

JMP[®] Software: Statistical Data Exploration

Course Notes

JMP® Software: Statistical Data Exploration Course Notes was developed by Monica Beals, Marissa Langford, Paul Marovich, Lori Rothenberg, Laura Ryan, and Catherine Truxillo. Additional contributions were made by Mark Bailey, Paul Terrill, and Melinda Thielbar. Editing and production support was provided by the Curriculum Development and Support Department.

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

JMP® Software: Statistical Data Exploration Course Notes

Copyright © 2007 SAS Institute Inc. Cary, NC, USA. All rights reserved. Printed in the United States of America. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, or otherwise, without the prior written permission of the publisher, SAS Institute Inc.

Book code E1068, course code LWJDEX7, prepared date 10Aug2007.

LWJDEX7_001

Table of Contents

Course Description	v
Prerequisites	vi
Chapter 1 Data Retrieval.....	1-1
1.1 Orientation	1-3
1.2 Accessing Data.....	1-17
1.3 The Data Table.....	1-34
1.4 Chapter Summary	1-48
Chapter 2 Data Exploration and Manipulation	2-1
2.1 The Columns and Rows Menus	2-3
2.2 Creating New Columns.....	2-27
2.3 Creating New Tables.....	2-39
2.4 Using the Tabulate Option to Create Summary Tables	2-53
2.5 Chapter Summary	2-64
Chapter 3 Graphical Data Exploration	3-1
3.1 Exploring Relationships Between Continuous Columns	3-3
3.2 Examining Relationships Between Attribute Data Columns	3-23
3.3 Exploratory Analysis Using Recursive Partitioning	3-28
3.4 Recursive Partitioning Details (Optional).....	3-38
3.5 Chapter Summary	3-44
Chapter 4 Reporting and Presenting Results.....	4-1
4.1 Saving Results.....	4-3

4.2	Creating a Project Journal.....	4-8
4.3	Creating a New Project.....	4-15
4.4	Chapter Summary	4-21

Appendix A Exercises and Solutions..... **A-1**

Exercises.....	A-2
Chapter 1 Exercises	A-2
Chapter 2 Exercises	A-2
Chapter 3 Exercises	A-4
Chapter 4 Exercises	A-5
Solutions to Exercises	A-6
Chapter 1 Solutions to Exercises	A-6
Chapter 2 Solutions to Exercises	A-12
Chapter 3 Solutions to Exercises	A-23
Chapter 4 Solutions to Exercises	A-31

Appendix B Index..... **B-1**

Course Description

The JMP® Software: Statistical Data Exploration course introduces JMP, its philosophy, and its extensive graphical data exploration capabilities. Students learn to perform elementary exploratory data analysis (EDA) and discover natural patterns in data. Students use JMP to learn to navigate through menus, dialog boxes, and results; manage data; save and present results in various formats; use decision trees in the EDA context; and obtain and interpret descriptive statistics.

To learn more...



A full curriculum of general and statistical instructor-based training is available at any of the Institute's training facilities. Institute instructors can also provide on-site training.

For information on other courses in the curriculum, contact the SAS Education Division at 1-800-333-7660, or send e-mail to training@sas.com. You can also find this information on the Web at support.sas.com/training/ as well as in the Training Course Catalog.



For a list of other SAS books that relate to the topics covered in this Course Notes, USA customers can contact our SAS Publishing Department at 1-800-727-3228 or send e-mail to sasbook@sas.com. Customers outside the USA, please contact your local SAS office.

Also, see the Publications Catalog on the Web at support.sas.com/pubs for a complete list of books and a convenient order form.

Prerequisites

Before selecting this course, you should be familiar with a Macintosh or Microsoft Windows operating system.

Chapter 1 Data Retrieval

1.1 Orientation	1-3
1.2 Accessing Data.....	1-17
1.3 The Data Table	1-34
1.4 Chapter Summary.....	1-48

1.1 Orientation

Objectives

- Describe some key components and features of JMP software.
- Identify user-assistance options.



Navigating with JMP Software

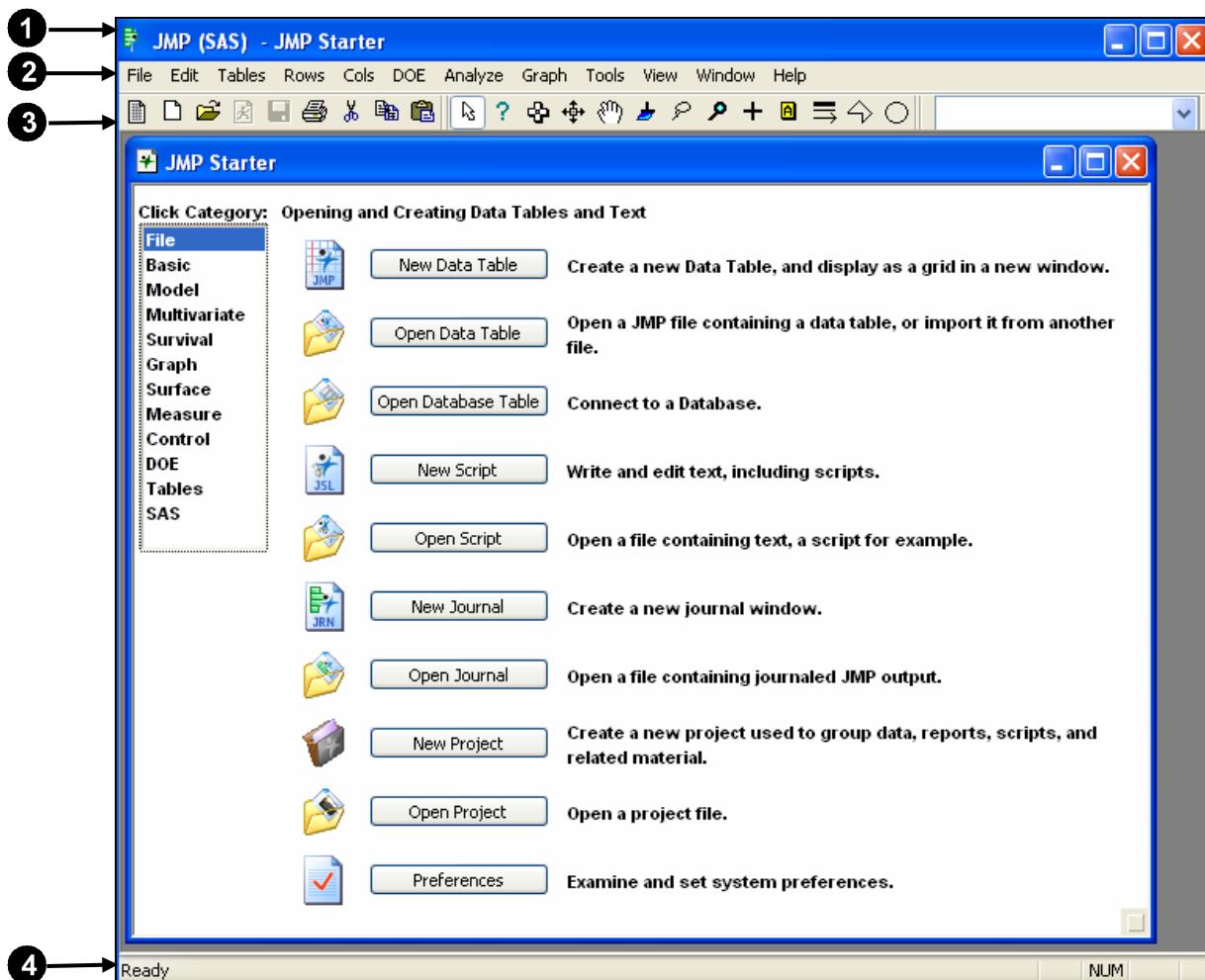
This demonstration illustrates opening JMP and navigating the software interface.

There are several ways to launch JMP. Choose one of the following methods:

- Select **Start** ⇒ **Programs** ⇒ **JMP 7** ⇒ **JMP 7**.
- Right-click the JMP icon on the Windows desktop and select **Open** in the context menu.
- Double-click the JMP icon.

The program opens with a brief animation and information about your copy of JMP. After this initial screen closes, you will see the Tip of the Day window. Close the Tip of the Day window for now by selecting **Close**. (You will return to this later.) Now the most apparent feature is the JMP Starter window, but there are a number of other important components that must be identified.

The JMP Window

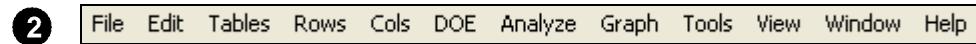


JMP presents an application window with typical Windows features and controls, such as a title bar, a menu bar, buttons on toolbars, and a status bar.

The title bar, JMP Starter in this case, is at the top of the screen. It tells you the name of the open, selected window.



The menu bar has several choices of lists for more options. For example, File, Edit, and Tables are the first three menus on the menu bar.



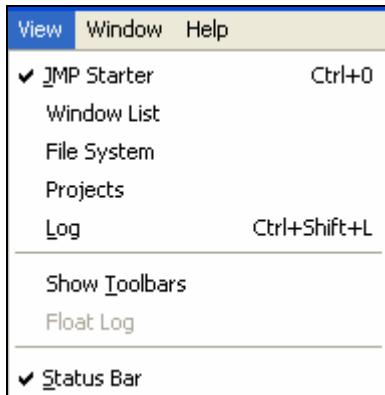
Buttons on toolbars are shortcuts for performing certain data table operations. The first three buttons on the toolbar in JMP are for creating a new data table, for creating a new script, and for opening an existing file.



The status bar is the area along the bottom edge of the JMP application window. The status bar displays messages (for example, closing database connections) and indicates on the right side whether modifier keys are active (for example, NUM, to indicate number lock).



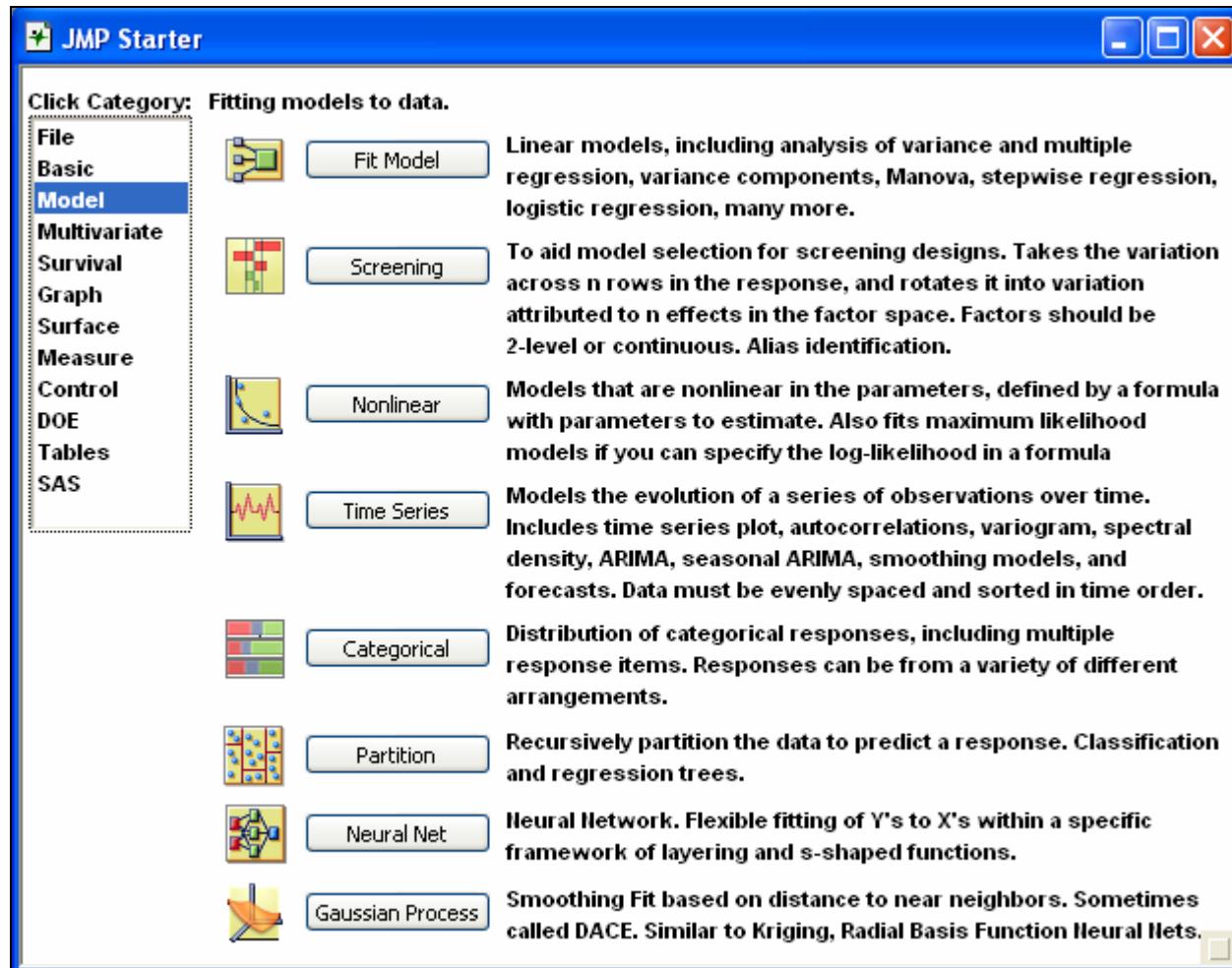
You are not required to use JMP Starter. You can close it and use commands on the menu bar or buttons on the toolbars if you prefer. You can always return to the JMP Starter window by opening it from the View menu. Select View \Rightarrow JMP Starter and the check mark appears, as shown below.



JMP Starter

The JMP Starter window is one way to navigate in JMP. It presents a collection of starting points, grouped by categories that organize platforms and commands by task with descriptions to guide your selections.

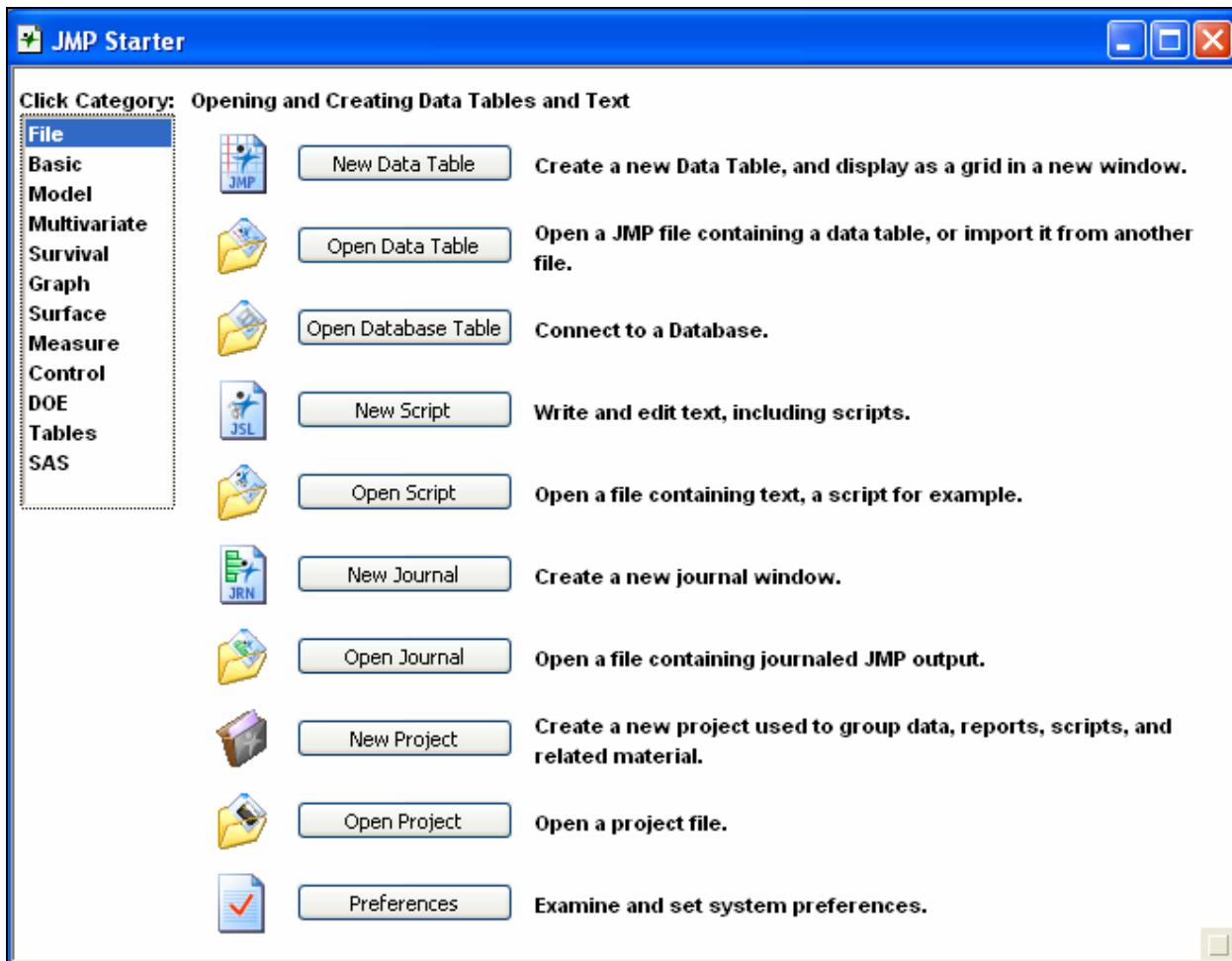
For example, if you select the **Model** category, then JMP Starter presents the following:



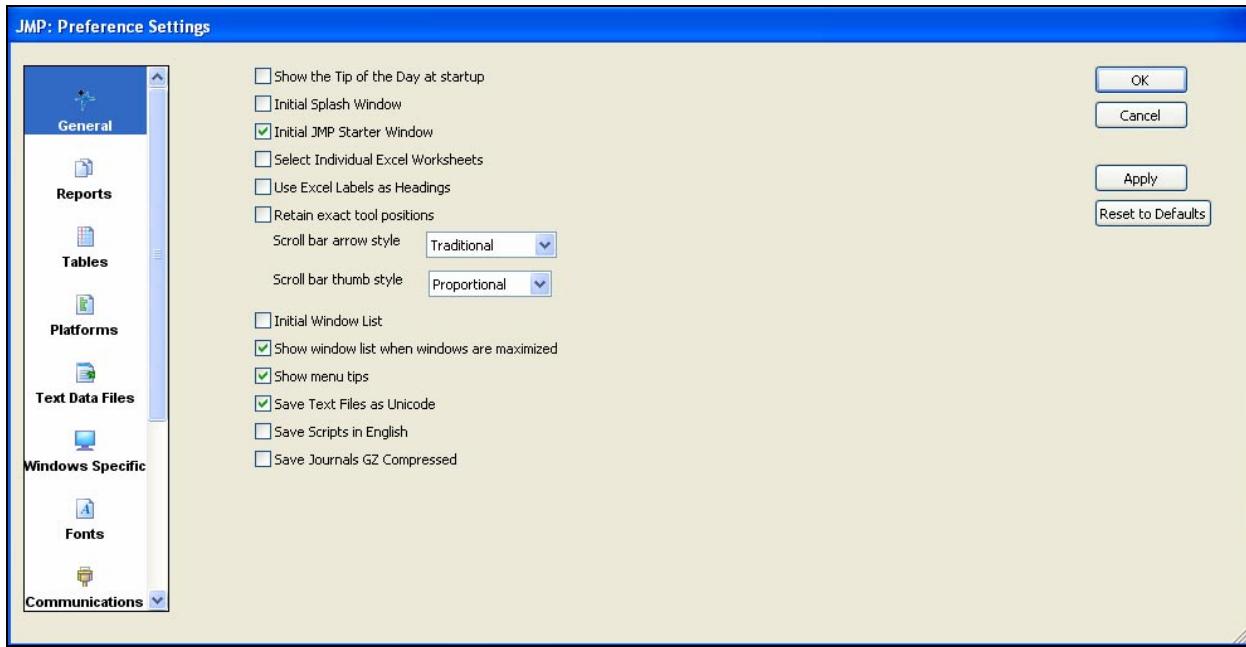
Notice that an icon, a button to activate the platform, and a description of the platform are presented.

Preferences

1. Click the **File** category and select **Preferences**.



The Preference Settings window also uses categories to organize the numerous choices that enable you to tailor components of your JMP environment to suit your specific needs or particular situation.



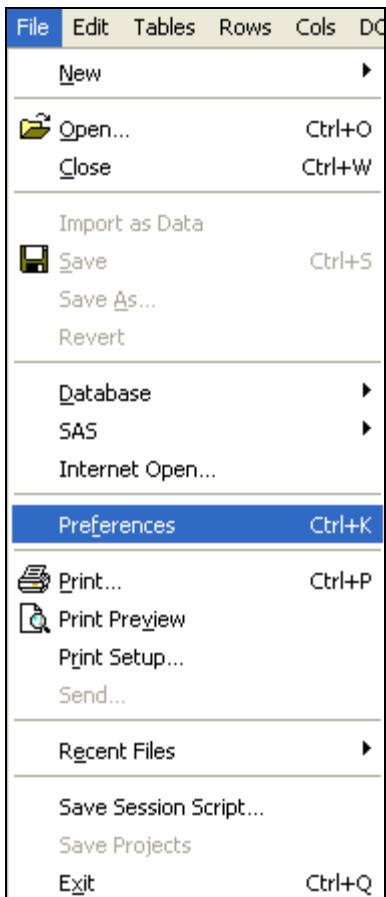
These numerous preferences include the following:

- general operation and appearance of JMP
- background color of windows and graphs
- type, style, and size of fonts
- copy/drag graphic formats results and graphic formats for RTF and HTML files
- communications settings
- default directory paths for file locations
- results initially presented by each analysis or graph platform
- settings for importing and exporting data

2. Select **OK** in the Preference Settings window to close it.



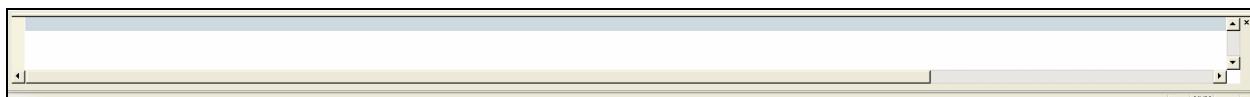
Another method to access your user preferences is to select **File** \Rightarrow **Preferences**.



The JMP Log

Another key component of JMP is the Log, which notifies you of important information such as system or error messages.

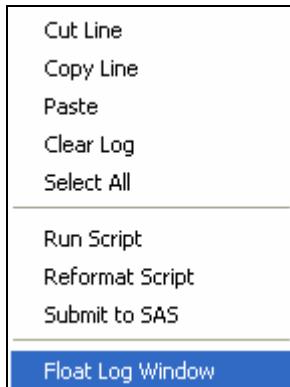
1. To open the JMP Log from the menu bar, select **View** \Rightarrow **Log**. By default, the Log is located directly above the Status Bar at the bottom of the JMP window.



You can resize the pane in which the Log appears by dragging the upper edge of the Log window up or down.

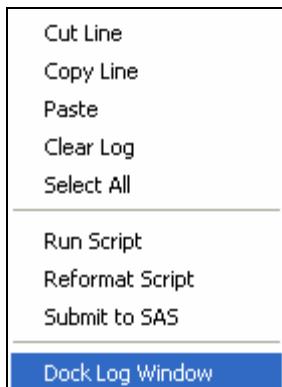
As an alternative to resizing the Log and leaving it at the bottom of the JMP window, you can also detach the Log from its default position and place it anywhere you want within the JMP window.

2. Right-click in the Log window.
3. Select **Float Log Window** from the menu that opens.

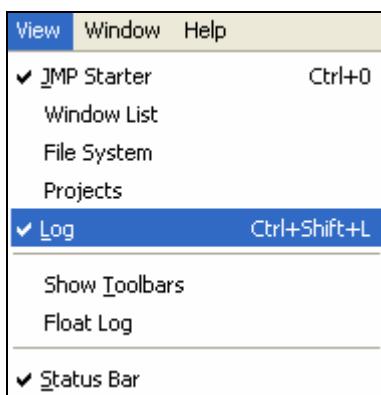


You can change the window back to a pane in the default position at the bottom of the JMP window.

4. Right-click in the Log window.
5. Select **Dock Log Window** from the menu that opens.



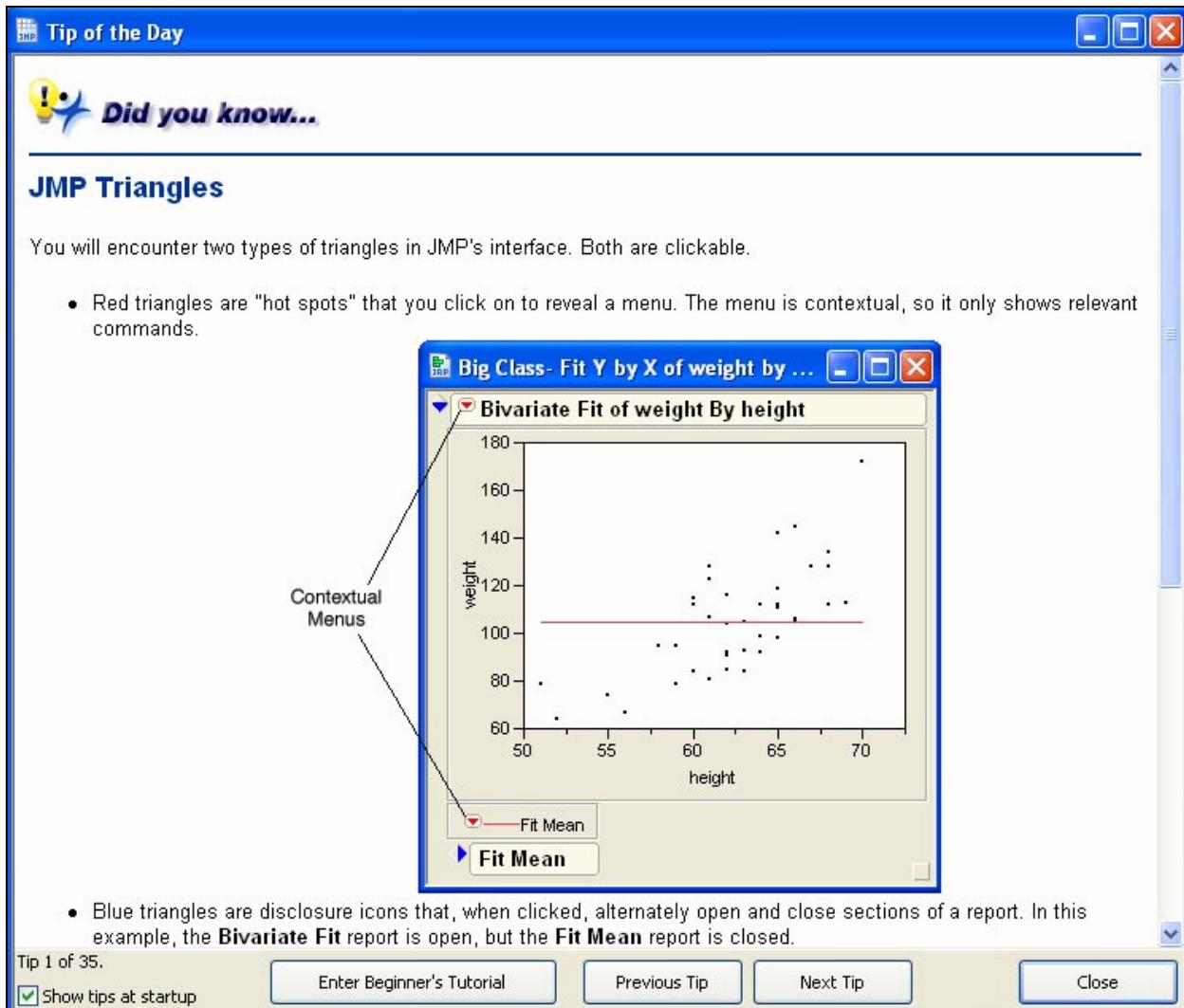
6. You can close the Log if you do not want it taking up space in your JMP window. Select **View** \Rightarrow **Log**, as shown below, to close the Log window.



User-Assistance Options

This demonstration explores user-assistance options available in JMP®7.

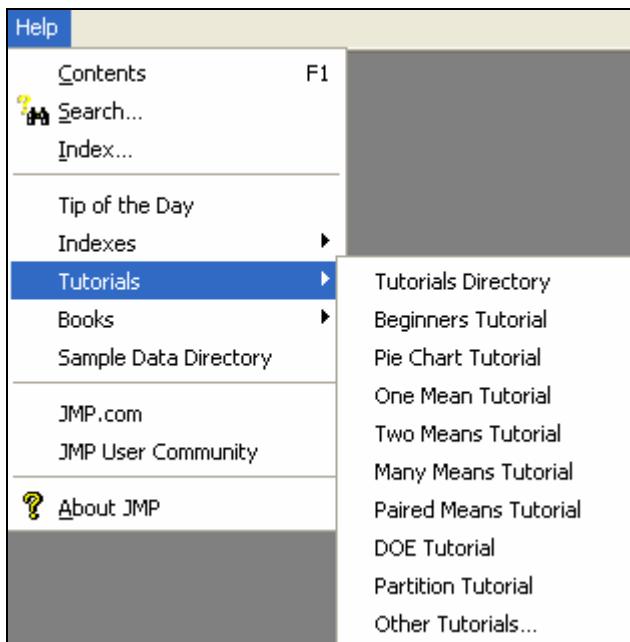
As you might recall, when you launched JMP, a window called Tip of the Day opened. To open the Tip of the Day window again, select **Help** ⇒ **Tip of the Day**.



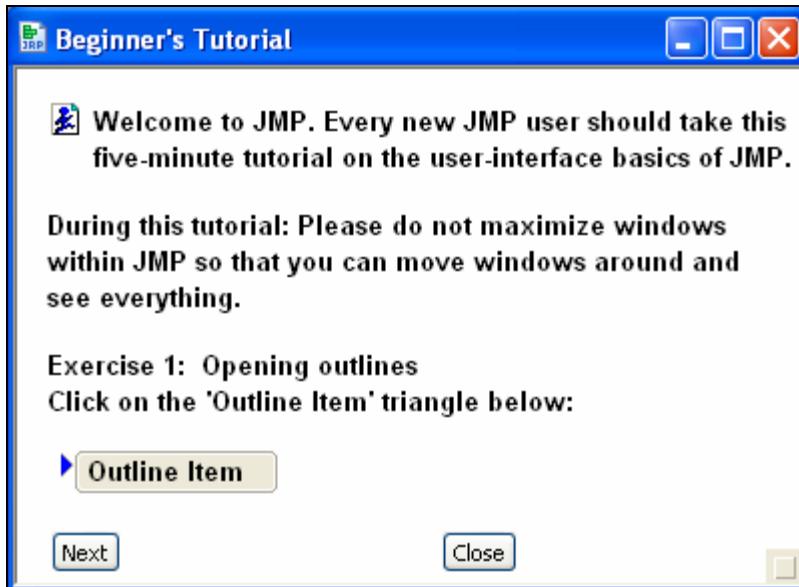
The software includes 35 tips. You can view the tips one at a time when you open JMP by consecutively using the Previous Tip and Next Tip buttons at the bottom of the Tip of the Day window. You can also access all 35 tips through the Help menu, as was demonstrated.

In addition to the Tip of the Day, several tutorials are provided in JMP®7.

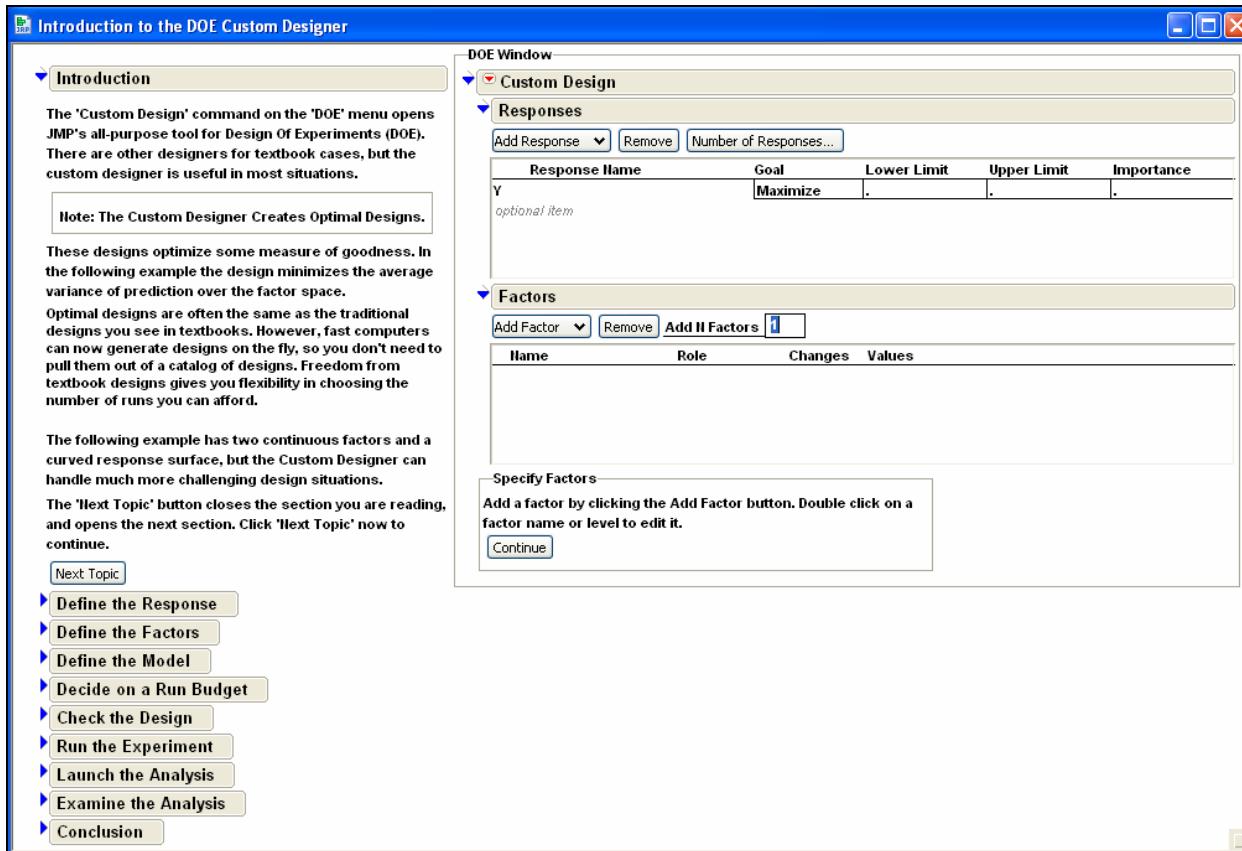
1. Select **Help** ⇒ **Tutorials**.



New users can choose the Beginners Tutorial to quickly learn how to navigate the JMP interface.

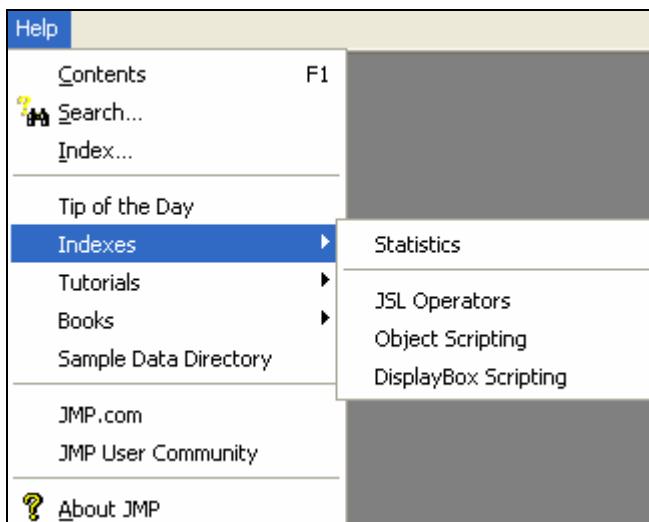


More advanced users can choose the DOE Tutorial for a lesson on the custom design platform available in JMP.



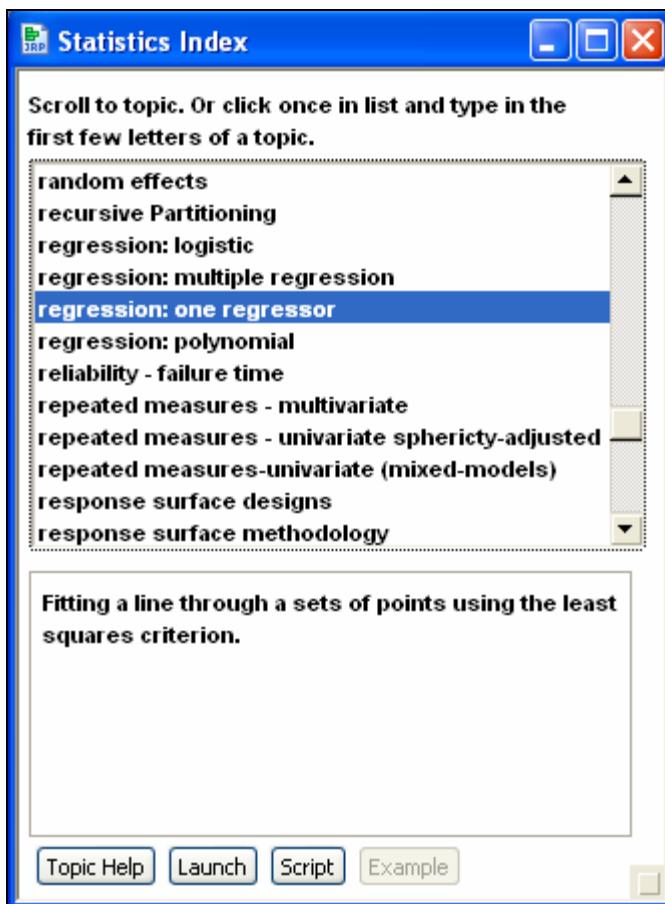
The Indexes option on the Help menu is used for browsing specific analytical procedures or the JMP Scripting Language (JSL). It also provides an entry point for JMP Help.

2. Select Help ⇒ Indexes.



3. Select Statistics from the pop-up menu to see a list of available topics.

4. Scroll down the list of topics and select **regression: one regressor**.



5. Select **Topic Help**.

The screenshot shows the JMP Statistics and Graphics Guide page for the Bivariate Scatterplot and Fitting platform. The top navigation bar includes the JMP logo, "Statistics and Graphics Guide", "Bivariate Scatterplot and Fitting", and "Prev | Next" buttons. The main content area is titled "The Fit Y by X Platform" and describes the Bivariate platform's functionality. It lists several types of fits available: simple linear regression, polynomial regression, smoothing spline, bivariate normal density ellipses, bivariate nonparametric density contours, and multiple fits over groups. A note at the bottom explains how to launch the platform from the main menu or the JMP Starter.

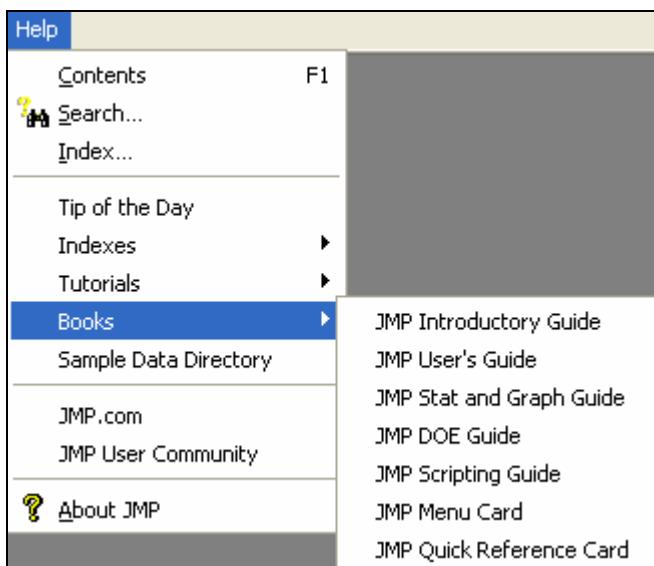
Note: You can launch the Bivariate platform from the **Fit Y by X** main menu command (or tool bar) or with the **Bivariate** button on the JMP Starter. When you choose **Fit Y by X** from the **Analyze** menu, JMP performs an analysis by context, which is the analysis appropriate for the modeling types you specify for the response and factor variables. When you use the **Bivariate** button on the JMP Starter, the launch dialog expects both factor and response variables to be continuous and advises you if you specify other modeling types.

This action opens the JMP Help window, based on the specific topic you requested.

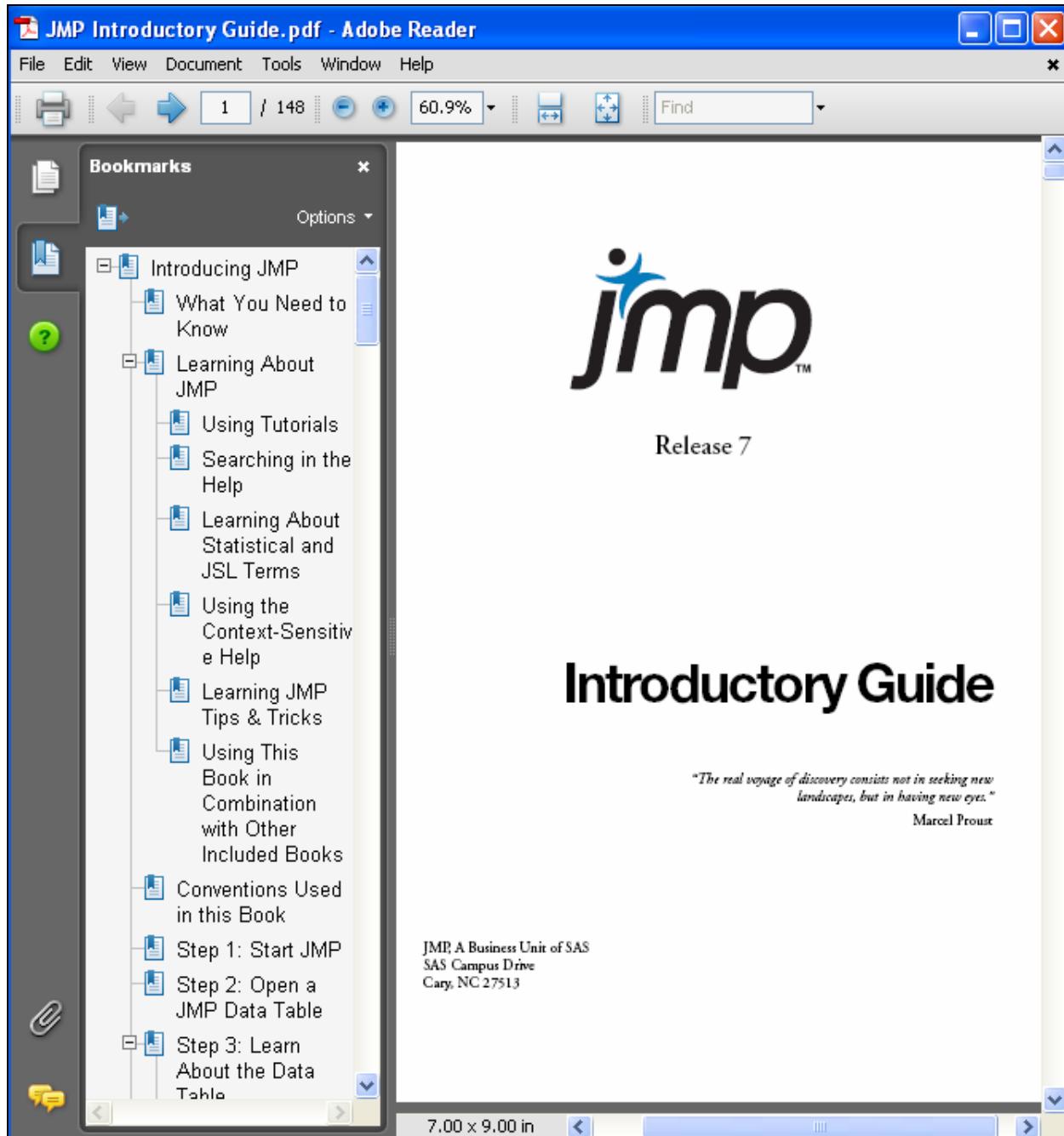
6. Close the Help window by clicking on the **X** in the upper-right corner of the window.

Electronic copies of the JMP software documentation can also be opened from the Help menu.

7. Select **Help** \Rightarrow **Books**.



8. Select **JMP Introductory Guide** from the pop-up menu.



A PDF of the book is opened in Adobe Reader.

9. Close all windows.

1.2 Accessing Data

Objectives

- Identify the types of files that JMP can open.
- Define ODBC drivers.
- Import a Microsoft Excel file into JMP.

9

JMP can access and read a number of different data files directly.

In this section, you learn to import Microsoft Excel files. You also learn what ODBC is and its purpose in JMP.

ODBC Drivers

- ODBC stands for Open DataBase Connectivity.
- An ODBC driver enables you to access files in that particular database from JMP.
- If you do not know which ODBC drivers you have, ask your software support personnel.

10

ODBC drivers come with the database software product (for example, Microsoft Access, and dBase). The ODBC driver for Microsoft Access is purchased with Microsoft Access through Microsoft. Each database software product has its own ODBC driver.

When JMP is first opened, it searches behind the scenes for any ODBC drivers you have. When you attempt to open a file, the database software product associated with each detected ODBC driver is displayed as a type of file that can be opened in JMP.

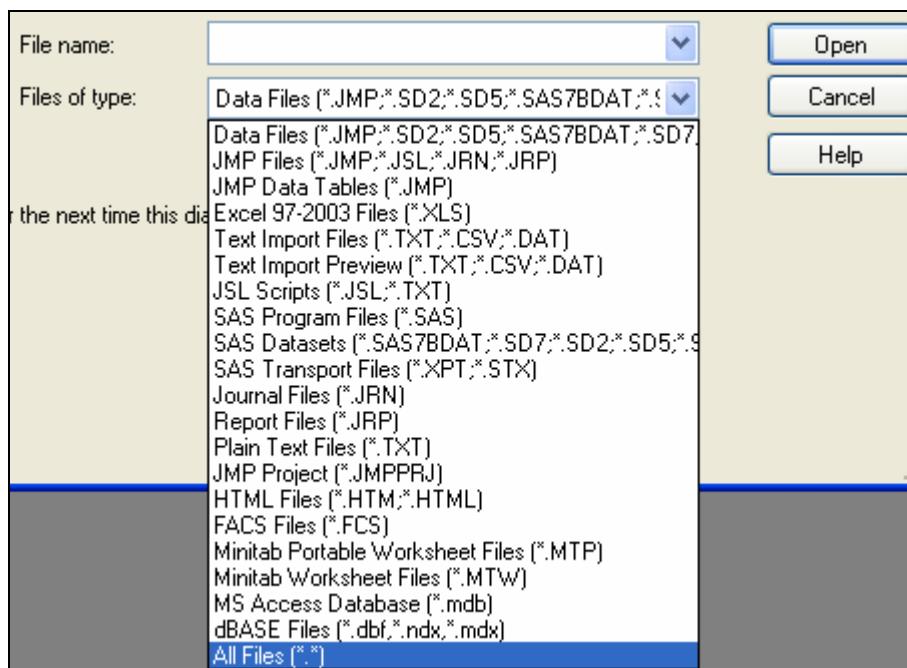
 The Database option is used to access tables stored in ODBC-compliant databases on your PC.



Importing Data with JMP

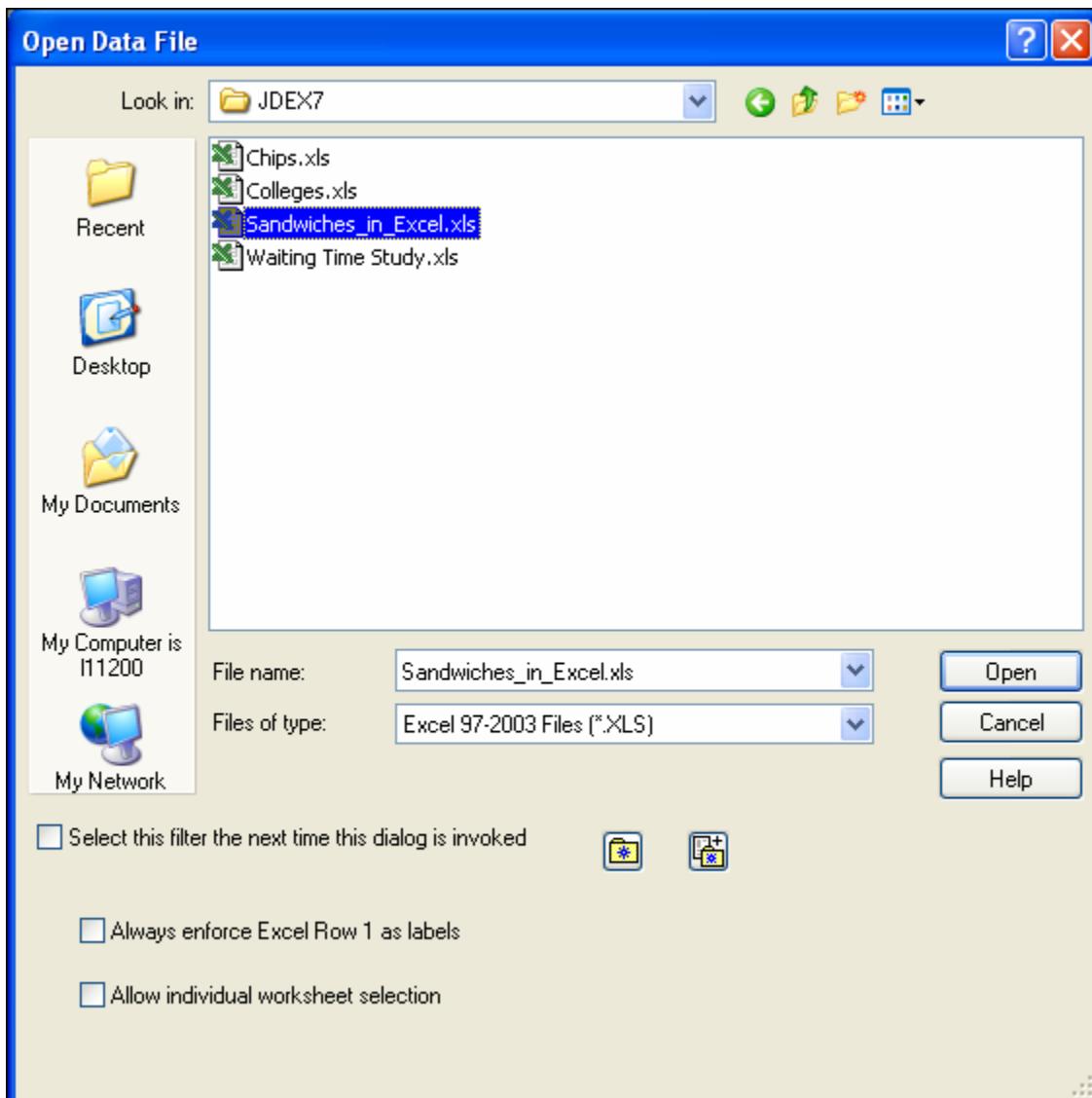
This demonstration illustrates how to import a Microsoft Excel file both directly and using ODBC.

1. Select **File** ⇒ **Open** or select **Open Data Table** from the File category in the JMP Starter window.
An Open Data File dialog box opens initially with a list of files in the JMP directory.
2. Navigate to the **JDEX7** data folder.
3. Notice the **Files of type** field located below the list of files. Your choice in this field applies a filter to the contents of the current directory so that you are only presented with certain kinds of files with each Open command.
4. Click the black triangle at the end of the field to see a menu of all of the types of files that JMP can open directly. A number of file types can be read. The following menu shows the end of a long list of possibilities. All Files (*.*) has been chosen to reveal all types of files in the folder.



Opening an Excel File in JMP

1. In the Open Data File dialog box, navigate to the **JDEX7** data folder.
2. Scroll down the list of Files of type and select **Excel Files 97-2003 (*.XLS)**.
3. Select **Sandwiches in Excel.xls**.



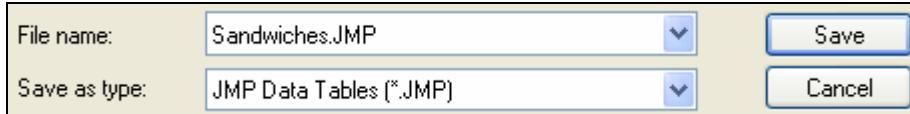
The two check boxes in the lower-left corner can be used to specify details about the Excel file that you are importing. The first can be used to enforce that the spreadsheet has labels in row 1. If there are multiple worksheets in the file, the second check box enables users to choose which worksheets to open. Each worksheet opens as a separate JMP data table.

4. Select **Always enforce Excel Row 1 as labels**.
5. Select **Open**.

JMP imports and displays the Excel file **Sandwiches_in_Excel**.

	Brand	Name	Category	Calories	TFat	Protein
1	C	Big Fish	Fish	565	33	
2	J	Spinach & Cheese Pocket	Frozen	223	5	
3	C	Turkey Club	Turkey	518	23	
4	G	Tuna on Wheat	Tuna	378	12	
5	A	Baby Beef	Beef	339	16	
6	K	Ham & Cheese	Frozen	339	16	
7	E	Grilled Chic	Chicken	400	18	
8	C	BBQ Chic	Chicken	286	5	
9	L	Lite Reuben	Frozen	254	5	
10	L	Lite Ham & Cheese	Frozen	260	9	
11	D	Grilled Chic	Chicken	297	8	
12	I	Fried Platter	Fish	855	46	
13	A	Deluxe Chic	Chicken	440	18	
14	K	Pepperoni Rollup	Frozen	354	16	

- Save this file as a JMP file with a more meaningful name. Select File \Rightarrow Save As....
- Type **Sandwiches.JMP** as the file name, and then select Save.



Notice that, by default, JMP redisplays the saved file with the new name.

	Brand	Name	Category	Calories	TFat	Protein	Carb	Fiber	Sodium	Weight
1	C	Big Fish	Fish	565	33	23	45	5	1006	9
2	J	Spinach & Cheese Pocket	Frozen	223	5	13	34	2	794	6
3	C	Turkey Club	Turkey	518	23	30	48	•	1494	10.7
4	G	Tuna on Wheat	Tuna	378	12	25	44	3	1024	9.7
5	A	Baby Beef	Beef	339	16	13	33	0	573	6
6	K	Ham & Cheese	Frozen	339	16	15	33	4	607	6
7	E	Grilled Chic	Chicken	400	18	14	39	0	975	9
8	C	BBQ Chic	Chicken	286	5	25	39	3	1118	8.2
9	L	Lite Reuben	Frozen	254	5	18	39	5	569	6.5
10	L	Lite Ham & Cheese	Frozen	260	9	5	40	3	946	5.5
11	D	Grilled Chic	Chicken	297	8	19	35	2	1163	8
12	I	Fried Platter	Fish	855	46	45	61	•	2043	14.7
13	A	Deluxe Chic	Chicken	440	18	21	45	3	1411	9.5
14	K	Pepperoni Rollup	Frozen	354	16	22	32	2	435	6
15	A	Deluxe Turkey	Turkey	511	24	19	47	•	1046	9.7
16	C	Deluxe Beef	Beef	534	24	33	44	•	1564	10.7
17	E	Spicy Fried Chic	Chicken	568	32	23	44	0	1185	9.5
18	F	Bacon Chic	Chicken	483	22	22	47	4	1346	6.5

- Close this file by clicking on the X in the upper-right corner of the Sandwiches window.

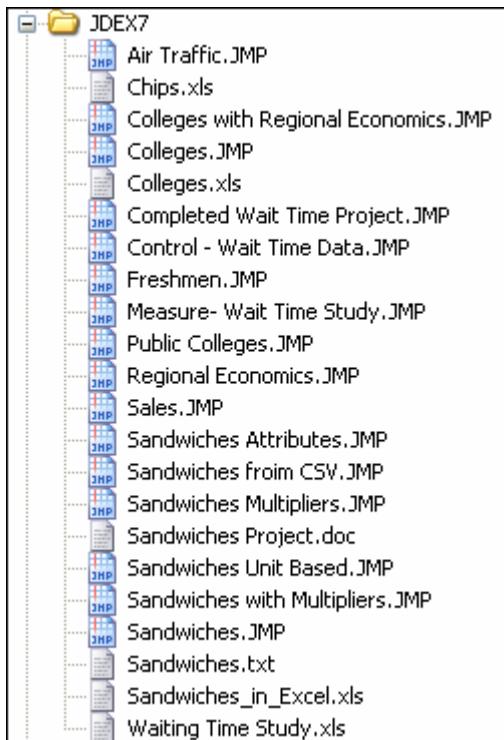
Using the File System

The File System command shows the files and directories of your local PC. The files and directories in the File System are shown in the form of a tree structure. You can use this toolbar to navigate into any directory on your computer to locate a file.

1. Select **View** ⇒ **File System**.

 The default location for the File System is on the left side of the JMP session. It can be undocked by clicking its title area and dragging it to its preferred location.

2. Navigate to **JDEX7**. (Your instructor will help with these steps because this location will vary.)



The files associated with the class can be viewed in the File System. Furthermore, after a file has been located, it can be opened in JMP by double-clicking the file name.

3. Double-click **Sandwiches in Excel.xls**.

The Excel file is opened as a JMP file.

4. Close this window without saving.

In addition, a file can be opened by right-clicking on the file name. The options in the menu will vary depending on the type of file that is selected. If the file selected is a JMP file, then the option for opening the file is naturally to open it in JMP. If the file is not a JMP file, such as an Excel file, then an option is available to open the file outside of JMP. In such a case, JMP will ask the application that “owns” that file extension to open the file. As long as the selected file is one that JMP is able to import, an option will also be available to open that file in JMP.

5. Right-click **Sandwiches in Excel.xls** and select **Open outside JMP**.

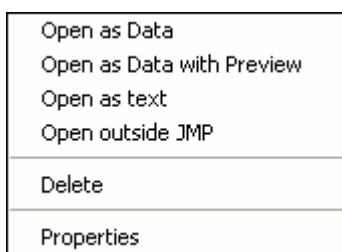


The file is opened in Microsoft Excel.

6. Close Microsoft Excel.

There are four methods for opening a text file.

7. Right-click **Sandwiches.txt** to view the options for opening this text file.



If you select the Open as Data option, JMP will attempt to import the file into a JMP data table by applying the default text import preferences. The Open as Data with Preview option treats the text in the file as data and the first few lines of the file are presented in the Text Import Preview window. The Open as text option loads the text into a script editor window. The Open outside JMP option will open the text file using your default text editor.

- These options are discussed further in the Optional portion of this section.

Finally, the File System List, by default, presents the file system on the hard drive. If you are connected to other file servers, you can add these servers to the tree.

8. Click the red triangle at the upper-right corner of the File System window and select **Add Folder**.

In the Browse for Folder window, navigate to any server on your network, if applicable. After a server is selected, select **OK**. This adds a new folder for your server at the top of the File System List. JMP will remember that the folder was added, so it will display this folder each time JMP is launched.

- The folder can be removed from the File System in JMP by right-clicking on the folder name and selecting **Delete**.

9. Close the Browse for Folder window.

Close the File System list.

10. Select **View** ⇒ **File System**.

Using the Database Command

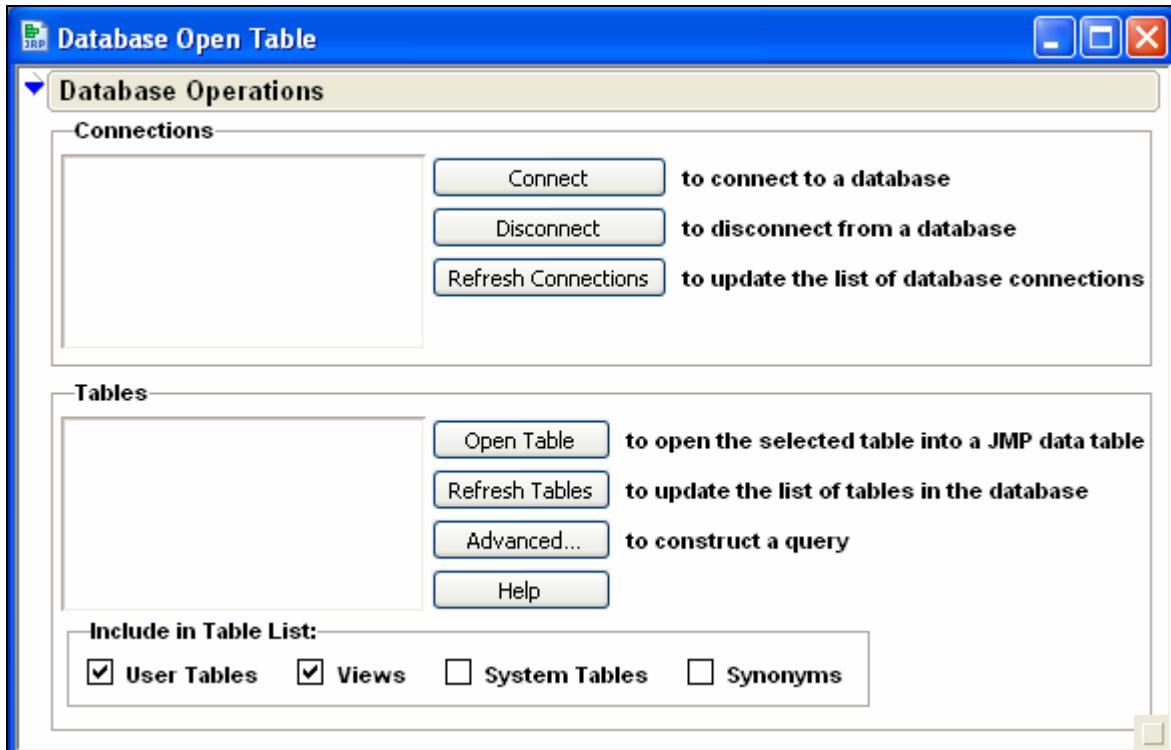
There is data stored in an Excel file named **Chips.xls**. The first ten rows are shown below.

