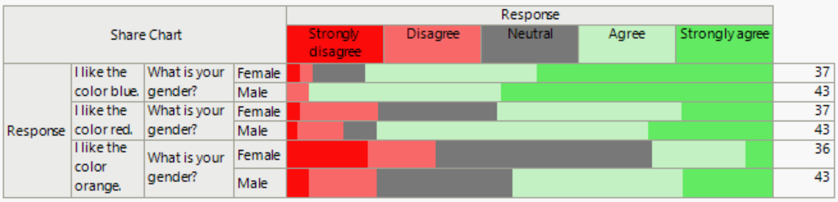
Related Tab Report

The second section of the report shows the results for three related questions that asked respondents to rank how much they liked a color. The responses to these questions are aligned, because the questions used the same rating scale of Strongly agree to Strongly disagree.

Figure 3.6 Aligned Response: Ranking Colors by Gender

Aligned Responses(I like the color blue., I like the color red., I like the color orange.) By What is your gender?

Freq Share			Response: Strongly disagree	Disagree	Neutral	Agree	Strongly agree	Total Responses
Response: I like the color blue.	What is your gender?	Female	1 2.7%	1 2.7%	4 10.8%	13 35.1%	18 48.6%	37
		Male	0 0.0%	2 4.7%	0 0.0%	17 39.5%	24 55.8%	43
I like the color red.	What is your gender?	Female	1 2.7%	6 16.2%	9 24.3%	14 37.8%	7 18.9%	37
		Male	1 2.3%	4 9.3%	3 7.0%	24 55.8%	11 25.6%	43
I like the color orange.	What is your gender?	Female	6 16.7%	5 13.9%	16 44.4%	7 19.4%	2 5.6%	36
		Male	2 4.7%	6 14.0%	12 27.9%	15 34.9%	8 18.6%	43



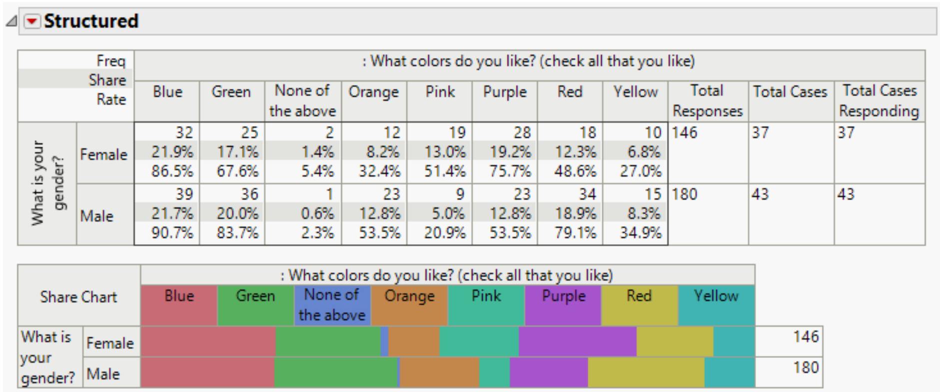
The analysis shows that blue is the color with the highest rankings for both males and females. Orange has a high number of neutral results for both genders as compared to red and blue. The value colors column property is used to define the colors used in the share chart.

Note: If you analyze aligned questions individually using the Simple tab, you obtain three individual reports containing the same information as the related report. Using the Related tab aligns the reports so that the results are easier to compare to one another.

Multiple Tab Report

The third section in the report shows the results from a multiple choice survey question grouped by gender. The question was "What colors do you like? (check all that you like)". Each possible response was collected in an individual column. If the subject selected a response, there is a value in the corresponding column. Otherwise, the column is empty.

Figure 3.7 Multiple Response: Colors Liked by Gender


Structured

What is your gender?	Freq / Share / Rate	Blue	Green	None of the above	Orange	Pink	Purple	Red	Yellow	Total Responses	Total Cases	Total Cases Responding
		: What colors do you like? (check all that you like)										
Female	Freq	32	25	2	12	19	28	18	10	146	37	37
	Share	21.9%	17.1%	1.4%	8.2%	13.0%	19.2%	12.3%	6.8%			
	Rate	86.5%	67.6%	5.4%	32.4%	51.4%	75.7%	48.6%	27.0%			
Male	Freq	39	36	1	23	9	23	34	15	180	43	43
	Share	21.7%	20.0%	0.6%	12.8%	5.0%	12.8%	18.9%	8.3%			
	Rate	90.7%	83.7%	2.3%	53.5%	20.9%	53.5%	79.1%	34.9%			

From the results, we observe that blue and green are highly liked by both genders. Females like pink and purple at higher rates than males.

In a multiple response question, more than one answer is allowed. A case represents a single responder, and Total Cases is the total number of responders. In our data, we have 37 female cases and 43 male cases. The total number of cases responding are the number of cases who selected one or more responses (which in our data, was everyone).

The Total Responses column lists the total number of selections made. For females, there were 146 colors selected by the 37 females. For males, there were 180 colors selected by the 43 male responders. The tabulation includes the number of responders selecting each color, the share or the percent of the total responses, and the rate or the percent of the total cases.

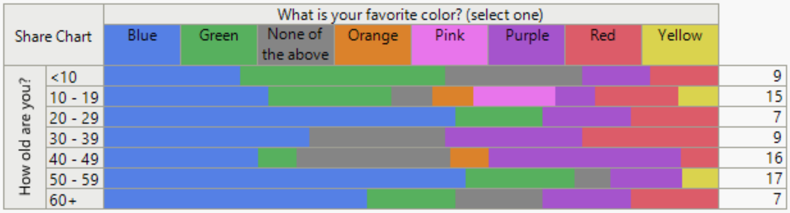
Structured Tab Report

The final section shows the table generated from the structured tab where you set the favorite color question as the response of interest grouped by the age group. In the structured format, the levels of the primary response of interest form the rows of the table and the levels of the grouping variable form the columns.

Figure 3.8 Structured: Favorite Color by Age

What is your favorite color? (select one) By How old are you?

What is your favorite color? (select one)	Freq / Share	<10	10 - 19	20 - 29	30 - 39	40 - 49	50 - 59	60+
Blue	Freq	2	4	4	3	4	10	3
	Share	22.2%	26.7%	57.1%	33.3%	25.0%	58.8%	42.9%
Green	Freq	3	3	1	0	1	3	1
	Share	33.3%	20.0%	14.3%	0.0%	6.3%	17.6%	14.3%
None of the above	Freq	2	1	0	2	4	1	1
	Share	22.2%	6.7%	0.0%	22.2%	25.0%	5.9%	14.3%
Orange	Freq	0	1	0	0	1	0	0
	Share	0.0%	6.7%	0.0%	0.0%	6.3%	0.0%	0.0%
Pink	Freq	0	2	0	0	0	0	0
	Share	0.0%	13.3%	0.0%	0.0%	0.0%	0.0%	0.0%
Purple	Freq	1	1	1	2	5	2	1
	Share	11.1%	6.7%	14.3%	22.2%	31.3%	11.8%	14.3%
Red	Freq	1	2	1	2	1	0	1
	Share	11.1%	13.3%	14.3%	22.2%	6.3%	0.0%	14.3%
Yellow	Freq	0	1	0	0	0	1	0
	Share	0.0%	6.7%	0.0%	0.0%	0.0%	5.9%	0.0%
Total Responses		9	15	7	9	16	17	7

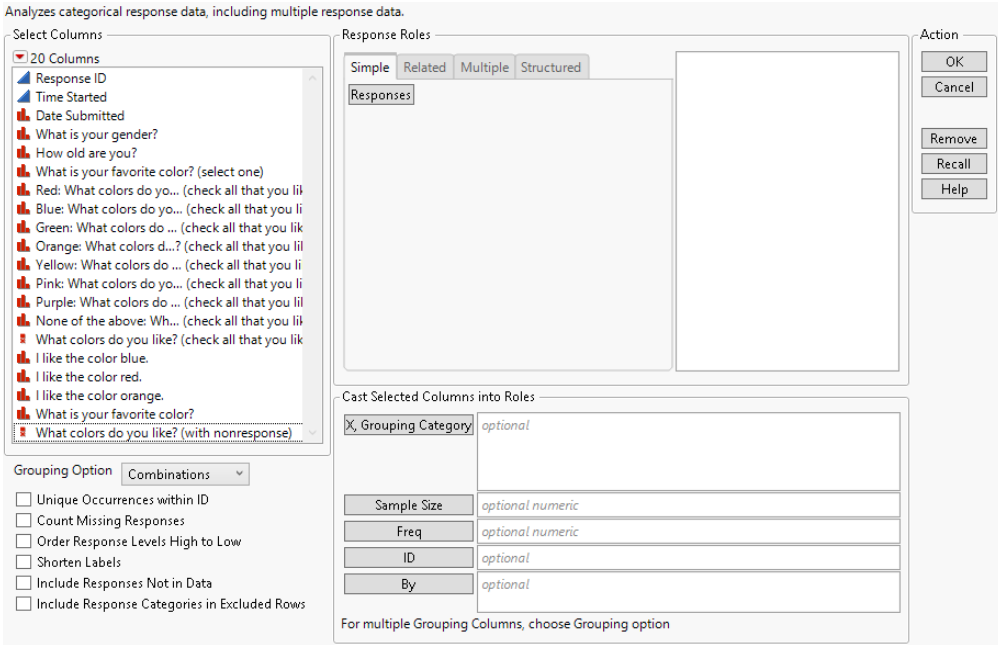


The 10 to 19 age group has two responders with pink as their favorite color. No other age groups have responders with pink as their favorite color. You can see from the bottom row of the table that the number of responders in each age group is small. You could gather more data to draw conclusions about the favorite colors across age groups.

Launch the Categorical Platform

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

Figure 3.9 Categorical Platform Launch Window for the Simple Tab



For more information about the options in the Select Columns red triangle menu, see the Get Started chapter in *Using JMP*.

Response Roles

The launch window includes tabs for three groups 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

Use the Simple tab to analyze a single response, such as a survey question where only a single answer is allowed. The data to be analyzed is contained in a single column.

Responses Summarizes data from a single column. If multiple columns are selected, the categorical report contains a separate report for each individual column.

Related Tab

Use the Related tab to analyze a set of related responses across multiple columns.

Figure 3.10 Categorical Platform Launch Window Related Tab

Aligned Responses Summarizes data from multiple columns that have the same response levels in a single report. This option is useful for survey data where 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 to the same question made at different time points. If 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 or item, but is given by a different individual (rater).

Multiple Tab

Use the Multiple tab to analyze multiple responses where the responses are recorded in one or more columns. The options on the Multiple tab are specific to how the data is organized in your data table. 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), or from a set of defect data where an item can have multiple defects.

Figure 3.11 Categorical Platform Launch Window Multiple Tab

Multiple Response Summarizes data from multiple responses where each possible response is recorded in its own individual column. Each column can contain blanks, which correspond to the item not being selected.

Figure 3.12 Multiple Response Column Format

	Red: What colors do you like? (check all that you like)	Blue: What colors do you like? ...	Green: What colors do you ...	Orange: What colors do you ...	Yellow: What colors do you ...	Pink: What colors do you like? ...	Purple: What colors do you ...	None of the above: What colors do you like? (check all that you like)
1	Red	Blue				Pink	Purple	
2	Red	Blue	Green					
3	Red	Blue	Green	Orange			Purple	
4								None of the above
5		Blue	Green	Orange			Purple	
6	Red	Blue	Green				Purple	
7	Red	Blue		Orange				
8	Red		Green	Orange				
9	Red	Blue	Green				Purple	
10	Red	Blue	Green	Orange	Yellow	Pink	Purple	

Multiple Response by ID Summarizes data from multiple responses where the data are recorded in a stacked format. There is a single column of responses with a second column containing an ID for the subject.

Figure 3.13 Multiple Response by ID Format

	Response ID	What colors do you like?
1	2	Red
2	2	Blue
3	2	
4	2	
5	2	
6	2	Pink
7	2	Purple
8	2	
9	3	Red
10	3	Blue

Multiple Delimited Summarizes data from multiple responses where the responses are in a single column and each response is separated by a comma, semicolon, or tab.

Figure 3.14 Delimited Multiple Response Format

	What colors do you like? (check all that you like)
1	Red, Blue, , , , Pink, Purple,
2	Red, Blue, Green, , , , ,
3	Red, Blue, Green, Orange, , , Purple,
4	, , , , , , , None of the above
5	, Blue, Green, Orange, , , Purple,
6	Red, Blue, Green, , , , Purple,
7	Red, Blue, , Orange, , , ,
8	Red, , Green, Orange, , , ,
9	Red, Blue, Green, , , , Purple,
10	Red, Blue, Green, Orange, Yellow, Pink, Purple,

Tip: Use the Multiple Responses column property for columns that contain delineated multiple responses. See The Column Info Window chapter in *Using JMP*.

Indicator Group Summarizes multiple responses that are stored in indicator columns. The data table contains a column for each possible response, and each column is an indicator (for example, 0 and 1). A blank value indicates a missing response. See "Indicator Group" on page 71.

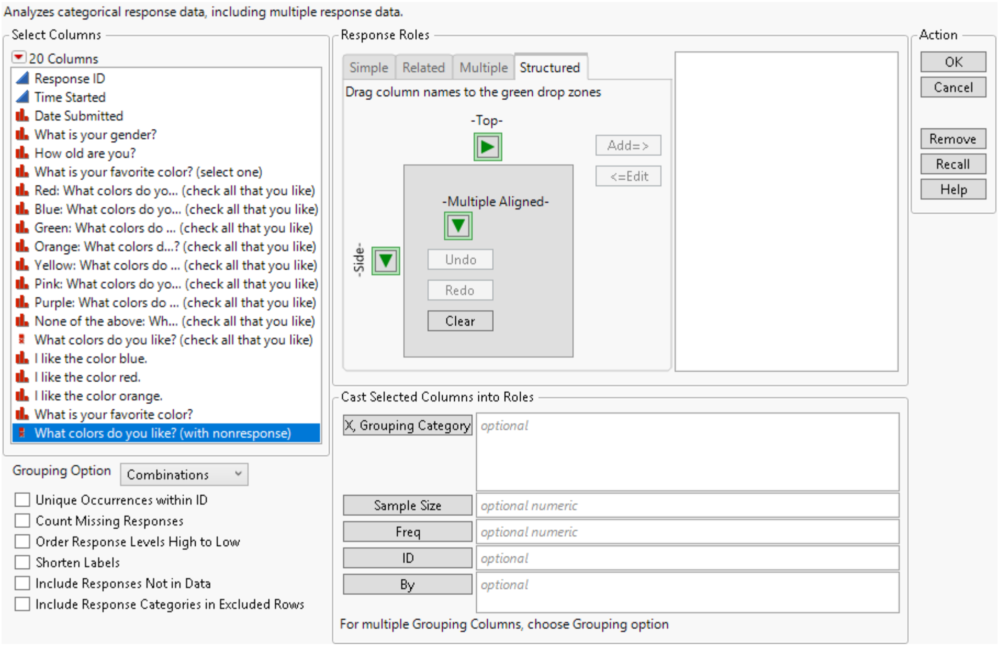
Response Frequencies Summarizes multiple responses that are stored in columns that contain frequency counts. This data format is the summarized version of the Indicator Group format. See "Response Frequencies" on page 72.

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 *Basic Analysis*.

Structured Tab

Use the Structured tab to construct custom cross tabulations.

Figure 3.15 Categorical Platform Launch Window for the Structured Tab



The Structured tab has three drop zones for assigning columns to roles. Drag column names to the green drop zone arrows. Alternatively, click the green arrows for a column list to select columns. The resulting structured table considers the innermost terms on the side of the table as responses and all other terms as grouping factors.

Drop Zones

Top Assigns one or more data table columns to the columns of the cross tabulation table. After one column has been assigned, additional columns can be nested within already assigned columns or added as additional column groups.

Side Assigns one or more data table columns to the rows of the cross tabulation table. After one column has been assigned, additional columns can be nested within already assigned rows or assigned as additional row groups.

Multiple Aligned Assigns two or more columns with aligned response.

Controls

Undo Click to undo column assignments.

Redo Click to redo the most recent assignment that was undone.

Clear Click to remove all assignments in the drop zones.

Add=> Click to add the constructed table to the structured table list.

<=Edit Click to make changes to the selected table from the structured table list.

Columns Roles

The following roles are available:

X, Grouping Category (Not applicable for the Structured tab.) Assigns a column as a grouping category. The responses are summarized for each group. If more than one grouping column is used, then the tabulation is nested by default. Use the Grouping Option in the launch window to change the summarization.

Sample Size Assigns a column whose values define the number of individual units in the group to which that frequency is applicable. The sample size is used for multiple response roles with summarized data. See "Example of the Multiple Responses" on page 68

Freq Assigns a column whose values define a frequency to each row for the analysis. The frequency role is used for summarized data.

ID Assigns a column that identifies the respondent. This option is required when Multiple Response by ID is selected, and it is not used for any 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 in the platform report window.

Grouping Option Defines how to use grouping variables in the 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 in the cross tabulation table.

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 Limits the counts to unique response levels within a participant. Specify a column as the ID using the ID role. This option is applicable only when Multiple Response by ID is selected, and it is not used for any other response types.

Count Missing Responses Specifies that missing values be included 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. For multiple responses, a response is considered missing if all response categories are missing or for indicators, if all are zero.

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, and does not apply to grouping categories.

Tip: Use the Value Order 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, and does not apply to column names.

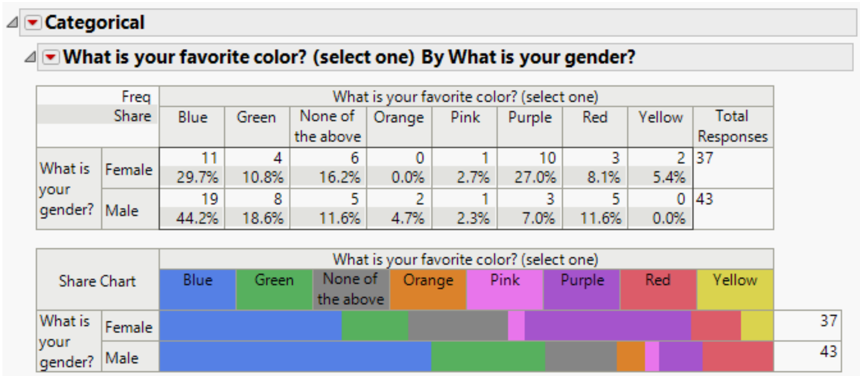
Include Responses Not in Data Specifies that categories with no responses be included in the report. The categories with no responses must be specified in the Value Labels column property. This option applies only to responses. For grouping variables tabulations, include only categories with responses.

Include Response Categories in Excluded Rows Specifies that response categories that appear only in excluded rows be included in the report. The counts for these categories are zero.

The Categorical Report

The initial Categorical report shows a cross tabulation and a share chart for each set of selected responses.

Figure 3.16 The Initial Categorical Report



Categorical

What is your favorite color? (select one) By What is your gender?

	Freq / Share	Blue	Green	None of the above	Orange	Pink	Purple	Red	Yellow	Total Responses
		What is your favorite color? (select one)								
What is your gender?	Female	11 29.7%	4 10.8%	6 16.2%	0 0.0%	1 2.7%	10 27.0%	3 8.1%	2 5.4%	37
	Male	19 44.2%	8 18.6%	5 11.6%	2 4.7%	1 2.3%	3 7.0%	5 11.6%	0 0.0%	43

The upper left corner of the table lists the quantities (Freq, Share, and Rate when applicable) that are included in each cell of the table. Remove or add these quantities using the options in the Categorical red triangle menu.

- The Frequencies count (labeled 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 (Freq) by the total number of responses.
- The Rate is the frequency of response (Freq) divided by the total number of cases (Total Cases). This quantity appears only for multiple responses and is not shown in Figure 3.16.

In Figure 3.16, the number of responses to the question What is your favorite color? are tabulated by gender. Consider the first row of the table with the results for females:

- The first cell of the table contains 11 responses. This is the count of the female respondents who selected blue as their favorite color.
- There are 37 total responses for females. Of the 37 female responses, 29.7% (11/37) selected blue as their favorite color.

Categorical Platform Options

The Categorical red triangle menu contains options that enable you to customize the report and request various statistical tests. The specific options that are available are determined by the response roles, the use of grouping categories, and the options selected in the launch window.

- "Report Options"
- "Statistical Testing Options"
- "Additional Categorical Platform Options"
- "Crosstab Table Options"

Report Options

The following options enable you to customize the appearance of the report:

Cross Tabulation Options

Frequencies Shows or hides the frequency (labeled Freq) in the Crosstab table. The frequency is the count of the responses in each category.

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

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

Rate per Case Responding (Available only for multiple responses.) Shows or hides the 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.

Chart Options

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.

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 represent the total number of responses in each grouping category.

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.

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*.

Crosstab Viewing Options

Crosstab Shows or hides the Crosstab table. The Crosstab table displays the response categories as column headings and 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, Rate, and Rate per Case Responding). If the report contains a transposed Crosstab table, this option also removes the transposed Crosstab table from the report.

Crosstab Transposed Shows or hides a transposed Crosstab table. The transposed Crosstab table displays the response categories as row labels and 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, Rate, and Rate per Case Responding). If the report contains a Crosstab table, this option also 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 for multiple responses, response homogeneity, association, relative risk, and agreement. Options depend on both the response data type and the grouping variable (X, Grouping Category) data type.

Test Multiple Response (Available only for multiple response data with one or more grouping categories.) See "Example of the Multiple Response Test" on page 52. 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).

Exclude Nonresponses Excludes nonresponses for count and homogeneity tests of multiple response categories. If a row is missing data in all response categories, it is treated as a nonresponse.

Test Response Homogeneity (Available only for response variables that do not have a multiple response modeling type and when one or more groupings variables are specified.) Shows or hides a report that contains tests for response homogeneity that depend on your grouping variable:

– When the grouping variable is not a multiple response, then one tests for independence of the response across the grouping variable levels. The likelihood ratio and Pearson chi-square tests are provided. See "Example of the Test for Response Homogeneity" on page 51.

– When the grouping variable is a multiple response, then one tests for independence of the response across *each* of the grouping variable levels. 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 57.

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 58.

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 59.

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 64.

Conditional Association (Available only when the Unique occurrence 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 64.

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 66.

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 67.

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 warnings for small sample sizes in 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

The following options enable you to add summary statistics to the report, save reports, and set report formats.

Summary Statistics Options

Total Responses Shows or hides the sum of the frequency counts for the response in crosstab tables and share charts. When a grouping variable is specified, the total is across each grouping category.

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 49.

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. Alternatively, 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. Participants 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 73.

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 the TTEST Procedure chapter in SAS Institute Inc. (2018). The results of the comparison are shown using letters. For more information about comparison letters, see "Comparison Letters" on page 48. For more information about specifying comparison groups, see "Example of User-Specified Comparison with Comparison Letters" on page 61.

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 more than one response is specified and there are no grouping variables in the analyses.) Orders the response reports by the mean score.

Save Options

Save Tables Contains options to save 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. The saved tables all include a Source script.

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 (Available only on Windows.) Creates a Microsoft Excel spreadsheet with the structure of the crosstab table. This option maps all of the tables to one sheet, with the response categories as rows, the sample levels as columns, and 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.

Report Format Options

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

Contents Summary Shows or hides a Contents Summary report at the top of the Categorical report. The Contents Summary report contains all of the tests and mean scores in a summary that has links to the associated report.

Show Columns Used in Report (Available only with SPSS or SAS names.) 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.

Tip: 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, Rates, and Means.

Arrange in Rows Enables you to arrange multiple reports across the window. 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 and in future JMP sessions. See "Set Preferences" on page 79.

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 available 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 38.

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 *Using JMP* 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.

Crosstab 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 61.

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 comparison letters to identify sample levels. For more than 26 levels, numbers are appended to the letters. The letters are shown to the right of the sample level headings of the crosstab table when a comparison option is turned on. These letters are used to identify levels in the Compare column.

