MINITAB® MANUAL

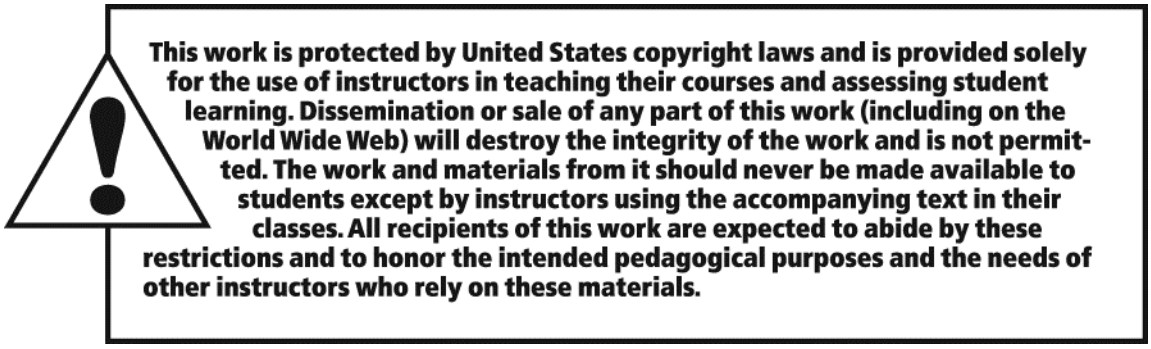
PRACTICING STATISTICS:
GUIDED INVESTIGATIONS FOR THE SECOND COURSE

SHONDA KUIPER
Grinnell College

Jeffrey Sklar
California Polytechnic State University

PEARSON

Boston Columbus Indianapolis New York San Francisco Upper Saddle River
Amsterdam Cape Town Dubai London Madrid Milan Munich Paris Montreal Toronto
Delhi Mexico City Sao Paulo Sydney Hong Kong Seoul Singapore Taipei Tokyo



Reproduced by Pearson from electronic files supplied by the author.

Publishing as Pearson, 75 Arlington Street, Boston, MA 02116.

ISBN-13: 978-0-321-78459-9
ISBN-10: 0-321-78459-6

www.pearsonhighered.com

PEARSON

Contents

Getting Started with Minitab

The following questions are designed to introduce you to Minitab software. Detailed instructions and appropriate data for each question are provided with the questions. Within each graph, be sure to include a title and proper labels. Copy and paste the graphs into Microsoft Word, and type in any necessary comments. Do not include any excess data or information when submitting your answers.

1. Create a Pie Chart and Bar Chart of the Education data. (Data are listed with the following examples).

2. Create a Histogram for Babe Ruth's home run data.

3. Create a Histogram for Babe Ruth's data using cut points of 20, 30, 40, 50, 60, and 70.

4. Create Side by Side Boxplots of all 3 players. (All 3 boxplots need to be on the same graph.)

- What is the overall pattern? Look for Shape Center and Spread. Is the data symmetric or skewed, is it unimodal?
- Are there any deviations from the overall pattern?
- Who is the best home run hitter? Explain.

5. Find the mean, variance, std. dev., and 5 number summaries for each player. (**Stat >Basic Statistics > Display Descriptive Stat**)

6. Create Stem and Leaf plots for Maris and McGwire (create 2 separate plots).

7. The Old Faithful geyser is one of the most popular tourist attractions at Yellowstone National Park. Naturally, tourists and park rangers would appreciate knowing how long they will need to wait for the next eruption to occur. Does there appear to be a linear relationship between the wait time (the time between geyser eruptions) and the eruption (the length of time of the actual eruption)? Create a scatterplot and regression line using eruption to predict the wait time.

Charts for Categorical Data

Pie Chart

1. Type the data shown into column C1. The first shaded row is used for labels {i.e. education}. Note: Anytime text is input into a column, C1 changes to C1-T (text data), which limits how that column can be used.