	A	B	C	D
1	ID	Wafer	Position	Line Width
2	1	1	Top	2.582194025
3	2	1	Bottom	1.042097041
4	3	1	Left	1.487603889
5	4	1	Right	3.259965631
6	5	2	Top	1.686510124
7	6	2	Bottom	1.984672782
8	7	2	Left	1.2378369
9	8	2	Right	2.112185121
10	9	3	Top	2.561254726

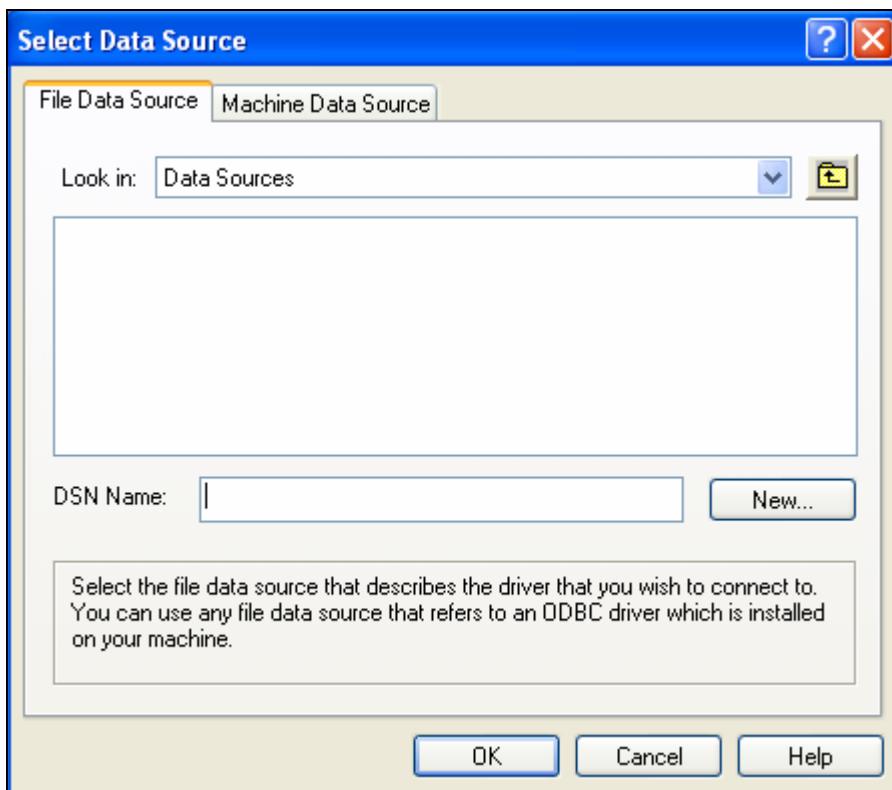
You decide that you want to bring only three of the original columns into JMP: **Wafer**, **Position**, and **Line Width**. Also, you only want rows where the value of **Position** is **Top**.

If you bring in a subset of an Excel file, you need to use the Database option, and the Excel ODBC driver must be installed.

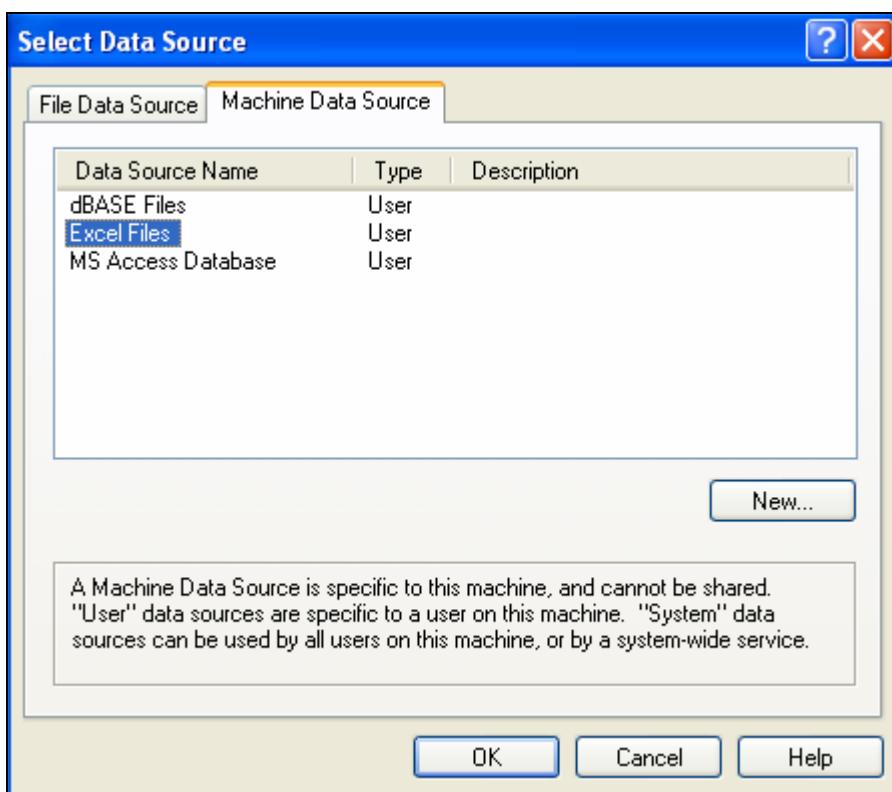
1. Select File \Rightarrow Database \Rightarrow Open Table.... The Database Open Table window opens.



2. Select Connect. The Select Data Source window opens.

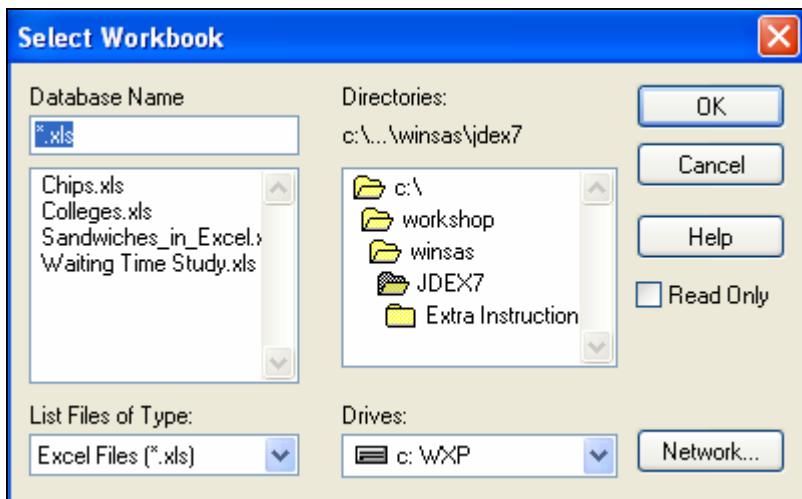


3. Select the Machine Data Source tab.
4. Under Data Source Name, select Excel Files.



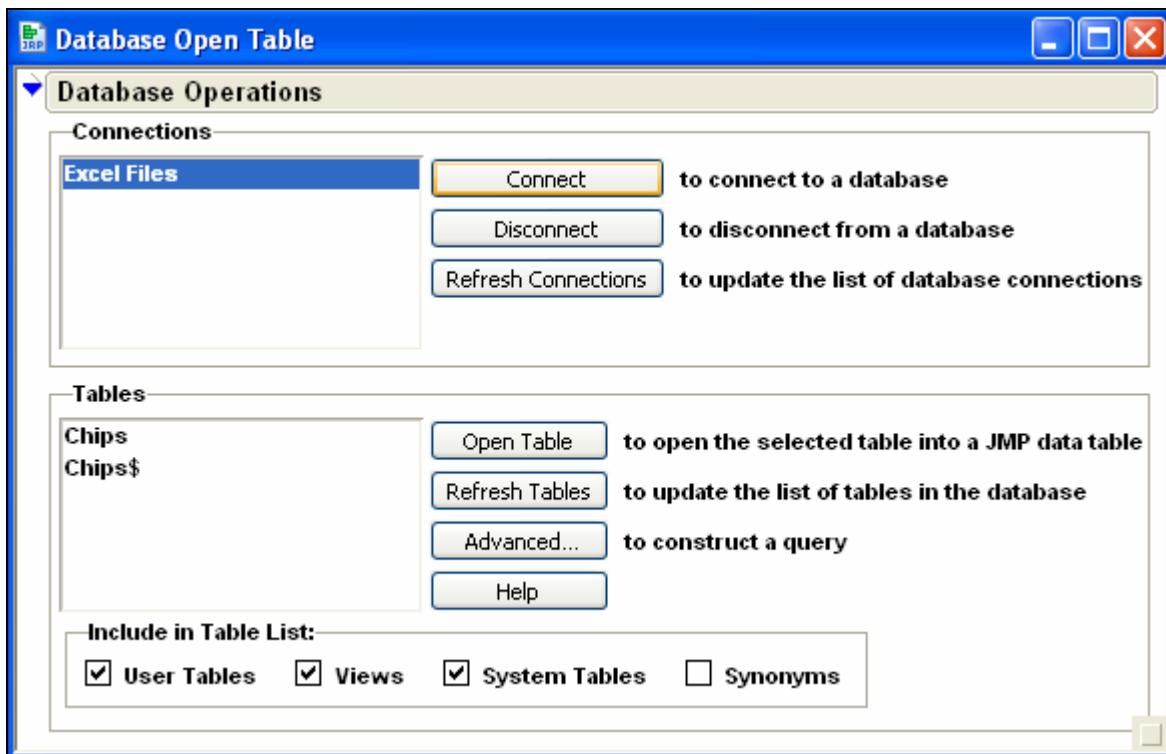
5. Select **OK**.

The Select Workbook window opens.



6. Select **Chips.xls** \Rightarrow **OK**.

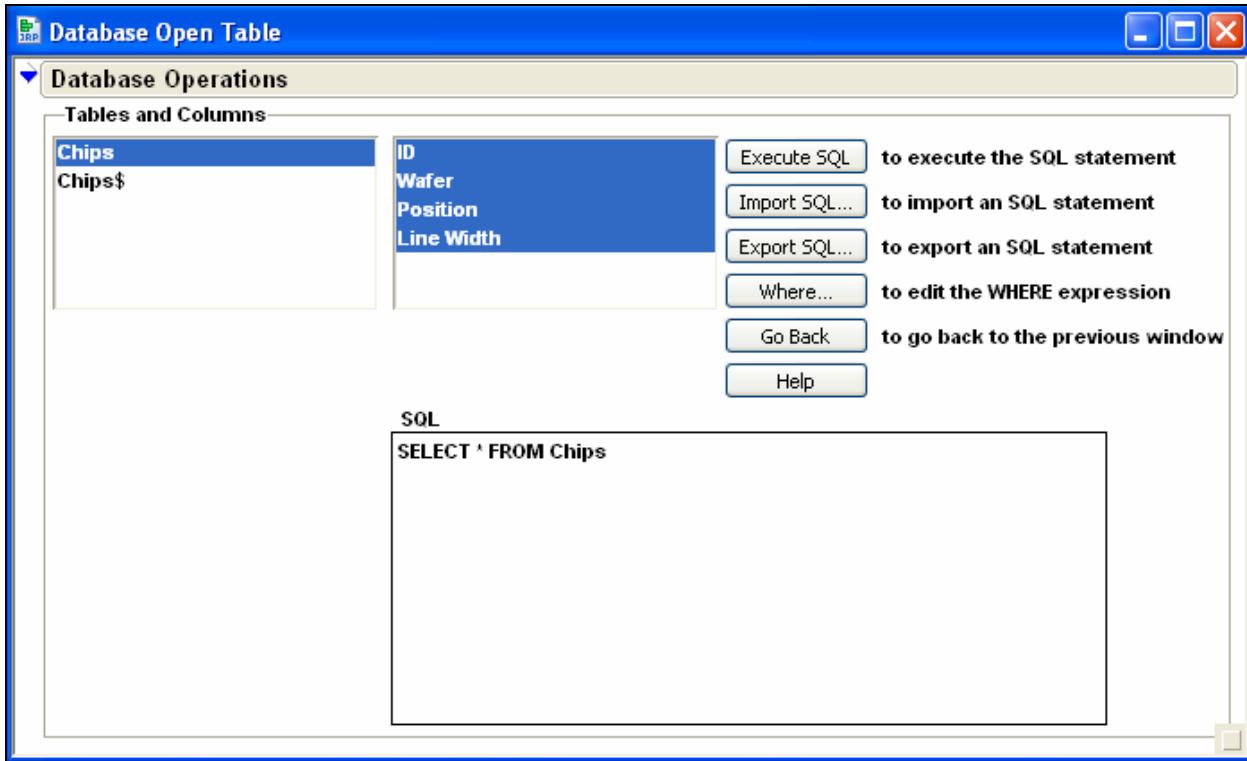
The Database Open Table window now shows the connection to the Excel database and the selected table.



- If the Excel workbook contains more than one spreadsheet, all of the worksheets are listed under Tables. The Excel ODBC driver will display the name of each worksheet twice, with the second listing containing a \$ at the end of the name. You can select either listing for the **Chips** worksheet to open.

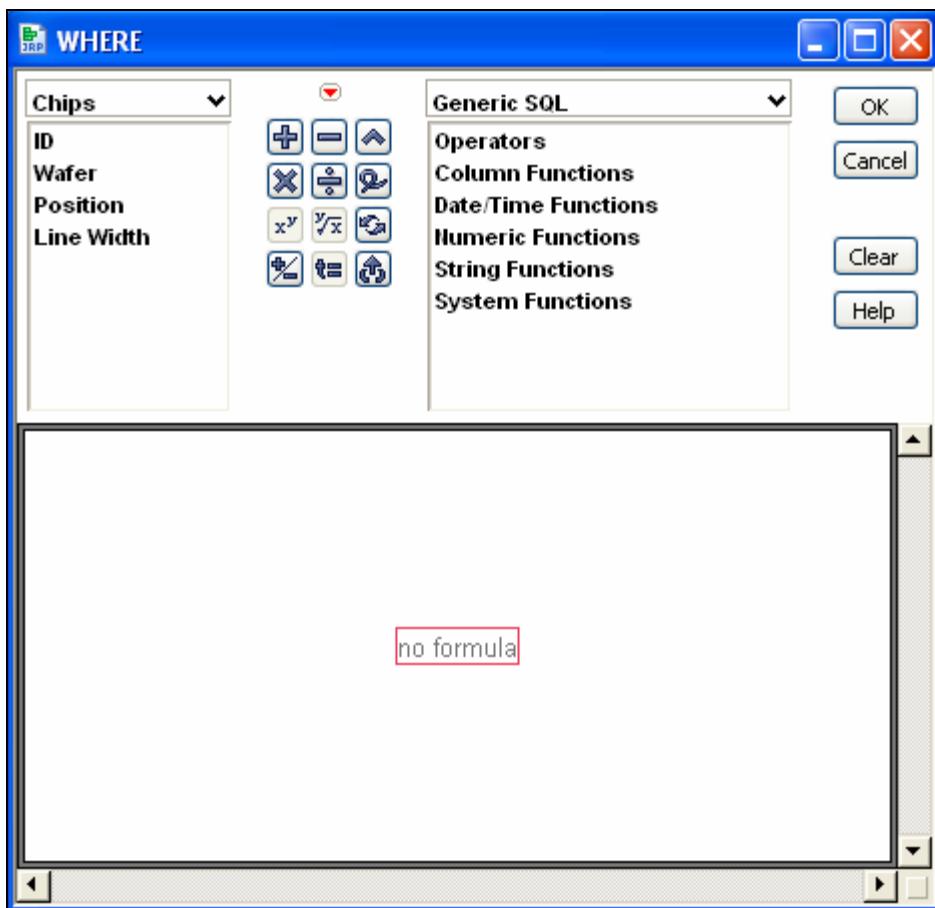
7. Select **Chips** ⇒ **Advanced....**

The SQL (Structured Query Language) area of the Database Open Table window is used to enter valid SQL statements. You can also import previously existing SQL statements with the Import SQL... button.



8. Hold down the CTRL key and select **ID**. The resulting SQL statement is shown below.

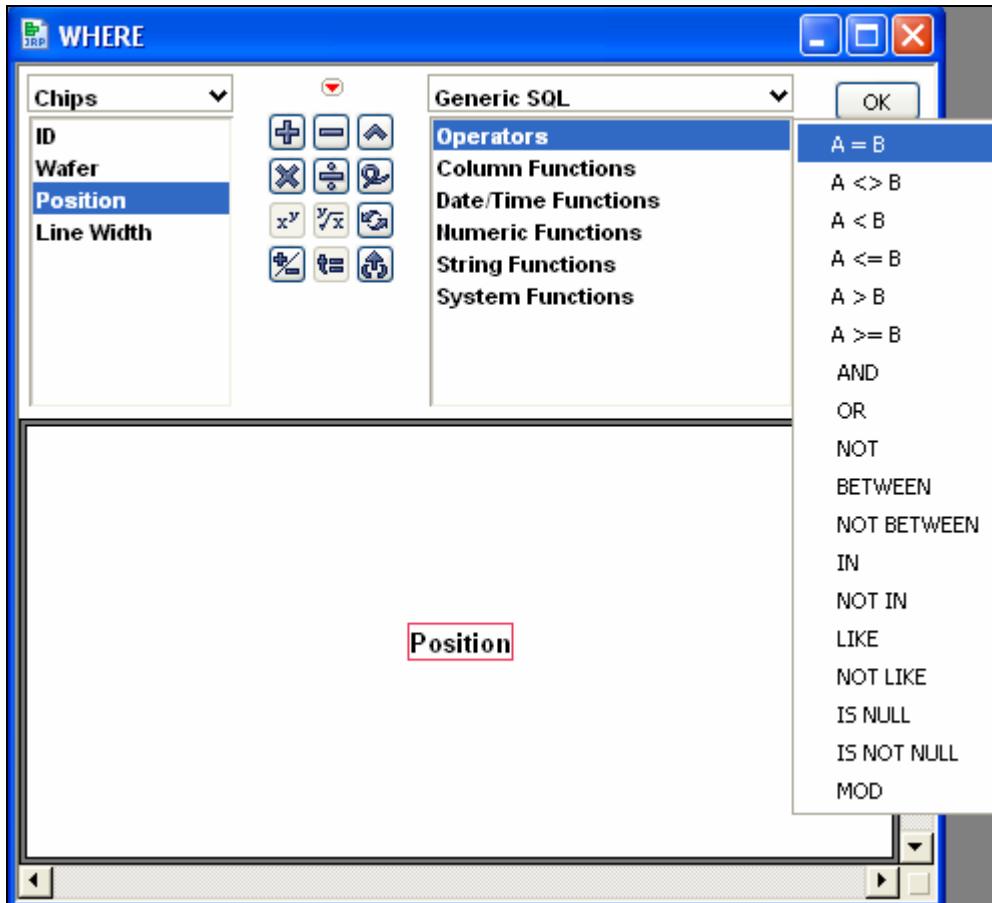
```
SQL
SELECT Wafer, Position, "Line Width" FROM Chips
```

9. Select **Where....**

A window opens that enables the user to build a Where clause interactively.

10. Select Position.

11. Select Operators \Rightarrow A = B.



12. Click the red box on the right.



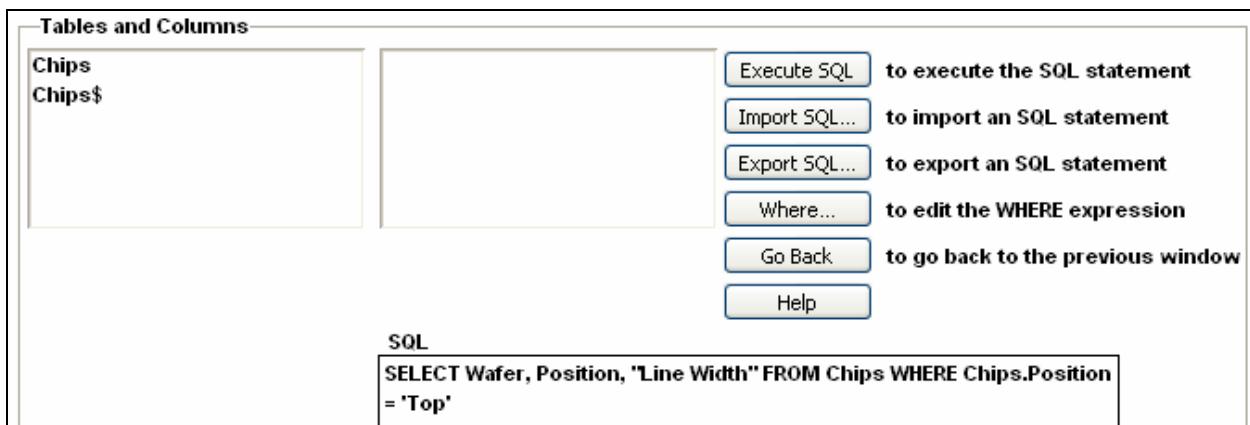
13. Type '**Top**' and press ENTER.

SQL requires single quotation marks around character values.



14. Select OK.

The completed query is in the SQL box in the Database Open Table window.



- The SQL statement constructed in JMP can be exported as a text file for use in subsequent queries.

15. Select Execute SQL.

CHIPS			
SQL: SELECT Wafer, Position JSL: Open Database('DSII: Columns (3/0)			
Wafer Position Line Width			
1	1	Top	2.58219402
2	2	Top	1.68651012
3	3	Top	2.56125473
4	4	Top	2.89926129
5	5	Top	1.8840566
6	6	Top	2.54233663
7	7	Top	2.00524101
8	8	Top	1.70418942
9	9	Top	2.44489778
10	10	Top	2.22526834

16. Select File \Rightarrow Save As...

17. Type **Chips.JMP** as the file name.

18. Select Save.

Now that you have saved **Chips .JMP**, disconnect from the Excel database.

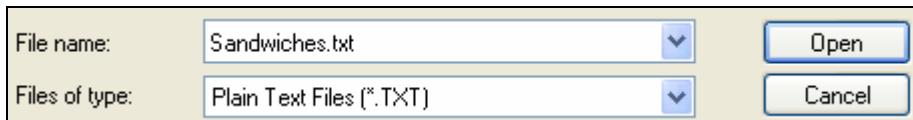
1. Return to the Database Open Table window by selecting **Window** \Rightarrow **Database Open Table**.
2. Select **Go Back**.
3. Select **Disconnect**.
4. Select **Yes** in the JMP - Caution Alert window.
5. Close the Database Open Table window.
6. Close **Chips .JMP**.

Text Files (Optional)

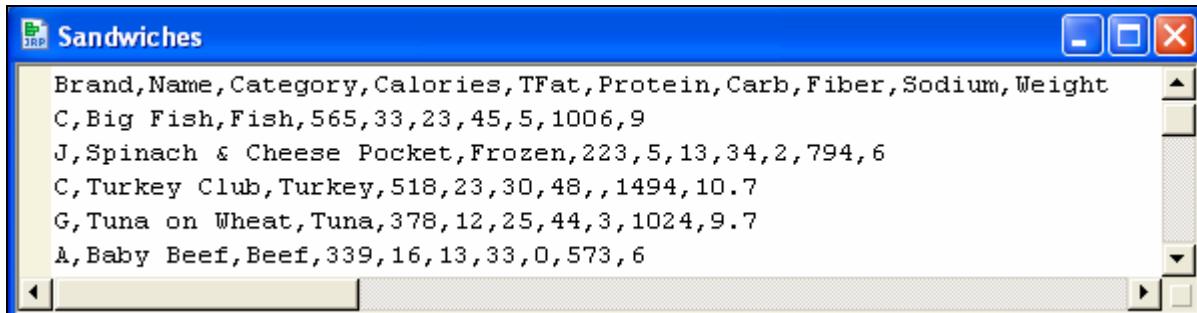
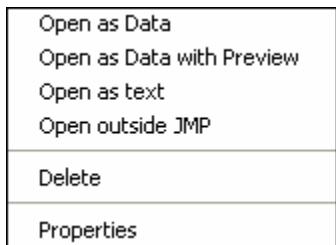
1. Select **File** \Rightarrow **Open**

In the **Files of type** field, one of the types of files is **Plain Text Files (*.TXT)**.

2. Scroll up or down in the list provided to **Plain Text Files (*.TXT)**. Use this format to investigate the contents of **.txt** files.
3. Select **Sandwiches.txt** to determine its contents, and then select **Open**.

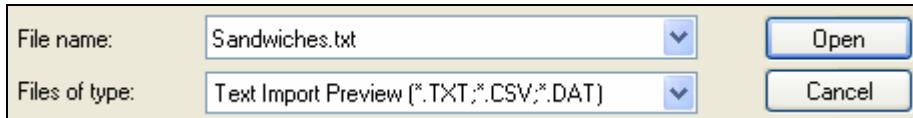


Alternately, right-click **Sandwiches.txt** in the File System list and select **Open as text**.



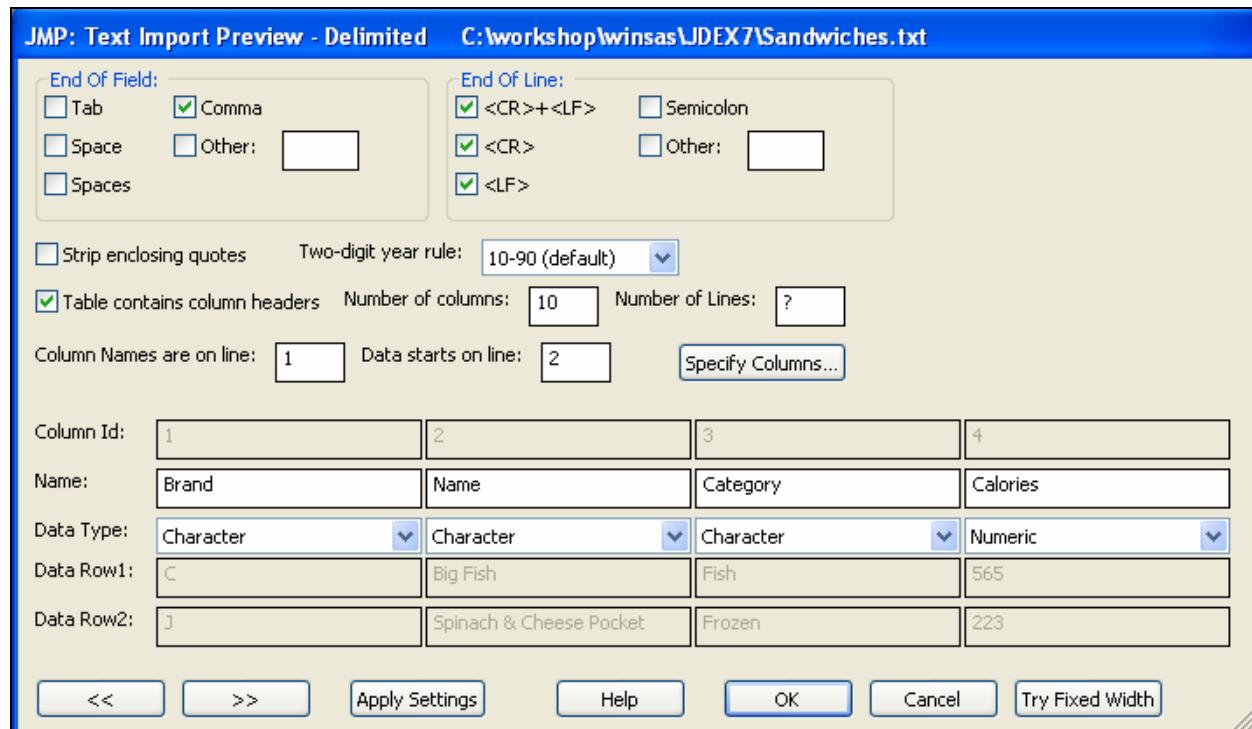
Sandwiches.txt uses a comma as a delimiter to separate values. Use this information to import this file into JMP.

4. Close the Sandwiches window and reopen the Open Data File dialog box.
5. Select **Text Import Preview (*.TXT, *.CSV, *.DAT)** as the type of file and select **Sandwiches.txt**. Select **Open**.



Alternately, you can right-click **Sandwiches.txt** in the File System list and select **Open as Data with Preview**.

Because only commas have been used to separate values, ensure that only Comma is selected under the End Of Field. If any other character has been selected, deselect it by clicking on the appropriate check mark. Select **Apply Settings**.



The default two-digit year rule appends 20 to the front of two-digit years indicated as 00-10. A 19 is appended otherwise.

Your initial review of **Sandwiches.txt** confirms that the Column names are located on line 1 and the data starts on line 2. JMP has displayed the name, the data type, and the first two rows of the data.

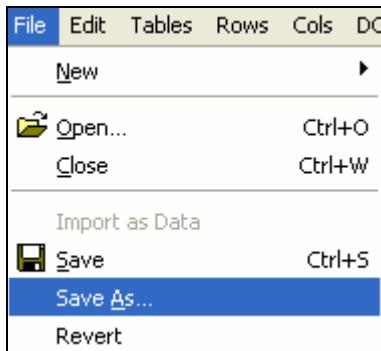
6. This dialog box shows that JMP can import this data in its current form. Select **OK**.

JMP displays the **Sandwiches.txt** file imported with the default name **Sandwiches.JMP**.

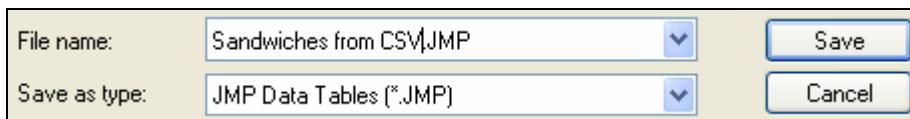
Sandwiches											
Sandwiches		Brand	Name	Category	Calories	TFat	Protein	Carb	Fiber	Sodium	Weight
1	C	Big Fish	Fish	565	33	23	45	5	1006	9	
2	J	Spinach & Cheese Pocket	Frozen	223	5	13	34	2	794	6	
3	C	Turkey Club	Turkey	518	23	30	48	•	1494	10.7	
4	G	Tuna on Wheat	Tuna	378	12	25	44	3	1024	9.7	
5	A	Baby Beef	Beef	339	16	13	33	0	573	6	
6	K	Ham & Cheese	Frozen	339	16	15	33	4	607	6	
7	E	Grilled Chic	Chicken	400	18	14	39	0	975	9	
8	C	BBO Chic	Chicken	286	5	25	39	3	1118	8.2	
9	L	Lite Reuben	Frozen	254	5	18	39	5	569	6.5	
10	L	Lite Ham & Cheese	Frozen	260	9	5	40	3	946	5.5	
11	D	Grilled Chic	Chicken	297	8	19	35	2	1163	8	
12	I	Fried Platter	Fish	855	46	45	61	•	2043	14.7	
13	A	Deluxe Chic	Chicken	440	18	21	45	3	1411	9.5	

A new window opens and presents both the data and its metadata (on the left side of the window). Metadata describes the characteristics of the data, so metadata can be described as "data about the data."

- For now, save this file with a more meaningful name. Select **File** \Rightarrow **Save As...**



- Type **Sandwiches from CSV.JMP** as the file name, and then select **Save**.



Notice that by default, JMP redisplays the saved file with the new name.

The image shows the 'Sandwiches from CSV' data table in JMP. The table has 10 columns: Brand, Name, Category, Calories, TFat, Protein, Carb, Fiber, Sodium, and Weight. The data includes various sandwich types like Big Fish, Spinach & Cheese Pocket, Turkey Club, etc., with their respective nutritional values. On the left, a tree view shows 'Sandwiches from CSV' expanded, with 'Source' and 'Columns (10/0)' listed. A legend on the far left identifies the column icons.

- Close this file by clicking on the X in the upper-right corner of the Sandwiches from CSV window.

1.3 The Data Table

Objectives

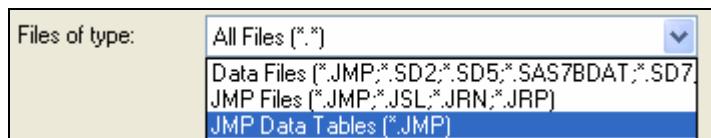
- Describe the four components of the JMP data table: the data grid, the Columns panel, the Rows panel, and the Table panel.
- Learn the basics of the JMP table using its four components.



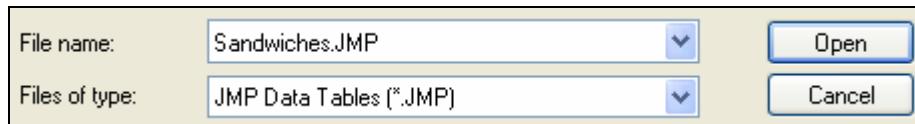
The JMP Data Table

This demonstration opens an existing JMP data table, **Sandwiches.JMP**, and examines the four components of the data table: the data grid, the Columns panel, the Rows panel, and the Table panel.

1. Select **File** \Rightarrow **Open**.
2. In the Open Data File dialog box, navigate to the **JDEX7** data folder.
3. Scroll down the list of Files of type. Select **JMP Data Tables (*.JMP)**.



4. Select **Sandwiches.JMP**.

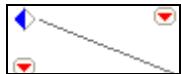


5. Select **Open**.

This entire window is known as the *data table*. On the right side is the *data grid*. You use the data grid to directly enter, view, select, and change your data. On the left side are three vertically stacked *panels*: one for the table, one for the columns, and one for the rows. With these panels, you enter, view, change, or use the metadata.

Sandwiches											
Sandwiches		Brand	Name	Category	Calories	TFat	Protein	Carb	Fiber	Sodium	Weight
Notes C:\workshop\winsas		1 C	Big Fish	Fish	565	33	23	45	5	1006	9
		2 J	Spinach & Cheese Pocket	Frozen	223	5	13	34	2	794	6
		3 C	Turkey Club	Turkey	518	23	30	48	•	1494	10.7
		4 G	Tuna on Wheat	Tuna	378	12	25	44	3	1024	9.7
		5 A	Baby Beef	Beef	339	16	13	33	0	573	6
		6 K	Ham & Cheese	Frozen	339	16	15	33	4	607	6
		7 E	Grilled Chic	Chicken	400	18	14	39	0	975	9
		8 C	BBQ Chic	Chicken	286	5	25	39	3	1118	8.2
		9 L	Lite Reuben	Frozen	254	5	18	39	5	569	6.5
		10 L	Lite Ham & Cheese	Frozen	260	9	5	40	3	946	5.5
		11 D	Grilled Chic	Chicken	297	8	19	35	2	1163	8
		12 I	Fried Platter	Fish	855	46	45	61	•	2043	14.7
		13 A	Deluxe Chic	Chicken	440	18	21	45	3	1411	9.5
		14 K	Pepperoni Rollup	Frozen	354	16	22	32	2	435	6
		15 A	Deluxe Turkey	Turkey	511	24	19	47	•	1046	9.7
		16 C	Deluxe Beef	Beef	534	24	33	44	•	1564	10.7
		17 E	Spicy Fried Chic	Chicken	568	32	23	44	0	1185	9.5
		18 E	Ranch Chic	Chicken	483	23	22	47	1	1346	9.5
		19 K	Beef and Cheese	Frozen	353	14	17	43	1	386	6

 You can alternately hide and show the three panels by clicking on the left-pointing diamond located in the upper-left corner of the data grid.



The Data Grid

The data grid organizes data into columns of *variables* and into rows of *observations*. This arrangement reflects the statistical nature of JMP and the kind of work for which it is best suited. Now that you have accessed the data, in order to make informed business decisions, you usually want to be able to answer questions about the data, such as the following:

- How many brands are represented?
- What is the 90th percentile of **Calories**?
- Is there a relationship between the **Calories** and **TFat**?

You can easily answer these questions because the data is stored in an accessible column-based format.

Likewise, you often want to address questions about or describe a subset of the sample:

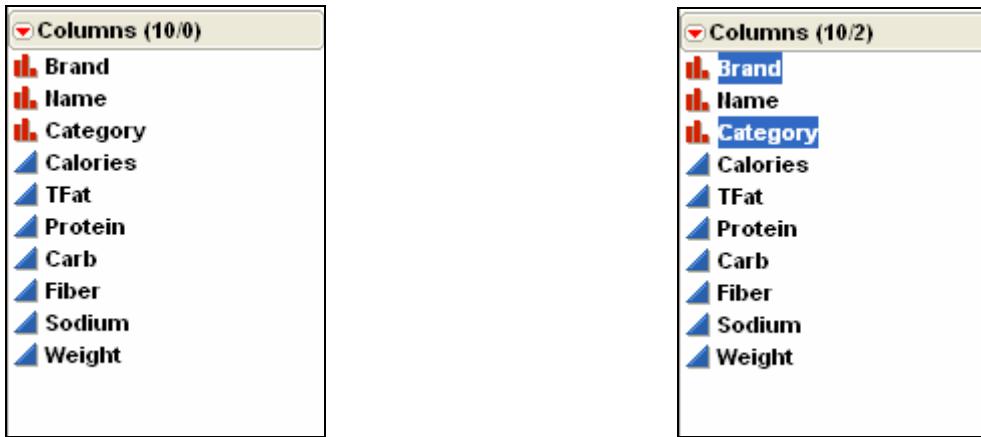
- What is the average calorie per sandwich by brand?
- Identify the different brands with different colors.

In these cases, you want to focus on an individual row or a group of rows. This is easily done when all of the data that is related to a single observation is organized row-wise.

The data grid does not open with the maximum number of rows and columns, as some financial spreadsheet applications do. This approach is not possible because there is no limit on the size of a data table imposed by JMP software. Instead, you create columns to store new variables as you need them and add rows when you have new observations. JMP determines the *sample size* (the number of rows) and the number of *missing values* (number of rows with empty cells) directly from the data table.

The Columns Panel

The Columns panel indicates the number of total columns (for example, 10) and selected columns (for example, 0) in the data table. It also presents a list of the columns with their properties, such as the column name. The second example below has ten total columns and two are selected (10/2).

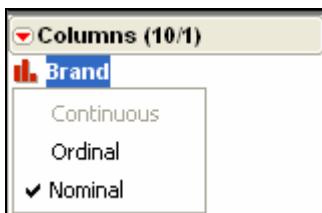


An icon precedes each column name to indicate the *modeling type*. There are two basic data modeling types: discrete and continuous. In general, discrete data takes on a limited number of values. In statistics, discrete data is often referred to as categorical data. There are two types of discrete data:

Nominal discrete values with no implicit order, used for a classification or grouping column (variable).

Ordinal discrete values that have a logical order but the intervals between values are not clear.

1. Click the modeling type icon in front of the **Brand** column.



Nominal is selected, but you have the ability to change it to Ordinal. The third choice, Continuous, is grayed out to indicate that this choice is not available for this column.

Nominal data can be numeric or character. If you designate 1, 2, and 3 as values, they might indicate group identification and would be considered nominal. However, they might also indicate some sort of order, in which case this variable would be considered ordinal.

In the example above, the columns **Brand**, **Name**, and **Category** are all categorical, specifically nominal.

The second data modeling type is continuous. It is always numeric.

Continuous numeric values with a logical order where the relative distances between the values are meaningful. These values are used directly in computations.

- Click the modeling type icon in front of the **Calories** column.



Calories has been designated as a continuous variable; it could be changed to one of the categorical types of data, provided the analyst considered the change necessary and correct.

In the example above, the columns between and inclusive of **Calories** and **Sodium** are continuous variables.

The initial modeling type displayed is the one that is automatically chosen by JMP. This choice can be changed by clicking on the modeling type for the appropriate column. By default JMP assigns all character data the nominal modeling type and all numeric data the continuous modeling type.

Having Some Fun with Columns

Selecting columns can be done in either the Columns panel or the data grid. After a column is chosen from one location, it is highlighted in the other.

- Click the column name **Category** from the data grid. **Category** is automatically highlighted in the Columns panel.

The screenshot shows the JMP interface with the 'Sandwiches' data table. The 'Category' column is selected in the data grid, which highlights it in the 'Columns' panel on the left. The 'Columns' panel lists all columns: Brand, Name, Category, Calories, TFat, and Protein. The 'Category' column is highlighted with a blue background in the 'Columns' panel.

		Brand	Name	Category	Calories	TFat	Protein
1	C	Big Fish		Fish	565	33	23
2	J	Spinach & Cheese Pocket		Frozen	223	5	13
3	C	Turkey Club		Turkey	518	23	30
4	G	Tuna on Wheat		Tuna	378	12	25
5	A	Baby Beef		Beef	339	16	13
6	K	Ham & Cheese		Frozen	339	16	15
7	E	Grilled Chic		Chicken	400	18	14
8	C	BBQ Chic		Chicken	286	5	25
9	L	Lite Reuben		Frozen	254	5	18
10	L	Lite Ham & Cheese		Frozen	260	9	5

2. To deselect selected column(s), hold down the CTRL key and click on the selected column(s).
3. To deselect all columns, find the rectangle in the upper-left corner of the data grid. Click in the upper-right triangle.



	Brand	Name	Category
1	C	Big Fish	Fish
2	J	Spinach & Cheese Pocket	Frozen
3	C	Turkey Club	Turkey

 An alternative way is to click in the bottom of the Columns panel, below the last column.



4. You can select **adjacent** columns by clicking on the first column, holding down the SHIFT key, and then clicking on the last column. All columns between the two selected are chosen. Click Brand, hold down the SHIFT key, and click Protein.




	Brand	Name	Category	Calories	TFat	Protein	Carb
1	C	Big Fish	Fish	565	33	23	45
2	J	Spinach & Cheese Pocket	Frozen	223	5	13	34
3	C	Turkey Club	Turkey	518	23	30	48
4	G	Tuna on Wheat	Tuna	378	12	25	44
5	A	Baby Beef	Beef	339	16	13	33
6	K	Ham & Cheese	Frozen	339	16	15	33
7	E	Grilled Chic	Chicken	400	18	14	39
8	C	BBQ Chic	Chicken	286	5	25	39
9	L	Lite Reuben	Frozen	254	5	18	39
10	L	Lite Ham & Cheese	Frozen	260	9	5	40
11	D	Grilled Chic	Chicken	297	8	19	35
12	I	Fried Platter	Fish	855	46	45	61

The columns between **Brand** and **Protein** have also been selected.

5. You can also select **non-adjacent** columns. In the Data Grid, click Brand, press and hold down the CTRL key, and then click Category and Fiber.



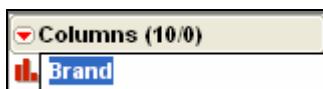
	Brand	Name	Category	Calories	TFat	Protein	Carb	Fiber	Sodium
1	C	Big Fish	Fish	565	33	23	45	5	1006
2	J	Spinach & Cheese Pocket	Frozen	223	5	13	34	2	794
3	C	Turkey Club	Turkey	518	23	30	48	*	1494

Changing the column name is done using the Columns panel or the data grid.

1. Click the column name **Brand** in the Columns panel.



2. Click the column name **Brand** in the Columns panel a second time. Notice that the background color changes and the column name is still highlighted.



3. Type the new name of the column, **Company**, and then press the ENTER key.



An alternative method for renaming columns using the data grid is presented below:

1. Click the column name **Brand** in the data grid.

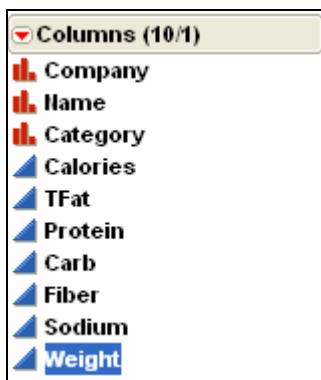
	Brand	Name	Category
1	C	Big Fish	Fish
2	J	Spinach & Cheese Pocket	Frozen
3	C	Turkey Club	Turkey

2. Type the new name of the column, **Company**, in the highlighted column name and press the ENTER key.

	Company	Name	Category
1	C	Big Fish	Fish
2	J	Spinach & Cheese Pocket	Frozen
3	C	Turkey Club	Turkey

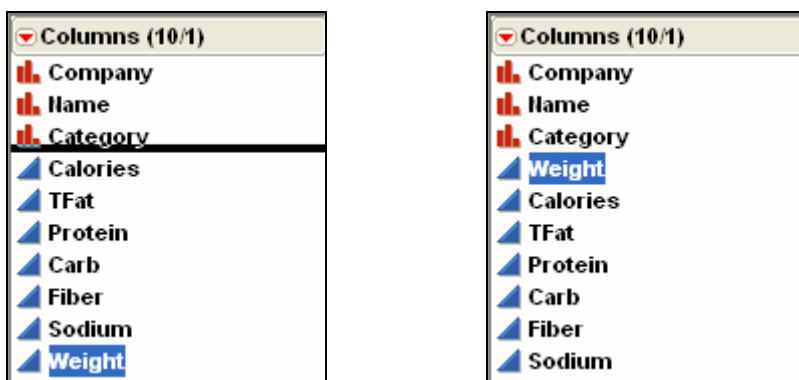
Moving columns can also be accomplished using the Columns panel or the data grid.

1. Click **Weight** in the Columns panel and continue to hold down the mouse button.



When you move the cursor, a bold black line appears.

2. Slide the black line to the desired position and release the mouse button. Notice that you can move more than one column concurrently. For this example, move **Weight** between **Category** and **Calories**.



An alternative method for moving columns using the data grid is presented below:

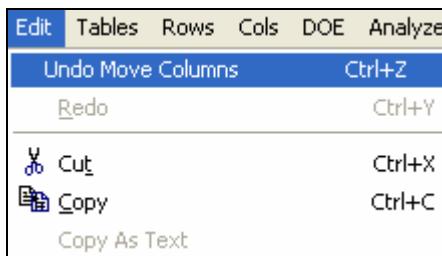
1. Click the column name **Weight** in the data grid. Click and continue to hold down the mouse button. As you slide the cursor, a black box appears.

	Brand	Name	Category	Calories	TFat	Protein	Carb	Fiber	Sodium	Weight
1	C	Big Fish	Fish	565	33	23	45	5	1006	9
2	J	Spinach & Cheese Pocket	Frozen	223	5	13	34	2	794	6
3	C	Turkey Club	Turkey	518	23	30	48	*	1494	10.7

2. Slide the box over the column where you want the highlighted column to be and release the mouse button. In this example, you want to move **Weight** in front of **Calories**, so **Calories** is highlighted with the box.

	Company	Name	Category	Weight	Calories
1	C	Big Fish	Fish	9	565
2	J	Spinach & Cheese Pocket	Frozen	6	223
3	C	Turkey Club	Turkey	10.7	518

When you move column(s), there is an option to restore the column(s) immediately to the previous position. To do this, select Edit \Rightarrow Undo Move Columns.



On to the Rows Panel

The Rows panel indicates the number of rows in the data table and how many are currently Selected, Excluded, Hidden, or Labelled.

All rows	64
Selected	0
Excluded	0
Hidden	0
Labelled	0

These conditions are known as *row states*. The counts of the row states are given in the Rows panel. In JMP, a row is synonymous with an observation. The row states of individual observations are indicated in the left margin of the data grid.

The Selected state is the easiest to use and understand. The other three row states are discussed in the next chapter. There are several ways to select specific rows or a range of rows.

Selecting a Specific Observation

To select a specific observation, click in the left margin next to the row number.

1. Click on row 15.
2. Notice that the values contained in this specific row have a different background color for emphasis.

Rows	13	A	Deluxe Chic	Chicken	9.5	440	18	21	45	3	1411
All rows	64	14	K	Pepperoni Rollup	Frozen	6	354	16	22	32	2 435
Selected	1	15	A	Deluxe Turkey	Turkey	9.7	511	24	19	47	• 1046
Excluded	0	16	C	Deluxe Beef	Beef	10.7	534	24	33	44	• 1564
Hidden	0	17	E	Spicy Fried Chic	Chicken	9.5	568	32	23	44	0 1185
Labelled	0	18	E	Ranch Chic	Chicken	9.5	483	23	22	47	1 1346

Selecting Contiguous Rows

To select contiguous rows, click at the start of the selection and drag the cursor to the opposite end.

1. Click on row 16 and continue to hold down the mouse button.
2. Drag the cursor to row 19.
3. Release the mouse button on row 19. Four rows have been selected.

Rows	15	A	Deluxe Turkey	Turkey
All rows	64	16	C	Beef
Selected	4	17	E	Chicken
Excluded	0	18	E	Chicken
Hidden	0	19	K	Frozen
Labelled	0	20	A	Beef Sub

-  An alternative is to click at the start of the desired rows and then hold down the SHIFT key and click at the opposite end of the selection. In the example above, click on row 16, hold down the SHIFT key and click on row 19.

Deselecting Row(s)

To deselect a row, hold down the CTRL key and click on the specific selected row(s).

- If rows 16 through 19 are selected, and you want to deselect row 18, hold down the CTRL key and click on row 18. The result of this action is shown below:

Rows		15	A	Deluxe Turkey	Turkey
All rows	64	16	C	Deluxe Beef	Beef
Selected	3	17	E	Spicy Fried Chic	Chicken
Excluded	0	18	E	Ranch Chic	Chicken
Hidden	0	19	K	Beef and Cheese	Frozen
Labelled	0	20	A	Beef Sub	Beef

-  As with columns, if you want to **deselect all rows**, you can find the rectangle in the upper-left corner of the data grid. Click in the lower-left triangle.

	Brand	Name	Category
1	C	Big Fish	Fish
2	J	Spinach & Cheese Pocket	Frozen
3	C	Turkey Club	Turkey

Combining Row Selection Actions

You can also combine these described actions. Suppose you want to add rows 25 through 28 to the selected rows of 16, 17, and 19. Hold down the CTRL key and click to extend the selection without including the rows in between.

- Hold down the CTRL key and click on row 25 and continue to hold down the mouse button.
- Drag down to row 28 and release the mouse button. Seven observations have been selected.

Rows		15	A	Deluxe Turkey	Turkey
All rows	64	16	C	Deluxe Beef	Beef
Selected	7	17	E	Spicy Fried Chic	Chicken
Excluded	0	18	E	Ranch Chic	Chicken
Hidden	0	19	K	Beef and Cheese	Frozen
Labelled	0	20	A	Beef Sub	Beef
		21	H	Chic Club	Chicken
		22	L	Lite Chic Broccoli	Frozen
		23	L	Lite Veggie Egg	Frozen
		24	A	Ham & Cheese	Ham
		25	G	Veggie Fever	Veggie
		26	A	Chic Breast	Chicken
		27	G	Chic Breast	Chicken
		28	A	Hamburger	Beef
		29	L	Lite Muffin	Frozen

- Deselect all rows.

The Table Panel

The Table panel is the top panel in the JMP window. It indicates the current name of the data table and presents a list of the current table variables and properties. A *table variable* has a constant character value that is always available in the data table for which it is created. It is commonly used to document tables. In this example, the data table **Sandwiches.JMP** has a table variable (**Notes**) that gives the location of the file.



1. To access the Table panel, click the red triangle next to **Sandwiches**.



- A red triangle in JMP always indicates that there are additional actions that are available.
2. To add a new table variable, select **New Table Variable**.

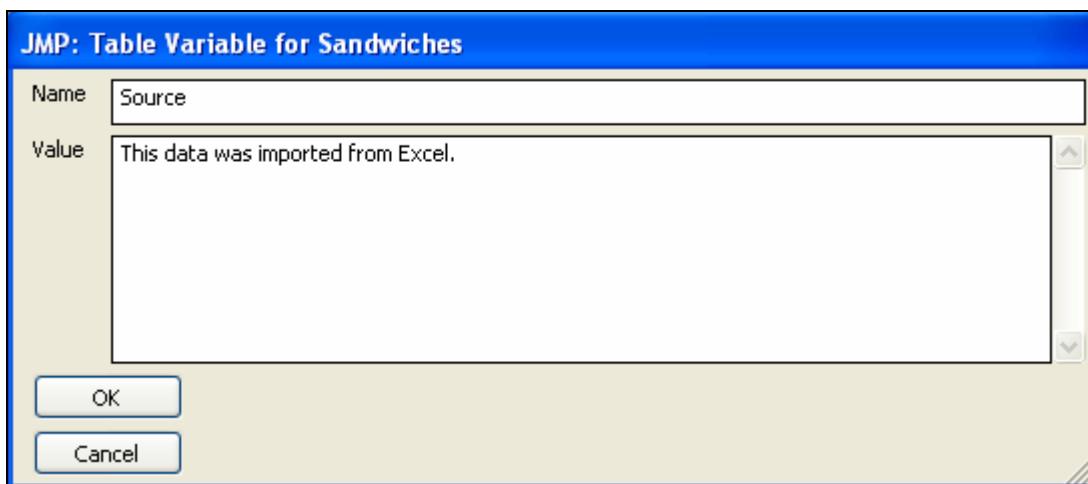


JMP provides a pop-up text window.

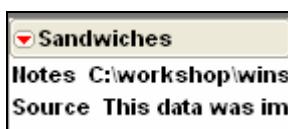


Add a table variable that describes the origin of this JMP data table.

3. Type **Source** as the name.
4. Type **This data was imported from Excel.** as the value.



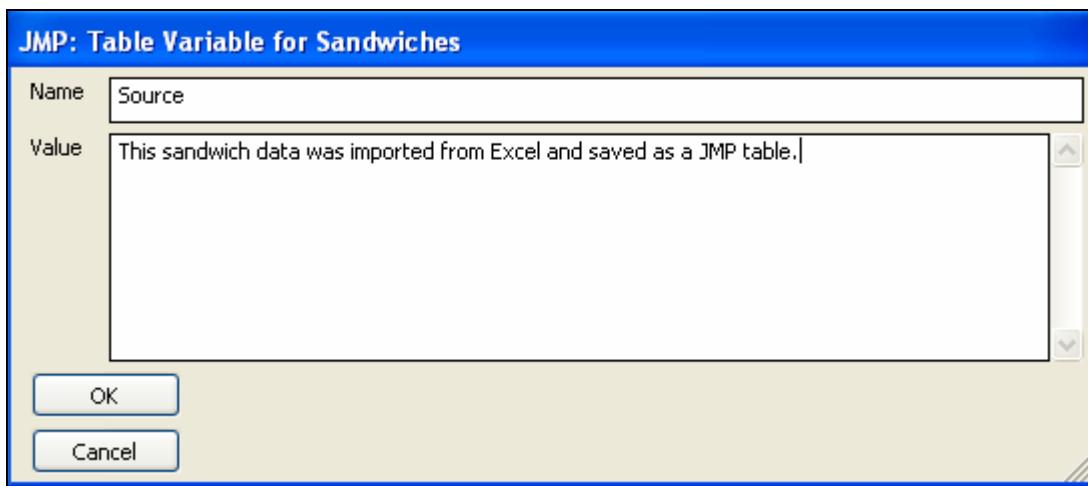
5. Select **OK**.
6. The Table panel now has two table variables.



7. To edit the contents of a table variable, right-click its name, **Source**, and select **Edit**.



You can change the name or the current value of this variable.



8. Select **OK**.
9. Save the file and close it.

1.4 Chapter Summary

All of the features of JMP are intended to function together as a concerted set of tools that support a unique approach to data analysis. You can appreciate why JMP was developed this way.

JMP is interactive, with more than one way of accomplishing the same task. The menu bar and the JMP Starter window organize a number of the various types of tasks.

The JMP Starter window provides an easy-to-use starting point to the many attributes of JMP. Within it, you find many grouped tasks, access to JMP Help, and the Preferences menu, which helps you tailor components of your individual JMP environment.

JMP can read data stored in a number of different formats. JMP is able to directly access data stored in ODBC compatible products.

The JMP data table has four components: the data grid, Columns panel, Rows panel, and Table panel.

- The data grid is arranged in columns (variables) and rows (observations).
- In the Columns panel, you can select a column or columns, change the name of column(s), and move column(s) to a different location in the data grid.
- In the Rows panel, you can select a row or rows.
- In the Table panel, you can add a table variable that can store explanations or notes about the data stored in the data grid.

Chapter 2 Data Exploration and Manipulation

2.1 The Columns and Rows Menus	2-3
2.2 Creating New Columns	2-27
2.3 Creating New Tables	2-39
2.4 Using the Tabulate Option to Create Summary Tables	2-53
2.5 Chapter Summary.....	2-64

2.1 The Columns and Rows Menus

Objectives

- Use the Columns menu to change attributes of existing columns.
- Use the Rows menu to logically select rows and/or change attributes of rows.

3

The Table and Rows panels provide some basic functionality in manipulating their respective elements of the data grid. With the introduction of their associated menus, more complex and diverse functions are added for both columns and rows.



Using the Columns and Rows Menus

You can change the column name, modeling type, or virtually anything about the column via the Columns menu. You can also add new properties. (The Rows menu provides similar properties for the rows of the data table.)

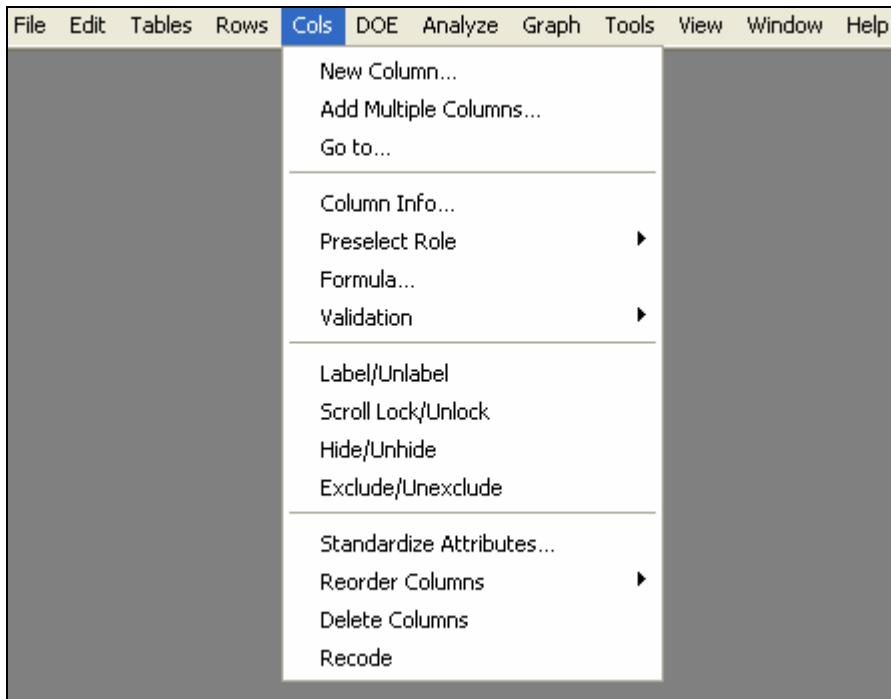
1. If **Sandwiches.JMP** is not open, then use the File category in the JMP Starter window to select **Open Data Table**.
2. Select **Sandwiches.JMP** \Rightarrow **Open**.

You can invoke the Columns menu from three locations within the JMP table. The most obvious is from the menu bar. In order to see all of the functionality of the Columns menu, select an existing column.

3. Select **Company** from the data grid or from the Columns panel.



4. Select **Cols** from the menu bar. The following menu is displayed.

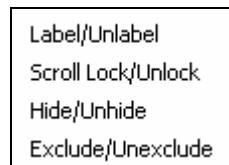




There are two other locations to invoke the Columns menu. Find the rectangle in the upper-left corner of the data grid and click the red triangle in the upper-right triangle, or click the red triangle on the left side of the Columns panel.

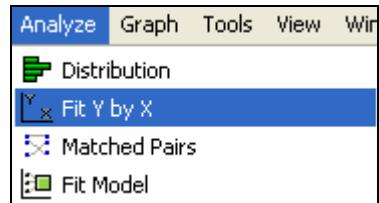


You are presented with a number of possible actions in the Columns menu, but start with the following subset:



To examine these options, generate a scatterplot.

1. Select **Analyze** \Rightarrow **Fit Y by X**.

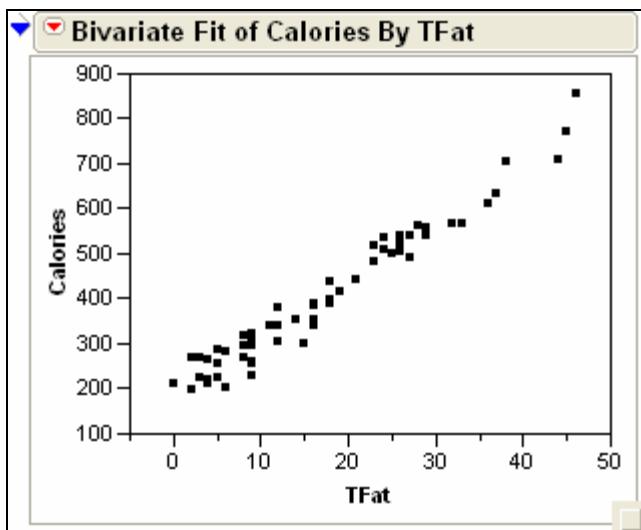


2. Select **Calories** \Rightarrow **Y, Response**.
3. Select **TFat** \Rightarrow **X, Factor**.



4. Select **OK**.

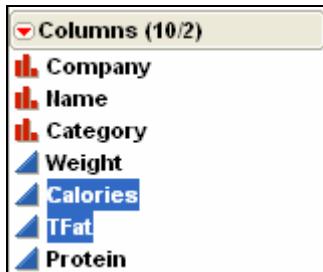
The following scatterplot is created:



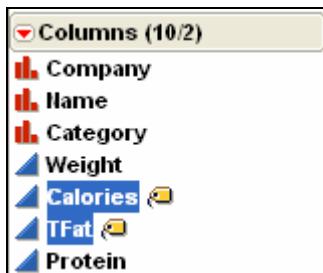
5. Select a point in the graph, such as the point in the upper-right corner of the graph. The row number in the data table corresponding to this point is displayed.

