# Contents

## Acknowledgments

## Part 1  The Basics  1

### Chapter 1  Introduction  3

- Purpose  4
- Audience  4
- What This Book Is and Isn’t  4
- How This Book Is Organized  5
- How to Use This Book  9
- Using JMP or JMP SE  9

### Chapter 2  Getting Started  11

- JMP and Your Computer  12
- Explaining JMP Terms  13
- Starting JMP  14
- Displaying a Simple Example  14
- Getting Help  17
- Exiting JMP  19

### Chapter 3  Using Data Tables  21

- Using Data Tables  22
- What Is a Data Table?  22
- Creating the Speeding Ticket Data Table  24
- Opening an Existing JMP Data Table  29
- Importing Data  29
- Understanding Data Tables  36
- Working with Columns  41
- Working with Rows  48
- Understanding Data Type and Modeling Type  50

From *Elementary Statistics Using JMP®*. Full book available for purchase [here](#).
Ordering Values 53
Adding Value Labels 56
Printing a Data Table 57
Sorting a Data Table 60
Summary 62
Exercises 70

Chapter 4 Summarizing Data 71
Checking Data for Errors 73
Using Distribution for a Continuous Variable 73
Using Distribution for Multiple Variables 89
Interacting with Distribution Results 94
Customizing Reports 100
Summaries 104
Exercises 111

Chapter 5 Graphing Data and Printing Results 113
Creating Custom Bar Charts 114
Creating Treemaps 122
Printing Results 127
Saving Results to Journals 129
Summaries 134
Exercises 138

Part 2 Statistical Background 139

Chapter 6 Understanding Fundamental Statistical Concepts 141
Populations and Samples 142
The Normal Distribution 146
Parametric and Nonparametric Statistical Methods 150
Testing for Normality 150
Building a Hypothesis Test 164
Statistical and Practical Significance 166
Summaries 170
Exercises 172
Chapter 7  Estimating the Mean  173
   Using One Number to Estimate the Mean  174
   Effect of Sample Size  175
   Effect of Population Variability  178
   The Distribution of Sample Averages  179
   Getting Confidence Intervals for the Mean  184
   Summaries  189
   Exercises  190

Part 3  Comparing Groups  191

Chapter 8  Comparing Two Groups  193
   Deciding between Independent and Paired Groups  195
   Summarizing Data from Two Independent Groups  196
   Summarizing Data from Paired Groups  199
   Building Hypothesis Tests to Compare Two Groups  205
   Performing the Two-Sample $t$-test  207
   Performing the Wilcoxon Rank Sum Test  220
   Enhancing the Two-Sample Graph  224
   Performing the Paired-Difference $t$-test  228
   Performing the Wilcoxon Signed Rank Test  235
   Summaries  239
   Exercises  243
   Special Topic: Paired Data in a Single Column  246

Chapter 9  Comparing More Than Two Groups  249
   Summarizing Data from Multiple Groups  251
   Building Hypothesis Tests to Compare More Than Two Groups  257
   Performing a One-way Analysis of Variance  259
   Analysis of Variance with Unequal Variances  266
   Performing a Kruskal-Wallis Test  270
   Enhancing JMP Graphs  273
   Multiple Comparison Procedures  274
Contents

Summarizing with an Example 293
Summary 299
Exercises 302

Part 4 Fitting Lines to Data 305

Chapter 10 Correlation and Regression 307
Summarizing Multiple Continuous Variables 309
Calculating Correlation Coefficients 316
Performing Straight-Line Regression 320
Fitting a Straight Line Using JMP 324
Summarizing Straight-Line Regression 335
Fitting Curves 336
Regression with Two or More Independent Variables 346
Summaries 353
Exercises 357

Chapter 11 Basic Regression Diagnostics 359
Concepts in Plotting Residuals 360
Creating Residuals Plots for the Energy Data 363
Creating Residuals Plots for the Engine Data 375
Using the Lack Of Fit Report 384
Testing the Regression Assumption for Errors 388
Summaries 392
Exercises 396
Special Topic: Leverage Plots 397

Part 5 Data in Summary Tables 401

Chapter 12 Creating and Analyzing Contingency Tables 403
Defining Contingency Tables 404
Summarizing Raw Data in Tables 405
Creating a JMP Contingency Table from an Existing Summary Table 411
Creating Contingency Tables for Several Variables 414
Performing Tests for Independence 419
Measures of Association with Ordinal Variables 424
Chapter 12
Creating and Analyzing Contingency Tables

Defining Contingency Tables  404
Summarizing Raw Data in Tables  405
   Understanding the Mosaic Plot  407
   Understanding the Contingency Table Report  409
   Enhancing the Report  410
Creating a JMP Contingency Table from an Existing Summary Table  411
Creating Contingency Tables for Several Variables  414
   Revising Colors in Mosaic Plots  417
Performing Tests for Independence  419
   Understanding Chi Square Test Results  422
   Understanding Fisher’s Exact Test Results  423
   Enhancing the Tests Report  423
Measures of Association with Ordinal Variables  424
   Understanding the Plot  427
   Understanding the Report  428
Summaries  429
   Key Ideas  429
   JMP Steps  430
Exercises  432
Special Topic: Statistical Summary Tables  434
Do Democrats and Republicans have the same responses to a survey question asking about campaign reforms? Do rural and urban residents own different types of vehicles (sports cars, sedans, SUVs, and trucks)? Do graduate and undergraduate students differ in whether they share an apartment, have their own apartment, or live in a dorm?

These questions involve looking at two classification variables, and testing if the variables are related. Neither variable contains quantitative measurements; instead, both variables identify the respondent as belonging to a group. For example, one respondent might be “rural, truck”, and another respondent might be “urban, sedan”. The two variables can be character or numeric, and are either nominal or ordinal. This chapter discusses:

- summarizing classification data using tables and plots
- testing for independence between the classification variables
- using measures of association between classification variables

The methods in this chapter are appropriate for nominal and ordinal variables.

**Defining Contingency Tables**

Suppose you have nominal or ordinal variables that classify the data into groups. You want to summarize the data in a table. To make discussing summary tables easier, this section introduces some notation.

Tables that summarize two or more classification variables are called *contingency tables*. These tables are also called *crosstabulations*, *summary tables*, or *pivot tables* (in Microsoft Excel). Tables that summarize two variables are called *two-way tables*, tables that summarize three variables are called *three-way tables*, and so on. A special case of the two-way table occurs when both variables have only two levels. This special case is called a *2×2 table*. Although this chapter shows how to create contingency tables involving several variables, the analyses are appropriate only for two-way tables. Figure 12.1 shows the parts of a contingency table.
Figure 12.1 Parts of a Contingency Table

<table>
<thead>
<tr>
<th>Rows</th>
<th>Columns</th>
</tr>
</thead>
<tbody>
<tr>
<td>row 1</td>
<td>cell_{11}</td>
</tr>
<tr>
<td>row 2</td>
<td>cell_{21}</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>row r</td>
<td>cell_{r1}</td>
</tr>
</tbody>
</table>

The table consists of rows and columns. Figure 12.1 contains \( r \) rows and \( c \) columns, and is an \( r \times c \) table. The rows and columns form *cells*. Each cell of a table is uniquely indexed by its row and column. For example, the cell in the second row and first column is \( \text{cell}_{21} \). A contingency table usually shows the number of observations in each cell, or the *cell frequency*. The total number of observations in the table is \( n \). The number of observations in each cell follows the same notation pattern as the cells. For example, the number of observations in cell \( \text{cell}_{21} \) is \( n_{21} \).

The contingency table in Figure 12.1 is a two-way table because it summarizes two variables. The phrase “two-way” does not refer to the number of rows or columns; it refers to the number of variables that are included in the table.

Sometimes, you have the raw data and you want to summarize the data in a table and analyze it. Other times, you already have a summary table and you want to analyze it. The next three sections discuss summarizing data in tables. The rest of the chapter discusses analyses for two-way tables.

---

**Summarizing Raw Data in Tables**

Suppose you have raw data that you want to summarize in a table. Table 12.1 shows data from an introductory statistics class. The instructor collected data on the gender of each student, and whether the student was majoring in Statistics.

---

1 Data is from Dr. Ramon Littell, University of Florida. Used with permission.
Table 12.1 Class Data

<table>
<thead>
<tr>
<th>Student</th>
<th>Gender</th>
<th>Major</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Male</td>
<td>Statistics</td>
</tr>
<tr>
<td>2</td>
<td>Male</td>
<td>Other</td>
</tr>
<tr>
<td>3</td>
<td>Female</td>
<td>Statistics</td>
</tr>
<tr>
<td>4</td>
<td>Male</td>
<td>Other</td>
</tr>
<tr>
<td>5</td>
<td>Female</td>
<td>Statistics</td>
</tr>
<tr>
<td>6</td>
<td>Female</td>
<td>Statistics</td>
</tr>
<tr>
<td>7</td>
<td>Male</td>
<td>Other</td>
</tr>
<tr>
<td>8</td>
<td>Male</td>
<td>Other</td>
</tr>
<tr>
<td>9</td>
<td>Male</td>
<td>Statistics</td>
</tr>
<tr>
<td>10</td>
<td>Female</td>
<td>Statistics</td>
</tr>
<tr>
<td>11</td>
<td>Male</td>
<td>Other</td>
</tr>
<tr>
<td>12</td>
<td>Female</td>
<td>Statistics</td>
</tr>
<tr>
<td>13</td>
<td>Male</td>
<td>Statistics</td>
</tr>
<tr>
<td>14</td>
<td>Male</td>
<td>Statistics</td>
</tr>
<tr>
<td>15</td>
<td>Male</td>
<td>Other</td>
</tr>
<tr>
<td>16</td>
<td>Female</td>
<td>Statistics</td>
</tr>
<tr>
<td>17</td>
<td>Male</td>
<td>Statistics</td>
</tr>
<tr>
<td>18</td>
<td>Male</td>
<td>Other</td>
</tr>
<tr>
<td>19</td>
<td>Female</td>
<td>Other</td>
</tr>
<tr>
<td>20</td>
<td>Male</td>
<td>Statistics</td>
</tr>
</tbody>
</table>

This data is available in the Stat Majors data table in the sample data for the book.

To summarize the data table in JMP:

1. Open the Stat Majors data table.
2. In the JMP Starter window, click Basic ➔ Contingency.
3. Click Major ➔ Y, Response Category. This variable forms the columns of the summary table.
4. Click Gender ➔ X, Grouping Category. This variable forms the rows of the summary table.
5. Click OK.
Figure 12.2 shows the results, with the Tests report hidden. (“Performing Tests for Independence” in this chapter discusses this hidden report.) JMP displays both a graph and a text report. The title of the report identifies the column variable (Major) and the row variable (Gender). The next two topics discuss the graphs and reports in Figure 12.2.