2. Type data into column C2 (type any label {i.e. count}in the very first shaded row.)

3. **Graph** **Pie Chart**

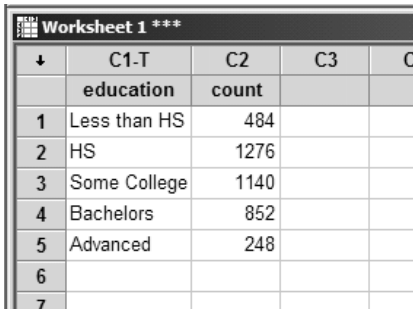
4. Click [Chart values from a table]

5. Click in space under "Categorical variable" to show list of columns

6. Double click C1 to move it to Categorical variable: and Double click C2 to move it to Summary variable.

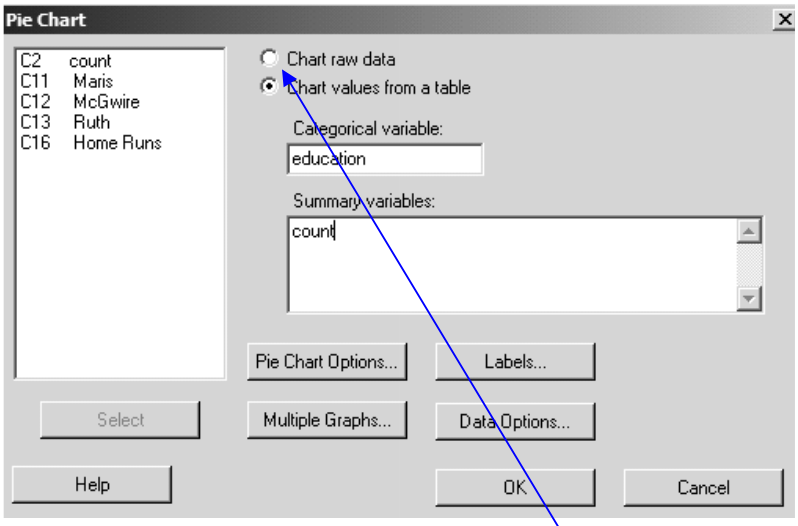
7. Click "Labels", then click "Slice Labels" and select the top three options.

8. Click [**OK**] [**OK**]



	C1-T	C2	C3	C
	education	count		
1	Less than HS	484		
2	HS	1276		
3	Some College	1140		
4	Bachelors	852		
5	Advanced	248		
6				

After the chart is made, you can right click on the graph and select "edit pie" to make changes to the graph.



If the data are not counted and simply listed in one column, for example: HS HS Bachelors HS Less than HS etc., use **Chart raw data**.

Bar Chart

1. Using the education and counts data in columns C1 and C2 (as described in the Pie Chart example above),

2. **Graph Bar Chart**

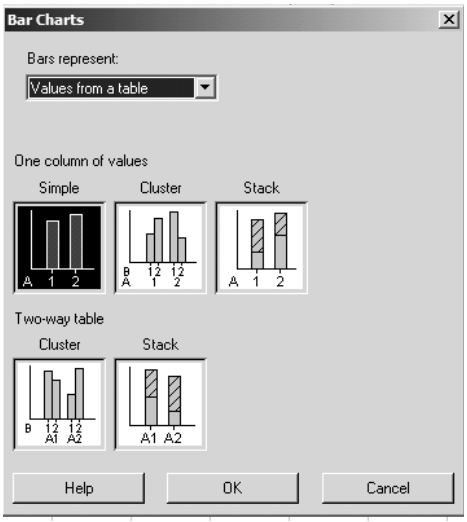
3. Bar represents "Values from a table" and click "Simple" {If data were not already summarized, you would select "counts of unique values"}