Suppose that you do not want the row number to appear, but instead you want to see the values of **Calories** and **TFat**. To do this, label **Calories** and **TFat**.

6. Select **Window** \Rightarrow **Sandwiches**.
7. Select the two columns, **Calories** and **TFat**.



8. Select **Cols** \Rightarrow **Label/Unlabel**.



If a column is being used as a label, individual points in a JMP graph are displayed with the value of the labeled column instead of the row number. If more than one column is used as a label, the values for both are displayed. The label icon (tag), located in the Columns panel beside the column name, indicates that the column is labeled.

9. Select **Window** \Rightarrow **Sandwiches – Fit Y by X of Calories by TFat**.

10. Select a point in the graph, such as the point in the upper-right corner of the graph.

Now, instead of the row number, the values for **Calories** and **TFat** are displayed. Remove these two columns as labels.

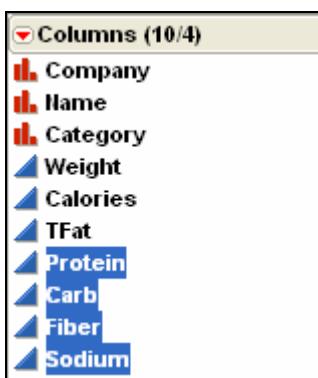
11. Select **Cols** \Rightarrow **Label/Unlabel**.

The Hide/Unhide command controls the display of column(s) in the data grid. Currently all ten columns are visible in the data grid.

Hide all of the columns between **Protein** and **Sodium**, inclusive.

1. Select **Window** \Rightarrow **Sandwiches**.

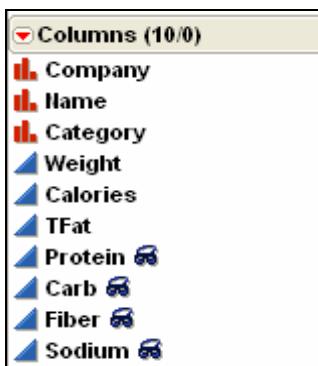
2. Select the columns that you want to hide: click **Protein** and hold the SHIFT key down and click **Sodium**.



3. Select **Cols** \Rightarrow **Hide/Unhide**.

4. Examine the data grid and Columns panel.

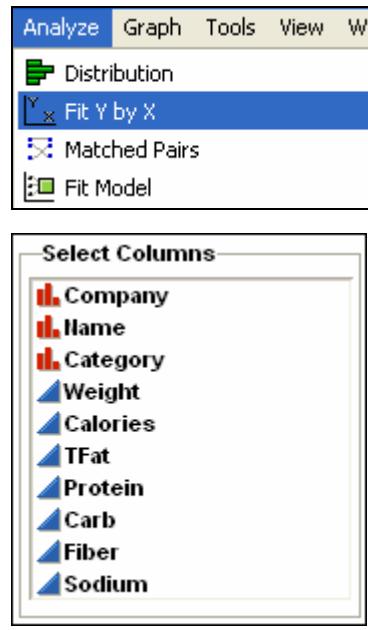
	Company	Name	Category	Weight	Calories	TFat	
1	C	Big Fish	Fish	9	565	33	
2	J	Spinach & Cheese Pocket	Frozen	6	223	5	
3	C	Turkey Club	Turkey	10.7	518	23	
4	G	Tuna on Wheat	Tuna	9.7	378	12	
5	A	Baby Beef	Beef	6	339	16	



In the Columns panel, notice the hidden icon (mask) beside the hidden columns. The hidden columns are not displayed in the data grid. The 6/0 Cols in the right triangle of the upper-left corner of the data grid indicates that six columns are displayed and none are selected.

If a column is hidden, it can still participate in an analysis. To see this, return to the Fit Y by X platform.

5. Select **Analyze** \Rightarrow **Fit Y by X**.

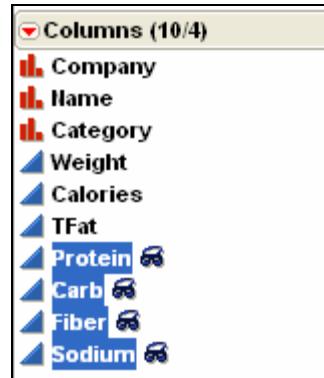


Notice that each of the ten columns is available to use in the scatterplot. Therefore hiding a column does not prevent it from participating in an analysis. The Exclude/Unexclude command is used to eliminate column(s) from any analyses.

1. Select **Cancel** to close the Fit Y by X platform.

Cancel

2. Select **Protein** through **Sodium**. These columns are hidden, so they must be selected in the Columns panel.



3. Select **Cols** \Rightarrow **Exclude/Unexclude**.

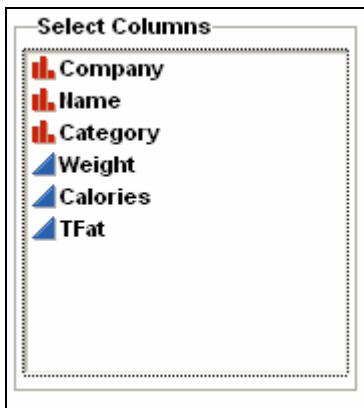
Notice the exclusion icon (crossed circle) beside the excluded columns, found in the Columns panel.



The excluded columns cannot be used in any sort of analyses. To see this, return to the Fit Y by X platform.

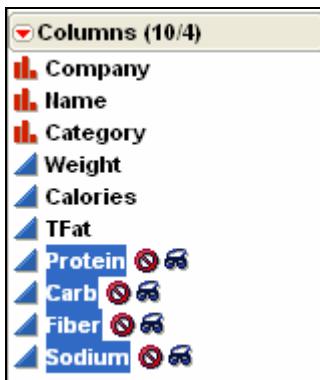
4. Select **Analyze** ⇒ **Fit Y by X**.

Notice that only **Company** through **TFat** can be selected for this scatterplot. The other columns are excluded.

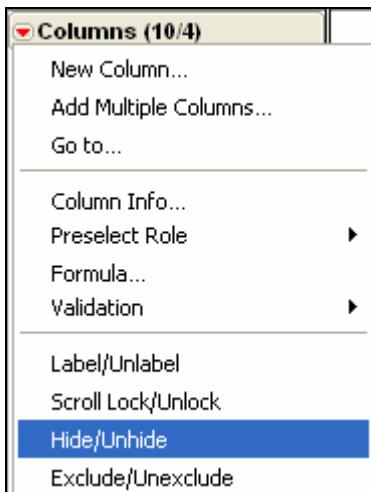


6. Select **Cancel** to close the Fit Y by X platform.

7. Be sure that **Protein** through **Sodium** are selected in the Columns panel.



8. Click the red triangle in the Columns panel and select **Hide/Unhide**.



The columns now appear in the data grid.

Sandwiches											
	Company	Name	Category	Weight	Calories	TFat	Protein	Carb	Fiber	Sodium	
1	C	Big Fish	Fish	9	565	33	23	45	5	1006	
2	J	Spinach & Cheese Pocket	Frozen	6	223	5	13	34	2	794	
3	C	Turkey Club	Turkey	10.7	518	23	30	48	*	1494	
4	G	Tuna on Wheat	Tuna	9.7	378	12	25	44	3	1024	
5	A	Baby Beef	Beef	6	339	16	13	33	0	573	
6	K	Ham & Cheese	Frozen	6	339	16	15	33	4	607	
7	E	Grilled Chic	Chicken	9	400	18	14	39	0	975	
8	C	BBQ Chic	Chicken	8.2	286	5	25	39	3	1118	
9	L	Lite Reuben	Frozen	6.5	254	5	18	39	5	569	
10	L	Lite Ham & Cheese	Frozen	5.5	260	9	5	40	3	946	
11	D	Grilled Chic	Chicken	8	297	8	19	35	2	1163	
12	I	Fried Platter	Fish	14.7	855	46	45	61	*	2043	
13	A	Deluxe Chic	Chicken	9.5	440	18	21	45	3	1411	

9. Click the red triangle in the Columns panel and select **Exclude/Unexclude**.

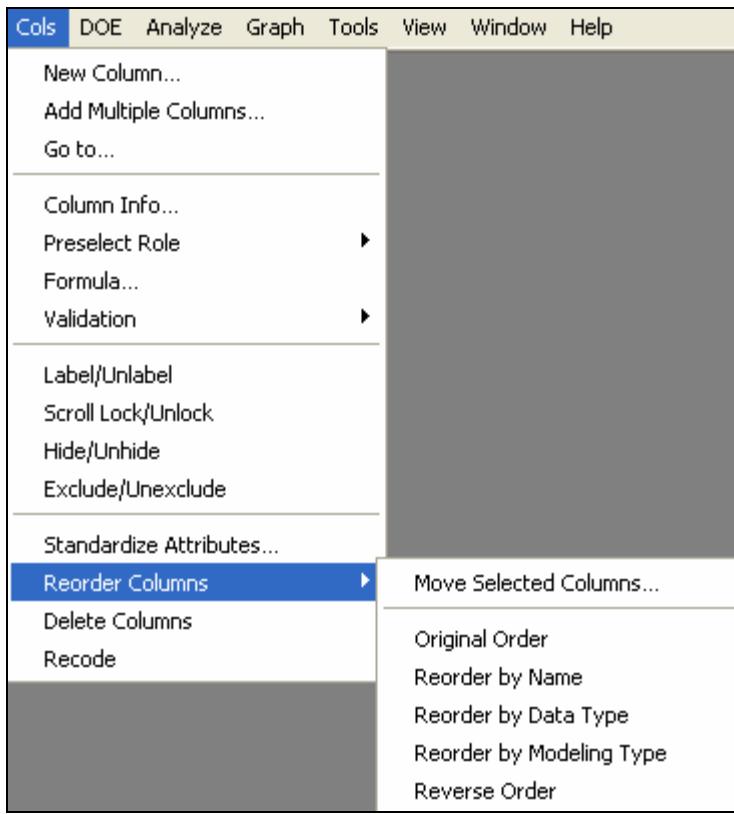
The Scroll Lock/Unlock command is used to lock the selected column(s) on the left of the data grid. If selected, the names of the column(s) will be italicized in the Table panel.

The four commands that have been described are simple switches. They are simply turned off or on. For example, if a column is hidden, then it can only be unhidden. There are no complex logical combinations of these commands available.

Portions of the remaining commands accessible from the Columns menu have extended submenus. Explore them in the following order: Reorder Columns, Preselect Role, and Validation.

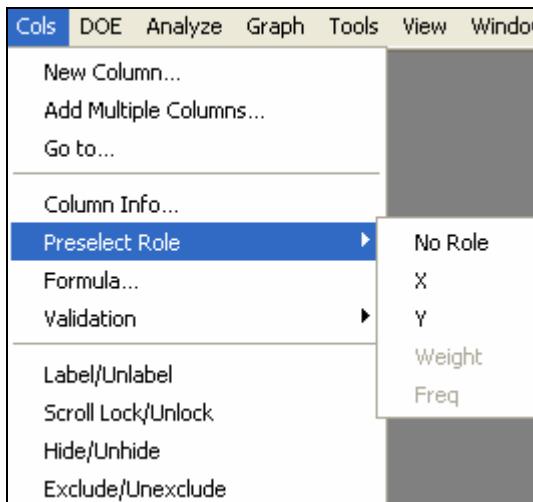
1. Select **Company**.

2. Select Cols \Rightarrow Reorder Columns.



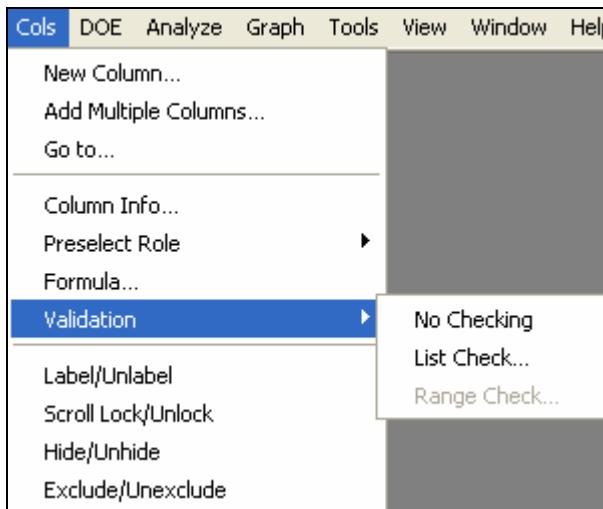
The Reorder Columns command enables you to move columns according to the selection that you make from its submenu. You can change the order of the columns in the data table by name, data type, modeling type, reverse order, or the original order. You have already learned that you can also manually select and drag columns to a new location. This operation works in the data grid or the Columns panel.

3. Select Cols \Rightarrow Preselect Role.



The Preselect Role command assigns a role to the selected column. *Role* refers to how a column participates in a JMP analysis or graph.

4. Select **Cols** \Rightarrow **Validation**.



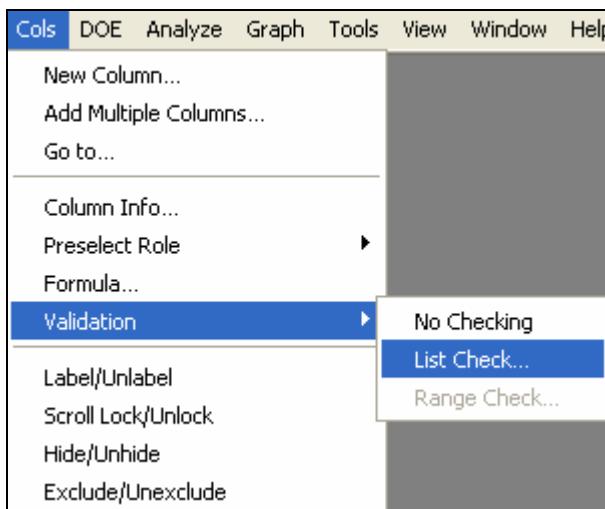
The Validation command lets you enter a list of valid values or valid range limit conditions for a column, which then restricts the values that can be entered into the columns. By default, you can enter any data into a column as long as it agrees with the data type (character or numeric). You can impose further restrictions by using one of the following characteristics:

No Checking	no restrictions whatsoever.
List Check...	usually used with nominal data to ensure that all values entered for a column correspond to values in a validated list.
Range Check...	used only with continuous data when only a limited range of values is permitted. Inclusion rules are defined with respect to the end points.

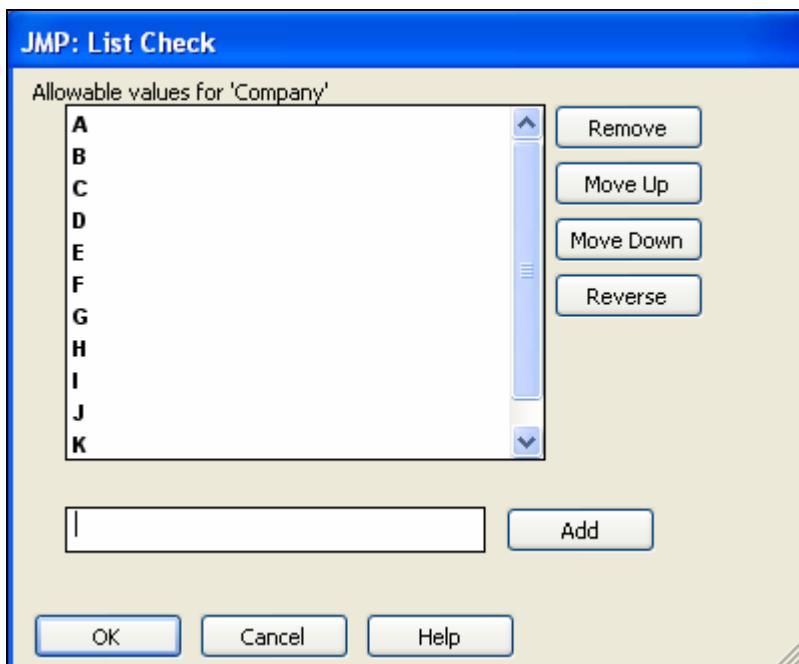
You can define or change the validation of a column through these commands. Notice that **Company** is a nominal variable, so Range Check... is grayed out, indicating that it is not a valid choice for this column.

Suppose you want to restrict the values of **Company** to the existing list. Adding a List Check... to this column saves typing and eliminates future entry errors in this column.

1. Select **Cols** \Rightarrow **Validation** \Rightarrow **List Check...**



JMP presents the valid list in a pop-up dialog box.



Notice that the list is ordered alphanumerically and that it is case sensitive. You can also add acceptable values via this dialog box by typing the value and selecting **Add**.

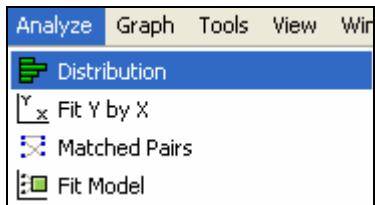
2. Select **OK** and locate the Table panel.



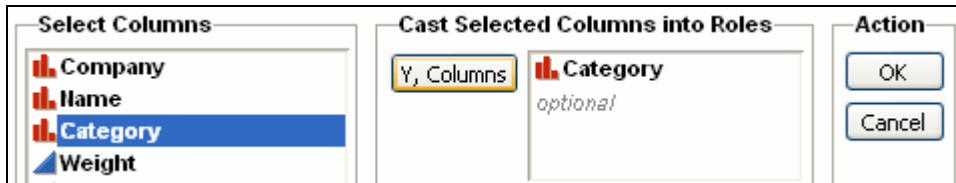
Notice the list check icon that indicates that this column has restricted values.

Data entry errors can be recoded quickly to ensure that the data conforms to a consistent format. As an example, examine a distribution of **Category**.

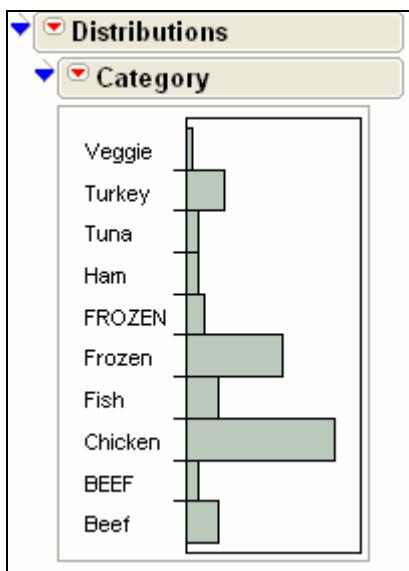
1. Select **Analyze** ⇒ **Distribution**.



2. Select **Category** ⇒ **Y, Columns**.



3. Select **OK**.

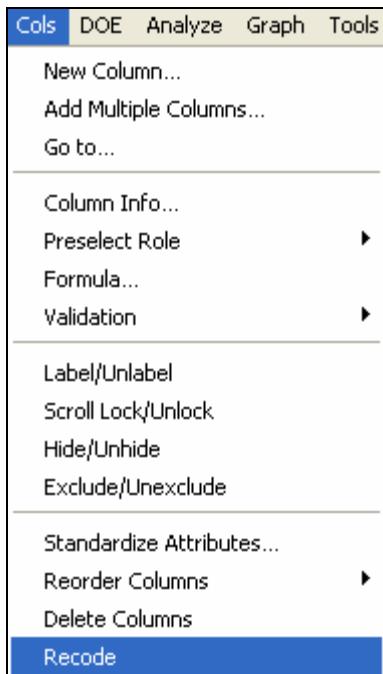


Notice that the values in **Category** have been entered inconsistently, with some entered in all uppercase and some in mixed case. You decide that you prefer mixed case.

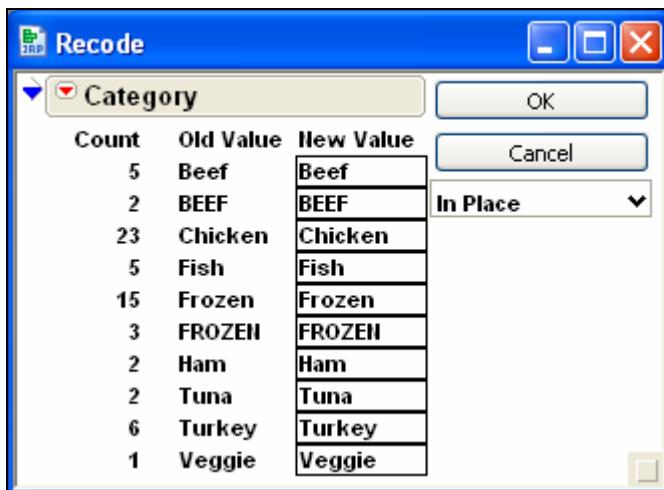
4. Select **Window** ⇒ **Sandwiches**.

5. Select **Category**.

6. Select **Cols** \Rightarrow **Recode**.

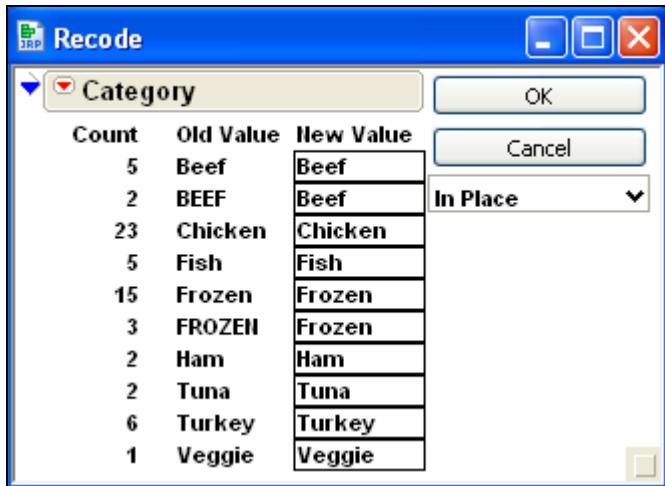


The Recode window opens and lists the values currently recorded in the selected column.



7. Click the text box next to BEEF and type **Beef**.
8. Click the text box next to FROZEN and type **Frozen**.

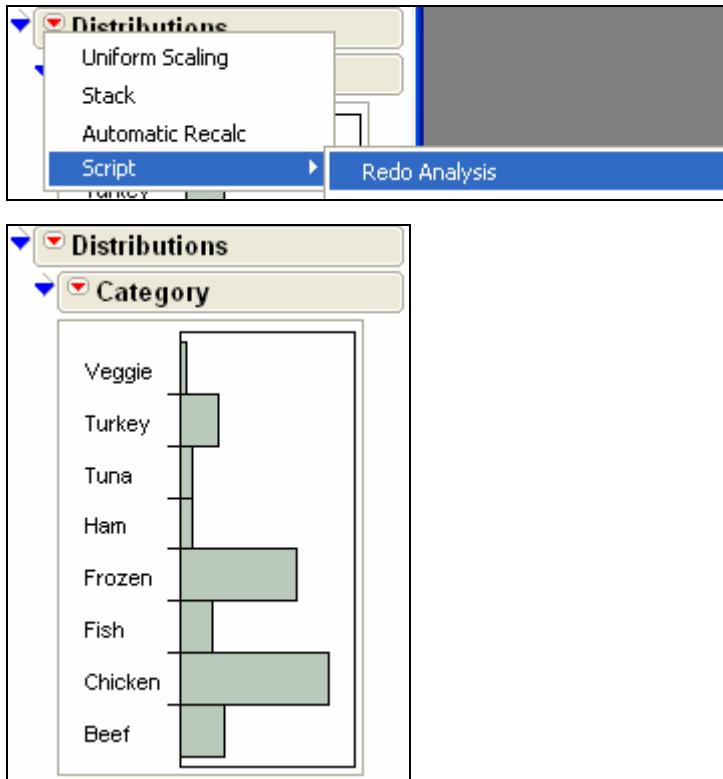
The correct or preferred values can be copied and pasted from text box to text box instead of typing.



9. Select OK.
10. Select Continue.

The values in **Category** are now consistent. To see this in the distribution, update the histogram.

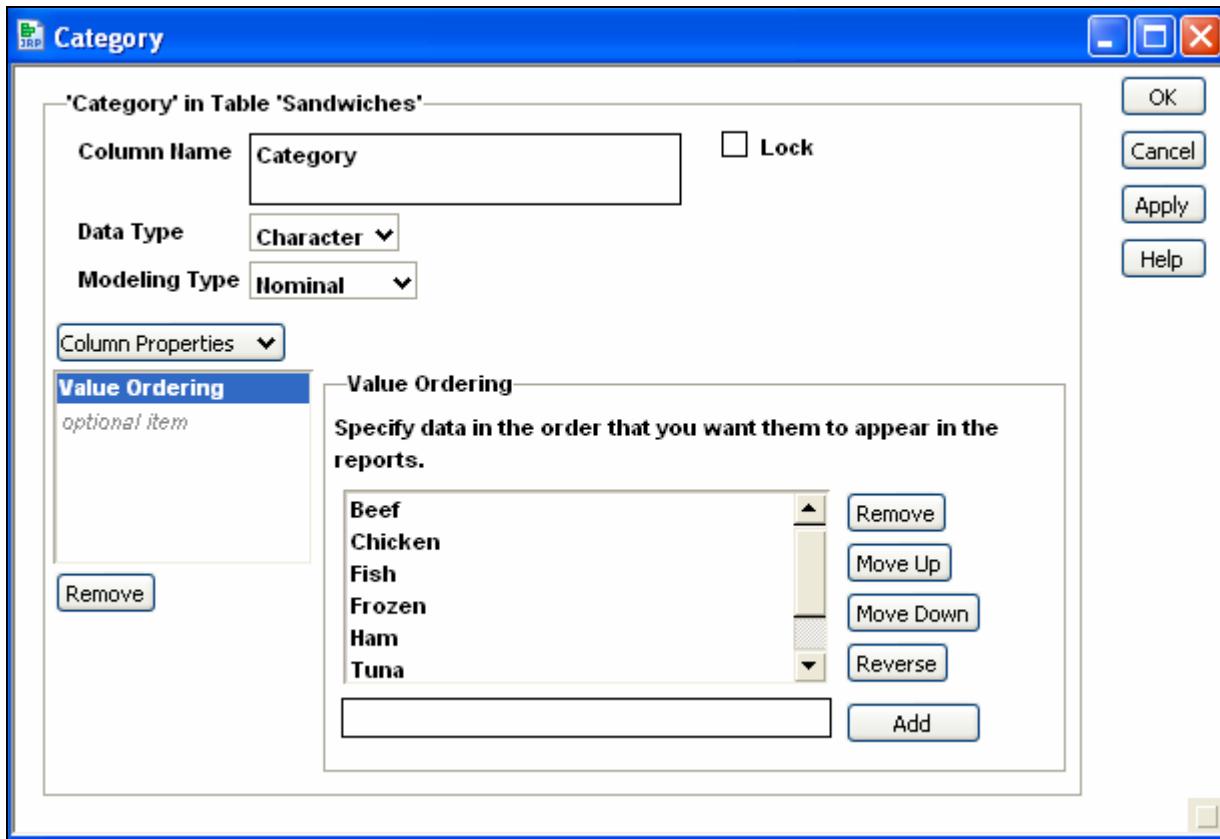
11. Select Window \Rightarrow Sandwiches – Distribution of Category.
12. Click the red triangle next to **Distributions** and select Script \Rightarrow Redo Analysis.



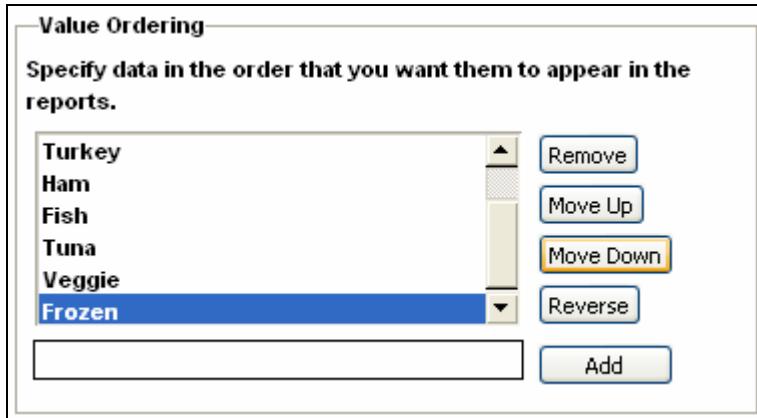
There is now one bar for all of the values for **Beef** and one bar for all of the values for **Frozen**.

By default, when creating charts using nominal columns, JMP orders the bars in the charts alphabetically from the bottom of a chart, which might not be the logical order of the values in the column. The order of the bars in a chart can be controlled using the Value Ordering property.

1. Close the two distribution windows.
2. In the **Sandwiches** data table, select Category.
3. Select Cols \Rightarrow Column Info....
4. Select Column Properties \Rightarrow Value Ordering.



5. Select **Turkey** and then select **Move Up** four times.
6. Select **Fish** and then select **Move Down** two times.
7. Select **Frozen** and select **Move Down** four times.



8. Select **OK**.
9. Click the asterisk next to **Category** in the Columns panel.



The Value Ordering property has been assigned to the Category column.

10. Deselect **Category**.

- The following values will be ordered correctly automatically by JMP:
- January, February, March, April, May, June, July, August, September, October, November, December
 - Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec
 - Sunday, Monday, Tuesday, Wednesday, Thursday, Friday, Saturday
 - Very Low, Low, Medium Low, Medium, Medium High, High, Very High
 - Strongly Disagree, Disagree, Neutral, Indifferent, Agree, Strongly Agree
 - Failing, Unacceptable, Very Poor, Poor, Bad, Acceptable, Average, Good, Better, Very Good, Excellent, Best

The remaining Column commands, New Column..., Add Multiple Columns..., Column Info..., and Formula... are examined and used later in this chapter.

And Now the Rows Menu

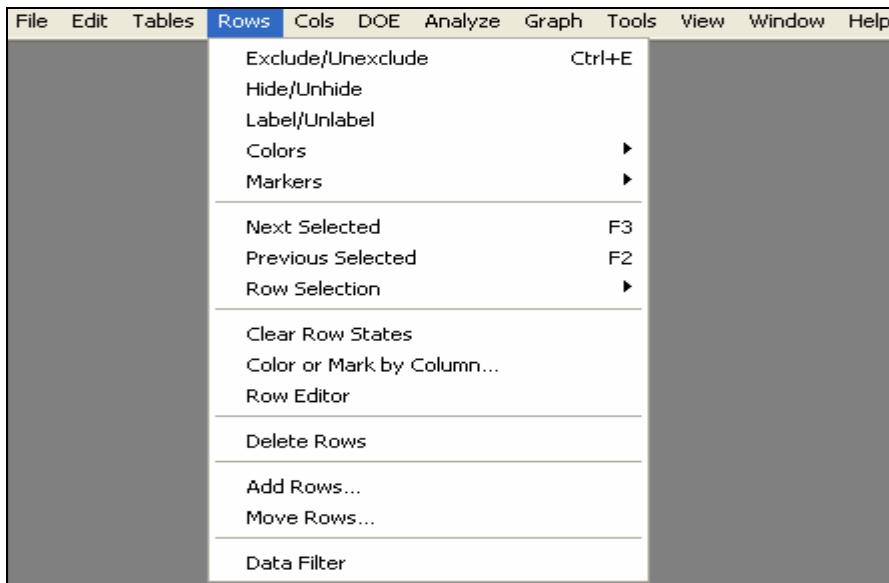
In the previous chapter you learned how to select rows using the mouse. The Rows menu adds more functionality, especially the ability to logically select rows. Later you will learn to select data by clicking graphical elements (for example, points or bars).

1. Select the first five rows and notice the contents of the Rows panel.

	Company	Name	Category
1	C	Big Fish	Fish
2	J	Spinach & Cheese Pocket	Frozen
3	C	Turkey Club	Turkey
4	G	Tuna on Wheat	Tuna
5	A	Baby Beef	Beef
6	K	Ham & Cheese	Frozen

Rows	
All rows	64
Selected	5
Excluded	0
Hidden	0
Labelled	0

2. Select **Rows** from the menu bar. The following menu is displayed:



There are two other locations to invoke the Rows menu from a JMP data table. Find the rectangle in the upper-left corner of the data grid and click the red triangle in the lower-left triangle, or click the red triangle on the left side of the Rows panel.



Notice that the first three commands in the Rows menu are directly related to the same three states in the Columns menu. They are also analogous to the Columns commands. These three commands control how the data participates in an analysis or graph:

Exclude/Unexclude alternately excludes or includes the selected rows in calculations.

Hide/Unhide alternately hides or shows the selected rows in graphs.

Label/Unlabel alternately labels or hides the label for data in graphs.

3. Deselect the first five rows.

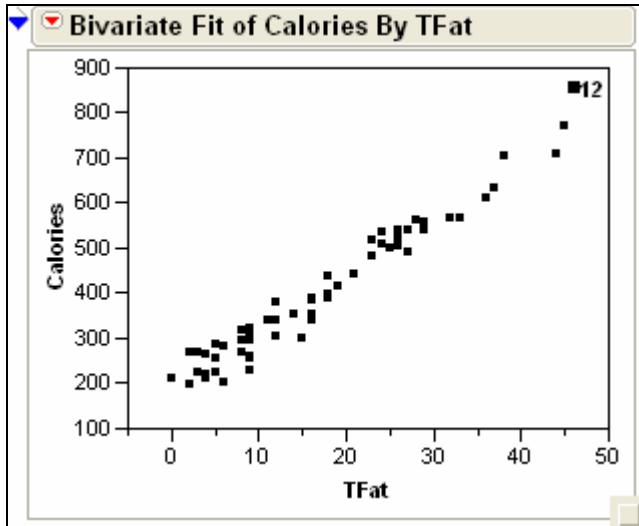
As you learned from the Columns menu, these commands are on/off switches. To examine these features in more detail, return to the scatterplot of **Calories** by **TFat**.

1. Select **Window** \Rightarrow **Sandwiches – Fit Y by X of Calories by TFat**.

Recall that the row number is displayed when a point is selected, unless one or more columns are labeled. It is necessary to hover over a point or select a point to see this information. To print this information on the graph, label the row.

2. Select a point in the graph, such as the point in the upper-right corner of the graph.
3. Select **Rows** \Rightarrow **Label/Unlabel**.

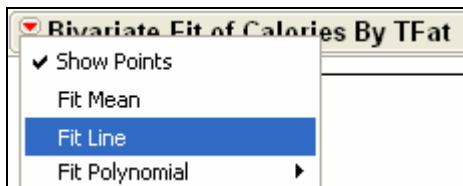
The row number is now printed on the graph.



4. Be sure that the point is still selected, and select **Rows** \Rightarrow **Hide/Unhide**.

Notice the point is now hidden from the graph. This does not mean the point will be excluded from any calculations or analyses. The Exclude/Unexclude command is used to eliminate row(s) from any analyses. To see this, fit a line to the data.

5. Click the red triangle next to **Bivariate Fit of Calories By TFat** and select **Fit Line**.



This is the line of best fit, based on all of the data points, even the hidden point. Exclude this point and refit the line.

6. Be sure that the point is still selected (you might need to check the data table) and select **Rows** \Rightarrow **Exclude/Unexclude**.

11	D	Grilled Chic	Chicken
12	I	Fried Platter	Fish
13	A	Deluxe Chic	Chicken

7. Click the red triangle next to **Bivariate Fit of Calories By TFat** and select **Fit Line**.

The location of the line changed slightly due to the exclusion of one of the data points.

8. To remove each line, click on the red triangle next to **Linear Fit** and select **Remove Fit**.

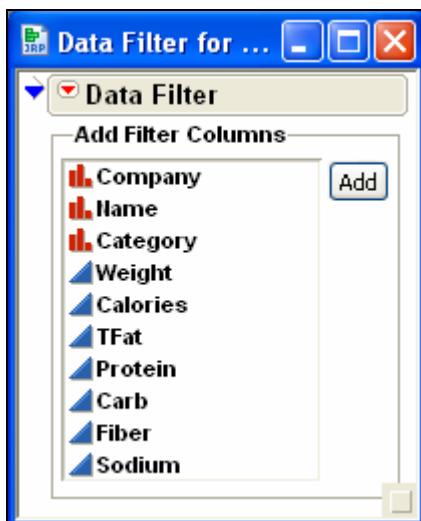
9. Select **Window** \Rightarrow **Sandwiches Unit Based**.

10. To unlabel, unexclude, and unhide this data point, select **Rows** \Rightarrow **Clear Row States**.

The Data Filter gives the freedom to create more complicated logical selections.

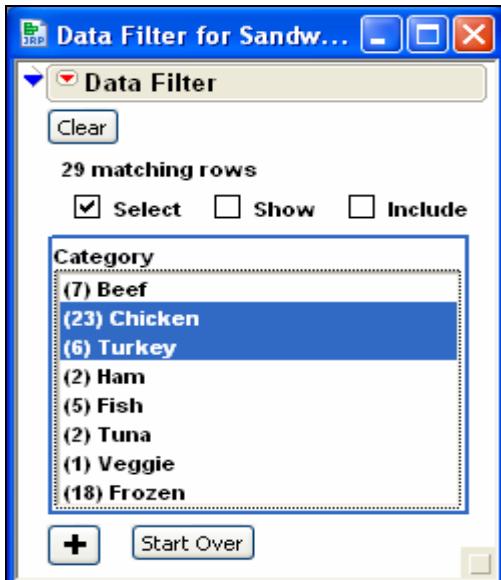
Suppose that you want to examine **Calories** and **TFat** for the turkey and chicken sandwiches.

1. Select **Rows** \Rightarrow **Data Filter**.



2. Select **Category** in the list of columns.
3. Select **Add**.

4. Select **(23) Chicken**.
5. While holding down the CTRL key, select **(6) Turkey**.



- In the data filter, the Show option means to show only the selected rows and hide the non-selected rows, and the Include option means to include only the selected rows and exclude the non-selected rows.
- The 29 rows (out of the 64 total rows) corresponding to poultry sandwiches are now selected in the data table.

Rows	
All rows	64
Selected	29
Excluded	0
Hidden	0
Labelled	0

14	K	Pepperoni Rollup	Frozen
15	A	Deluxe Turkey	Turkey
16	C	Deluxe Beef	Beef
17	E	Spicy Fried Chic	Chicken
18	E	Ranch Chic	Chicken

- An alternative method for selecting specific rows is to select **Rows** \Rightarrow **Row Selection** \Rightarrow **Select Where...**.

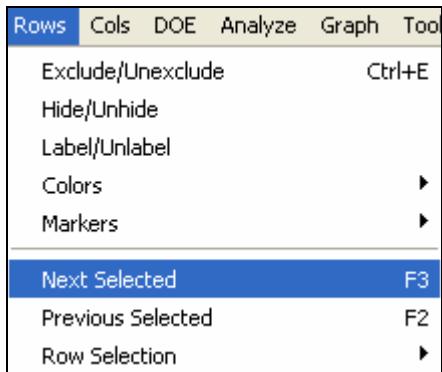


6. Close the Data Filter window.

You can scroll through the data grid to find each row that has been selected. However, the Next Selected and Previous Selected commands are convenient ways to examine the selection in a large data table. Both commands scroll the data grid to show the adjacent row in the selection. Next Selected scrolls downward; Previous Selected moves upward.

Explore how JMP displays the adjacent rows.

1. Select **Rows** \Rightarrow **Next Selected**.

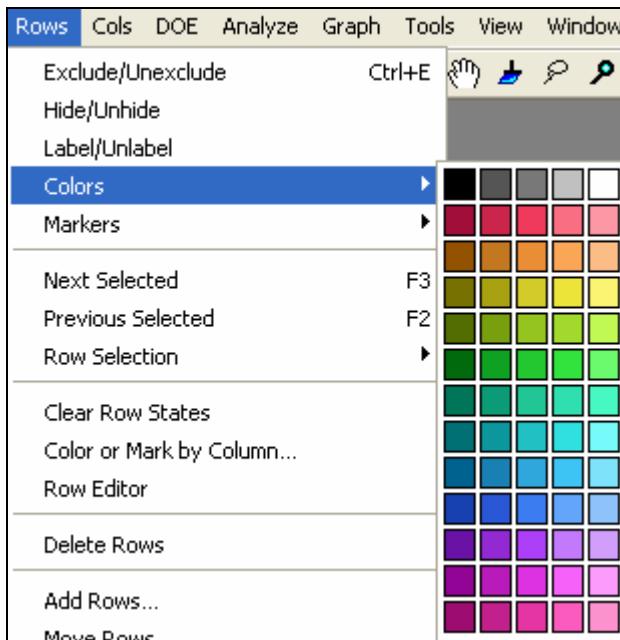


JMP responds to the request by flashing the row number, 3 in this case.

2. Select **Rows** \Rightarrow **Next Selected** a second time and JMP flashes row number 7.
3. Select **Rows** \Rightarrow **Next Selected** a third time and JMP flashes row 8.

To further identify this logical selection of rows in the data table and the scatterplot, assign these 29 selected values a specific color and marker to be used in JMP graphics and analyses.

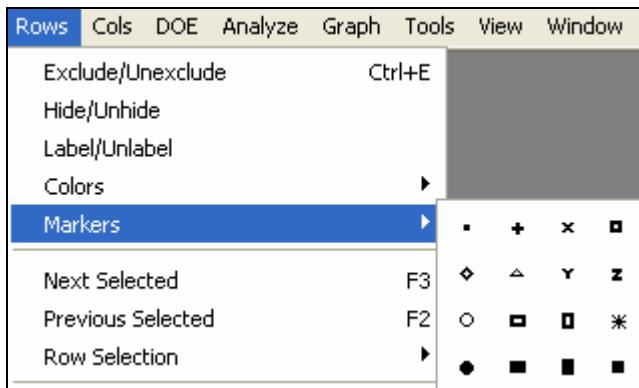
1. Select **Rows** \Rightarrow **Colors** and select any of the 65 possible colors. For illustrative purposes, choose a shade of red.



JMP changes the default marker color of black to the shade of red that you selected. Notice the change in the data grid to any of the selected rows. A red dot now appears beside the row number.

▪	15	A	Deluxe Turkey	Turkey
	16	C	Deluxe Beef	Beef
▪	17	E	Spicy Fried Chic	Chicken
▪	18	E	Ranch Chic	Chicken
	19	K	Beef and Cheese	Frozen

- To identify the selected rows by assigning them a unique marker, instead of the default dot, select **Rows** \Rightarrow **Markers** and choose the plus sign from the list of 16 available markers.

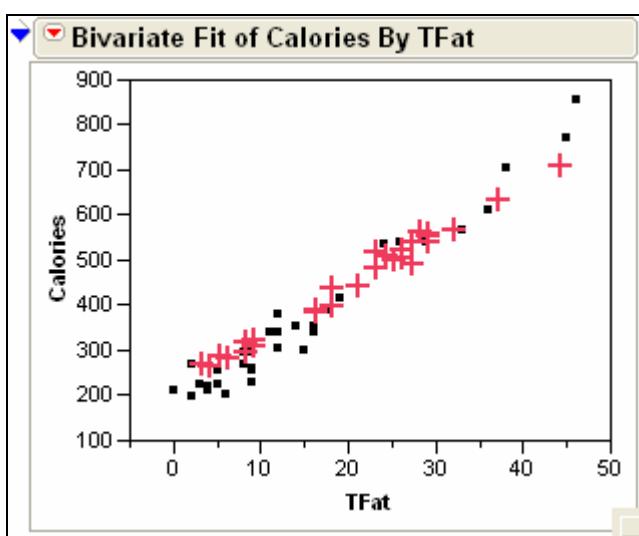


JMP now displays a red plus sign for those rows where **Category** is Turkey or Chicken.

+	15	A	Deluxe Turkey	Turkey
---	-----------	---	----------------------	--------

Navigate to the scatterplot to examine the color and marker changes.

- Select **Window** \Rightarrow **Sandwiches – Fit Y by X of Calories by TFat**.



The data points that correspond to the poultry sandwiches are distinguished from the remaining points.

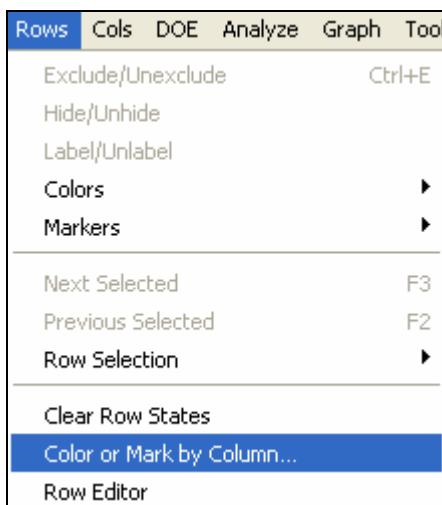
- Close the scatterplot window.

Clear the color and marker assigned to these rows.

3. Select **Rows** \Rightarrow **Clear Row States**.

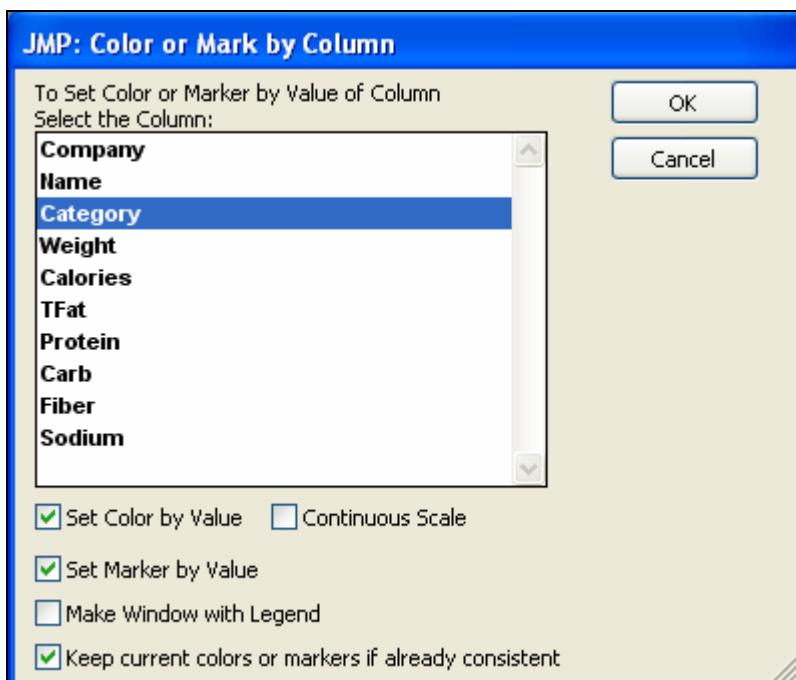
Finally, suppose that you want to identify each level of a categorical variable with a unique color and marker. Using **Category**, you could use the data filter for each value, but there are eight values of **Category**. JMP has another way to do it that is easier.

1. Select **Rows** \Rightarrow **Color or Mark by Column...**.



2. Select **Category** in the list of columns. **Set Color by Value** is automatically selected.

3. Select **Set Marker by Value**.



4. Select **OK**.

	Company	Name	Category
1	C	Big Fish	Fish
2	J	Spinach & Cheese Pocket	Frozen
3	C	Turkey Club	Turkey
4	G	Tuna on Wheat	Tuna
5	A	Baby Beef	Beef
6	K	Ham & Cheese	Frozen
7	E	Grilled Chic	Chicken
8	C	BBQ Chic	Chicken
9	L	Lite Reuben	Frozen

JMP has given each value of **Category** a unique color and marker.

5. Select **File** ⇒ **Save** or select the Save icon on the menu bar.

2.2 Creating New Columns

Objectives

- Use the JMP Formula Editor to create new numeric columns.
- Create a new nominal character column from an existing nominal or ordinal column.

10

In many analyses, the data that you collect is inadequate in its current form for your purpose(s); possibly it is in an unusable scale. You must change or transform the data to extract the most information from it.

In **Sandwiches . JMP**, the column **Weight** varies for each row, indicating that some sandwiches are heavier than others. A reasonable approach in this situation is to divide all the numeric columns by **Weight** in order to generate a "unit" value. These new columns would, therefore, compensate for larger values in heavier sandwiches and smaller values in lighter sandwiches.

Could there be a difference between frozen and fresh sandwiches in terms of **Calories** or **TFat**? Create a new nominal column that makes the distinction between frozen and fresh sandwiches.



Creating New Columns

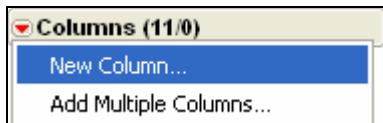
This demonstration illustrates the methods to create both numeric and character columns. It also makes use of the JMP Formula Editor. Create a new column that will be the quotient of **Calories** and **Weight**, called **Unit Calories**.

1. Select **Cols** \Rightarrow **New Column...**

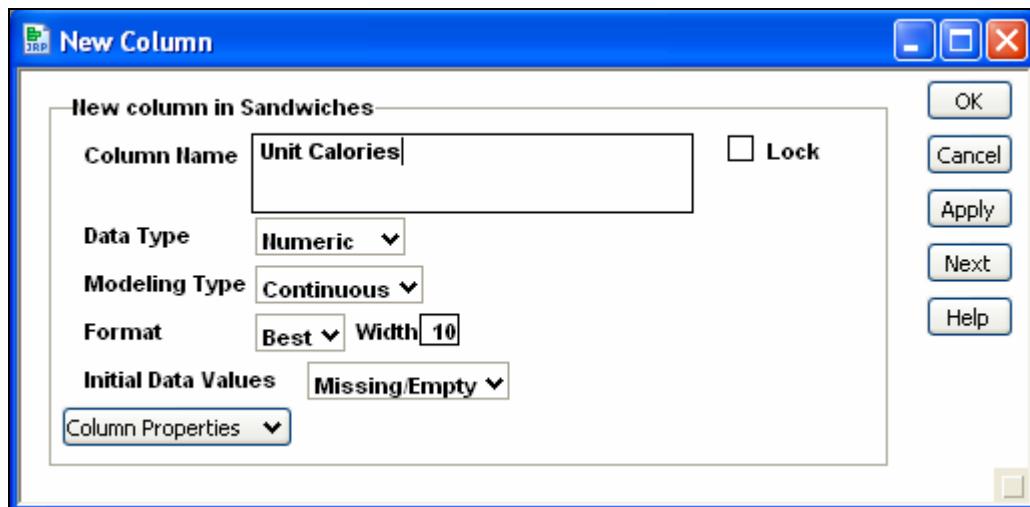
- There are a number of methods to add a new column. You can double-click to the right of the last column in the data grid.

Fiber	Sodium	Column 11	
5	1006	*	
2	794	*	

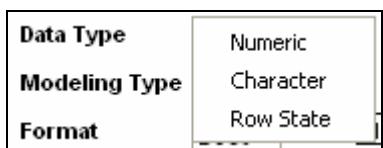
Or access the Columns menu using the Columns panel.

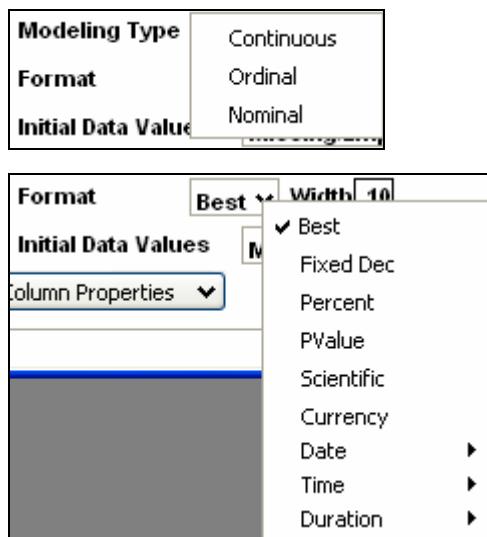


2. Type **Unit Calories** as the column name.

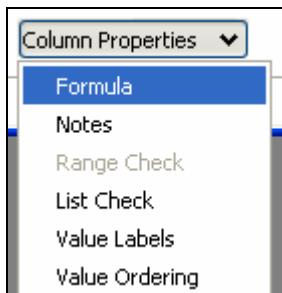


- The data type is Numeric, the modeling type is Continuous, and JMP enables the Best format with a column width of 10. To examine any of the other options in these fields, click on the check mark beside the automatic selection menu.

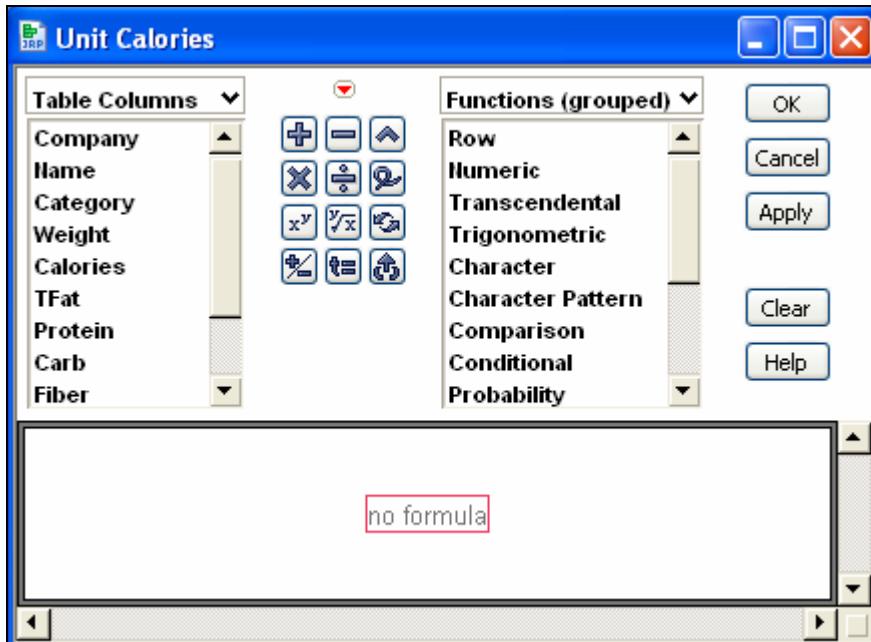




3. Select Column Properties \Rightarrow Formula.



A new window, the Formula Editor, opens. Use it to define and, if necessary, later change your formula.



The window is divided into an upper work panel that consists of three components: formula element browser, keypad, and function browser. The lower panel is the formula editing area.

4. Select **Calories** from the list of columns.



5. Select from the keypad and notice the changes in the formula editing area. A red box outlines the denominator of the equation. You now have two of the three components of your new column **Unit Calories**.



6. Select **Weight** from the list of columns.

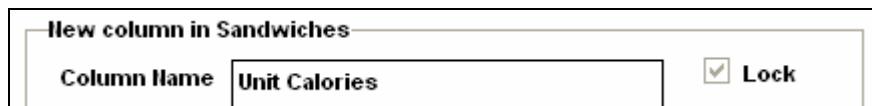


7. Select **Apply**.

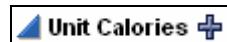
8. Select **OK** to close the Unit Calories window. Your new column has now been created. Remember that you can change the appearance of your column using the Format command.

Weight	Calories	TFat	Protein	Carb	Fiber	Sodium	Unit Calories
9	565	33	23	45	5	1006	62.7777778
6	223	5	13	34	2	794	37.1666667
10.7	518	23	30	48	.	1494	48.411215

- A column is automatically locked when you add a formula to a column. The values in a locked column cannot be edited. The check box next to the Lock in the Column Info window in this example is grayed out to indicate that the column cannot be unlocked.



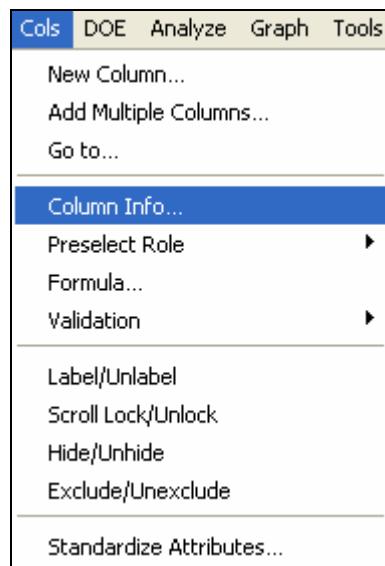
9. Select **OK** to close the New Column window. Notice the formula icon (a plus sign) next to **Unit Calories** in the Columns panel. This icon indicates that this column is defined by a formula.



10. Select **File** \Rightarrow **Save**.

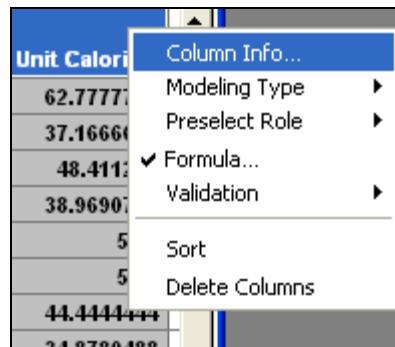
11. Select the **Unit Calories** column.

Suppose that you want to examine the formula of this column or change the name of this column, or change the format. The command for these actions can be found by selecting **Cols** \Rightarrow **Column Info....**



You can change anything about a column in this dialog box. You can change more than one column in the same way at the same time using the Standardize Attributes... command.

As a shortcut to Column Info..., you can also right-click at the top of any column to get an abbreviated version of the Columns menu.

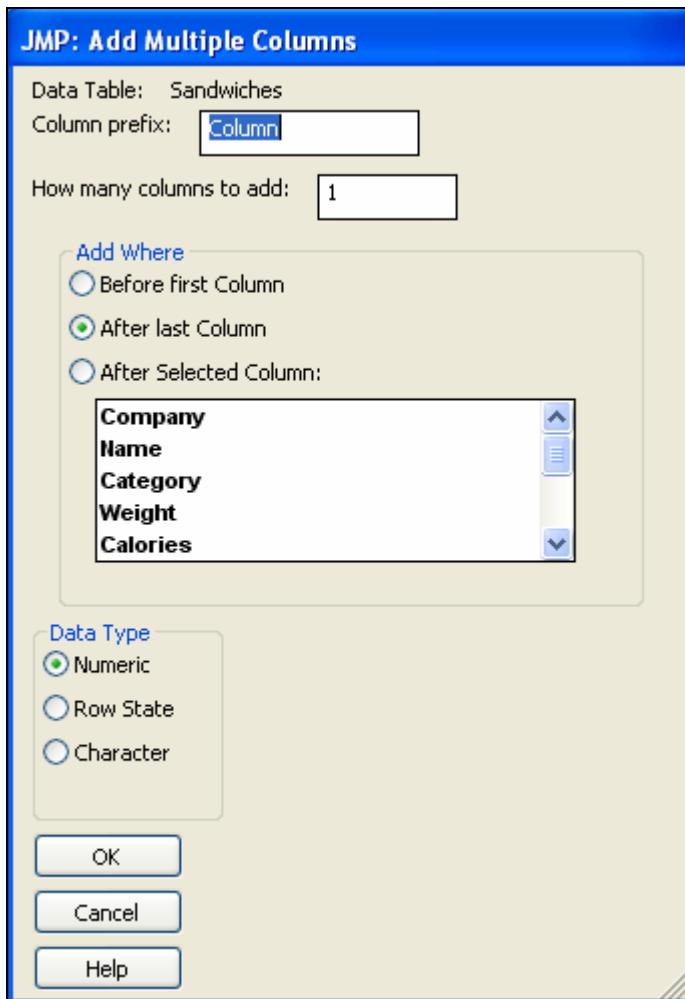


12. Close **Sandwiches.JMP**.

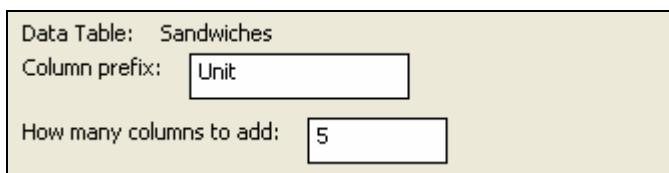
Creating the Desired Numeric Columns (Optional)

Recall that one of your goals was to take the six numeric columns in **Sandwiches.JMP**, **Calories** through **Sodium** inclusive, and divide them by **Weight**. There are five remaining columns to create.

1. Open **Sandwiches.JMP**.
2. Select **Cols** \Rightarrow **Add Multiple Columns....**



3. Type **Unit** as the column prefix.
4. Type **5** to indicate how many columns to add.



5. Leave the default **After last Column**.
6. Leave **Numeric** as the data type.

7. Select **OK**.

Five columns, named **Unit 1** through **Unit 5**, have been added to **Sandwiches.jmp**.