Understanding the Mosaic Plot

Figure 12.2 shows a two-way mosaic plot. JMP documentation calls this a contingency analysis mosaic plot or a mosaic plot.

The left axis of the plot identifies proportions. The right axis shows color coding for the column variable (Major). The right axis also shows the relative proportions for the column variable, combined across the levels of the row variable. In Figure 12.2, the bar at the right axis shows the relative proportions of Statistics and Other majors, combined across the Male and Female students.

For variables with two levels, JMP uses red and blue color coding. For variables with more levels, JMP selects a color scheme.

JMP uses the data to scale the width of the columns in the mosaic plot. For the Stat Majors data, 8 of the 20 students are female, and JMP scales the Female column accordingly.

Look at the left side of the plot. The left side shows that over 75% of the Female students are majoring in Statistics (the blue rectangle). The right side of the two-way mosaic plot shows that half of the Male students are majoring in Statistics. This difference gives a visual hint that the pattern of majoring in Statistics or Other differs for Male and Female students. Later, “Performing Tests for Independence” shows how to check this visual hint to see whether it leads to a statistical conclusion.

The two-way mosaic plot provides a visual summary of the two variables. If the pattern for the row variable is the same across the values of the column variable, then the color patterns—and the relative proportion for each color—is the same from left to right across the mosaic plot. If the pattern differs, then the relative proportion for each color differs also.
Figure 12.2  Contingency Results for Stat Majors Data

Contingency Analysis of Major By Gender

Mosaic Plot

<table>
<thead>
<tr>
<th>Major</th>
<th>Female</th>
<th>Male</th>
<th>Other</th>
<th>Statistics</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>1</td>
<td>7</td>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>Total %</td>
<td>5.00</td>
<td>30.00</td>
<td>35.00</td>
<td></td>
</tr>
<tr>
<td>Col %</td>
<td>12.50</td>
<td>50.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Row %</td>
<td>14.29</td>
<td>85.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Count</td>
<td>7</td>
<td>13</td>
<td>6</td>
<td>20</td>
</tr>
<tr>
<td>Total %</td>
<td>35.00</td>
<td>65.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col %</td>
<td>87.50</td>
<td>50.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Row %</td>
<td>53.85</td>
<td>46.15</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Tests
### Understanding the Contingency Table Report

The Contingency Table report summarizes the data table. The top row shows the values for the column variable, and the leftmost column shows the values for the row variable.

The outside edges of the table give *row totals* and *column totals*. Looking at the row totals, 7 students are **Female**, and 13 are **Male**. Looking at the column totals, 12 students are majoring in **Statistics**, and 8 are majoring in something else (**Other**). The outside edges also give the *row percentages* and *column percentages*. Looking at the row percentages, 35% (calculated from 7/20) of the students are **Female**. Looking at the column percentages, 60% (calculated from 12/20) of the students are majoring in **Statistics**.

The lower-right corner cell gives the overall total for the table, which is 20. Because there are no missing values, this number matches the number of observations in the data table. However, if the data table contained the **Gender** and not the **Major** for a student, the summary table would contain only 19 observations.

The top-left corner cell gives a key to understanding the main body of the table. The list gives details:

- **Count**: Number of observations in each cell. This class has 1 female student who is majoring in **Other**.
- **Total %**: Percentage of the total number of observations represented by the cell count. The single **Female** student majoring in **Other** represents 5% of the class. The class has 20 students, so 1/20=5%.
- **Col %**: Percentage of observations in the column represented by the cell frequency. The single **Female Other** student represents 12.5% of the **Other** majors in the class. The class has 8 students majoring in **Other**, so 1/8=12.5%.
- **Row %**: Percentage of observations in the row represented by the cell frequency. The single **Female Other** student represents 14.29% of the females in the class. The class has 7 female students, so 1/7=14.29%.

The **Col %** values sum to 100% for each column.

The **Row %** values sum to 100% for each row.
Enhancing the Report

The Contingency Table report has a hot spot that provides options for hiding the summary statistics that JMP automatically displays, and for adding more statistics. In JMP, click the hot spot for Contingency Table. Select a checked item to hide it. Select an unchecked item to display it. JMP uses the three unchecked items (Expected, Deviation, and Cell Chi Square) when performing statistical tests.

The Contingency Table report differs from most other reports in JMP. You cannot double-click on the cells in the summary table to change their appearance. JMP displays percentages to two decimal places.

The Contingency Analysis of Major By Gender report title (at the top of the window) has a hot spot. It provides options for closing the Mosaic Plot, the Contingency Table report, and the Tests report that JMP automatically displays. This hot spot also has options for adding more analyses. See the JMP documentation for more details.

The Contingency Analysis of Major By Gender hot spot provides a feature for changing the display. Suppose you want to rotate the mosaic plot so that the row variable displays as rows in the plot. Some people find this easier to view, because the rows of the mosaic plot are the same as the rows in the Contingency Table report. In JMP:

1. Click the hot spot for the Contingency Analysis of Major By Gender report.
2. Select Display Options→Horizontal Mosaic.

Figure 12.3 shows the results.
Sometimes, you already have a summary table of the data. You can create a JMP data table that matches the summary table, and then analyze it. It is not necessary to expand the summary table into a data table that has one row for every observation in the summary table. Table 12.2 shows frequency counts for several court cases. The columns show the defendant’s race, and the rows identify whether the death penalty was imposed after the defendant was convicted of homicide.²

Table 12.2 Death Penalty Data

**Defendant’s Race**

<table>
<thead>
<tr>
<th>Decision</th>
<th>Black</th>
<th>White</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>149</td>
<td>141</td>
</tr>
<tr>
<td>Yes</td>
<td>17</td>
<td>19</td>
</tr>
</tbody>
</table>

This data is available in the Penalty data table in the sample data for the book.

To create the data table in JMP, follow the steps in Chapter 3. Your data table will have four rows, one for each cell in the table. It will have three columns, one for decision, one for race, and one for count.

The data table itself summarizes the data because it displays a spreadsheet-like view. However, the results from Contingency provide more features, including the two-way mosaic plot, percentages, and statistical tests. In JMP:

1. Open the Penalty data table.
2. In the JMP Starter window, click Basic$\rightarrow$Contingency.
3. Click Defendant Race$\rightarrow$Y, Response Category.
4. Click Decision$\rightarrow$X, Grouping Category.
5. Click Count$\rightarrow$Freq.
6. Click OK.

As with the raw data table, the $y$ variable forms the columns of the summary table, and the $x$ variable forms the rows of the summary table. The Freq variable tells JMP the frequency count for each cell. Without a Freq variable, JMP shows a single observation for each cell.

Figure 12.4 shows the results, with the Tests report hidden.
Figure 12.4 Contingency Results with a Freq Variable

Contingency Analysis of Defendant Race By Decision

Mosaic Plot

Freq: Count

Contingency Table

<table>
<thead>
<tr>
<th>Defendant Race</th>
<th>Black</th>
<th>White</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>149</td>
<td>141</td>
<td>290</td>
</tr>
<tr>
<td></td>
<td>45.71</td>
<td>43.25</td>
<td>88.96</td>
</tr>
<tr>
<td></td>
<td>89.76</td>
<td>88.13</td>
<td></td>
</tr>
<tr>
<td></td>
<td>51.38</td>
<td>48.62</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>17</td>
<td>19</td>
<td>36</td>
</tr>
<tr>
<td></td>
<td>5.21</td>
<td>5.83</td>
<td>11.04</td>
</tr>
<tr>
<td></td>
<td>10.24</td>
<td>11.88</td>
<td></td>
</tr>
<tr>
<td></td>
<td>47.22</td>
<td>52.78</td>
<td></td>
</tr>
<tr>
<td></td>
<td>166</td>
<td>160</td>
<td>326</td>
</tr>
<tr>
<td></td>
<td>50.92</td>
<td>49.08</td>
<td></td>
</tr>
</tbody>
</table>

Tests
JMP produces the same plots and reports with a Freq variable as it does when creating a summary table from raw data. Look at Figure 12.4, between the Mosaic Plot and the Contingency Table report, and you see the only difference. JMP identifies the Freq variable with the text Freq: Count.

Compare the mosaic plots in Figures 12.2 and 12.4. For the Penalty data, the proportion of red and blue blocks (black and white defendants) is roughly the same for both the Yes and No decisions. Figure 12.4 might lead you to initially conclude that the two variables are not related. “Performing Tests for Independence” uses a statistical test to check this initial conclusion. Figure 12.4 scales the width of the columns according to the values of the X, Grouping Category variable. The Yes column is much narrower than the No column, reflecting the small percentage of Yes decisions.

When using a Freq variable, you can change the statistics displayed in the Contingency Table report just as you can for summary tables that are created from raw data.

Creating Contingency Tables for Several Variables

To create contingency tables for several variables, start with the approach for creating contingency tables for two variables. Select a row variable and a column variable, and add By variables to create multiple contingency tables. Chapter 8 shows how to use a By variable in Distribution. The same approach works in Contingency. For an example, use the Cars 1993 data. In JMP:

1. Click HelpÆSample Data Directory.
2. Click the disclosure diamond for Exploratory Modeling.
3. Select **Cars 1993** and JMP opens the data table.

4. Click the X in the upper-right corner of the Sample Data Directory window to close the window.

5. In the JMP Starter window, click **Basic ➔ Contingency**.

6. Click **Passenger Capacity ➔ Y, Response Category**.

7. Click **Vehicle Category ➔ X, Grouping Category**.

8. Click **Domestic Manufacturer ➔ By**.

9. Click **OK**.

   JMP displays a warning message, informing you that **Passenger Capacity** is not a nominal or ordinal variable. The message asks you to consider canceling the action. In this case, **Passenger Capacity** is defined as a continuous variable in the data table, but it is actually an ordinal variable. As a result, the warning message can be ignored. In contrast, the multiple **Cost** and **Mileage** variables are correctly defined as continuous variables, and using these variables in a summary table is inappropriate.

10. Click **Continue**.

   JMP creates a summary table for each level (0 and 1) of **Domestic Manufacturer**. This variable follows a coding convention where 0 means ‘no’ and 1 means ‘yes’. When a row in the data table has **Domestic Manufacturer=0**, that row corresponds to a foreign car. Figure 12.5 displays results, with the Contingency Table and Tests reports hidden.
Figure 12.5 Contingency Results Using a By Variable
JMP displays the summaries for the two levels of Domestic Manufacturer in separate reports. Figure 12.5 displays the mosaic plots for each level of Domestic Manufacturer. The mosaic plots show that the domestic cars (1) differ from foreign cars (0). As one example of a difference, there are no Large foreign cars.

JMP uses multiple colors in the mosaic plot because the variables have more than two levels. Using a By variable has no effect on the colors in the mosaic plot. See “Revising Colors in Mosaic Plots” for a way to enhance the mosaic plot.