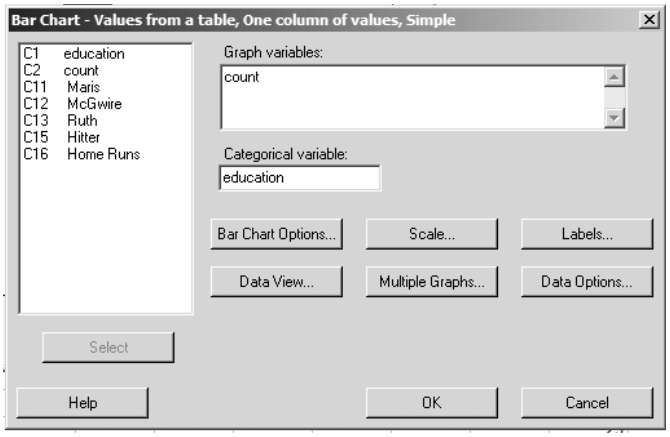
4. Click **OK**.

5. Click in space under "Catagorical variable" to show list of columns.

6. Double click C1 to move it to **Categorical variable:** box, then double click C2 to move it to **Graph variables:** box. Note: Since in this example C1 is text, it is not allowed as a Graph variable.

7. Click **OK** {You can also double click text in the chart to make editing changes}





Charts for Quantitative Data

Histogram (for Ruth) **1.** Enter the data into a column **2.** **Graph** **Histogram** **3.** Select "Simple" [**OK**] **4.** Select the column named Ruth [**OK**]	**Histogram with Specified Classes** **1.** Do steps 1-4 of the general histogram **2.** Double Click the bars on the histogram **3.** Select the binning tab **4.** List each Midpoint/Cutpoint position {separated by a space}
Boxplot (for Ruth) **1.** Enter the data into a column **2.** **Graph** **Boxplot** **3.** Select " One Y Simple" [**OK**] **4.** Select the column named Ruth [**OK**]	**Side-by-Side Boxplots** (Maris, McGwire, and Ruth) **1.** Enter the data into three columns **2.** **Graph** **Boxplot** **3.** Select "Multiple Y's Simple" **4.** Select the columns named Maris, McGwire, and Ruth [**OK**]
Time Plot (for Ruth) **1.** Enter the ordered data into a column **2.** Graph Time Series Plot "Simple" **3.** Double click C3 to move it to Y **4.** Select Time/Scale **5.** Enter 1920 into Index	**Stem-and-Leaf Plot** **1.** Enter the data into a column **2.** **Graph** **Stem-and-Leaf** **3.** Double click columns to move it to "Graph Variables" **4.** Click increment to select class size

Normal Quantile Plots

1. **Graph** **Probability Plot**
2. Double click appropriate Column to Variables
3. Distribution: Normal

Regression Plots/Equations

1. **Stat** **Regression** **Fitted Line Plot**
2. Double Click appropriate Columns to X (eruption) and Y (Wait Time)
3. Linear

Wait Time	eruption
78	4.4
74	3.9
68	4
76	4
80	3.5
84	4.1
50	2.3
93	4.7
55	1.7
76	4.9
58	1.7
74	4.6
75	3.4

Worksheet 1 ***

	C1	C2	C3
	Maris	McGwire	Ruth
1	8	9	22
2	13	9	25
3	14	22	34
4	16	32	35
5	23	33	41
6	26	39	41
7	28	39	46
8	33	42	46
9	39	49	46
10	61	52	47
11		58	49
12		65	54
13		70	54
14			59
15			60
16			

Chapter 1
Randomization Tests: Schistosomiasis

Activities

2. Copy and paste the data from the `Mice` data set into the first four columns of a new Minitab worksheet.

From the menu bar, select **Stat > Basic Statistics > Display Descriptive Statistics**. Select and transfer all four data columns into the **Variables:** box. Click **OK**.

7. First, stack the female mice data. In Minitab, click anywhere on the session window. From the menu bar, select **Editor > Enable commands**. Note that when the editor command is on, the programming code will show in the session window.