Figure 3.17 Crosstab Table with Comparison Letters

country By type

Freq Share			country				
			American	European	Japanese	Total Responses	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

A letter in the Compare column indicates a difference between two levels. The row containing the letter is one level and the letter in the Compare column indicates the second level. When 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. For example, in Figure 3.17, notice the following:

- The first row of the table for Family cars has a B in the Compare column. The letter B is associated with Sporty cars. This 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 Sporty cars (100) is less than the total responses for Family cars (155).

- The c in the Compare column of the Sporty row indicates a significant difference at the 0.10 level between the country of origin when comparing Sporty to Work cars. The c is in the row for Sporty cars because the total responses for Sporty cars (100) is greater than the total responses for Work cars (48).

Warnings for small counts are indicated by asterisks in the comparison cells. One asterisk indicates that the level has fewer than 100 responses. Two asterisks indicate fewer than 30 responses. In Figure 3.17, notice that the row for Work has an asterisk in the Compare column. This warning that the count of responses is small appears because the total responses for Work cars (48) is less than100.

You can change the alpha levels and thresholds for the warning counts in the Categorical platform preferences. For more information about changing preferences, see "Set Preferences" on page 79.

Tip: If you want only one set of comparison letters in your report, set the **Lowercase Alpha Level** to 0 in the preferences.

The following examples illustrate the use of comparison letters:

- "Example of Compare Each Sample with Comparison Letters" on page 58
- "Example of Compare Each Cell with Comparison Letters" on page 59
- "Example of Mean Score with Comparison Letters" on page 73

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 from the top two ratings (top two boxes). Use the Supercategories column property to define groups of responses.

When you add a Supercategories column property to a response column, no additional columns are added to your data table. Instead, the Supercategories column property adds an additional category column to crosstab tables and Frequency Charts in the Categorical platform report. You can create multiple supercategories for a single response column. Share Charts do not show supercategories, and supercategories are not applied to grouping columns.

To create a supercategory:

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 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.

Additional Examples of the Categorical Platform

- "Example of the Test for Response Homogeneity"
- "Example of the Multiple Response Test"
- "Example of Supercategories"
- "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 Aligned Responses"
- "Example of Conditional Association and Relative Risk"
- "Example of Rater Agreement"
- "Example of Repeated Measures"
- "Example of the Multiple Responses"
- "Example of Mean Score with Comparison Letters"
- "Example of a Structured Report"
- "Example of a Multiple Response with Nonresponse"

Example of the Test for Response Homogeneity

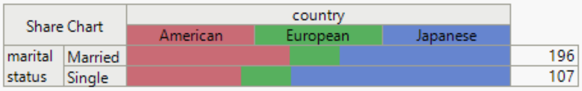
This example uses the Car Poll.jmp sample data table, which contains data collected from a survey about car ownership. The data include demographics about the individuals polled and information about their car. You want to explore the relationship between marital status and the origin of the car. You 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.

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 3.18 Test Response Homogeneity

country By marital status

		country			
	Freq Share	American	European	Japanese	Total Responses
marital status	Married	83 42.3%	26 13.3%	87 44.4%	196
	Single	32 29.9%	14 13.1%	61 57.0%	107



Test Response Homogeneity

Response Dimension Label	Sample Dimension Label	LR Chisq	LR PValue	Pearson Chisq	Pearson PValue
country	marital status	5.14007	0.0765	5.08146	0.0788

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 provides results from two versions of the test. The Pearson Test and the Likelihood Ratio Test both have chi-square test statistics and associated *p*-values. The test for response homogeneity has a *p*-value of about 0.08 for either method.

Example of the Multiple Response Test

This example uses the Consumer Preferences.jmp sample data table. This table contains data from a survey about 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) are 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. Scroll to the right until you see the Brush Delimited column.

Figure 3.19 Consumer Preferences Data Table

	Brush After Waking Up	Brush After Meal	Brush Before Sleep	Brush Another Time	Brush Other	Brush Delimited
1	1	0	0	0		Wake,
2	0	1	0	0		After Meal,
3	1	0	1	0		Wake,Before Sleep,
4	0	1	0	0		After Meal,
5	1	0	1	0		Wake,Before Sleep,
6	1	1	1	0		Wake,After Meal,Before Sleep,
7	1	1	1	0		Wake,After Meal,Before Sleep,

Note that the Brush Delimited column contains the responses to a multiple response question. Each response is separated by a comma. The four columns preceding Brush Delimited contain the same information in a different data format. Each column (Brush after Waking Up, Brush After Meal, Brush Before Sleep, and Brush Another Time) contains one response. If the response was selected the column value is a 1, and otherwise it is 0.

3. Select **Analyze > Consumer Research > Categorical**.
4. Select the **Multiple** tab.
5. Select Brush Delimited and click **Multiple Delimited** on the Multiple tab.

Tip: Alternatively, you can select Brush after Waking Up, Brush After Meal, Brush Before Sleep, and Brush Another Time and click **Indicator Group** on the Multiple tab

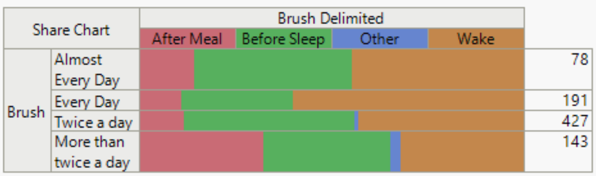
---

6. Select Brush and click **X, Grouping Category**.
7. Click **OK**.
8. Click the Categorical red triangle and select **Test Multiple Response > Count Test, Poisson**.

Figure 3.20 Test Multiple Response, Poisson

Multiple Delimited(Brush Delimited) By Brush

	Freq / Share / Rate	Brush Delimited: After Meal	Before Sleep	Other	Wake	Total Responses	Total Cases	Total Cases Responding
Brush	Almost Every Day	11	32	0	35	78	48	48
		14.1%	41.0%	0.0%	44.9%			
		22.9%	66.7%	0.0%	72.9%			
	Every Day	21	55	0	115	191	142	142
		11.0%	28.8%	0.0%	60.2%			
		14.8%	38.7%	0.0%	81.0%			
	Twice a day	49	189	4	185	427	209	209
		11.5%	44.3%	0.9%	43.3%			
		23.4%	90.4%	1.9%	88.5%			
	More than twice a day	46	47	4	46	143	49	49
		32.2%	32.9%	2.8%	32.2%			
		93.9%	95.9%	8.2%	93.9%			



Test Each Response, Poisson

Brush, Brush Delimited

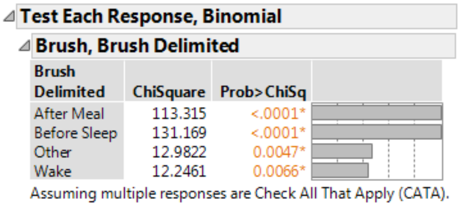
Brush Delimited	ChiSquare	Prob>ChiSq
After Meal	59.5436	<.0001*
Before Sleep	39.1097	<.0001*
Other	12.7131	0.0053*
Wake	1.8793	0.5978

Chi-squared tests use Poisson rates.

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. The bar graph to the right of the Prob>ChiSq column plots the *p*-values on a -Log10(p) scale. 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.

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

Figure 3.21 Test Multiple Response, Binomial



Test Each Response, Binomial

Brush, Brush Delimited

Brush Delimited	ChiSquare	Prob>ChiSq
After Meal	113.315	<.0001*
Before Sleep	131.169	<.0001*
Other	12.9822	0.0047*
Wake	12.2461	0.0066*

Assuming multiple responses are Check All That Apply (CATA).

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) differs across the age groups. The *p*-value for each response is less than 0.05.

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

Example of Supercategories

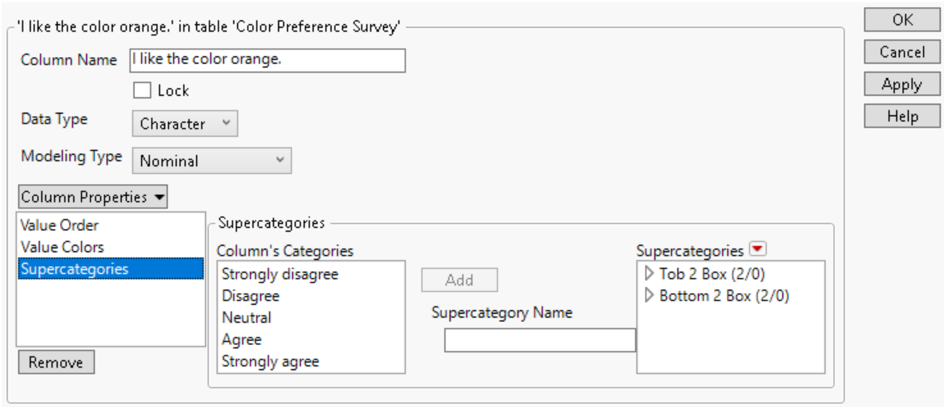
This example uses the Color Preference Survey.jmp sample data table, which contains survey data on people’s color preferences. This example illustrates the use of supercategories.

1. Select **Help > Sample Data Library** and open Color Preference Survey.jmp.
2. Select the column I like the color orange. Right-click the column heading and select **Column Properties > Supercategories**.
3. Under Supercategories, select Agree and Strongly agree for the column’s categories.
4. Under Supercategory Name, enter Top 2 Box.
5. Click **Add**.

Tip: Click on the triangle to the left of the name to show the categories included in the supercategory.

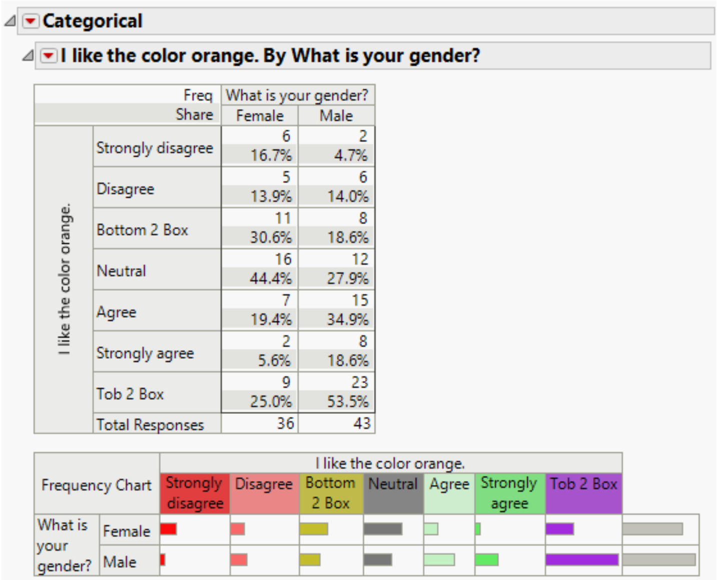
6. Under Supercategories, select Disagree and Strongly disagree for the column's categories.
7. Under Supercategory Name, enter Bottom 2 Box.
8. Click **Add**.

Figure 3.22 Supercategory Column Property



9. Click **OK**.
10. Select **Analyze > Consumer Research > Categorical**.
11. Select the **Structured** tab.
12. Click the Side green triangle and select I like the color orange.
13. Click the Top green triangle and select What is your gender?
14. Click **Add=>** and then click **OK**.
15. Click the Categorical red triangle and select **Transposed Freq Chart**.

Figure 3.23 Structured Categorical Report with Supercategories

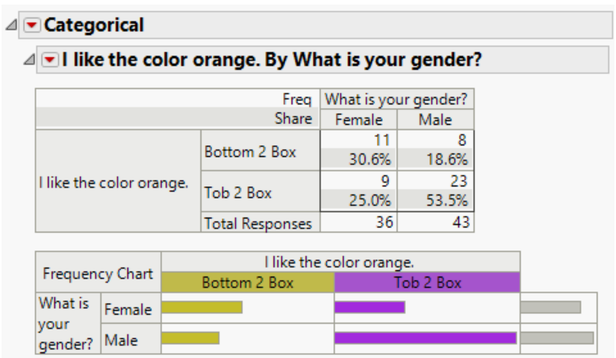


	Freq / Share	What is your gender? Female	Male
I like the color orange.	Strongly disagree	6 16.7%	2 4.7%
	Disagree	5 13.9%	6 14.0%
	Bottom 2 Box	11 30.6%	8 18.6%
	Neutral	16 44.4%	12 27.9%
	Agree	7 19.4%	15 34.9%
	Strongly agree	2 5.6%	8 18.6%
	Tob 2 Box	9 25.0%	23 53.5%
	Total Responses	36	43

The crosstab table includes two additional rows, one for each supercategory. The supercategories are also included in the Frequency Chart. Note that the frequency counts in the Top 2 Box row are the sums of the counts in the Agree and Strongly agree categories. The frequency counts in the Bottom 2 Box row are the sums of the counts in the Disagree and Strongly disagree categories.

16. Click the Categorical red triangle and deselect **Response Levels**.

Figure 3.24 Structured Categorical Report with Supercategories and No Response Levels



	Freq / Share	What is your gender? Female	Male
I like the color orange.	Bottom 2 Box	11 30.6%	8 18.6%
	Tob 2 Box	9 25.0%	23 53.5%
	Total Responses	36	43

By removing the response levels, your output now contains only the supercategories. Note that the totals are for all levels. In this example, the neutral responses are not included in the supercategories.

Example of the Cell Chisq Test

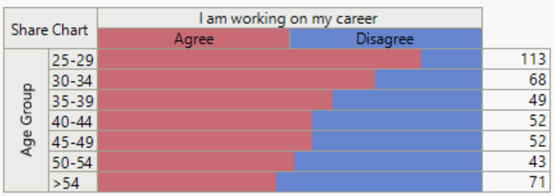
This example uses the Consumer Preferences.jmp sample data table. This table contains survey data about people's attitudes and opinions, as well as questions concerning oral hygiene. You explore the distribution of responses to the statement "I am working on my career" across age groups.

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 3.25 Cell Chisq

I am working on my career By Age Group

I am working on my career	Freq / Share / Cell Chisq PVal	Age Group 25-29	30-34	35-39	40-44	45-49	50-54	>54
	Agree	95 84.1% 0.00788	49 72.1% 0.41003	30 61.2% 0.80398	29 55.8% 0.45495	29 55.8% 0.45495	22 51.2% 0.29058	33 46.5% 0.06415
	Disagree	18 15.9% 0.00039	19 27.9% 0.27136	19 38.8% 0.74035	23 44.2% 0.31848	23 44.2% 0.31848	21 48.8% 0.15823	38 53.5% 0.01345
	Total Responses	113	68	49	52	52	43	71



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. The expected cell count is calculated as the row total times the column total divided by the overall count.

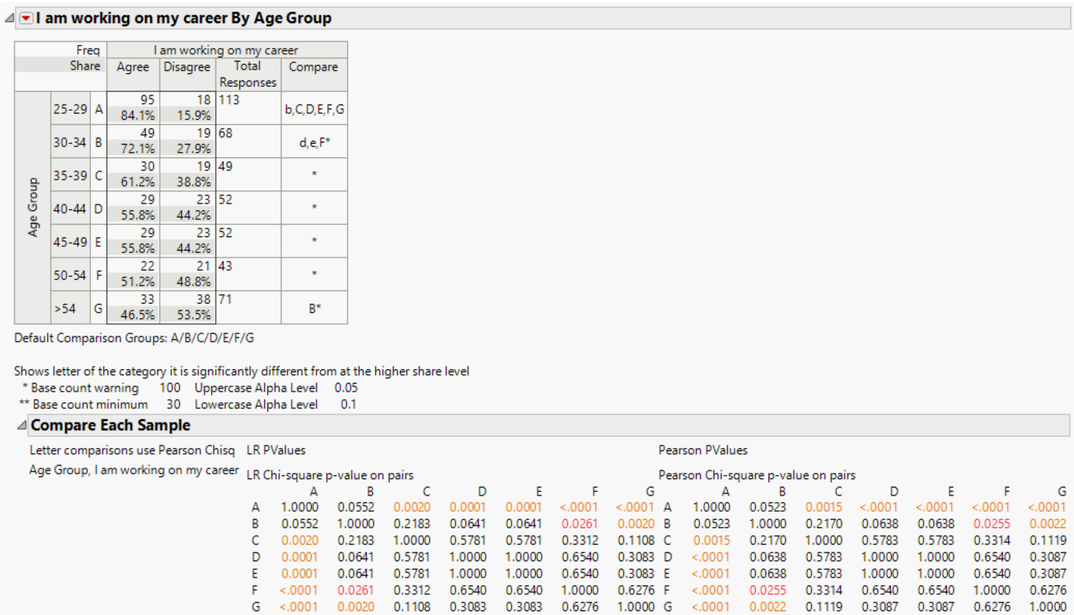
For example, the expected number of responses under the null hypothesis that the rates are equal in the 25 - 29 group who agree is (287*113)/448 = 72.4; the observed value was 95. This observed value, with a *p*-value of 0.00788, is significantly larger than the expected value. The number of responses in the 25 - 29 group who agree with "I am working on my career" is higher than expected if 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. This table contains survey data about 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 3.26 Compare Each Sample



I am working on my career By Age Group

		Freq / Share	I am working on my career: Agree	Disagree	Total Responses	Compare
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
Age Group, I am working on my career

LR PValues
LR Chi-square p-value on pairs

	A	B	C	D	E	F	G
A	1.0000	0.0552	0.0020	0.0001	0.0001	<.0001	<.0001
B	0.0552	1.0000	0.2183	0.0641	0.0641	0.0261	0.0020
C	0.0020	0.2183	1.0000	0.5781	0.5781	0.3312	0.1108
D	0.0001	0.0641	0.5781	1.0000	1.0000	0.6540	0.3083
E	0.0001	0.0641	0.5781	1.0000	1.0000	0.6540	0.3083
F	<.0001	0.0261	0.3312	0.6540	0.6540	1.0000	0.6276
G	<.0001	0.0020	0.1108	0.3083	0.3083	0.6276	1.0000

Pearson PValues
Pearson Chi-square p-value on pairs

	A	B	C	D	E	F	G
A	1.0000	0.0523	0.0015	<.0001	<.0001	<.0001	<.0001
B	0.0523	1.0000	0.2170	0.0638	0.0638	0.0255	0.0022
C	0.0015	0.2170	1.0000	0.5783	0.5783	0.3314	0.1119
D	<.0001	0.0638	0.5783	1.0000	1.0000	0.6540	0.3087
E	<.0001	0.0638	0.5783	1.0000	1.0000	0.6540	0.3087
F	<.0001	0.0255	0.3314	0.6540	0.6540	1.0000	0.6276
G	<.0001	0.0022	0.1119	0.3087	0.3087	0.6276	1.0000

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 the Compare column, which is to the right of the group label. The letters in the Compare column enable you to interpret the outcomes of the statistical test of independence among groups.

The Compare Each Sample report 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 through 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, which indicate differences that are 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 number 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. This table contains survey data about people’s attitudes and opinions, as well as questions concerning oral hygiene. You explore the distribution of the responses about job satisfaction 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 click **Responses** on the Simple tab.
4. Select Employee Tenure and click **X, Grouping Category**.
5. Click **OK**.
6. Click the Categorical red triangle and select **Compare Each Cell**.
7. Click the **Compare Each Cell - Details** gray disclosure icon.

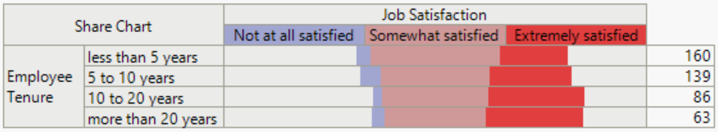
Figure 3.27 Compare Each Cell

Job Satisfaction By Employee Tenure

Freq Share Comparisons			Job Satisfaction Not at all satisfied	 Somewhat satisfied	 Extremely satisfied	 Total Responses
Employee Tenure	less than 5 years	A	11 6.9%	97 60.6% d	52 32.5%	160
	5 to 10 years	B	13 9.4%	72 51.8%	54 38.8%	139
	10 to 20 years	C	4 4.7% *	43 50.0% *	39 45.3% a*	86
	more than 20 years	D	4 6.3% *	30 47.6% *	29 46.0% a*	63

Default Comparison Groups: A/B/C/D

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 Cell - Details

Letter comparisons use Fisher's Exact Test
Employee Tenure, Job Satisfaction

LR Pairs

		AB	BB	AC	BC	CC	AD	BD	CD	DD
Not at all satisfied	1.0000	0.4323	1.0000	0.4781	0.1806	1.0000	0.8872	0.4658	0.6515	1.0000
Somewhat satisfied	1.0000	0.1246	1.0000	0.1092	0.7931	1.0000	0.0783	0.5820	0.7739	1.0000
Extremely satisfied	1.0000	0.2526	1.0000	0.0476	0.3366	1.0000	0.0607	0.3377	0.9341	1.0000

Pearson Pairs

		AB	BB	AC	BC	CC	AD	BD	CD	DD
Not at all satisfied	1.0000	0.4316	1.0000	0.4870	0.1948	1.0000	0.8878	0.4763	0.6496	1.0000
Somewhat satisfied	1.0000	0.1246	1.0000	0.1086	0.7931	1.0000	0.0774	0.5820	0.7740	1.0000
Extremely satisfied	1.0000	0.2524	1.0000	0.0465	0.3360	1.0000	0.0585	0.3364	0.9341	1.0000

Fisher Exact Pairs

		AB	BB	AC	BC	CC	AD	BD	CD	DD
Not at all satisfied	1.0000	0.5235	1.0000	0.5853	0.2990	1.0000	1.0000	0.5912	0.7221	1.0000
Somewhat satisfied	1.0000	0.1302	1.0000	0.1373	0.8909	1.0000	0.0983	0.6493	0.8685	1.0000
Extremely satisfied	1.0000	0.2764	1.0000	0.0531	0.4034	1.0000	0.0649	0.3574	1.0000	1.0000

The *p*-values for pairwise likelihood ratio chi-square, Pearson 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 between 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, the employees with fewer than 5 years of tenure are Extremely satisfied at a lower rate than the group with 20 years of tenure. This finding is 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 highest share of responses.

Example of User-Specified Comparison with Comparison Letters

This example uses the Consumer Preferences.jmp sample data table. This table contains survey data about people's attitudes and opinions. 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 click **Responses** on the Simple tab.
4. Select Employee Tenure and click **X, Grouping Category**.
5. Click **OK**.
6. Click the red triangle next to I am working on my career By Employee Tenure and select **Show Letters**.
7. Click the red triangle next to I am working on my career By Employee Tenure and select **Specify Comparison Groups**.
8. Enter A/B, B/C, C/D and click **OK**.
9. Click the Categorical red triangle and select **Test Response Homogeneity**.

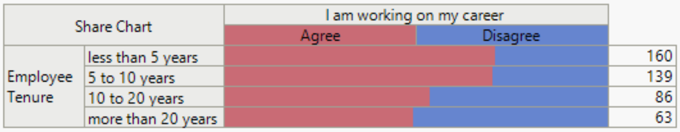
Figure 3.28 Specify Comparison Example

I am working on my career By Employee Tenure

Freq / Share / Comparisons			I am working on my career: Agree	Disagree	Total Responses
Employee Tenure	less than 5 years	A	113 70.6%	47 29.4%	160
	5 to 10 years	B	97 69.8% C	42 30.2%	139
	10 to 20 years	C	46 53.5% *	40 46.5% B*	86
	more than 20 years	D	31 49.2% *	32 50.8% *	63

Specified Comparison Groups: A/B, B/C, C/D

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



Test Response Homogeneity

Response Dimension Label	Sample Dimension Label	LR Chisq	LR PValue	Pearson Chisq	Pearson PValue
I am working on my career	AB, Employee Tenure	0.02515	0.8740	0.02515	0.8740
I am working on my career	BC, Employee Tenure	6.04171	0.0140*	6.09112	0.0136*
I am working on my career	CD, Employee Tenure	0.267	0.6054	0.26699	0.6054

The test of response homogeneity compares Group A to Group B, Group B to Group C, and Group C to Group D. Group B (5 to 10 years) respondents agree with the statement "I am working on my career" more often than those in Group C (10 to 20 years). The Pearson *p*-value for this difference in agreement rates is 0.0136.

Example of Aligned Responses

This example uses the Consumer Preferences.jmp sample data table. This table contains survey data about attitudes and opinions.

1. Select **Help > Sample Data Library** and open Consumer Preferences.jmp.
2. Scroll to see the column I am working on my career.