You can enhance reports with a By variable using the same JMP features discussed earlier.

At the end of this chapter, see “Special Topic: Statistical Summary Tables” for a way to summarize several variables in a single table. See “Further Reading” at the end of the book or refer to JMP documentation for references on analyses involving more than two classification variables.

**JMP Hint:**
As you become familiar with JMP, you might want to select AnalyzeÆFit Y by X instead of selecting Contingency from the JMP Starter window. Both choices launch the same JMP platform.

One advantage of Contingency is that JMP provides a warning window when you select a continuous variable. With Fit Y by X, JMP assumes that you want to perform a different analysis and does not create the contingency table.

### Revising Colors in Mosaic Plots

JMP uses a red and blue color scheme for variables with two levels in mosaic plots. For variables with multiple levels, JMP uses multiple colors. When printing these mosaic plots on a gray scale printer, the colors can be difficult to distinguish. You can revise the colors to use shades of gray. In JMP:

1. Right-click on the mosaic plot and select Set Colors. If the Set Colors option does not display, carefully place your mouse pointer so that it is on the mosaic plot.
2. JMP shows the old and new colors in the Select Colors for Values window. When the window first appears, the two colors are the same because you have not changed the colors yet, as shown in the following window.

![Select Colors for Values Window](image1)

3. Click the right colored oval for 2 and select the lightest shade of gray from the color palette, as shown in the following window.

![Select Color Window](image2)

4. Click the right colored oval for 7 and select black from the color palette.

5. Click the Macros button and select Gradient between ends, as shown in the following window.

![Macro Selection Window](image3)
6. When you are finished changing colors, click OK.

JMP updates the mosaic plot to use the new colors. When changing colors and using a By variable, JMP changes the colors for one mosaic plot at a time. To change the colors for all of the mosaic plots, change them for each level of the By variable.

**JMP Hints:**
- If you want to return to the automatic color choices, click the Macros button and select Revert to old colors.
- If you find that selecting the colored oval is difficult, select the row and select the color button at the far-left side of the window. This displays a color palette and you can change colors from it.

---

**Performing Tests for Independence**

When you collect classification data, you want to know whether the variables are related in some way. For the Penalty data, is the defendant’s race related to the verdict? Does knowing the defendant’s race tell you anything about the likelihood that the defendant will receive the death penalty?
In statistical terms, the null hypothesis is that the row and column variables are independent. The alternative hypothesis is that the row and column variables are not independent. To test for independence, you compare the observed cell frequencies with the cell frequencies that would occur in the situation where the null hypothesis is true. (See “Technical Details: Expected Cell Frequencies” at the end of this section.)

One commonly used test is a Chi Square test, which tests the hypothesis of independence. A test statistic is calculated and compared with a critical value from a Chi-Square distribution. Suppose you want to test the hypothesis of independence at the 10% significance level for the Penalty data. (For more information, see the general discussion about building hypothesis tests in Chapter 6.)

The steps for analysis are the same as the steps for comparing groups:

1. Create a JMP data table.
2. Check the data table for errors.
3. Choose the significance level for the test.
4. Check the assumptions for the test.
5. Perform the test.
6. Make conclusions from the test results.

To check the data table for errors, use Distribution and the tools from Chapter 4.

To test for independence with two classification variables, the assumptions are:

- Data are counts. Practically, this requires that the variables are nominal or ordinal. JMP performs these tests in the Contingency platform for nominal and ordinal variables.
- Observations are independent. The values for one observation are not related to the values for another observation. To check the assumption of independent observations, think about your data and whether this assumption is reasonable. This is not a step where using JMP will answer this question.
Observations are a random sample from the population. You want to make conclusions about a larger population, not about just the sample. For the statistics class, you want to make conclusions about statistics classes in general, not for just this single class. To check this assumption, think about your data and whether this assumption is reasonable. This is not a step where using JMP will answer this question.

Sample size is large enough for the tests. As a general rule, the sample size should be large enough to expect five responses in each cell of the summary table. JMP prints warning messages when this assumption is not met.

JMP automatically performs appropriate tests when creating results from Contingency. Figure 12.6 shows the Tests report for the Penalty data. Figure 12.4 shows the other results, with the Tests report hidden.

**Figure 12.6 Testing for Independence with the Penalty Data**

<table>
<thead>
<tr>
<th>Tests</th>
<th>Source</th>
<th>DF</th>
<th>-LogLike</th>
<th>RSquare (U)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>1</td>
<td>0.11073</td>
<td>0.0005</td>
<td></td>
</tr>
<tr>
<td>Error</td>
<td>324</td>
<td>225.80004</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C. Total</td>
<td>325</td>
<td>225.91076</td>
<td></td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>326</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test</th>
<th>ChiSquare</th>
<th>Prob&gt;ChiSq</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood Ratio</td>
<td>0.221</td>
<td>0.6379</td>
</tr>
<tr>
<td>Pearson</td>
<td>0.221</td>
<td>0.6379</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fisher's Exact Test</th>
<th>Prob</th>
<th>Alternative Hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>0.7412</td>
<td>Prob(Defendant Race=White) is greater for Decision=No than Yes</td>
</tr>
<tr>
<td>Right</td>
<td>0.3843</td>
<td>Prob(Defendant Race=White) is greater for Decision=Yes than No</td>
</tr>
<tr>
<td>2-Tail</td>
<td>0.7246</td>
<td>Prob(Defendant Race=White) is different across Decision</td>
</tr>
</tbody>
</table>

The first report in Figure 12.6 (with Source as the first column heading) is similar to an analysis of variance table for a continuous response variable. See the JMP documentation for more details. The test for independence does not use this report.
Before making decisions about the test, think about the assumptions. The data are counts, so the first assumption is reasonable. The observations are independent, because the race of and decision for a defendant is unrelated to the race of and decision for another defendant. The second assumption is reasonable. The observations are a random sample from the population of defendants convicted of homicide, so the third assumption is reasonable. In the Contingency Table report, use the hot spot to display the Expected values, and confirm that the fourth assumption is reasonable. Also, when this assumption is not met, JMP prints a warning message. Now, look at the results from the test for independence.

**Understanding Chi Square Test Results**

In the mosaic plot, the proportion of white and black defendants was roughly the same for the Yes and No decisions. The mosaic plot led to an initial conclusion that the variables were unrelated. The Chi Square test for independence quantifies the initial conclusion.

Look under the heading Test in Figure 12.6. JMP displays two Chi Square tests. The Pearson and Likelihood Ratio tests both have the same assumptions. The Pearson test uses the observed and expected cell frequencies, and the Likelihood Ratio test uses a more complex formula.

Look at the number under Prob>ChiSq. This value is 0.6379, which is greater than the significance level of 0.10. You conclude that there is not enough evidence to reject the null hypothesis of independence between the defendant’s race and the decision. (Refer to Agresti in “Further Reading” for an additional analysis of this data that considers the race of the victim.) Although the p-values for the two tests are identical in Figure 12.6, this is not always true. Typically, the p-values are similar, but are not identical.

In general, to interpret JMP results, look at the p-value under Prob>ChiSq, in the line for Pearson. If the p-value is less than the significance level, reject the null hypothesis that the two variables are independent. If the p-value is greater, you fail to reject the null hypothesis.
Understanding Fisher’s Exact Test Results

Fisher’s exact test was developed for the special case of a $2 \times 2$ table. This test is very useful when the assumptions for a Chi Square test are not reasonable, and is especially useful for tables with small cell frequencies. JMP automatically performs this test for $2 \times 2$ tables, but it isn’t available for larger tables.

Look under the heading Fisher’s Exact Test in Figure 12.6. JMP displays results for both one-sided tests and the two-sided test. Look at the line labeled 2-Tail. The $p$-value is $0.7246$, so you fail to reject the null hypothesis that the two variables are independent.

In general, to interpret JMP results, look at the 2-Tail $p$-value, which tests for independence between the two variables. The one-sided Left and Right $p$-values might be useful in specific situations. Interpret the $p$-value the same way you do for the Chi Square test.

Enhancing the Tests Report

You can double-click on a column in the Tests report and change the format of the numbers that are displayed. As with other reports in JMP, you can change the report titles.

Position the mouse pointer near a $p$-value and move it around in a very small circle. JMP displays a pop-up window, which includes text explaining the $p$-value. To close the window, move your mouse pointer again. JMP highlights $p$-values less than 0.05 with an asterisk.

To hide the Tests report, click the disclosure diamond next to the report title.

To remove the Tests report, click the hot spot at the top of the Contingency report. Click the Tests selection so that it is deselected, and JMP removes the report and its title.
Technical Details: Expected Cell Frequencies

The Chi Square test compares the observed cell frequencies with the expected cell frequencies, assuming the null hypothesis that the variables are independent. Click the hot spot for the Contingency Table report to display the Expected frequencies. To calculate the expected cell frequencies, multiply the row total and column total, and divide by the total number of observations. For the Black-No cell, the formula is:

\[
\frac{[(\text{row total for No}) \times (\text{column total for Black})]}{\text{total } N}
\]

\[
= \frac{(290 \times 166)}{326}
\]

\[
= 147.67
\]

The Chi Square test is always valid if there are no empty cells (no cells with a cell frequency of 0), and if the expected cell frequency for all cells is 5 or greater.

Because these conditions are true for the Penalty data, the Chi Square test is a valid test. If these two conditions are not true, JMP prints a message warning that the Chi Square test might not be valid.

There is some disagreement among statisticians about exactly when the test should not be used, and what to do when the test is not valid. Two practical recommendations are to collect more data, or to combine low-frequency categories.

Combine low-frequency categories only when it makes sense. For example, consider a survey that asked people how often they called a Help Desk in the past month. Suppose the categories are: 0 (no calls), 1-2, 3-5, 6-8, 9-11, 12-15, 16-20, and “over 20”. Now, suppose the expected cell counts for the last four categories are less than 5. It makes sense to combine these last four categories into a new category of “9 or more”.

Measures of Association with Ordinal Variables

When you reject the null hypothesis for either the Chi Square test or Fisher’s exact test, you conclude that the two variables are not independent. But, you do not know how the two variables are related. When both variables are ordinal, measures of association give more insight into your data. As defined in Chapter 3, ordinal variables have values that provide names and an implied order.
Two measures of association are Kendall’s tau and Spearman’s rank correlation coefficient. Both measures range from -1.0 to 1.0. Values close to 1.0 indicate a positive association, and values close to -1.0 indicate a negative association. (The Spearman’s rank correlation coefficient is similar to correlation coefficients discussed in Chapter 10. This correlation coefficient is essentially the Pearson correlation, applied to ranks instead of to actual values.)