From the menu bar, select **Data > Stack > Columns**. Select and transfer C1 and C2 into the **Stack the following columns:** box. Click **Column of Current worksheet:** and enter C5 into the box. In the **Store subscripts in:** box, enter C6. Check the **Use variable names in subscript column** box. Click **OK**.

Next, randomly allocate mice worm counts to a group. From the menu bar, select **Calc > Random Data > Sample from columns**. In the dialog box, **sample** 10 rows from C5 and **store samples** in C7. Leave the **Sample with replacement box** unchecked. Click **OK**.

8. From the menu bar, select **Data > Unstack Columns**. **Unstack the data in**: C7 **using subscripts in:** C6. Select the **After last column in use** option, then click **OK**. This should place the unstacked data into C8 and C9.

In the Session window, enter the command `MTB > let C10 = mean(C8)-mean(C9)`, then hit the **Enter** key. C10 now contains the difference between the mean of two randomly allocated groups.

9. Open the program Notepad (**Programs > Accessories > Notepad**) and enter the following six lines:

```
sample 10 C5 C7
unstack C7 C8 C9;
subs C6;
varnames.
let C10(k1) = mean(C8)-mean(C9)
let k1=k1+1
```

Click **File > Save as.** Name the file schistosome.mtb, and select **All File Types** from the dropdown menu before clicking **Save**.

In the Minitab Session window, enter the command `MTB> let k1=1`, then hit **Enter**. Note: this command must be entered before you run any exec file.

From the menu bar, select **File > Other Files > Run an exec**. In the **Number of times to execute:** box, enter 1000, then click on the **Select File** button. Locate the schistosome.mtb file on your computer and click **Open**. C10 now represents 1000 simulated differences between two randomly selected groups.

Next, count the number of simulations that result in a mean difference greater than or equal to 7.6, divide that count by 1,000 and report the resulting empirical p-value. In the Session window, enter the command `MTB > let C11 = (C10 >= 7.6)`. Hit **Enter**, enter the command `MTB > sum C11`, then hit **Enter** again.

10. From the menu bar, select **Graph > Histogram > OK**. Select and transfer C10 into the **Graph variables:** box, then click **OK**. To change the bin widths, double click the horizontal axis of the histogram, select the **Binning** tab, click the **Number of intervals** option and enter 10.

13. In the Session window, enter the command `MTB > let C11 = (C10 <= -7.6 or C10 >= 7.6).` Hit **Enter**, then enter the command `MTB > sum C11`.

Alternatively, select **Stat > Tables >Tally Individual Variables** from the menu bar to count the frequency of each outcome.

Extended Activities

17. Copy and paste the data from the `Age` data set into the Minitab worksheet. Use C5 to list the age column and C6 for whether the person was laid off or not. C1-C4 will not be needed for these data.

Find the schistosome.mtb file on your computer and rename it age.mtb.

In the Session window, enter the command `MTB > let k1=1.`

From the menu bar, select **File > Other Files > Run an exec**. In the **Number of times to execute** box, enter 1000, then click on the **Select File** button. Locate the age.mtb file on your computer and click **Open**.

In this study the alternative hypothesis is Ha: mean (yes) – mean (no) > 0, so you may choose to change the command from `MTB > let C10(k1) = mean(C8)-mean(C9)` to `MTB > let C10(k1) = mean(C9)-mean(C8)`, since C9 represents the "yes" data and C8 represents the "no" data.

In the Session window, enter the command `MTB > let C11 = (C10 >= 16.57)`. Hit **Enter**, then enter the command `MTB > sum C11`.

18. In the Session window, modify the appropriate line in the code to

`MTB > let C10(k1) = median(C8)-median(C9)` and hit **Enter**.

20. Copy and paste the Fastdiff-Slowdiff column from the `Music` data set into C1 in the Minitab worksheet. In C2, alternate "1" and "-1" (to indicate the two possible orderings for listening) for a total of 14 each.