Figure 3.29 Consumer Preferences Data Table (Partial Table)

	I am working on my career	I want to see the world	My home needs some major ...	I have vast interests outside of work	I want to get my debt under control	I come from a large family
1	Agree	Disagree	Disagree	Agree	Disagree	Agree
2	Agree	Agree	Agree	Agree	Disagree	Agree
3	Disagree	Agree	Agree	Agree	Agree	Agree
4	Agree	Agree	Agree	Agree	Agree	Agree
5	Disagree	Agree	Disagree	Agree	Agree	Disagree

Note that there are six columns, all with the same responses: Agree and Disagree. The responses for these six columns are aligned. For this example, we analyze the columns I am working on my career and I want to see the world. First, use standardize attributes to set the value colors, value order, and modeling type for these two columns:

3. Select the column headings I am working on my career and I want to see the world. Right-click and select **Standardize Attributes**.
4. Select **Column Properties > Value Colors.** Right-click the Agree color oval and set it to green and set Disagree to red.
5. Select **Column Properties > Value Order.** Click **Reverse**.
6. Select **Attributes > Modeling Type** and **Modeling Type > Ordinal**.
7. Click **OK**.

Next, use the Categorical platform to analyze these two columns.

8. Select **Analyze > Consumer Research > Categorical**.
9. Select I am working on my career and I want to see the world.
10. Select the **Related** tab, and then click **Aligned Responses** on the Related tab.

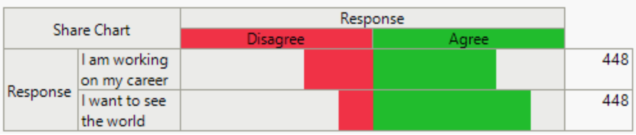
Tip: To do the same thing using the Structured tab, drag the two columns into the Multiple Aligned drop zone (green arrow).

11. Click **OK**.

Figure 3.30 Aligned Response Report

Aligned Responses(I am working on my career, I want to see the world)

Freq Share		Response: Disagree	Agree	Total Responses
Response	I am working on my career	161 35.9%	287 64.1%	448
	I want to see the world	78 17.4%	370 82.6%	448



Notice that the Share Chart is a directional bar chart. The type of bar chart you see depends on the modeling type of the columns:

- Ordinal columns result in a directional bar chart. The columns in this example are both ordinal.
- Nominal columns result in a stacked bar chart.

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.

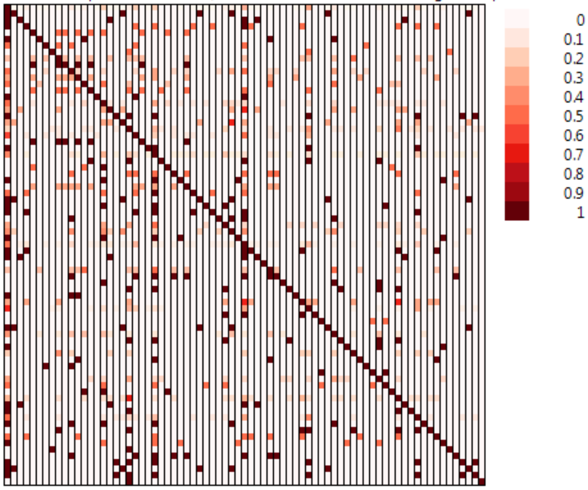
1. Select **Help > Sample Data Library** and open AdverseR.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select the Multiple tab.
4. Select ADVERSE REACTION and click **Multiple Response by ID** on the Multiple tab.
5. Select TREATMENT GROUP and click **X, Grouping Category**.
6. Select PATIENT ID and click **ID**.
7. Under the other launch window options, select **Unique Occurrences within ID** and click **OK**.
8. Click the Categorical red triangle and select **Conditional Association**.

Figure 3.31 Conditional Association Report (Partial Report)

Multiple Response by ID(ADVERSE REACTION) By TREATMENT GROUP

Conditional Association: ADVERSE REACTION

Given the response in each row, this shows the rate of also having the response in a column.



Conditional Assoc

	ABDOMINAL PAIN	ABNORMAL VISION	ALOPECIA	AMBLYOPIA	ANEMIA
ABDOMINAL PAIN	1.0000	0.0323	0.0323	0.0645	0.0000
ABNORMAL VISION	1.0000	1.0000	0.0000	0.0000	0.0000
ALOPECIA	1.0000	0.0000	1.0000	0.0000	0.0000
AMBLYOPIA	1.0000	0.0000	0.0000	1.0000	0.0000
ANEMIA	0.0000	0.0000	0.0000	0.0000	1.0000
ANGINA PECTORIS	1.0000	0.0000	0.0000	0.0000	0.0000
ANOREXIA	0.5000	0.0000	0.0000	0.0000	0.0000
ANXIETY	0.0000	0.0000	0.0000	0.0000	0.0000
ASTHENIA	0.2500	0.0000	0.0000	0.0000	0.2500
ASTHMA	0.0000	0.0000	0.0000	0.0000	1.0000
BACK PAIN	0.5000	0.0000	0.0000	0.0000	0.0000
BRONCHITIS	0.3333	0.0000	0.0000	0.0000	0.3333
CHEST PAIN	0.5000	0.0000	0.0000	0.0000	0.0000

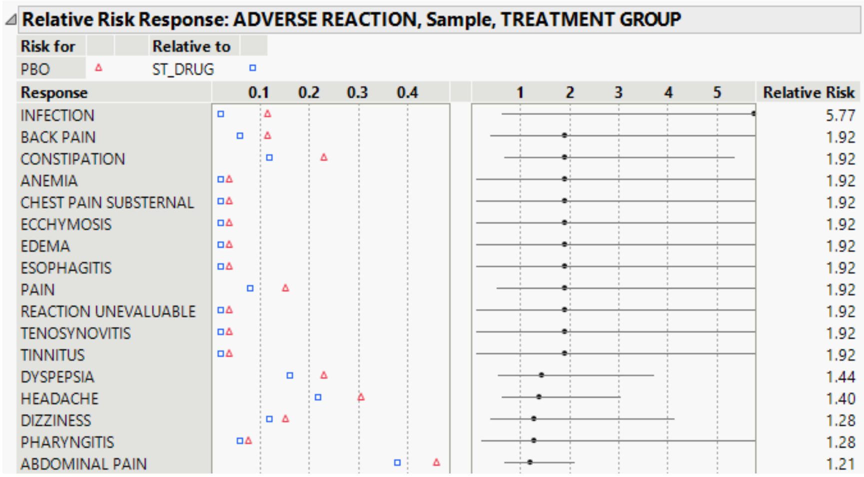
The conditional association matrix provides the conditional probability of one adverse reaction given the presence of another reaction. The probabilities are across all groups. The probability of abnormal vision given that a patient has abdominal pain is 0.0323.

Tip: Place your cursor over the heat map for conditional probabilities.

9. Click the Categorical red triangle and select **Relative Risk.**
10. Select PBO in the window and click **OK**.

    Right-click the Relative Risk Report in the window and select **Sort by Column**.
11. Select Relative Risk and click **OK**.

Figure 3.32 Relative Risk Report (Partial Report)



The Relative Risk option computes relative risks of 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

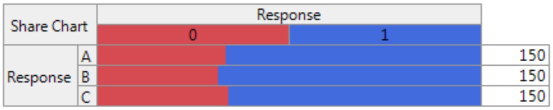
This example uses the Attribute Gauge.jmp sample data, which has the ratings (0/1) from three operators rating 50 parts three times.

1. Select **Help > Sample Data Library** and open Attribute Gauge.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select the Related tab.
4. Select A, B, and C, and click **Rater Agreement** on the Related tab.
5. Click **OK**.

Figure 3.33 Rater Agreement Report

Rater Agreement(A, B, C)

Freq / Share		Response 0	Response 1	Total Responses
Response	A	50 33.3%	100 66.7%	150
	B	47 31.3%	103 68.7%	150
	C	51 34.0%	99 66.0%	150



Rater Agreement Statistics

Degree of Agreement

Degree of Agreement

Response1	Response2	Kappa	Std Err	Bowker Symmetry	Bowker P-Value
A	B	0.862944	0.044198	1	0.3173
A	C	0.776119	0.054671	0.066667	0.7963
B	C	0.788007	0.053702	1.142857	0.2850

Two by two tables are the same as Bowker's Test

Details

Rater Row	Rater Col	Level	0	1
B	A	0	44	3
B	A	1	6	97
C	A	0	43	8
C	A	1	7	92
C	B	0	42	9
C	B	1	5	94

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, indicating no strong evidence of asymmetry 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 table to explore repeated measures where we consider the election results as repeated measures.

1. Select **Help > Sample Data Library** and open Presidential Elections.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select the Related tab.
4. Select 1980 Winner through 2012 Winner, and click **Repeated Measures** on the Related tab.
5. Click **OK**.

6. Near the bottom of the report window, click the gray Transition Report disclosure icon to open the Transition Report.

Figure 3.34 Repeated Measures Transition Report

Transition Report

All From	to	Transition Counts			Transition Rates			
All 1980 Winner	1984 Winner		**Democrat**	**Republican**		**Democrat**	**Republican**	**1984 Winner**
		Democrat	1	5	Democrat	0.1667	0.8333	
		Republican	0	44	Republican	0.0000	1.0000	
All 1984 Winner	1988 Winner		**Democrat**	**Republican**		**Democrat**	**Republican**	**1988 Winner**
		Democrat	1	0	Democrat	1.0000	0.0000	
		Republican	9	40	Republican	0.1837	0.8163	
All 1988 Winner	1992 Winner		**Democrat**	**Republican**		**Democrat**	**Republican**	**1992 Winner**
		Democrat	10	0	Democrat	1.0000	0.0000	
		Republican	22	18	Republican	0.5500	0.4500	
All 1992 Winner	1996 Winner		**Democrat**	**Republican**		**Democrat**	**Republican**	**1996 Winner**
		Democrat	29	3	Democrat	0.9063	0.0938	
		Republican	2	16	Republican	0.1111	0.8889	
All 1996 Winner	2000 Winner		**Democrat**	**Republican**		**Democrat**	**Republican**	**2000 Winner**
		Democrat	20	11	Democrat	0.6452	0.3548	
		Republican	0	19	Republican	0.0000	1.0000	
All 2000 Winner	2004 Winner		**Democrat**	**Republican**		**Democrat**	**Republican**	**2004 Winner**
		Democrat	18	2	Democrat	0.9000	0.1000	
		Republican	1	29	Republican	0.0333	0.9667	
All 2004 Winner	2008 Winner		**Democrat**	**Republican**		**Democrat**	**Republican**	**2008 Winner**
		Democrat	19	0	Democrat	1.0000	0.0000	
		Republican	9	22	Republican	0.2903	0.7097	
All 2008 Winner	2012 Winner		**Democrat**	**Republican**		**Democrat**	**Republican**	**2012 Winner**
		Democrat	26	2	Democrat	0.9286	0.0714	
		Republican	0	22	Republican	0.0000	1.0000	

The Transition Report is unique to the repeated measures analysis. This report includes counts and rates of differences between subsequent time points. Between 1980 and 1984, there were 5 Democratic states that transitioned to Republican states at a rate of 0.8333 or 5 out of 6 states. In 1980, they voted Democratic but voted Republican in 1984. Between 2008 and 2012, there were 2 out of 28 Democratic states that transitioned to Republican at a rate of 0.0714. All other states voted the same way in both the 2008 and 2012 elections.

Example of the Multiple Responses

The following examples use sample data tables that contain the same information organized in five 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

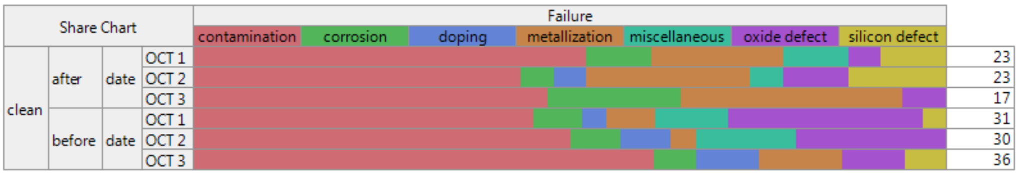
The Failure3MultipleField.jmp sample data table has a row for each unit and multiple columns for defects, where defects are entered one per column. In this example, there are three columns for defects. Thus, any one unit had at most three defects.

1. Select **Help > Sample Data Library** and open Quality Control/Failure3MultipleField.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select the Multiple tab.
4. Select Failure1, Failure2, and Failure3, and click **Multiple Response** on the Multiple tab.
5. Select clean and date and click **X, Grouping Category**.
6. Click **OK**.

Figure 3.35 Multiple Response Report

Multiple Response(Failure1, Failure2, Failure3) By clean*date

			Freq / Share / Rate	Failure: contamination	corrosion	doping	metallization	miscellaneous	oxide defect	silicon defect	Total Responses	Total Cases	Total Cases Responding
clean	after	date	OCT 1	12	2	0	4	2	1	2	23	50	23
				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	23
				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	17
				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	28
				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	30
				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	36
				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%			



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 the defects for that batch. The 12 contamination defects were from 50 units. For each clean and date combination, there were 50 total units. Each unit could have one or more defects. Therefore, the rate of contamination per unit was 24%. For the batch before cleaning on OCT 1, the Total Cases Responding is 28. The Total Responses count is 31, because three of the cases reported two defects.

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 occurrences of each defect type, and an ID column for each batch.

Figure 3.36 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 the Multiple tab.
4. Select failure and click **Multiple Response by ID** on the Multiple tab.
5. Select clean and date and click **X, Grouping Category**.
6. Select SampleSize and click **Sample Size**.
7. Select N and click **Freq**.
8. Select ID and click **ID**.
9. Click **OK**.

The resulting report is the same as the report shown in Figure 3.35 with the exception of the Total Cases Responding column in the crosstab table. Here, the defect counts were 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.

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 3.37, that some units did not have any observed defects, so the failures column is empty.

Figure 3.37 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 the Multiple tab.
4. Select failures and click **Multiple Delimited** on the Multiple tab.
5. Select clean and date and click **X, Grouping Category**.
6. Click **OK**.

When you click **OK**, you also get the report shown in Figure 3.35.

Note: If you specify more than one delimited column, separate analyses are produced for each column.

Indicator Group

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

Figure 3.38 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

1. Select **Help > Sample Data Library** and open Quality Control/Failures3Indicators.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select the Multiple tab.
4. Select contamination through silicon defect and click **Indicator Group** on the Multiple tab.
5. Select clean and date and click **X, Grouping Category**.
6. Click **OK**.

When you click **OK**, you get the report shown in Figure 3.35.

Response Frequencies

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

Figure 3.39 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

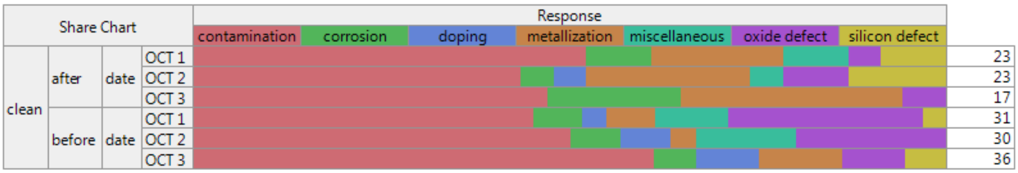
1. Select **Help > Sample Data Library** and open Quality Control/Failure3Freq.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Select the Multiple tab.
4. Select the frequency variables (contamination through silicon defect).
5. On the Multiple tab, click **Response Frequencies**.
6.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

Figure 3.40 Defect Rate Output

Sample Size: SampleSize

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

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



The resulting output is the same as that in Figure 3.35 with the exception of the Total Cases Responding column in the crosstab table. Here, the defect counts were 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 an average employee tenure using the Mean Score option in the categorical platform, assign Value Scores to the column values. For more information about column properties, see The Column Info Window chapter in *Using JMP*.

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 Comparisons**.

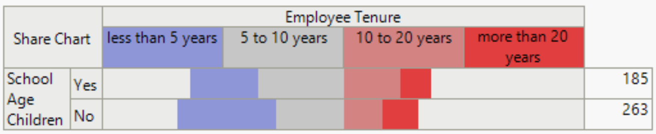
Figure 3.41 Categorical Report with Mean Scores

Employee Tenure By School Age Children

Freq / Share			Employee Tenure: less than 5 years	5 to 10 years	10 to 20 years	more than 20 years	Total Responses	Mean	Compare Means
School Age Children	Yes	A	53 28.6%	65 35.1%	44 23.8%	23 12.4%	185	10.170	
	No	B	107 40.7%	74 28.1%	42 16.0%	40 15.2%	263	9.529	

Default Comparison Groups: A/B

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



The mean employee tenure for those with school age children is 10.17 and 9.53 for those without school age children. Because the means are not statistically different, the Compare Means column in the Crosstab table is empty. If there were a difference, a letter would indicate the difference. If you had not used value scores, then the mean for those with school age children would be 2.20 and 2.057 for those without school age children.

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

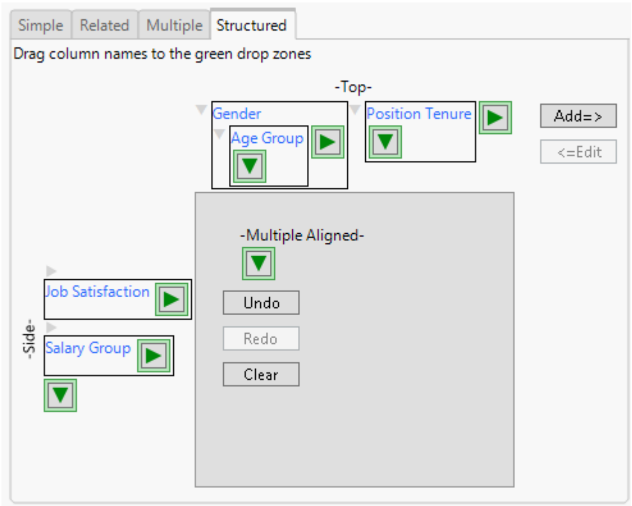
Example of a Structured Report

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

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.

Tip: Click on the green drop zone arrows to select columns.

Figure 3.42 Structured Tab Report Setup



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

Figure 3.43 Structured Tab Report Example

Job Satisfaction+Salary Group By Gender*Age Group+Position Tenure

	Freq	Gender																	
	Share	M							F										
		Age Group							Age Group							Position Tenure			
		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	3	1	3	3	2	2	3	1	0	5	2	1	3	15	11	6	0
		5.5%	10.3%	3.3%	8.8%	7.9%	8.3%	4.1%	5.2%	2.6%	0.0%	27.8%	14.3%	5.3%	13.6%	7.2%	8.2%	6.8%	0.0%
	Somewhat satisfied	32	14	15	19	19	10	26	34	30	12	6	5	11	9	122	70	40	10
		58.2%	48.3%	50.0%	55.9%	50.0%	41.7%	53.1%	58.6%	76.9%	63.2%	33.3%	35.7%	57.9%	40.9%	58.7%	52.2%	45.5%	55.6%
	Extremely satisfied	20	12	14	12	16	12	21	21	8	7	7	7	7	10	71	53	42	8
		36.4%	41.4%	46.7%	35.3%	42.1%	50.0%	42.9%	36.2%	20.5%	36.8%	38.9%	50.0%	36.8%	45.5%	34.1%	39.6%	47.7%	44.4%
	Total Responses	55	29	30	34	38	24	49	58	39	19	18	14	19	22	208	134	88	18
Salary Group	less than 40000	20	6	4	8	6	5	10	35	19	6	6	2	7	4	80	40	15	3
		36.4%	20.7%	13.3%	23.5%	15.8%	20.8%	20.4%	60.3%	48.7%	31.6%	33.3%	14.3%	36.8%	18.2%	38.5%	29.9%	17.0%	16.7%
	40000 to 60000	16	14	11	12	14	7	13	15	10	8	6	7	6	12	67	38	37	9
		29.1%	48.3%	36.7%	35.3%	36.8%	29.2%	26.5%	25.9%	25.6%	42.1%	33.3%	50.0%	31.6%	54.5%	32.2%	28.4%	42.0%	50.0%
	60000 to 80000	8	5	7	7	8	8	10	5	5	3	3	3	4	5	32	25	19	5
		14.5%	17.2%	23.3%	20.6%	21.1%	33.3%	20.4%	8.6%	12.8%	15.8%	16.7%	21.4%	21.1%	22.7%	15.4%	18.7%	21.6%	27.8%
	80000 to 120000	9	3	6	4	7	3	7	1	2	1	2	1	1	0	19	17	10	1
		16.4%	10.3%	20.0%	11.8%	18.4%	12.5%	14.3%	1.7%	5.1%	5.3%	11.1%	7.1%	5.3%	0.0%	9.1%	12.7%	11.4%	5.6%
	greater than 120000	2	1	2	3	3	1	9	2	3	1	1	1	1	1	10	14	7	0
		3.6%	3.4%	6.7%	8.8%	7.9%	4.2%	18.4%	3.4%	7.7%	5.3%	5.6%	7.1%	5.3%	4.5%	4.8%	10.4%	8.0%	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 females, there is a difference in job satisfaction across age groups (Pearson *p*-value = 0.0149). The middle-aged females tend to be the least satisfied with their jobs. Share and frequency charts can be added to your report to visualize your results.

Example of a Multiple Response with Nonresponse

This example uses the Color Preference Survey.jmp sample data table, which contains survey data about people's color preferences. You can use the Categorical platform to summarize results from a multiple response question with nonresponses. A nonresponse is a participant who did not answer the question. Surveys often use "none of the above" or "not applicable" as a response choice. This provides a method for distinguishing responders who do not see an appropriate response from those who did not answer the question.

1. Select **Help > Sample Data Library** and open Color Preference Survey.jmp.
2. Select **Analyze > Consumer Research > Categorical**.
3. Click the **Structured** tab.
4. Select What colors do you like? (with nonresponse) and drag it to the green arrow drop zone at the **Side** of the table.

Note: This column contains multiple response data where the response levels are delimited with commas. There are three rows that contain no responses. This column was constructed for this example and does not align with the responses in the individual "What colors do you like" columns.

---

5. Click the **Top** green arrow and select What is your gender?
6. Click **Add=>**.
7. Click **OK**.

Figure 3.44 Initial Cross Tabulation

Categorical

What colors do you like? (with nonresponse) By What is your gender?

What colors do you like? (with nonresponse)	Freq / Share / Rate	What is your gender? Female	Male
Blue	Freq	31	38
	Share	21.5%	21.5%
	Rate	83.8%	88.4%
Green	Freq	24	36
	Share	16.7%	20.3%
	Rate	64.9%	83.7%
None of the above	Freq	2	0
	Share	1.4%	0.0%
	Rate	5.4%	0.0%
Orange	Freq	12	23
	Share	8.3%	13.0%
	Rate	32.4%	53.5%
Pink	Freq	19	9
	Share	13.2%	5.1%
	Rate	51.4%	20.9%
Purple	Freq	28	23
	Share	19.4%	13.0%
	Rate	75.7%	53.5%
Red	Freq	18	33
	Share	12.5%	18.6%
	Rate	48.6%	76.7%
Yellow	Freq	10	15
	Share	6.9%	8.5%
	Rate	27.0%	34.9%
Total Responses		144	177
Total Cases		37	43
Total Cases Responding		36	41

Note that the number of total cases is higher than the number of total cases responding. There are 37 Females who filled out the survey but only 36 responded to this question.

8. Click the Categorical red triangle and select **Test Multiple Responses > Exclude Nonresponses.**
9. Click the Categorical red triangle and select **Test Multiple Responses > Homogeneity Test, Binomial.**

Figure 3.45 Binomial Homogeneity Test Results

Test Each Response, Binomial

Note: Nonresponse rows are excluded.

What is your gender?, What colors do you like? (with nonresponse)

What colors do you like? (with nonresponse)	ChiSquare	Prob>ChiSq
Blue	0.8918	0.3450
Green	5.0612	0.0245*
None of the above	3.1020	0.0782
Orange	4.0516	0.0441*
Pink	7.9930	0.0047*
Purple	4.1132	0.0426*
Red	8.0999	0.0044*
Yellow	0.6819	0.4089

Assuming multiple responses are Check All That Apply (CATA).