An animal epidemiologist tested dairy cows for the presence of a bacterial disease. The disease is detected by analyzing blood samples, and the disease severity for each animal was classified as none (0), low (1), or high (2). The size of the herd that each cow belonged to was classified as small (10), medium (100), or large (1000). Table 12.3 shows the number of cows in each herd size, and in each disease severity category. Both the herd size and disease severity variables are ordinal. The disease is transmitted from cow to cow by bacteria, so the epidemiologist wanted to know whether disease severity is affected by herd size. As the herd size gets larger, is there either an increasing or a decreasing trend in disease severity?

The epidemiologist tested for independence using the Chi Square test. However, she still does not know whether there is a trend in disease severity that is related to an increase in herd size. Kendall’s tau or Spearman’s rank correlation coefficient can answer this question.

### Table 12.3 Cow Disease Data

<table>
<thead>
<tr>
<th>Herd Size</th>
<th>Disease Severity</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>None (0)</td>
</tr>
<tr>
<td>Small (10)</td>
<td>9</td>
</tr>
<tr>
<td>Medium (100)</td>
<td>18</td>
</tr>
<tr>
<td>Large (1000)</td>
<td>11</td>
</tr>
</tbody>
</table>

This data is available in the Cow Disease data table in the sample data for the book. The icons for Herd Size and Disease indicate that these two variables are ordinal, as shown in the following window. The variables are also numeric, which JMP requires for the Multivariate platform.

---

3 Data is from Dr. Ramon Littell, University of Florida. Used with permission.
To create the data table in JMP, follow the steps in Chapter 3. Your data table will have 9 rows, one for each cell in the table. It will have 3 columns, one for herd size, one for disease severity, and one for number.

JMP requires numeric variables to create Kendall’s $\tau$ and Spearman’s rank correlation coefficient. JMP provides these statistics in the Multivariate platform. This platform automatically creates a scatterplot matrix that is useful for investigating multiple continuous variables. To create the measures of association in JMP:

1. Open the Cow Disease data table.
2. In the JMP Starter window, click Multivariate $\rightarrow$ Multivariate.
3. Click Herd Size and Disease and then Y, Columns.
4. Click Number $\rightarrow$ Freq.
5. Click OK.

JMP displays a warning message, informing you that Herd Size and Disease are not continuous variables. The message asks you to consider canceling the action. In this case, both variables are ordinal. As a result, the warning message can be ignored. Using these two variables to create the measures of association is appropriate.

6. Click Continue. JMP creates the results.
7. Hold down the ALT key and click the hot spot for the Multivariate report.
8. Select Correlations Multivariate to close this report. It makes sense for continuous variables.
9. Select **Scatterplot Matrix** to close the plot. It makes sense for continuous variables.

10. Select **Nonparametric Correlations** → **Spearman’s \( \rho \)**. (The \( \rho \) is the Greek letter rho, which is how statisticians often refer to Spearman’s rank correlation coefficient.)

11. Select **Nonparametric Correlations** → **Kendall’s \( \tau \)**. (The \( \tau \) is the Greek letter tau.)

12. Click **OK**.

Figure 12.7 shows the results.

**Figure 12.7 Measures of Association for the Cow Disease Data**

```

<table>
<thead>
<tr>
<th>Multivariate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Freq: Number</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nonparametric: Spearman's ( \rho )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable by Variable</td>
</tr>
<tr>
<td>Disease</td>
</tr>
<tr>
<td>---------</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Nonparametric: Kendall's ( \tau )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable by Variable</td>
</tr>
<tr>
<td>Disease</td>
</tr>
<tr>
<td>---------</td>
</tr>
</tbody>
</table>
```

Figure 12.7 highlights the **Freq** variable by showing **Freq: Number** just below the **Multivariate** report title. JMP performs exactly the same analyses with raw data. The only difference in the results is that the report does not show a **Freq** variable. The plots and reports are similar to the **Pairwise Correlations** results, discussed in Chapter 10.

**Understanding the Plot**

For both measures of association, JMP creates a bar chart scaled from -1 to 1. JMP colors from the value 0 to the bar segment that matches the value of the measure of association. Both Spearman’s rank correlation coefficient and Kendall’s \( \tau \) have a value of about 0.2, so the same area is colored for each plot.
Understanding the Report

For each measure of association, JMP lists the two variables involved, the measure, and a p-value.

A positive measure of association indicates an increasing trend between the two variables. As the ordinal levels of one variable increase, so do the ordinal levels of the other variable. For the Cow Disease data, the epidemiologist can conclude that disease severity increases with increasing herd size.

A negative measure of association indicates a decreasing trend between the two variables. As the ordinal levels of one variable increase, the ordinal levels of the other variable decrease.

The closer a measure of association is to 0, the weaker the strength of the relationship. A value of -0.8 or 0.8 indicates a stronger relationship than a value of -0.1 or 0.1.

The p-value is the result of a test that the measure of association is significantly different from 0. For this test, the null hypothesis is that the measure of association is 0. The alternative hypothesis is that the measure of association is different from 0. JMP highlights p-values less than 0.05 with an asterisk.

For the Cow Disease data, you conclude that the measures of association are significantly different from 0.

In general, if the p-value is less than the significance level, reject the null hypothesis and conclude that the measure of association is significantly different from 0. If the p-value is greater than the significance level, fail to reject the null hypothesis. Conclude that there is not enough evidence to say that the measure of association is significantly different from 0.

JMP Hint:

As you become familiar with JMP, you might want to select Analyze→Multivariate Methods→Multivariate instead of selecting Multivariate→Multivariate from the JMP Starter window. Both choices launch the same JMP platform.

One advantage of selecting Multivariate in the JMP Starter window is that JMP provides a warning window for an ordinal variable. Click Continue and create reports with measures of association for ordinal variables. When you select Analyze from the menu, JMP requires continuous variables in the Multivariate platform.
Chapter 12: Creating and Analyzing Contingency Tables

Summaries

Key Ideas

- Contingency tables are tables that summarize two or more classification variables. The rows and columns of a contingency table form cells, and the number of observations in a cell is the cell frequency for that cell.

- The Basic→Contingency choice in the JMP Starter window creates contingency tables and a two-way mosaic plot, and performs statistical tests on the two variables. Use nominal or ordinal variables in the Contingency platform. As you become familiar with JMP, you might want to select Analyze→Fit Y by X from the menu instead of selecting Contingency in the JMP Starter window.

- Use By variables to summarize more than two variables. JMP creates separate reports for each level of the By variable. With multiple By variables, JMP creates a separate report for each combination of levels of the multiple By variables.

- Both the Chi Square test and Fisher’s exact test are used to test for independence between two classification variables. Generally, the Chi Square test should not be used if there are empty cells or if cells have expected cell frequencies of less than 5. One option is to collect more data, which increases the expected cell frequency. Another option is to combine levels that have only a few observations.

- Regardless of the test, the steps for analysis are:
  1. Create a JMP data table.
  2. Check the data table for errors.
  3. Choose the significance level for the test.
  4. Check the assumptions for the test.
  5. Perform the test.
  6. Make conclusions from the test results.
Regardless of the test, to make conclusions, compare the \( p \)-value for the test with your significance level.

- If the \( p \)-value is less than the significance level, reject the null hypothesis that the two variables are independent. JMP displays an asterisk next to \( p \)-values that are less than 0.05.
- If the \( p \)-value is greater than the significance level, fail to reject the null hypothesis.

Kendall’s \( \tau \) and Spearman’s rank correlation coefficient are both measures of association that provide information about how strongly related ordinal variables are. Use these statistics to decide whether there is an increasing or a decreasing trend in the two variables, or if there is no trend at all. JMP tests whether the measure of association is significantly different from 0, and reports the \( p \)-value.

The Multivariate \( \rightarrow \) Multivariate choice in the JMP Starter window creates measures of association for the two variables. Use numeric ordinal variables in the Multivariate platform. As you become familiar with JMP, you might want to select Analyze \( \rightarrow \) Multivariate Methods from the menu instead of selecting Multivariate in the JMP Starter window, but remember that it requires numeric continuous variables.

---

**JMP Steps**

To summarize two nominal or ordinal variables in a contingency table:

1. In the JMP Starter window, click Basic \( \rightarrow \) Contingency.
2. Click the column variable and then \( Y, \) Response Category.
3. Click the row variable and then \( X, \) Grouping Category.
4. If your data already form a summary table, click Count \( \rightarrow \) Freq.
5. Click OK.

To summarize three or more variables in a contingency table:

1. Complete steps 1 through 4 above.
2. Before step 5, click the additional variables and then By. JMP creates a contingency table for each combination of levels of the additional variables.
3. Click OK.
To rotate the mosaic plot:

1. Click the hot spot for the Contingency Analysis report.
2. Click Display Options→Horizontal Mosaic.

To revise colors to gray scale in the mosaic plot:

1. Right-click on the mosaic plot and select Set Colors. If the Set Colors option does not display, carefully place your mouse pointer so that it is on the mosaic plot.
2. JMP shows the old and new colors in the Select Colors for Values window. When the window first appears, the two colors are the same because you have not changed colors yet.
3. Click the right colored oval for the top color in the list and select the lightest shade of gray.
4. Click the right colored oval for the bottom color in the list and select black from the color palette.
5. Click the Macros button and select Gradient between ends.
6. Click OK.

To perform tests for independence:

Follow the steps above for creating a contingency table. JMP automatically creates the Tests report, which contains results for the Chi Square test. For 2×2 tables, JMP includes results from Fisher’s exact test. However, see the steps for analysis in “Key Ideas.” Although JMP automatically performs the tests, you still need to check assumptions and think about your data.

To generate measures of association:

The measures of association discussed in this chapter require two ordinal variables. JMP requires numeric variables.

1. In the JMP Starter window, click Multivariate→Multivariate.
2. Click the row variable and column variable and then Y, Columns.
3. If the data already form a summary table, click the Count→Freq.
4. Click OK.

If your variables are ordinal, JMP displays a warning message, informing you that the variables are not continuous variables. The message asks you to
consider canceling the action. When the variables are numeric ordinal variables, creating measures of association is appropriate.

5. Click **Continue**. JMP creates the results.

6. Hold down the ALT key and click the hot spot for the **Multivariate** report.

7. Select **Correlations Multivariate** to close this report. It makes sense for continuous variables.

8. Select **Scatterplot Matrix** to close the plot. It makes sense for continuous variables.

9. Select **Nonparametric Correlations**→**Spearman’s**ρ.

10. Select **Nonparametric Correlations**→**Kendall’s**τ.

11. Click **OK**.

---

### Exercises

1. Perform a test for independence for the **Stat Majors** data. Are the assumptions for the test reasonable? Define the null and alternative hypotheses. Use a 10% alpha level. Discuss your conclusions.