The following macro, when entered into the Session window, will randomly assign each observed difference to an order (Fastdiff-Slowdiff or Slowdiff-Fastdiff) by multiplying by 1 or -1. "1" will represent Fastdiff-Slowdiff and "-1" will represent Slowdiff-Fastdiff:

```
MTB > sample 28 C1 C5
let C6 = C2*C5
let C10(k1) = sum(C6)/28
let k1=k1+1
```

21. Copy and paste the data from the `ChiSq` data set into the first Minitab column, C1.

Open the program Notepad (**Programs > Accessories > Notepad**) and enter the following short macro:

```
sample 40 C1 C2
let C3(k1) = mean(C2)
let k1=k1+1
```

Click **File > Save as.** Name the file chisquare.mtb, and select **All File Types** from the dropdown menu before clicking **Save**.

In the Session window, enter the command `MTB > let k1=1` and hit **Enter**.

From the menu bar, select **File > Other Files > Run an exec**. In the **Number of times to execute** box, enter 1000, then click on the **Select File** button. Locate the chisquare.mtb file on your computer and click **Open**.

From the menu bar, select **Graph > Histogram > OK**. Select and transfer C3 into the **Graph variables:** box, then click **OK**.

22. Copy and paste the data from the `ChiSq` data set into the first Minitab column, C1.

In the Session window, enter the command `MTB > Sample 40 C1 C2`.

Open the program Notepad (**Programs > Accessories > Notepad**) and enter the following short macro:

```
Sample 40 C2 C3;
Replace.
let C4(k1) = mean(C3)
let k1=k1+1
```

Click **File > Save as.** Name the file chisquare.mtb, and select **All File Types** from the dropdown menu before clicking **Save**.

In the Session window, enter the command `MTB > let k1=1` and hit **Enter**.

From the menu bar, select **File > Other Files > Run an exec**. In the **Number of times to execute** box, enter 1000, then click on the **Select File** button. Locate the chisquare.mtb file on your computer and click **Open**.

From the menu bar, select **Graph > Histogram > OK**. Select and transfer C4 into the **Graph variables:** box, then click **OK**.

23. Open the chisquare.mtb file in Notepad and change the third line to `let C3(k1) = stdev(C2)`

24. Copy and paste the `MedSalaries` data into the first Minitab column, C1.

Open the program Notepad (**Programs > Accessories > Notepad**) and enter the following short macro:

```
Sample 100 C1 C2;
Replace.
let C3(k1) = mean(C2) {Note: Use stdev(C2) for Activity (24B)}
let k1=k1+1
```

Click **File > Save as.** Name the file medsalaries.mtb, and select **All File Types** from the dropdown menu before clicking **Save**.

In the Session window, enter the command `MTB > let k1=1` and hit **Enter**.

From the menu bar, select **File > Other Files > Run an exec**. In the **Number of times to execute** box, enter 1000, then click on the **Select File** button. Locate the medsalaries.mtb file on your computer and click **Open**. This will generate the bootstrap percentile confidence interval

25. Copy and paste the 'Pitcher' and 'First Baseman' data from the `NLBB Salaries` data set into two columns in Minitab.

From the menu bar, select **Stat > Nonparametrics > Mann Whitney**. In the **First Sample** box, select the Pitcher column; in the **Second Sample** box, select the First Baseman column. Click **OK**.

27. Copy and paste the 'Pitcher' and 'First Baseman' data from the `NLBB Salaries` data set into two columns in Minitab.

From the menu bar, select **Stat > Basic Statistics > 2-sample t**. In the **First** box, select the Pitcher column; in the **Second** box, select the First Baseman column. Click **OK**.

From the menu bar, select **Graph > Individual Value Plot > Multiple Y's Simple > OK**. In the **Graph variables:** box, select and transfer the Pitcher and First Baseman columns. Click **OK**.