The analysis supports that blue and yellow have similar preference levels across genders while other colors tend to be favored more or less across genders. For example, Red appears to be liked more by males than by females. The *p*-value for the test is 0.0044, which supports a conclusion that there is a difference in preference across genders. From the crosstab in Figure 3.44 you find that 76.7% of the males liked the color red as compared to 48.6% of the females.

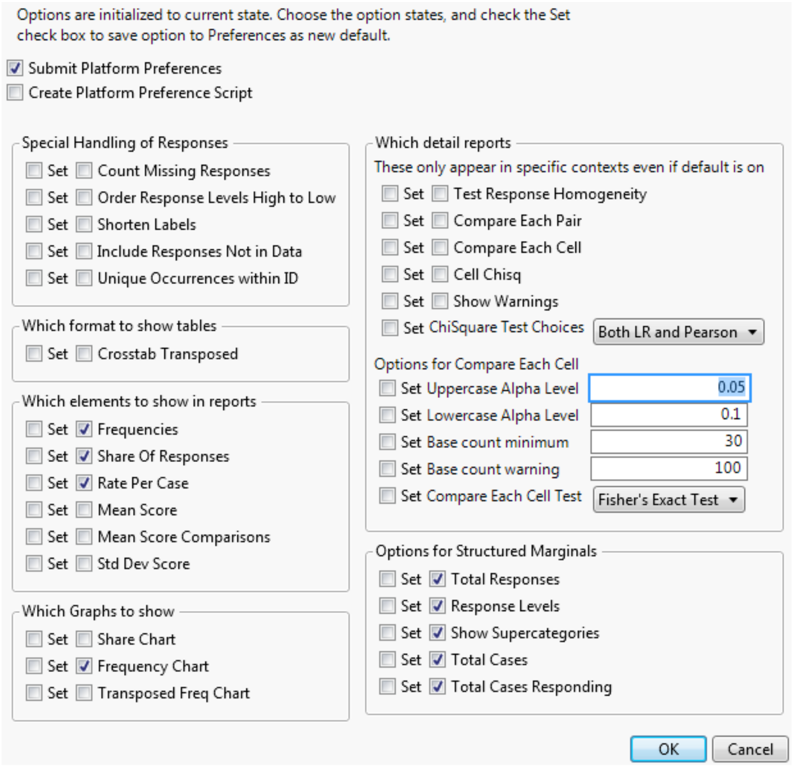
The note at the top of the output indicates that nonresponse rows are excluded from the analysis. The note at the bottom of the output indicates that it is assumed that the multiple response is a check all that applies type question. This means that a respondent can select more than one response to the question. The tests are performed using only those rows with responses. The three rows that did not have responses for the question are not included.

Tip: You can obtain the same output using the Multiple tab. Select What colors do you like? (with nonresponse) as a Multiple Delimited response and What is your Gender? as a grouping variable.

Set Preferences

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

Figure 3.46 Set Preferences Window



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 changes 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.

Running the saved 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.

Statistical Details for the Categorical Platform

This section contains 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).

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.

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

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

where

χ^2 is the standard Pearson Chi-squared statistic

$\bar{\delta} = 1 - \frac{m_{++}}{n_+ C}$ is the correction factor

The correction factor contains the following quantities:

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 are $(R\text{-}1)C$ or the number of rows minus 1 times the number of columns.

Chapter 4

Choice Models

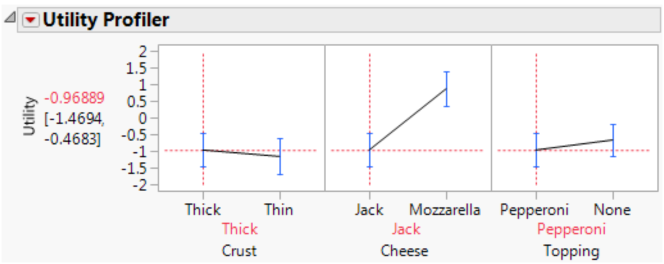
Fit Models for Choice Experiments

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 . . . 83
Examples of the Choice Platform . . . 85
 One Table Format with No Choice . . . 85
 Multiple Table Format . . . 88
Launch the Choice Platform . . . 95
 Launch Window for One Table, Stacked . . . 96
 Launch Window for Multiple Tables, Cross-Referenced . . . 98
Choice Model Report . . . 104
 Effect Summary . . . 104
 Parameter Estimates . . . 105
 Likelihood Ratio Tests . . . 106
 Bayesian Parameter Estimates . . . 106
Choice Platform Options . . . 108
 Willingness to Pay . . . 111
Additional Examples . . . 114
 Example of Making Design Decisions . . . 114
 Example of Segmentation . . . 123
 Example of Logistic Regression Using the Choice Platform . . . 127
 Example of Logistic Regression for Matched Case-Control Studies . . . 130
 Example of Transforming Data to Two Analysis Tables . . . 132
 Example of Transforming Data to One Analysis Table . . . 137
Statistical Details for the Choice Platform . . . 139
 Special Data Table Rules . . . 139
 Utility and Probabilities . . . 140
 Gradients . . . 141

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 Choice Designs chapter in the *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. See 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 123. For more information about the gradient values, see “Gradients” on page 141.

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). These parameters i can be analyzed with 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 85 shows how to analyze a subset of the available data in a one table format.
- "Multiple Table Format" on page 88 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.

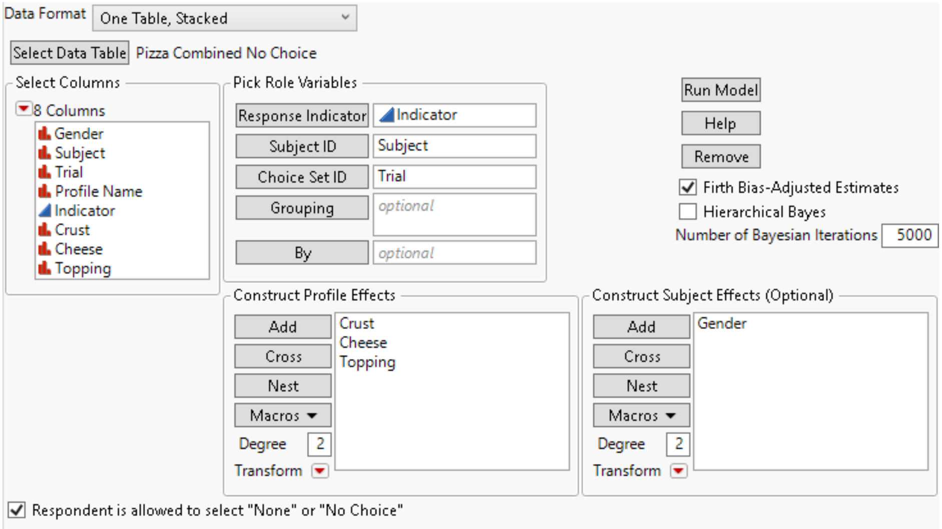
   Choice sets are defined by the combination of Subject and Trial. Notice that there are missing values in the Indicator column for some choice sets.

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

   The One Table, Stacked data format is the default.

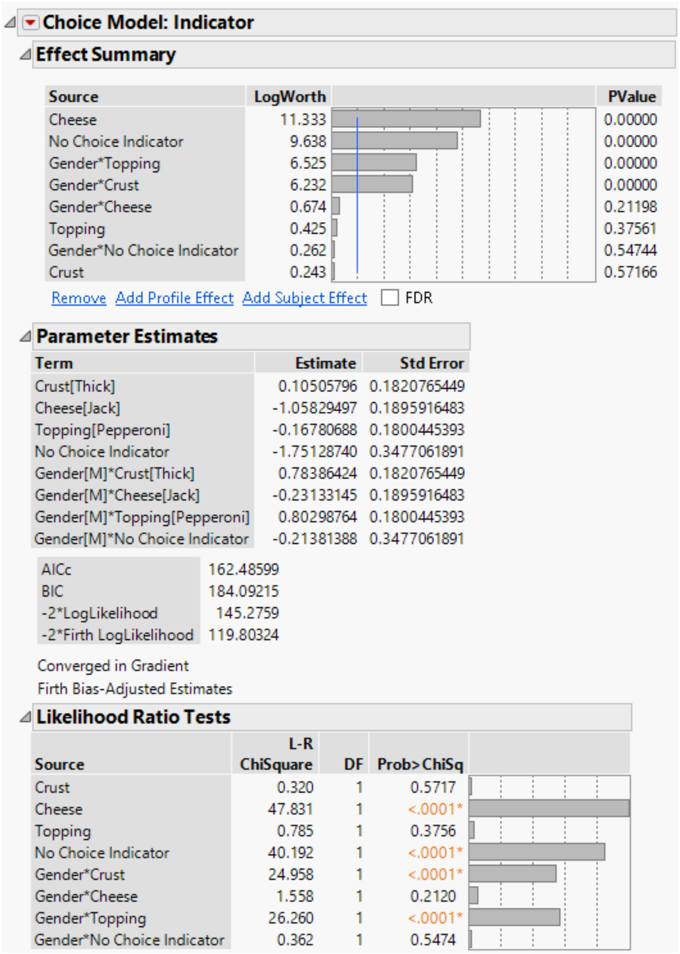
3. Click **Select Data Table.**
4. Select **Pizza Combined No Choice** and click **OK**.
5. Complete the launch window as follows:
   – Select Indicator and click **Response Indicator**.
   – Select Subject and click **Subject ID**.
   – Select Trial and click **Choice Set ID**.
   – Select Crust, Cheese, 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



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



Choice Model: Indicator

Effect Summary

Source	LogWorth	PValue
Cheese	11.333	0.00000
No Choice Indicator	9.638	0.00000
Gender*Topping	6.525	0.00000
Gender*Crust	6.232	0.00000
Gender*Cheese	0.674	0.21198
Topping	0.425	0.37561
Gender*No Choice Indicator	0.262	0.54744
Crust	0.243	0.57166

Remove Add Profile Effect Add Subject Effect FDR

Parameter Estimates

Term	Estimate	Std Error
Crust[Thick]	0.10505796	0.1820765449
Cheese[Jack]	-1.05829497	0.1895916483
Topping[Pepperoni]	-0.16780688	0.1800445393
No Choice Indicator	-1.75128740	0.3477061891
Gender[M]*Crust[Thick]	0.78386424	0.1820765449
Gender[M]*Cheese[Jack]	-0.23133145	0.1895916483
Gender[M]*Topping[Pepperoni]	0.80298764	0.1800445393
Gender[M]*No Choice Indicator	-0.21381388	0.3477061891

AICc	162.48599
BIC	184.09215
-2*LogLikelihood	145.2759
-2*Firth LogLikelihood	119.80324

Converged in Gradient
Firth Bias-Adjusted Estimates

Likelihood Ratio Tests

Source	L-R ChiSquare	DF	Prob>ChiSq
Crust	0.320	1	0.5717
Cheese	47.831	1	<.0001*
Topping	0.785	1	0.3756
No Choice Indicator	40.192	1	<.0001*
Gender*Crust	24.958	1	<.0001*
Gender*Cheese	1.558	1	0.2120
Gender*Topping	26.260	1	<.0001*
Gender*No Choice Indicator	0.362	1	0.5474

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.

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

9. 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	TrimOni	•	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	ThickOni	•	Thick	Mozzarella	Pepperoni
10	M	11	4	ThickJackoni	•	Thick	Jack	Pepperoni
11	F	14	2	TrimOni	•	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	TrimOni	•	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 93.

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 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**.

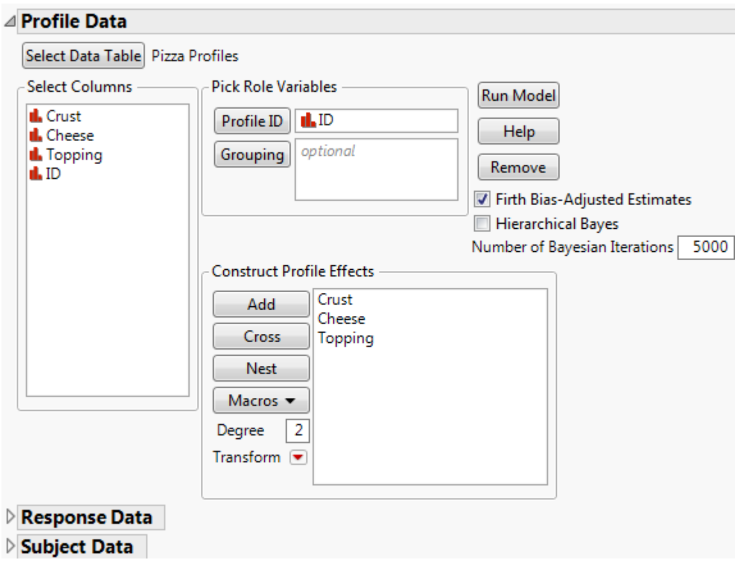
   There are three separate sections, one for each of the data sources.

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.

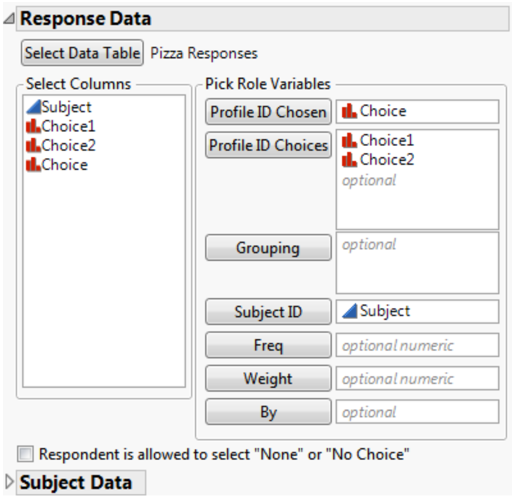
5. Select Pizza Profiles.jmp and click **OK**.
6. Select ID and click **Profile ID**.
7. Select Crust, Cheese, and Topping and click **Add**.

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**.

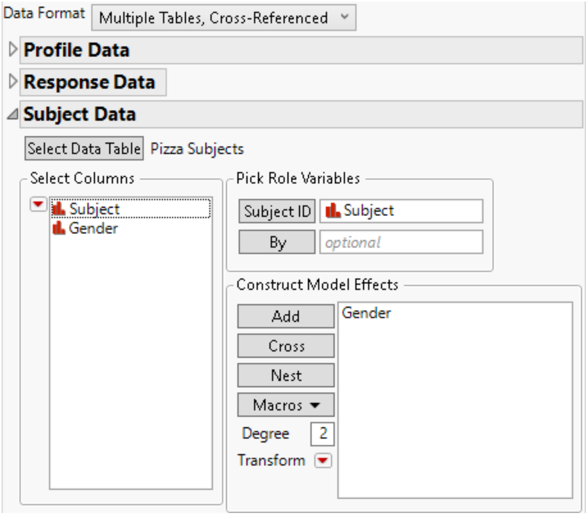
Figure 4.6 Response Data Window



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

Choice Model: Choice

Effect Summary

Source	LogWorth	PValue
Cheese	15.716	0.00000
Gender*Crust	7.701	0.00000
Gender*Topping	7.329	0.00000
Topping	0.297	0.50483
Gender*Cheese	0.281	0.52380
Crust	0.032	0.92949

Remove Add Profile Effect Add Subject Effect ☐ FDR

Parameter Estimates

Term	Estimate	Std Error
Crust[Thick]	0.08314124	0.2277722879
Cheese[Jack]	-1.44147673	0.2522310248
Topping[Pepperoni]	-0.17845527	0.2244231434
Gender[F]*Crust[Thick]	-1.00638844	0.2277722879
Gender[F]*Cheese[Jack]	0.18490649	0.2522310248
Gender[F]*Topping[Pepperoni]	-0.98323564	0.2244231434

AICc	87.777628
BIC	104.19559
-2*LogLikelihood	75.083413
-2*Firth LogLikelihood	56.079805

Converged in Gradient
Firth Bias-Adjusted Estimates

Likelihood Ratio Tests

Source	L-R ChiSquare	DF	Prob>ChiSq
Crust	0.008	1	0.9295
Cheese	67.682	1	<.0001*
Topping	0.445	1	0.5048
Gender*Crust	31.504	1	<.0001*
Gender*Cheese	0.406	1	0.5238
Gender*Topping	29.842	1	<.0001*

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

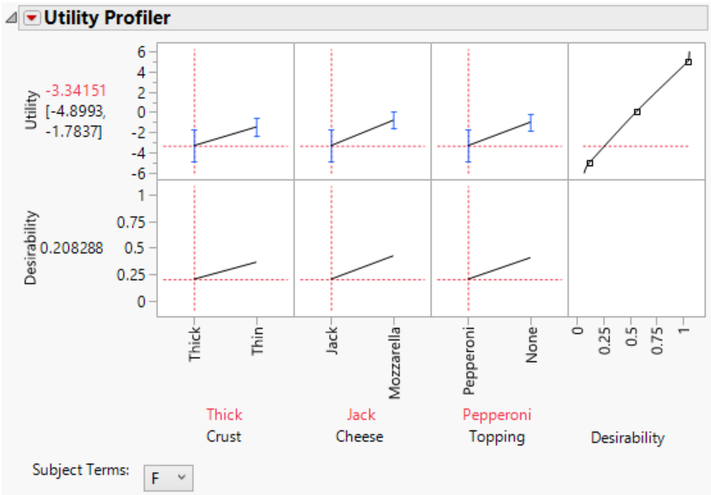
Next, you use the Utility Profiler to explore your results and find optimal settings for the attributes.

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

   The Subject Terms menu beneath the profiler indicates that it is showing results for females.

2. Click the Utility Profiler red triangle and select **Optimization and Desirability > Desirability Functions**.

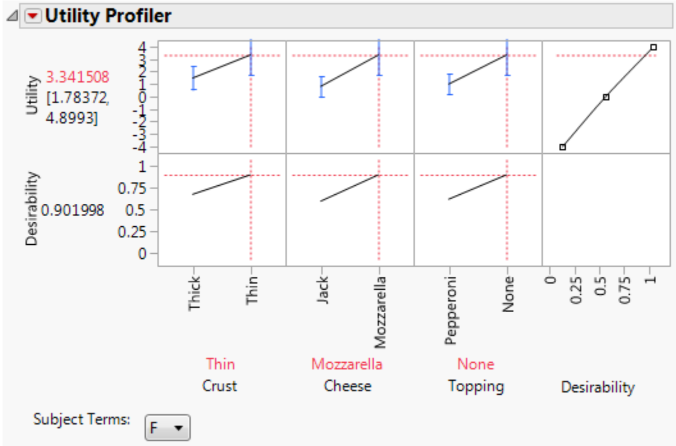
Figure 4.9 Utility Profiler with Desirability Function



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

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

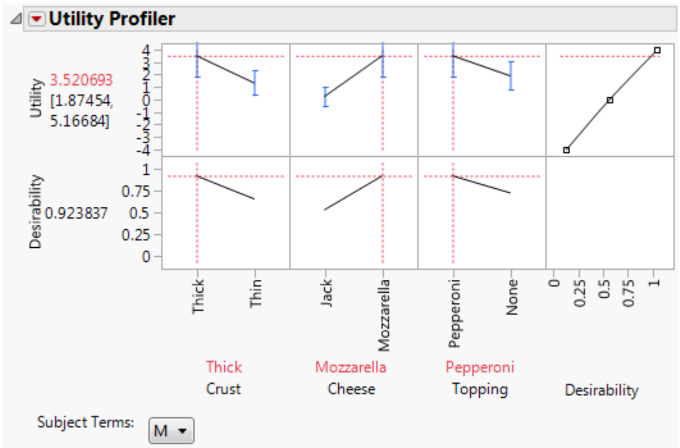
Figure 4.10 Utility Profiler with Optimal Settings for Females



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

4. From the Subject Terms menu, select **M**.
5. Click the Utility Profiler red triangle and select **Optimization and Desirability > Maximize Desirability.**

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. In the Choice launch window specify whether you are using one or multiple data tables in the Data Format list.

One Table, Stacked

For the One Table, Stacked format, the data are in a single data table. There is a row for every profile presented to a subject and an indicator of whether that profile was selected. The Pizza Combined No Choice.jmp sample data table contains the results of a choice experiment in a single table format. See "One Table Format with No Choice" on page 85.

For more information about the launch window for this format, see "Launch Window for One Table, Stacked" on page 96.

Multiple Tables, Cross-Referenced

For the Multiple Tables, Cross-Referenced format, the data are in two or three separate data tables. A profile data table and a response data table are required. A subject data table is optional. Note the following:

- The profile data table must contain a column with a unique identifier for each profile and columns for the profile level variables. The profile identifier is used in the response data table to identify the profiles presented and the profile selected.
- The optional subject data table must contain a column with a unique subject identifier for each subject and columns for the subject level variables. The subject identifier is used in the response table to identify the subjects.

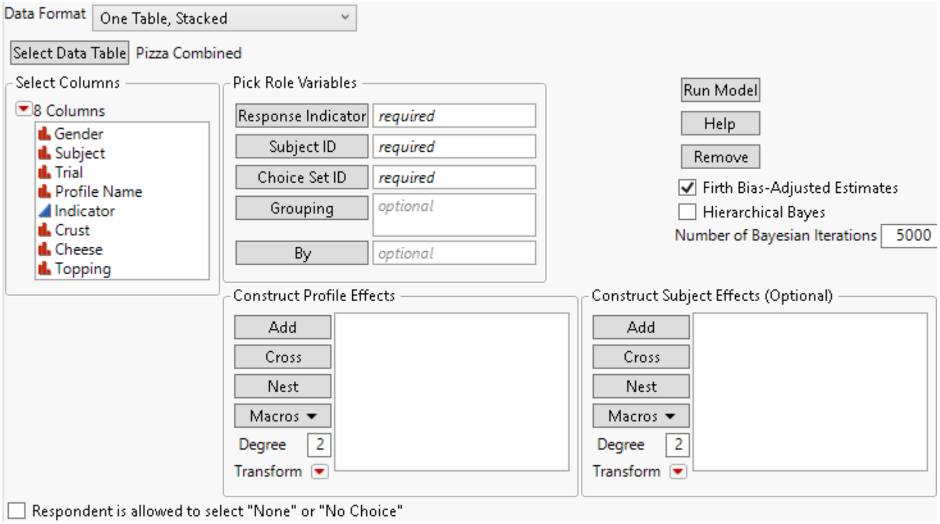
The launch window for this format contains three sections: Profile Data, Response Data, and Subject Data. Each section corresponds to a different data table. You can expand or collapse each section as needed.

The Pizza Profiles.jmp, Pizza Responses.jmp, and Pizza Subjects.jmp sample data tables contain the results of a choice experiment using three tables. There is one table for the profiles, one for the responses, and one for the subject information. See "Multiple Table Format" on page 88.

For more information about the launch window for this format, see "Launch Window for Multiple Tables, Cross-Referenced" on page 98.

Launch Window for One Table, Stacked

Figure 4.12 Launch Window for One Table, Stacked Data Format



For more information about the options in the Select Columns red triangle menu, see the Get Started chapter in *Using JMP*.

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 97.

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.

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.

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

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

Note: The choice model observes the column coding property of continuous profile and subject effects.

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

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

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. See 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 106.

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

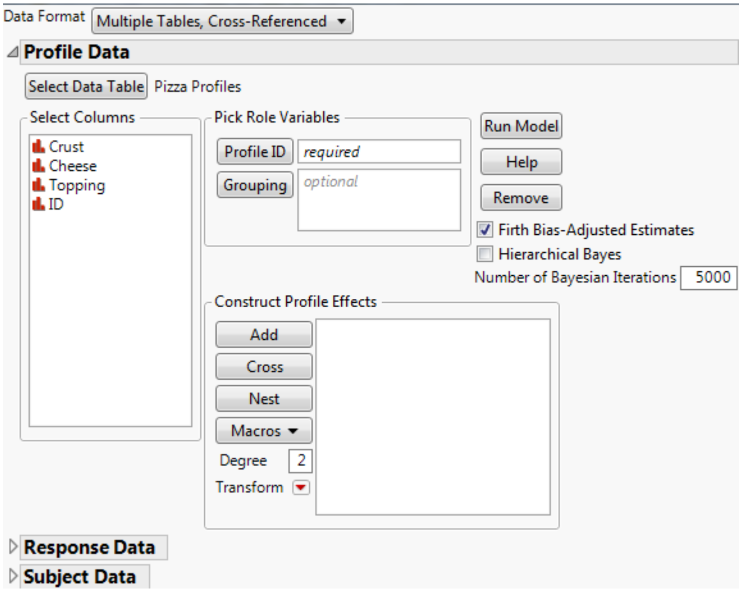


Figure 4.13 shows the launch window for Multiple Tables, using Pizza Profiles.jmp as the Profile table.