2. Test for independence with the **Cow Disease** data. Are the assumptions for the test reasonable? Define the null and alternative hypotheses. Use a 5% alpha level. Discuss your conclusions. Does it make a difference in the reports and graphs if you make the variables ordinal? Do the results of the statistical test change?

3. Create a new data table for the **Cow Disease** data. Use 10, 20, and 30 for the **Small, Medium**, and **Large** values of **Herd Size**. Use 0, 4, and 8 for the **None, Low**, and **High** values of **Disease Severity**. Repeat the analysis in the chapter, creating Kendall’s tau and Spearman’s rank correlation coefficient. Do the results differ? Do the values used for a variable have an effect on these statistics?

4. From the **Sample Data Directory**, click the disclosure diamond for **Categorical Models** and open the **Alcohol** data table. Create a summary table of **Relapsed** by **Alcohol Consumption**. Rotate the mosaic plot.
5. From the Sample Data Directory, click the disclosure diamond for Business and Demographic and open the Movies data table. Create a summary table of Type by Rating. Are the assumptions for the test for independence reasonable? If so, define the null and alternative hypotheses and use a 5% alpha level. Discuss your conclusions. Does it make sense to create measures of association for this data? If so, generate them and discuss your conclusions.

6. From the Sample Data Directory, click the disclosure diamond for Business and Demographic and open the Titanic data table. Create a summary table of Survived by Class. Are the assumptions for the test for independence reasonable? If so, define the null and alternative hypotheses and use a 5% alpha level. Discuss your conclusions. Does it make sense to create measures of association for this data? If so, generate them and discuss your conclusions.

7. From the Sample Data Directory, click the disclosure diamond for Exploratory Modeling and open the Cars 1993 data table. Create a summary table of Vehicle Category by Passenger Capacity. Are the assumptions for the test for independence reasonable? If so, define the null and alternative hypotheses and use a 5% alpha level. If not, explain whether it makes sense to combine categories and discuss which categories you would combine.

8. With the Cars 1993 data table from exercise 7, create a summary table of Vehicle Category by Domestic Manufacturer. Are the assumptions for the test for independence reasonable? If so, define the null and alternative hypotheses and use a 5% alpha level. Discuss your conclusions. Does it make sense to create measures of association for this data? If so, generate them and discuss your conclusions.

9. From the sample data for the book, open the Cereal Revised data table. Create a summary table of Enriched by Fiber Gr. Are the assumptions for the test for independence reasonable? If so, define the null and alternative hypotheses and use a 5% alpha level. Perform the test. Create measures of association for these variables. Discuss your conclusions.

10. From the Sample Data Directory, click the disclosure diamond for Exploratory Modeling and open the Mushroom data table. Create a summary table of Odor by Habitat for each level of Edibility. (Use Edibility as a By variable.) Are the assumptions for the test for independence reasonable? If so, define the null and alternative hypotheses and use a 5% alpha level. Discuss your conclusions, and whether results differ for edible and poisonous mushrooms.
Special Topic: Statistical Summary Tables

The Contingency platform with By variables summarizes three or more classification variables. JMP provides two other tools that summarize multiple variables in a single table. This section gives a brief introduction to the TablesÆSummary platform. JMP also provides the TablesÆTabulate platform, which has a drag-and-drop interface for creating tables. The Tabulate platform has some similarities with creating pivot tables in Microsoft Excel.

This example uses the Cars 1993 data and creates a summary table. The table condenses the two tables from the Contingency platform (used with a By variable earlier in the chapter) into a single table. In JMP:

1. Click HelpÆSample Data Directory.
2. Click the disclosure diamond for Exploratory Modeling.
3. Click Cars 1993.jmp. JMP opens the data table.
4. Click the X in the upper-right corner of the Sample Data Directory window to close it.
5. Click TablesÆSummary.
6. Click Domestic ManufacturerÆGroup.
7. Click Vehicle CategoryÆGroup.
8. Click Passenger CapacityÆGroup. Compare your window with Figure 11S.1 to confirm your choices.
9. Click OK.
Figure 11S.1 Window for Summary Table

Figure 11S.2 shows the results.

The N Rows column gives the frequency count for the combination of levels for the three variables in that row. For example, the first row in the data table reports a single car that is from a foreign manufacturer (Domestic Manufacturer equal to 0), is Compact, and can carry 4 passengers.
Figure 11S.2 Summary Table for Cars 1993 Data

<table>
<thead>
<tr>
<th>Cars 1993 By (Domestic Manufacturer, Vehicle Category, Passenger Capacity)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domestic Manufacturer</td>
</tr>
<tr>
<td>------------------------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>11</td>
</tr>
<tr>
<td>12</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>15</td>
</tr>
<tr>
<td>16</td>
</tr>
<tr>
<td>17</td>
</tr>
<tr>
<td>18</td>
</tr>
<tr>
<td>19</td>
</tr>
<tr>
<td>20</td>
</tr>
<tr>
<td>21</td>
</tr>
</tbody>
</table>

JMP Hint:
The order of variables is important in the Tables → Summary platform. JMP uses the order to create the columns in the new data table. If you hold down the CTRL key and click to select variables in the Select Columns area, JMP uses the columns in the order they appear in the data table. This might not be what you want.

One advantage of using Tables → Summary is that the results are in a data table. You can create bar charts, or use the data in any appropriate JMP platform.

This special topic is an extremely brief introduction to summarizing multiple variables in tables. The Summary platform can create very complex tables, with means, standard deviations, and other statistics in columns. The Tabulate platform can create similarly complex tables, with the distinction that these tables are in a report window, rather than in a data table. Discussing all of the features of these two platforms would require another book!

Index

A

“After row” option (Move Rows)  49
All Graphs option (Display Options)  227, 243
alpha level (α-level)  166–167, 242, 256
  means diamonds  265
  See also specific test by name
ALT key, and hot spots  225
alternative hypotheses  164, 168
  See also specific test by name
  comparing multiple groups  256
  comparing two independent groups  205
  independence tests for contingency plots  420
  lack of fit tests  384
analysis of variance
  See ANOVA
Analysis of Variance report
  multiple regression  351
  polynomial regression  343
  straight-line regression  330
analysis platforms  13
Annotate tool
  See sticky notes
ANOVA (analysis of variance)
  means diamonds  265
  multiple comparison tests and  293
  one-way  258, 259–265
  unequal variances  266–270
  Welch Anova test vs.  269–270
arithmetic average of random samples
  174–175, 179–183
assumption of equal variances  260, 295
  multiple groups  266–269
  two groups  208–212, 242
assumption of independent observations  260
assumption of normality
  See normality testing
“At Start” option (Move Rows)  49
Attempt to Discern Format button  32
automatic value orders  53
average
  See also population mean
  sample average  145
axes, scaling  98–100, 134
  bar charts  119–120
  uniform scaling  99–100, 199
Axis Specification window  120–121, 134, 331

B

balanced data  258
bar charts  114–121, 136
  See also histograms
  colors  119–120, 136
  continuous variables in  114, 136
  histograms vs.  114
  resizing  119–120
  scaling axes  119–120
  sorting data tables for  114–115
Bartlett test  269
Basic Category, JMP Starter window  73–74
bell-shaped curves
  See normal distributions
bitmaps, pasting results as  128, 135
Bivariate platform  324–335, 340–345, 364, 366, 376
  See also regression
diagnostics
  See also regression diagnostics
blue disclosure diamonds  13, 37, 75
Bonferroni approach  281–285, 293
box plots  109–110
  See also graphs
  colors  82
  for continuous variables  81–82
  median in  81
  normality testing  154–155, 162
  selecting and deselecting features  82
  shortest half  81
  whiskers in  81
Box Plots option (Display Options) 227, 243
Brown-Forsythe test 269
By column, Distribution platform 197–199, 254
  contingency tables for several variables 414–419

C
Cancel option (Close All Windows) 20
causation vs. correlation 319
cell frequency, contingency tables 405, 424
Centered Polynomial option 341
centering polynomials 341, 344
Central Limit Theorem 179–182, 189
  confidence interval and 184
  Empirical Rule and 182–183
character variables 51
charts
  See also specific chart type by name
  printing 127–129
Chi Square tests 420–422, 424
ChiSquare statistic 272
CI
  See confidence intervals for mean
Clear Row States option 325
Close All Below option 96
Close All Like This option 95, 107
Close All Windows window 19–20
closing JMP 19–20
closing reports and graphs 97–98
coefficient of variation 85
colors
  bar charts 119–120, 136
  box plots 82
  histograms 78, 109
  mosaic plots 407, 417–419, 431
  scatterplots 313–314
  treemaps 123–127, 137
Cols menu 37, 43
column information 67
Column Information window 47, 67
column totals (Contingency Table report) 409
columns in data tables 41–47
  creating data tables manually 24–28, 64
data types and modeling types 50–52
  excluding and unexcluding 46–47
  headings for 101–102, 107
  hiding and unhiding 45–47, 66, 67, 93, 108
  hot spots and 37, 39
  naming 25
  ordering in treemaps 122–123, 137
  ordering values in 53–55, 67–68
  paired data in 246–248
  resizing 42
  scroll locking 43, 66
  selecting for summaries 90
  viewing information on 47, 67
columns in reports, hiding 93, 108
columns panel 37–38
commands, JMP 13
comments
  See sticky notes
comparing multiple groups 249–302
  ANOVA with unequal variances 266–270
  Bonferroni approach 281–285, 293
  building hypothesis tests 257–259
  comparison circles 275–277
  Dunnett’s test 290–292
  Each-Pair, Student’s t test 277–282
  enhancing graphs for 273–274
  equal variances 266–269
  Hsu’s MCB test 288–290
  Kruskal-Wallis tests 270–273, 297
  one-way ANOVA 259–265
  summarizing data 251–257, 300
  Tukey-Kramer HSD test 281, 286–288, 293, 298
  Welch ANOVA 266–270
comparing two groups 193–243
  building hypothesis tests 205–207
enhancing two-sample graphs 224–228, 242–243
independent vs. paired groups 195
paired data in single column 246–248
paired-difference $t$-tests 206, 228–235
summarizing data from independent groups 196–199, 241
summarizing data from paired groups 199–204, 241
two-sample $t$-tests 206, 207–220, 242
Wilcoxon rank sum tests 206, 220–224, 242
Wilcoxon signed rank tests 206, 235–238, 242
comparison circles 275–277
deselecting 291
comparisonwise error rate 275, 281
Confid Quantile report 280
  Bonferroni approach 285
  Hsu’s MCB test 288
  Tukey-Kramer HSD test 288
confidence intervals for mean 83, 184–189, 189–190
See also alpha level ($\alpha$-level)
Central Limit Theorem and 184
choosing (changing in JMP) 187–188
Empirical Rule and 184
one-sided 188
polynomial regression 344–345
straight-line regression 334–335
two-sample $t$-tests 219–220, 242
confidence limits on predicted values
  multiple regression 351–352, 357
  polynomial regression 344–345
  straight-line regression 332
Connect Means option (Display Options) 227
Connecting Letters Report 279, 280
  Bonferroni approach 285
  example of 298
  Tukey-Kramer HSD test 286
contingency analysis mosaic plots 407
colors in 407, 417–419, 431
  rotating to horizontal 410–411, 431
Contingency platform
  See contingency tables
Contingency Table report 409–410
contingency tables 403–432
  cell frequency 405, 424
  creating for several variables 414–419
  creating from existing summary tables 411–414
defining 404–405
  measures of association (ordinal variables) 424–428, 431–432
  performing tests for independence 419–424, 431
statistical summary tables 434–436
summarizing raw data 405–411, 430
continuous variables 50–52, 86–88
  box plots 81–82
  checking for errors 73, 88
  Distribution platform 73–88
  histograms 76–80
  in bar charts 114, 136
  missing values 88
  moments reports 83–86
  multiple, summarizing 309–316
  quantiles reports 82–83
  stem-and-leaf plots 86–88
  summarizing data, in general 73–76
copying results to Microsoft applications 128–133, 135
correlation coefficients 316–320
  missing values and 317
  $p$-value 318–320
  significance levels 318–320
  Spearman’s rank correlation coefficient 425
correlation vs. causation 319
Correlations report 316–320, 354
Corresponding Letters Report 281
Count (stem-and-leaf plots) 87–88
Count value (Frequencies reports) 92
crosstabulations
  See contingency tables
Cum Prob value (Frequencies reports) 93
curve fitting (polynomial regression) 336–345
  adding confidence curves 344–345
  centering polynomials 341, 344
  customizing two-sample graphs 224–228, 242–243
CV 85