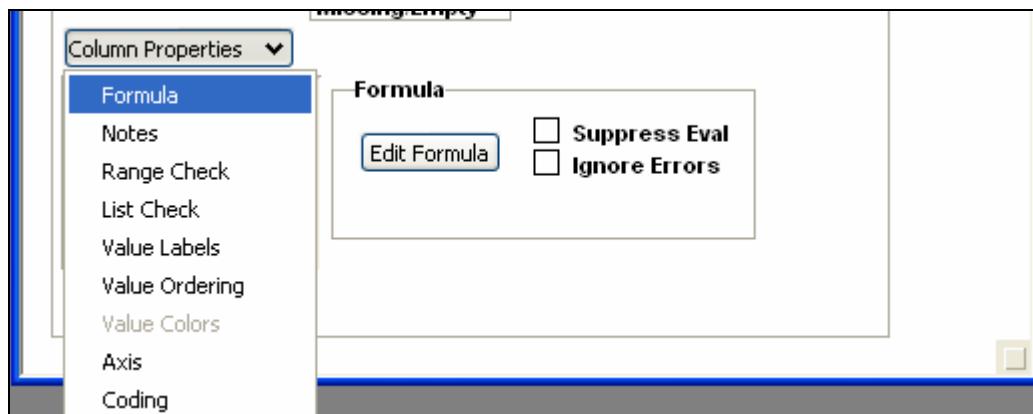
Sodium	Unit Calories	Unit 1	Unit 2	Unit 3	Unit 4	Unit 5
1006	62.7777778	*	*	*	*	*
794	37.1666667	*	*	*	*	*
1494	48.411215	*	*	*	*	*

Now add a formula for each of these columns that divides the original column by **Weight**.

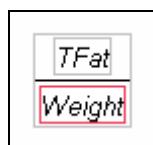
1. Select **Unit 1**.
2. Select **Cols** \Rightarrow **Column Info....**
3. Rename the column as **Unit TFat**.



4. Select **Column Properties** \Rightarrow **Formula** \Rightarrow **Edit Formula**.



5. Add the formula **TFat** \div **Weight**.

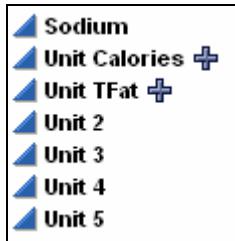


6. Select **Apply** \Rightarrow **OK**.
7. Select **OK**.

8. JMP displays the renamed, calculated column **Unit TFat** adjacent to the column **Unit Calories**.

Sodium	Unit Calories	Unit TFat	Unit 2	Unit 3	Unit 4	Unit 5
1006	62.7777778	3.66666667
794	37.1666667	0.83333333
1494	48.411215	2.14953271

9. Notice the formula icons in the Columns panel.



10. Select **File** \Rightarrow **Save**.

11. Close **Sandwiches.JMP**.

Each of the four remaining new Unit columns was generated in a similar manner and saved in a new JMP table named **Sandwiches Unit Based.JMP**.

Creating a New Categorical Column

As you might expect, it is also possible to create a categorical column. Notice that one of the values of **Category** is **Frozen**. Suppose that you want to create a new column that indicates whether the sandwich is fresh or frozen.

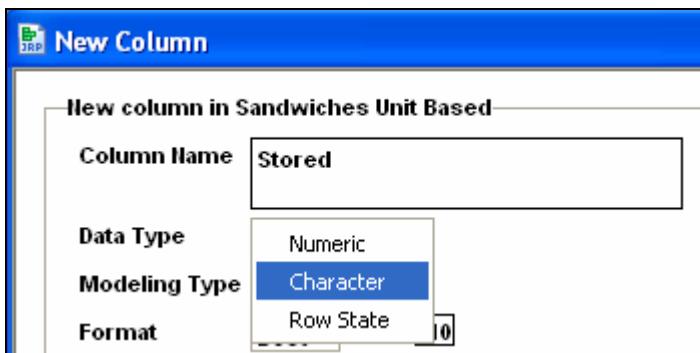
1. Select **File** \Rightarrow **Open**.
2. Select **Sandwiches Unit Based.JMP** \Rightarrow **Open**.

Category	Unit Calories	Unit TFat	Unit Protein	Unit Carb	Unit Fiber	Unit Sodium
Fish	62.7778	3.6667	2.5556	5.0000	0.5556	111.78
Frozen	37.1667	0.8333	2.1667	5.6667	0.3333	132.33
Turkey	48.4112	2.1495	2.8037	4.4860	.	139.63
Tuna	38.9691	1.2371	2.5773	4.5361	0.3093	105.57

Notice that the original columns have been hidden (indicated in the Columns panel), and there is a missing value for **Unit Fiber** (indicated with a period) because the value of **Fiber** is also missing. The “Unit” columns are displayed with the Fixed Decimal format with four decimal places.

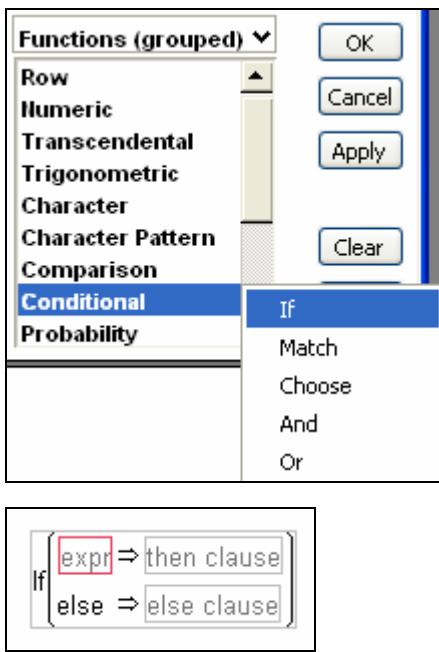
3. Select **Cols** \Rightarrow **New Column...**

4. Type **Stored** as the column name and select **Character** as the data type.

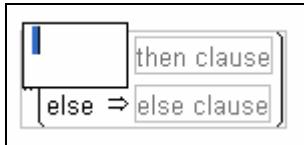


Modeling Type is automatically set as Nominal when the Data Type is Character.

5. Select **Column Properties** \Rightarrow **Formula**.
 6. From Functions (grouped), select **Conditional** \Rightarrow **If**, and notice the red box outlining expr.



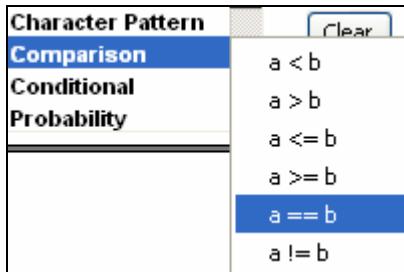
7. Double-click in the red box labeled expr.



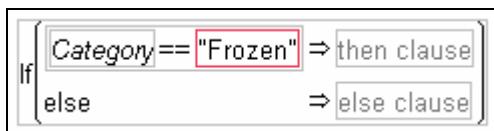
8. Select **Category** from the list of Table Columns.

9. From Functions (grouped), select **Comparison**.

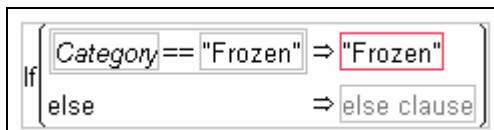
10. Select a == b.



11. Double-click in the red box, type "**Frozen**", and press ENTER



12. Double-click in the box labeled `then clause`, type "**Frozen**", and press ENTER.



13. Double-click in the box labeled `else clause`, type "**Fresh**", and press ENTER. The formula is now complete.

14. Select Apply ⇒ OK.

15. Select OK.

16. Move the newly created **Stored** column beside **Category**.

Name	Category	Stored	Unit Calories	Unit TFat	Unit Protein
Big Fish	Fish	Fresh	62.7777778	3.66666667	2.55555556
Spinach & Cheese Pocket	Frozen	Frozen	37.1666667	0.83333333	2.16666667
Turkey Club	Turkey	Fresh	48.411215	2.14953271	2.80373832

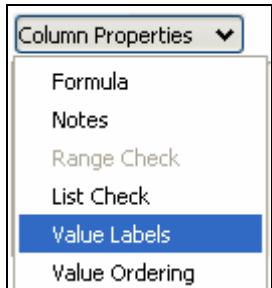
17. Select File ⇒ Save.

Using Value Labels

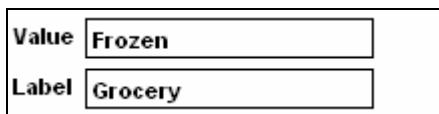
Value labels enable you to display a label instead of the original value in each instance that the value appears. When you assign value labels, the **labels** appear in the data table instead of the original values, but the original values are not lost. They can be displayed by double-clicking the label.

In **Sandwiches Unit Based.JMP**, rows where the value of **Stored** is **Frozen** represent sandwiches purchased from a grocery store, and those with a value of **Fresh** represent sandwiches purchased from a fast-food restaurant. To create a report comparing the number of sandwiches purchased from grocery stores to those from fast-food restaurants, assign the value labels **Grocery** and **Fast Food** to the stored column.

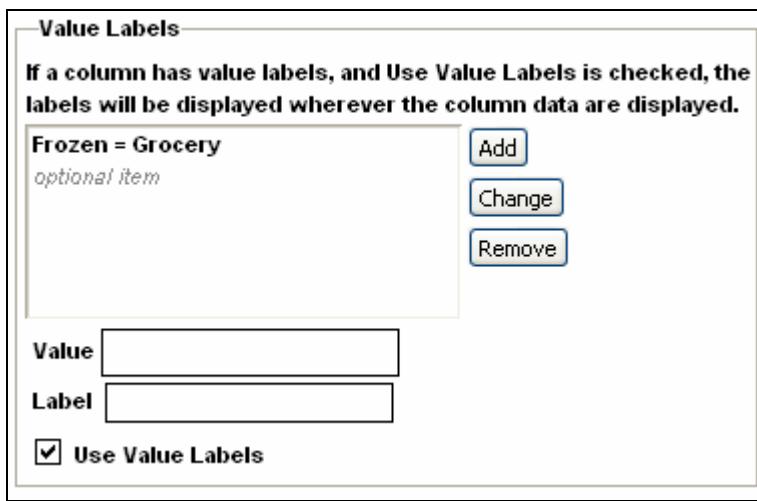
1. Select **Stored**.
2. Select **Cols** \Rightarrow **Column Info**.
3. Select **Column Properties** \Rightarrow **Value Labels**.



4. Type **Frozen** as the value.
5. Type **Grocery** as the label.



6. Select **Add**.



7. Type **Fresh** as the value.
8. Type **Fast Food** as the label.
9. Select **Add**.

10. Select Apply.

11. Select OK.

Name	Category	Stored	Unit Calories	Unit TFat
Big Fish	Fish	Fast Food	62.7777778	3.66666667*
Spinach & Cheese Pocket	Frozen	Grocery	37.1666667	0.83333333*
Turkey Club	Turkey	Fast Food	48.411215	2.14953271*
Tuna on Wheat	Tuna	Fast Food	38.9690722	1.2371134*
Baby Beef	Beef	Fast Food	56.5	2.66666667*
Ham & Cheese	Frozen	Grocery	56.5	2.66666667*

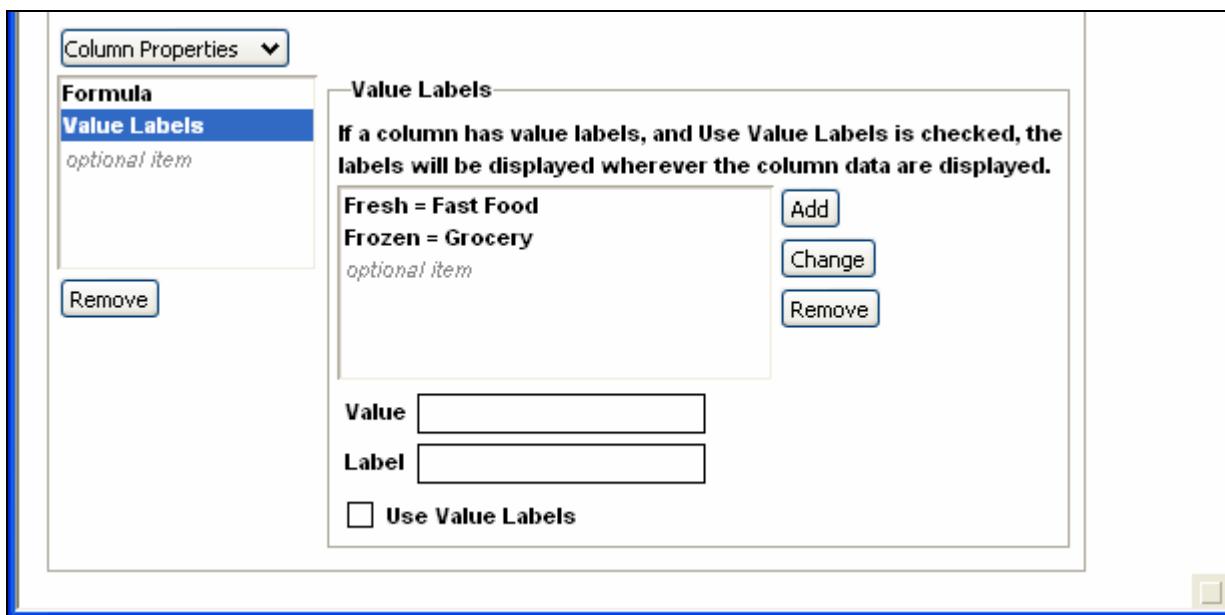
The value labels are now displayed in the **Stored** column. The asterisk beside **Stored** in the Columns panel indicates a property has been assigned to that column.

12. Click the asterisk next to **Stored**.

13. Select Value Labels.



14. Deselect Use Value Labels.



15. Select Apply.

16. Select OK.

The Value Labels property is still assigned to **Stored**, but the labels will not be used in reports.

2.3 Creating New Tables

Objectives

- Explore the numerous submenus beneath the Tables menu to create new JMP tables.

18

The previous sections have shown a number of ways to create new JMP tables. Now, explore how to create new JMP tables from existing JMP tables, using some of the Tables submenus.

The Tables Menu

- Summary
- Subset
- Sort
- Stack
- Split
- Transpose
- Concatenate
- Join
- Update
- Tabulate
- Missing Data Pattern



19

One method of exploring data is grouping like values together to determine what kind of relationship exists between the groups. In the example, summarize the unit data by **Company** and by **Stored** and **Company** to explore possible relationships. This type of analysis, in turn, might prompt you to choose a subset of the data for further investigation.

Several important and useful operations are available in the Tables menu to arrange the data differently. You can

- join two or more data tables together side by side. You can specify columns to match values and specify whether you want to keep or drop other columns.
- concatenate two or more columns together end to end. You can specify whether you want to drop columns or repeated values.
- sort the data table based on the values in one column (simple sort) or more than one column (nested sort). You can replace the original data table with the sorted data table or create a new data table for the sorted data and leave the original table unsorted.
- stack two or more columns into a single column. In addition, another column is created, which contains the name of the original column next to the data value.
- split a column into multiple columns based on the levels in another column. The levels become the column names of the splits.
- update one table with data from a second table. You can specify individually which columns to update in the current table.

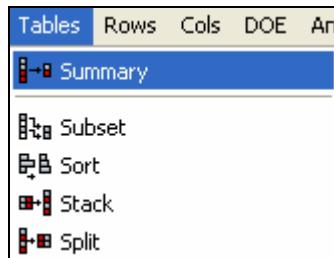


Creating New Tables

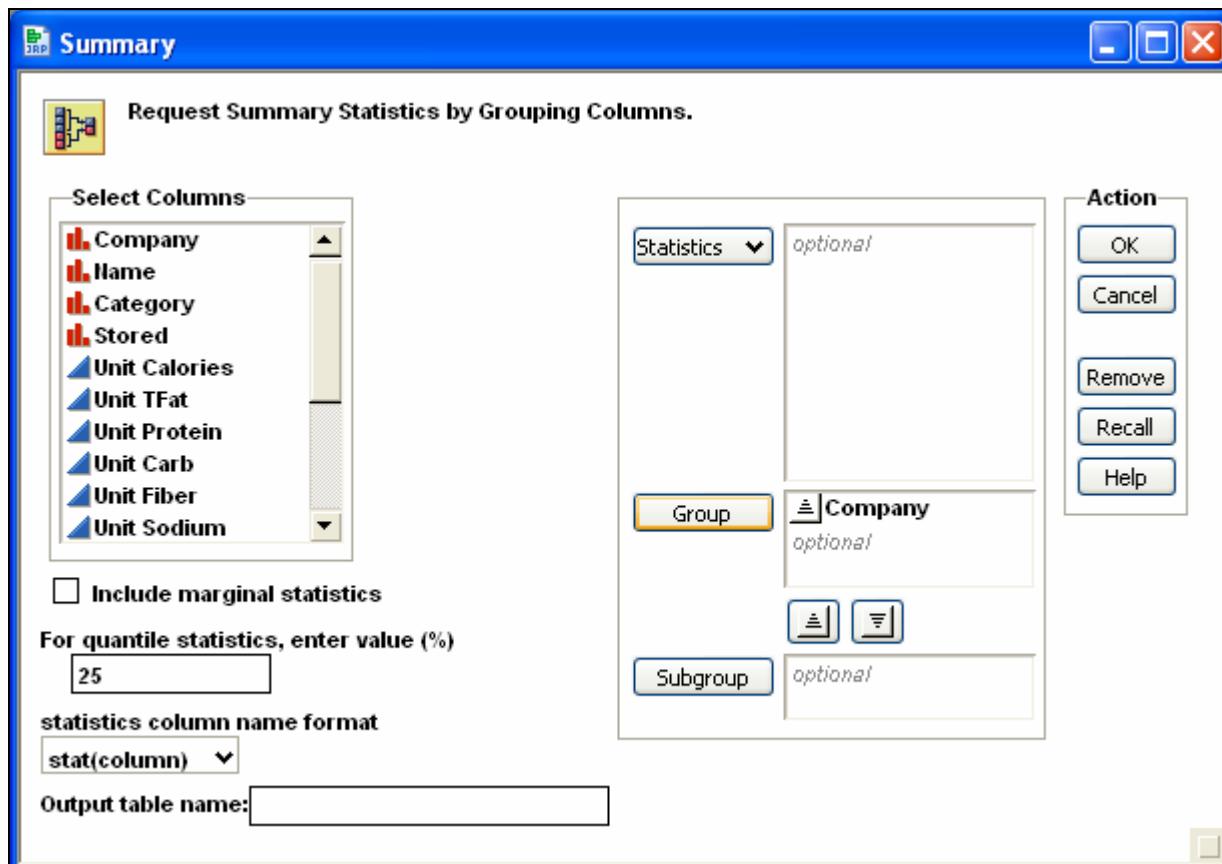
This demonstration illustrates the methods used to create new tables by summarizing tables, sorting and subsetting existing tables, and joining two tables by common column(s).

Examine potential relationships between companies. Make sure that **Sandwiches Unit Based.JMP** is your current window.

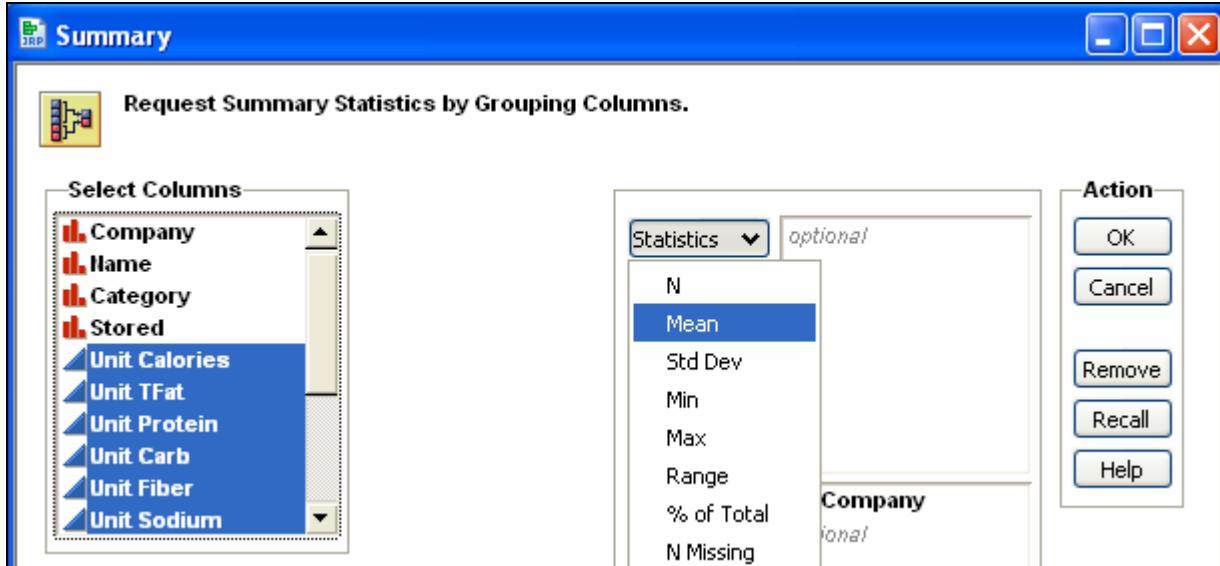
1. Select **Tables** \Rightarrow **Summary**.



2. JMP displays the following window. Select **Company** \Rightarrow **Group**.



3. Select Unit Calories through Unit Sodium.
4. Select Statistics \Rightarrow Mean.
5. Select OK.



JMP creates a new table with the title **Sandwiches Unit Based By (Company)**. Keep this table as the active window.

	Company	II Rows	Mean(Unit Calories)	Mean(Unit TFat)	Mean(Unit Protein)
1	A	15	50.0018363	2.21299898	2.37911314
2	B	2	64.619711	3.7373581	2.83178535
3	C	6	51.2396633	2.42745249	2.84656497
4	D	2	47.7174296	1.83802817	2.31426056
5	E	8	55.6622663	2.86497434	2.66839175
6	F	1	65.7333333	3.6	2.53333333
7	G	7	32.4886481	0.68309606	2.58933773
8	H	3	47.6008857	2.00636695	2.72018212
9	I	2	53.2959184	2.42176871	3.03061224
Z	J	3	41.5555556	1.27777778	3.11111111
Z	K	4	55.9166667	2.25	3.08333333
Z	L	11	41.774246	1.14244543	2.30558734

The table generated by the Summary command will not be saved unless you specifically perform a Save As... command. The naming convention is the original table name, the word By, and in parentheses the column or columns that were used as group(s).

Notice that **Company F** only has one product and **Company A** has the most products with 15. What **Company** has the highest and lowest average for **Unit Calories**?

You could search through this column manually, but using another of the JMP Table commands, Sort, makes your analysis easier.

1. Select **Tables** \Rightarrow **Sort**.
2. Select **Mean(Unit Calories)** \Rightarrow **By**.



- The default direction of the sort is ascending. (The triangle is pointing upward.) You could select the downward-pointing triangle to change the sort to descending.

3. Select **OK**.

Company	II Rows	Mean(Unit Calories)
G	7	32.4886481
J	3	41.5555556
L	11	41.774246
H	3	47.6008857
D	2	47.7174296
A	15	50.0018363
C	6	51.2396633
I	2	53.2959184
E	8	55.6622663
K	4	55.9166667
B	2	64.619711
F	1	65.7333333

- JMP can replace the values in the original table with the sorted values instead of opening a new window. To do this, check the box next to **Replace table** in the Sort dialog box.

Company G has the smallest average for Unit Calories, and the one sandwich of **Company F** has the highest value.

4. Close the summary data tables without saving.

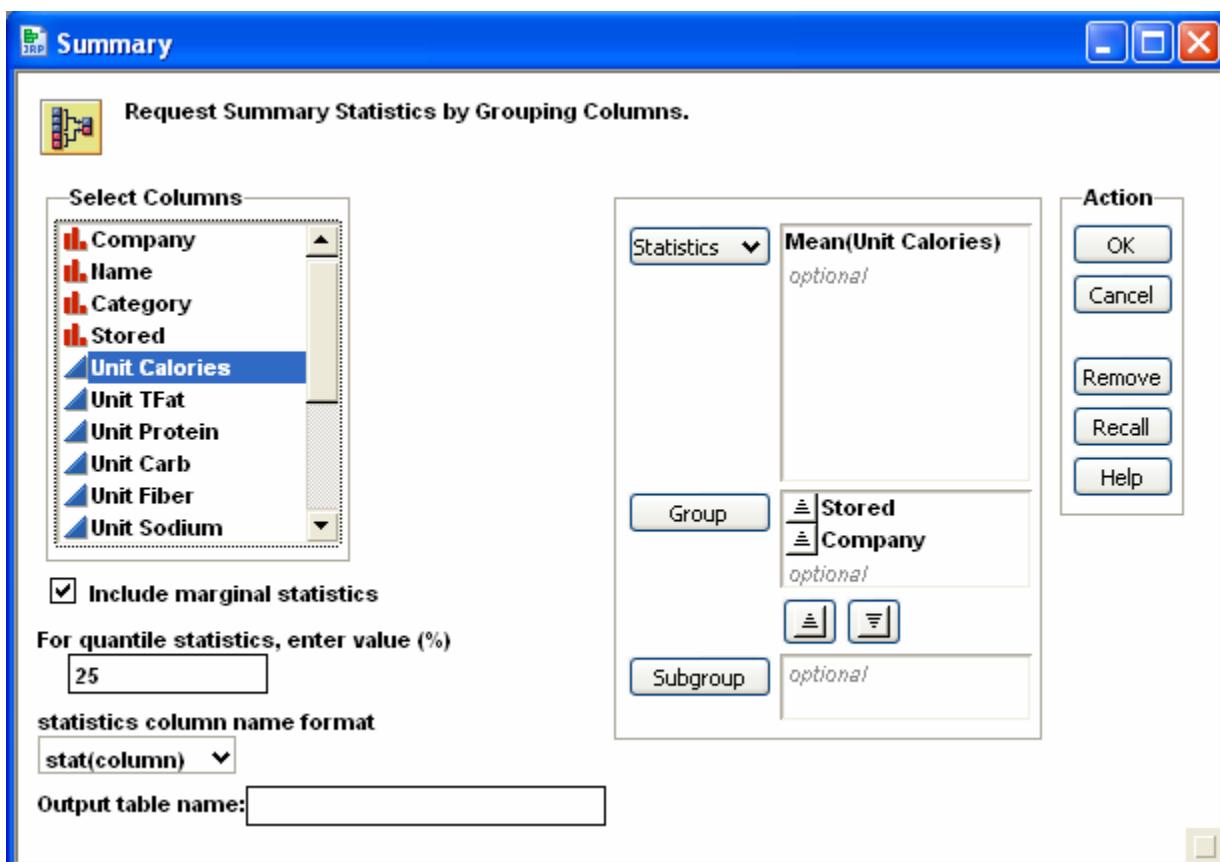
Adding a Second Level of Summary (Optional)

Most of the companies represented have more than one product. It might be of interest to determine which of the two storage techniques, Fresh and Frozen, has the smallest average value for **Unit Calories** and which company within the two techniques has the smallest average value for **Unit Calories**.

1. Select **Tables** \Rightarrow **Summary**.
2. Select **Stored** \Rightarrow **Group**.
3. Select **Company** \Rightarrow **Group**.



4. Select **Unit Calories** \Rightarrow **Statistics** \Rightarrow **Mean** and check **Include marginal statistics**. Select **OK**.



JMP creates a new table with the title **Sandwiches Unit Based By (Stored, Company)**. Examine this table to answer the questions that have been posed.

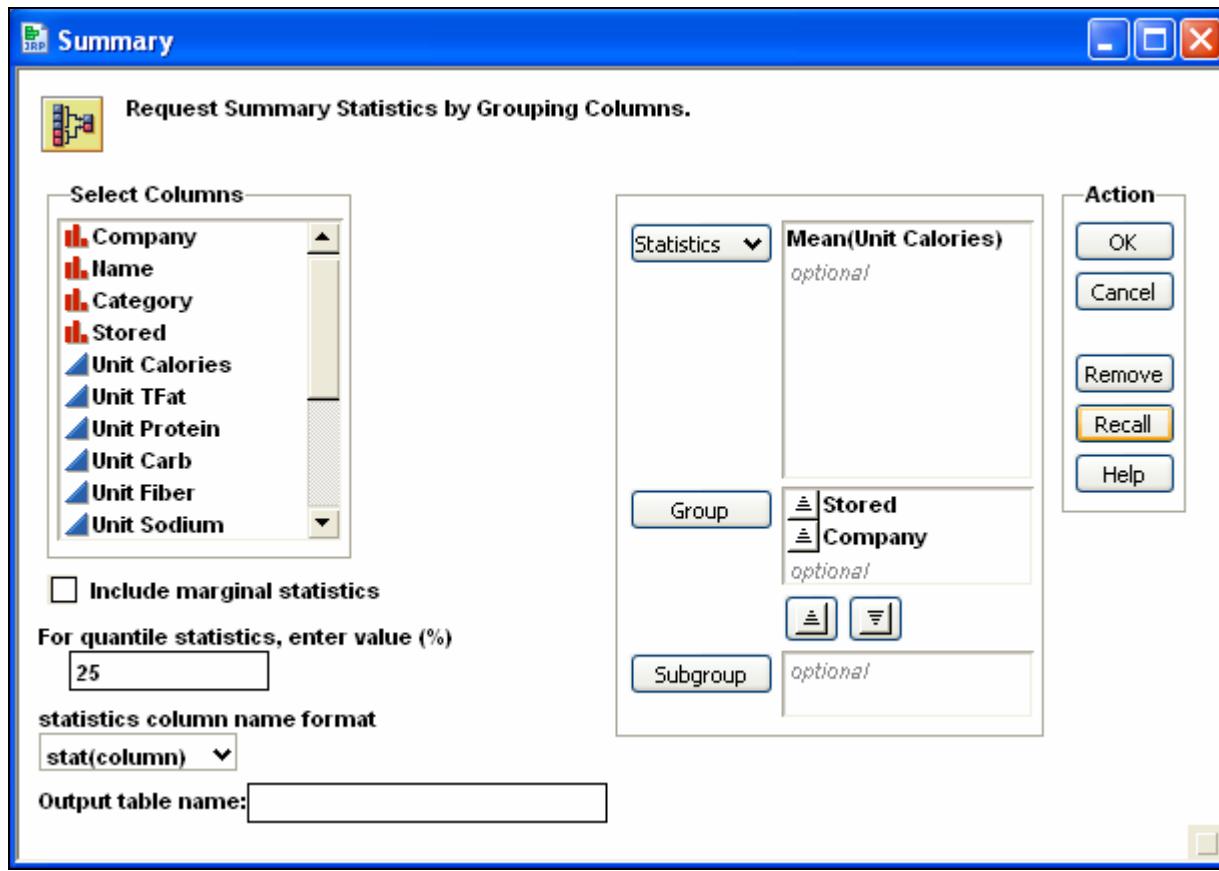
	Stored	Company	II Rows	Mean(Unit Calories)
1	Fresh	A	15	50.0018363
2	Fresh	B	2	64.619711
3	Fresh	C	6	51.2396633
4	Fresh	D	2	47.7174296
5	Fresh	E	8	55.6622663
6	Fresh	F	1	65.7333333
7	Fresh	G	7	32.4886481
8	Fresh	H	3	47.6008857
9	Fresh	I	2	53.2959184
10	Fresh		46	49.3475283
Z	11	Frozen J	3	41.5555556
Z	12	Frozen K	4	55.9166667
Z	13	Frozen L	11	41.774246
Z	14	Frozen	18	44.8805577
	15		64	48.0911928

Using the **Include marginal statistics** option has generated three new rows in this table, along with the 12 combinations of **Stored** and **Company**. Rows 10 and 14 provide the information to answer your first question: row 10 indicates that there are 46 fresh sandwiches with an average of 49.3475, which is greater than the 18 frozen sandwiches with an average of 44.8806. Row 15 states that there are 64 sandwiches, in total, with an average of 48.0911 unit calories.

5. Close **Sandwiches Unit Based Summarized (Stored, Company)**.

To answer the second part of the question, you must realize that the three marginal total rows in the current table are not needed. You could select and delete the three rows, but the following steps are easier.

1. Select **Tables** \Rightarrow **Summary**.
2. Select **Recall**.
3. Verify that **Include marginal statistics** does **not** have a check mark in the box beside it.
4. Select **OK**.

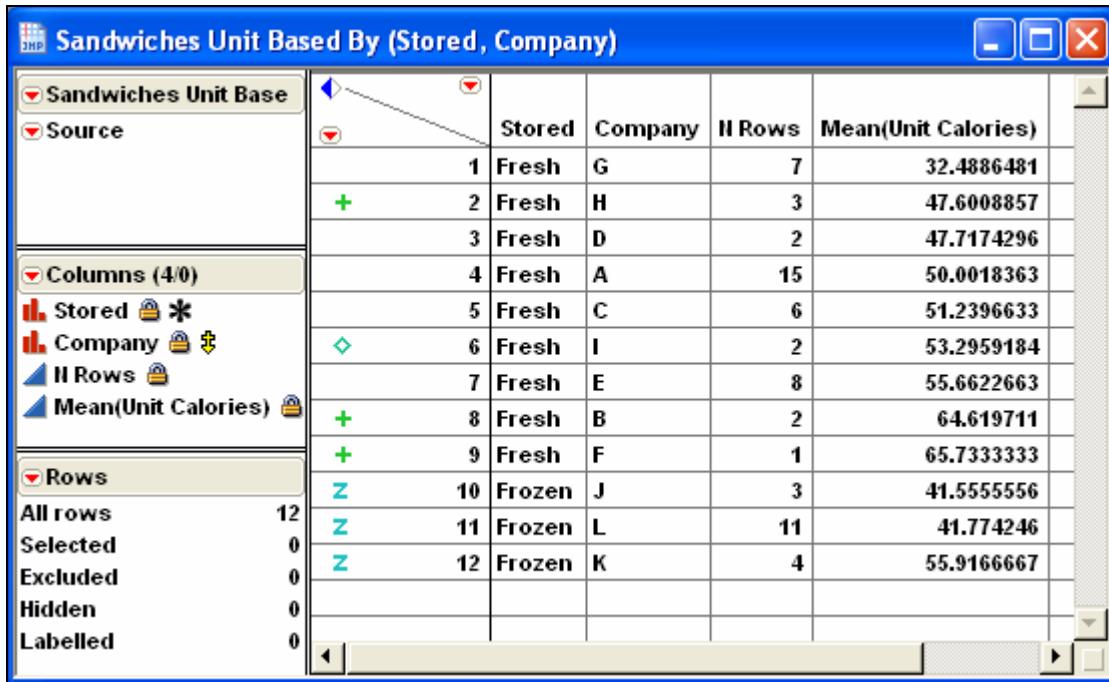


JMP creates a new table with the title **Sandwiches Unit Based By (Stored, Company) . JMP**.

Sandwiches Unit Based By (Stored, Company)					
Sandwiches Unit Base		Stored	Company	II Rows	Mean(Unit Calories)
1	Fresh	A		15	50.0018363
2	Fresh	B		2	64.619711
3	Fresh	C		6	51.2396633
4	Fresh	D		2	47.7174296
5	Fresh	E		8	55.6622663
6	Fresh	F		1	65.73333333
7	Fresh	G		7	32.4886481
8	Fresh	H		3	47.6008857
9	Fresh	I		2	53.2959184
10	Frozen	J		3	41.5555556
11	Frozen	K		4	55.9166667
12	Frozen	L		11	41.774246

Source: Columns (4/0): Stored, Company, II Rows, Mean(Unit Calories). Rows: All rows 12, Selected 0, Excluded 0, Hidden 0, Labelled 0.

5. Select **Tables** \Rightarrow **Sort**.
6. Select **Stored** and **Mean(Unit Calories)** \Rightarrow **By**.
7. Select **Replace Table**.
8. Select **OK**.



Company G has the smallest average for Fresh sandwiches (32.4886) and **Company J** has the smallest for Frozen sandwiches (41.5556).

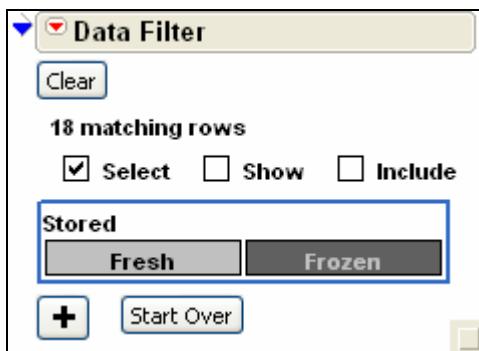
9. Close **Sandwiches Unit Based By (Stored, Company) .JMP**. Do not save.

Subsetting a Table

Suppose you would like the 18 rows identified as **Stored**=Frozen in a separate JMP table. You have learned how to select these rows, but not how to subset these chosen values into a new table.

1. Select **Rows** \Rightarrow **Data Filter**.
2. Select **Stored**.
3. Select **Add**.

4. Select **Frozen** and close the Data Filter window.

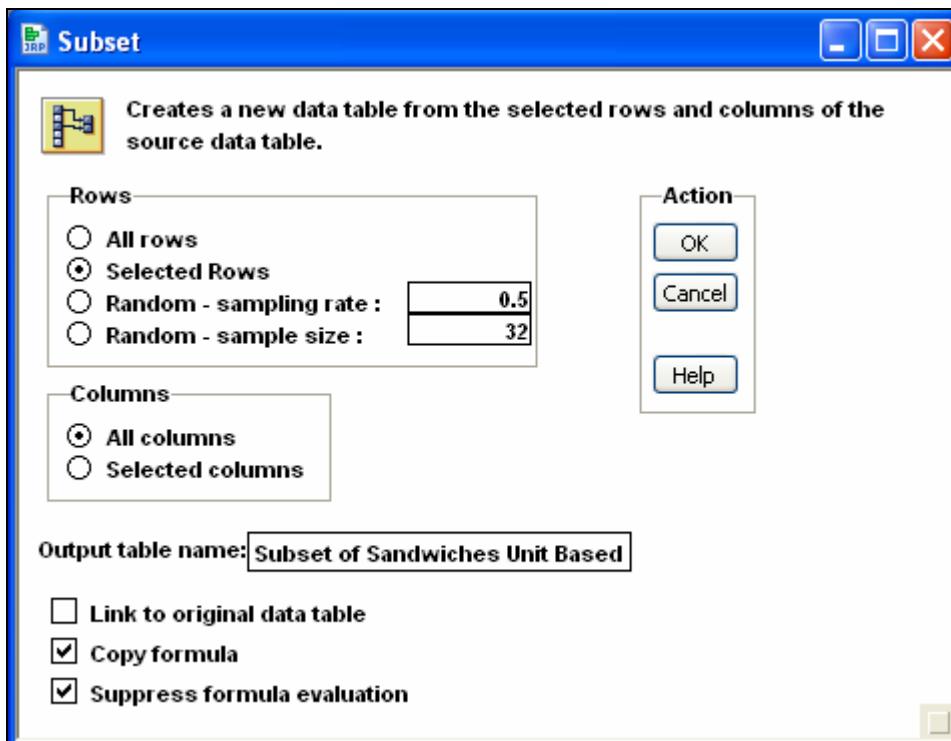


The rows for **Stored=Frozen** are selected.

Ham & Cheese	Frozen	Frozen	56.5	2.66666667
Grilled Chic	Chicken	Fresh	44.444444	2
BBQ Chic	Chicken	Fresh	34.8780488	0.6097561
Lite Reuben	Frozen	Frozen	39.0769231	0.76923077
Lite Ham & Cheese	Frozen	Frozen	47.2727273	1.63636364
Grilled Chic	Chicken	Fresh	37.125	1

As above, specific rows can also be selected by selecting **Rows** \Rightarrow **Row Selection** \Rightarrow **Select Where....**

5. Select **Tables** \Rightarrow **Subset**.



If you check the **Link to original data table** box, JMP highlights the corresponding rows in the original table and in all its plots and graphs.

6. Type **Frozen Sandwiches** as the output table name. Leave Selected Rows as the rows option.
7. Select **OK**.

	Company	Name	Category	Stored	Unit Calories	Unit TFat
Z	1 J	Spinach & Cheese Pocket	Frozen	Frozen	37.1666667	0.833333333
Z	2 K	Ham & Cheese	Frozen	Frozen	56.5	2.666666667
Z	3 L	Lite Reuben	Frozen	Frozen	39.0769231	0.76923077
Z	4 L	Lite Ham & Cheese	Frozen	Frozen	47.2727273	1.63636364
Z	5 K	Pepperoni Rollup	Frozen	Frozen	59	2.666666667
Z	6 K	Beef and Cheese	Frozen	Frozen	58.8333333	2.33333333
Z	7 L	Lite Chic Broccoli	Frozen	Frozen	41.3846154	1.23076923
Z	8 L	Lite Veggie Egg	Frozen	Frozen	46	1.8
Z	9 L	Lite Muffin	Frozen	Frozen	38.1818182	0.72727273
Z	10 K	Club Rollup	Frozen	Frozen	49.3333333	1.33333333
Z	11 L	Lite Turkey Mustard	Frozen	Frozen	40.1818182	0.72727273
Z	12 L	Lite Ham	Frozen	Frozen	39.2307692	1.38461538
Z	13 L	Lite Pizza	Frozen	Frozen	46.6153846	1.84615385
Z	14 L	Lite Omelet	Frozen	Frozen	38.4615385	0.38461538
Z	15 L	Lite Bagel	Frozen	Frozen	45.1111111	1.33333333
Z	16 J	Cheese Pocket	Frozen	Frozen	50	2.5
Z	17 L	Lite Chic	Frozen	Frozen	38	0.72727273
Z	18 J	Veggie Pocket Pie	Frozen	Frozen	37.5	0.5

8. Close **Frozen Sandwiches.JMP**. Do not save.
9. Deselect the selected rows and column from **Sandwiches Unit Based.JMP**.

Joining Two Tables

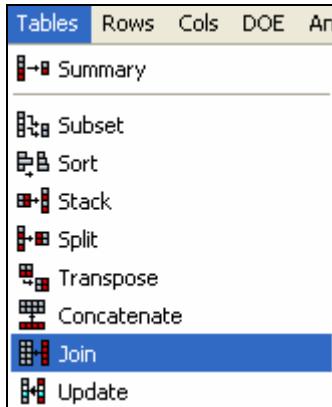
Often, data is compiled and stored in different tables. You might need to combine two different tables to create new columns for analysis. If the data in the second file is additional data, saved in the same data structure and format, you can use the Concatenate command, also found under the Tables menu.

Another possibility is that the two tables have common column(s) by which you could combine the two tables into one. For example, price-related data has been stored in **Sandwiches Multipliers.JMP**, which contains two columns: **Category** and **Unit Cost Multipliers**. Notice that the column **Category** is also a column in **Sandwiches Unit Based.JMP**.

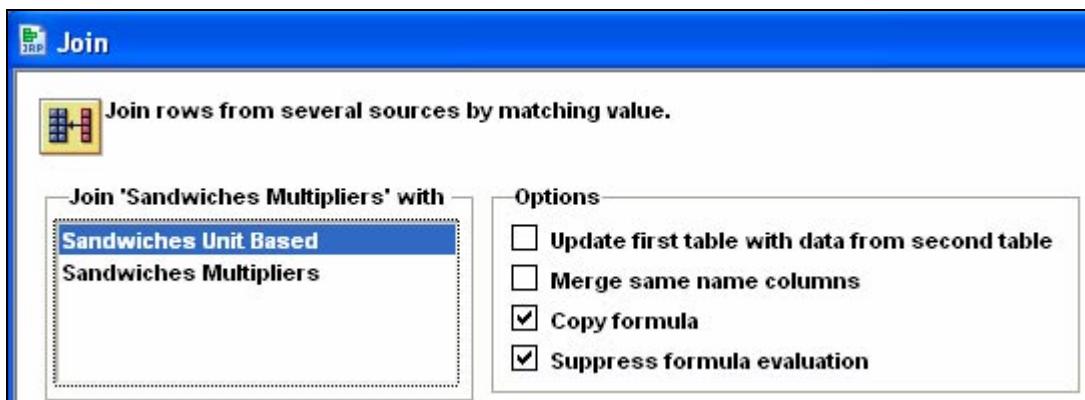
 When you join tables, the names of the common columns do not need to be the same. In this example, if **Category** were instead named **Sandwich Type** in **Sandwiches Multipliers.JMP**, JMP would still be able to match the data in the columns from the two tables.

In order to create a report that includes both the price-related information and the sandwich unit values, combine these two tables based on their shared column.

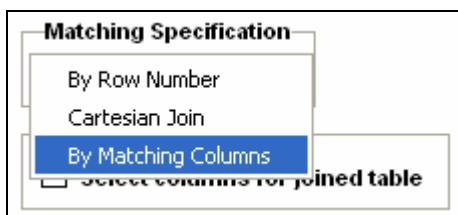
1. Select **File** \Rightarrow **Open**.
2. Select **Sandwiches Multipliers.JMP** \Rightarrow **Open**. Keep this as the active window.
3. Select **Tables** \Rightarrow **Join**.



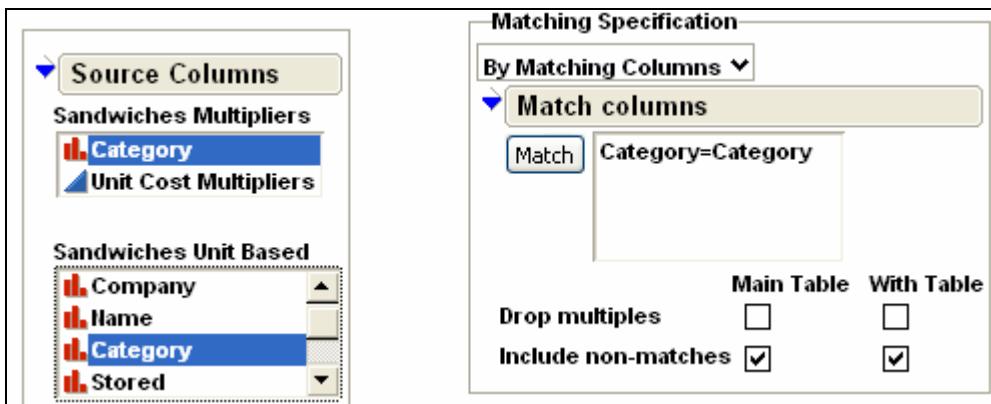
4. Select **Sandwiches Unit Based.JMP**.



5. Under Matching Specification, select **By Row Number** \Rightarrow **By Matching Columns**.



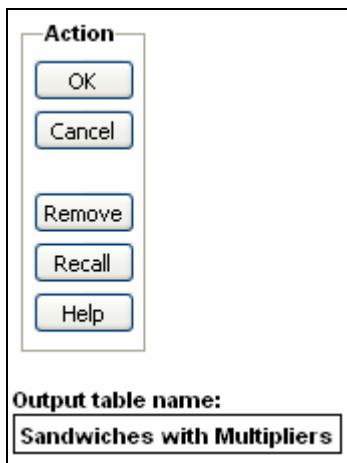
6. Under **Source Columns**, select **Category** from both tables and select **Match**.
7. Select **Include non-matches** for both Main Table and With Table.



- Checking Include non-matches provides good data management in the event that not all the values are matched in each table.
8. Select **Select columns for joined table**.
 9. Under **Source Columns**, from **Sandwiches Multipliers**, select **Category** and **Unit Cost Multipliers**. Select **Select**.
 10. From **Sandwiches Unit Based**, select **Company** and **Name**. Select **Select**.
 11. Select **Stored** through **Weight**. Select **Select**.



12. Type **Sandwiches with Multipliers** as the output table name and select **OK**.



JMP opens the new output table.

13. Scroll down the table to row 31.

11.0 Cols	Category	Unit Cost Multipliers	Company	Name	Stored	Unit Calories
28	Chicken	3.8014	H	Grilled Chic	Fresh	36.8831169
29	Chicken	3.8014	E	Supreme Chicken	Fresh	63.4
30	Chicken	3.8014	E	Chic Caesar	Fresh	53.6082474
31	Deli	7.5644				*
32	Fish	3.1131	C	Big Fish	Fresh	62.7777778
33	Fish	3.1131	I	Fried Platter	Fresh	58.1632653
34	Fish	3.1131	I	Fried Fish	Fresh	48.4285714
35	Fish	3.1131	A	Fish Filet	Fresh	58.6956522
36	Fish	3.1131	E	Supreme Fish	Fresh	59.8039216
37	Frozen	0.5677	J	Spinach & Cheese Pocket	Frozen	37.1666667
38	Frozen	0.5677	K	Ham & Cheese	Frozen	56.5

How do you interpret this row? It has missing values for **Company**, **Name**, **Stored**, **Unit Calories**, and so on. This row has missing values because **Category=Deli** does not have any matches in **Sandwiches Unit Based.JMP**. However, it is still included in the joined table because you checked **Include non-matches**. Upon further investigation, Row 57 has the same condition, where **Category=Kosher**.

14. Close **Sandwiches with Multipliers**. Do not save.

15. Close **Sandwiches Multipliers**.

Save **Sandwiches Unit Based**.

16. Select **File** \Rightarrow **Save**.

2.4 Using the Tabulate Option to Create Summary Tables

Objectives

- Discuss the features of tables created with the Tabulate option.
- Use the Tabulate option to create a summary table.

26

The Tabulate Option

The Tabulate option can be used to

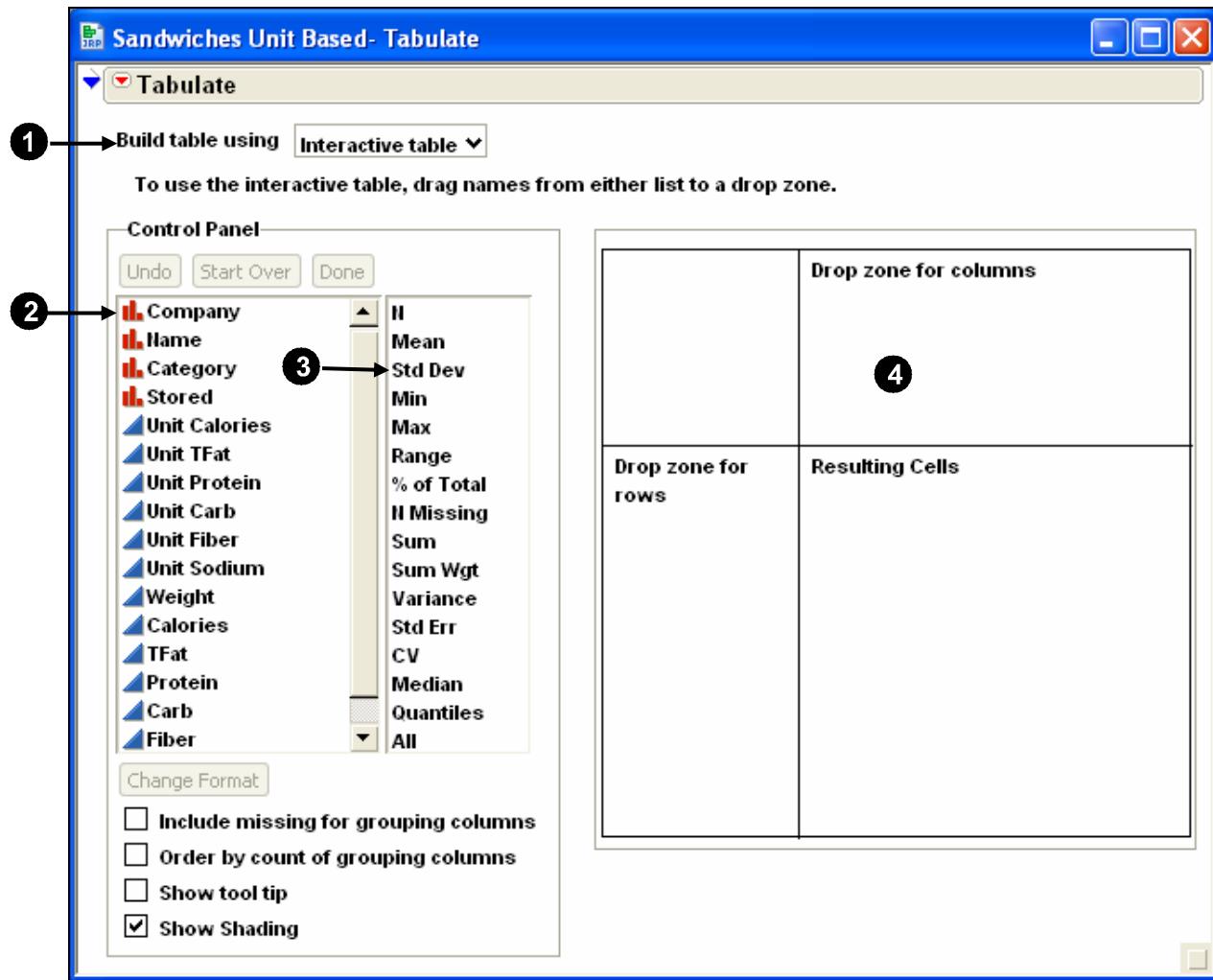
- summarize and organize data to easily explore patterns and relationships in data
- create tables interactively or through dialog boxes
- swap rows and columns
- change and add statistics
- save the table.

27



Using the Tabulate Option

- With **Sandwiches Unit Based** as the active table, select **Tables** \Rightarrow **Tabulate**.

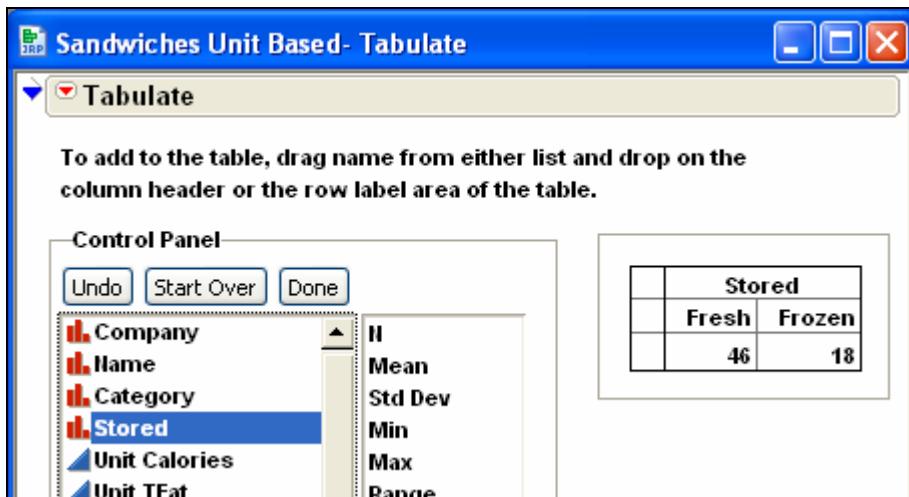


The pivot table can be built interactively or through a dialog box. The default method ① is Interactive table.

A list of columns in the active data table ② is displayed along with a list of statistics ③ that can be added to the table.

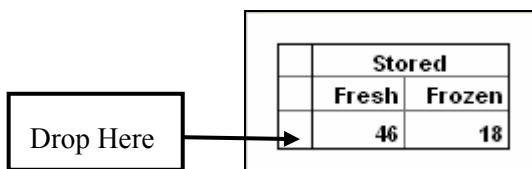
Elements are selected and dragged into the drop zone ④. After items are in the drop zone, they can be rearranged or deleted.

2. Select **Stored** and drag it to the upper-right drop zone (**Drop zone for columns**).



The default statistic in the row of the table is N, the number of rows in each level of **Stored**.

3. Select **Company**. Drag and drop it in the lower-left corner of the table.



Each level of **Company** is listed in a separate row of the table. Statistics are added next.

Company	Stored	
	Fresh	Frozen
A	15	0
B	2	0
C	6	0
D	2	0
E	8	0
F	1	0
G	7	0
H	3	0
I	2	0
J	0	3
K	0	4
L	0	11

4. Select **Unit Calories**. Drag it into the drop zone to the immediate right of **Company**, and select **Add Analysis Columns**.

Company	Stored	
	Fresh	Frozen
A	15	0
B		Add Grouping Columns
C		Add Analysis Columns
D	2	0
E	8	0

Company	Stored		
	Fresh	Frozen	
A	Unit Calories Sum	750.027545	.
B	Unit Calories Sum	129.239422	.
C	Unit Calories Sum	307.43798	.
D	Unit Calories Sum	95.4348592	.
E	Unit Calories Sum	445.29813	.
F	Unit Calories Sum	65.7333333	.
G	Unit Calories Sum	227.420537	.
H	Unit Calories Sum	142.802657	.
I	Unit Calories Sum	106.591837	.
J	Unit Calories Sum	.	124.666667
K	Unit Calories Sum	.	223.666667
L	Unit Calories Sum	.	459.516706

The sum of the calories for each **Company** within **Stored** is the default statistic. The sum can be replaced and additional statistics can be added to the table.

5. Select **Mean**. Drag and drop it on **Sum**. The sum is replaced by the mean in all rows of the table.

Company	Stored		
	Fresh	Frozen	
A	Unit Calories Mean	50.0018363	.
B	Unit Calories Mean	64.619711	.
C	Unit Calories Mean	51.2396633	.

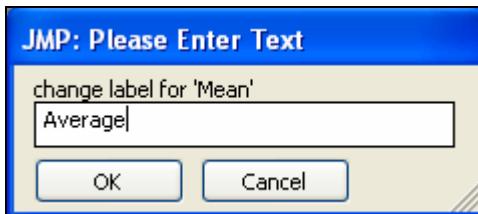
6. Select **Std Dev**. Drag and drop it on **Mean**.

Company	Stored		
	Fresh	Frozen	
A	Unit Calories Mean	50.0018363	.
	Std Dev	10.4248128	.
B	Unit Calories Mean	64.619711	.
	Std Dev	14.1589194	.
C	Unit Calories Mean	51.2396633	.
	Std Dev	9.53373668	.

7. Right-click **Mean** and select **Change Item Label**.

Company	Unit Calories	Stored		
		Fresh	Frozen	Range
A	Unit Calories	Mean		
B	Unit Calories	Mean		
C	Unit Calories	Mean	51.2396633	
		Std Dev	9.53373668	

8. Type **Average** in the text box.



9. Select **OK**.

10. Right-click **Std Dev** and select **Change Item Label**.

11. Type **Standard Deviation** in the text box.

12. Select **OK**.

Company	Unit Calories	Stored		
		Fresh	Frozen	Range
A	Unit Calories	Average	50.0018363	
		Standard Deviation	10.4248128	
B	Unit Calories	Average	64.619711	
		Standard Deviation	14.1589194	
C	Unit Calories	Average	51.2396633	
		Standard Deviation	9.53373668	

The interactivity of the tool makes it easy to swap columns and rows to see if other orientations present the information more clearly.

13. Select **Stored**. Drag and drop it on **Company**.

Company	Stored	Unit Calories	Average	50.0018363
A	Fresh	Mean	Standard Deviation	10.4248128
B	Fresh	Unit Calories	Average	64.619711
		Std Dev	Standard Deviation	14.1589194
C	Fresh	Unit Calories	Average	51.2396633
		Min	Standard Deviation	9.53373668
		Max		
		Range		

14. Select **Edit** \Rightarrow **Undo**.

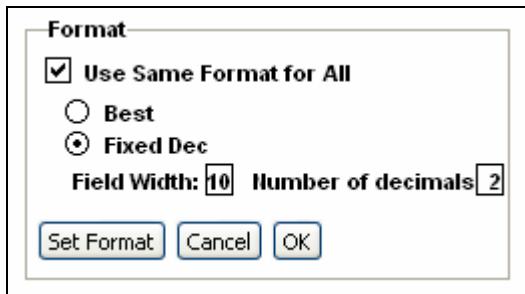
15. To change numeric cell formatting, select **Change Format** in the Control Panel.

Change Format

16. Select **Use Same Format for All**.



17. Select **Fixed Dec**.



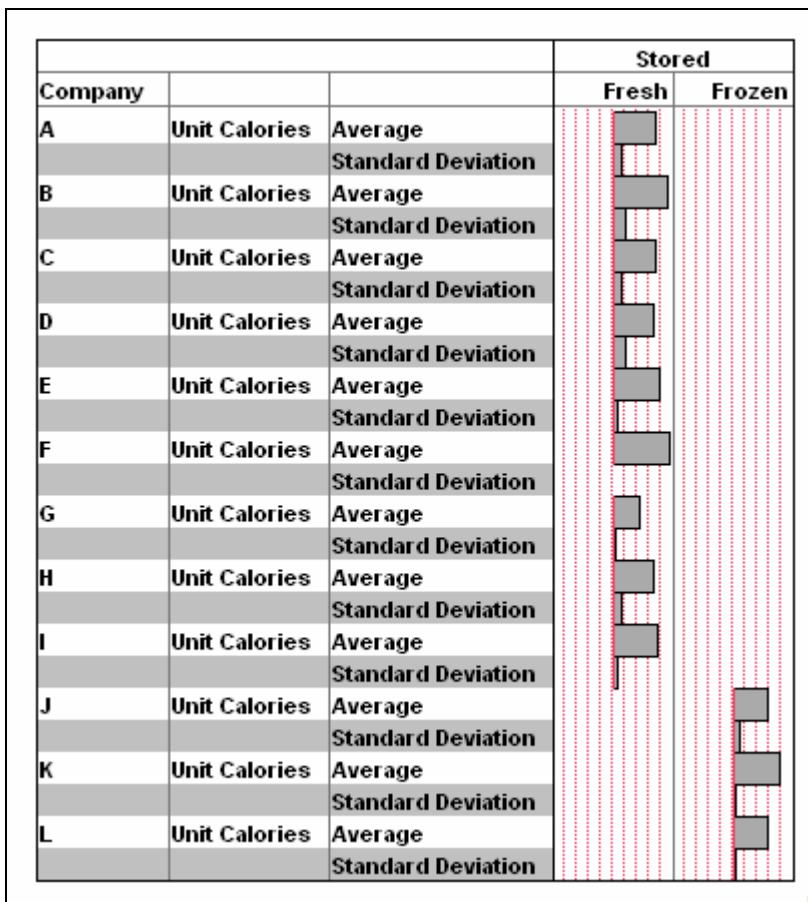
18. Select **Set Format**.