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

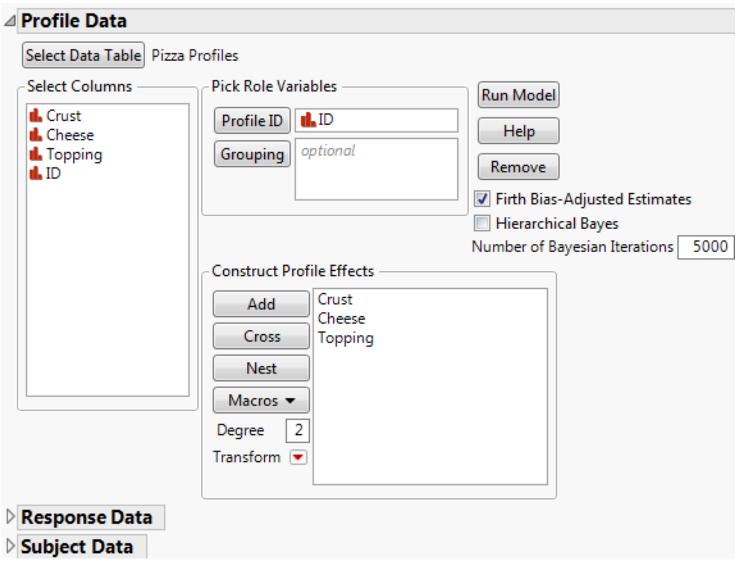
- "Profile Data" on page 98
- "Response Data" on page 100
- "Subject Data" on page 102

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



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 Model Specification chapter in *Fitting Linear Models*.

Note: The choice model observes the column coding property of continuous profile and subject effects.

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. See 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 106.

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.

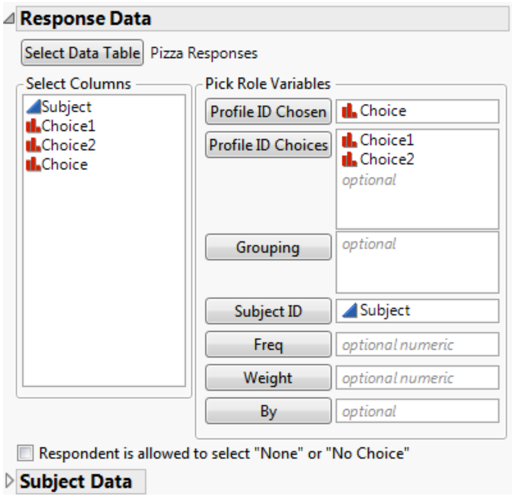
Response Data

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



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. There must be at least two 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 Produces a separate report for each level of the By Variable. If more than one By variable is assigned, a separate analysis is produced for each possible combination of the levels of the By variables.

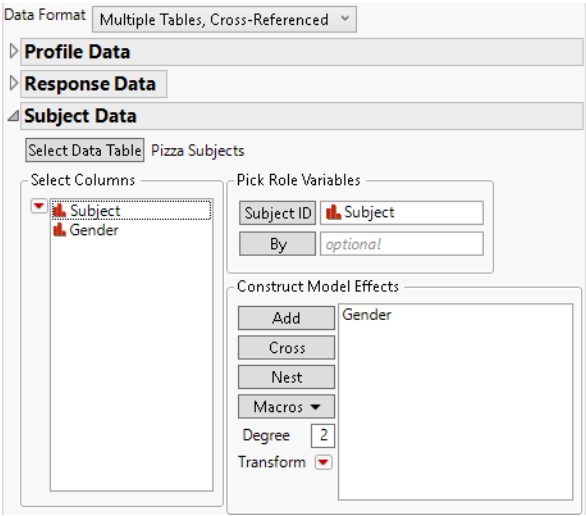
Respondent is allowed to select "None" or "No Choice" Enters a No Choice Indicator into the model for response rows containing missing values. For the Multiple Tables, Cross-Referenced data format, the No Choice rows must contain (categorical) missing values in the Profile ID Chosen column in the Response 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

	Subject	Gender
1	1	M
2	2	F
3	3	M
4	4	F



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.

By Produces a separate report for each level of the By variable. If more than on By variable is assigned, a separate report is produced for each possible combination of the levels of the By variables.

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

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

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. See the Standard Least Squares Report and Options chapter in *Fitting Linear Models*.

Note: The Effect Summary report is not applicable to models fit with Hierarchical Bayes.

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*-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 $-\log_{10}$(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 more information about the FDR correction, see Benjamini and Hochberg (1995). For more information about the false discovery rate, see the Response Screening chapter in *Predictive and Specialized Modeling* 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 more information about utility, see "Utility and Probabilities" on page 140.

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), –2Loglikelihood, and –2Firth Loglikelihood. See the Statistical Details appendix in *Fitting Linear Models*.

The –2Firth 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 –2Loglikelihood 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.

JMP PRO Bayesian Parameter Estimates

(Available only for Hierarchical Bayes.) 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 110.

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 multinominal 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: Choice

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

Choice Model red triangle menu contains the following options.

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 106.

JMP PRO **Show MLE Parameter Estimates** (Available for Hierarchical Bayes) 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 for Hierarchical Bayes) 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 *Fitting Linear Models*.

Confidence Intervals (Not available for Hierarchical Bayes) Shows or hides a confidence interval for each parameter in the Parameter Estimates report.

Confidence Limits (Available for Hierarchical Bayes) Shows or hides 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.

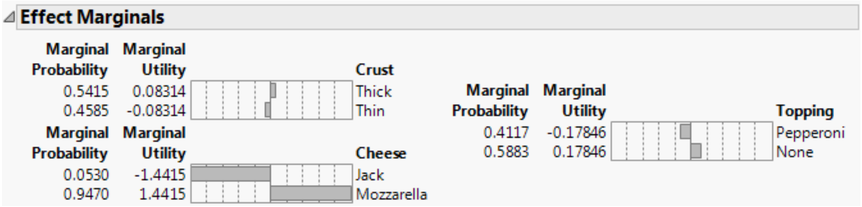
Correlation of Estimates If Hierarchical Bayes was not selected, shows the correlations between the maximum likelihood parameter estimates.

For 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 or hides 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 set to 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 or hides the predicted utility for different factor settings. The utility is the value predicted by the linear model. See "Find Optimal Profiles" on page 93 for an example of the Utility Profiler. For more information about utility, see "Utility and Probabilities" on page 140. For more information about the Utility Profiler options, see the Profiler chapter in *Profilers*.

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

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

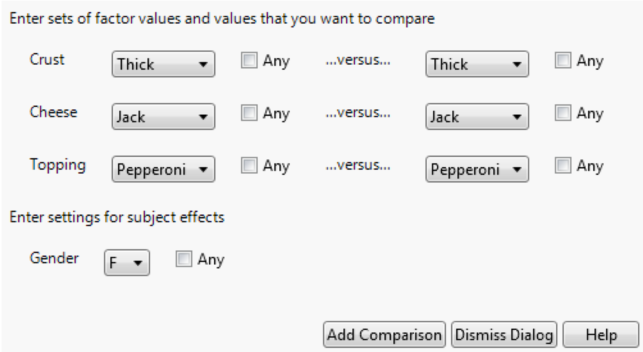
where U is the utility for the current settings and U_b is the utility for the baseline settings. This implies that the probability for the baseline settings is 0.5. See "Utility and Probabilities" on page 140.

See "Comparisons to Baseline" on page 119 for an example of using the Probability Profiler. For more information about the Probability Profiler options, see the Profiler chapter in *Profilers*.

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 121 for an example of using the Multiple Choice Profiler. For more information about the Multiple Choice Profiler options, see the Profiler chapter in *Profilers*.

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 model includes a continuous price column. Calculates the maximum price increase (decrease) that a customer is willing to pay for a new feature over the baseline feature cost. 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 (Not available for Hierarchical Bayes.) 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. See “Gradients” on page 141. For an example, see “Example of Segmentation” on page 123.

Save Subject Estimates (Available for Hierarchical Bayes.) 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 106.

Save Bayes Chain (Available for Hierarchical Bayes.) Creates a table that gives information about the chain of iterations used in computing subject-specific Bayesian estimates. See “Save Bayes Chain” on page 113.

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 *Using JMP* 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.

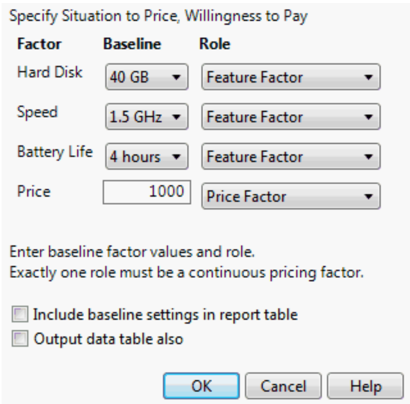
Background Variable A factor that you want to hold constant, at each of its levels, in the Willingness to Pay calculation. Generally, these are subject-level factors. Specifying a subject factor as a Background Variable rather than a Background Constant provides Willingness to Pay estimates for all levels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



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

Willingness to Pay

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

Baseline Utility
-3.4736

Factor	Feature Setting	Price Change	Std Error	Lower 95%	Upper 95%	New Price
Hard Disk	80 GB	$959.67	381.743	$211.47	$1,707.87	$1,959.67
Speed	2.0 GHz	$496.86	218.515	$68.58	$925.14	$1,496.86
Battery Life	6 hours	$387.88	191.088	$13.35	$762.40	$1,387.88

Standard deviations for Price Change calculated by Delta method.

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 114
- "Analyze the Model" on page 116
- "Comparisons to Baseline" on page 119
- "Multiple Choice Comparisons" on page 121

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 116.

2. Click the green triangle next to the **Open Profile and Subject Tables** script.

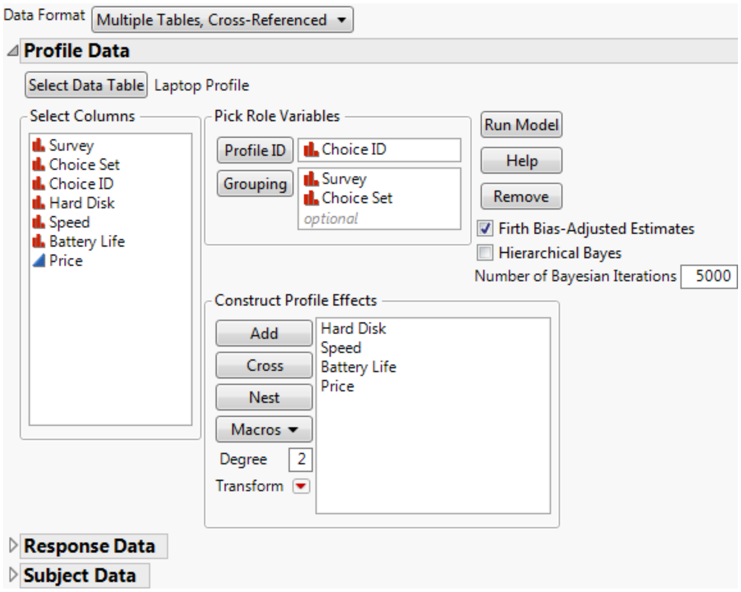
   The script opens the Laptop Profile.jmp and Laptop Subjects.jmp data tables.

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

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

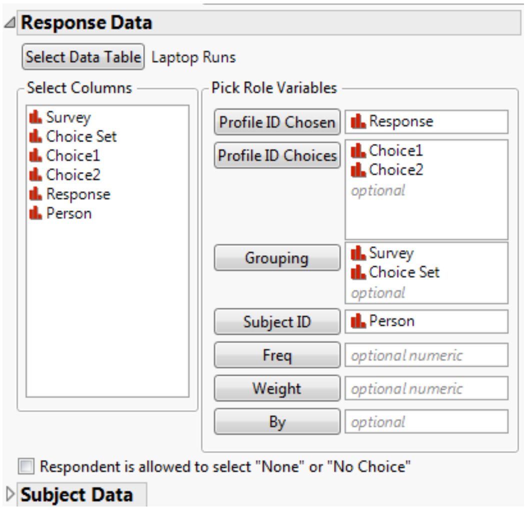
4. From the Data Format list, select **Multiple Tables, Cross-Referenced**.
5. Click **Select Data Table** under Profile Data and select Laptop Profile.jmp. Select Choice ID and click Profile ID.
6. Select Hard Disk, Speed, Battery Life, and Price and click **Add**.
7. Select Survey and Choice Set and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

Figure 4.22 Profile Data Window for Laptop Study



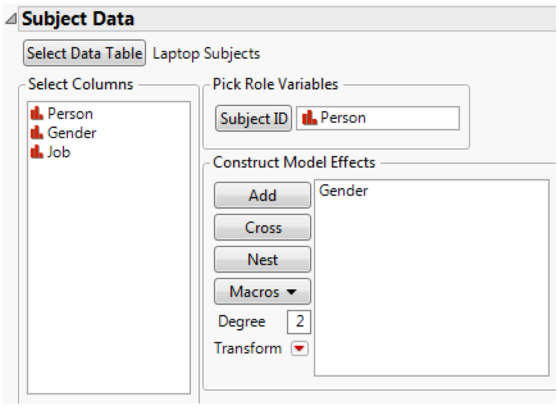
8. Open the **Response Data** outline.
9. From the **Select Data Table** list, select Laptop Runs.jmp.
10. Complete the Response Data table as follows:
    – Select Response and click **Profile ID Chosen**.
    – Select Choice1 and Choice2 and click **Profile ID Choices**.
    – Select Survey and Choice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

Figure 4.23 Response Data Window for Laptop Study



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**.

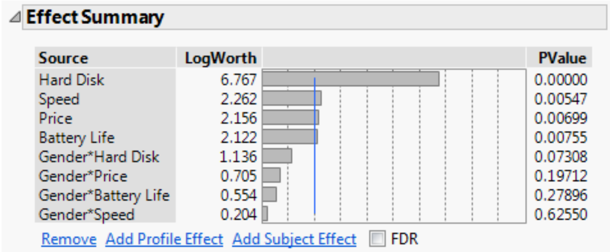
Figure 4.24 Subject Data Window for Laptop Study



Analyze the Model

1. Click **Run Model**.

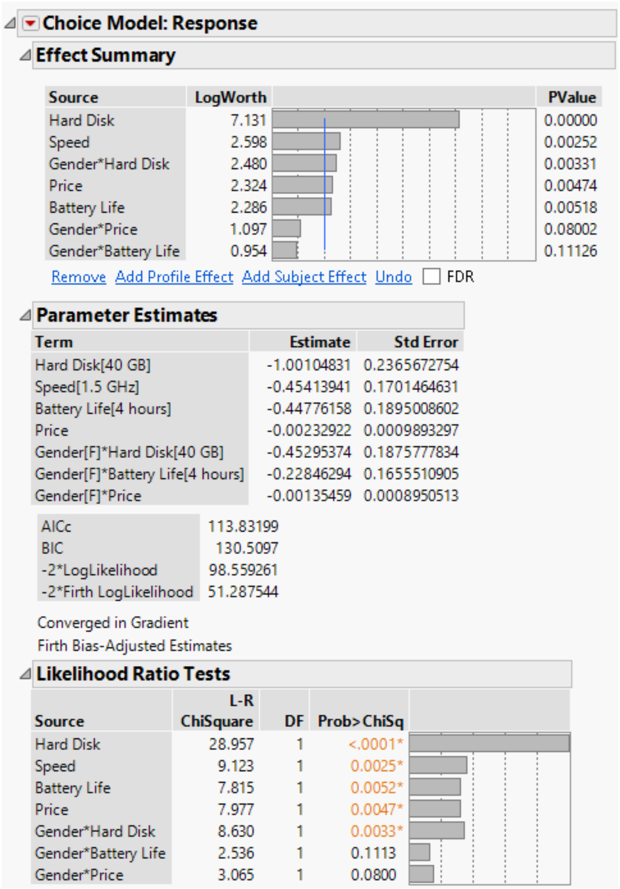
Figure 4.25 Laptop Effect Summary



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



Choice Model: Response

Effect Summary

Source	LogWorth	PValue
Hard Disk	7.131	0.00000
Speed	2.598	0.00252
Gender*Hard Disk	2.480	0.00331
Price	2.324	0.00474
Battery Life	2.286	0.00518
Gender*Price	1.097	0.08002
Gender*Battery Life	0.954	0.11126

Remove Add Profile Effect Add Subject Effect Undo FDR

Parameter Estimates

Term	Estimate	Std Error
Hard Disk[40 GB]	-1.00104831	0.2365672754
Speed[1.5 GHz]	-0.45413941	0.1701464631
Battery Life[4 hours]	-0.44776158	0.1895008602
Price	-0.00232922	0.0009893297
Gender[F]*Hard Disk[40 GB]	-0.45295374	0.1875777834
Gender[F]*Battery Life[4 hours]	-0.22846294	0.1655510905
Gender[F]*Price	-0.00135459	0.0008950513

AICc	113.83199
BIC	130.5097
-2*LogLikelihood	98.559261
-2*Firth LogLikelihood	51.287544

Converged in Gradient
Firth Bias-Adjusted Estimates

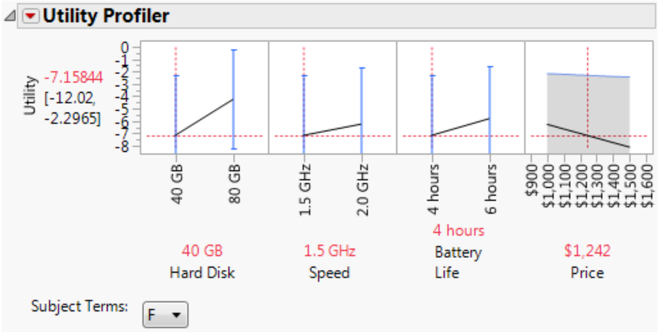
Likelihood Ratio Tests

Source	L-R ChiSquare	DF	Prob>ChiSq
Hard Disk	28.957	1	<.0001*
Speed	9.123	1	0.0025*
Battery Life	7.815	1	0.0052*
Price	7.977	1	0.0047*
Gender*Hard Disk	8.630	1	0.0033*
Gender*Battery Life	2.536	1	0.1113
Gender*Price	3.065	1	0.0800

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.

3. Click the Choice Model: Response red triangle and select **Utility Profiler**.

Figure 4.27 Laptop Profiler Results for Females

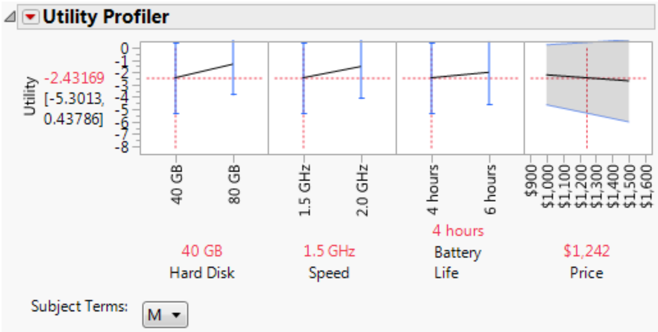


Tip: If your utility profiler does not look like Figure 4.27, click the Utility Profiler red triangle and select **Appearance > Adapt Y Axis**.

---

4. 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 (Figure 4.26 on page 118). 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

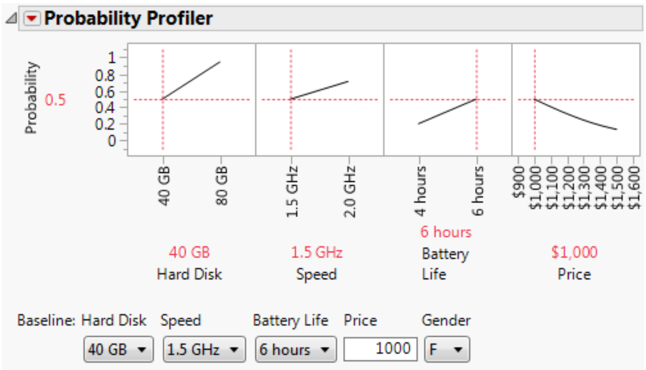
Suppose you are developing a new product. You want to explore the likelihood that a customer selects the new product over the old product, or over a competitor's product. Use the Probability Profiler to compare profiles to a baseline profile.

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 114. Then complete step 1 and step 2 in "Analyze the Model" on page 116.
   – In the Laptop Runs.jmp sample data table, click the green triangle next to the **Choice Reduced Model** script.
2. Click the Choice Model: Response 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.

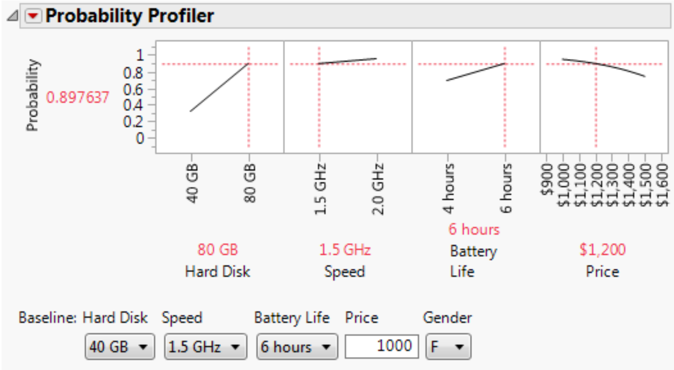
Figure 4.29 Probability Profiler with Text Entry Area for Price



   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 114. Then complete step 1 and step 2 in "Analyze the Model" on page 116.
   – In the **Laptop Runs.jmp** sample data table, click the green triangle next to the **Choice Reduced Model** script.
2. Click the Choice Model: Response red triangle and select **Multiple Choice Profiler**.

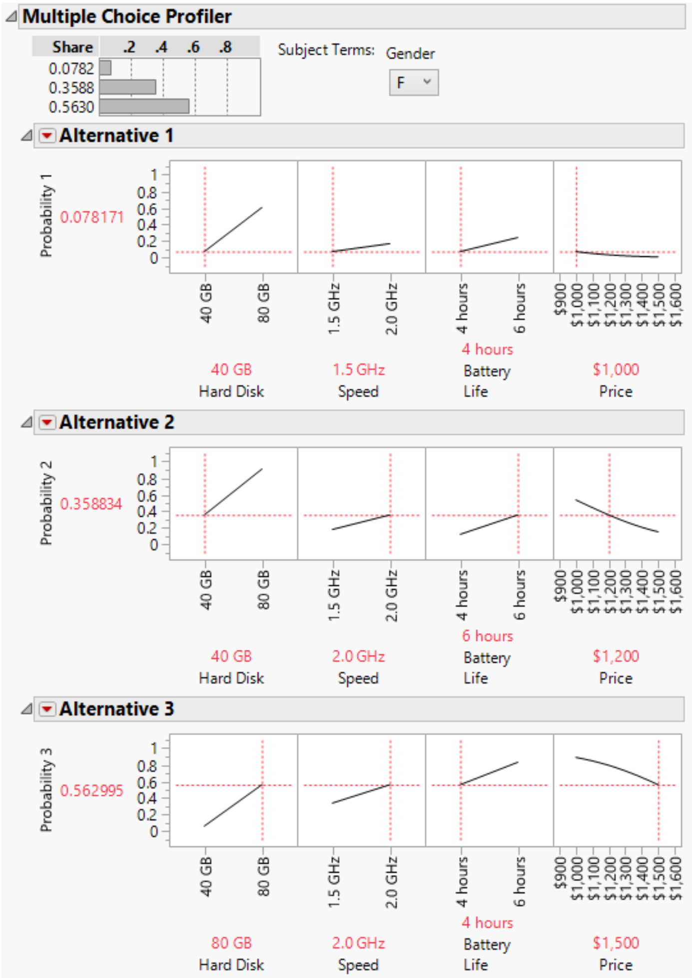
   A window appears, asking for the number of alternative choices to profile. Accept the default number of 3.
3. Click **OK**.