D

data, fitting lines to
  See also regression
  See also regression diagnostics
correlation 316–320, 354, 425
data, importing 29–36
  file preferences 32–34, 65–66
  from Excel spreadsheets 30–31, 64–65
  from text files 32–36, 64–65
data errors, checking for 73, 315–316
  continuous variables 73, 88
  least squares regression and 324
  nominal or ordinal variables 73, 93
  scatter plots for 315–316
data files, importing
  See data, importing
data statistics
  See Moments reports
data subsets 161–162, 171
data summary
  See summarizing data
data tables 22–69
  See also columns in data tables
  creating contingency tables from 411–414
  creating manually 24–28, 48–49, 64
  creating subsets of 161–162, 171
  data types and modeling types 50–52
  defined 13, 22–23
  labeling values 56–57, 68
  locking 41
  opening 29, 64
  ordering values 53–55, 67–68
  paired data in single column 246–248
  printing 57–60, 68
  resizing 37, 38, 66
rows in 24–28, 48–49, 64
saving 28, 381
saving scripts to 271
sorting 60–62, 69, 114–115
viewing and using 36–41
viewing in journals 129
data types 50–52
decimal places in reports 86, 108
Decrease Font Size option 80, 106
degrees of freedom
  comparing two groups 205–206, 216, 232
  for equal and unequal variances tests 216, 269
  lack of fit error and 385
  one-way ANOVA 263
density ellipses (scatterplots) 313–314, 325
deselecting
  See selecting and deselecting
Detailed Comparisons Report 281
  Bonferroni approach 285
  Tukey-Kramer HSD test (not available) 288
DF
  See degrees of freedom
DFNum and DFDen values 211
diamonds
  disclosure diamonds 13, 37, 75
  means diamonds 81, 265
Difference Matrix report 281
  Bonferroni approach 285
  Hsu’s MCB test 290
  Tukey-Kramer HSD test 288
differences between paired groups, finding 199–204
disclosure diamonds 13, 37
  in reports 75
Display Options 224–228, 243, 265, 273–274
distribution of sample averages 179–183
Distribution platform 108
  See also Moments reports
  By column for independent groups 197–199, 254, 414–419
continuous variables 73–88
differences for paired groups 204
interacting with results of 94–100
nominal or ordinal variables 89–93
normality test 151–153
paired differences tests 230–232
plotting sample averages 175–176
summarizing independent groups 196–197
testing error assumptions for residuals 388–392
Wilcoxon signed rank tests 236–237, 242
documentation on JMP 12
Dunnett’s test 290–292

E

Each-Pair, Student’s t test 277–282
Empirical Rule 149
  Central Limit Theorem and 182–183
  confidence interval and 184
End Of Field settings 32
End Of Line settings 32
equal variances (pooled) tests 242, 260, 295
  multiple groups 266–269
  two groups 208–212, 242
equation for fitted lines, calculating 326–328
error rate
  comparisonwise 275, 281
  experimentwise 275, 281–285, 293
Error Sum of Squares statistic 330
errors
  lack of fit error 384
  Root Mean Square Error 268, 281, 388
  standard error of the mean 83, 182, 264
Type I error 166, 168
Type II error 167–168
variance assumption for regression, testing 388–392, 395
variation due to error 257–258
errors in data, checking for 73
continuous variables 73, 88
least squares regression and 324
nominal or ordinal variables 93

scatterplots for 315–316
estimating mean 173–190
See also population mean
certainty intervals for mean 184–189
distribution of sample averages 179–183
point estimates 174–175, 179–183
population variability effects 178–179, 189
sample size effects 175–177, 189
Excel spreadsheets, importing 30–31, 64–65
excluding columns in data tables 46–47
exiting JMP 19–20
expected cell frequencies 424
experimentwise error rate 275, 281
controlling with Bonferroni approach 281–285, 293
exponential distributions 180
exporting tables to Microsoft Word 58–59, 68

F

F Ratio value
  one-way ANOVA 264
two-sample t-tests 211
Field Width button (Text Import Preview) 36
file preferences for importing files 32–34, 65–66
files, importing
  See importing data
Fisher’s exact test 423
Fit Model platform 347
  leverage plots 397–400
  plotting residuals against independent variables 394
  plotting residuals against predicted values 372–373
  plotting residuals against time sequence 373–374, 394–395
residuals from 370
Fit Y by X platform  272, 343, 347
   See also ANOVA
   See also comparing two groups
   contingency tables for several variables  417
fitted lines, equation for calculating  326–328
fitting lines to data
   See also regression
   See also regression diagnostics
correlation  316–320, 354, 425
font size  80, 106
formula window  202
Frequencies reports  92–93

G
Goodness-of-Fit Test report  152
grabber tool (hand icon)  77
Grand Mean option (Display Options)  227, 243
Graph Category, JMP Starter window  116–117
Graph platform for residuals plots
   plotting residuals against independent
      variables  364–366, 371–372
   plotting residuals against predicted values
      366–368
   plotting residuals against time sequence
      368–369
graph size, changing
   See resizing
graphs
   See also resizing
   See also specific graph type by name
closing  97–98
copying to Microsoft applications  128–129, 135
   for multiple groups, enhancing  273–274
   for two samples, enhancing  224–228, 242–243
   hiding  95–96, 107
   printing  127–129
   scaling axes  98–100
groups, comparing multiple
   See comparing multiple groups
   See comparing two groups
   See comparing two groups

H
hand icon (grabber tool)  77
headings for columns  101–102, 107
help on JMP  12, 17–19
hiding and showing columns  45–47, 66, 67, 93, 108
   in Contingency Table reports  410
   reports and graphs  95–96, 107, 110, 331
highlighting
   See selecting and deselecting
histograms
   See also bar charts
   See also box plots
   See also graphs
   See also reports
   bar charts vs.  114
   checking for errors  88
colors  78, 109
customizing appearance of  76–80, 109
   for continuous variables  76–80
   for nominal or ordinal variables  91–92
   normality testing  154–155, 162
   percents in  79
   resizing  77, 80, 106, 109
   scaling axes  98–100
   selecting and deselecting bars  76–77
Histograms option (Display Options)  228, 243
honestly significant difference
   See Tukey-Kramer HSD test
horizontal mosaic plots  410–411, 431
hot spots  13
   ALT key for multiple choices  225
columns and rows  37, 39
reports  76
scatter plot matrices  313
tables  39
HSD
   See Tukey-Kramer HSD test
Hsu’s MCB test 288–290
HTM files, saving journals as 133
hypotheses testing 164–166
See also specific test by name
  comparing multiple groups 257–259
  comparing two groups 205–207
  independence tests for contingency plots 420
  lack of fit tests 384
  one-sided 219

I
images, journals as 133
importing data 29–36
  file preferences 32–34, 65–66
  from Excel spreadsheets 30–31, 64–65
  from text files 32–36, 64–65
  .xls files 30–31
Increase Font Size option 80, 106
independence, testing in contingency plots 419–424, 431
independent groups
  alpha level for confidence intervals 242
  By column for 197–199, 254, 414–419
  paired groups vs. 195
  statistical significance 206–207
  statistical tests for comparing 206
  summarizing data from 196–199
independent observations 260
independent variables, plotting residuals
  against 361–362
  multiple regression 369–372
  polynomial regression 380–381
  straight-line regression 364–366, 377
  with Bivariate platform 393
  with Fit Model platform 394
indexes, in JMP Help 18
influential points (leverage plots) 398
interquartile range 81
  See also box plots
interval estimates
  See confidence intervals for mean
  interval variables 52

J
JMP data tables
  See data tables
JMP 12–20
  commands 13
  documentation 12
  exiting 19–20
  help 12, 17–19
  releases 12
  saving sessions 20
  starting 14
JMP Starter window 14
  Graph Category 116–117
  modeling type selection 89
  new journals 129
  with Basic category selected 73–74
journal files (.JRN) 132
journals 129–133, 135
  adding text to 131
  images as 133
  new journals 129
  printing 133, 135
  saving 131, 133, 135
  sticky notes 131
  viewing data tables in 129
JPG files, saving journals as 133
JRN (journal files) 132

K
Kendall’s tau statistic 425
Kolmogorov-Smirnoff-Lillifors (KSL) test 153
Kruskal-Wallis tests 270–273, 297
kurtosis 85, 148
  normality testing 155–156, 162

L
labeling values in data tables 56–57, 68
Lack of Fit report 351, 384–388
“large,” defined
  Central Limit Theorem 182
correlation coefficients 317
residuals 361–362
Lasso tool 94, 313–314
least significance difference
  See LSD Threshold Matrix report
least squares regression
  See also curve fitting
  See also multiple regression
  See also straight-line regression
checking for data errors and 324
normality testing 324
random samples and 323
Level value (Frequencies reports) 92
Levene test 269
leverage plots 397–400
likelihood ratio tests 422
line fitting
  See also regression
  See also regression diagnostics
correlation 316–320, 354, 425
Linear Fit report 326–328
  enhancing 331, 356
  means and confidence limits 332,
  334–335
linear relationships, finding
  See correlation coefficients
  See straight-line regression
Lock Data Table menu choice 41
locking data tables 41
locking scroll for columns 43, 66
lower 95% mean 83
LSD Threshold Matrix report 280, 281
  Bonferroni approach 285
  example of 299
  Hsu’s MCB test 289–290
  Tukey-Kramer HSD test 288