19. Select **OK**.

Company			Stored	
			Fresh	Frozen
A		Average	50.00	.
		Standard Deviation	10.42	.
B		Average	64.62	.
		Standard Deviation	14.16	.
C		Average	51.24	.
		Standard Deviation	9.53	.

The resulting table can also be accompanied by a chart.

20. Click the red triangle next to **Tabulate** and select **Show Chart**.



21. To complete the table and turn off the interactive building tools, select **Done**.



If you choose to save the table and chart, by default they are saved as a .JRP file. The data table must exist and be in the original location in order for the JRP script to work. Deleting or moving the table will break the JRP script. Also, moving a JRP file to another computer (for example, via an e-mail attachment) might not work as expected. To easily share tables and saved analyses, save the reports and the original data tables to a journal (see Chapter 4). The data tables can be recreated from the journal by right-clicking on the data table in the journal and selecting **Make Data Table**.

Close all windows.

22. Select **Window** ⇒ **Close All**.



Creating a Custom Menu (Optional)

Sales data is saved in **Sales.JMP**. Suppose this file is regularly updated with additional data, and each update requires a new summary table to be created. Create a custom menu item, saved to the Tables menu, which can be used to generate this summary table.

1. Select **File** \Rightarrow **Open**.
2. Select **Sales.JMP** \Rightarrow **Open**.
3. Select **Tables** \Rightarrow **Tabulate**.
4. Select **Product** and drag it to the upper-right drop zone (**Drop zone for columns**).

	Product	
	A	B
	8	7

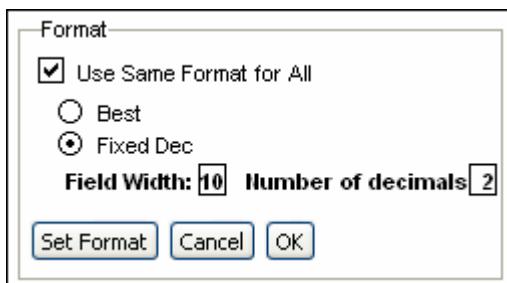
5. Select **Sales**. Drag and drop it in the lower-left corner of the table, and select **Add Analysis Columns**.
6. Select **Mean**. Drag and drop it on **Sum**.
7. Select **Sum**. Drag and drop it on **Mean**.

		Product	
		A	B
Sales	Sum	16400	12650
	Mean	2050	1807.14286

8. Select **Division**. Drag it into the zone of **Sales**, but underneath **Sales**, and select **Add Grouping Columns**.

		Product	
		A	B
Sales			
		Add Grouping Columns	Add Analysis Columns

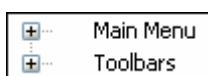
9. Select **Change Format**.
10. Select **Use Same Format for All**.
11. Select **Fixed Dec**.
12. Select **Set Format** \Rightarrow **OK**.



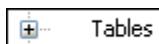
13. Select **Done**.
14. Click the red triangle next to **Tabulate** and select **Script** \Rightarrow **Copy Script**.
15. Select **Edit** \Rightarrow **Customize** \Rightarrow **Menus and Toolbars**.

The Menus window appears docked on the left side of the JMP session.

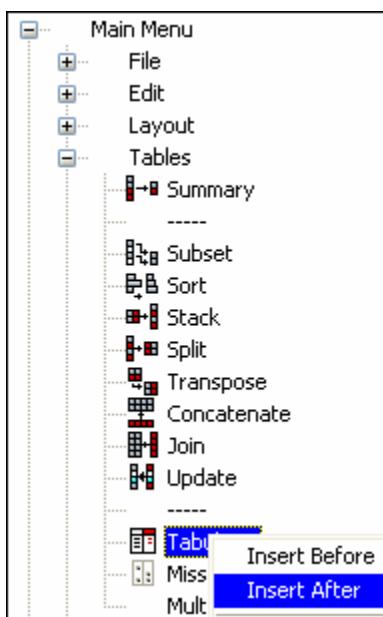
16. Click the plus sign to expand Main Menu.



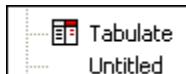
17. Click the plus sign to expand Tables.



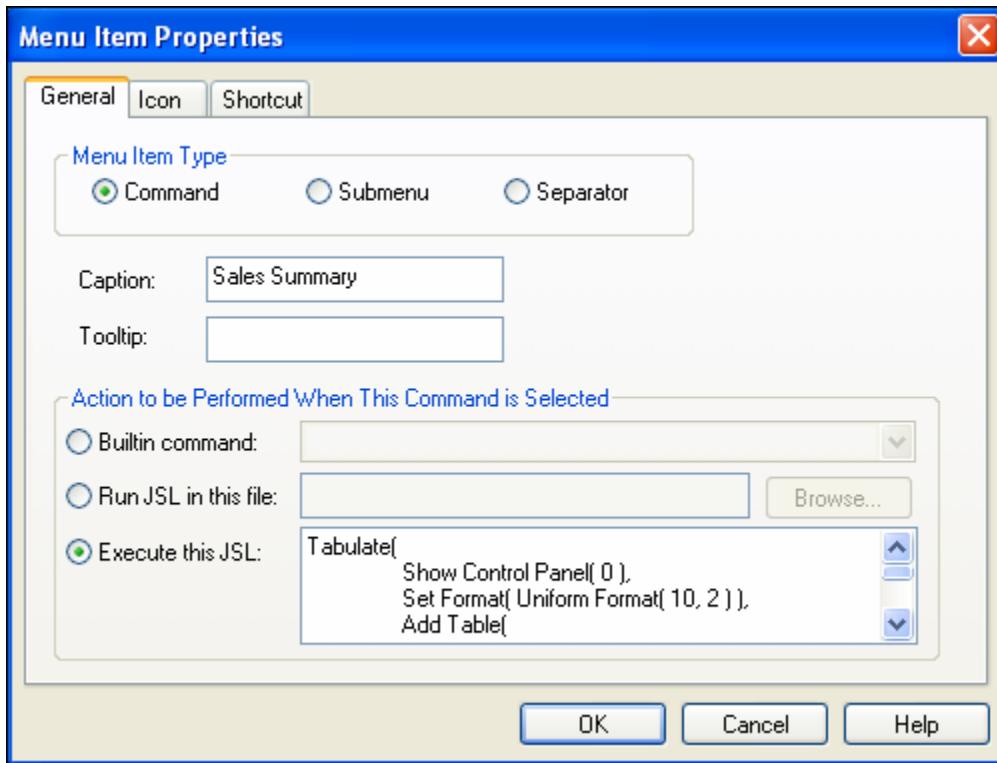
18. Right-click **Tabulate** and select **Insert After**.



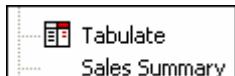
JMP creates a new, untitled custom menu option.



19. Right-click **Untitled** and select **Properties**.
20. In the Execute this JSL field, delete `print("Not implemented.");`
21. Right-click in the Execute this JSL field and select **Paste**.
22. Type **Sales Summary** as the caption.
23. Select **OK**.



The Sales Summary menu item appears under Tabulate.



24. Select **Apply** in the Menu window.

The Sales Summary menu item now appears in the Tables menu.

25. Select **Window** \Rightarrow **Sales**.

Add two additional rows of data.

26. Double-click under row 15 in the data table.

27. Type **B**, **1**, and **1100** and press ENTER.

28. Double-click under row 16 in the data table.

29. Type **A**, **1**, and **3300** and press ENTER.

16	B	1	1100
17	A	1	3300

30. Select **Tables** \Rightarrow **Sales Summary**.

JMP produces the same summary table that you previously produced, but updates the values to reflect the additional data.

			Product	
	Division		A	B
Sales	1	Sum	7600.00	6850.00
		Mean	1900.00	1370.00
	2	Sum	12100.00	6900.00
		Mean	2420.00	2300.00

31. Close all open windows. Do not save.

2.5 Chapter Summary

Using the Columns and Rows menus, you can perform a number of actions to manipulate, change, add, or delete columns and rows from the data grid. Three actions, Label/Unlabel, Hide/Unhide, and Exclude/Unexclude, are simple switches, indicating whether the attribute is present or not.

Columns can be moved, reordered, and created. Values in the columns can be limited using the Validation command, and either List or Range Checking, based on the type of column. Columns are created using the Formula Editor and can use stored constants, current columns, or values stored in Table variables.

Rows can be given specific colors and markers to enhance any graphical presentations of the data. Rows can be Selected, Excluded, Hidden, or Labeled.

JMP provides a platform to summarize tables by specific columns, which creates a new table.

New tables can be created from existing tables using the subset, sort, and join operations as well as stack, split, and transpose. Summary tables can be created using the Tabulate option.

Chapter 3 Graphical Data Exploration

3.1 Exploring Relationships Between Continuous Columns	3-3
3.2 Examining Relationships Between Attribute Data Columns	3-23
3.3 Exploratory Analysis Using Recursive Partitioning.....	3-28
3.4 Recursive Partitioning Details (Optional).....	3-38
3.5 Chapter Summary.....	3-44

3.1 Exploring Relationships Between Continuous Columns

Objectives

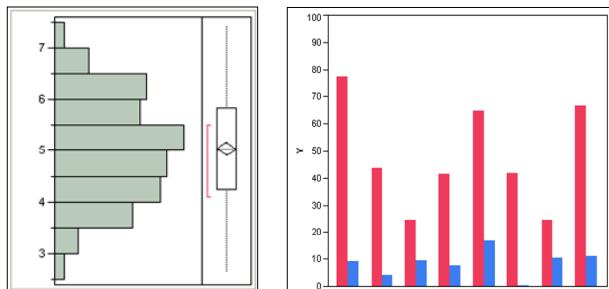
- Create, examine, and interact with JMP graphs to begin data discovery.
- Review options and commands to improve images for user understanding.
- Save scripts to a data table.

3

Pictures make your data come to life. The best way to begin to get answers to your questions using the data is through pictures that JMP calls graphs.

There are many different kinds of graphs, some of which are very specialized. Focus on three general kinds of graphs that will give you a good start to understanding your data. Generate, examine, and learn to modify their appearance.

Graphing Continuous Values

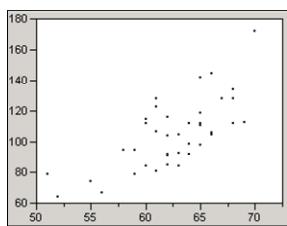


4

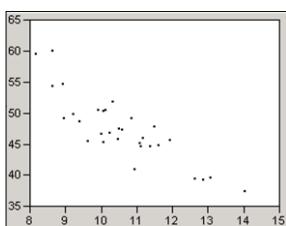
You need to know about the location and the variation of your data, and any unusual data values.

- Continuous data can be forced into bins (categories) and plotted in special bar charts called *histograms*. These graphs display the relative amounts or frequencies of certain values or ranges of values.
- *Box plots* display graphics to represent the data. Modifications of the basic box plot (whiskers) identify outlying or far outlying data or assess the shape.
- *Bar charts* are used to make graphical comparisons between groups.

Trends Between Values



Scatterplots show how your variables are associated.



5

You want to know whether the change in one of your variables is associated with changes in another variable.

Scatterplots display the data in two coordinates formed by the values of each variable.

Querying and Modifying Graphs

- Graphics tools
 - grabber, brush, lasso, magnifier, cross-hairs
- Change row states
 - colors, markers
- Modify axes
 - formats, line styles



6

The graphics tools work in any graph to accomplish specific tasks.

- The *Grabber tool* is used to grab the axes and interactively scroll vertically or horizontally. You can also freely rotate the three-dimensional Spinning Plot.
- The *Brush tool* is used to dynamically select the data. As you move the brush over data, you select or deselect it.
- The *Lasso tool* selects data in a non-rectangular shape. You draw the outline of the data with the lasso to select it.
- The *Magnifier tool* zooms in closer to the data you click. You can drag a region to interactively change the axis scales. Hold down the ALT key and click to return to the original scales.
- The *Crosshairs tool* displays the coordinates when you click and hold the mouse button. You can read exact values from either axis.

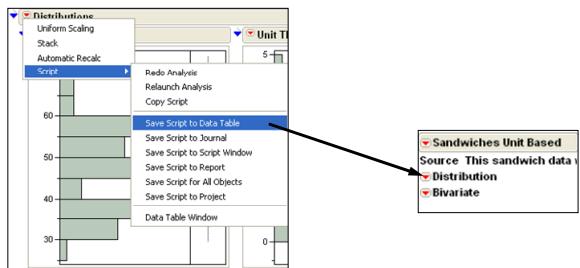
Using JSL Scripts to Save Results

- JMP scripting language (JSL) can be used to reproduce results without performing the analysis using menus.
- JSL scripts can be saved to your data table, keeping a record of the analyses you have performed.
- JSL scripts attached to a data table enable you to efficiently share your analyses with others.

7

JSL scripts make it possible to record, repeat, program, automate, and customize tasks.

Scripts in the Table Panel



8

Take advantage of JSL scripts stored in the Table panel to generate analyses without having to know how they were created.

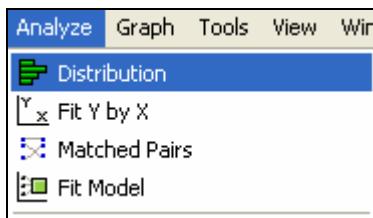


Generate and Explore Graphs of Continuous Columns

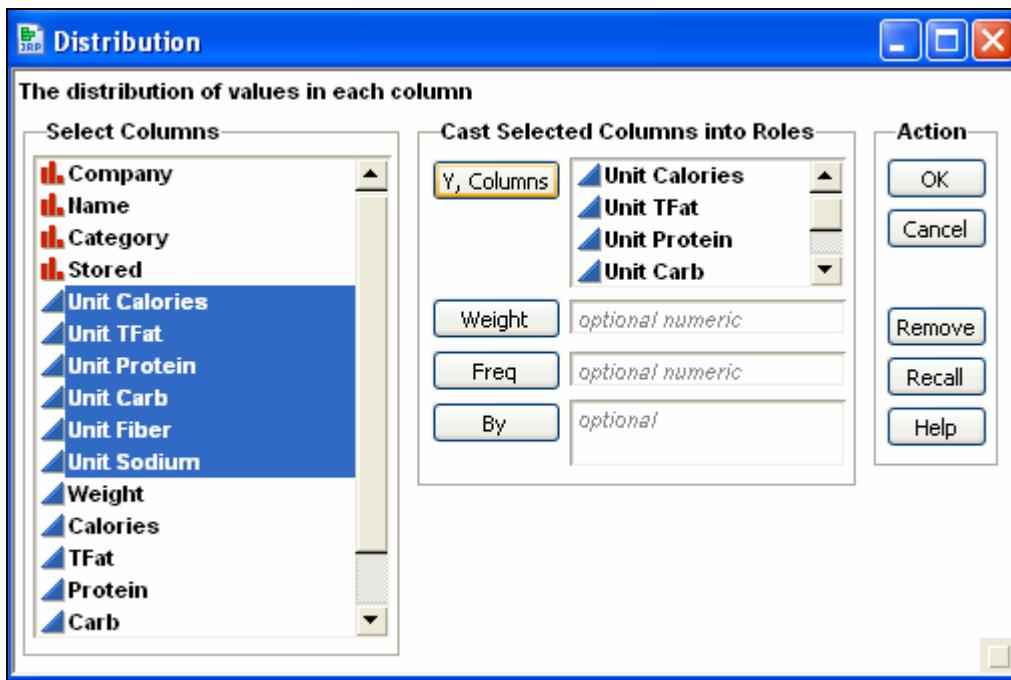
Relationships between continuous columns can be explored using histograms and scatterplots.

Examining Histograms

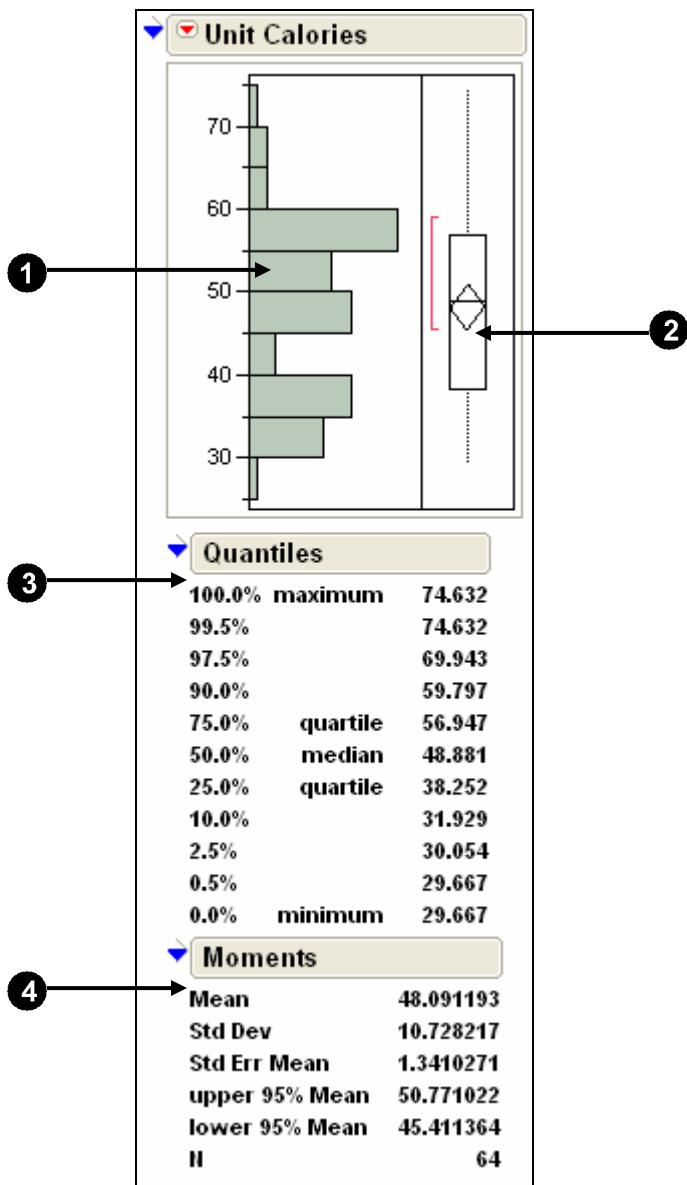
1. Select File \Rightarrow Open.
2. Select Sandwiches Unit Based.JMP \Rightarrow Open.
3. Select Analyze \Rightarrow Distribution. The Distribution dialog window opens.



4. Select Unit Calories through Unit Sodium \Rightarrow Y, Columns.



5. Select OK.



The portion of the report corresponding to

- ① is a histogram of the values for **Unit Calories**. Each bar in a histogram shows the relative percentage of observations that fall within a specified range of values.
- ② is an outlier box plot, which plots quartile values and highlights points suspected of being outliers.
- ③ is the Quantiles report, where the median and other selected percentiles are reported. (Although JMP does not list the range, this table reports the minimum and maximum so you can calculate the range.)
- ④ is the Moments report, where you can obtain the mean and standard deviation. The values in the Quantiles and Moments reports are beyond the scope of this course.

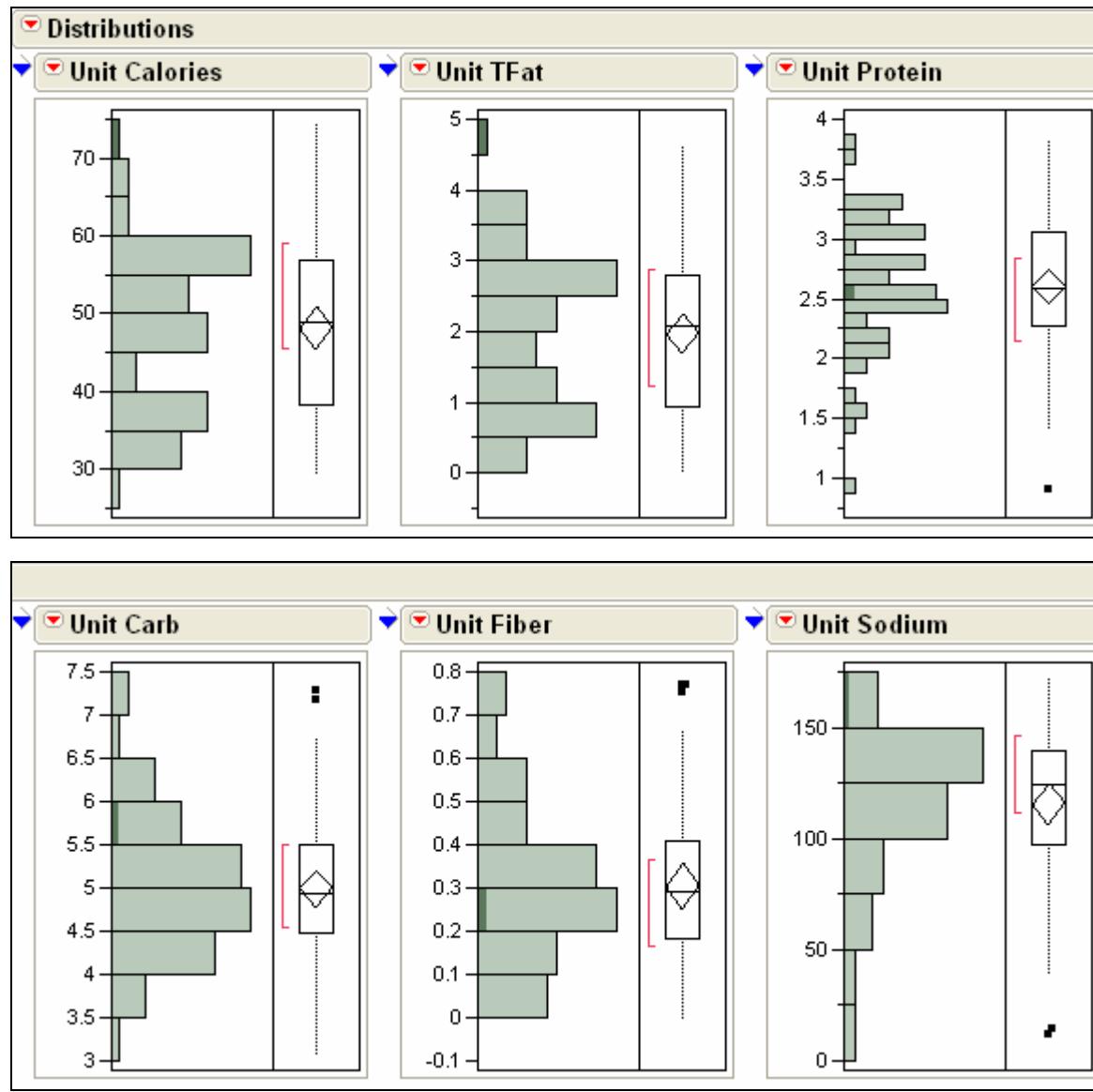
Scroll bars provide the ability to scroll horizontally and vertically. If you scroll to the right, you will find this analysis window is displaying statistics and graphs for the continuous columns that begin with **Unit**.

The blue triangles (expand/collapse icons) that are pointing down indicate that the results for the specific column can be collapsed. To collapse the output, click on the blue triangle. Notice that the blue triangle points to the right. To expand the collapsed output, click the blue triangle a second time.

What kind of statements can you make regarding the distributions of these continuous variables? Again, another piece has been added to the data exploration puzzle.

6. Click the top bar in the histogram for **Unit Calories**.

JMP simultaneously selects and highlights the corresponding rows in the data grid and in any other open analysis windows. The Rows panel in the data table indicates that only one row was selected.



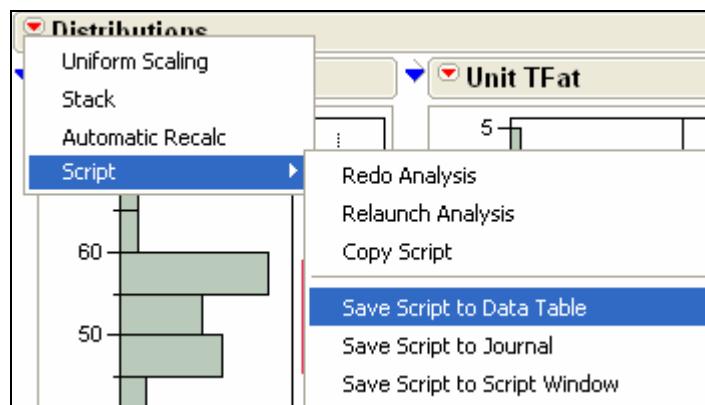
- Continue to click down the bars in the **Unit Calories** distribution to examine the relationship between **Unit Calories** and the remaining unit columns.

When the value of **Unit Calories** is high, so are the values of **Unit TFat** and **Unit Sodium**, but the values of **Unit Protein**, **Unit Carb**, and **Unit Fiber** are all approximately in the middle of their respective values.

-  To select multiple bars in the Distribution results, hold down the SHIFT key and click your other choices.

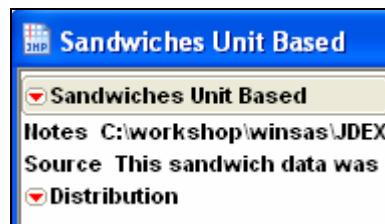
Save the script to recreate this analysis to the data table.

- Click the red triangle next to **Distributions** and select Script \Rightarrow Save Script to Data Table.



- Select Window \Rightarrow Sandwiches Unit Based.

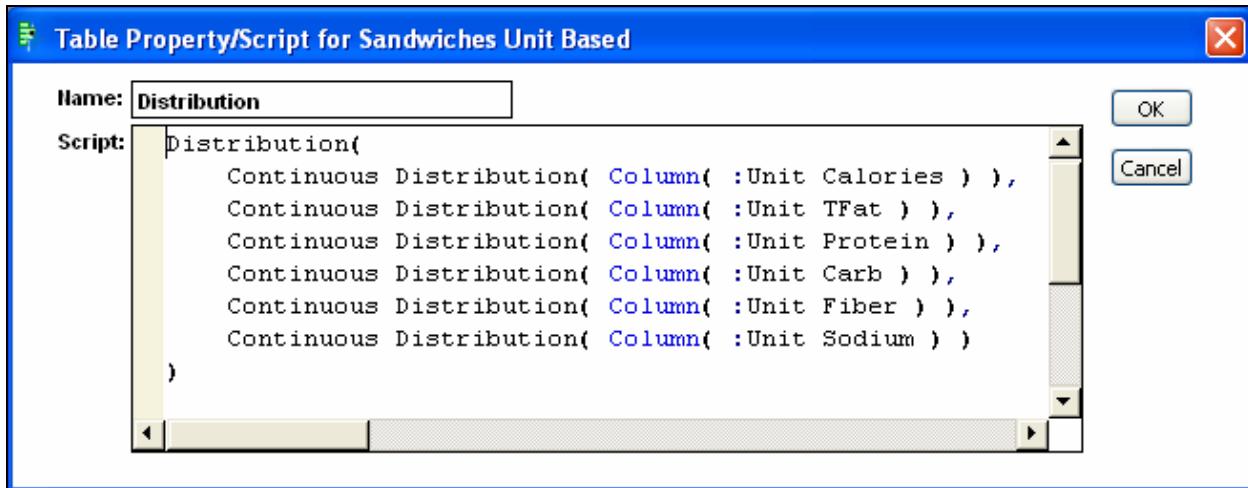
- Deselect all rows.



The script is saved in the Table panel.

- To view the script, click the red triangle next to **Distribution** and select Edit.





12. Select **OK** to close the script window.

Using Context-Sensitive Help

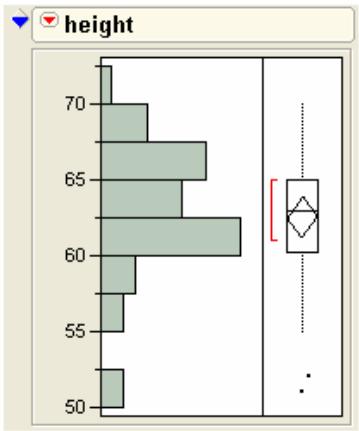
If you want more information about the outlier box plot, use the context-sensitive Help feature in JMP.

1. Select **Window** \Rightarrow **Sandwiches Unit Based- Distribution**.
2. Select **Tools** \Rightarrow **Help** and click on the outlier box plot.

A Help window opens and describes how to interpret the plot.

Outlier Box Plot

The outlier box plot is a schematic that lets you see the sample distribution and identify points with extreme values, sometimes called *outliers*. Box Plots show selected quantiles of continuous distributions.

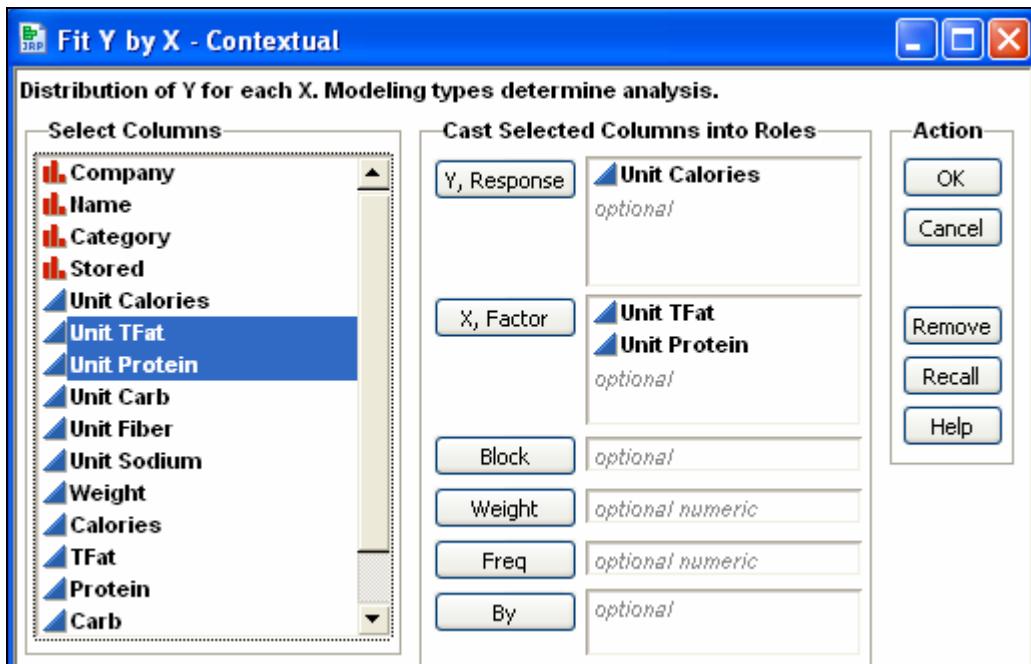


The ends of the box are the 25th and 75th quantiles, also called the *quartiles*. The difference between the quartiles is the *interquartile range*. The line across the middle of the box identifies the median sample value and the means diamond indicates the sample mean and 95% confidence interval.

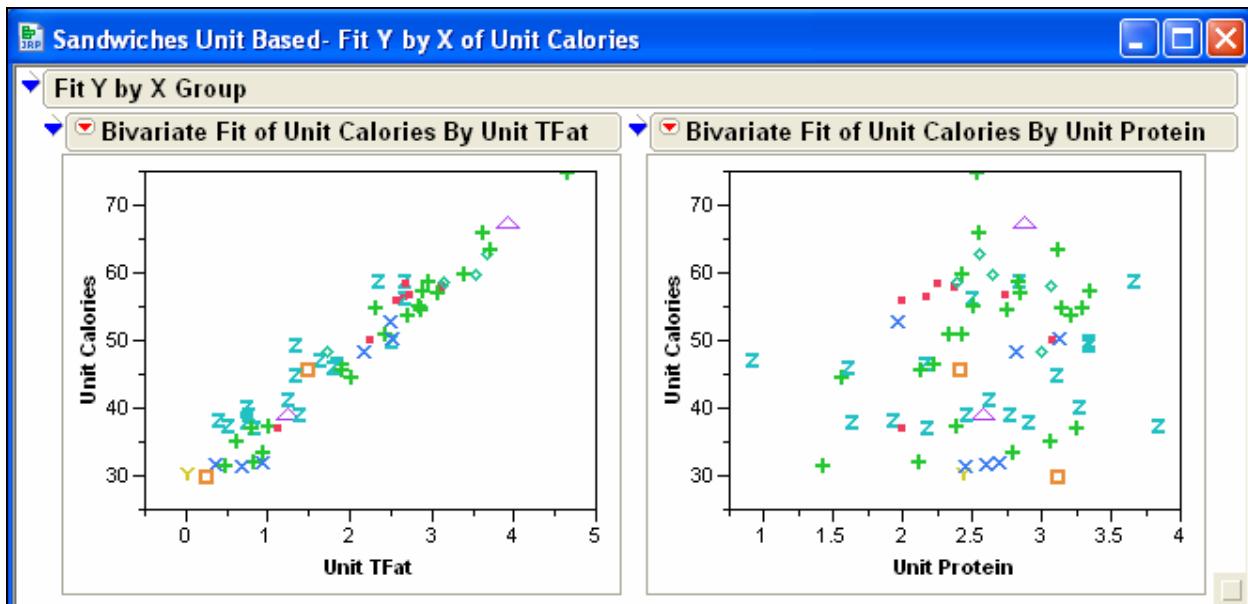
3. Close the JMP Help window and **Sandwiches Unit Based- Distribution**.

Examining Scatterplots

1. Select Analyze \Rightarrow Fit Y by X.
2. Select Unit Calories \Rightarrow Y, Response.
3. Select Unit TFat and Unit Protein \Rightarrow X, Factor.



4. Select OK.



The report contains two scatterplots that can be used to explore the relationship between **Unit Calories** and **TFat** as well as **Unit Calories** and **Unit Protein**.

Do you see any relationship, or trend, between these two columns? Answering this question would expand the data exploration phase of your analysis.

JMP generates the results with default scatterplot names. You can change these names to something more meaningful.

5. You can change the titles of individual plots and display tables in report windows. Double-click **Bivariate Fit of Unit Calories By Unit TFat**.

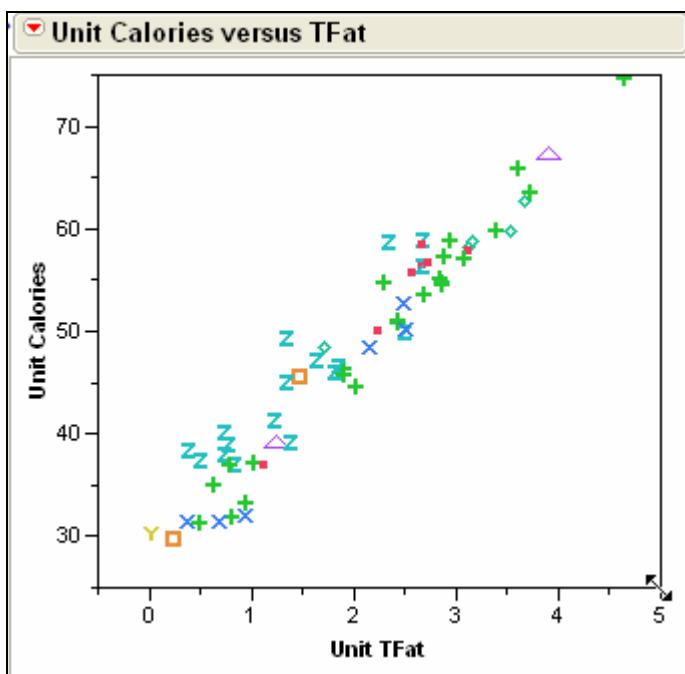


6. Type **Unit Calories versus TFat** and press ENTER.

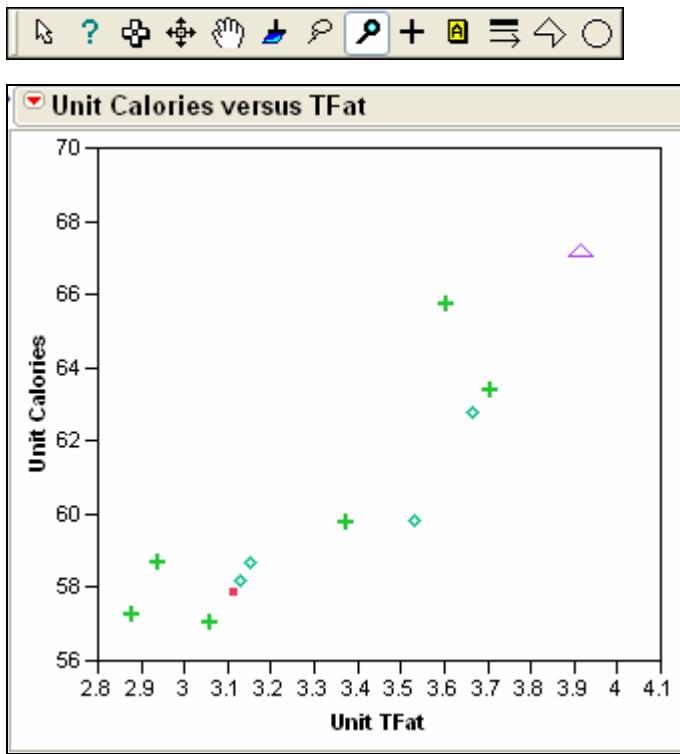
Graphics Tools

Explore the scatterplot of **Unit Calories** and **TFat** more closely with the use of the graphics tools.

1. To make the graph larger, drag it by using the double-headed arrow found by hovering the cursor in the lower-right corner of the graph.



2. Choose the Magnifier tool and draw a rectangle that captures the values in the plot above where **TFat** is between 3 and 4 and **Unit Calories** is between 56 and 70.



The magnified image is displayed. Notice that the scale of the axes is different.

Investigate the maximum value in this magnified window of **Unit TFat**=4.

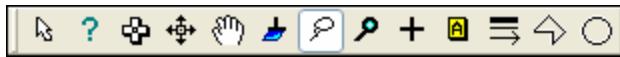
3. Choose the Crosshairs tool from the menu.



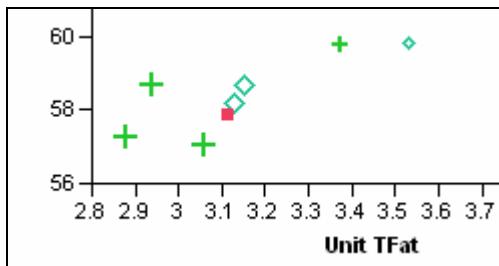
4. Place the Crosshairs tool on the data point near **Unit TFat** = 4 and **Unit Calories** = 68.
 5. Click and hold to read the approximate value of **Unit Calories**, which is 67.

Now select the values in the lower-left quadrant of the graph using the Lasso tool.

6. Choose the Lasso tool from the menu.



7. Hold the mouse button down and draw a circle around the values in the lower-left quadrant of the graph.



Because the rows have been selected, their markers are larger. Examining the Rows panel, you find that six rows have been selected and could be placed in their own data table for further analyses.

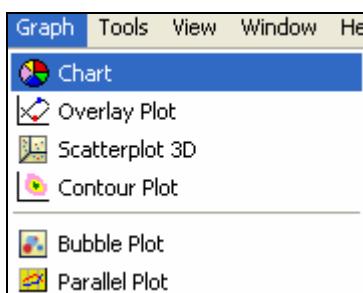
Restore the scatterplot to its original scale.

8. Select the Magnifier tool and double-click anywhere in the graph. JMP displays the scatterplot in its original scale.
9. Save the script for this analysis to the data table. Click the red triangle next to **Unit Calories versus TFat** and select **Script** \Rightarrow **Save Script to Data Table**.
10. Close the scatterplot window.
11. Save **Sandwiches Unit Based**.

Numerical and Graphical Results

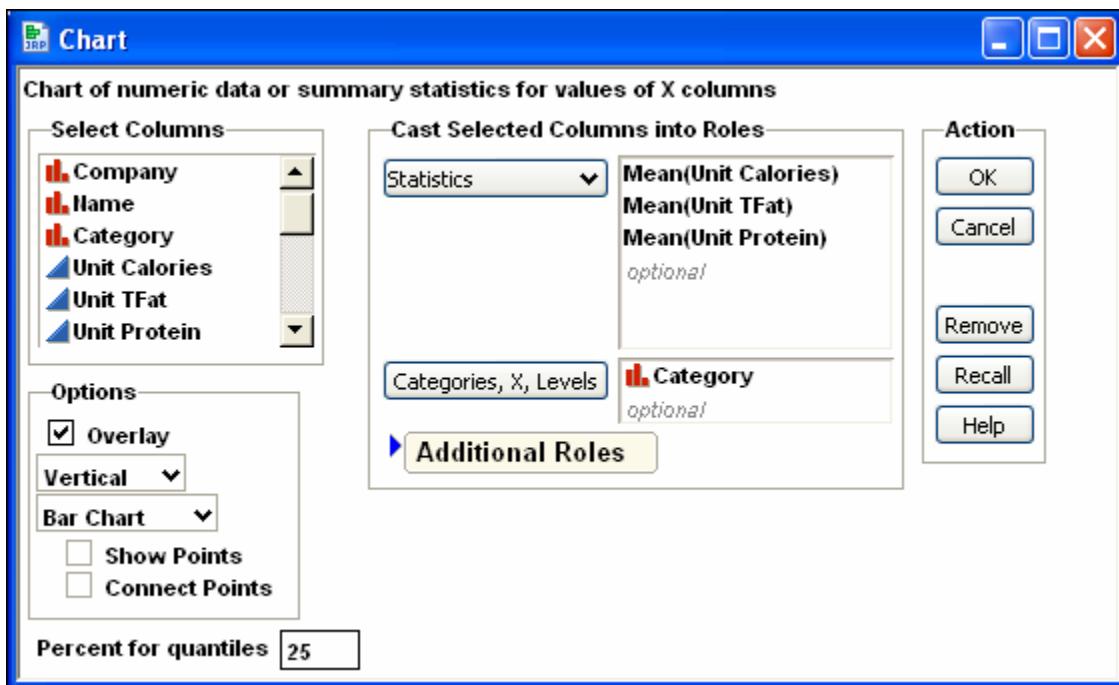
You have already generated numeric JMP tables for summary purposes. JMP also generates graphical summaries, which might be easier to understand.

1. Deselect the six selected rows.
2. Select **Graph** \Rightarrow **Chart**.

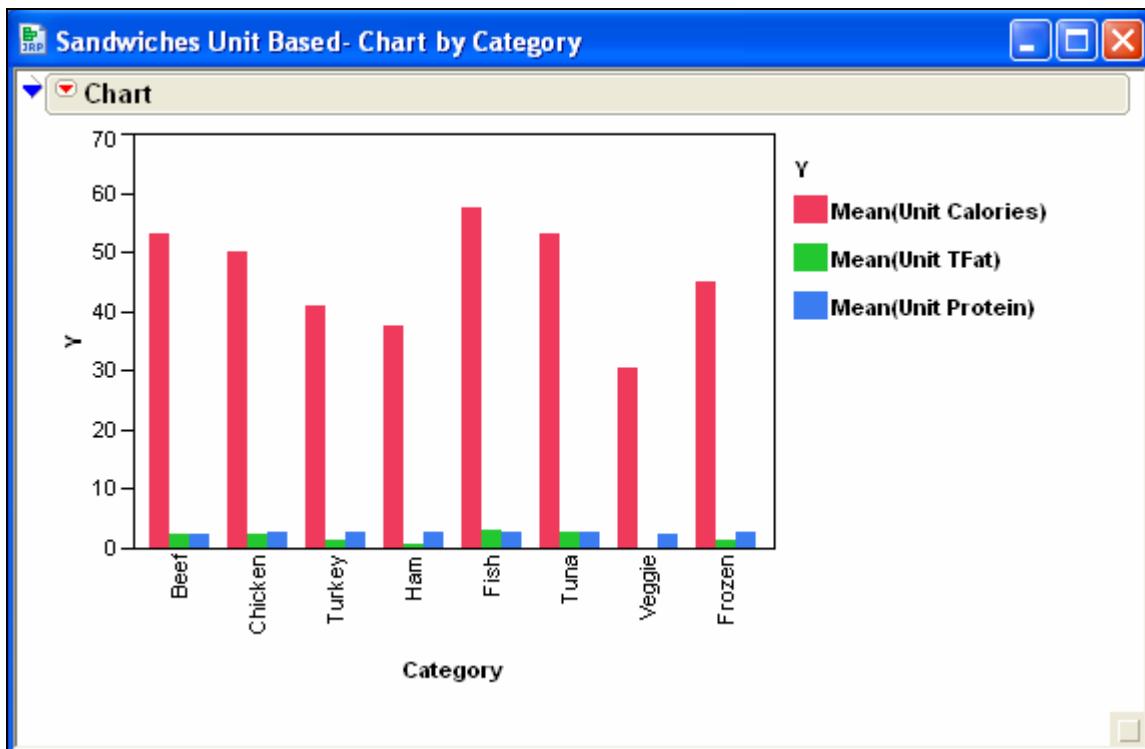


3. Select **Category** \Rightarrow **Categories, X, Levels**.

4. Select Unit Calories through Unit Protein \Rightarrow **Statistics** \Rightarrow **Mean**.

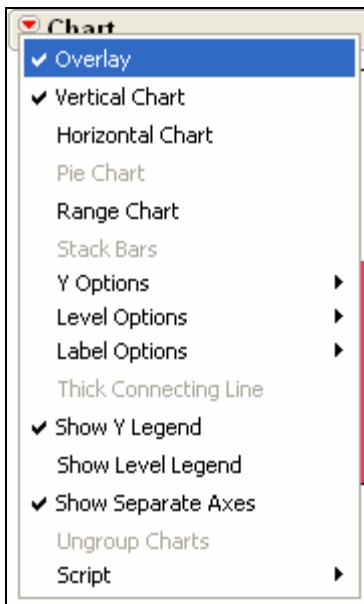


5. Select OK.



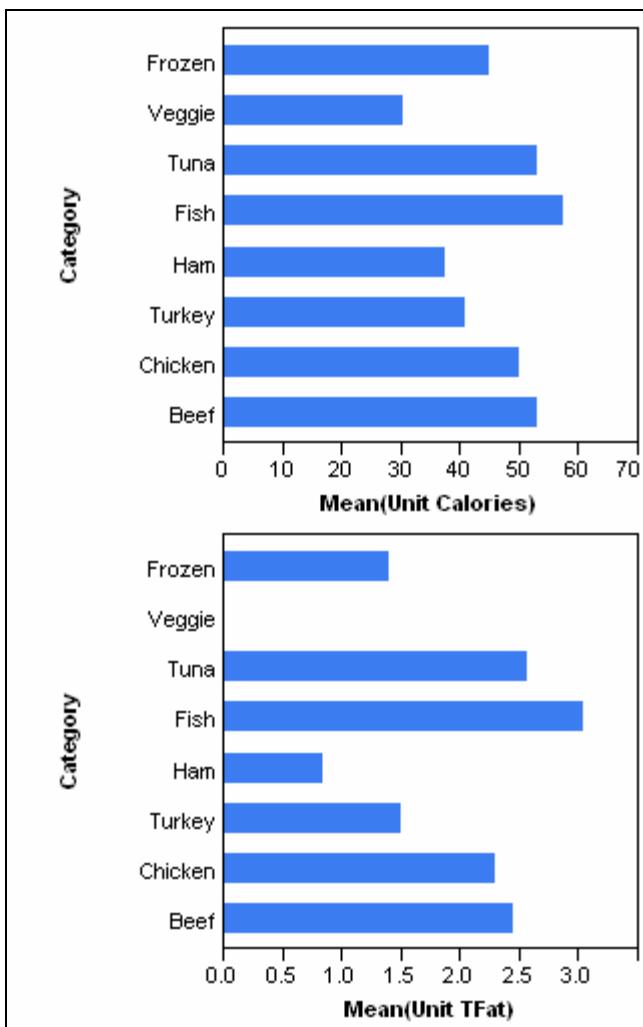
An overlay plot opens by default. It might be easier to compare the unit values by category if the plot is separated into three bar charts.

6. Click the red triangle next to **Chart** and select **Overlay**.



Several options are available that will make the charts more readable.

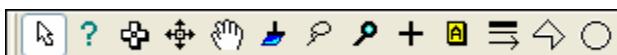
7. Click the red triangle next to **Chart** and select **Horizontal Chart**.



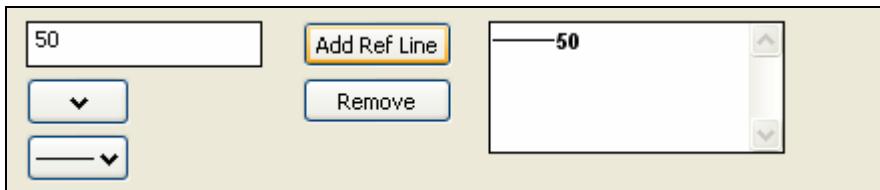
JMP presents three graphs with the averages (means) of the **Unit Calories**, **Unit TFat**, and **Unit Protein** columns summarized by the column **Category**.

Notice that the values of **Category** on the vertical axis are ordered according to the values assigned in an earlier chapter. It appears that the value for **Unit TFat** in a Veggie sandwich is zero.

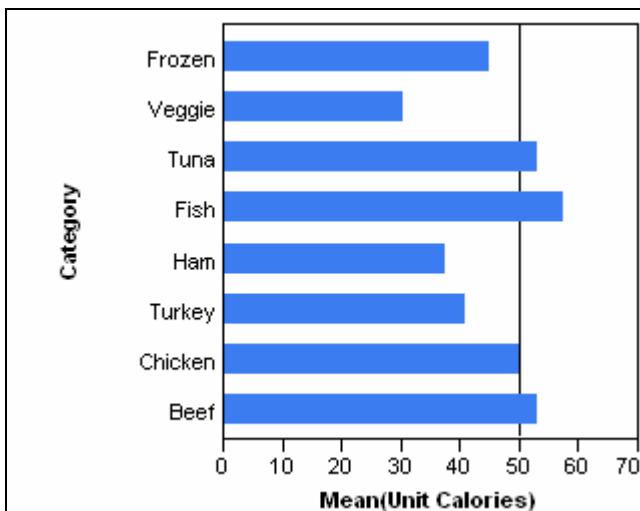
8. Use the default Arrow cursor and starting at the first bar in the graph, move the cursor down toward the axes. Notice that after you approach the axes, the Arrow cursor changes to the Grabber cursor.



9. Double-click on the **Mean(Unit Calories)** axis.
10. Modify the dialog box as shown below. Type **50** in the box on the left and select **Add Ref Line**.



11. Select **OK** to apply the changes.



A black solid reference line has been added at the value of **Mean (Unit Calories)=50.0**.

- If column labels are turned off, the average value of each **Category** level can be determined by clicking and holding on the bar of your choice.
12. Close all windows.

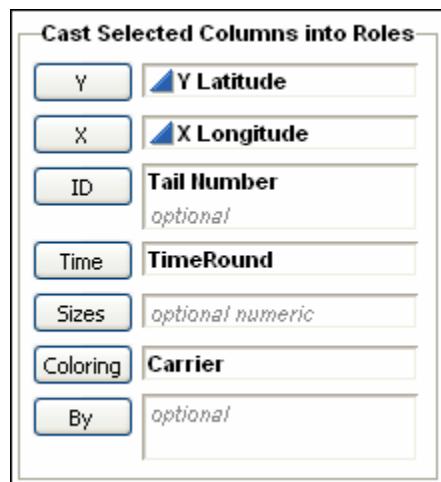
Examining Bubble Plots

A *bubble plot* is a scatterplot that represents its points as circles (bubbles). Optionally, the bubbles can be sized according to another column, colored by another column, aggregated across groups defined by one or more other columns, and dynamically indexed by a time column. With the opportunity to see up to five dimensions at once (x position, y position, size, color, and time), bubble plots can produce dramatic visualizations and make interesting discoveries easy.

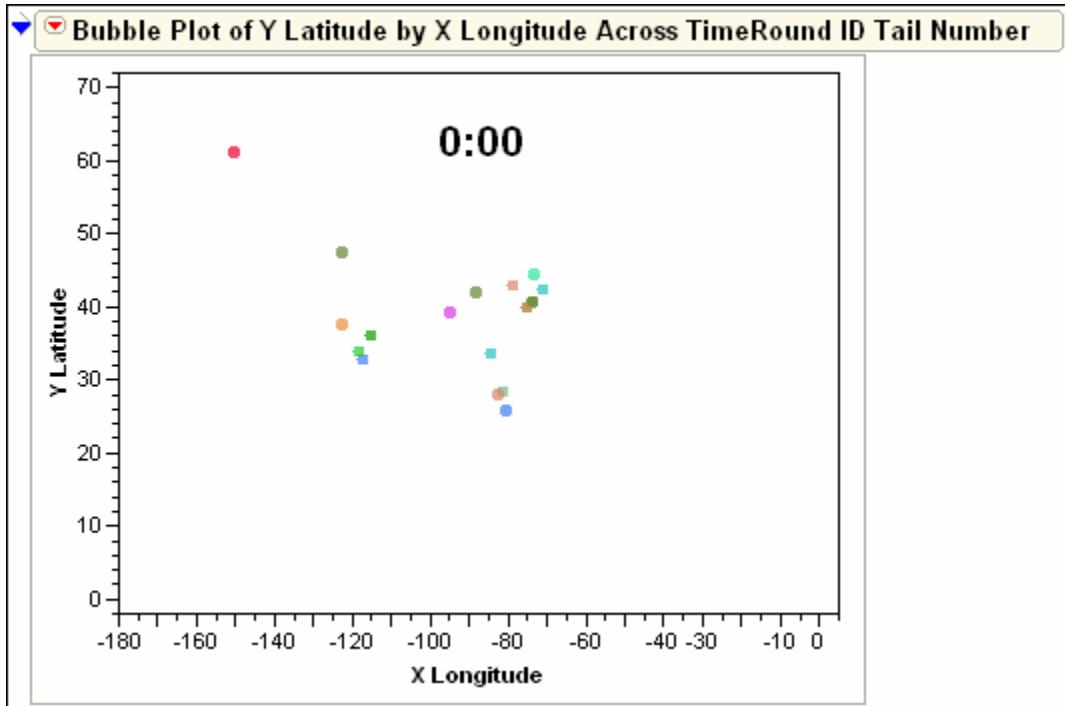
Air Traffic.JMP contains information on all of the domestic flights on January 6, 2005. Use a bubble plot to examine the locations of the flights (**Latitude** and **Longitude**), the airlines (**Carrier**), the flight number (**Tail Number**), and the departure and arrival times for each flight (**TimeRound**). Specifically, **TimeRound** is only a change in the representation of the original time values (**Time**).

 **Time** is an integer in the form HHMM, representing HH:MM. The **TimeRound** column converts **Time** to a true JMP value of time, a clock time, always rounded down to the last five minutes. There are two values of **TimeRound** for each flight, one for departure and one for arrival. The bubble plot will automatically linearly interpolate the points between the departure and destination locations over the interval of time.

1. Open **Air Traffic.JMP**.
2. Select **Graph** \Rightarrow **Bubble Plot**.
3. Select **Y Latitude** \Rightarrow **Y**.
4. Select **X Longitude** \Rightarrow **X**.
5. Select **Tail Number** \Rightarrow **ID**.
6. Select **TimeRound** \Rightarrow **Time**.
7. Select **Carrier** \Rightarrow **Coloring**.



8. Select **OK**.

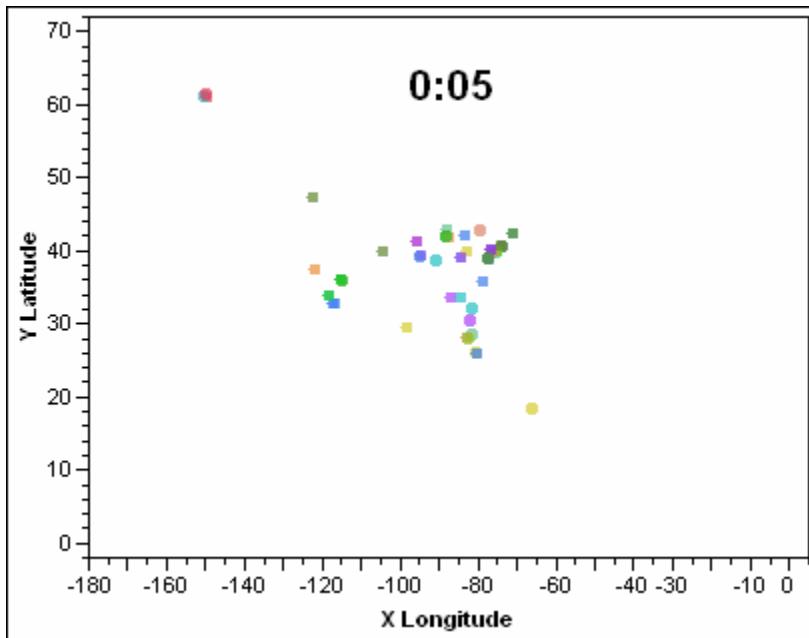


The bubble plot is generated, and the outline item describes most of the roles of the variables. A key is available that shows all roles.

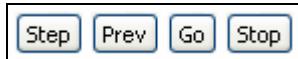
- Click the red triangle next to **Bubble Plot** and select Show Roles.



- Select Step to move forward one unit in time (which is, in this case, five minutes).



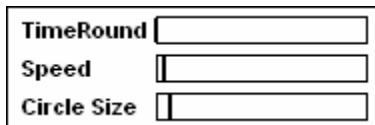
Instead of using Step to move forward in time, you can use Go to have JMP move forward automatically.



11. Select Go.

The bubbles move as time advances in increments of five minutes. The progress through time can be noted by the location of the slider bar next to TimeRound.

The location in time can be changed by adjusting this same slider bar. It is possible to change the rate at which the bubbles move through time by adjusting the slider bar next to Speed. In addition, the default circle size can be modified using the slider bar next to Circle Size.



A point can be selected in the bubble plot by clicking on it. After a point is selected, the Rows panel indicates how many actual observations are encompassed in that selected point. However, the default is that a point does not maintain its selection status after it is selected (the Rows panel still indicates that it is selected in the data table, but it does not remain selected in the bubble plot). It is possible to change this default setting.

12. Select a point in the bubble plot.

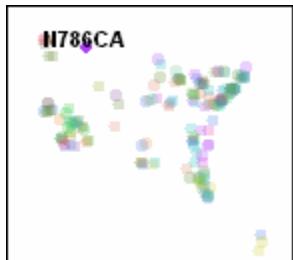
Notice the updated Rows panel. Notice, too, that the bubble does not maintain its selected status, making it difficult to track a particular bubble.

<input checked="" type="checkbox"/> Rows	
All rows	38118
Selected	14
Excluded	0
Hidden	0
Labelled	0

13. Click the red triangle next to **Bubble Plot** and select Selectable Across Gaps.

14. Select a point in the bubble plot.

Notice the ID variable (**Tail Number**) is printed on the graph and remains printed on the graph as the bubble moves. The ID variable's value remains printed on the graph until it is removed.



15. Deselect the selected bubble by clicking on a blank area of the plot.

Specific values of a variable can also be selected using the Data Filter. Select all of the JetBlue flights.

16. Select **Rows** \Rightarrow **Data Filter**.

17. Select **Carrier** \Rightarrow **Add**.

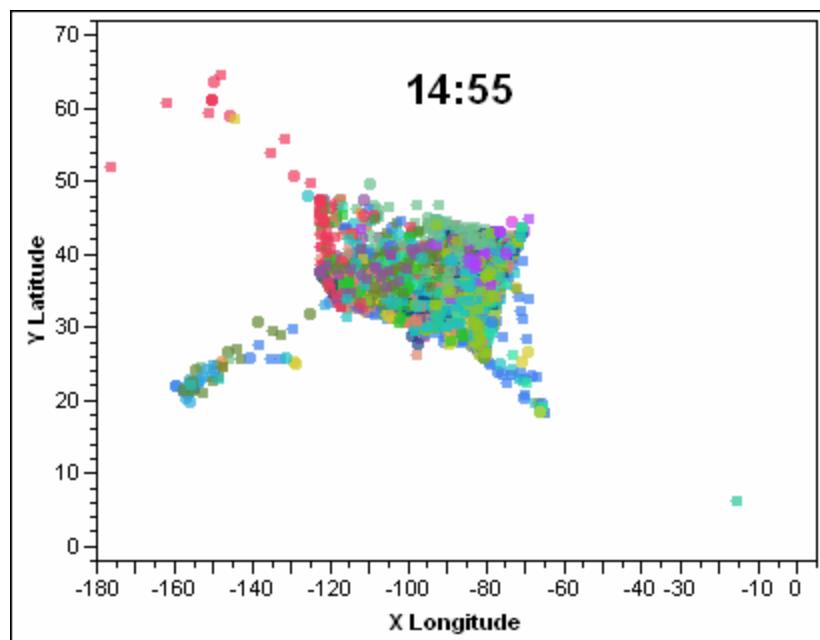
18. Select **JetBlue**.

The bubbles that correspond to JetBlue flights are labeled and selected. This makes it easy to track them as they move through time.

19. Select **Clear** in the Data Filter to clear the row selection.

20. Close the Data Filter.

21. Select **Stop**.



Notice the rough outline of the United States visible in the bubble plot.