   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 4hours, 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

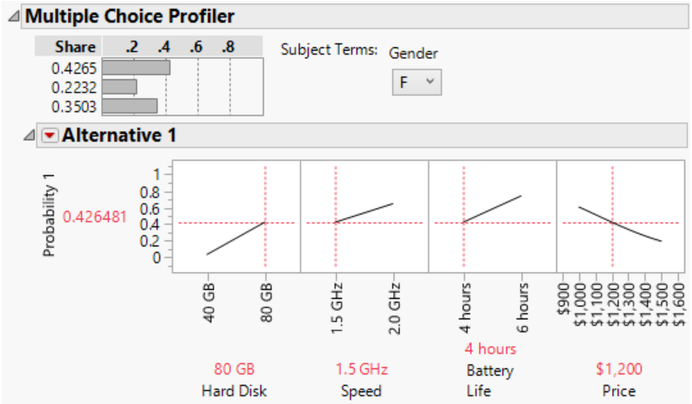


You can see that Company B has the greatest Share of 0.5630. It is obvious that with your company's settings, very few females buy your product.

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 88. 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: Choice red triangle and select **Save Gradients by Subject**.

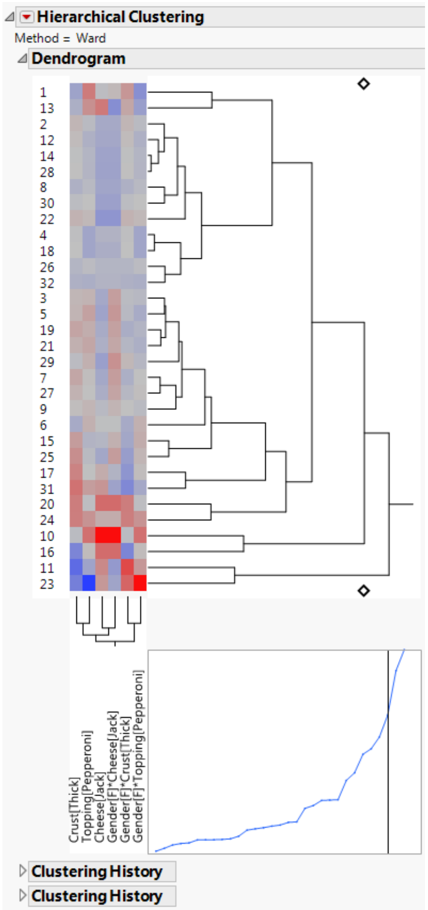
A data table appears with gradient forces saved for each main effect and subject interaction.

Figure 4.33 Gradients by Subject for Pizza Data, Partial View

	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 either diamond to see that the rows have been placed into three clusters.

5. Click the Hierarchical Clustering red triangle 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 chapter in *Multivariate Methods* for a discussion of other Hierarchical Clustering options.

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

6. Select all columns except Subject and Cluster. Right-click the selected columns and select **Delete Columns**.

7. 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.

Figure 4.35 Subject Data with Cluster Column

	Subject	Crust[Thick]	Cheese[Jack]	Topping[Pepperoni]	Gender[F]*Crust[Thick]	Gender[F]*Cheese[Jack]	Gender[F]*Topping[Pepperoni]	Cluster
1	1	-0.00959	-0.00168	0.014876	0.009585	0.001685	-0.01488	1
2	2	0.002373	-0.00758	-0.00239	0.002373	-0.00758	-0.00239	1
3	3	0.002129	-0.0079	0.003031	-0.00213	0.007899	-0.00303	1
4	4	-0.00106	-0.00485	-0.00901	-0.00106	-0.00485	-0.00901	1
5	5	0.002828	-0.00945	0.00725	-0.00283	0.009453	-0.00725	1
6	6	-0.0073	-0.00089	0.003761	-0.0073	-0.00089	0.003761	1
7	7	0.006003	-0.00815	0.000308	-0.006	0.008151	-0.00031	1
8	8	-0.0055	-0.00887	-0.00274	-0.0055	-0.00887	-0.00274	1
9	9	0.000438	-0.00271	0.002402	-0.00044	0.00271	-0.0024	1
10	10	-0.00217	0.032641	0.017043	-0.00217	0.032641	0.017043	2

This table can now be used for further analysis.

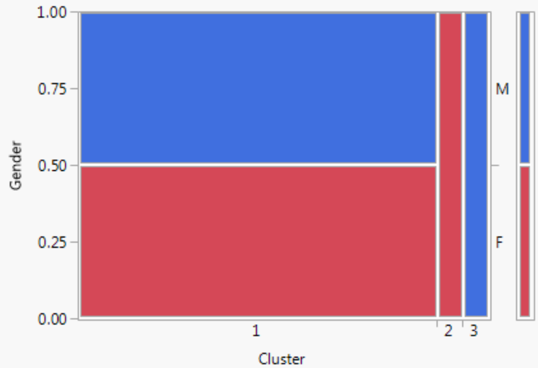
Explore the Clusters

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

Figure 4.36 Contingency Analysis of Gender by Cluster

Contingency Analysis of Gender By Cluster

Mosaic Plot



Contingency Table

Gender

Count / Total % / Col % / Row %	F	M	
Cluster 1	14 43.75 87.50 50.00	14 43.75 87.50 50.00	28 87.50
Cluster 2	2 6.25 12.50 100.00	0 0.00 0.00 0.00	2 6.25
Cluster 3	0 0.00 0.00 0.00	2 6.25 12.50 100.00	2 6.25
	16 50.00	16 50.00	32

Tests

N	DF	-LogLike	RSquare (U)
32	2	2.7725887	0.1250

Test	ChiSquare	Prob>ChiSq
Likelihood Ratio	5.545	0.0625
Pearson	4.000	0.1353

Warning: 20% of cells have expected count less than 5, ChiSquare suspect.

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

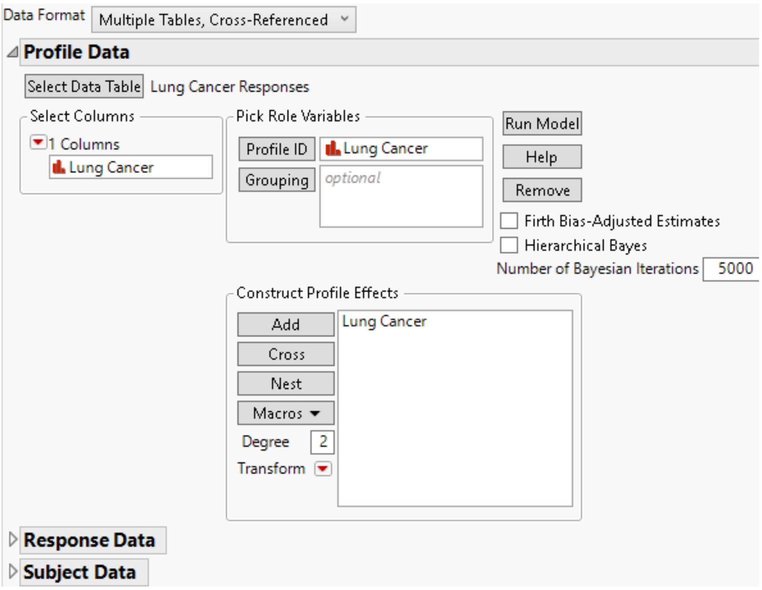
Use the Choice Platform

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.

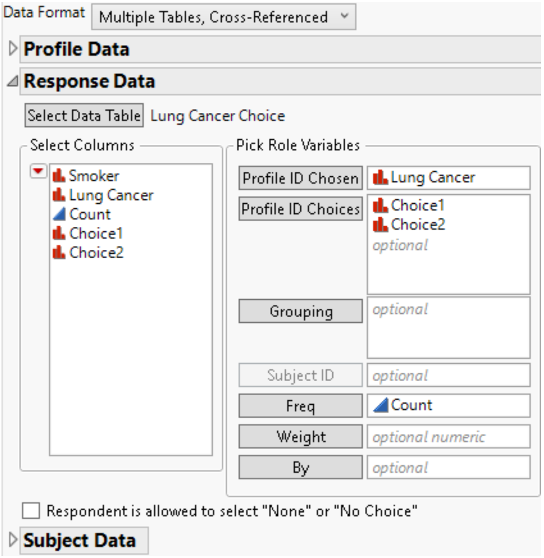
Figure 4.37 Completed Profile Data Panel



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**.

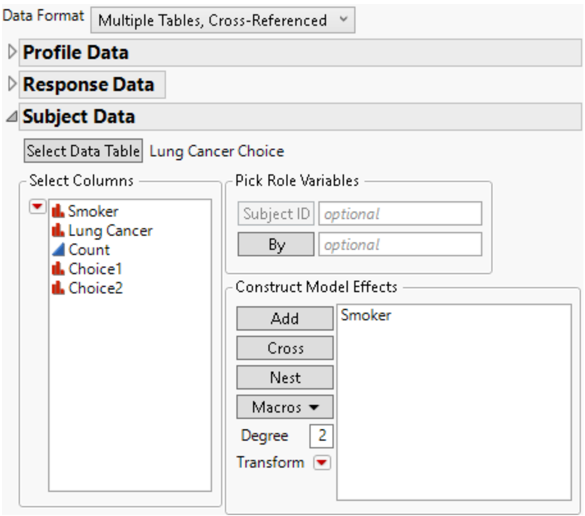
– Select Count and click **Freq**.

Figure 4.38 Completed Response Data Panel



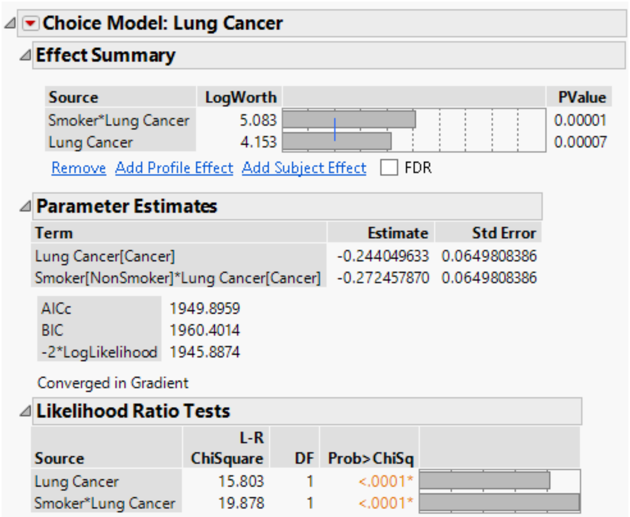
11. Open the **Subject Data** outline.
12. Click **Select Data Table**, select Lung Cancer Choice.jmp, and click **OK**.
13. Select Smoker and click **Add**.

Figure 4.39 Completed Subject Data Panel



14. Click **Run Model**.

Figure 4.40 Choice Modeling Logistic Regression Results



Choice Model: Lung Cancer

Effect Summary

Source	LogWorth	PValue
Smoker*Lung Cancer	5.083	0.00001
Lung Cancer	4.153	0.00007

Remove Add Profile Effect Add Subject Effect FDR

Parameter Estimates

Term	Estimate	Std Error
Lung Cancer[Cancer]	-0.244049633	0.0649808386
Smoker[NonSmoker]*Lung Cancer[Cancer]	-0.272457870	0.0649808386

AICc	1949.8959
BIC	1960.4014
-2*LogLikelihood	1945.8874

Converged in Gradient

Likelihood Ratio Tests

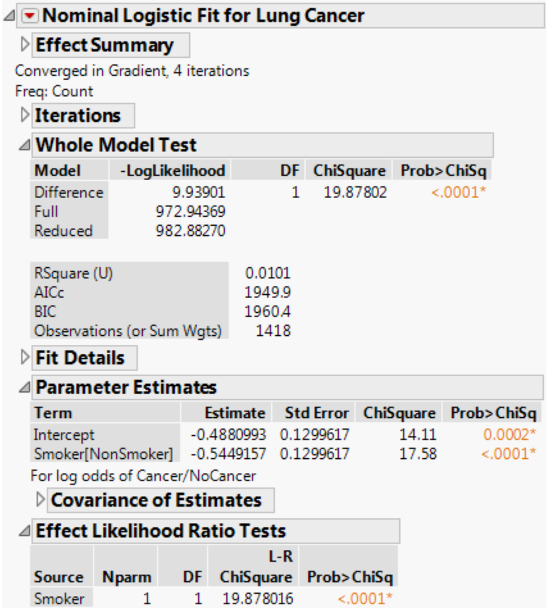
Source	L-R ChiSquare	DF	Prob>ChiSq
Lung Cancer	15.803	1	<.0001*
Smoker*Lung Cancer	19.878	1	<.0001*

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


Nominal Logistic Fit for Lung Cancer

Effect Summary

Converged in Gradient, 4 iterations

Freq: Count

Iterations

Whole Model Test

Model	-LogLikelihood	DF	ChiSquare	Prob>ChiSq
Difference	9.93901	1	19.87802	<.0001*
Full	972.94369			
Reduced	982.88270			

RSquare (U)	0.0101
AICc	1949.9
BIC	1960.4
Observations (or Sum Wgts)	1418

Fit Details

Parameter Estimates

Term	Estimate	Std Error	ChiSquare	Prob>ChiSq
Intercept	-0.4880993	0.1299617	14.11	0.0002*
Smoker[NonSmoker]	-0.5449157	0.1299617	17.58	<.0001*

For log odds of Cancer/NoCancer

Covariance of Estimates

Effect Likelihood Ratio Tests

Source	Nparm	DF	L-R ChiSquare	Prob>ChiSq
Smoker	1	1	19.878016	<.0001*

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 Models chapter in *Fitting Linear Models*.

Example of Logistic Regression for Matched Case-Control Studies

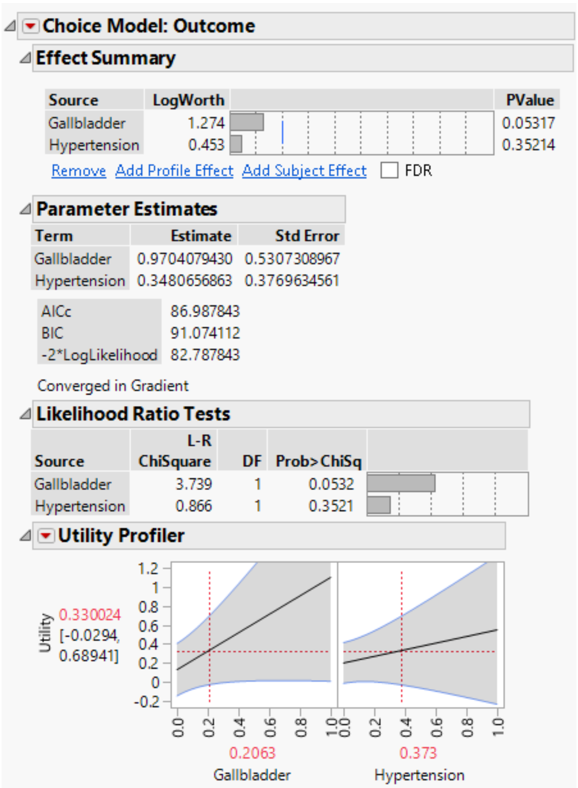
This section provides an example using the Choice platform to perform logistic regression on the results of a study of endometrial cancer with 63 matched pairs. The data are from the Los Angeles Study of the Endometrial Cancer Data reported in Breslow and Day (1980). The goal of the case-control analysis 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 more information about performing logistic regression using the Choice platform, see “Logistic Regression” on page 140.

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: Outcome red triangle and select **Utility Profiler**.

Figure 4.42 Logistic Regression on Endometrial Cancer Data



Choice Model: Outcome

Effect Summary

Source	LogWorth	PValue
Gallbladder	1.274	0.05317
Hypertension	0.453	0.35214

Remove Add Profile Effect Add Subject Effect FDR

Parameter Estimates

Term	Estimate	Std Error
Gallbladder	0.9704079430	0.5307308967
Hypertension	0.3480656863	0.3769634561

AICc	86.987843
BIC	91.074112
-2*LogLikelihood	82.787843

Converged in Gradient

Likelihood Ratio Tests

Source	L-R ChiSquare	DF	Prob>ChiSq
Gallbladder	3.739	1	0.0532
Hypertension	0.866	1	0.3521

Utility Profiler

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).

Figure 4.44 Choose Function for Choice Mode Column of Daganzo Data

$$\text{Choose}\left(\textit{Choice}\right)\begin{pmatrix}1 & \Rightarrow & \text{"Subway"}\\ 2 & \Rightarrow & \text{"Bus"}\\ \text{else} & \Rightarrow & \text{"Car"}\end{pmatrix}$$

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

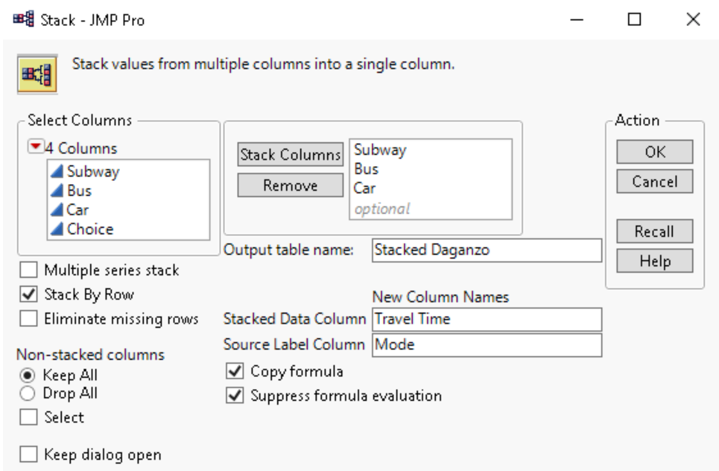
	Subway	Bus	Car	Choice	Choice Mode	Subject
1	16.481	16.196	23.89	2	Bus	1
2	15.123	11.373	14.182	2	Bus	2
3	19.469	8.822	20.819	2	Bus	3
4	18.847	15.649	21.28	2	Bus	4
5	12.578	10.671	18.335	2	Bus	5
6	11.513	20.582	27.838	1	Subway	6
7	10.651	15.537	17.418	1	Subway	7

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**.

Figure 4.49 Partial Subset Table of Daganzo Data with Choice Set

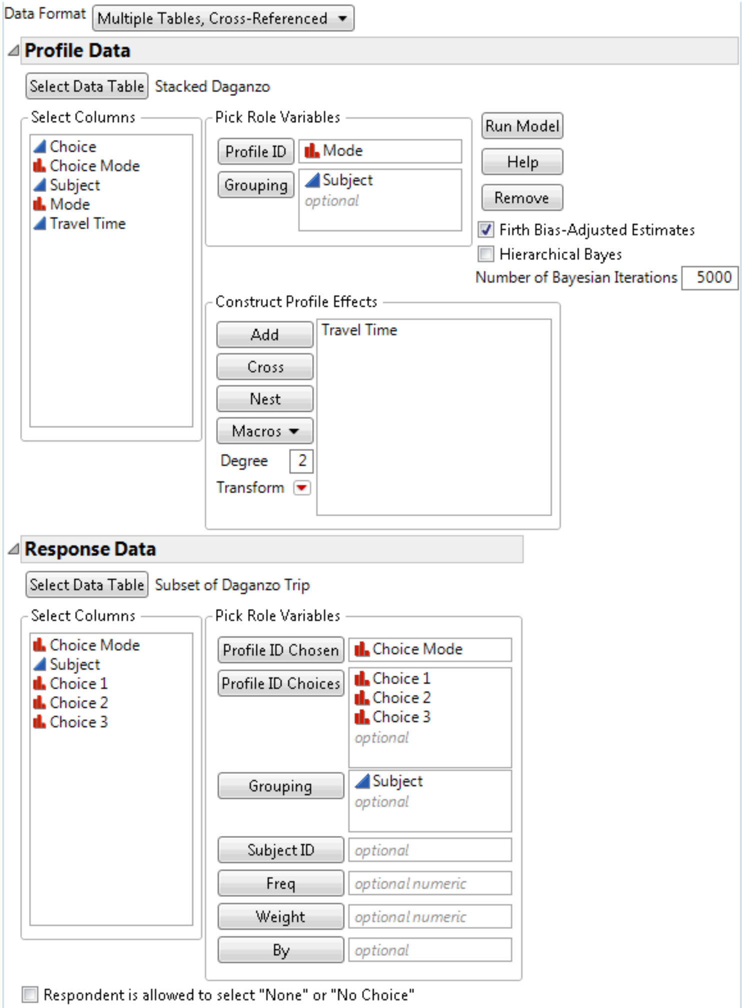
	Choice Mode	Subject	Choice 1	Choice 2	Choice 3
1	Bus	1	Bus	Subway	Car
2	Bus	2	Bus	Subway	Car
3	Bus	3	Bus	Subway	Car
4	Bus	4	Bus	Subway	Car
5	Bus	5	Bus	Subway	Car
6	Subway	6	Bus	Subway	Car
7	Subway	7	Bus	Subway	Car

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.

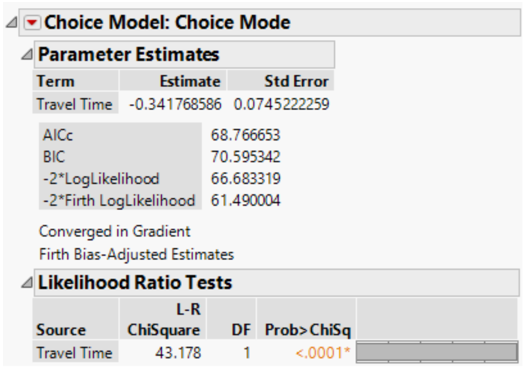
Figure 4.50 Choice Dialog Box for Subset of Daganzo Data



4. Click **Run Model**.

   The resulting parameter estimate now expresses the utility coefficient for Travel Time.

Figure 4.51 Parameter Estimate for Travel Time of Daganzo Data



Choice Model: Choice Mode

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	L-R 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 132.
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

$$\text{If}\begin{pmatrix} \text{Choice Mode} == \text{Mode} & \Rightarrow 1 \\ \text{else} & \Rightarrow 0 \end{pmatrix}$$

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

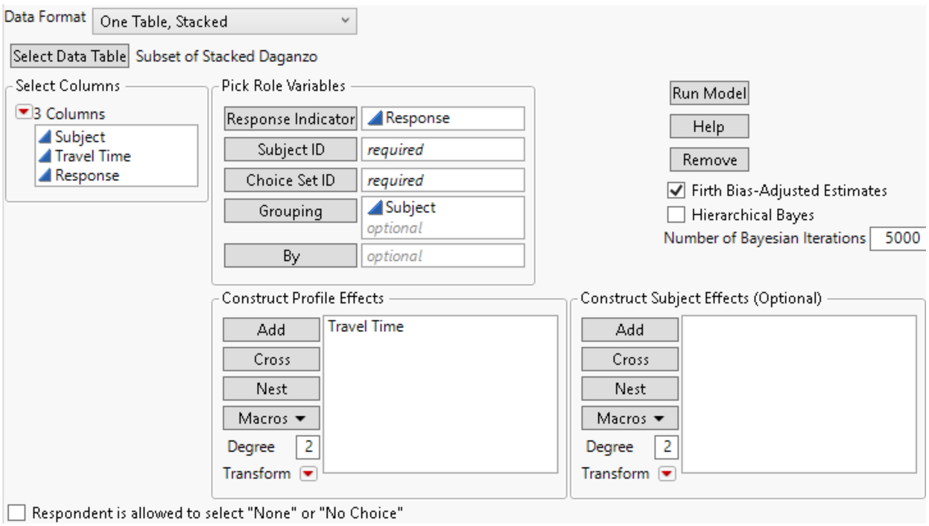
    A partial listing of the new data table is shown in Figure 4.53.

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



13. Click **Run Model.**

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

Choice Model: Response

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	L-R 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 127 and "Example of Logistic Regression for Matched Case-Control Studies" on page 130.