M
Make Text Bigger/Smaller options 80, 106
Mann-Whitney U test
  See Wilcoxon rank sum tests
  Matched Pairs platform
  order of selection 235
  ordinal variables 238
  paired differences tests 232–235
  Wilcoxon signed rank tests 237–238,
  242
maximum, quantiles reports 82
maximum value, checking 88
MCB 288–290
mean
  See also confidence intervals for mean
  See also estimating mean
  See also population mean
  See also sample averages
  upper 95% 83
Mean CI Lines option (Display Options) 227
means diamonds 81, 265
Mean Error Bars option (Display Options) 227
Mean Lines option (Display Options) 227
Mean of Means option (Display Options) 227
Mean of Response value 329
Mean Square value, one-way ANOVA 264
Means and Std Deviations report 256
measures of association 424–428, 431–432
median
  in box plots 81
  of normal distributions 148
Microsoft Excel spreadsheets, importing
  30–31, 64–65
Microsoft PowerPoint, copying JMP results
to 128–129, 135
Microsoft Word
  copying JMP results to 128–129, 135
  printing to tables in 58–59, 68
minimum, quantiles reports 82
minimum value, checking 88
missing values 88
  See also errors in data, checking for
  continuous variables 316
  correlation coefficient calculations 317
mode, normal distributions 148
Index

modeling types 50–52, 89
Moments reports 110
  checking for errors 88
  for continuous variables 83–86
  normality testing 155–156
mosaic plots 407
  colors in 407, 417–419, 431
  rotating to horizontal 410–411, 431
multiple comparison procedures 259, 274–293
  Bonferroni approach 281–285, 293
  comparison circles 275–277
  Dunnett’s test 290–292
  Each-Pair, Student’s t test 277–282
  Hsu’s MCB test 288–290
  Tukey-Kramer HSD test 281, 286–288, 293, 298
multiple regression 346–353, 356
  confidence limits 351–352, 357
  Lack of Fit report 386
  leverage plots 398–400
  residuals plots for 369–375
Multivariate platform 355
  Correlations report 316–320, 354
  measures of association 425–428, 431–432
  Multivariate Simple Statistics report 316, 355
  Pairwise Correlations report 317–319, 355
  scatterplots 311–315
  Univariate Simple Statistics report 315, 355

N

N Missing statistic 85
N statistic 83
names
  column headings 101–102
  data table columns 25
  report titles 101–102, 107
New Table Variable menu choice 40
Next Page option (Print Preview) 58
nominal variables 50–51
  bar charts 114, 136
  checking for errors 73, 93
  histograms 91–92
  one-way Kruskal-Wallis tests 272–273
  summarizing data with Distribution window 89–93
  summarizing in contingency tables 405–411, 430
nonparametric statistical methods 150, 258
  for comparing two groups 206
normal distributions 146–149
  Empirical Rule 149, 182–184
  median of 148
  mode 148
  of sample averages 180–182
  outliers and 160, 162
  population mean of 148
  smoothness of 148
  symmetry of 148
normal quantile plots 156–159, 162
normality testing 150–163, 171, 261, 296, 324, 388–389
  box plots 154–155, 162
  Distribution platform 151–153
  kurtosis 155–156, 162
  Moments reports 155–156
  null hypothesis 166
  p-value 166
  skewness and 155–156, 162
  statistical test 150–153
notes in reports 40–41, 66
  See also sticky notes
null hypotheses 164–165, 168
  See also specific test by name
  comparing multiple groups 256
  comparing two independent groups 205
  independence tests for contingency plots 420
  lack of fit tests 384
  normality testing 166
  p-value and 167
numeric variables 50–51
O'Brien test 269
observations
See also rows in data tables
independent 260
omitting
See hiding and showing
1-Alpha option 187–188
One Page option (Print Preview) 58
one-sided confidence intervals 188
one-sided hypothesis tests 219
one-way analysis of variance 258, 259–265
Oneway platform 261–273
enhancing graphs for 273–274
summarizing data 251–257, 300
online help 12, 17–19
Open All Below option 96
Open All Like This option 96
opening data tables 29, 64
opinion polls 143
Ordered Differences Report 280–281
Bonferroni approach 285
example of 299
Tukey-Kramer HSD test 288
ordering
values in data tables 53–55, 67–68
variables in Summary platform 436
variables in treemaps 122–123, 137
ordinal variables 50–51
bar charts 114, 136
checking for errors 73, 93
histograms 91–92
Matched Pairs platform 238
measures of association 424–428, 431–432
one-way Kruskal-Wallis tests 272–273
summarizing data with Distribution
window 89–93
summarizing in contingency tables 405–411, 430
outlier box plots
See box plots
outliers
excluding 161–162, 171
influential points (leverage plots) 398
normal distribution and 160, 162
residual plots and 361–362
outline item icons, defined 13
overall risk 281
overlap marks (mean diamonds) 265
Overlay Plot option 368
p-value 150, 164–165
See also significance levels
Chi-Square tests 422
correlation coefficient testing 318–320
measures of association 427–428
normality testing 166
null hypotheses and 167
practical significance vs. 168–169
statistical significance 166–169
p-value, comparing multiple groups 258–259
for equal variances (pooled) tests 268
one-way ANOVA 263
Kruskal-Wallis tests 272
p-value, comparing two groups 206–207
for equal variances (pooled) tests 211–212, 216
for unequal variances (unpooled) tests 214, 216
paired difference tests 231–232, 233–234
Wilcoxon rank sum tests 223
Wilcoxon signed rank tests 237, 238
page breaks, inserting in reports 128, 134
paired data in single column 246–248
paired-difference t-tests 206, 228–235, 242
Distribution platform for 230–235
Matched Pairs platform for 232–235
paired groups
independent groups vs. 195
statistical significance 206–207
statistical tests for comparing 206
summarizing data from 199–204
Pairwise Correlations report 317–319, 355
Index

Parameter Estimates report 153, 154
- multiple regression 350–351
- polynomial regression 343
- straight-line regression 328–329
parameters of populations 144–145
parametric statistical methods 150, 258
- for comparing two groups 206
Paste Special option 128, 135
pasting results to Microsoft applications 128–129
Pearson correlation coefficients 317, 425
Pearson tests 422
percentiles, plotting
  See box plots
  See Quantiles reports
percents in histogram counts 79
pictures, pasting results as 128, 135
pivot tables
  See contingency tables
platforms 13
  See also Distribution platform
- Bivariate 324–325, 370, 393
- Fit Model 347, 370, 372–374, 397–400
- Fit Y by X 272, 343, 347, 417
- Graph 364–369, 371–372
- Oneway 273–274, 251–257, 300
- Summary 434–436
- Tabulate 434
plot size, changing
  See resizing
plots
  See specific plot type by name
point estimates of population mean 174–175
- distribution of 179–183
Points Jittered option (Display Options) 226, 228, 243
Points option (Display Options) 227, 243
Points Spread option (Display Options) 228
polynomial regression 355
- centering polynomials 341, 344
- Lack of Fit report, understanding 387–388
- residuals plots for 380–383
polynomial regression (curve fitting) 336–345
- adding confidence curves 344–345
- centering polynomials 341, 344
pooled t-tests 212–217
population, defined 142–143
population mean 83, 145, 147
  See also confidence intervals for mean
  See also estimating mean
  See also standard deviation of population
  of normal distributions 148
- point estimates of 174–175
- standard error of 83, 182, 264
- straight-line regression 334–335
population parameters 144–145
population variability
  See standard deviation of population
PowerPoint 128–129, 135
practical significance 168–169
predicted values, confidence limits on
  See confidence limits on predicted values
predicted values, plotting residuals against 360–361
  See also regression diagnostics
- multiple regression 372–373
- polynomial regression 381–382
- straight-line regression 366–368, 377–378
  with Bivariate platform 393
  with Fit Model platform 394
predicted values, saving 332, 351, 356, 357
Prediction Expression report 350
Prev Page option (Print Preview) 58
previewing imported text files 34–36, 66
Print Preview option 58
printing
  data tables 57–60, 68
  journals 133, 135
  results 127–129
  to Microsoft Word tables 58–59, 68
prior significant F-test 275
Prob value (Frequencies reports) 92
Prob<W column (Goodness-of-Fit) 152, 153
pure error 384
   See also Lack of Fit report

Q
quadratic polynomials 336
   See also polynomial regression
Quantiles reports
   See also reports
      checking for errors 88
      for continuous variables 82–83
      normal quantile plots 156–159, 162
? tool (context-sensitive help) 19
quitting JMP 19–20