Your graph might look slightly different because you might have stopped it at a different time.

22. Click the red triangle next to **Bubble Plot**.

By selecting **Filled**, the bubbles are filled with color, rather than only being a colored outline. By selecting **Trails**, the bubble plot shows the past history of selected bubbles as a semi-transparent trail. This option is only available when a **Time** column is specified. Finally, by selecting **All Labels**, each bubble in the plot has a label printed on the graph.

23. Close the bubble plot.

24. Close **Air Traffic.JMP**.

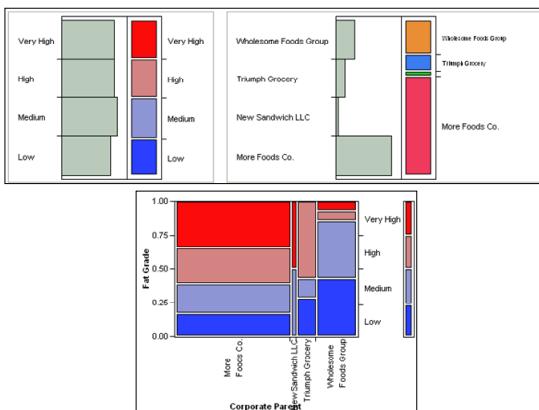
3.2 Examining Relationships Between Attribute Data Columns

Objectives

- Examine the distribution reports of nominal and ordinal columns.
- Explore relationships between nominal and ordinal columns with the mosaic plot.

15

Relationships Between Attribute Data Columns



16

Histograms and mosaic charts can be used to explore relationships between nominal and ordinal columns.

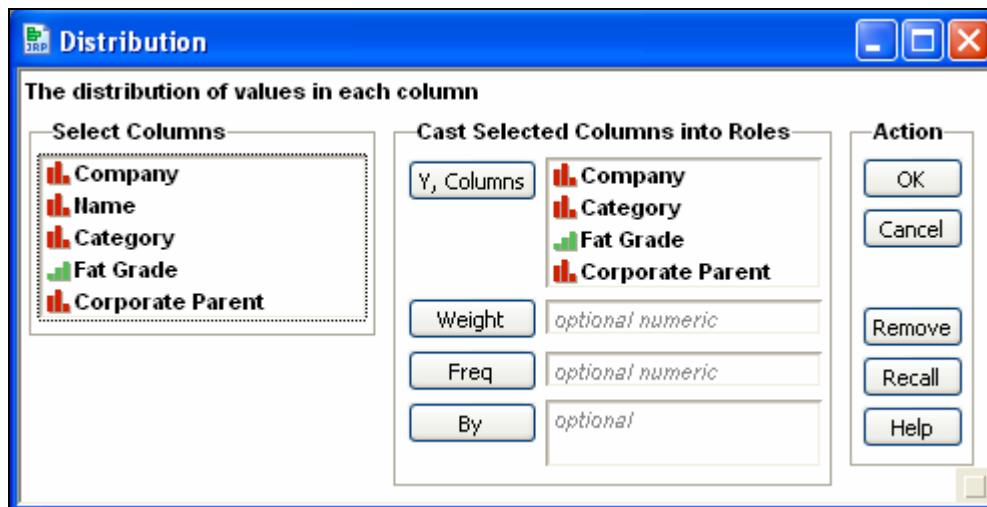


Exploring Distributions of Attribute Data Columns

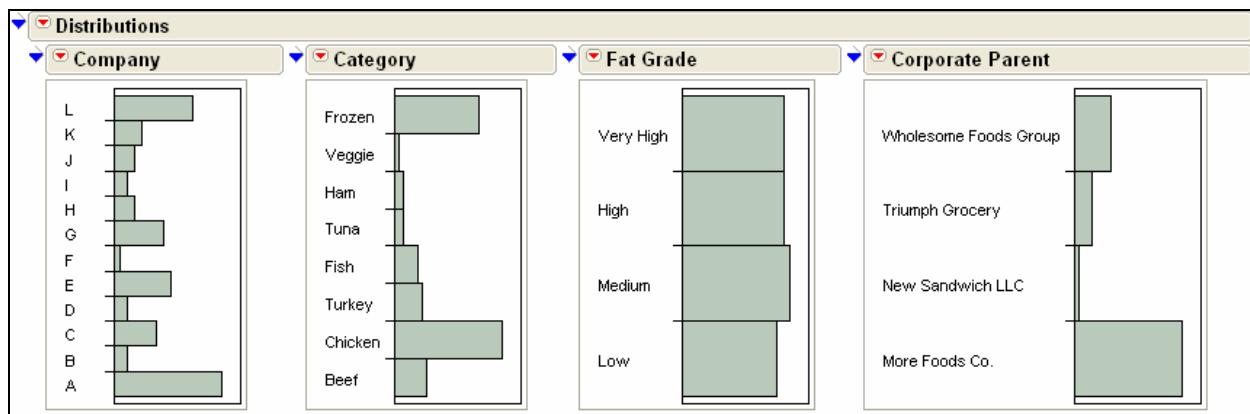
Histograms and mosaic charts graphically assist users in determining whether a relationship or an association exists between nominal and ordinal columns in a data table.

Histograms

1. Open **Sandwiches Attributes.JMP**.
2. Select **Analyze** \Rightarrow **Distribution**.
3. Select **Company**, **Category**, **Fat Grade**, and **Corporate Parent** \Rightarrow **Y, Columns**.



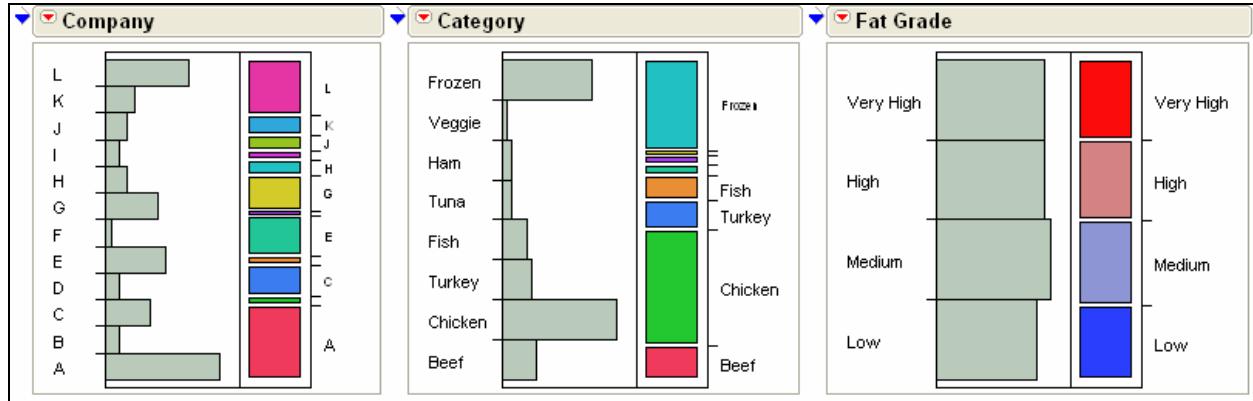
4. Select **OK**.



A histogram is displayed for each nominal or ordinal column. The levels and count for each column are listed in a table below the graphs.

 The values in the **Fat Grade** histogram are ordered from **Low** to **Very High** automatically by JMP.

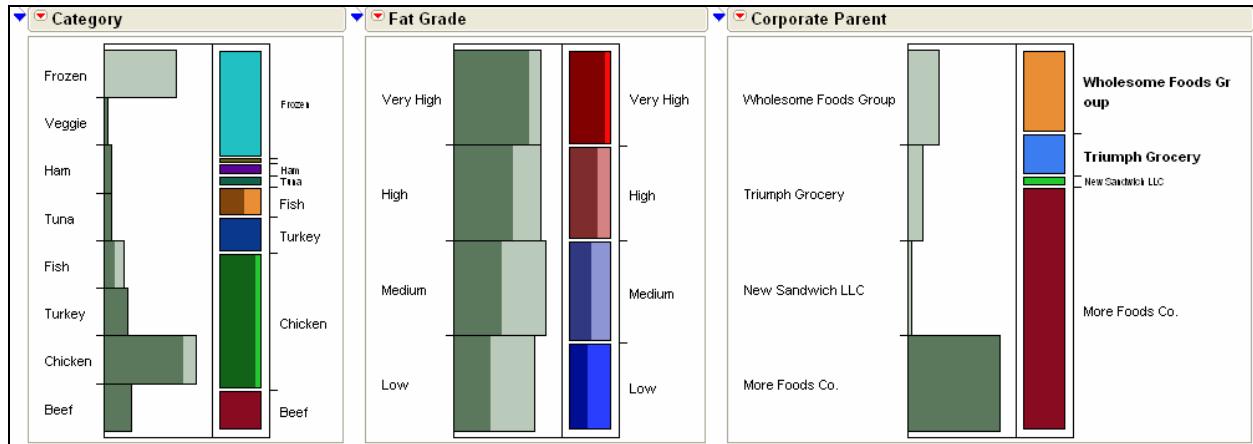
5. Hold the CTRL key, click the red triangle next to **Company**, and select **Mosaic Plot**.



A mosaic plot displays an attribute column as a stacked bar chart where each segment is proportional to the group's sample size. You can visualize mosaic bars as the bars of the corresponding histogram stacked end to end.

Using the histograms and mosaic plots, it can be easily determined that the majority of the sandwiches in the table are Chicken or Frozen and that **Company A** makes the most sandwiches.

6. Look for potential relationships or associations in the data by clicking the bar corresponding to **More Foods Co.** in the **Corporate Parent** analysis.



JMP automatically highlights the selected rows in all open analysis windows and in the corresponding data table.

Visually, you can make statements about the products of **More Foods Co.** They are all **Fresh** sandwiches and the majority of their sandwiches are **Beef**, **Chicken**, and **Turkey**.

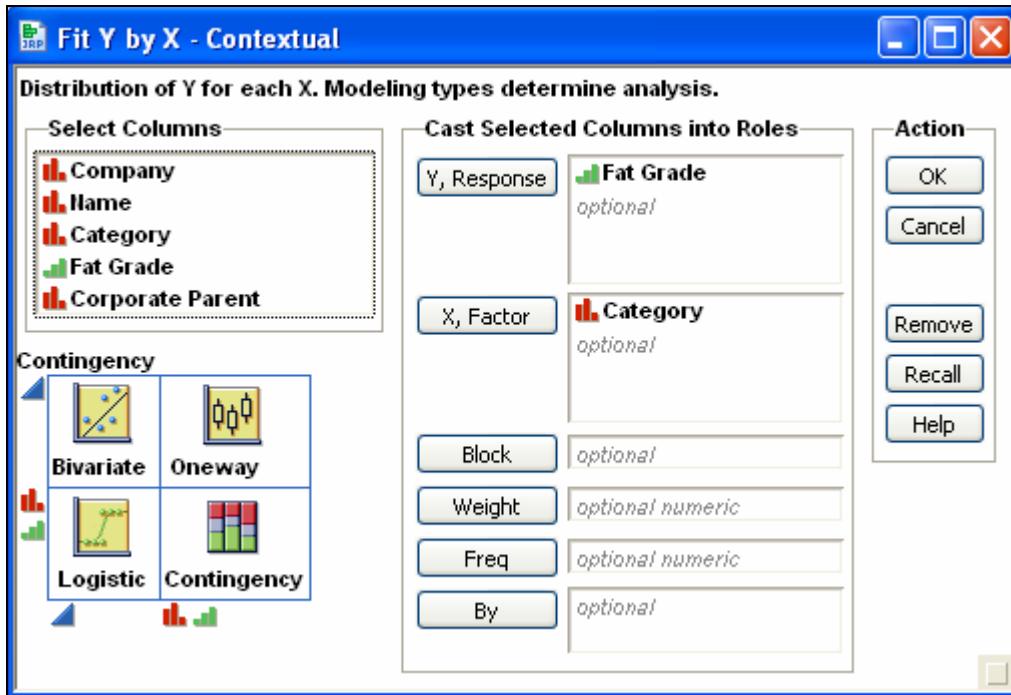
 The interactive capability of JMP enables you to quickly create a new data table from observations selected on a graph. Select **Tables** \Rightarrow **Subset** to create a new table with these 41 rows. Alternately, right-click on the selected bar and select **Subset** to create a new table.

7. Click the red triangle next to **Distributions** and select **Script** \Rightarrow **Save Script to Data Table**.

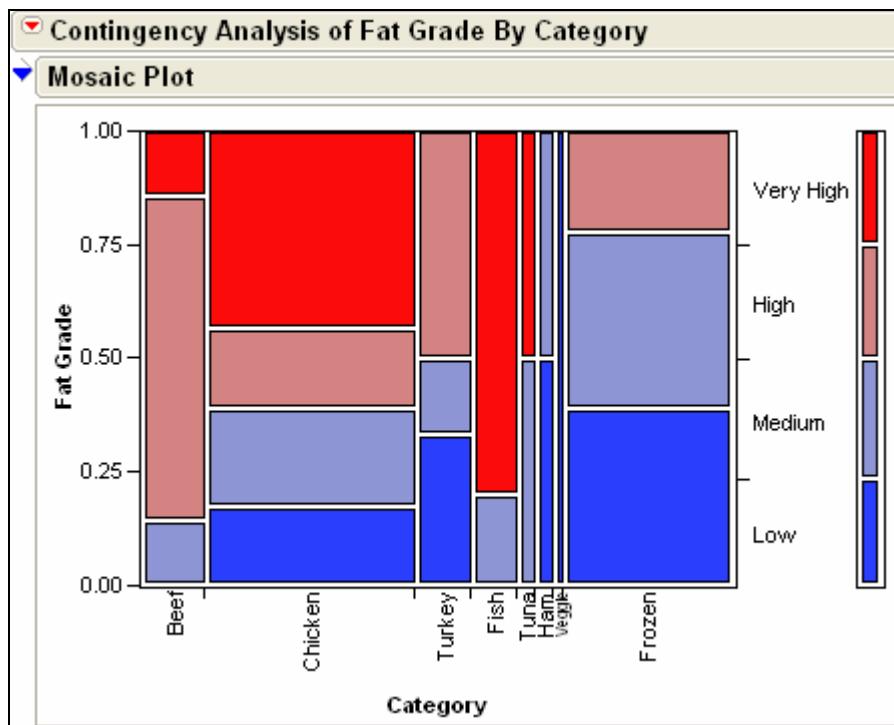
Examining Associations Using Mosaic Plots

In the Fit Y by X platform, relationships between nominal and ordinal columns can be examined using a mosaic plot.

1. Select Analyze ⇒ Fit Y by X.
2. Select Fat Grade ⇒ Y, Response.
3. Select Category ⇒ X, Factor.



4. Select OK.



Mosaic plots display panels whose heights and widths correspond to percentages. The height of the panels represents the percentage of each value of the Y-axis variable (**Fat Grade**) for each value of the X-axis variable (**Category**). For example, over seventy-five percent of fish sandwiches are very high in fat.

The width of the panels is proportional to the percentage of each value of the X-axis variable for each value of the Y-axis variable. In this example, chicken sandwiches make up the largest category, while veggie and ham sandwiches make up a small proportion of the data.

If the height of the bars for each level of the Y-axis variable is approximately equal, there is probably no association between the two columns.

Graphically, there appears to be an association between **Category** and **Fat Grade**.

- A contingency table and formal tests for association are provided beneath the mosaic plot. The discussion of these reports is beyond the scope of this course.
5. Click the red triangle next to **Contingency Analysis of Fat Grade By Company** and select Script \Rightarrow Save Script to Data Table.
 6. Close all windows without saving.

3.3 Exploratory Analysis Using Recursive Partitioning

Objectives

- Define partitioning.
- Use the Partition platform in JMP.

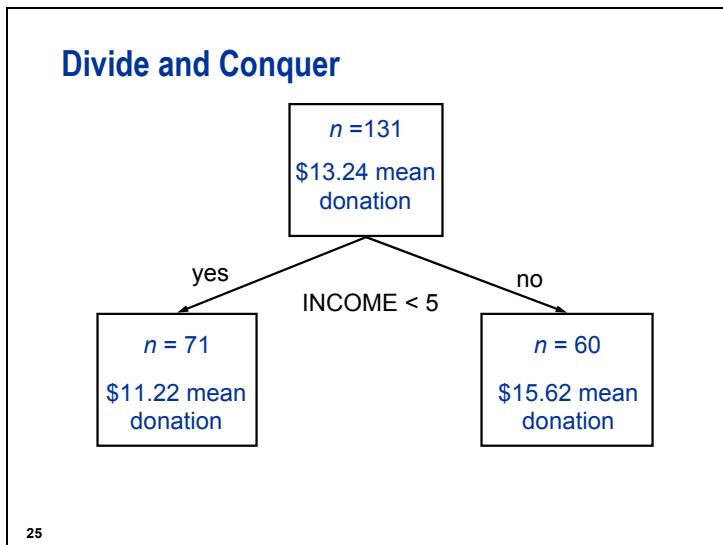
23

In this section, you learn what partitioning is and how to use the Partition platform in JMP.

Recursive Partitioning

Partitioning refers to segmenting the data into subgroups that are as homogeneous as possible with respect to the dependent variable (Y).

24



The tree of partitions is fitted to the data by recursive partitioning. In this case, the binary split, **INCOME** < 5, was chosen. (**INCOME** is an ordinal variable divided into levels 0-9.) The 131 cases are split into two groups, one with a mean donation of \$11.22 and the other with a mean donation of \$15.62.

The method is recursive because each subgroup results from splitting a subgroup from a previous split. Thus, the 71 cases in the left box and the 60 cases in the right box are split again in similar fashion.



Applications of trees include the following:

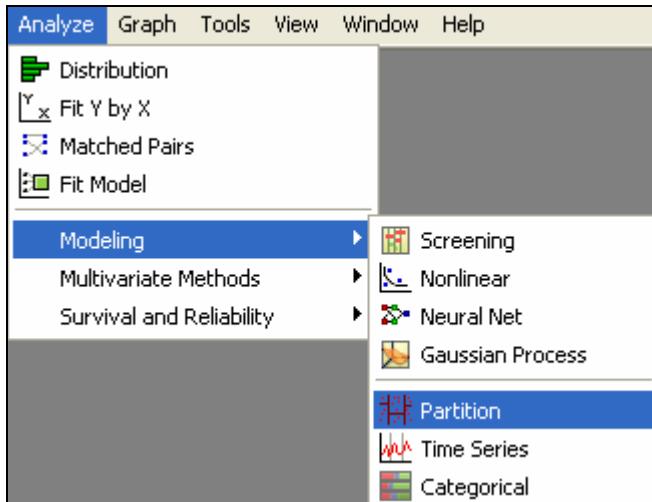
- missing value imputation (replacing missing values with reasonable choices)
- categorical variable consolidation (combining category levels into a smaller number)
- variable selection



Recursive Partitioning

This demonstration illustrates how to explore data using the Partition platform in JMP. **Freshmen.JMP** contains information on 100 rising college sophomores. The university hopes to understand what variables help predict student success in college. Specifically, the university wants to know what variables help predict whether a student will fail during freshman year and leave the school or succeed and return for sophomore year. To measure student success, **GPA** is used as the response. The predictors are **Miles from Home**, **College** (within the university), **Accommodations** (dormitory or off-campus housing), **Years Off** (time off between high school and college), **Part-time Work Hours**, **Attends Office Hours**, and **High School GPA**.

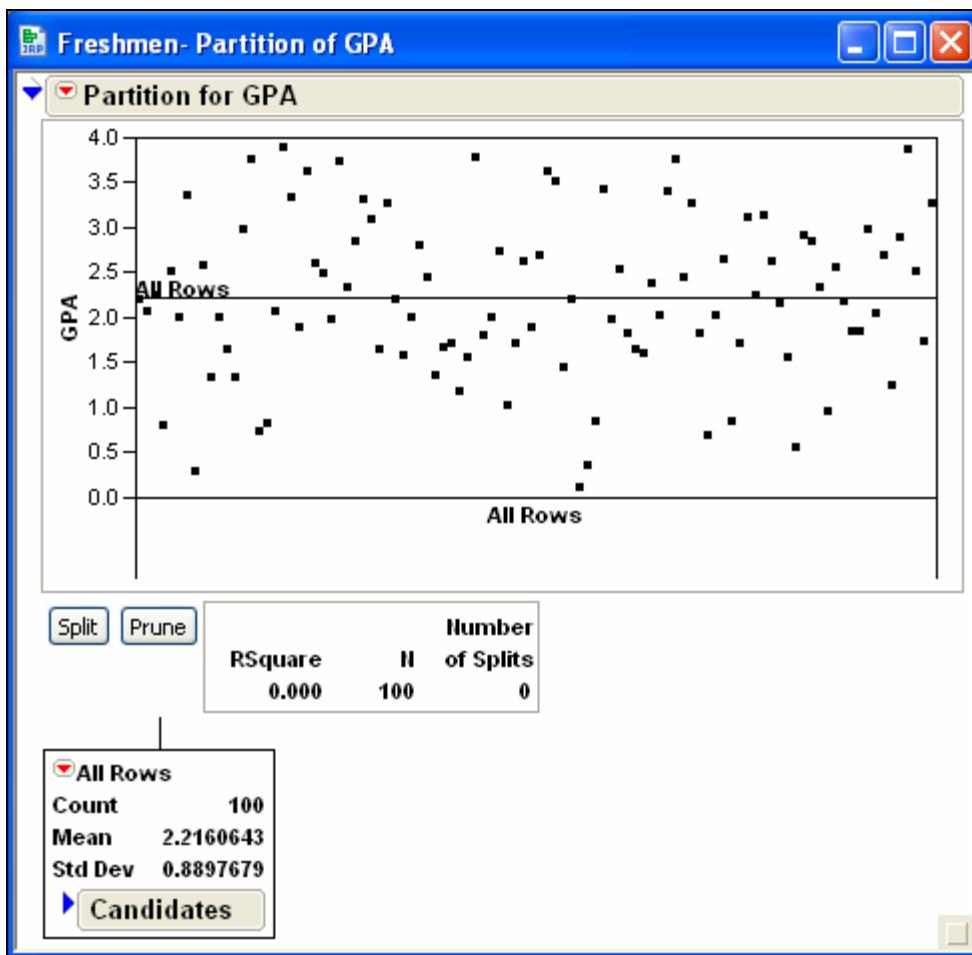
1. Open **Freshmen.JMP**.
2. Select **Analyze** \Rightarrow **Modeling** \Rightarrow **Partition**.



3. Select **GPA** \Rightarrow **Y, Response**.
4. Select **Miles from Home** through **High School GPA** \Rightarrow **X, Factor**.
5. Select **OK**.



JMP produces a partition graph and output.



The vertical axis on the plot represents **GPA**. The average **GPA** is plotted as the horizontal line in the graph. The horizontal axis is not significant; JMP simply spreads the data points so that they are as easy to see as possible. At this point, JMP has all 100 rows together in one node, and the average **GPA** for all 100 points is 2.21.

In order to determine the first, optimal split, each independent variable and its range of values must be considered.

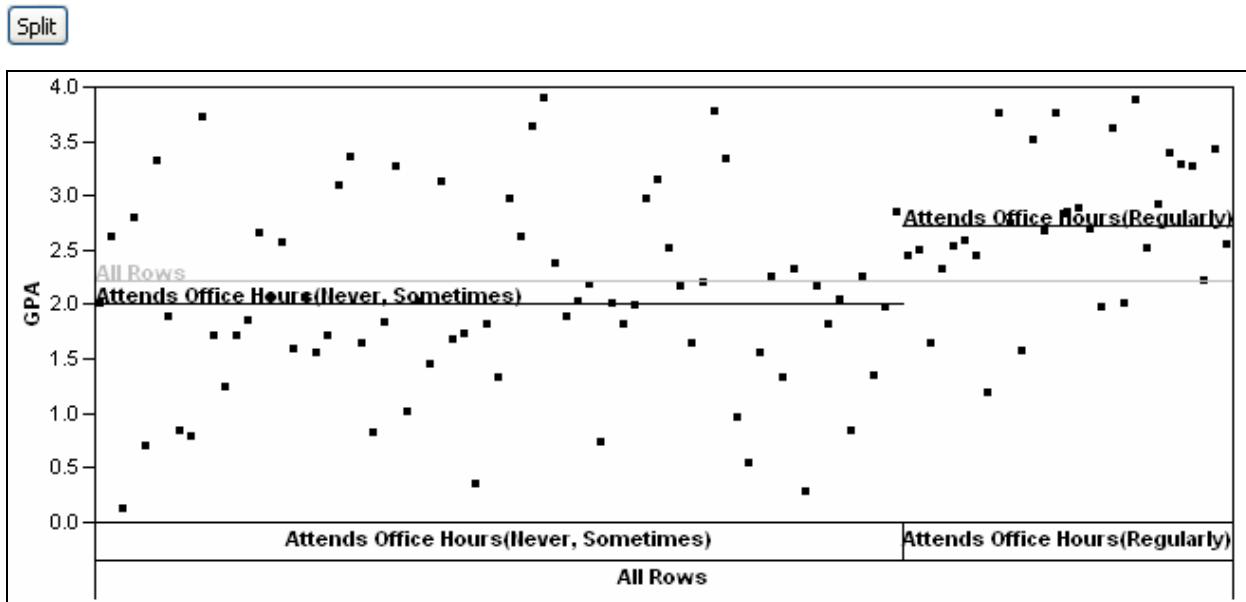
6. Open the Candidates report in the All Rows box by clicking on the blue expand/collapse icon next to **Candidates**.

<input checked="" type="checkbox"/> All Rows		
Count	100	
Mean	2.2160643	
Std Dev	0.8897679	
▼ Candidates		
Term	Candidate SS	LogWorth
Miles from Home	12.22467098	< 3.711709577
College	1.80815853	0.239548632
Accommodations	2.85103798	0.938789863
Years Off	0.93335382	0.164484416
Part-Time Work Hours	6.66309130	1.815383906
Attends Office Hours	10.71194431	> 3.740178086
High School GPA	3.39879906	0.497201634

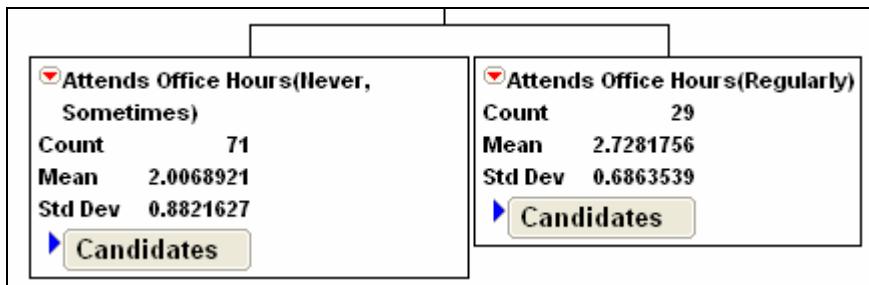
The candidate variables are shown in the Candidates report. You can examine the Candidate SS column to see which split to make first. In general, the independent variable with the highest SS value is the one that JMP uses for the first split. In this example, however, **Miles from Home** has the highest Candidate SS, but JMP uses **Attends Office Hours** instead because **Attends Office Hours** has a larger LogWorth value. JMP indicates the discrepancy in this case with arrow markers. (When the two values are in agreement, JMP indicates the best choice with an asterisk.)

-  The LogWorth value is the negative log of the p -value for a Pearson chi-square test for independence between groups. Larger LogWorth values indicate better splits.

7. Select **Split**.



After this first split, a horizontal line appears on the plot for each group involved in the split. These lines are labeled with the variable and the level of the split. Notice that the line for All Rows is now shown as a light gray line; only the lowest levels of the splits (the leaves) remain in dark print. The most homogeneous groups (or the lowest levels of the tree) are called the *leaves*.



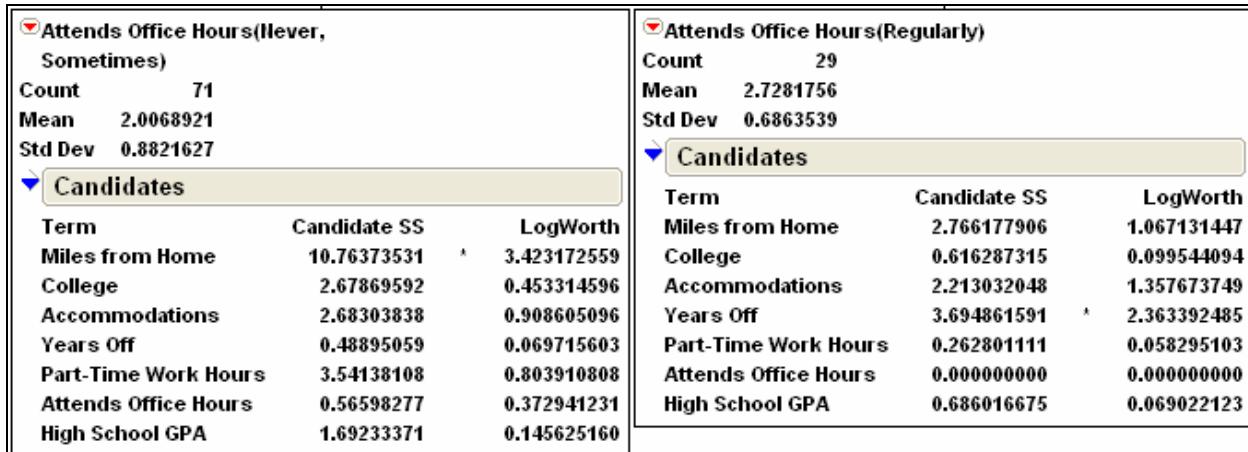
JMP splits the 100 points into two groups. The split variable and value used, **Attends Office Hours** with Regularly in one group and Sometimes and Never in the other group, is the one that produces the greatest difference in the two group averages with respect to **GPA**.

Examine the report under the graph.

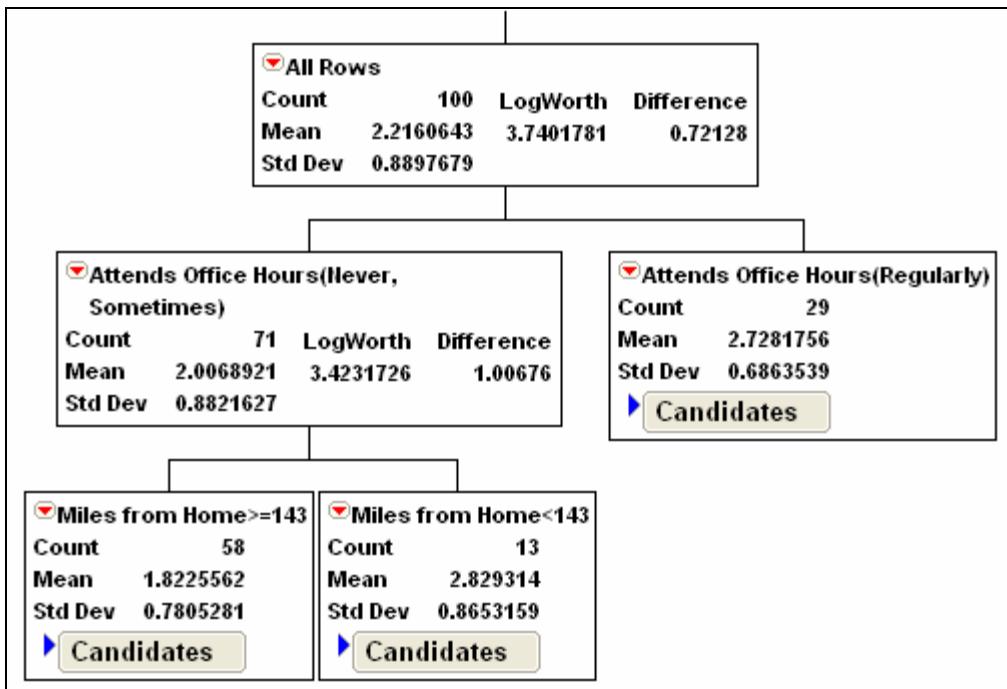
RSquare	Number of Splits
0.137	100 1

The RSquare for the partition model given is .137. About 14% of the variation in **GPA** is accounted for by **Attends Office Hours**.

- Open the Candidate reports for the two leaves by clicking on the blue expand/collapse icons next to **Candidates** in each box.



What will be the next split made? **Miles from Home** in the **Attends Office Hours(Never, Sometimes)** leaf will be the next split because it has the highest Candidate SS value.

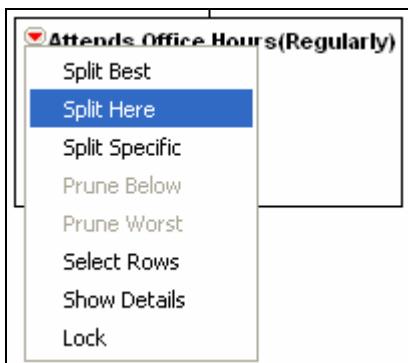
9. Select Split.

JMP splits the Never and Sometimes leaf of **Attends Office Hours** into two leaves for **Miles from Home**. Again, examine the RSquare report.

RSquare	Number of Splits	Number
0.274	100	2

The RSquare for the partition model given is .274. About 27% (or over one quarter) of the variation in **GPA** is accounted for by **Attends Office Hours** and **Miles from Home**.

- ✍ Users can select a specific leaf to split by clicking the red triangle next to the split level in a leaf and selecting **Split Here**. The same menu can be used to select a specific candidate variable within a leaf as the split to make by selecting **Split Specific**.

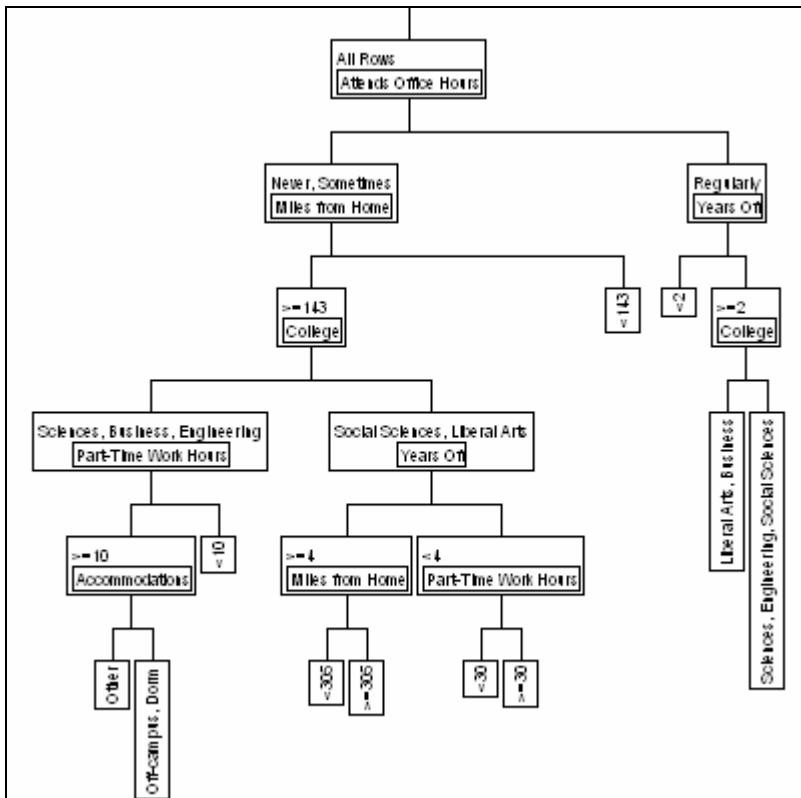
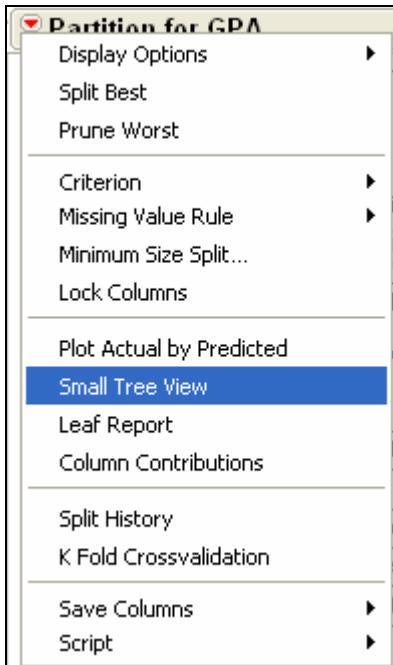


Continue to split until you have performed a total of ten splits.

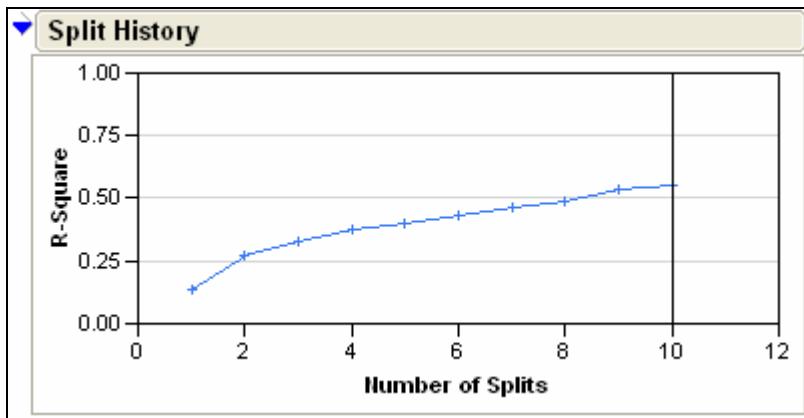
10. Select Split eight times.

The tree is too large to see easily, so add a small tree graph to your output.

11. Click the red triangle next to **Partition for GPA** and select **Small Tree View**.



With each split, RSquare increases. To see the added benefit of each split with respect to the other splits, click the red triangle next to **Partition for GPA** and select **Split History**.



It is easy to see from this graph that the first two splits explained a large portion of the variation in **GPA**. After the first two splits, the next three splits explained a reasonably large additional amount of variation in **GPA**. However, after about the fifth split, the additional variation in **GPA** that is explained is not particularly large. Therefore, there does not seem to be any benefit to splitting more than five times.

12. Select **Prune** five times.

RSquare	Number of Splits
0.427	100

After further examination, it appears that the five variables that explain approximately 43% of the variation in **GPA** are **Attends Office Hours**, **Miles from Home**, **Years Off**, **College**, and **Part-Time Work Hours**.

13. To see a summary of the leaves, click the red triangle next to **Partition for GPA** and select **Leaf Report**.

Leaf Label	Mean	Count
Attends Office Hours(Never, Sometimes)&Miles from Home>=143&College(Sciences, Business, Engineering)&Part-Time Work Hours>=10	1.44025304	28
Attends Office Hours(Never, Sometimes)&Miles from Home>=143&College(Sciences, Business, Engineering)&Part-Time Work Hours<10	2.42094871	5
Attends Office Hours(Never, Sometimes)&Miles from Home>=143&College(Social Sciences, Liberal Arts)	2.13105718	25
Attends Office Hours(Never, Sometimes)&Miles from Home<143	2.82931398	13
Attends Office Hours(Regularly)&Years Off<2	2.1498602	8
Attends Office Hours(Regularly)&Years Off>=2	2.9484862	21

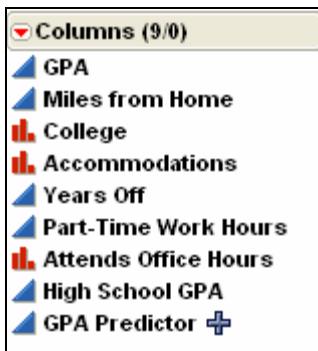
The number of observations and the average **GPA** of the observations in each group are summarized in the report.

A number of columns can be saved back to the data table, including the prediction formula.

14. Click the red triangle next to **Partition for GPA** and select **Save Columns** \Rightarrow **Save Prediction Formula**.

15. Select **Window** \Rightarrow **Freshmen**.

16. Click the plus sign next to **GPA Predictor** in the Columns panel.



The prediction equation is displayed.

17. Select File \Rightarrow Save As....

18. Type **GPA Partition**.

19. Select Save.

20. Close the data table and the Partition platform output.

 When a data set has missing values, the results of the Partition Analysis can differ depending on which method JMP uses to assign missing values to groups. The option to choose the method is in the Preferences menu. Select File \Rightarrow Preferences \Rightarrow Platforms \Rightarrow Partition \Rightarrow Missing Value Rule.

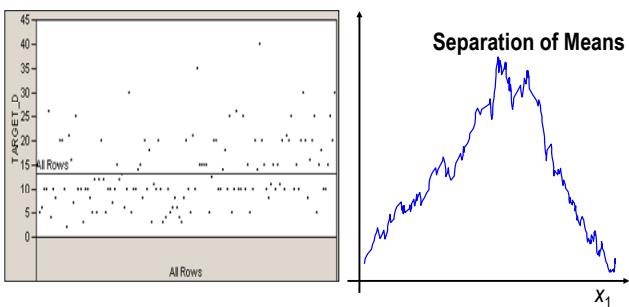
3.4 Recursive Partitioning Details (Optional)

Objectives

- Describe the splitting criteria used in JMP.
- Review algorithm parameters available in JMP.

32

Tree Algorithm: Calculate Separation of Means



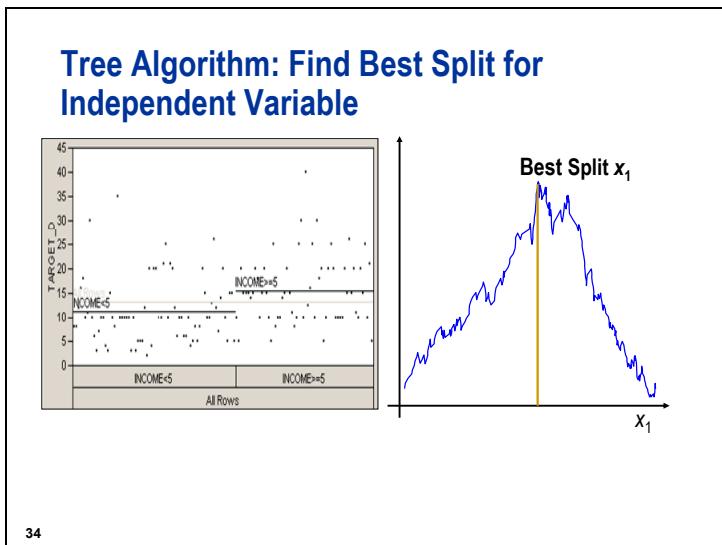
33

Which splits are to be considered? Which split is the best? The number of splits is enormous in all but the simplest cases. No split-search algorithm exhaustively examines all possible partitions. Instead, various restrictions are imposed to limit the possible splits to consider. The most common restriction, and the one used in JMP, is to look at only binary splits.

Other restrictions involve binning continuous inputs, stepwise search algorithms, and sampling.

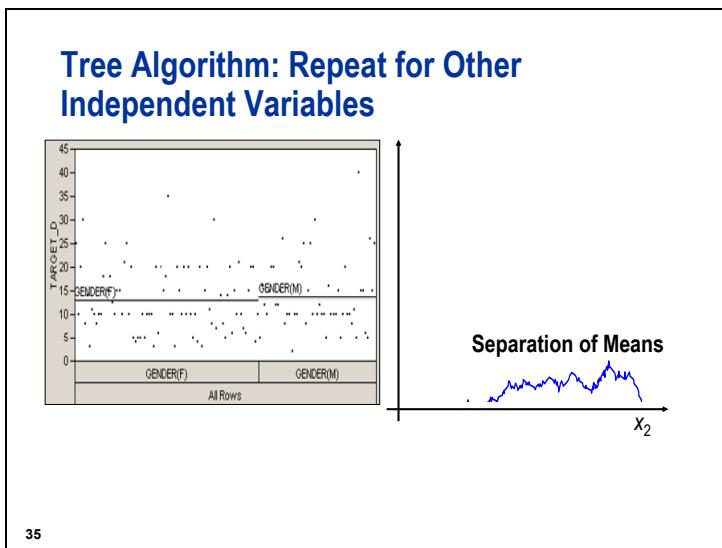
When Y is continuous, the Partition platform fits means and creates binary splits that most significantly separate the means. The graph on the right shows the separation of means for all possible binary splits of the dependent variable values using values of X_1 . The graph goes from using the lowest values of X_1 on the left side of the graph to the highest values on the right side of the graph for the splits.

For continuous responses, the criterion is the Sum of Squares. That is, this can be thought of as maximizing the difference between the means.

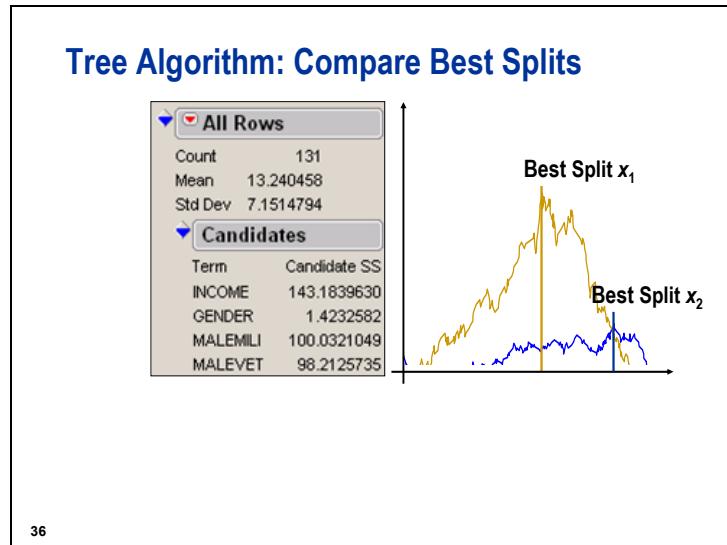


After all of the possible binary splits have been calculated, the split that creates the largest separation between the two group means is chosen as the best split for that variable.

In this case, the continuous response was the amount of money that a person donated to a nonprofit organization. The first independent variable considered was the income level of the household. **INCOME** is an ordinal variable with levels 0-9. All of the possible binary splits using income levels are considered. The income level that creates the largest mean separation is chosen as the best split for that variable. In this example, the best split is at 5.

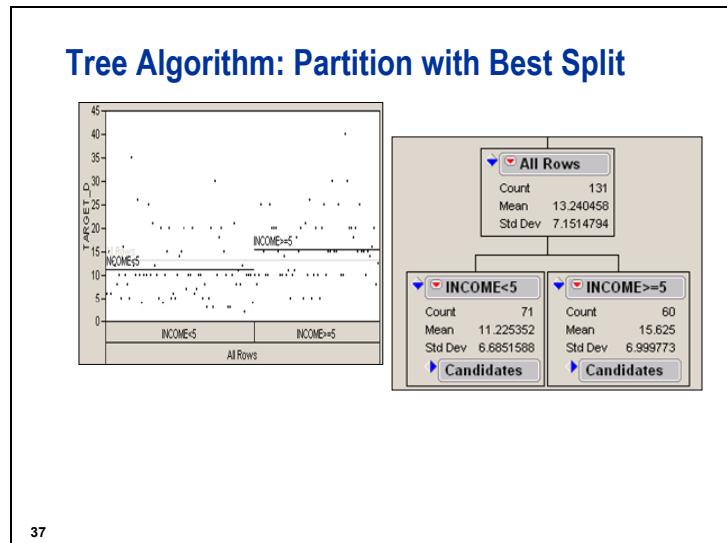


The same process is repeated for each of the other independent variables (Xs). For example, **GENDER** is considered next.

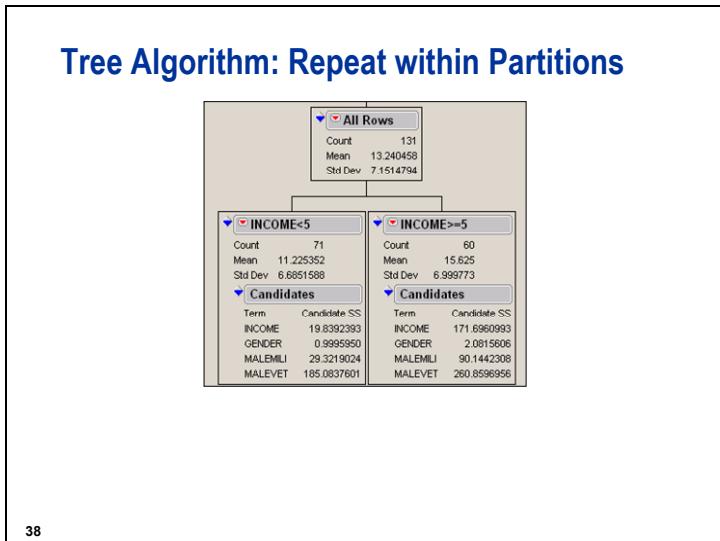


After the best split is determined for each of the independent variables, compare these best splits. In the general case depicted in the graph on the right, the best split for X_1 results in a greater separation of the means than the best split for X_2 .

In JMP, there is a report of the best split values for each independent variable. These values are in the Candidates report in the column labeled Candidate SS. In the example, the best split for **INCOME** results in a greater separation of the means than does any other best split.

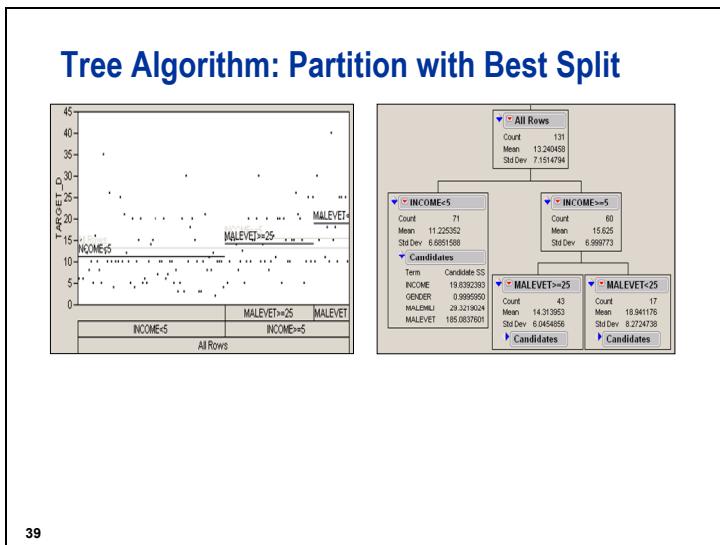


Because the split for **INCOME** at a value of 5 provides the greatest separation of the two group means, the first split of the data is done using **INCOME<5** and **INCOME>=5**.

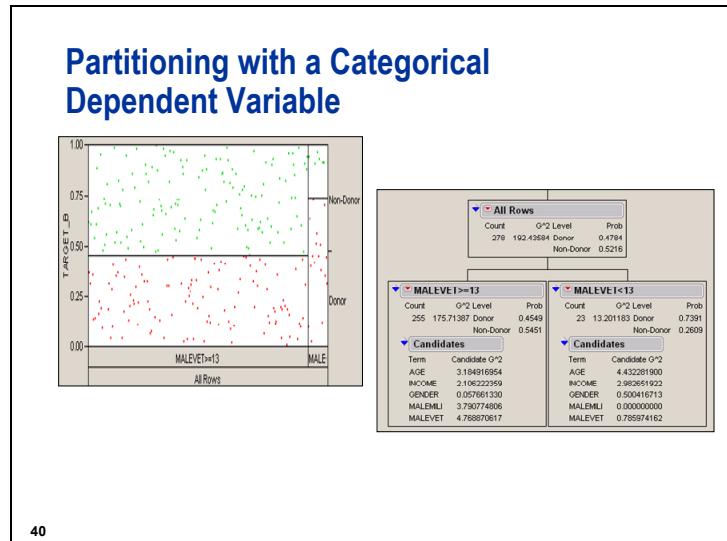


Now the process begins again within each partition. Within each partition, each possible binary split is examined for the greatest separation of the two group means. This is done for each independent variable. Then, the best splits for all of the independent variables are compared within each partition. Finally, the best splits for each of the two partitions are compared.

In JMP, there are Candidate reports within each partition.



Because the split for **MALEVET** provides the greatest separation of the two group means, the next split of the data is done using **MALEVET**<25 and **AGE** >=25. (**MALEVET** is the percent of household male veterans from census data.)



40

For categorical responses, the criterion is the likelihood-ratio chi-square, which is reported as G^2 in the report. This is actually twice the [natural log] entropy or change in the entropy. *Entropy* is a measure of variability for categorical data. It measures the uncertainty of the outcome. Entropy is $\Sigma -\log(p)$ for each observation, where p is the probability attributed to the response that occurred.

- ✍ When the Partition platform calculates a G^2 on excluded data for a categorical response, it uses the rate value $0.25/m$ when it encounters a zero rate in a group with m rows. Otherwise, a missing statistic would be reported because the logarithm of zero is undefined.

Tree Algorithm Parameters	
	Maximum Branches
	Split Worth Criterion
	Stopping Options
	Pruning Method
	Missing Value Method
<i>Default Settings</i>	
	2
	Variance Reduction
	Entropy
	Interactive
	Interactive
	Closest

41

In summary, the Partition platform in JMP

- forms binary splits.
- determines splits with variance reduction for continuous dependent variables (variance reduction is the name for the maximum separation between group means) and with entropy for categorical dependent variables.
- does not use a stopping rule because currently, the platform is purely interactive. Continue clicking the **Split** button until the result is satisfactory.
- prunes by combining the most recent split back into one group. This is done interactively.
- places rows with missing values into the group that it best fits as determined by its Y values. This is the Closest option for the Missing Value Rule.

 If the Y is missing, the row is simply ignored. If an X is missing, the row is currently ignored for calculating the criterion values, but then the row becomes part of the group where it best fits, determined by Y's value. The Missing Value Rule Random option changes this to a random group rather than a best-fitting group.

3.5 Chapter Summary

Histograms, scatterplots, and bar charts are used to graphically assess the relationship between continuous columns.

JMP provides tools to graphically select a subset of data (Lasso), examine in finer detail (Magnifier), and determine horizontal and vertical values (Crosshairs). You can change the format of any JMP graph by selecting an axis.

The relationship or association between attribute data columns can be examined using histograms and mosaic plots.

The Partition platform in JMP

- forms binary splits
- determines splits with variance reduction for continuous dependent variables (that is the name for the maximum separation between group means) and with the likelihood ratio chi square test for categorical dependent variables
- does not use a stopping rule because the platform is purely interactive
- prunes by combining the most recent split back into one group
- places rows with missing values into the group that it best fits as determined by its Y values.

Chapter 4 Reporting and Presenting Results

4.1 Saving Results	4-3
4.2 Creating a Project Journal.....	4-8
4.3 Creating a New Project	4-15
4.4 Chapter Summary.....	4-21

4.1 Saving Results

Objectives

- Save results from analyses into a JMP journal.
- Save the JMP journal as an .HTM file.

3

Save the results from analyses into a special kind of JMP file called a *journal*, where you can choose to save its contents in different types of formats.

Journals



4

A *journal* is a static copy of a JMP report that enables users to easily share results, as well as data, with each other. Journals can be saved in a variety of file formats.



Using Saved Scripts to Generate Results

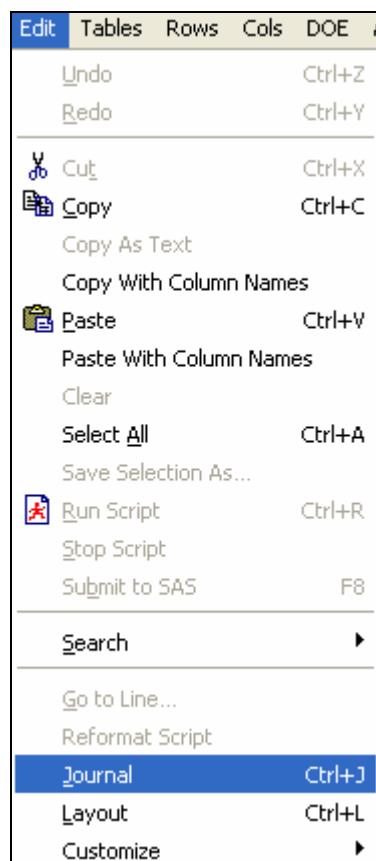
This demonstration uses saved JSL scripts to explore data, saves the results to a JMP journal file, and saves a journal in different formats.

The JMP Journal

You have created many windows, full of analyses and endless possibilities for exploration using JMP and its powerful interactive capabilities.

Suppose you want to collect two analyses in one window for possible printing, distribution, or publication. A JMP journal provides this type of functionality.

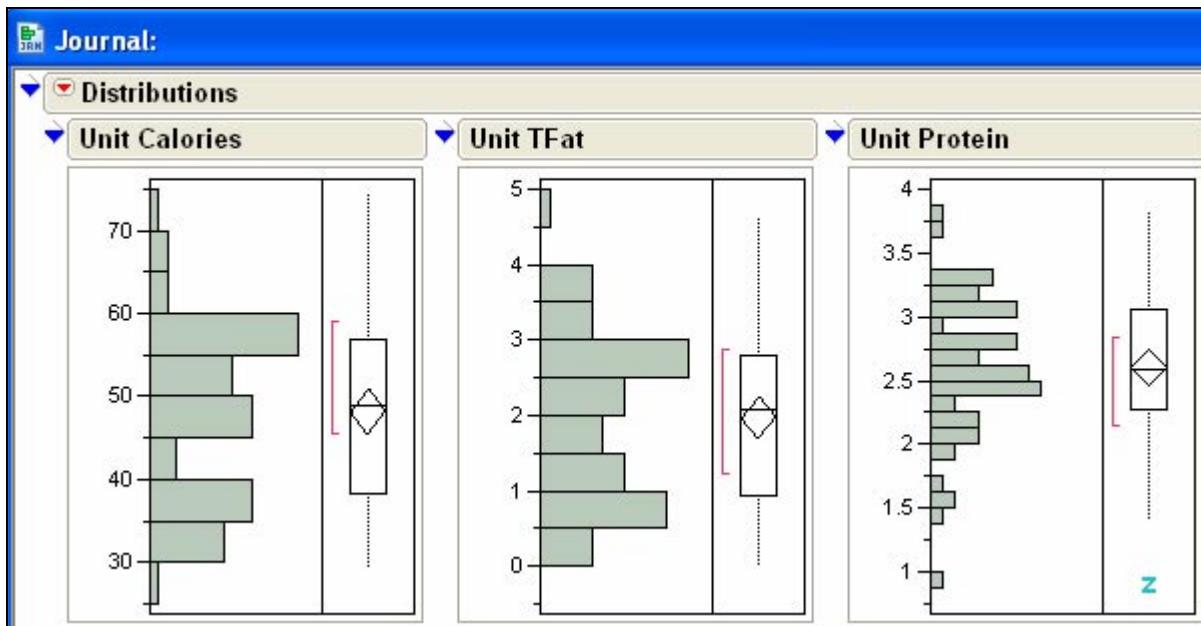
- ✍ A script can be run to generate the analysis, but a journal is a collection of already generated analyses.
1. Open **Sandwiches Unit Based.JMP**.
 2. Click the red triangle next to **Distribution** in the Table panel and select **Run Script**.
 3. Select **Edit** ⇒ **Journal**.





The Layout platform offers you many of the same advantages as the journal but you can also change the arrangement of the results into any layout that you want. You can use special tools to drag, group or ungroup, and align results as objects.

JMP opens a new window, Journal:, which has the same content as the Distribution window.

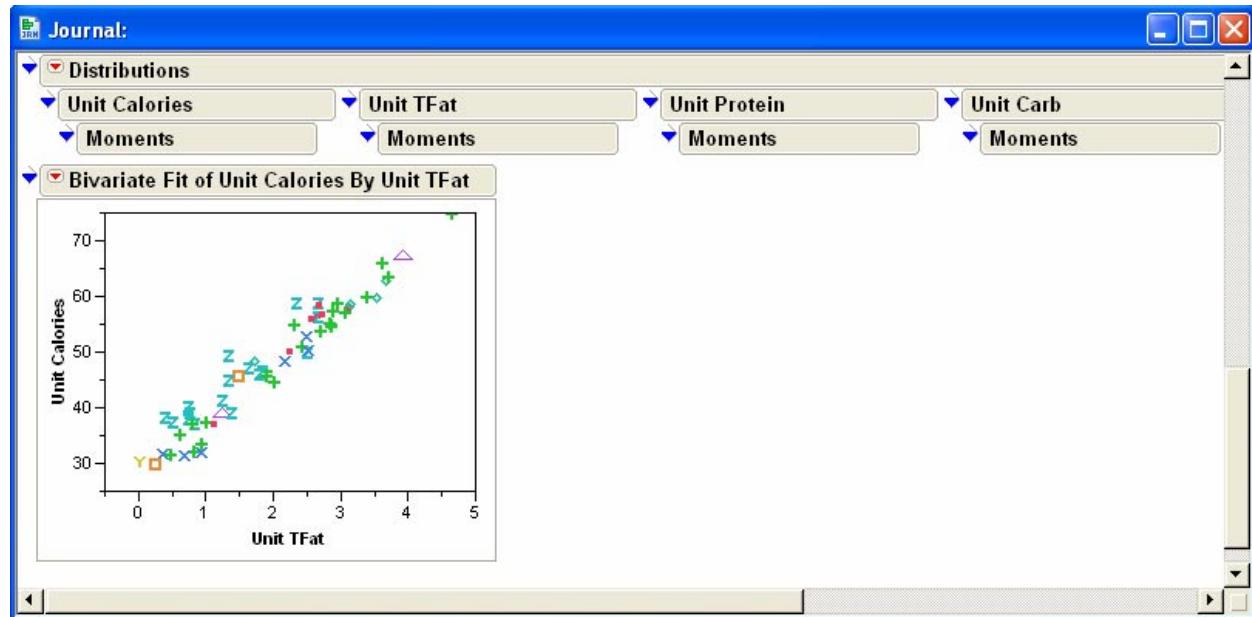


The journal file acts as a snapshot of the chosen analysis. You can still scroll within the window and collapse analysis results using the blue triangle, but you cannot click in any of the bars as you could, for example, in the Distribution window. Journals are not interactive with the data table.