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 Δ results in the following:

$$\Delta = \Sigma_{ij}\Delta_{ij} = \Sigma H^{-1}g_{ij} = 0,$$

Here i is the subject index, j is the choice response index for each subject, Δij 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:

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

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.

Chapter 5

MaxDiff

Fit Models for MaxDiff Experiments

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

Contents

Overview of the MaxDiff Modeling Platform 145
Examples of the MaxDiff Platform 146
  One Table Format 146
  Multiple Table Format 149
Launch the MaxDiff Platform 152
  Launch Window for One Table, Stacked 154
  Launch Window for Multiple Tables, Cross-Referenced 156
MaxDiff Model Report 160
  Effect Summary 160
  MaxDiff Results 161
  Parameter Estimates 162
  Bayesian Parameter Estimates 164
  Likelihood Ratio Tests 165
MaxDiff Platform Options 166
  Comparisons Report 168
  Save Bayes Chain 168

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 123 in the "Choice Models" chapter. For more information about the gradient values, see "Gradients" on page 141 in the "Choice Models" chapter.

JMP PRO 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.

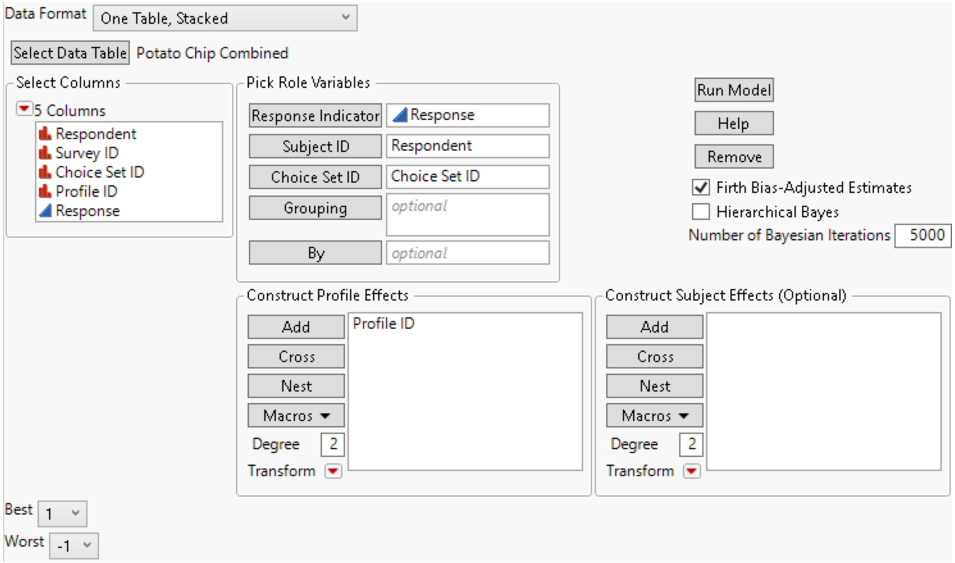
- "One Table Format" on page 146 shows how to analyze a subset of the available data in a one-table format. Note that you could add additional profile and subject data to the single table for a more complete analysis.
- "Multiple Table Format" on page 149 shows how to bring together information from 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.

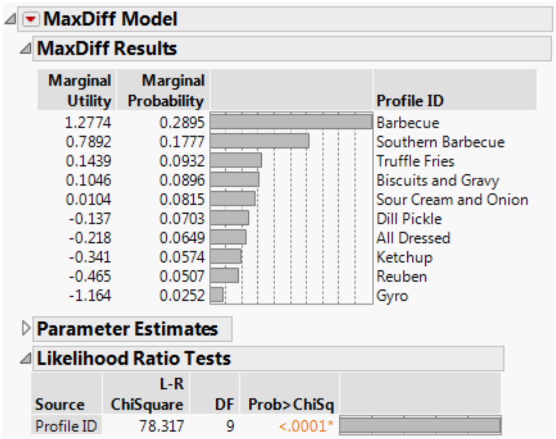
Figure 5.2 Completed MaxDiff Launch Window



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

6. Click **Run Model**.

Figure 5.3 MaxDiff Report for Potato Chip Combined.jmp



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.

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

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 151.

---

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 about 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.

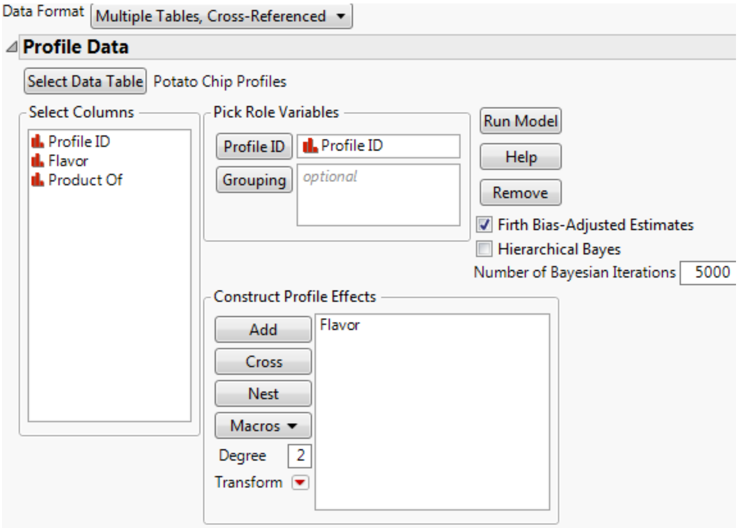
   There are three separate outlines, one for each of the data sources.
5. Click **Select Data Table** under Profile Data.

   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.**

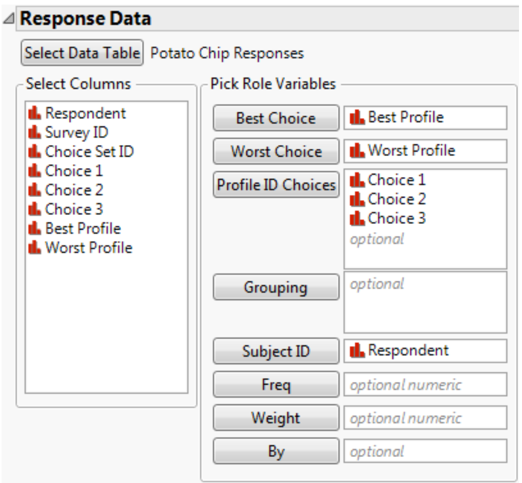
   Note that Product Of is another profile effect that you could add to the effects list.

Figure 5.5 Complete Profile Data Outline



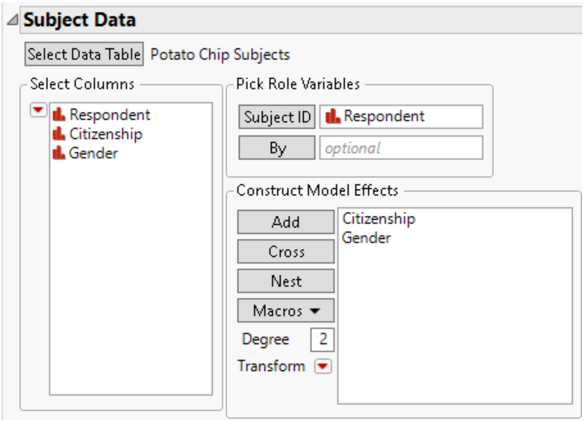
9. Open the Response Data outline. Click **Select Data Table**.
10. Select Potato Chip Responses.jmp and click **OK**.
11. Assign roles to columns as follows. The completed launch dialog is shown in Figure 5.6.
    – Select Best Profile and click **Best Choice**.
    – Select Worst Profile and click **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



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**.

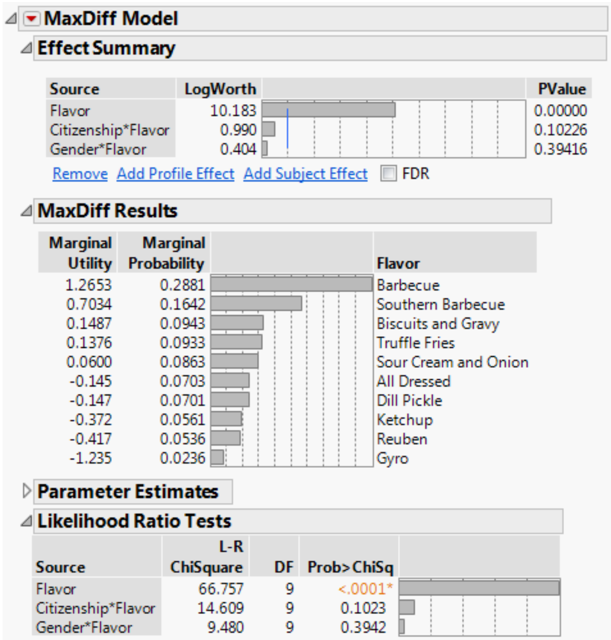
Figure 5.7 Completed Subject Data Outline



Explore the Model

1. Click **Run Model**.

Figure 5.8 MaxDiff Model Report



MaxDiff Model

Effect Summary

Source	LogWorth	PValue
Flavor	10.183	0.00000
Citizenship*Flavor	0.990	0.10226
Gender*Flavor	0.404	0.39416

Remove Add Profile Effect Add Subject Effect FDR

MaxDiff Results

Marginal Utility	Marginal Probability	Flavor
1.2653	0.2881	Barbecue
0.7034	0.1642	Southern Barbecue
0.1487	0.0943	Biscuits and Gravy
0.1376	0.0933	Truffle Fries
0.0600	0.0863	Sour Cream and Onion
-0.145	0.0703	All Dressed
-0.147	0.0701	Dill Pickle
-0.372	0.0561	Ketchup
-0.417	0.0536	Reuben
-1.235	0.0236	Gyro

Parameter Estimates

Likelihood Ratio Tests

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

The Effect Summary report 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. In the MaxDiff launch window, specify whether you are using one or multiple data tables in the Data Format list.

One Table, Stacked

For the One Table, Stacked format, the data are in a single data table. There is a row for every profile presented to a subject within a choice set and an indicator for the best and worst profiles in that choice set. The Potato Chip Combined.jmp sample data table contains the results of a MaxDiff experiment in a single table format. See "One Table Format" on page 146.

For more information about the launch window for this format, see “Launch Window for One Table, Stacked” on page 154.

Multiple Tables, Cross-Referenced

For the Multiple Tables, Cross-Referenced format, the data are in two or three separate data tables. A profile data table and a response data table are required. A subject data table is optional. Note the following:

- The profile data table must contain a column with a unique identifier for each profile and columns for the profile level variables. The profile identifier is used in the response data table to identify best and worst profile responses for each choice set.
- The optional subject data table must contain a column with a unique subject identifier for each subject and columns for the subject level variables. The subject identifier is used in the response data table to identify the subjects.

The launch window for this format contains three sections: Profile Data, Response Data, and Subject Data. Each section corresponds to a different data table. You can expand or collapse each section as needed.

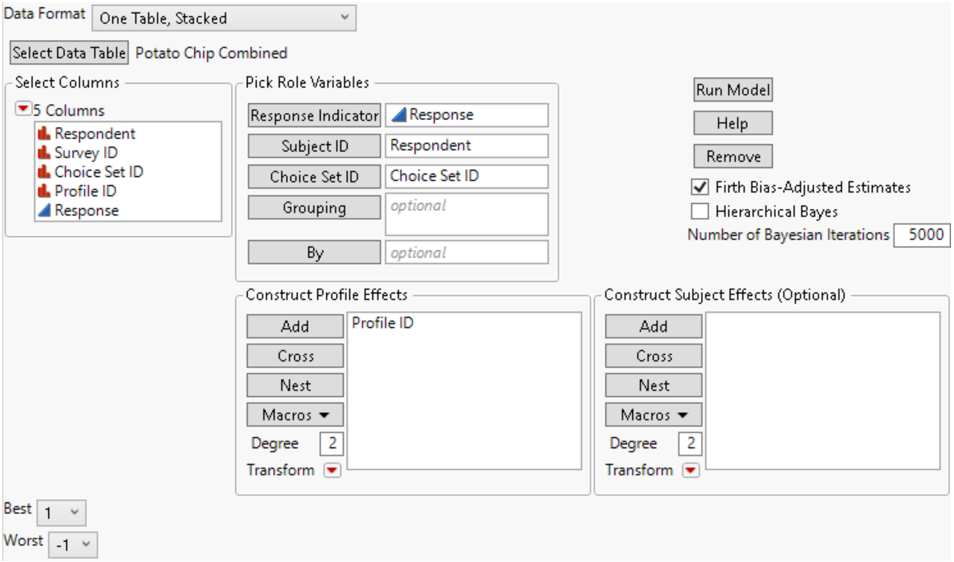
The Potato Chip Profiles.jmp, Potato Chip Responses.jmp, and Potato Chip Subjects.jmp sample data tables contain the results of a MaxDiff experiment using three tables. See “Multiple Table Format” on page 149.

For more information about the launch window for this format, see “Launch Window for Multiple Tables, Cross-Referenced” on page 156.

Launch Window for One Table, Stacked

Launch the MaxDiff platform by selecting **Analyze > Consumer Research > MaxDiff**. For one table select **One Table, Stacked** from the Data Format menu.

Figure 5.9 Launch Window for One Table, Stacked Data Format



For more information about the options in the Select Columns red triangle menu, see the Get Started chapter in *Using JMP*.

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, or 0 for the Best and Worst choices, and the third value for profiles that are not Best or Worst. The default coding is a 1 to indicate the Best choice and a -1 for 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.

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 By variables.

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

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

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

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

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. See Heinze and Schemper (2002) for a discussion of the separation problem in logistic regression.

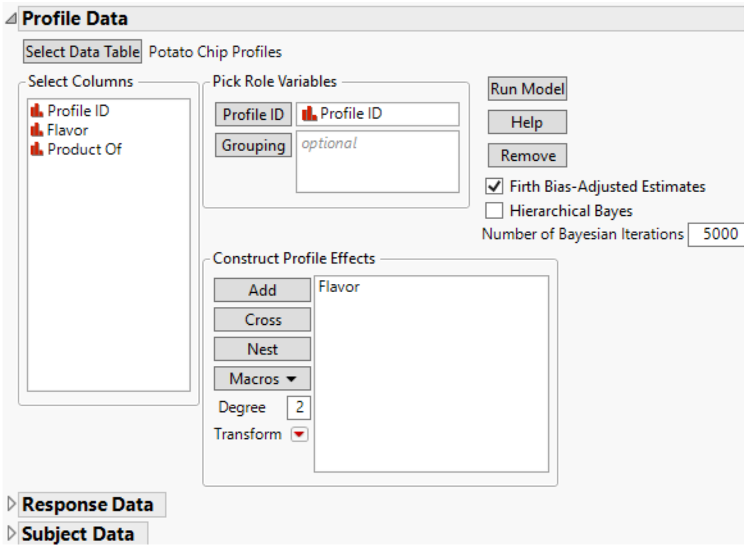
JMP PRO **Hierarchical Bayes** Uses a Bayesian approach to estimate subject-specific parameters. See "Bayesian Parameter Estimates" on page 164.

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.

Launch Window for Multiple Tables, Cross-Referenced

Launch the MaxDiff platform by selecting **Analyze > Consumer Research > MaxDiff**. For multiple tables select **Multiple Tables, Cross-Referenced** form the Data Format menu.

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



For more information about the options in the Select Columns red triangle menu, see the Get Started chapter in *Using JMP*.

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

- "Profile Data" on page 156
- "Response Data" on page 157
- "Subject Data" on page 158

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 Model Specification chapter in *Fitting Linear Models*.

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. See Heinze and Schemper (2002) for a discussion of the separation problem in logistic regression.

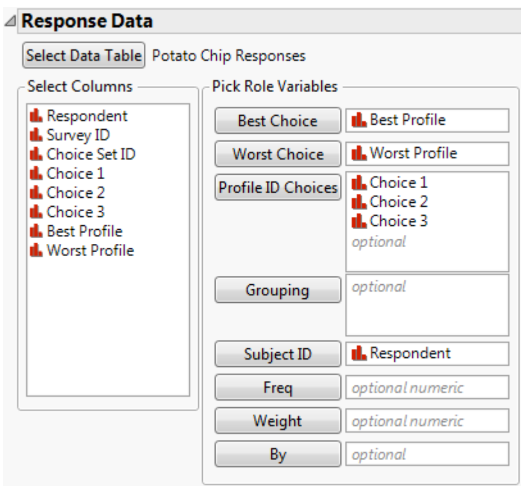
Hierarchical Bayes Uses a Bayesian approach to estimate subject-specific parameters. See “Bayesian Parameter Estimates” on page 164.

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 response 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 study participant designated as Best.

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

Profile ID Choices The columns that contain the **Profile ID**s of the set of possible choices for each choice set. There must be at least three profiles.

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.

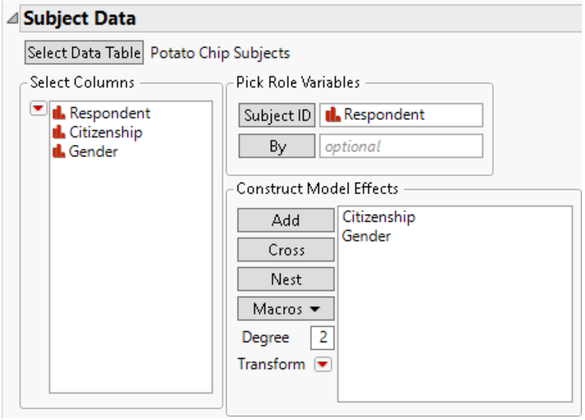
Weight A column containing a weight for each observation in the data table. The weight is included in analyses only when its value is greater than zero.

By 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 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.

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 By variables.

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

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

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 166.

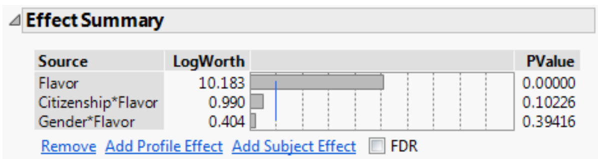
Effect Summary

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. See the Standard Least Squares Report and Options chapter in *Fitting Linear Models*.

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

Source	LogWorth	PValue
Flavor	10.183	0.00000
Citizenship*Flavor	0.990	0.10226
Gender*Flavor	0.404	0.39416

Remove Add Profile Effect Add Subject Effect FDR

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 $-\log_{10}$(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 more information about the FDR correction, see Benjamini and Hochberg (1995). For more information about the false discovery rate, see the Response Screening chapter in *Predictive and Specialized Modeling* 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.

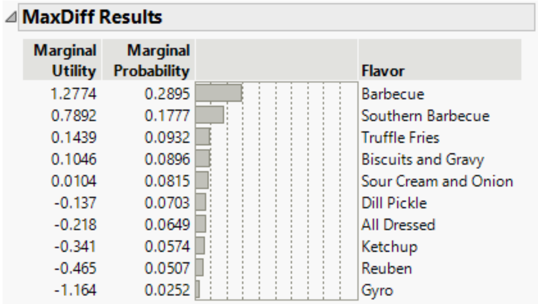
Add Profile Effect Opens a column dialog that contains a list of all columns in the data table for the OneTable, Stacked data format, and for the columns in the Profile Data table for the Multiple Tables, Cross-Referenced data format. Select columns that you want to add to the model, and then click Add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

Add Subject Effect Opens a column dialog 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



MaxDiff Results

Marginal Utility	Marginal Probability	Flavor
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

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 Chart Shows a bar chart 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), –2Loglikelihood, and –2Firth Loglikelihood. See the Statistical Details appendix in *Fitting Linear Models* for more information about the first three of these measures.

The –2Firth 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.

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 167.

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 multinominal 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]	-0.444803533	0.1857067541	0.2930798802
Citizenship[Canadian]*Product Of[Canada]	-0.103783512	0.2338637067	0.1945040440
Gender[Female]*Product Of[Canada]	-0.258577340	0.2205559716	0.1846953243

Total Iterations	5000
Burn-In Iterations	2500
Number of Respondents	30
Avg Log Likelihood After Burn-In	-196.5618

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.

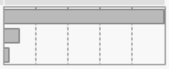
Likelihood Ratio Tests

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 ChiSquare	DF	Prob>ChiSq
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 Chart Shows a bar chart of the L-R ChiSquare values.

MaxDiff Platform Options

The MaxDiff Model red triangle menu contains the following options:

JMP PRO **Show MLE Parameter Estimates** (Available for Hierarchical Bayes.) 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 for Hierarchical.) 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 *Fitting Linear Models*.

Confidence Intervals (Not available for Hierarchical Bayes) Shows or hides a confidence interval for each parameter in the Parameter Estimates report.

Confidence Limits (Available for Hierarchical Bayes) Shows or hides 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 or hides the correlations between the maximum likelihood parameter estimates.

For Hierarchical Bayes, shows or hides 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 168.

All Levels Comparison Report Shows the All Levels Comparison Report, which contains a table with information about 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.

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 (Not available for Hierarchical Bayes.) 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. See "Example of Segmentation" on page 123 in the "Choice Models" chapter.

JMP PRO **Save Subject Estimates** (Available for Hierarchical Bayes.) 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 164.

JMP PRO **Save Bayes Chain** (Available for Hierarchical Bayes.) Creates a table that gives information about the chain of iterations used in computing subject-specific Bayesian estimates. See "Save Bayes Chain" on page 168.

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 *Using JMP* 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 1 Probability 1 divided by Probability 2.

Odds 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.

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.

Chapter 6

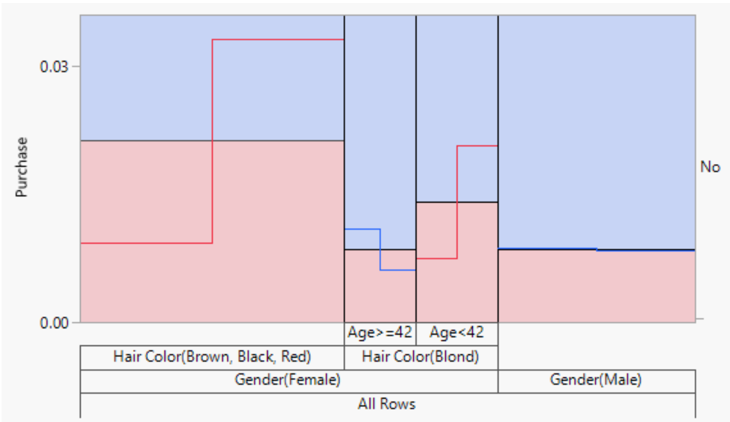
Uplift Models

Model the Incremental Impact of Actions on Consumer Behavior

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 173
Example of the Uplift Platform 174
Launch the Uplift Platform 175
The Uplift Model Report 176
    Uplift Model Graph 177
    Uplift Report Options 179

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. Although traditional partition models select splits to optimize classification, uplift models select splits to maximize treatment differences.

The uplift partition model accounts for the fact that individuals are grouped by a treatment factor. To determine splits, models are fit to all possible binary splits of each factor. The type of model that is fit is dependent on the type of response. 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 either case, the interaction term measures the difference in uplift between the groups of individuals in the two splits. The most significant split is selected and the process repeats.

The Uplift platform selects the most significant split based on the significance of interaction terms in each of the binary split models. However, predictor selection based solely on p-values introduces bias in favor of predictors with many levels that result in many models for the single predictor. For this reason, JMP adjusts p-values to account for the number of levels or models considered. The correction used is based on Monte Carlo simulation. See Sall (2002). The splits are determined by the minimum adjusted p-values for tests of the significance of the interaction effect across models for all possible binary splits across all predictors. The logworth for each adjusted p-value, namely $-\log_{10}$(adj p-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.