R
r values
   See correlation coefficients
random samples 144
   arithmetic average of 174–175, 179–183
   simple 144
   stratified random sampling 144
ratio variables 52
raw data, summarizing in tables 405–411, 430
Recall option 124, 134
records
   See rows in data tables
red line in box plots 81
red triangles, defined 13
regression
   See also multiple regression
   See also polynomial regression
   See also straight-line regression
   assumption for errors, testing 388–392, 395
   fitting straight lines with JMP 324–335
Lack of Fit report 351, 384–388
   regression diagnostics 360–395
      Lack of Fit report 384–388
      leverage plots 397–400
      plotting residuals, concepts of 360–363
      residuals plots for multiple regression 369–375
      residuals plots for polynomial regression 380–383
      residuals plots for straight-line regression 364–369, 375–379
      testing error assumptions 388–392, 395
releases of JMP 12
   “Replace table” check box (Sort window) 60
report comments
   See sticky notes
reports 13, 101–102, 107
   Analysis of Variance 330, 343, 351
   closing 97–98
   Confid Quantile 280
   Connecting Letters 279, 280
   Contingency Table 409–410
   copying to Microsoft applications 128–129, 135
   Correlations 316–320, 354
   Corresponding Letters 281
   decimal places in 86, 108
   Detailed Comparisons 281, 285, 288
   Difference Matrix 281, 285, 288, 290
   disclosure diamonds in 75
   Distribution window, continuous
      variables 73–88
   Distribution window, nominal and
      ordinal variables 89–93
   Frequencies 92–93
   Goodness-of-Fit Test 152
   hiding 95–96, 107
   hiding columns in 93, 108
   hot spots 76
   inserting page breaks 128, 134
   interacting with Distribution results 94–100
   Lack of Fit 351, 384–388
Linear Fit 326–328, 331–332, 334–335, 356
Means and Std Deviations 256
Moments 83–86, 88, 110, 155–156
Multivariate Simple Statistics 316, 355
notes in 40–41, 66
Ordered Differences 280–281, 285, 288, 299
ordering values in 53
Pairwise Correlations 317–319, 355
Parameter Estimates 153, 154, 328–329, 343, 350–351
Predicted Expression 350
printing 127–129
Quantiles 82–83, 88, 156–159, 162
removing 331
Summary of Fit 329, 343, 351
titles 101–102, 107
Univariate Simple Statistics 315, 355
residuals
defined 360
from Bivariate platform 370
from Fit Model platform 370
residuals plots
See also independent variables, plotting residuals against
See also regression diagnostics
See also time sequence, plotting residuals against
against predicted values 366–368, 372–373
basic concepts 360–363
multiple regression 369–375
outliers and 361–362
polynomial regression 380–383
straight-line regression 364–369, 375–379
resizing
axes 98–100, 119–120, 134
bar charts 119–120
columns in data tables 42
data tables 37, 38, 66
histograms 77, 80, 106, 109
treemaps 124, 137
resizing axes
See also independent variables, plotting residuals against
See also regression diagnostics
See also time sequence, plotting residuals against
against predicted values 366–368, 372–373
basic concepts 360–363
multiple regression 369–375
outliers and 361–362
polynomial regression 380–383
straight-line regression 364–369, 375–379
residuals
defined 360
from Bivariate platform 370
from Fit Model platform 370
residuals plots
See also independent variables, plotting residuals against
See also regression diagnostics
See also time sequence, plotting residuals against
against predicted values 366–368, 372–373
basic concepts 360–363
multiple regression 369–375
outliers and 361–362
polynomial regression 380–383
straight-line regression 364–369, 375–379
resizing
axes 98–100, 119–120, 134
bar charts 119–120
columns in data tables 42
data tables 37, 38, 66
histograms 77, 80, 106, 109
treemaps 124, 137
resizing axes
See also independent variables, plotting residuals against
See also regression diagnostics
See also time sequence, plotting residuals against
against predicted values 366–368, 372–373
basic concepts 360–363
multiple regression 369–375
outliers and 361–362
polynomial regression 380–383
straight-line regression 364–369, 375–379
reshaping
are treated as
in the same way
outside the scope of
the present text.

See also independent variables, plotting residuals against
See also regression diagnostics
See also time sequence, plotting residuals against
against predicted values 366–368, 372–373
basic concepts 360–363
multiple regression 369–375
outliers and 361–362
polynomial regression 380–383
straight-line regression 364–369, 375–379
reshaping
are treated as
in the same way
outside the scope of
the present text.

S
sample, defined 143
sample averages 145
distribution of 179–183
normal distributions 180–182
plotting 175–176
sample size
balanced vs. unbalanced data 258
effects on mean estimates 175–177, 189
large, defined 182
sample statistics 144–145
See also Moments reports;
See also specific statistic by name
scatterplots 315–316
sample variance 84, 145–146
of errors, testing assumption for regression 388–392, 395
point estimates and 175
samples, random 144
arithmetic average of 174–175, 179–183
least squares regression 323
Save All option (Close All Windows) 20
Save Individually option (Close All Windows) 20

Index 453
Save None option (Close All Windows) 20
Save Predicteds option 332, 351, 356, 357
Save Residuals option 364
Save Session window 20
saving
data tables 28, 381
journals 131, 133, 135
predicted values (regression) 332, 351, 356, 357
residuals (regression) 364, 369–370
scripts 271
sessions 20
scaling axes 98–100, 134
bar charts 119–120
uniform scaling 99–100, 199
scatterplot matrices 312–315, 354
hot spots 313
measures of association 426
scatterplots 309–315, 354
colors 313–314
density ellipses 313–314, 325
enhancing 331, 354
measures of association 425–428, 431–432
sample statistics 315–316
sticky notes 331
two-way 312
Score Sum, Score Mean values 223
scripts, saving 271
scroll lock for columns 43, 66
Select Columns section 90
Select Options window 225
selecting and deselecting
box plot features 82
columns in data tables 42–43, 90
comparison circles 291
histogram bars 76–77
multiple graphs 94–95
rows in data tables 48–49
sessions, saving 20
Set Colors option (mosaic plots) 417–419
Shapiro-Wilk test 153
shortest half (box plots) 81
Show Counts option (histograms) 78–79
Show Level Legend option 119
“Show tips at startup” checkbox 14
showing
See hiding and showing
sidedness
See skewness
Signed-Rank statistic
See Wilcoxon signed rank tests
Signif Prob column (Pairwise Correlations) 318
significance levels 166–167, 168–169
See also p-value
correlation coefficients 318–320
leverage plots 397–400
multiple regression parameters 350–351
simple random regression samples 144
sizing objects
See resizing
sizing text 80, 106
skewness 85
example of 87–88
normality testing 155–156, 162
smoothness of normal distributions 148
Sort window 59–60
sorting data tables 60–62, 69
for bar charts 114–115
Spearman’s rank correlation coefficient 425
Specify Columns button (Text Import Preview) 35
Specify Transformation or Constraint window 341
Split window 247
splitting columns of paired data 246–248
spreadsheets, importing 30–31, 64–65
square of standard deviation
See variance
standard deviation of mean, squared
See variance
standard deviation of population 83, 145, 147
estimating 188–189
mean estimates and 178–179, 189
normal distribution 149
standard error of the difference 215
standard error of the mean 83, 182
  one-way ANOVA 264
starting JMP 14
statistical methods 150, 206, 258
statistical significance 166–169, 206–207
statistical summary tables 434–436
statistical test for normality 150–153
Statistics Index 85
  statistics 144–145
Std Dev
  See standard deviation of population
Std Dev Lines option (Display Options) 227, 243, 271
Std Err Dif 215
Std Err Mean
  See standard error of the mean
StdErr Prob value (Frequencies report) 93
stem-and-leaf plots 110
  for continuous variables 86–88
  normality testing 159–160, 162
sticky notes 102–104, 107–108
  journals 131
  notes in reports 40–41, 66
  scatterplots 331
straight-line regression 320–324, 355
  confidence intervals 334–335
  confidence limits 332
  Lack of Fit report 384–385, 387
  population mean 334–335
  residuals plots for 364–369, 375–379
stratified random sampling 144
“Strip enclosing quotes” option 33
Student’s t test, Each-Pair 277–282
subsets of data, creating 161–162, 171
Sum of Squares statistic 264, 330
sum of weights 84
Sum statistic 84
Sum Wgt 84
summarizing data 71–110
  See also contingency tables
  See also errors in data, checking for
  continuous variables 73–88
multiple continuous variables 309–316
multiple groups 251–257, 300
nominal variables 89–93
raw data 405–411, 430
two independent groups 196–199, 241
two paired groups 199–204, 241
summary measures 144–145
  See also Moments reports
  See specific statistic by name
  scatterplots 315–316
Summary of Fit report
  multiple regression 351
  polynomial regression 343
  straight-line regression 329
Summary platform 434–436
summary tables
  See contingency tables
superscripts 343
symmetry of normal distributions 148

T

t distribution 185
t Ratio 216, 329
t-tests
  See also two-sample t-tests
  Each-Pair, Student’s t test 277–282
  graphs for 218
  one-sided hypothesis tests 219
  paired-difference 206, 228–235, 242
  pooled and unpooled t-tests 212–217
“Table contains column headers” option 33
table panel 39–41
tables
  See also contingency tables
  See also data tables
exporting to Microsoft Word 58–59, 68
  hot spots 39
  statistical summary tables 434–436
two-way 404–405
tables menu 39
Tabulate platform 434
Index

456

test statistic 150, 164–165
normality testing 166
testing errors assumption for regression 388–392, 395
testing for equal variances 295
   multiple groups 266–269
   two groups 208–212, 242
testing for normality
   See normality testing
testing hypotheses
   See hypotheses testing
tests for independence 419–424, 431
Tests report (Contingency platform) 421–424
text, adding to journals 131
text files, importing 32–36, 64–65
Text Import Preview window 34–36
text size in histograms 80, 106
three-way tables
   See contingency tables
time sequence, plotting residuals against 362–363
   multiple regression 373–374
   polynomial regression 382–383
   straight-line regression 368–369, 378–379
   with Bivariate platform 394
   with Fit Model platform 373–374, 394–395
Tip of the Day window 14
titles for reports 101–102, 107
Total Sum of Squares statistic 330
Total value (Frequencies reports) 92
treemaps 122–127, 137
triangle icons 13, 37
Try Delimited button 36
Try Fixed Width button 35–36
Tukey-Kramer HSD test 281, 286–288, 293, 298
2x2 tables
   See contingency tables
   “Two-digit year rule” option 33
two-group comparisons
   See comparing two groups
Two Page option (Print Preview) 58
two-sample t-tests 206, 207–220, 242
   alpha level for confidence intervals 219–220, 242
   comparing two means 212–219
   enhancing graphs for 243
   F Ratio value 211
   testing for equal variances 208–212, 242
two-way mosaic plots 407
   colors in 407, 417–419, 431
   rotating to horizontal 410–411, 431
two-way scatterplots 312
   See also scatterplots
two-way tables 404–405
   See also contingency tables
Type I error 166, 168
Type II error 167–168
U
unbalanced data 258
unequal variances
   See also testing for equal variances
   ANOVA with 266–270
   tests of (unpooled tests) 212–219, 242
   Welch Anova test 212, 269–270
unexcluding columns in data tables 46–47
unhiding
   See also hiding and showing
   columns in data tables 67
uniform scaling 99–100, 199
Univariate Simple Statistics report 315, 355
unpooled t-tests 212–217
upper 95% mean 83
Use Value Labels check box 56
V
value labels 56–57, 68
value orders, assigning 53–54, 67–68
values in data tables
   data types and modeling types 50–52
   ordering 53–55, 67–68
variables
  See also columns in data tables
  See also continuous variables
  See also independent variables, plotting residuals against
  See also nominal variables
  See also ordinal variables
  character  51
  interval  52
  numeric  50–51
  ordering in Summary platform  436
  ordering in treemaps  122–123, 137
  ratio  52
variance  84, 145–146
  of errors, testing assumption for regression  388–392, 395
  point estimates and  175
variance analysis
  See ANOVA
variance equality
  See equal variances (pooled) tests
  See unequal variances
variation due to error  257–258
viewing data tables  36–41

W
Welch Anova test  212, 269–270
whiskers in box plots  81
Wilcoxon rank sum tests  206, 220–224, 242
Wilcoxon signed rank tests  206, 235–238, 242
within-group variation  257–258
Word
  copying JMP results to  128–129, 135
  printing to tables in  58–59, 68

X
X Axis Proportional option (Display Options)  227, 243, 255
  mean diamonds and  265
Gain Greater Insight into Your JMP® Software with SAS Books.

Discover all that you need on your journey to knowledge and empowerment.

support.sas.com/bookstore
for additional books and resources.