Now add the scatterplot to the same Journal window.

4. Select **Window** \Rightarrow **Sandwiches Unit Based**.
5. Click the red triangle next to **Bivariate** in the Table panel and select **Run Script**.
6. Select **Edit** \Rightarrow **Journal**.

7. Select **Window** \Rightarrow **Journal**: Scroll to the bottom of the window.



Data tables can also be added to the journal. The output data in the journal can be converted back to a JMP data table by right-clicking and selecting **Make Into Data Table**.

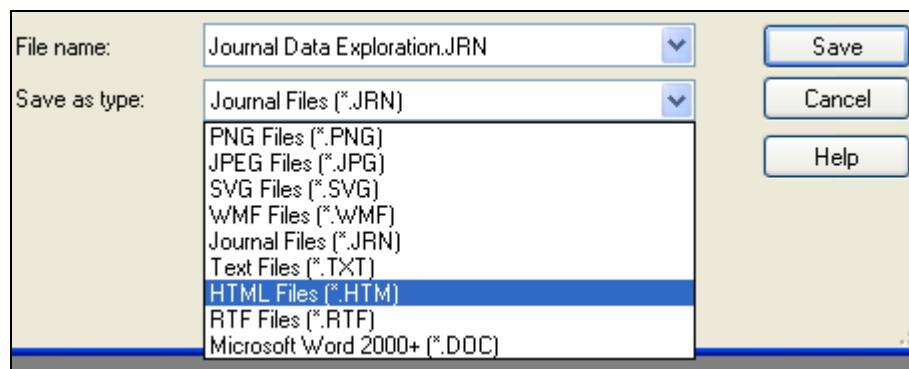
8. Select **File** \Rightarrow **Save As...**
9. Type **Journal Data Exploration** and select **Save**.
10. Select **Window** \Rightarrow **Sandwiches Unit Based - Distribution**.
11. Select **Window** \Rightarrow **Close All Reports**.

Saving a JMP Journal as an HTML File

The journal file, **Journal Data Exploration.JRN**, now contains the results of two analyses.

Suppose you want to publish these results in HTML format for distribution within your organization and save the results to be included in a professional journal.

1. Select **File** \Rightarrow **Save As...**. Select **HTML Files (*.HTM)** in the **Save as type** field.



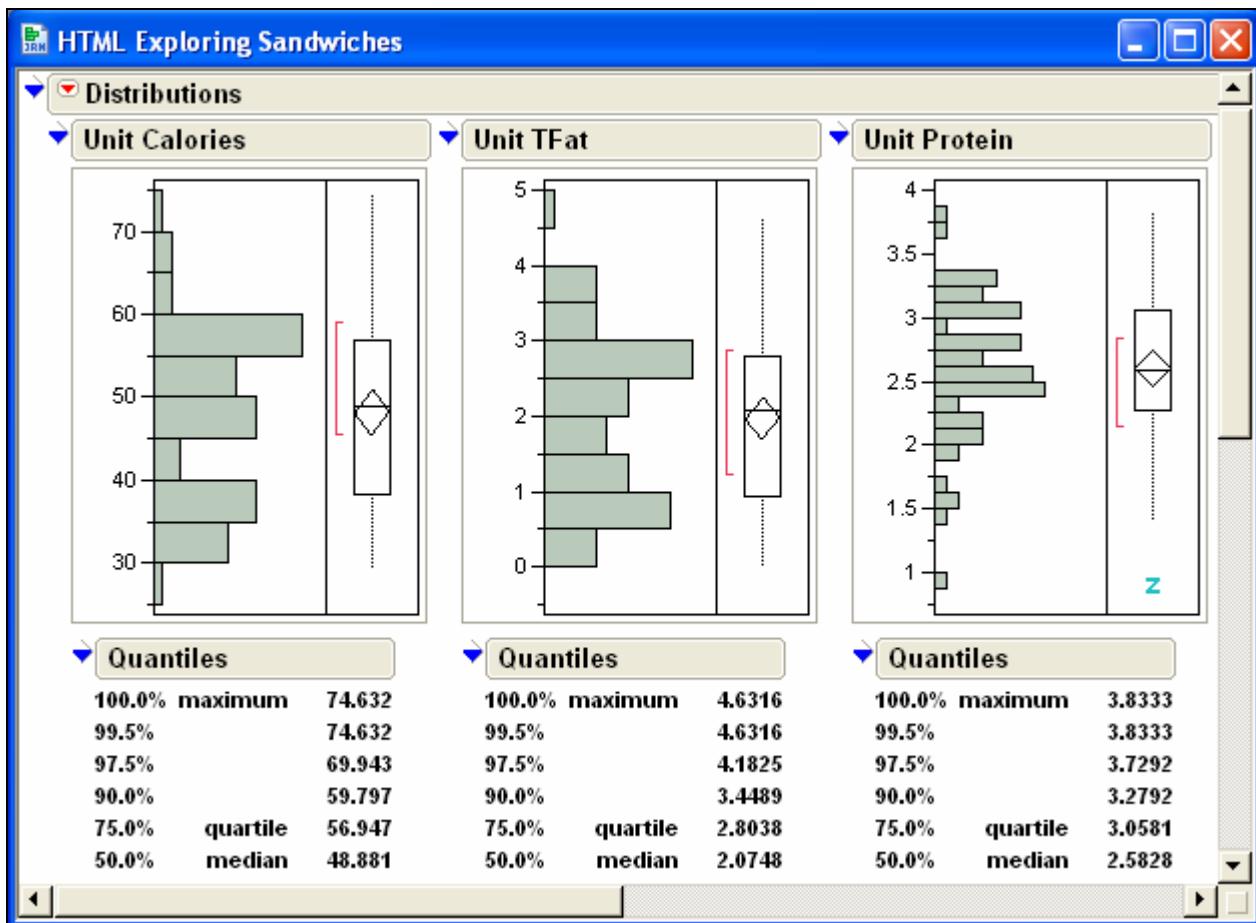
Notice that JMP automatically uses the same name as the journal file but adds the .HTM extension.

- Type **HTML Exploring Sandwiches** and select Save.



Notice that the default value (**PNG**) for Graphics File Format was used. The file name was chosen for future data management issues.

- JMP automatically displays the saved file.



- Close all reports and data tables.

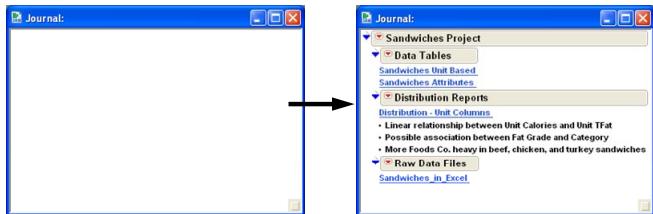
4.2 Creating a Project Journal

Objectives

- Create a project journal.
- Add open windows to a journal.

11

Project Journals



12

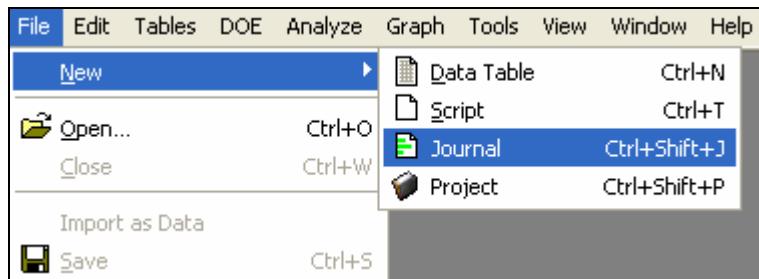
The New Journal feature in JMP can be used to package the many elements of a project together. Tables can be opened, scripts can be run, and the Internet can be accessed from the journal.



Creating a Project Journal

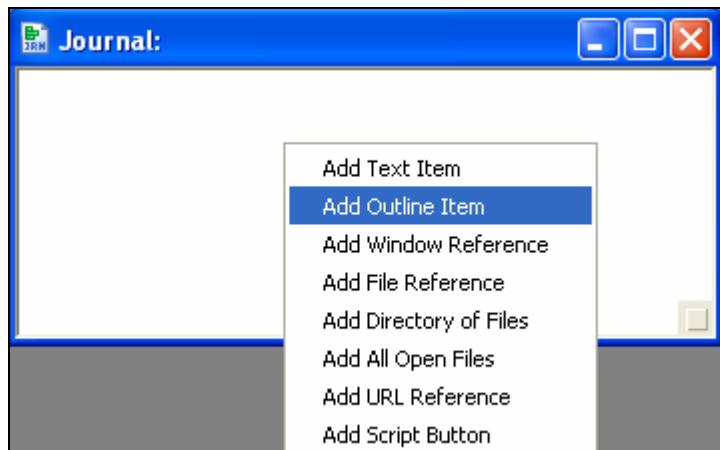
A journal can be built from scratch, starting with an empty Journal window and adding items to the window, or the journal can be created after data analysis is in progress. In this demonstration, a journal is built from scratch.

1. Select **File** \Rightarrow **New** \Rightarrow **Journal**.



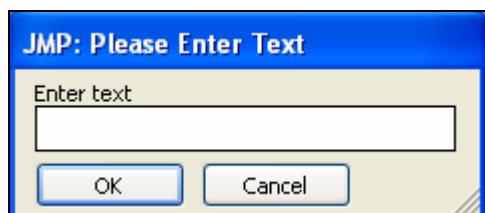
An empty Journal window opens.

2. Right-click in the empty Journal window.



A list of items that can be added to the journal is displayed. The chosen item is placed in the upper-left side in the Journal window.

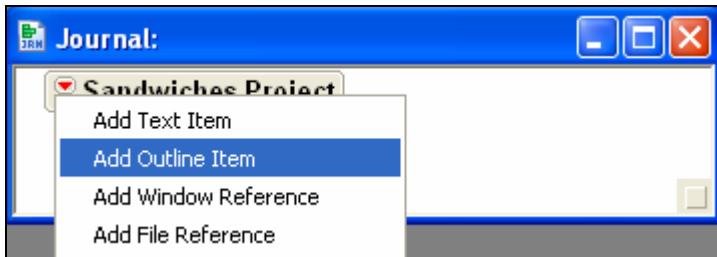
3. Select **Add Outline Item**.



4. Type **Sandwiches Project** and select **OK**.



5. Click the red triangle next to **Sandwiches Project** and select **Add Outline Item**.



6. Type **Data Tables** and select **OK**.



- After items are nested under an outline box, the blue expand/collapse icon appears next to the outline box.
7. Click the red triangle next to **Data Tables** and select **Add File Reference**.
8. Select [Sandwiches Unit Based.JMP](#) \Rightarrow [Open](#).

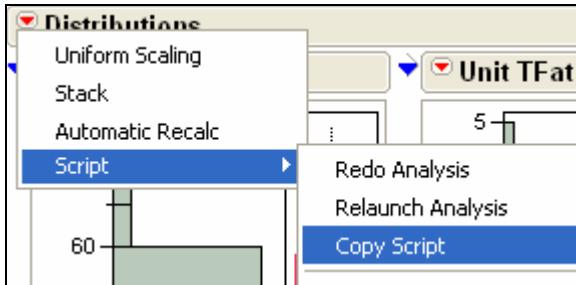


A link to the data table has been placed in the journal. Follow the same steps to add a link to one more data table.

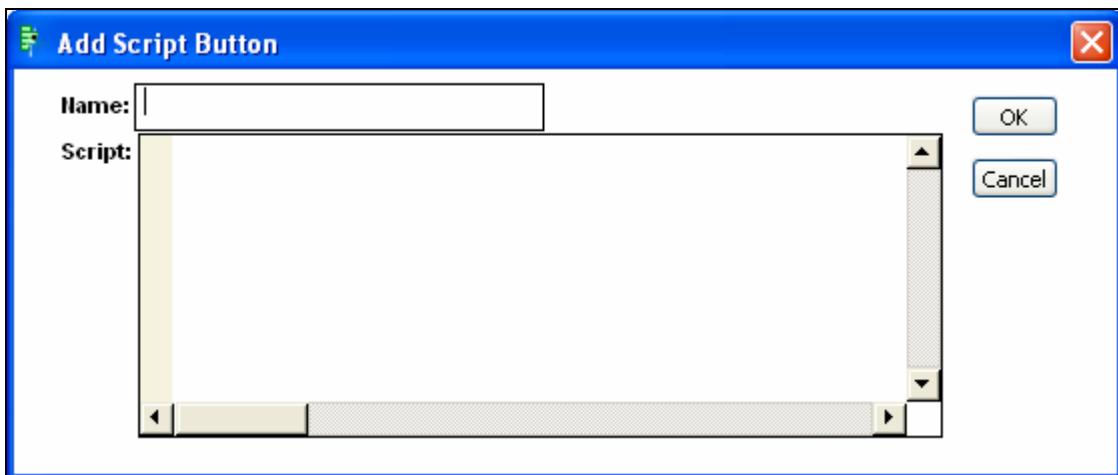
9. Click the red triangle next to **Data Tables** and select **Add File Reference**.
10. Select [Sandwiches Attributes.JMP](#) \Rightarrow [Open](#).

Scripts generated during the analysis of this data can also be added to the journal.

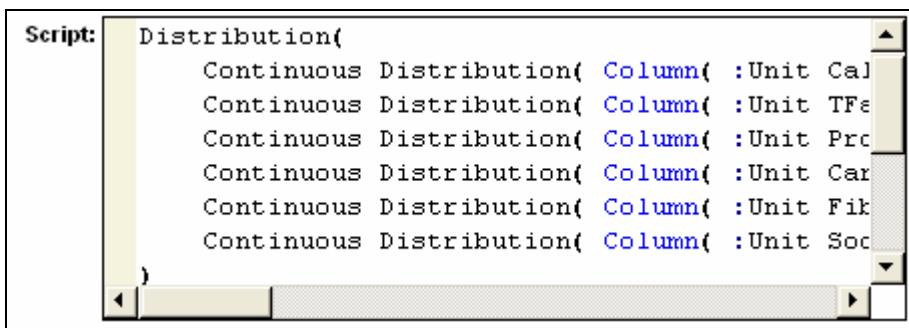
1. Select **Sandwiches Unit Based** to open the data table.
2. Click the red triangle next to **Distribution** in the Table panel and select **Run Script**.
3. Click the red triangle next to **Distributions** and select **Script** \Rightarrow **Copy Script**.



4. Select **Window** \Rightarrow **Journal**:
5. Click the red triangle next to **Sandwiches Project** and select **Add Outline Item**.
6. Type **Distribution Reports** in the text box and select **OK**.
7. Click the red triangle next to **Distribution Reports** and select **Add Script Button**.



8. Right-click in the Script box and select **Paste**.

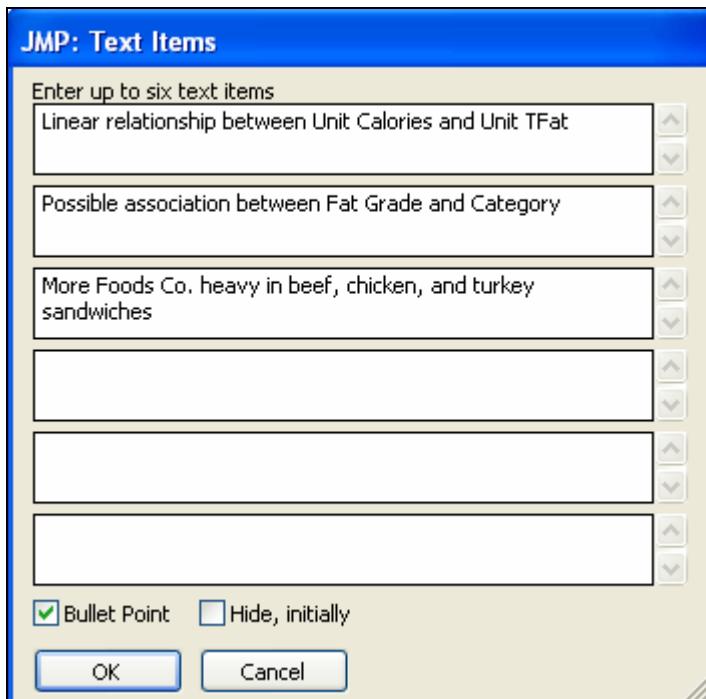


9. Type **Distribution - Unit Columns** in the box next to Name and select **OK**.



Text notes can also be added to a journal.

1. Click the red triangle next to **Distribution Reports** and select **Add Text Item**.
2. Type **Linear relationship between Unit Calories and Unit TFat** in the first text box.
3. Type **Possible association between Fat Grade and Category** in the second text box.
4. Type **More Foods Co. heavy in beef, chicken, and turkey sandwiches** in the third text box.
5. Select **Bullet Point**.



6. Select **OK**.



Links to spreadsheets used to create JMP tables can also be added.

1. Click the red triangle next to **Sandwiches Project** and select Add Outline Item.
2. Type **Raw Data Files** and select OK.
3. Click the red triangle next to **Raw Data Files** and select Add File Reference.
4. Change Files of Type to **Excel 97-2003 Files (*.xls)**. Select Sandwiches in Excel.xls \Rightarrow Open.

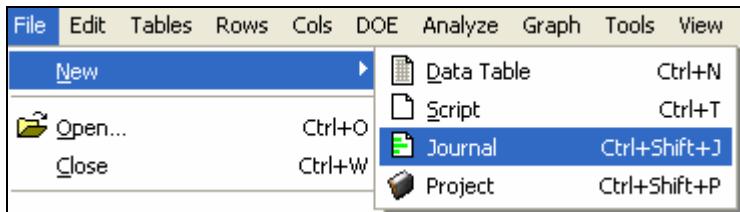


After the journal is complete, it can be saved.

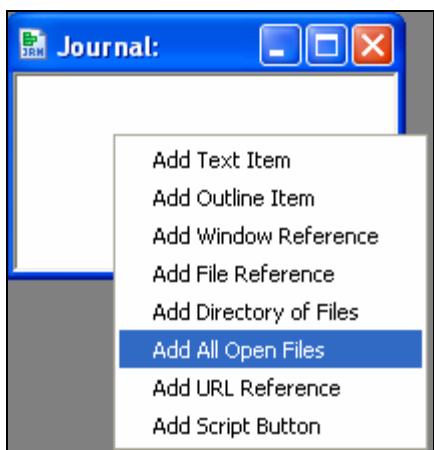
5. Select File \Rightarrow Save As... and type **Sandwich Project** as the name.
 6. Select Save.
- Unless the user modifies the scripts saved to this journal, the user must select a data table and then select the script(s) that corresponds to the analyses for that specific table. For example, to view the distribution report for the unit columns, the user will select Sandwiches Unit Based and then select Distribution – Unit Columns.
7. Click the **X** in the upper-right corner of the Sandwich Project window to close the window.

Project journals can also be created spontaneously. Links to open tables and report windows can be quickly added to a journal any time that a user is working in JMP.

1. Select **File** \Rightarrow **New** \Rightarrow **Journal**.



2. Right-click in the empty Journal window and select **Add All Open Files**.



Links to the open data table and output generated by the script are automatically added to the journal.

3. Close all windows without saving.

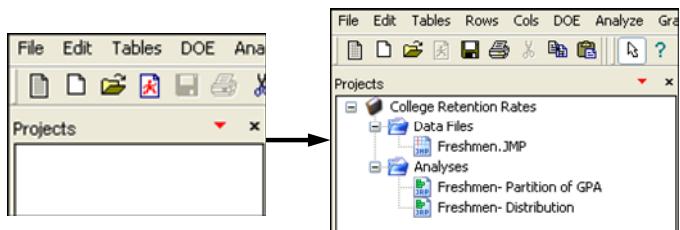
4.3 Creating a New Project

Objectives

- Create a JMP project.
- Examine options available for a JMP project.

19

JMP Projects



20

JMP can package the many elements of data analysis together in a JMP project. A JMP project enables you to store data tables, reports, journals, scripts, and so forth, in one file. By default, a project contains links to its contents. You can also save a project so it contains copies of the files within it, making it possible to share a project with others.

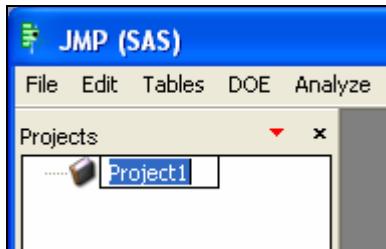
When a JMP project is created, it contains everything needed to re-open all of the included files. This is a major difference from saved session scripts. The scripts save the state of JMP, while the projects save the state along with the files.



Creating a Project

This demonstration illustrates creating a JMP project.

1. Select **File** \Rightarrow **New** \Rightarrow **Project**.

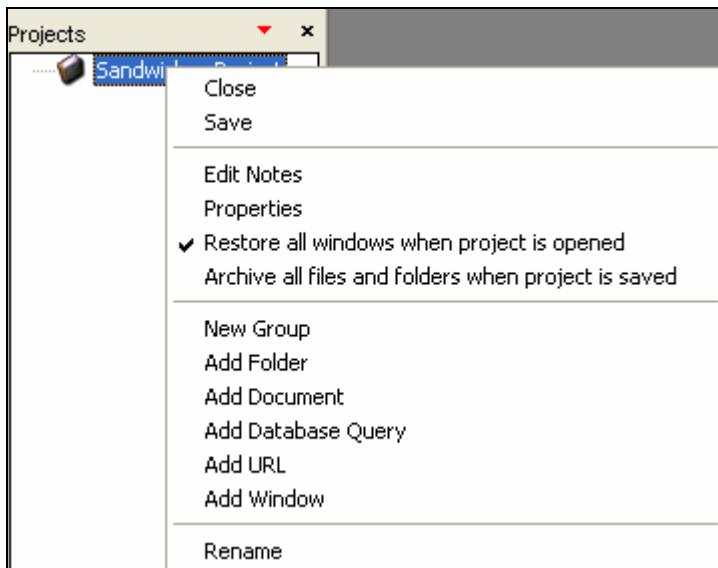


The Projects window appears on the left side of the JMP application window.

2. Type **Sandwiches Project** and press ENTER.



3. Right-click **Sandwiches Project**.



The Project menu opens and shows options for the project.

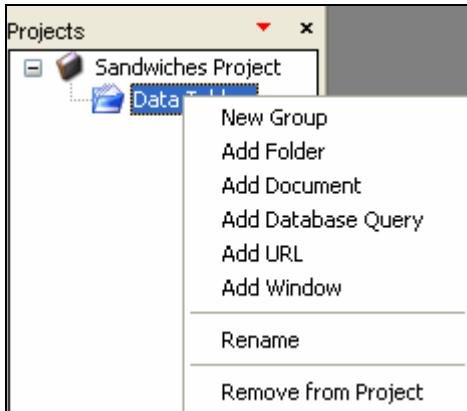
4. Select **New Group**.



5. Type **Data Tables** and press ENTER.



6. Right-click **Data Tables**.



A menu opens and shows options for the group.

7. Select **Add Document**.

8. Select **Sandwiches Unit Based.JMP** \Rightarrow **Open**.



Items can also be added to a project using the drag-and-drop method from the Window list, the File System list, or from Microsoft Outlook.

Now add a link to **Sandwiches Attributes**.

9. Right-click **Data Tables** and select **Add Document**.

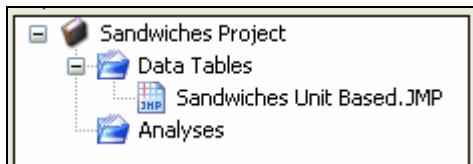
10. Select **Sandwiches Attributes.JMP** \Rightarrow **Open**.



Now add a new group where you will store reports.

11. Right-click **Sandwiches Project** and select **New Group**.

12. Type **Analyses** and press ENTER.



Generate the analyses to place in the project.

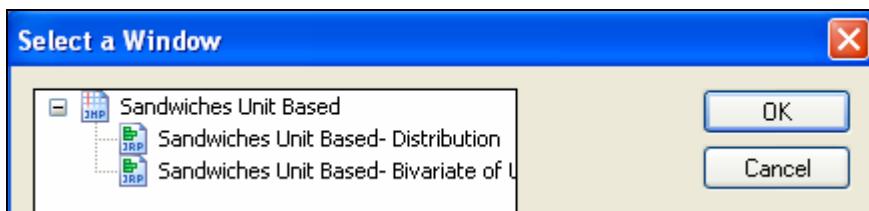
13. In the JMP project, double-click **Sandwiches Unit Based.JMP** to open the file.

14. Click the red triangle next to **Distribution** and select **Run Script**.

15. Select **Window** \Rightarrow **Sandwiches Unit Based**.

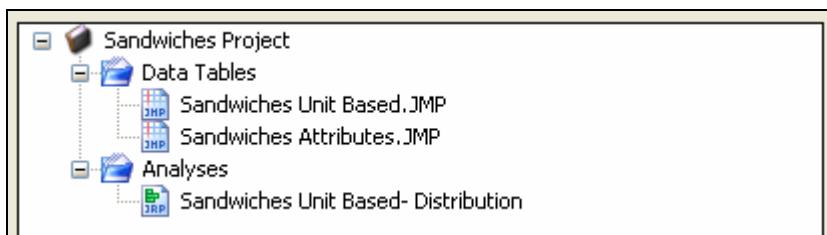
16. Click the red triangle next to **Bivariate** and select **Run Script**.

17. Right-click **Analyses** and select **Add Window**.



A list of the open windows appears.

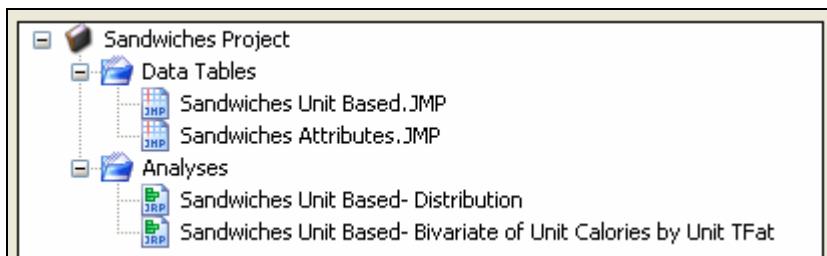
18. Select **Sandwiches Unit Based – Distribution** \Rightarrow **OK**.



A link to this report has been added to the project. Now add a link to the scatterplot.

19. Right-click **Analyses** and select **Add Window**.

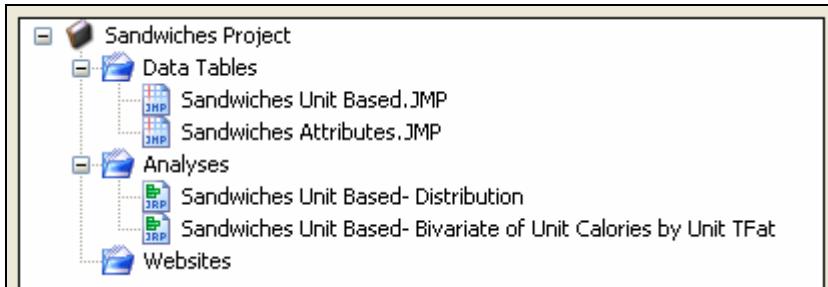
20. Select **Sandwiches Unit Based – Bivariate of Unit Calories by Unit TFat** \Rightarrow **OK**.



Now add a new group where you will store Websites.

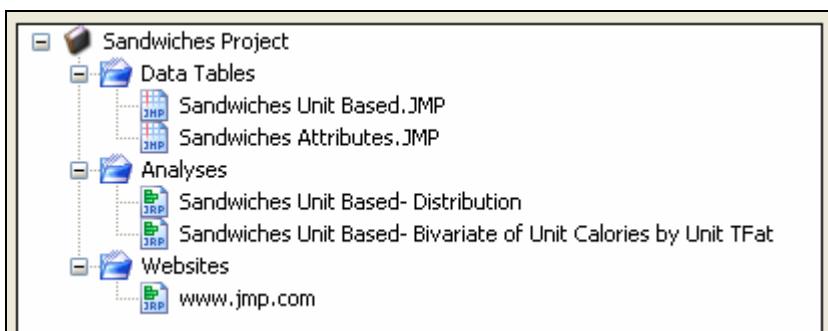
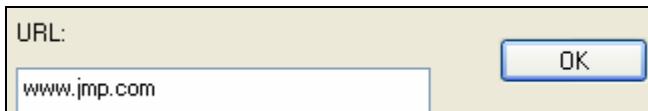
21. Right-click **Sandwiches Project** and select **New Group**.

22. Type **Websites** and press ENTER.



23. Right-click **Websites** and select **Add URL**.

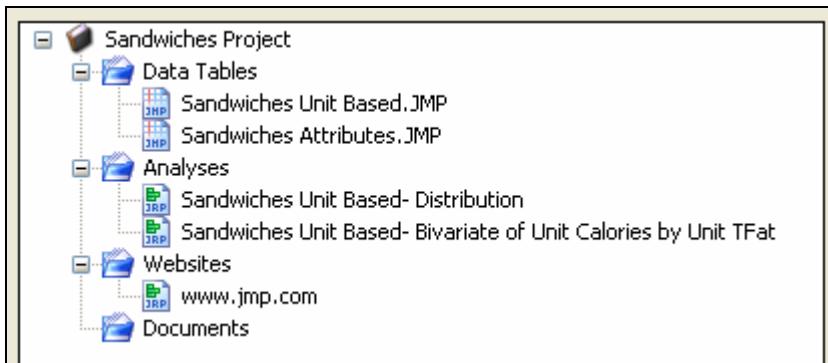
24. Type **www.jmp.com** in the URL window and select **OK**.



Now add a new group where you will store documents.

25. Right-click **Sandwiches Project** and select **New Group**.

26. Type **Documents** and press ENTER.



27. Right-click **Documents** and select **Add Document**.

28. Select Sandwiches Project.doc \Rightarrow Open.

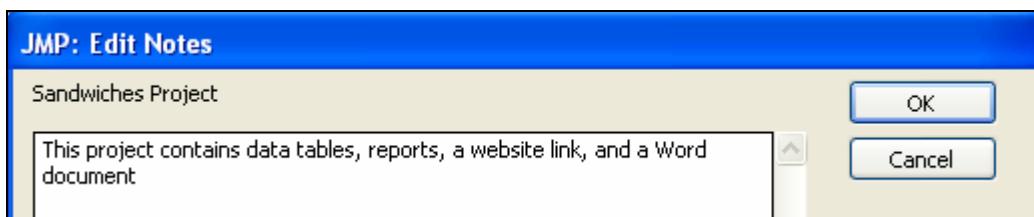
A link to the Word document Sandwiches Project has been added to the project.



Now add a note to the project.

29. Right-click Sandwiches Project and select Edit Notes \Rightarrow Yes.

30. Type **This project contains data tables, reports, a website link, and a Word document.** and select OK.



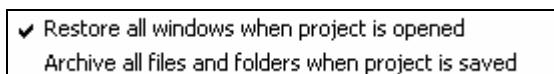
Notes can be viewed by right-clicking and selecting either Edit Notes or Properties.

The JMP project is complete. Save the project.

31. Right-click Sandwiches Project and select Save \Rightarrow Save.

Finally, examine the remaining properties of the project.

32. Right-click Sandwiches Project.



The property **Restore all windows when project is opened** is turned on. This means that each item in the project will be opened when the closed project is reopened.

The property **Archive all files and folders when project is saved** is turned off. When a project is saved, links to the items in the projects are saved rather than the items themselves. If this property is selected, then a copy of each item is copied into the project archive. From this point on, if you open the project and modify an item, you are modifying the archived copy rather than the originals.

33. Close all windows.

The remaining options for a JMP project are Add Folder and Add Database Query. Add Folder enables you to add a folder of items to the project, and Add Database Query opens the database query platform, enabling you to generate a query and save it to the project.

4.4 Chapter Summary

Results from JMP analyses can be saved as a journal file, with a default file extension of JRN. Journals can be saved as HTML files and RTF files; the latter is readable by word processing software.

Links to raw data files, JMP tables, and scripts can be added to a new journal to create a project journal. These journals are an efficient way to assemble the data and reports for a project.

Appendix A Exercises and Solutions

Exercises	A-2
Chapter 1 Exercises	A-2
Chapter 2 Exercises	A-2
Chapter 3 Exercises	A-4
Chapter 4 Exercises	A-5
 Solutions to Exercises	 A-6
Chapter 1 Solutions to Exercises	A-6
Chapter 2 Solutions to Exercises	A-12
Chapter 3 Solutions to Exercises	A-23
Chapter 4 Solutions to Exercises	A-31

Exercises

Chapter 1 Exercises

1. Importing an Excel File into JMP

A Microsoft Excel file **Colleges.xls** contains the 1991 graduation rates for the top 200 universities and colleges in the United States determined by *Money* magazine.

- a. Import the Excel file directly into JMP and save it as **Colleges.JMP**. (You **must** save this file in order to use it in the next chapter.)
- b. Use the Database command to bring into JMP only **NAME**, **TYPE**, **RATE**, and **TUITION** as the columns, and only the rows so that **TYPE** = **private**. Do not save this file.

2. Examining Data and Modeling Types, Moving Columns, and Saving a Table Variable

Return to **Colleges.JMP**.

- a. How many rows and columns are in the data table?
- b. What data type is used for each column?
- c. What modeling type is used for each column?
- d. Move **STATE** to the end of the data grid and make **TYPE** the first column.
- e. Create a new table variable **Data Source** with the following value: 1991 graduation rates for the top 200 American universities and colleges determined by *Money* magazine.
- f. Save the changes you have made to **Colleges.JMP**.

Chapter 2 Exercises

Open **Colleges.JMP**. (If you did not save this file previously, then open **Colleges_bkup.JMP**.)

1. Recoding and Logically Selecting Rows

- a. Verify that the values in **TYPE** are coded as **private** and **public**. Use the Recode option to correct any data entry errors discovered in the **TYPE** column.
- b. Select the rows in the data table where the value of **REGION** is **Northeast** and the value of **RATE** is greater than or equal to 90. How many rows are selected?
- c. Save this table if necessary.

2. Creating New Columns

Return to **Colleges.JMP**.

- a. The tuition inflation rate tends to increase about 8% per year. An 8% college inflation rate means that the cost of college doubles every nine years. Create a new column **2000 TUITION**, which gives the tuitions for the year 2000, nine years after this data was collected.
- b. Create a new nominal numeric column **GradRateGE80** to identify those colleges that have at least an 80-percent graduation rate. Notice that this column will contain only two values: 1 if the comparison is true, 0 if false. Use a formula with a conditional clause to compute the value.
- c. Add the value label property to the **GradRateGT80** column to insure that the values displayed in the data table are Greater than/equal to 80 and Less Than 80.
- d. Save this table.

3. Creating a Summary Table

- a. Find the average graduation **RATE** by **REGION**. What is the average for the Central region?
- b. Close this summary table without saving it.

4. Logically Selecting Rows and Creating Subsets

- a. Select the rows in the data table where the value of **TYPE** is public.
- b. How many rows are selected?
- c. Make a subset with the selected rows and save it as **Public Colleges.JMP**. Close the file.

5. Creating a New Table by Combining Two Tables That Have the Same Column

Return to **Colleges.JMP**. Open **Regional Economics.JMP**.

- a. Join **Regional Economics.JMP** to **Colleges.JMP** by matching rows with the same **REGION** value. Name the table **Colleges with Regional Economics.JMP**.
- b. How many rows are in the new data table?
- c. Save the new table as **Colleges with Regional Economics.JMP**.
- d. Close **Colleges with Regional Economics.JMP** and **Regional Economics.JMP**.

6. Using the Tabulate Platform

- a. Use the Tabulate platform to make a pivot table that summarizes the number of colleges and the median **RATE** and **TUITION** by **REGION** and **TYPE**.
- b. How many private colleges are in the Northeast region?
- c. Close the summary table. Do not save.

Chapter 3 Exercises

Open **Colleges . JMP**. (If you did not save this file previously, then open **Colleges_bkup . JMP**.)

1. Using a Scatterplot

- a. Use Fit Y by X to make a scatterplot of **RATE** versus **TUITION**.
- b. Is there any indication of a linear relationship between the two columns? If so, is it positive or negative?

2. Using Colors and Markers

- a. Color rows based on their **REGION** value and include a legend. Are there any patterns in the scatterplot with respect to **REGION**?
- b. Mark rows based on their **TYPE** value and include a legend. Are there any patterns in the scatterplot with respect to **TYPE**?

3. Creating a Bar Chart

- a. Make a chart of the average **RATE** by **REGION**.
- b. Save a script in the data table to reproduce this analysis. Close the chart output.

4. Using the Contingency Platform

- a. Use Fit Y by X to make a contingency table of **TYPE** versus **REGION**.
- b. Does the proportion of public schools appear the same in all regions?

5. Using the Partition Platform

- a. Use the Partition platform to explore **RATE** versus **TYPE**, **TUITION**, and **REGION**.
- b. Use the Candidates report to determine what the first split will be. Explain why.
- c. Make the first split. Where is it made? Describe the two groups.
- d. Use the Candidates report to determine what the next split will be. Explain why.
- e. Make the second split. Where is it made? Describe the two groups.
- f. Close all windows. Save **Colleges . JMP**.

Chapter 4 Exercises

1. Saving Results in Journal Files

Open **Colleges with Regional Economics.JMP**.

- a. Generate a Distribution analysis for **TYPE**, **RATE**, **TUITION**, and **REGION**.
- b. Place the results in a journal.
- c. Open **Colleges.JMP**. Use the previously saved script (from 3.b in Chapter 3 Exercises) to make another chart of the average **RATE** by **REGION**.
- d. Place the results in the journal.
- e. Save the journal as **College Analyses.JRN**. Close the chart and distribution outputs.

2. Saving Journals as HTM or RTF Files

- a. Save **College Analyses.JRN** as an HTML file **College Analyses as HTML.HTM**. View the file, if possible.
- b. Save the journal as an RTF file: **College Analyses as Rich Text.RTF**.
- c. Close **College Analyses as HTML** and **College Analyses as Rich Text**.

3. Creating a Project Journal

A bank initiated a quality improvement project to improve customer satisfaction scores and increase customer retention at a large downtown branch. It was determined that the time that a customer spends waiting in line is correlated to customer satisfaction scores. The quality improvement team collected data before, during, and after the project to establish a baseline customer wait time and to monitor improvement in customer wait time. The data is stored in **Measure - Wait Time Study.JMP**, **Control - wait Time Data.JMP**, and **Completed Wait Time Project.JMP**. Create a project journal that bundles the data and analyses used in this project.

- a. Close all active windows and create a new project journal.
- b. Add an outline item called **Bank Waiting Times Project**.
- c. Under **Bank Waiting Times Project**, add an outline item called **Data Tables**.
- d. Under the **Data Tables** outline item, add links to the three JMP data tables.
- e. Generate distribution reports for the **Time of Day** and **Wait Time** columns in the **Measure - Wait Time Study** table and the **Control - Wait Time Data** table. Store the scripts that create the analysis to the journal under an outline item called **Distribution Reports**. Name the two distribution reports **Distribution - Measure** and **Distribution - Control**, respectively.
- f. Add a link to the journal that connects to the spreadsheet **Waiting Time Study.xls**, which was used to create JMP tables. Add this link under an outline item called **Raw Data Files**.
- g. Save the journal as **Waiting Time Project**.

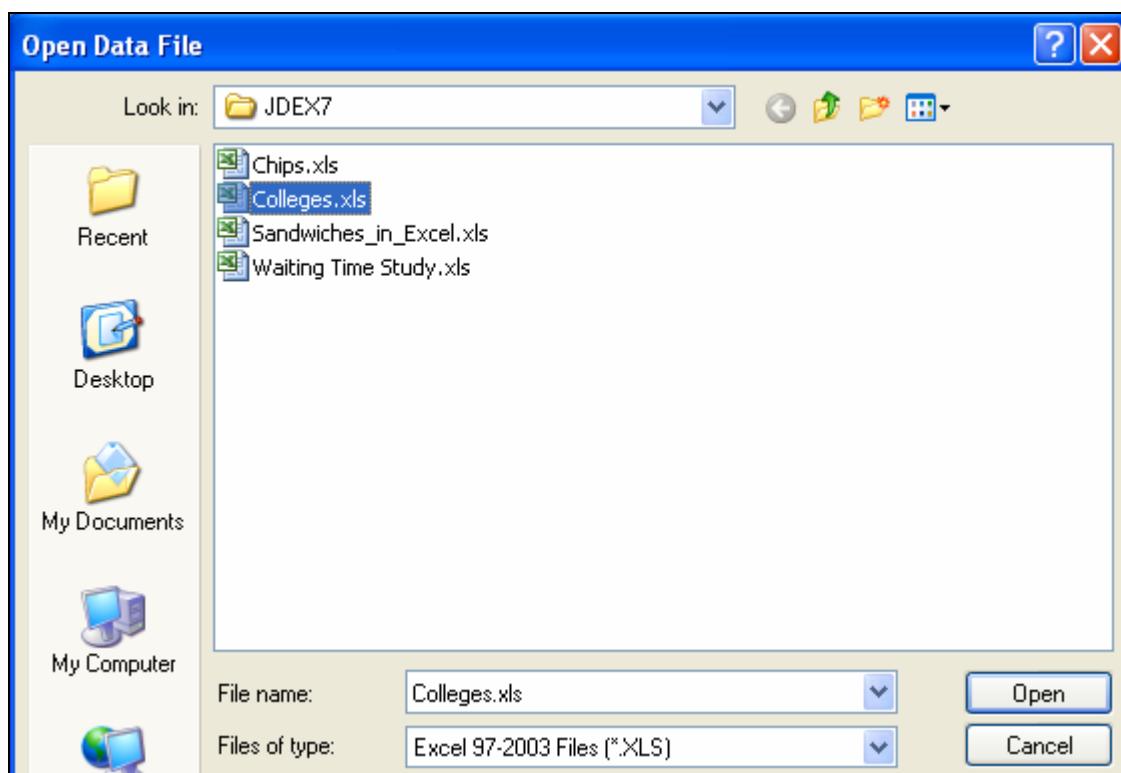
Solutions to Exercises

Chapter 1 Solutions to Exercises

1. Importing an Excel File into JMP

a. Import **Colleges.xls** into JMP and save it as **Colleges.JMP**.

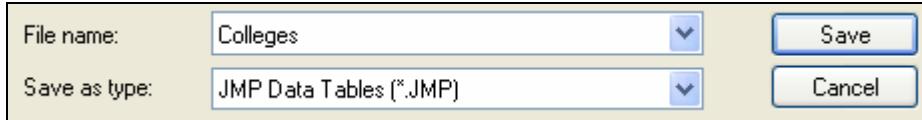
- 1) Select **File** \Rightarrow **Open**.
- 2) Select **Excel 97-2003 Files (*.XLS)** for the files of type.
- 3) Select **Colleges.xls**.



- 4) Select **Open**.

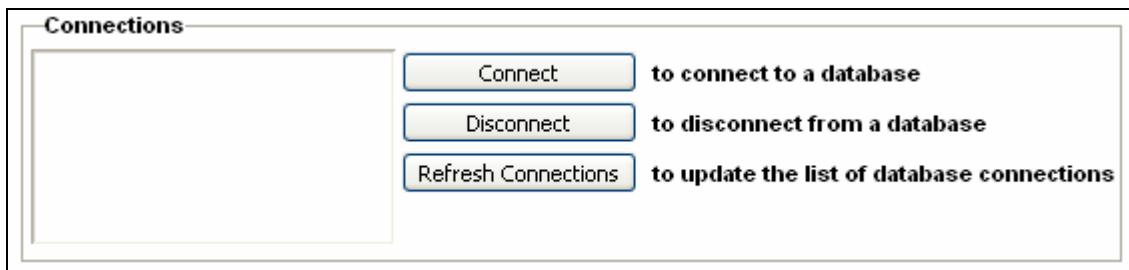
	NAME	STATE	TYPE	RATE	TUITION	REGION
1	Alfred	NY	private	67	6350	Northeast
2	Alvernia	PA	private	70	6202	Northeast
3	American International	MA	private	62	7734	Northeast
4	Atlantic Christian	NC	private	64	5690	South
5	Auburn	AL	public	63	4428	South
6	Austin	TX	private	61	8735	Southwest
7	Barker U.	KS	private	55	6410	Central
8	Barat	IL	private	73	7919	Midwest

- 5) Select **File** \Rightarrow **Save As...**
- 6) Type **Colleges** as the file name.
- 7) Select **Save**.

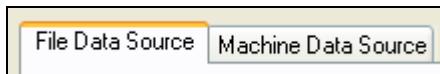


- b. Use the Database command to bring into JMP only **NAME**, **TYPE**, **RATE**, and **TUITION** as the columns, and only the rows so that **TYPE** = private.

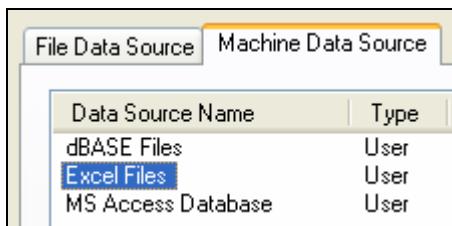
- 1) Select **File** \Rightarrow **Database** \Rightarrow **Open Table...**
- 2) Select **Connect**.



- 3) Select the **Machine Data Source** tab.

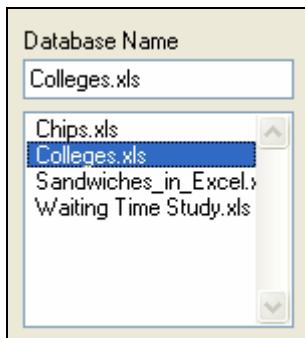


- 4) Under Data Source Name, select **Excel Files**.

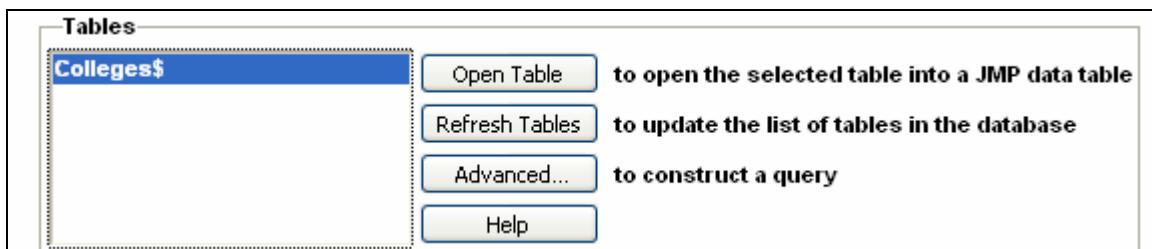


- 5) Select **OK**.

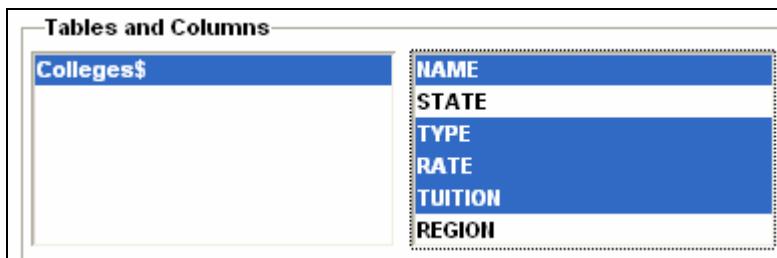
6) Select Colleges.xls \Rightarrow OK.



7) Select Colleges\$ \Rightarrow Advanced....



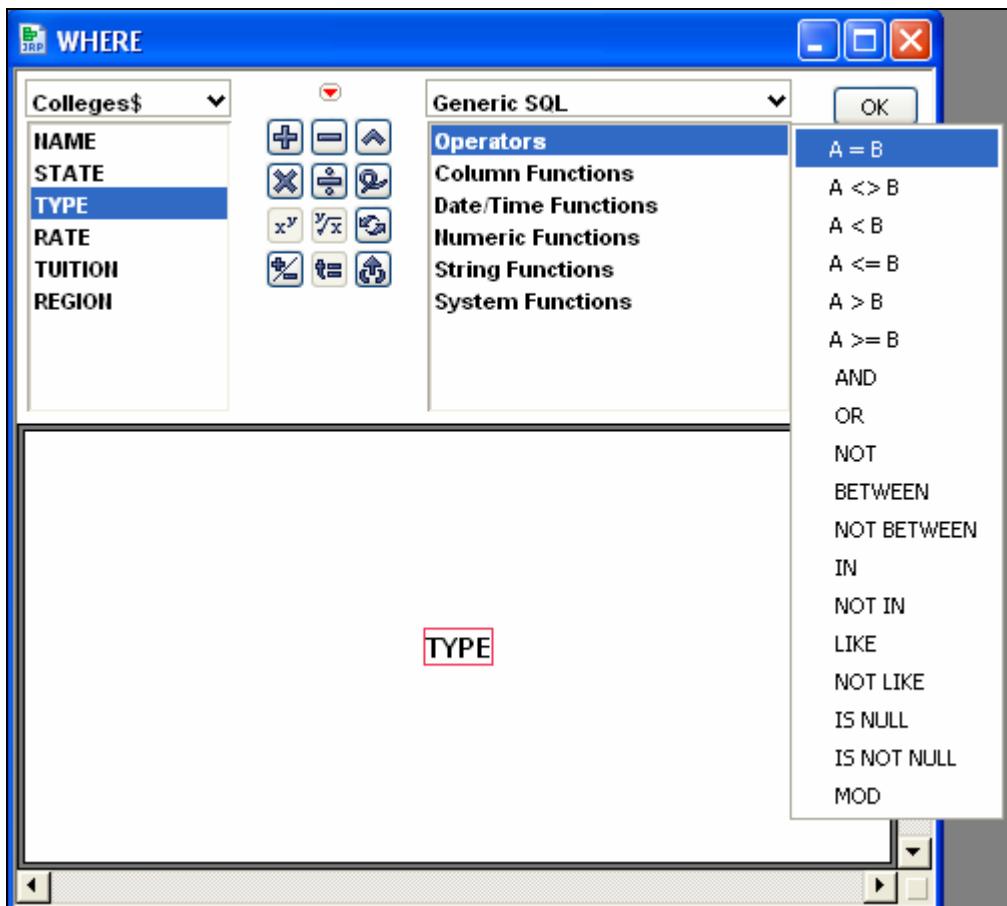
8) While holding down the CTRL key, select STATE and REGION.



9) Select Where....

10) Select TYPE.

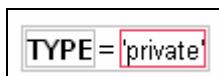
11) Select Operators \Rightarrow A = B.



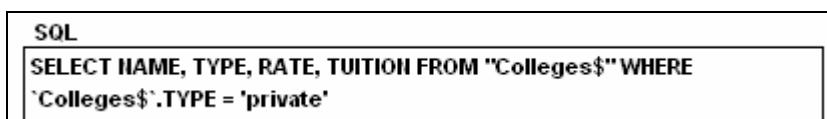
12) Click the red box on the right.



13) Type '**private**' and press ENTER.



14) Select OK.



15) Select Execute SQL.

The JMP data table opens with the specified columns and rows.

	NAME	TYPE	RATE	TUITION
1	Alfred	private	67	6350
2	Alvernia	private	70	6202
3	American International	private	62	7734
4	Atlantic Christian	private	64	5690
5	Austin	private	61	8735
6	Barker II	private	55	6440

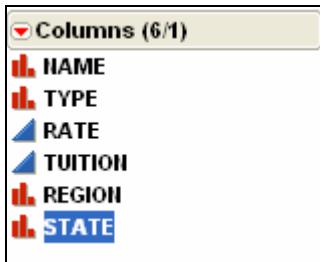
- 16) Close **Colleges\$.JMP**. Do not save.
- 17) Select **Go Back**.
- 18) Select **Disconnect**.
- 19) Select **Yes** in the JMP - Caution Alert window.
- 20) Close the Database Open Table window.

2. Examining Data and Modeling Types, Moving Columns, and Saving a Table Variable

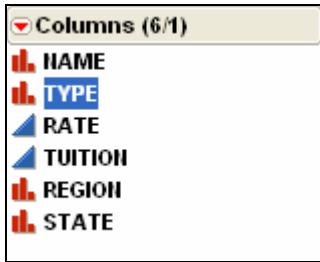
- a. How many rows and columns are in the data table?
 - 1) There are 200 rows, based on information from the Rows panel.
 - 2) There are 6 columns, based on information from the Columns panel.
- b. What data type is used for each column?
 - 1) Based on information from the Columns panel, **NAME**, **STATE**, **TYPE**, and **REGION** are character data columns, and **RATE** and **TUITION** are numeric data columns.
- c. What modeling type is used for each column?
 - 1) Based on information from the Columns panel, **NAME**, **STATE**, **TYPE**, and **REGION** use the nominal modeling type, and **RATE** and **TUITION** use the continuous modeling type.
- d. Move **STATE** to the end of the data grid and make **TYPE** the first column.
 - 1) Select **STATE** in the Columns panel.

NAME
STATE
TYPE
RATE
TUITION
REGION

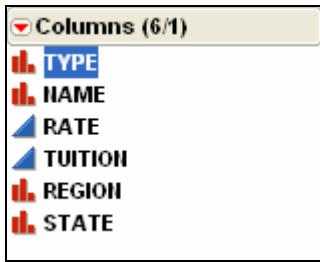
- 2) Drag **STATE** down the list until the solid black line is below **REGION**.



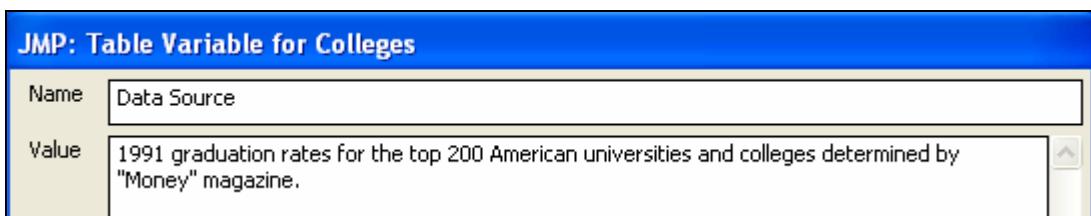
- 3) Select **TYPE** in the Columns panel.



- 4) Drag **TYPE** up the list until the solid black line is above **NAME**.



- e. Create a new table variable **Data Source** with the following value: 1991 graduation rates for the top 200 American universities and colleges determined by *Money* magazine.
- 1) Click the red triangle in the Table panel and select **New Table Variable**.
 - 2) Type **Data Source** as the name.
 - 3) Type **1991 graduation rates for the top 200 American universities and colleges determined by "Money" magazine.** as the value.



- 4) Select **OK**.
- f. Save the changes you made to **Colleges.JMP**.
- 1) Select **File** ⇒ **Save**.

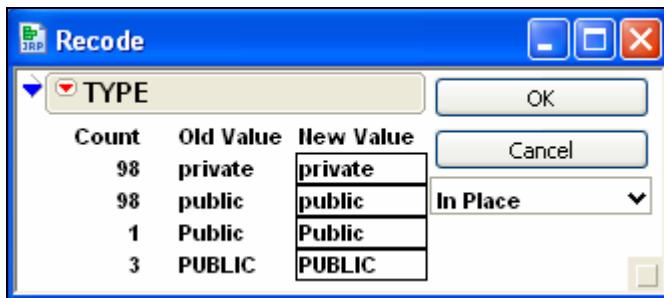
Chapter 2 Solutions to Exercises

Open **Colleges.jmp**. (If you did not save this file previously, then open **Colleges_bkup.jmp**.)

1. Recoding and Logically Selecting Rows

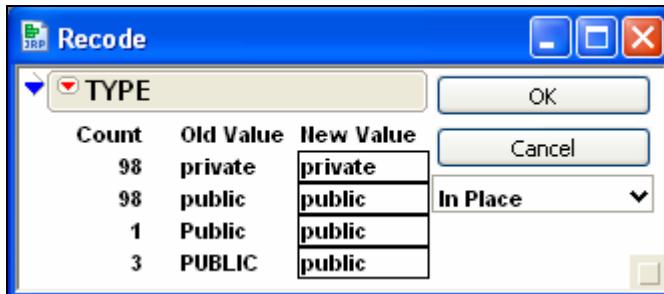
- a. Verify that the values in **TYPE** are coded as **private** and **public**. Use the Recode option to correct any data entry errors discovered in the **TYPE** column.

- 1) Select **TYPE**.
- 2) Select **Cols** \Rightarrow **Recode**.



Values have been entered incorrectly in 4 rows. Corrections can be made by typing the correct values in the New Value column or by copying and pasting the correct values over the incorrect values.

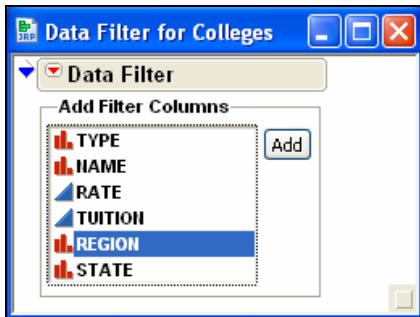
- 3) Change **PUBLIC** to **public**.
- 4) Change **Public** to **public**.



- 5) Select **OK**.
- 6) Select **Continue**.
- 7) Select **File** \Rightarrow **Save**.

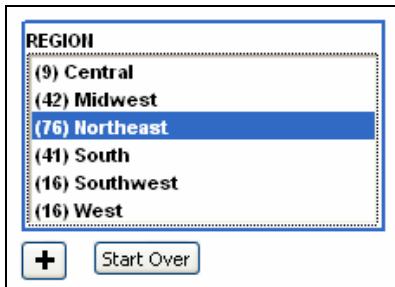
- b. Select the rows in the data table where the value of **REGION** is Northeast and the value of **RATE** is greater than or equal to 90. How many rows are selected?

- 1) Deselect all columns.
- 2) Select **Rows** \Rightarrow **Data Filter**.
- 3) Select **REGION** \Rightarrow **Add**.

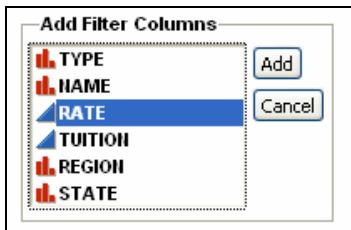


4) Select Northeast.

5) Select .



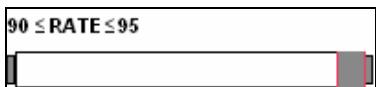
6) Select RATE \Rightarrow Add.



7) Select 28 in the inequality above the slider bar.

 $28 \leq \text{RATE} \leq 95$

8) Type **90** and press ENTER.

 $90 \leq \text{RATE} \leq 95$

 Alternately, the slider can be used to adjust the range of values in the inequality.

9) Close the Data Filter window.

Examine the Rows panel. Four rows are selected.



All rows	200
Selected	4

10) Deselect all rows.

c. Save this table if necessary.

1) Select **File** \Rightarrow **Save**.

2. Creating New Columns

Return to **Colleges.JMP** file.

a. The tuition inflation rate tends to increase about 8% per year. An 8% college inflation rate means that the cost of college doubles every nine years. Create a new column **2000 TUITION**, which gives the tuitions for the year 2000, nine years after this data was collected.

1) Select **Cols** \Rightarrow **New Column...**.

2) Type **2000 TUITION** as the name.

3) Select **Column Properties**.

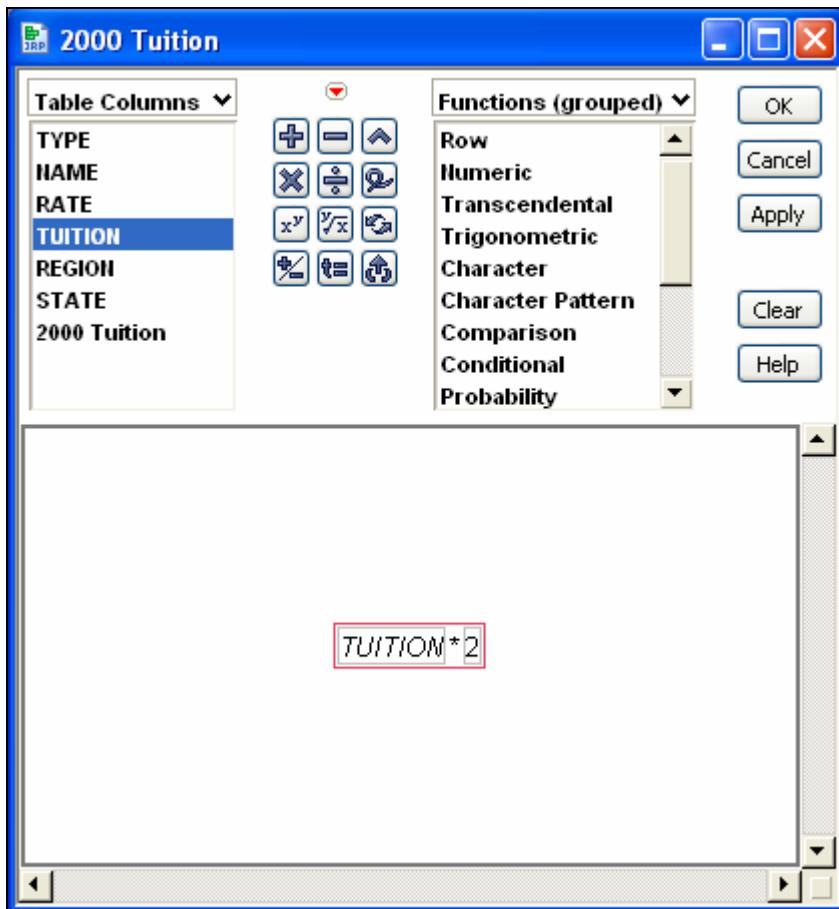
4) Select **Formula**.