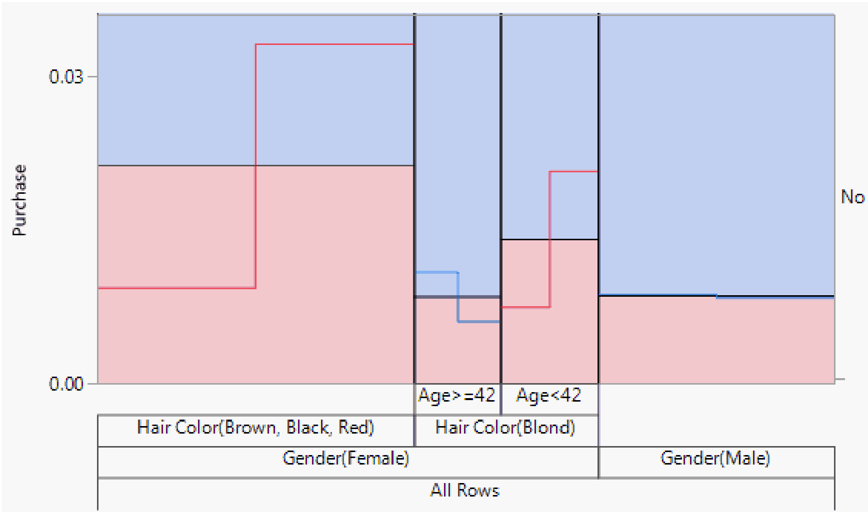
The data table shows a Promotion column, indicating whether the member received promotional material. The column Purchase indicates whether the member purchased the product over the test period. For each member, the following information was assembled: Gender, Age, Hair Color (natural), U.S. Region, 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 Order 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. Note that the vertical scale has been modified in order to show the detail.

Figure 6.2 Graph after Three Splits

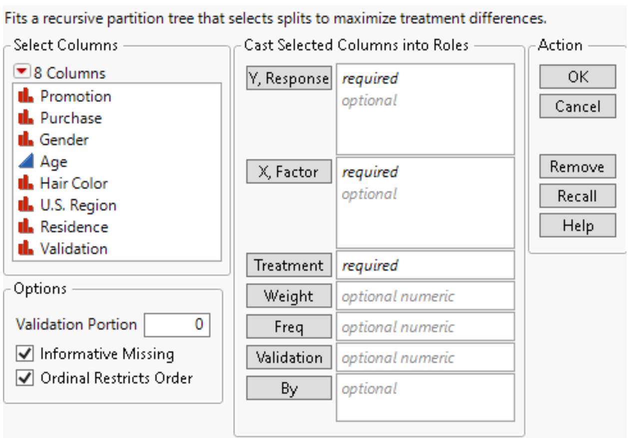


The graph indicates that uplift in purchases occurs for females with black, red, or brown hair and for younger females (Age < 42) with blond hair. For older blond-haired women (Age ≥ 42) and males, the promotion has a negative effect.

Launch the Uplift Platform

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

Figure 6.3 Uplift Platform Launch Window



For more information about the options in the Select Columns red triangle menu, see the Get Started chapter in *Using JMP*.

Y, Response Assigns one or more columns to be analyzed.

X, Factor Assigns one or more columns to be used as factors.

Treatment Assigns a categorical treatment column. If the treatment column contains more than two levels, the first level is treated as one treatment level and the remaining levels are combined into a second treatment level.

Weight Assigns a numeric column that contains a weight for each observation in the data table. A row is included in the analysis only when its weight is greater than zero.

Freq Assigns a frequency variable to this role. This is useful for summarized data.

JMP PRO **Validation** Assigns a numeric column that defines a validation set. This column should contain at most three distinct values. 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. See the Make Validation Column chapter in *Predictive and Specialized Modeling*.

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 variable.

The following options are also available:

Validation Portion The portion of the data to be used as a validation set. Enter a value between 0 and 1.

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"

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 appears only 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 levels except the first 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 might 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 can have a low RSquare value.

RMSE The root mean square error (RMSE) for the regression model associated with the tree. RMSE is given only for continuous responses. See *Fitting Linear Models*.

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 given only for continuous responses. See the Statistical Details appendix in *Fitting Linear Models*.

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 (Appears only for two-level categorical responses.) For each treatment level, the proportion of subjects in this node who responded.

Mean (Appears only 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.

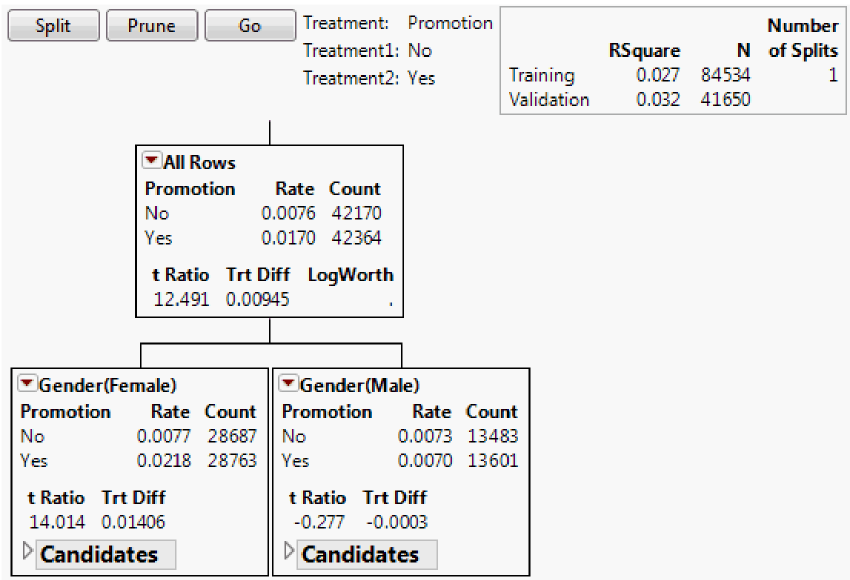
Trt Diff The difference in response means across the levels of Treatment. This is the uplift, with the following assumptions:

– 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 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 more information about these options, see the Partition Models chapter in *Predictive and Specialized Modeling*.

JMP 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 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 regression model.

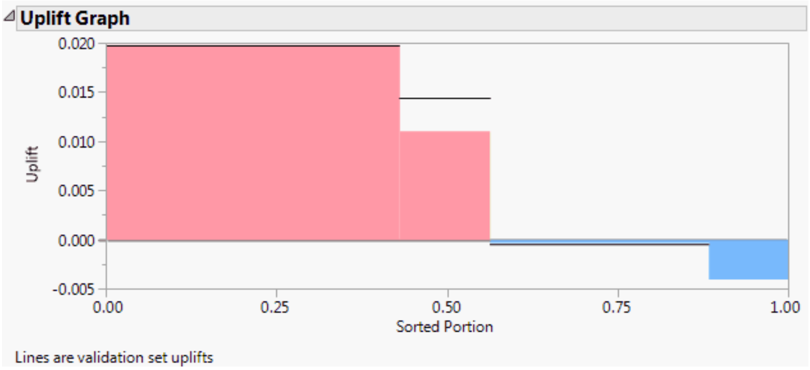
JMP PRO Uplift Graph

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 ≥ 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



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 *Predictive and Specialized Modeling*.

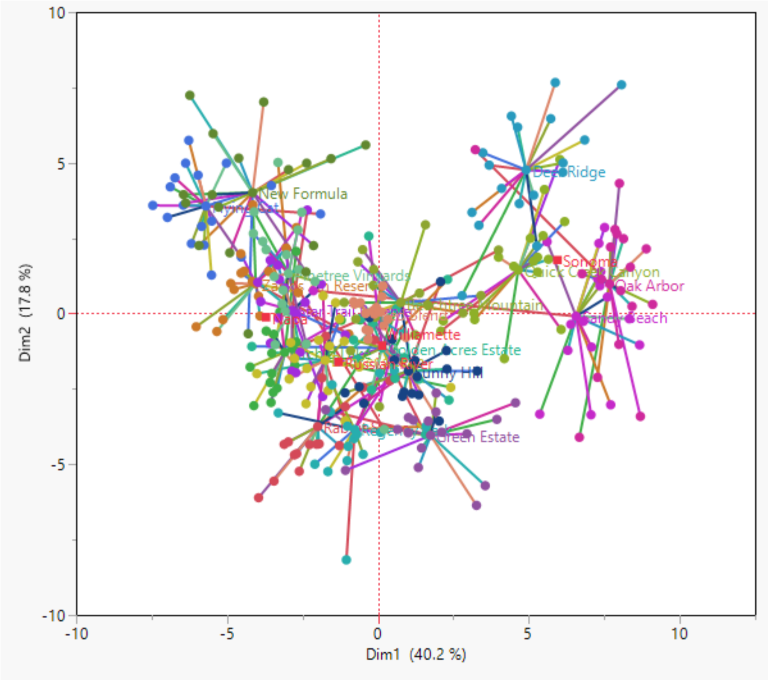
Chapter **7**

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 183
Example of Multiple Factor Analysis 184
Launch the Multiple Factor Analysis Platform 187
Data Format 188
The Multiple Factor Analysis Report 190
Summary Plots 190
Consensus Map 191
Multiple Factor Analysis Platform Options 191
Statistical Details for the Multiple Factor Analysis Platform 194

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

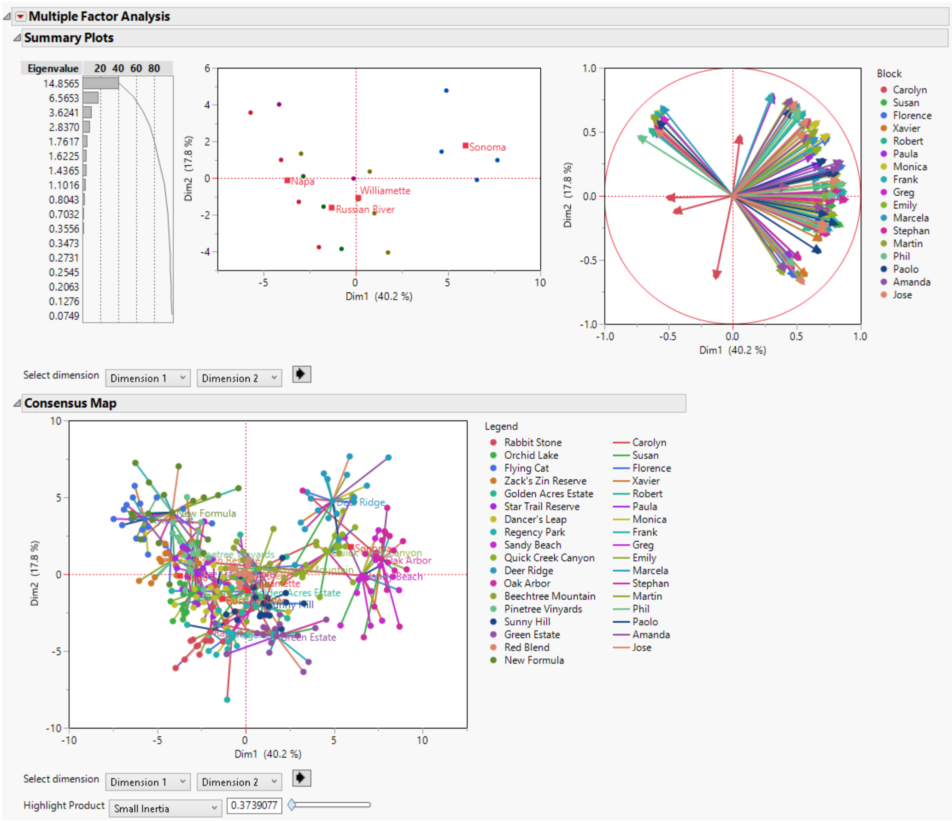
This example uses data from a simulated sensory panel study of wine characteristics. Participants rated 16 wines on a number of characteristics from 1 (no intensity) to 10 (prominent intensity). You want to better understand how the 16 wines are similar or different.

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 to Jose 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



Tip: To set the legend in the Consensus Map to two columns, double-click the legend and set the Item Wrap in the Legend Settings to 18.

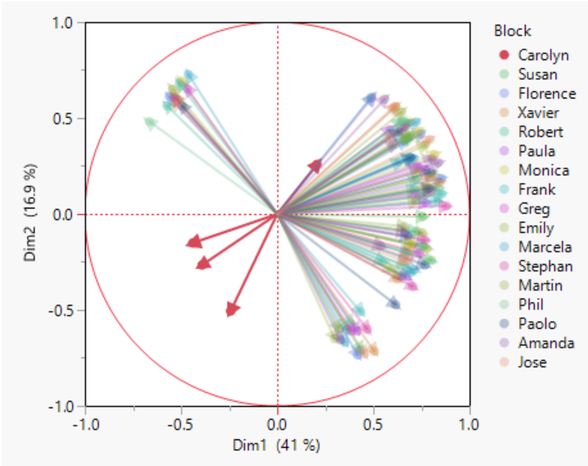
Notice the following in the Summary Plots:

– 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** to highlight her results.

Figure 7.3 Loading Plot with Results for Carolyn Highlighted

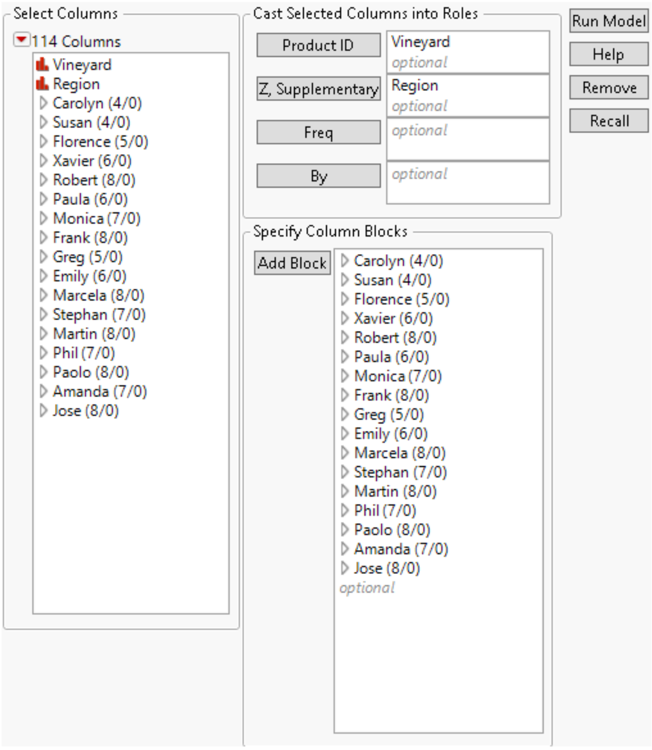


Carolyn's results differ from the other panelists. You might want to re-run the analysis without her results. See Figure 7.6 on page 190 for the results of the analysis without Carolyn.

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



For more information about the options in the Select Columns red triangle menu, see the Get Started chapter in *Using JMP*.

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.

By A column or columns whose levels define separate analyses. For each level of the specified column, the corresponding rows are analyzed using the other variables that you have specified. The results are presented in separate reports. If more than one By variable is assigned, a separate report is produced for each possible combination of the levels of the By variables.

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.

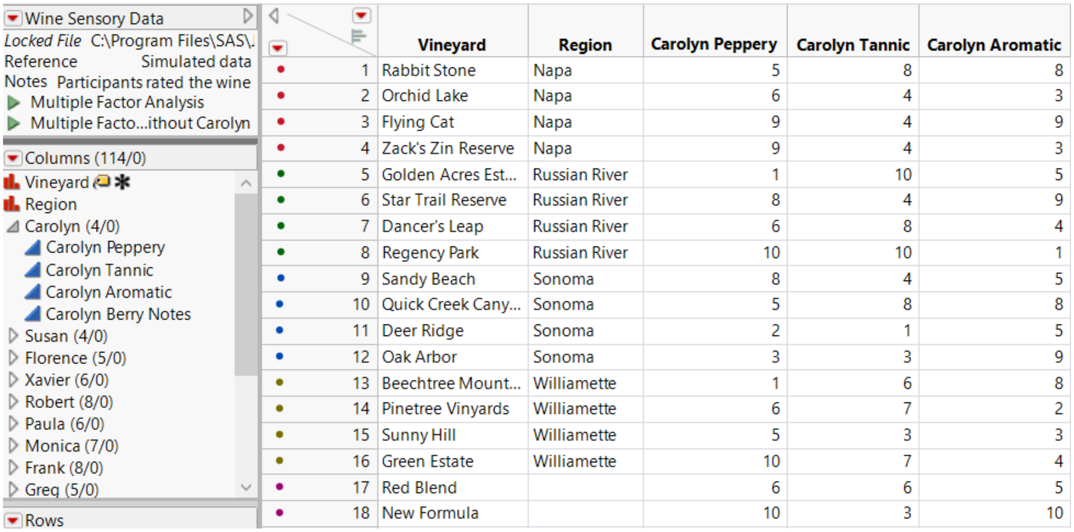
Tip: If you group the columns into blocks before running the platform, you can select multiple column groups and cast them as blocks in a single action. Otherwise, you must select each group of columns for each block and click Add Block, one block at a time. Double-click a block name to change it.

Data Format

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. The Vineyards in rows 17 and 18 are not assigned a Region. The analysis could be used to explore which region the vineyards are most aligned to.

Figure 7.5 Partial View of a Data Table for Multiple Factor Analysis



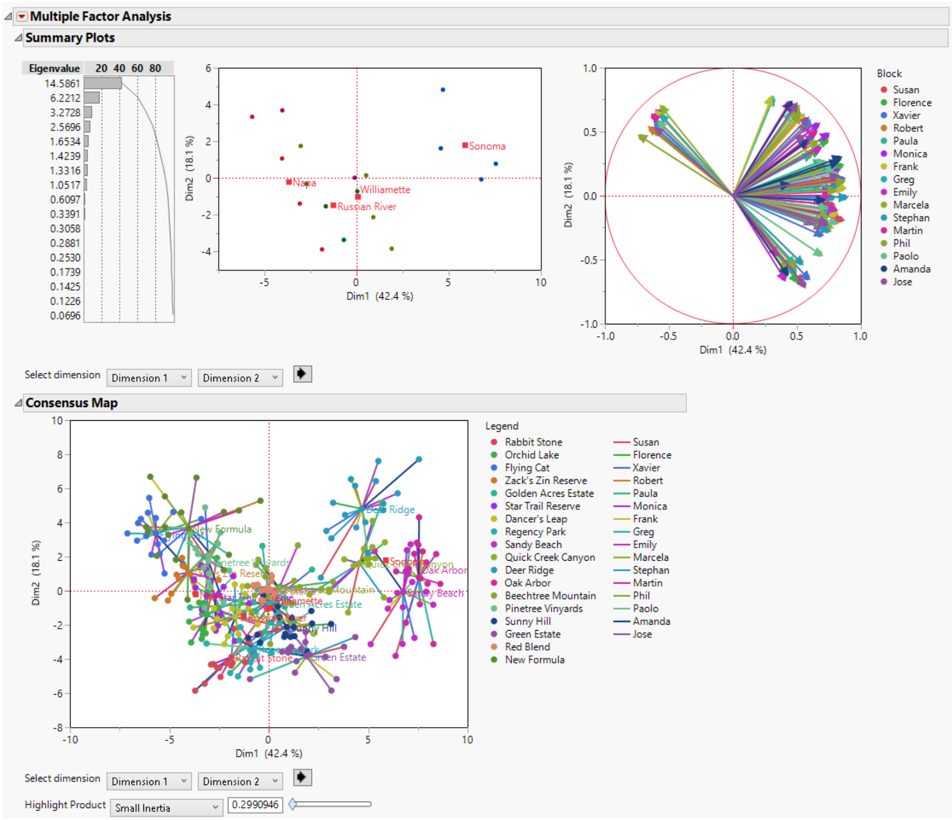
	Vineyard	Region	Carolyn Peppery	Carolyn Tannic	Carolyn Aromatic
1	Rabbit Stone	Napa	5	8	8
2	Orchid Lake	Napa	6	4	3
3	Flying Cat	Napa	9	4	9
4	Zack's Zin Reserve	Napa	9	4	3
5	Golden Acres Est...	Russian River	1	10	5
6	Star Trail Reserve	Russian River	8	4	9
7	Dancer's Leap	Russian River	6	8	4
8	Regency Park	Russian River	10	10	1
9	Sandy Beach	Sonoma	8	4	5
10	Quick Creek Cany...	Sonoma	5	8	8
11	Deer Ridge	Sonoma	2	1	5
12	Oak Arbor	Sonoma	3	3	9
13	Beechtree Mount...	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
17	Red Blend		6	6	5
18	New Formula		10	3	10

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 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. Place your cursor over a data point to view the block label for that point. Click a product ID in the legend to highlight it in the consensus map.

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 dimensions, in order, from largest to smallest.

Eigenvectors Shows or hides a table of the eigenvectors for each of the consensus dimensions, in order, from left to right. Using these coefficients to form a linear combination of the original variables produces the consensus principal component variables.

Variable Loadings Shows or hides the loadings for each column. As in principal components analysis, loadings represent correlations of variables with components. Values near zero indicate the variable has little effect on the consensus dimension.

Variable Partial Contributions Shows or hides a table that contains the partial contributions of variables. The partial contributions represent the percentage of variance that each variable contributes to the consensus dimension.

Variable Squared Cosines Shows or hides a table that contains the squared cosines of variables. The sum of the squared cosine values across consensus dimensions is equal to 1 (100%) for each variable. The squared cosines represent the overlap in variance between variables and dimensions.

Tip: For the variable loadings, variable partial contributions, and variable squared cosines, values near zero indicate the variable is weakly related to the consensus dimension. Values far from zero indicate a strong association. 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 191.

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 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 block name labels on all points in the consensus map and bi-plot. Shows or hides column name labels on all points in the partial axes plot.

Tip: Use row labels to identify centroids on the consensus map and data points on the loading 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. 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 partial contributions 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 *Using JMP* 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 **X** matrix can be defined as follows:

$$X = P\Delta Q^T$$ with the constraint $$P^T M P = Q^T A Q = I$$

The matrices use are as follows:

X is an n x p centered and normalized matrix of sub-tables. In consumer research there are n products and p panelists' ratings.

Q is a p x q matrix of right singular vectors, which are weighted by the MFA singular values to obtain the loadings on q principal components.

Δ is a q x 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 x q matrix of left singular vectors, which are weighted by the MFA singular values to obtain the q principal components of the compromise.

M is the n x n diagonal matrix of mass weights.

A is the p x 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.

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. Accessed April 25, 2016. https://eml.berkeley.edu/reprints/mcfadden/zarembka.pdf.

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. Accessed January 13, 2017. http://www.stochasticsolutions.com/pdf/sig-based-up-trees.pdf.

Rossi, P. E., Allenby, G. M., and McCulloch, R. (2005). *Bayesian Statistics and Marketing*. Chichester, UK: John Wiley & Sons.

Sall, J. (2002). "Monte Carlo Calibration of Distributions of Partition Statistics." SAS Institute Inc.,Cary, NC. Accessed July 29, 2015. https://www.jmp.com/content/dam/jmp/documents/en/white-papers/montecarlocal.pdf.

SAS Institute Inc. (2018). "The TTEST Procedure." In *SAS/STAT 15.1 User's Guide.* Cary NC: SAS Institute Inc. https://support.sas.com/documentation/onlinedoc/stat/151/ttest.pdf.

Train, K. E. (2001). "A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit." Department of Economics, University of California, Berkley. Accessed January 13, 2017. https://eml.berkeley.edu/~train/compare.pdf.

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.

Appendix B

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.

- Progressë® Telerik® UI for WPF: Copyright © 2008-2019 Progress Software Corporation. All rights reserved. Usage of the included Progress® Telerik® UI for WPF 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.
- 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

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.

- Pako - Copyright © 2014–2017 by Vitaly Puzrin and Andrei Tuputcyn.

  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.

- dmlc/xgboost is Copyright © 2019 SAS Institute.

  Licensed under the Apache License, Version 2.0 (the "License"); you may not use this file except in compliance with the License. You may obtain a copy of the License at

  http://www.apache.org/licenses/LICENSE-2.0

  Unless required by applicable law or agreed to in writing, software distributed under the License is distributed on an "AS IS" BASIS, WITHOUT WARRANTIES OR CONDITIONS OF ANY KIND, either express or implied. See the License for the specific language governing permissions and limitations under the License.

- libzip is Copyright (C) 1999-2019 Dieter Baron and Thomas Klausner.

  This file is part of libzip, a library to manipulate ZIP archives. The authors can be contacted at <libzip@nih.at>.

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

  THIS SOFTWARE IS PROVIDED BY THE AUTHORS ``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 AUTHORS 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.