5) Select **TUITION** in the Table Columns list.

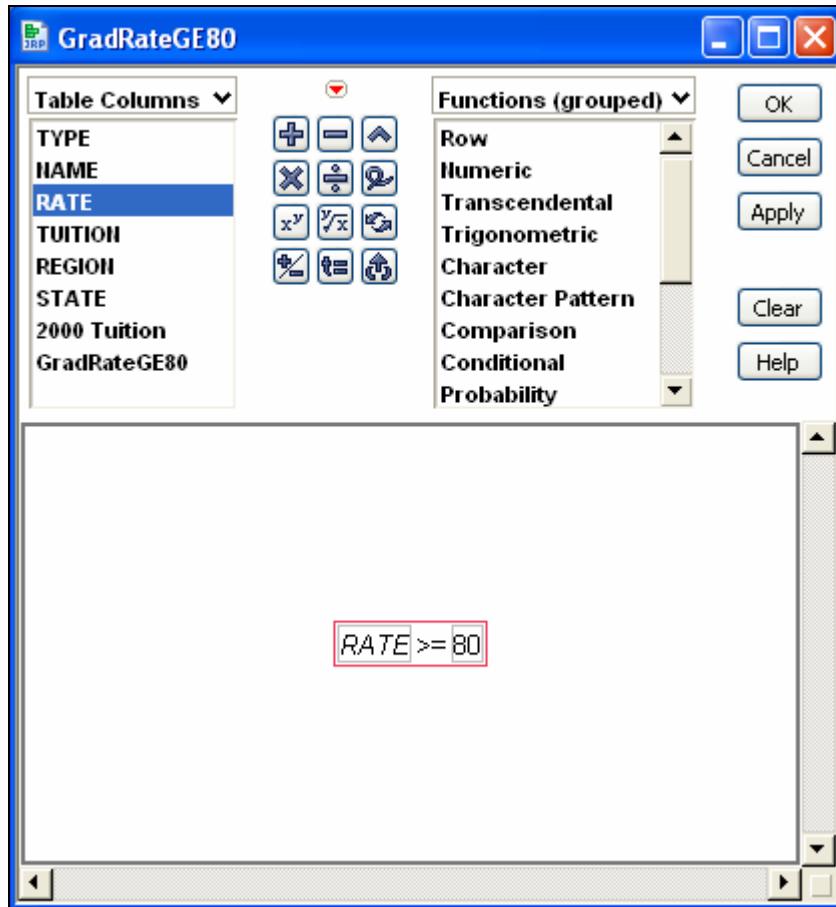
6) Select the multiplication symbol.

7) Type **2** and press ENTER.

8) Select **Apply**.



- 9) Select OK.
 - 10) Select OK.
- b. Create a new nominal numeric column **GradRateGE80** to identify those colleges that have at least an 80-percent graduation rate. Notice that this column will contain only two values: 1 if the comparison is true, 0 if false. Use a formula with a conditional clause to compute the value.
- 1) Select Cols \Rightarrow New Column...
 - 2) Type **GradRateGE80** as the name.
 - 3) Select Continuous \Rightarrow Nominal for the modeling type.
 - 4) Select Column Properties.
 - 5) Select Formula.
 - 6) Select RATE in the Table Columns list.
 - 7) Select Comparison \Rightarrow a >= b.
 - 8) Type **80** and press ENTER.
 - 9) Select Apply.



- 10) Select **OK**.
 - 11) Select **OK**.
- c. Add the value label property to the **GradRateGT80** column to ensure that the values displayed in the data table are Greater than/equal to 80 or Less than 80.
- 1) Select **GradRateGT80**.
 - 2) Select **Cols** \Rightarrow **Column Info...**.
 - 3) Select **Column Properties** \Rightarrow **Value Labels**.
 - 4) Type **0** as the value.
 - 5) Type **Less than 80** as the label.
 - 6) Select **Add**.
 - 7) Type **1** as the value.
 - 8) Type **Greater than/equal to 80** as the label.

9) Select Add.

Value Labels

If a column has value labels, and Use Value Labels is checked, the labels will be displayed wherever the column data are displayed.

0 = Less than 80	Add
1 = Greater than/equal to 80	Change
optional item	Remove
Value	<input type="text"/>
Label	<input type="text"/>
<input checked="" type="checkbox"/> Use Value Labels	

10) Select Apply \Rightarrow OK.

	TYPE	NAME	RATE	TUITION	REGION	STATE	2000 Tuition	GradRateGE80
1	private	Alfred	67	6350	Northeast	NY	12700	Less than 80
2	private	Alvernia	70	6202	Northeast	PA	12404	Less than 80
3	private	American International	62	7734	Northeast	MA	15468	Less than 80
4	private	Atlantic Christian	64	5690	South	NC	11380	Less than 80
5	public	Auburn	63	4428	South	AL	8856	Less than 80
6	private	Austin	61	8735	Southwest	TX	17470	Less than 80
7	private	Barker U.	55	6410	Central	KS	12820	Less than 80
8	private	Barat	73	7919	Midwest	IL	15838	Less than 80
9	private	Barber-Scotia	70	4000	South	NC	8000	Less than 80
10	private	Baylor	90	5620	Southwest	TX	11240	Greater than/equal to 80
11	private	Bellarmine	74	6000	Southwest	TX	12000	Less than 80

11) Deselect all columns.

d. Save this table.

1) Select File \Rightarrow Save.

3. Creating a Summary Table

a. Find the average graduation RATE by REGION.

1) Select Tables \Rightarrow Summary.

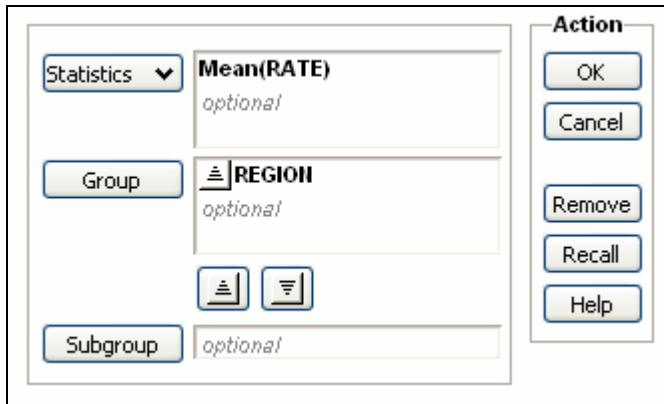
2) Select REGION.

3) Select Group.

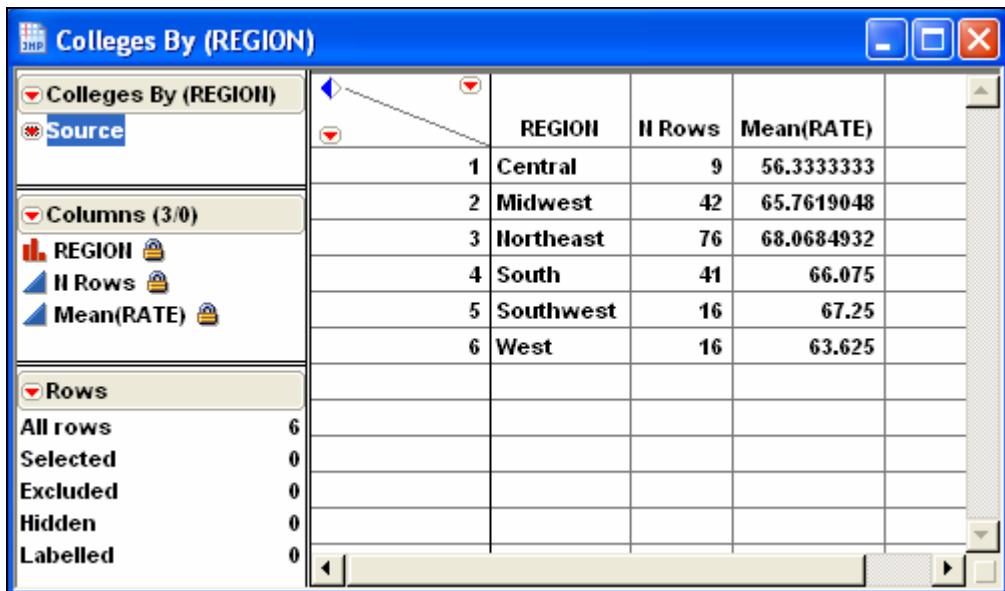
4) Select RATE.

5) Select Statistics.

6) Select Mean.

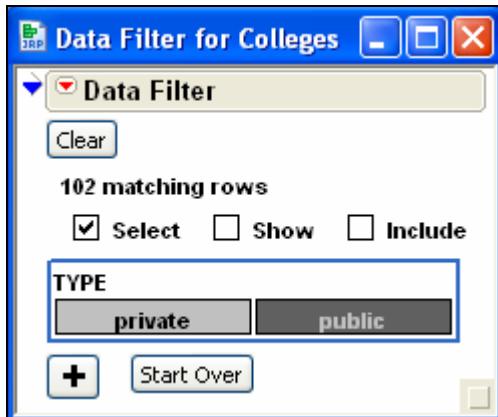


7) Select **OK**.



The average for Central is 56.3.

- Close this summary table without saving it.
 - Select **File** \Rightarrow **Close**.
- #### 4. Logically Selecting Rows and Creating Subsets
- Select the rows in the data table where the value of **Type** is **public**.
 - Select **Rows** \Rightarrow **Data Filter**.
 - Select **TYPE** \Rightarrow **Add**.
 - Select **public**.



The public school rows have been selected.

- b. How many rows are selected?

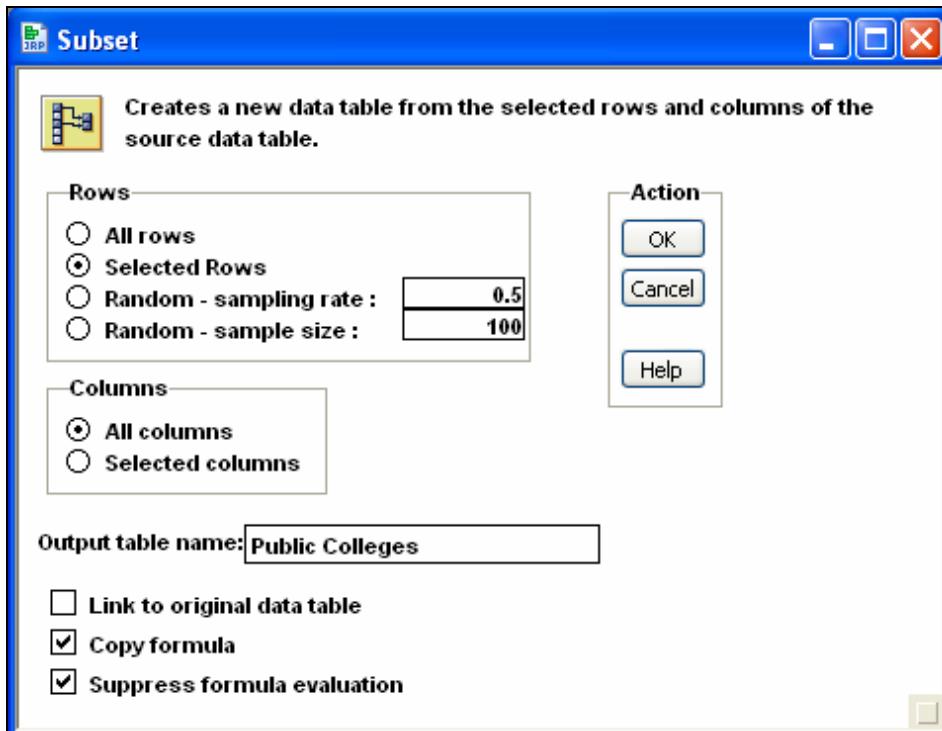
Based on the Rows panel in **Colleges.JMP**, 102 rows are selected.

Rows	
All rows	200
Selected	102

- 1) Close the Data Filter window.

- c. Make a subset with the selected rows and save it as **Public Colleges.JMP**. Close the file.

- 1) Select **Tables** \Rightarrow **Subset**.
- 2) Type **Public Colleges** as the name.



3) Select **OK**.

	TYPE	NAME	RATE	TUITION	REGI
+	1 public	Auburn	63	4428	South
+	2 public	Bloomsburg U. of Pennsylvania	68	4480	Northe
+	3 public	Bowling Green State	60	6158	Midwe

4) Select **File** \Rightarrow **Save**.

5) Select **Save**.

6) Select **File** \Rightarrow **Close**.

7) Deselect all rows in **Colleges.JMP**.

5. Creating a New Table by Combining Two Tables That Have the Same Column

Return to **Colleges.JMP**. Open **Regional Economics.JMP**.

	Region	Economic Cost Factor
1	Northeast	9.44
2	South	5.70
3	Southwest	4.88
4	Midwest	4.62
5	Central	3.95
6	West	8.79

a. Join **Regional Economics** to **Colleges.JMP** by matching rows with the same **REGION** value. Name the table **Colleges with Regional Economics.JMP**.

1) Select **Tables** \Rightarrow **Join**.

2) Select **Colleges**.

Change the Matching Specification.

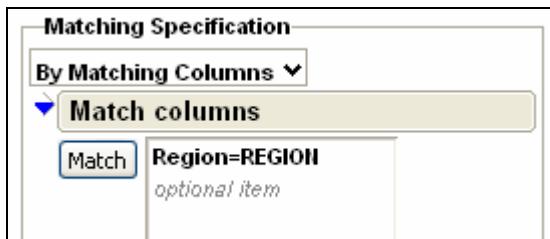
3) Select **By Row Number** \Rightarrow **By Matching Columns**.

Select the column in each data table that represents the same values.

4) Select **Region** in the **Regional Economics** data table.

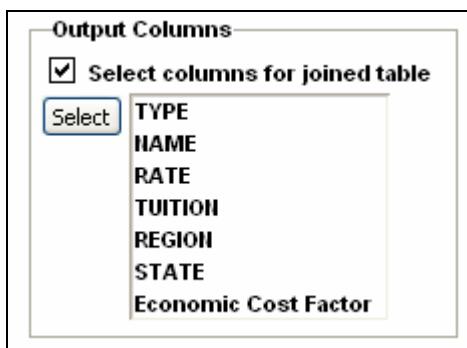
5) Select **REGION** in the **Colleges** data table.

6) Select **Match**.



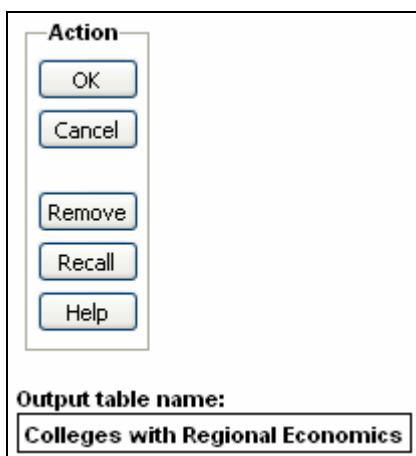
Choose the columns that should be included in the final combined data table.

- 7) Select Select columns for joined table.
- 8) Select TYPE through STATE (all columns) in the **Colleges** data table.
- 9) Select Select.
- 10) Select the Economic Cost Factor in the **Regional Economics** data table.
- 11) Select Select.



Name the combined data table.

- 12) Type **Colleges with Regional Economics** as the name.



- 13) Select OK.
- b. How many rows are in the new data table?

Based on the Rows panel in the joined table, there are 200 rows in the new table.

- c. Save the new table as **Colleges with Regional Economics.JMP**.

The file name was given already during the joining of the two original tables.

1) Select **File** \Rightarrow **Save As....**

2) Select **Save**.

- d. Close **Colleges with Regional Economics.JMP** and **Regional Economics.JMP**.

1) Select **File** \Rightarrow **Close** for each data table.

6. Using the Tabulate Platform

- a. Use the Tabulate platform to make a pivot table that summarizes the number of colleges and the median **RATE** and **TUITION** by **REGION** and **TYPE**.

1) Select **Tables** \Rightarrow **Tabulate**.

2) Drag **REGION** to the **Drop zone for columns** box.

3) Drag **TYPE** under **REGION** in the top box.

	REGION											
	Central		Midwest		Northeast		South		Southwest		West	
	TYPE		TYPE		TYPE		TYPE		TYPE		TYPE	
	private	public	private	public	private	public	private	public	private	public	private	public
	4	5	23	19	38	38	17	24	12	4	4	12

4) Drag **RATE** to the lower-left box.

5) Select **Add Analysis Column**.

6) Drag **TUITION** under **RATE**.

7) Select **Add Analysis Column**.

8) Drag **Median** under **RATE**.

9) Drag **N** under **Median**.

		REGION											
		Central		Midwest		Northeast		South		Southwest		West	
		TYPE		TYPE		TYPE		TYPE		TYPE		TYPE	
		private	public	private	public	private	public	private	public	private	public	private	public
RATE	Median	55	53	69	60	72	64	70	63.5	66.5	58.5	81	58
	II	4	5	23	19	38	35	16	24	12	4	4	12
TUITION	Median	6312.5	4688	8000	6240	8275	4865	6653	5555	7582.5	4493.5	13890	7542
	II	4	5	23	19	38	37	17	24	12	4	4	12

10) Select **Done**

- b. How many private colleges are in the Northeast region?

There are 38 private colleges in the Northeast region.

- c. Close the summary table. Do not save.

1) Select **File** \Rightarrow **Close**.

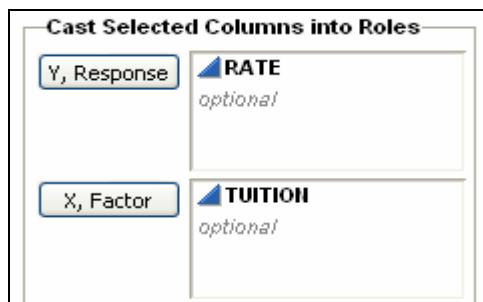
Chapter 3 Solutions to Exercises

Open **Colleges.JMP**. (If you did not save this file previously, then open **Colleges_bkup.JMP**).

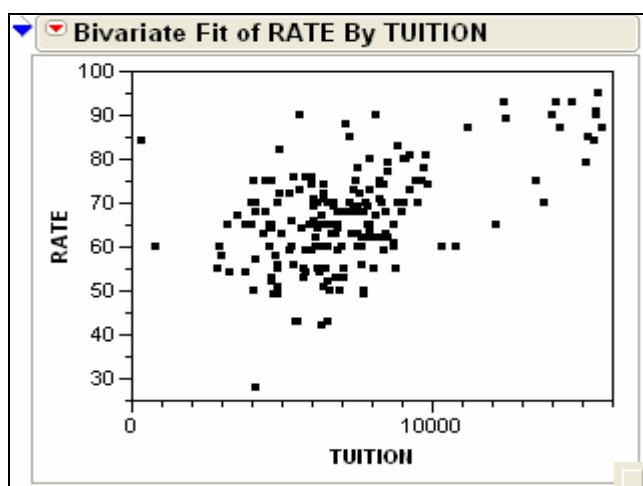
1. Using a Scatterplot

- a. Use Fit Y by X to make a scatterplot of **RATE** versus **TUITION**.

- 1) Select **Analyze** \Rightarrow **Fit Y by X**.
- 2) Select **RATE** \Rightarrow **Y, Response**.
- 3) Select **TUITION** \Rightarrow **X, Factor**.



- 4) Select **OK**.
- b. Is there any indication of a linear relationship between the two columns? If so, is it positive or negative?

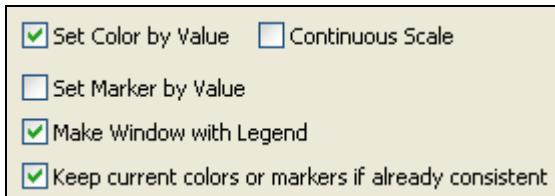


There appears to be a positive linear relationship between **RATE** and **TUITION**.

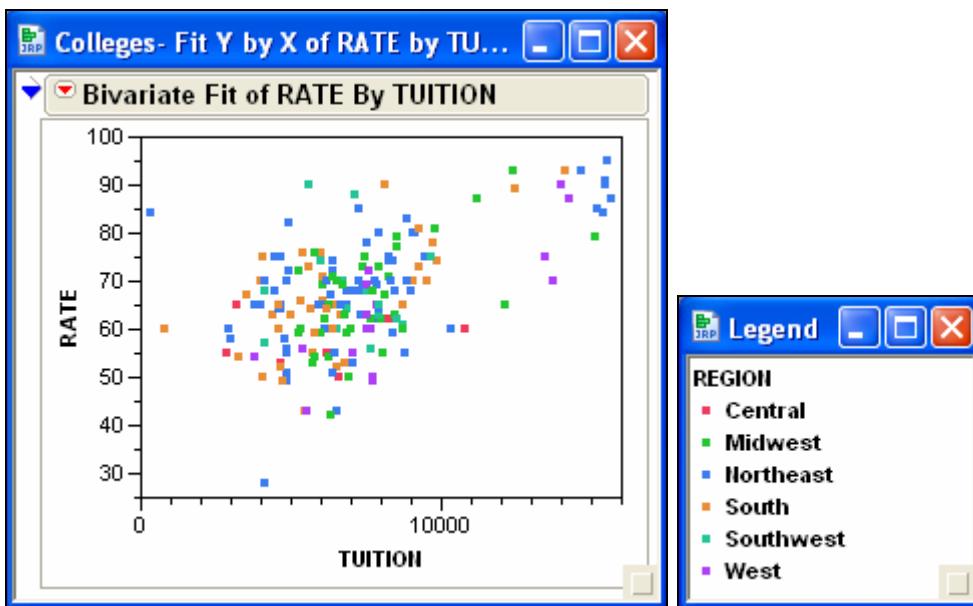
2. Using Colors and Markers

- a. Color rows based on their **REGION** value and include a legend. Are there any patterns in the scatterplot with respect to **REGION**?

- 1) Select **Rows** \Rightarrow **Color or Mark by Column...**
- 2) Select **REGION**.
- 3) Select **Set Color by Value**.
- 4) Select **Make Window with Legend**.



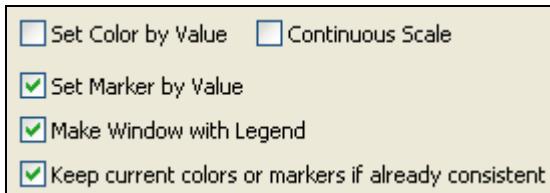
- 5) Select **OK**.



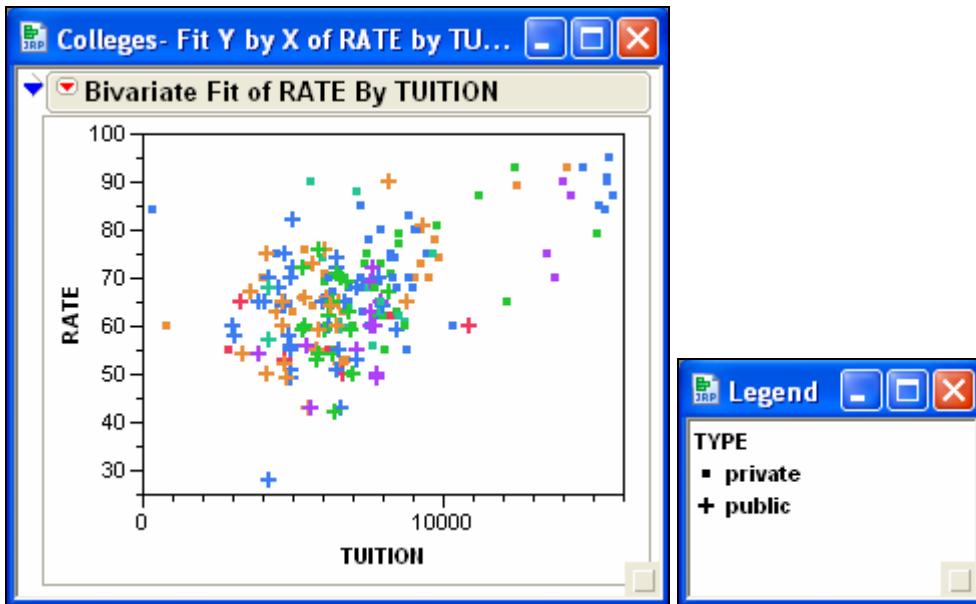
There is no apparent pattern with respect to **REGION**.

- b. Mark rows based on their **TYPE** value and include a legend. Are there any patterns in the scatterplot with respect to **TYPE**?

- 1) Select **Rows** \Rightarrow **Color or Mark by Column...**
- 2) Select **TYPE**.
- 3) Deselect **Set Color by Value**.
- 4) Select **Set Marker by Value**.
- 5) Select **Make Window with Legend**.



6) Select **OK**.



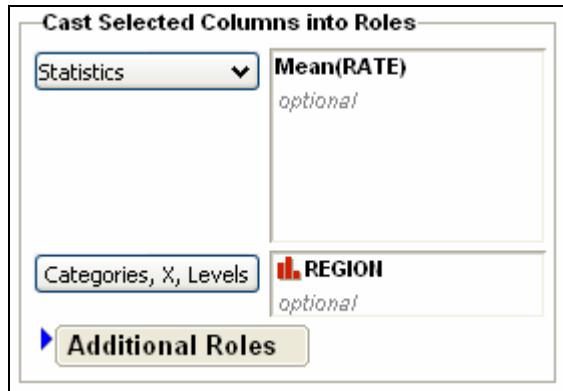
The public schools appear to cluster to the left, and the private schools span from one end of the plot to the other.

7) Close the report windows.

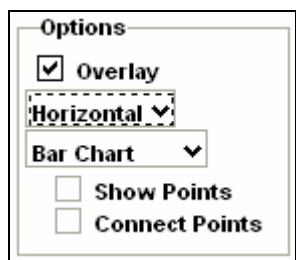
3. Creating a Bar Chart

a. Make a chart of the average **RATE** by **REGION**.

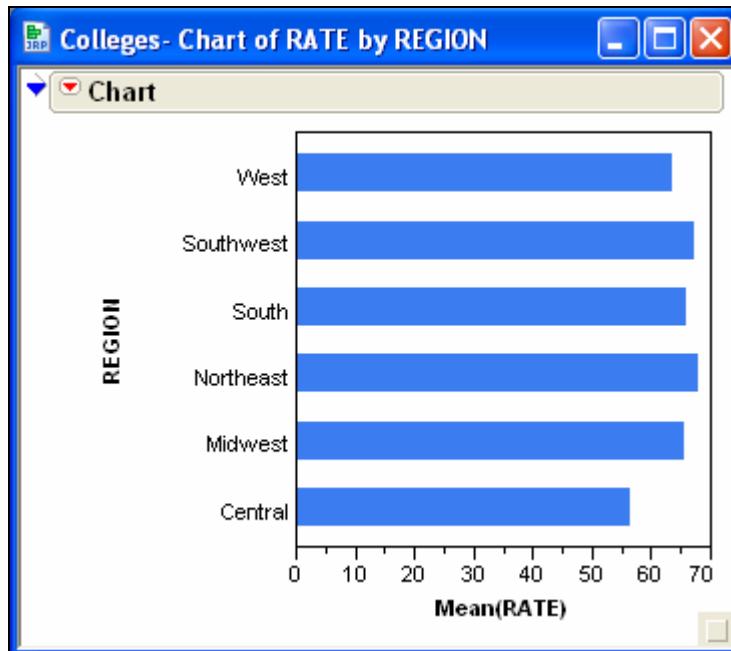
- 1) Select **Graph** \Rightarrow **Chart**.
- 2) Select **RATE**.
- 3) Select **Statistics** \Rightarrow **Mean**.
- 4) Select **REGION**.
- 5) Select **Categories, X, Levels**.



6) Select Vertical \Rightarrow Horizontal in the Options panel.

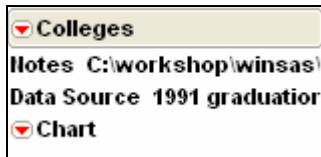


7) Select OK.



b. Save a script in the data table to reproduce this analysis.

- 1) Click the red triangle next to **Chart**.
- 2) Select Script \Rightarrow Save Script to Data Table.

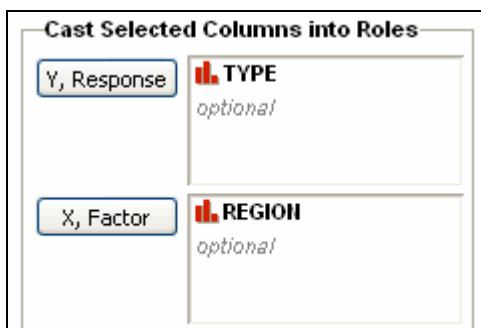


3) Close the Chart platform.

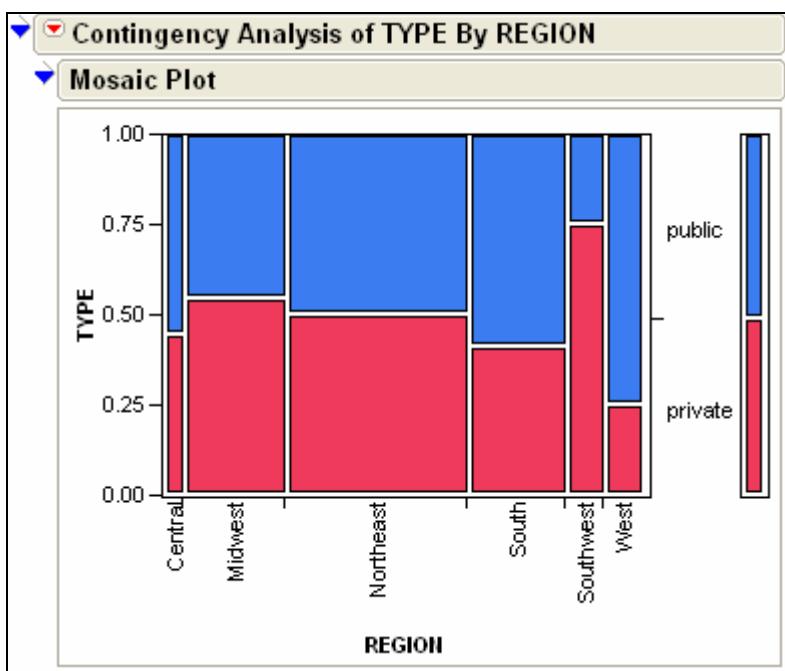
4. Using the Contingency Platform

a. Use Fit Y by X to make a contingency table of **TYPE** versus **REGION**.

- 1) Select **Analyze** \Rightarrow **Fit Y by X**.
- 2) Select **TYPE** \Rightarrow **Y, Response**.
- 3) Select **REGION** \Rightarrow **X, Factor**.



4) Select **OK**.



b. Does the proportion of public schools appear the same in all regions?

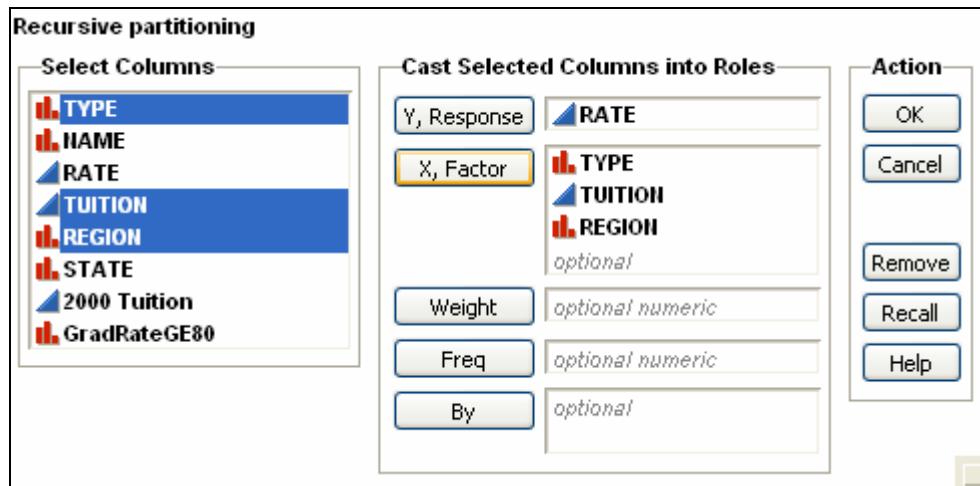
The proportions in the Southwest and West regions appear to differ the most from the other regions.

- 1) Close the Contingency platform.

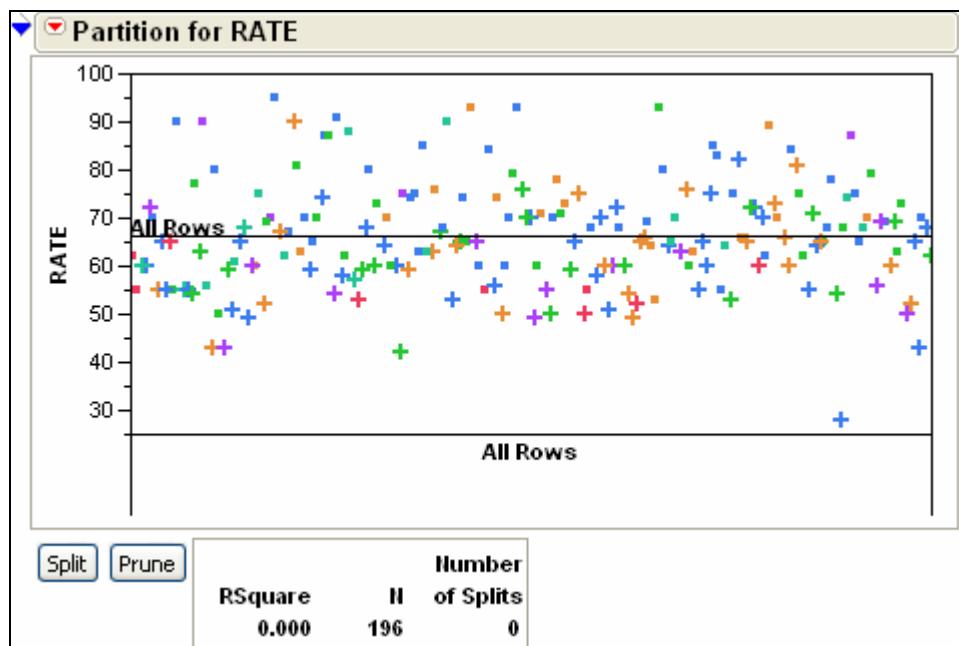
5. Using the Partition Platform

- a. Use the Partition platform to explore **RATE** versus **TYPE**, **TUITION**, and **REGION**.

- 1) Select **Analyze** \Rightarrow **Modeling** \Rightarrow **Partition**.
- 2) Select **RATE** \Rightarrow **Y, Response**.
- 3) Select **TYPE**, **TUITION**, and **REGION** \Rightarrow **X, Factor**.



- 4) Select **OK**.



- b. Use the Candidates report to determine what the first split will be. Explain why.

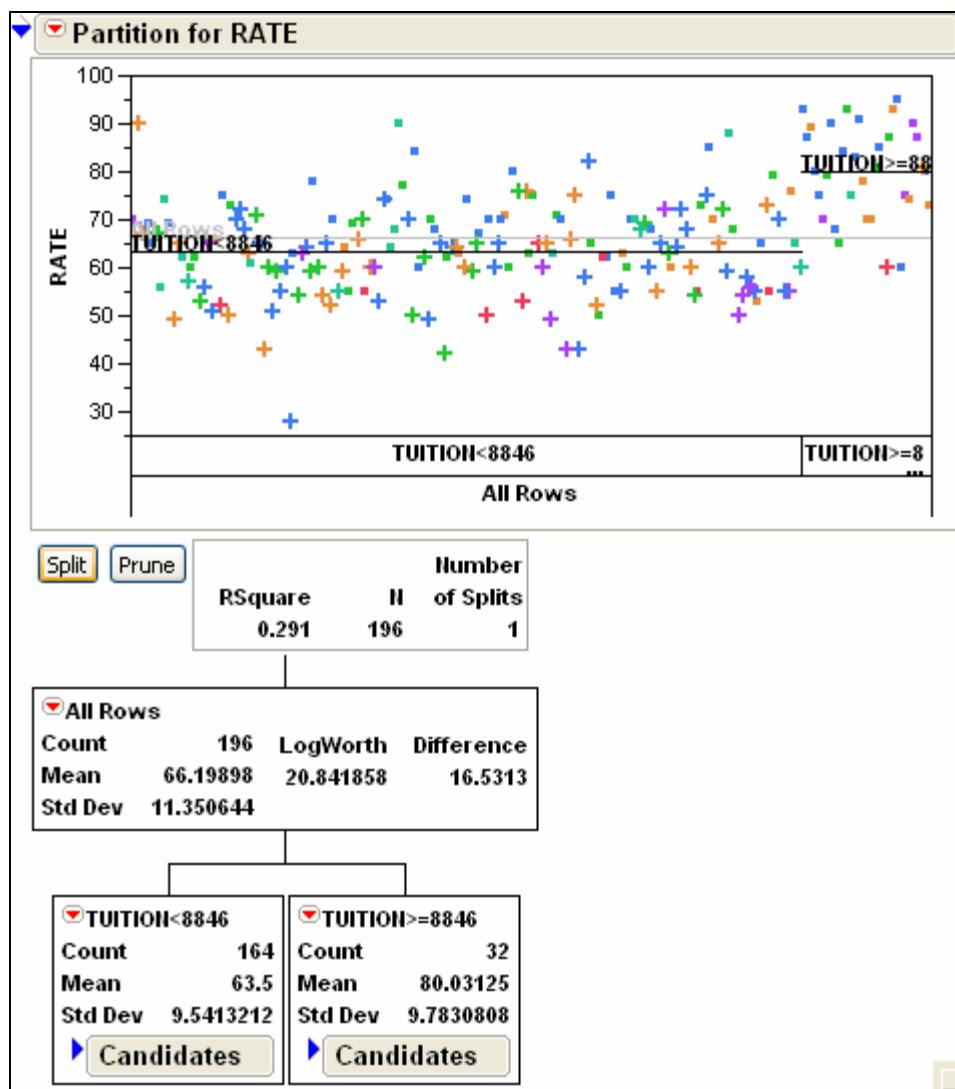
- 1) Click the blue triangle next to **Candidates**.

All Rows		
Count	196	
Mean	66.19898	
Std Dev	11.350644	
Candidates		
Term	Candidate SS	LogWorth
TYPE	5491.360175	12.62874936
TUITION	7317.271046	20.84185814
REGION	918.138192	0.97915305

JMP will split the rows into two groups based on **TUITION** because this split will yield the highest SS and LogWorth.

- c. Make the first split. Where is it made? Describe the two groups.

- 1) Select **Split**.



The larger group, which contains 164 rows with **TUITION** values less than 8846, has a mean **RATE** of 63.5. The smaller group, which contains 32 rows with **TUITION** values greater than or equal to 8846, has a mean **RATE** of 80.

- d. Use the Candidates reports to determine what the next split will be. Explain why.

<input checked="" type="checkbox"/> TUITION<8846			<input checked="" type="checkbox"/> TUITION>=8846		
Count	164	Count	32		
Mean	63.5	Mean	80.03125		
Std Dev	9.5413212	Std Dev	9.7830808		
Candidates					
Term	Candidate SS	LogWorth	Term	Candidate SS	LogWorth
TYPE	1778.796892	5.488872712	TYPE	0.000000	0.0000000000
TUITION	1154.446541	2.580191341	TUITION	1268.270711	4.879294789
REGION	918.456944	1.669728291	REGION	202.505114	0.129324316

The next split will occur in the larger leaf on the left between the two levels of **TYPE** because this split will yield the largest SS and LogWorth.

- e. Make the second split. Where is it made? Describe the two groups.

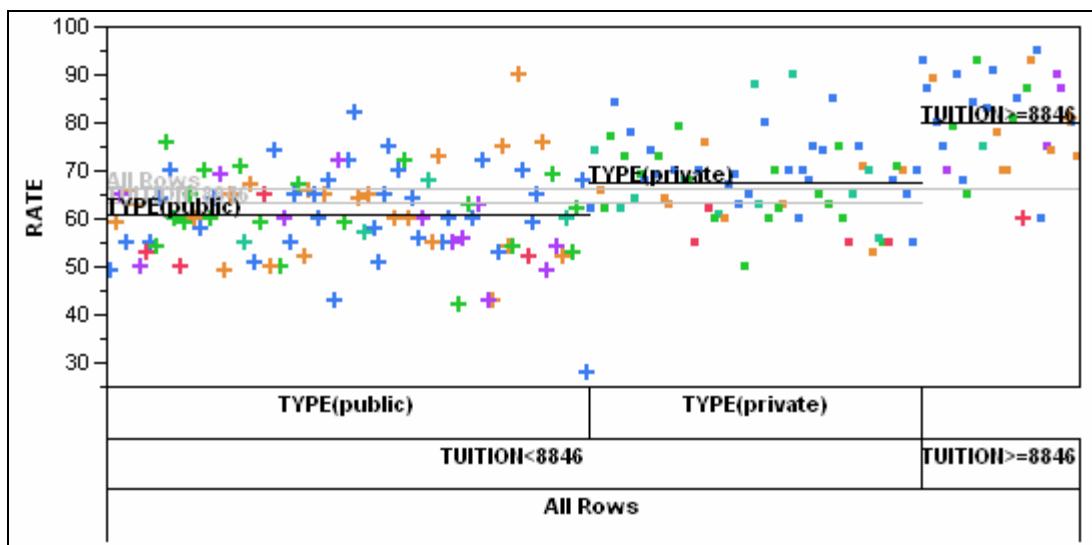
- 1) Select Split.

<input checked="" type="checkbox"/> TYPE(public)	<input checked="" type="checkbox"/> TYPE(private)
Count	97
Mean	60.762887
Std Dev	9.4083
Candidates	
Term	Candidate SS
TYPE	1778.796892
TUITION	1154.446541
REGION	918.456944

<input checked="" type="checkbox"/> TYPE(private)	<input checked="" type="checkbox"/> TYPE(public)
Count	67
Mean	67.462687
Std Dev	8.3145152
Candidates	
Term	Candidate SS
TYPE	0.000000
TUITION	1268.270711
REGION	202.505114

The two new leaves contain 97 and 67 rows for the public and private colleges, respectively. The mean **RATE** is 60.8 and 67.5 for the public and private colleges, respectively.

The scatterplot at the top of the Partition platform helps you to see how the groups are forming.



- f. Close all windows. Save **Colleges.JMP**.

- 1) Select **Window** \Rightarrow **Close All**.
- 2) Select **Yes** to save **Colleges.JMP** if prompted.

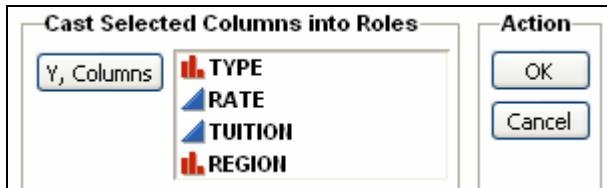
Chapter 4 Solutions to Exercises

1. Saving Results in Journal Files

Open **Colleges with Regional Economics.JMP**.

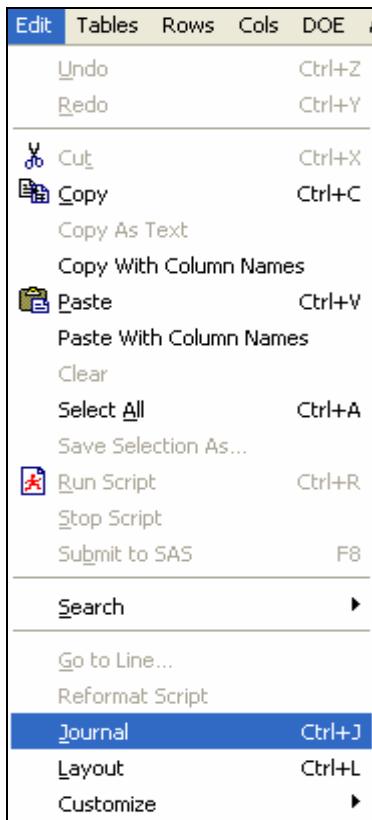
- a. Generate a distribution analysis for **TYPE**, **RATE**, **TUITION**, and **REGION**.

- 1) Select **Analyze** \Rightarrow **Distribution**.
- 2) Select **TYPE**, **RATE**, **TUITION**, and **REGION** \Rightarrow **Y, Columns**.



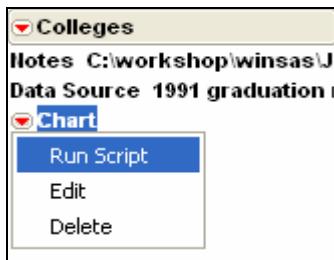
- 3) Select **OK**.
- b. Place the results in a journal.

- 1) Select **Edit** \Rightarrow **Journal**.



- c. Open **Colleges.JMP** file. Use the previously saved script (from 3.b) to make another chart of the average **RATE** by **REGION**.

1) Click the red triangle next to **Chart** in the Table panel and select **Run Script**.



- d. Place the results in the journal.

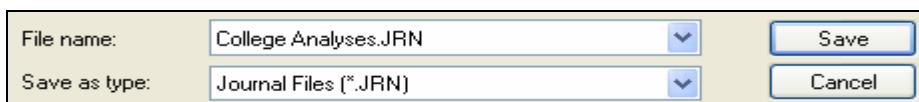
1) Select **Edit** ⇒ **Journal**.

- e. Save the journal as **College Analyses.JRN**.

1) Select **Window** ⇒ **Journal**:

2) Select **File** ⇒ **Save As...**

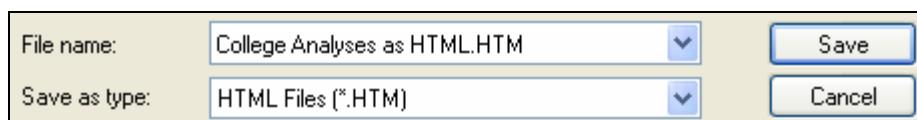
3) Type **College Analyses** as the file name.



- 4) Select Save.
- 5) Navigate to the Distribution and Chart outputs and close them.

2. Saving Journals as HTM or RTF Files

- a. Save **College Analyses.JRN** as an HTML file **College Analyses as HTML.HTM**. View the file, if possible.
 - 1) Select File \Rightarrow Save As...
 - 2) Type **College Analyses as HTML** as the name.
 - 3) Select **HTML Files (*.HTM)** for the file type.



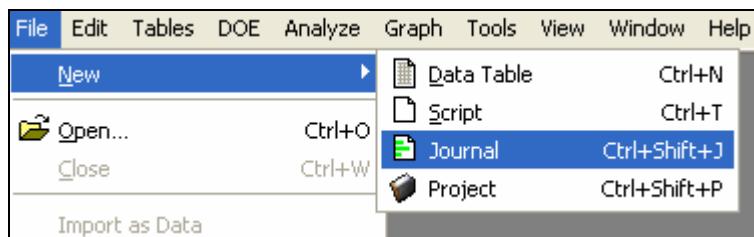
- 4) Select Save.
- b. Save the journal as an RTF file: **College Analyses as Rich Text.RTF**.
 - 1) Select File \Rightarrow Save As...
 - 2) Type **College Analyses as Rich Text** as the file name.
 - 3) Select **RTF Files (*.RTF)** for the file type.



- 4) Select Save.
- c. Close **College Analyses as Rich Text.RTF**.
 - 1) Select File \Rightarrow Close.

3. Creating a Project Journal

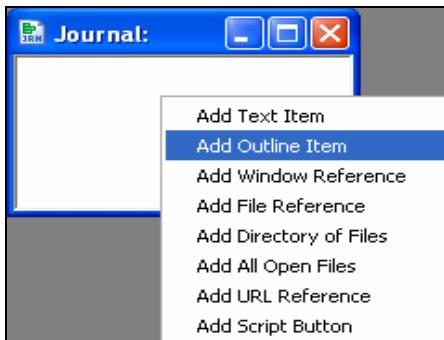
- a. Close all active windows and create a new project journal.
 - 1) Select Window \Rightarrow Close All.
 - 2) Select File \Rightarrow New \Rightarrow Journal.



An empty Journal window opens.

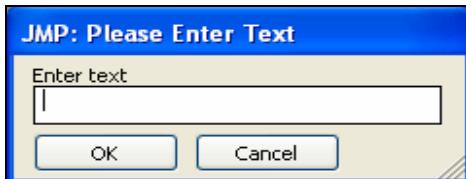
- b. Add an outline item called **Bank Waiting Times Project**.

- 1) Right-click in the empty Journal window.



A list of items that can be added to the journal is displayed. The chosen item is placed in the upper-left side in the Journal window.

- 2) Select Add Outline Item.

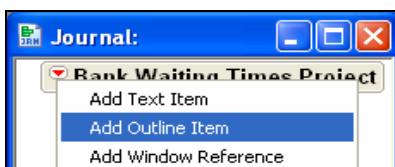


- 3) Type **Bank Waiting Times Project** and select OK.



- c. Under **Bank Waiting Times Project**, add an outline item called **Data Tables**.

- 1) Click the red triangle next to **Bank Waiting Times Project** and select Add Outline Item.

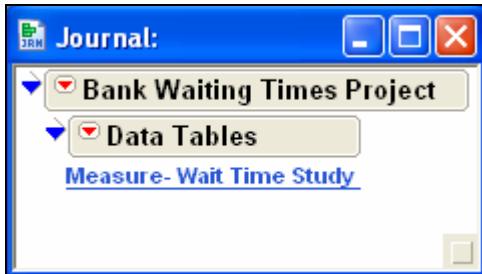


- 2) Type **Data Tables** and select OK.



After items are nested under an outline box, the blue expand/collapse diamond appears next to the outline box.

- d. Under the **Data Tables** outline item, add links to the three JMP data tables.
- 1) Click the red triangle next to **Data Tables** and select Add File Reference.
 - 2) Select Measure – Wait Time Study.JMP \Rightarrow Open.

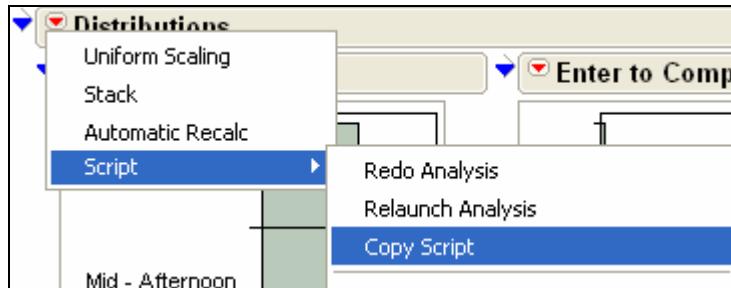


A link to the data table has been placed in the journal. Follow the same steps to add links to two additional tables.

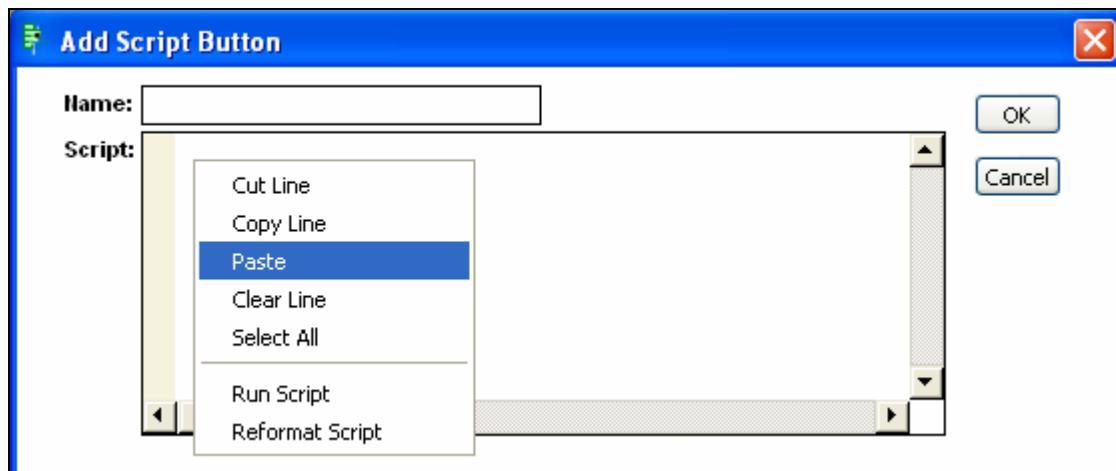
- 3) Click the red triangle next to **Data Tables** and select Add File Reference.
- 4) Select Control - Wait Time Data.JMP \Rightarrow Open.
- 5) Click the red triangle next to **Data Tables** and select Add File Reference.
- 6) Select Completed Wait Time Project.JMP \Rightarrow Open.



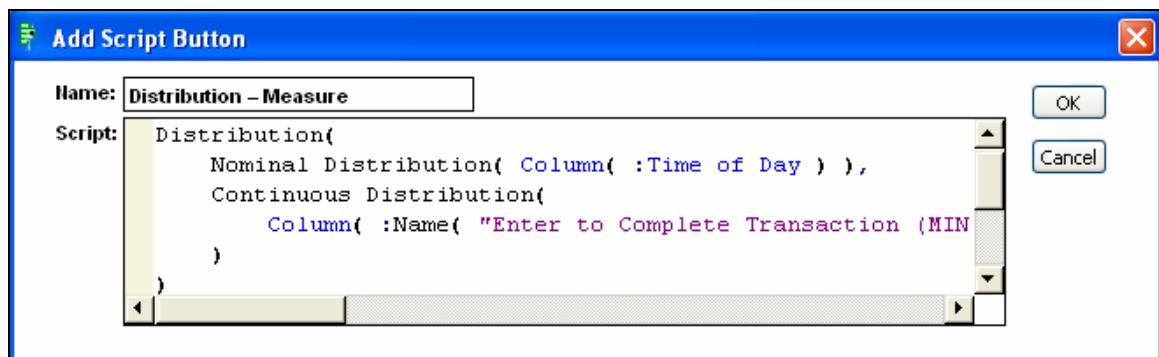
- e. Generate distribution reports for the **Time of Day** and **Wait Time** columns in the **Measure – Wait Time Study** table and the **Control - Wait Time Data** table. Store the scripts that create the analysis to the journal under an outline item called **Distribution Reports**. Name the two distribution reports **Distribution – Measure** and **Distribution – Control**, respectively.
- 1) Select Measure – Wait Time Study to open the data table.
 - 2) Select Analyze \Rightarrow Distribution.
 - 3) Select Time of Day and Enter to Complete Transaction (MIN). Select Y, Columns \Rightarrow OK.
 - 4) Click the red triangle next to **Distributions** and select Script \Rightarrow Copy Script.



- 5) Select **Window** \Rightarrow **Journal**:
- 6) Click the red triangle next to **Bank Waiting Times Project** and select **Add Outline Item**.
- 7) Type **Distribution Reports** in the text box and select **OK**.
- 8) Click the red triangle next to **Distribution Reports** and select **Add Script Button**.



- 9) Right-click in the Script box and select **Paste**.



- 10) Type **Distribution - Measure** as the name and select **OK**.



- 11) Select **Control - Wait Time Data** to open the data table.
- 12) Select Analyze \Rightarrow Distribution.
- 13) Select Time of Day and Enter to Complete Transaction (MIN). Select Y, Columns \Rightarrow OK.
- 14) Click the red triangle next to **Distributions** and select Script \Rightarrow Copy Script.
- 15) Select Window \Rightarrow Journal:.
- 16) Click the red triangle next to **Distribution Reports** and select Add Script Button.
- 17) Right-click in the Script box and select Paste.
- 18) Type **Distribution – Control** as the name and select OK.



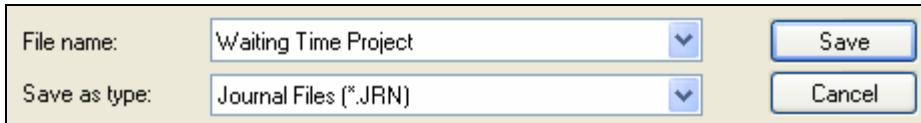
- f. Add a link to the journal that connects to the spreadsheet **Waiting Time Study.xls**, which was used to create JMP tables. Add this link under an outline item called **Raw Data Files**.
 - 1) Click the red triangle next to **Bank Waiting Times Project** and select Add Outline Item.
 - 2) Type **Raw Data Files** and select OK.
 - 3) Click the red triangle next to **Raw Data Files** and select Add File Reference.
 - 4) Change Files of Type to Excel 97-2003 Files (*.XLS). Select Waiting Time Study.xls and then select Open in the dialog box.



After the journal is complete, it can be saved.

g. Save the journal as **Waiting Time Project.**

- 1) Select **File** \Rightarrow **Save As...** and type **Waiting Time Project** as the file name.



- 2) Select **Save**.

 Unless the user modifies the scripts saved to this journal, the user must select a data table and then select the script(s) that correspond to the analyses for that specific table. For example, to view the distribution report for the measure portion of the project, the user will select **Measure – Wait Time Study** and then select **Distribution – Measure**.

- 3) Close all windows.

Appendix B Index

A

attribute data columns
exploring distributions, 3-24–3-27

B

bar charts
distribution of values, 3-3
Beginner's Tutorial, 1-12
box plots
distribution of values, 3-3
Brush tool
graphics tools, 3-4
bubble plots, 3-19–3-23

C

categorical columns, 2-34–2-36
columns
categorical, 2-34–2-36
creating, 2-19–2-26
numeric, 2-32–2-34
Columns menus, 2-4–2-18
Columns panels, 1-38–1-42, 2-4–2-18
concatenating columns
Tables menu, 2-40
context-sensitive help, 3-10
continuous, 1-37
continuous columns
exploring graphs, 3-6–3-18
Crosshairs tool
graphics tools, 3-4

D

data grids, 1-35–1-36
data tables, 1-18–1-36
Database command, 1-23–1-30
documentation, 1-15
DOE Tutorial, 1-13

F

file formats
HTML, 4-6–4-7
journals, 4-6–4-7

G

Grabber tool
graphics tools, 3-4
graphics tools, 3-4, 3-12–3-14
graphs
exploring continuous columns, 3-6–3-18
summaries, 3-14–3-18

H

Help
context-sensitive, 3-10
histograms, 3-24
distribution of values, 3-3
HTML
file formats, 4-6–4-7

I

importing data
Excel files, 1-19–1-20
text files, 1-18
J
JMP Log, 1-9–1-10
JMP scripting language (JSL)
saving results, 3-5
Table panels, 3-5
JMP Starter, 1-5–1-6
joining data tables
Tables menu, 2-40, 2-49–2-52
journals, 4-3–4-7, 4-8–4-20
file formats, 4-6–4-7

L

Lasso tool
graphics tools, 3-4
layouts, 4-5

M

Magnifier tool
graphics tools, 3-4
missing values, 1-36
modeling types, 1-38

- mosaic plots, 3-26
- N**
- nominal, 1-37
- numeric columns, 2-32–2-34
- numerical results, 3-14–3-18
- O**
- observations, 1-36
- ODBC drivers, 1-17
- ordinal, 1-37
- P**
- panels, 1-35
- Columns, 1-38–1-42, 2-4–2-18
 - Rows, 1-42–1-44, 2-19–2-26
 - Table, 1-45–1-47, 3-5
- Partition platform, 3-28–3-37, 3-43
- partitioning
 - with categorical dependent variable, 3-42
- preferences, 1-7–1-9
- project journals
 - creating, 4-9–4-20
- projects
 - creating, 4-16–4-20
- R**
- recode columns, 2-14
- recursive partitioning, 3-28–3-37
- roles, 2-11
- row states, 1-42
 - Selected, 1-42–1-44
- Rows menus, 2-19–2-26
- Rows panels, 1-42–1-44, 2-19–2-26
- S**
- sample sizes, 1-36
- saving results, 4-3–4-7
 - JMP scripting language (JSL), 3-5
- scatterplots
 - trends between values, 3-4
- scripts. See JMP scripting language (JSL)
- Selected
 - row states, 1-42–1-44
- sorting data tables
- Tables menu, 2-40, 2-43, 2-47
- splitting columns
 - Tables menu, 2-40
- stacking columns
 - Tables menu, 2-40
- statistics, 1-13
- subset
 - Tables menu, 2-47–2-49
- summary tables, 2-41–2-42, 2-44–2-47
- T**
- Table panels, 1-45–1-47
 - JSL scripts, 3-5
- table variables, 1-45
- tables
 - creating, 2-41–2-52
 - Tables menu
 - concatenate, 2-40
 - join, 2-40, 2-49–2-52
 - sort, 2-40, 2-43, 2-47
 - split, 2-40
 - stack, 2-40
 - subset, 2-47–2-49
 - summary, 2-41–2-42, 2-44–2-47
 - update, 2-40
- Tabulate option, 2-54–2-63
- Tip of the Day window, 1-11
- tree algorithms
 - best split for independent variable, 3-39
 - compare best splits, 3-40
 - parameters, 3-43
 - partition with best split, 3-40
 - repeat for other independent variables, 3-39
 - repeat within partitions, 3-41
 - separation of means, 3-38
- U**
- updating columns
 - Tables menu, 2-40
- user-assistance options, 1-11–1-16
- V**
- variables, 1-36