28. Copy and paste the 'Salary' and 'Position' columns from the `NLBB Salaries` data into two columns in Minitab.

From the menu bar, select **Stat > Nonparametrics > Kruskal-Wallis**. In the **Response** box, select the Salary column; in the **Factor** box, select the Position column. Click **OK**.

Chapter 2
The Two-Sample t-test, Regression, and ANOVA: Making Connections

Activities

4. Copy and paste the data from the `Games1` data set into a new Minitab worksheet.

From the menu bar, select **Graph > Boxplots > One Y > With Groups > OK**. Select and transfer the Time column into the **Graph variables:** box, then select and transfer the Type column into the **Categorical Variables:** box. Click **OK**.

From the menu bar, select **Stat > Basic Statistics > Display Descriptive Statistics**. Select and transfer the Time column into the **Variables:** box, then select and transfer the Type column into the **By variables:** box. Click the **Statistics** button and verify that mean and standard deviation are selected. Click **OK**, and then click **OK** again.

7. From the menu bar, select **Data > Code > Text to Numeric**. Select and transfer the Type column into the **Code data from:** box, then select and transfer the Means column into the **Into Columns:** box.

Enter the following information into the specified columns:

Original values:	**New:**
Color	38.1
Standard	35.55

Click **OK**.

From the menu bar, select **Calc > Calculator**. Enter 'Resid' into the **Store result in variable:** box, and then enter 'Time'-'Means' into the **Expression:** box. Click **OK**.

From the menu bar, select **Graph > Histogram > Simple**. Select and transfer the Resid column into the **Graph Variables:** box and click **OK**.

9. From the menu bar, select **Graph > Scatterplot > With Connect Line > OK**. Select and transfer the Resid column into the **Y variables:** box, then select and transfer the StudentID column into the **X variables:** box. Click **OK**.

10. From the menu bar, select **Stat > Basic Statistics > 2-Sample t**. Click the **Samples in one column** option, then select and transfer the Time column into the **Samples** box, and select and transfer the Type column into the **Subscripts** box. Check the **Assume equal variances** box and click **OK**.

11. From the menu bar, select **Calc > Make Indicator Variables**. Select and transfer the Type column into the

Indicator variables for: box. In the **Store results in:** box, enter C6-C7.

Note that regression can only be conducted when the explanatory variable is quantitative. So Type is treated as a quantitative variable by coding Color = 1 and Standard = 0.

From the menu bar, select **Stat > Regression > Fitted Line Plot**. Select and transfer the Time column into the **Response (Y):** box, and then select and transfer Type_Color into the **Predictor (X):** box. Click **OK**.

12. *Note: Minitab 15 requires a macro to create a confidence interval. The following steps work for Minitab 16.*

From the menu bar, select **Stat > Regression > General Regression**. Enter the desired values into the dialog box, and click the **Results** button. Under **Coefficient table**, check **Display confidence intervals** and click **OK**.

13. Refer to the instructions for Activities 11 and 12.

14. From the menu bar, select **Stat > Regression > Regression**. Select and transfer the Time column into the **Response (Y):** box, and then select and transfer Type_Color into the **Predictor (X):** box. Click the **Graphs** button and check the boxes for **Histogram of residuals** and **Residuals versus order.** Click **OK**, and then click **OK** again.

15. From the menu bar, select **Stat > Regression > Fitted Line Plot**. Select and transfer the Time column into the **Response (Y):** box, and then select and transfer Type_Color into the **Predictor (X):** box. Click **OK**.

19. From the menu bar, select **Stat > Basic Statistics > Display Descriptive Statistics**. Select and transfer the Time column into the **Variables:** box, then select and transfer the Type_Color into the **By variables:** box. Click **OK**.

21. From the menu bar, select **Stat > ANOVA > Main Effects Plot**. Select and transfer the Time column into the **Response:** box, and then select and transfer the Type column into the **Factors** box. Click **OK**.

23. & 26. From the menu bar, select **Stat > ANOVA > General Linear Models**. Select and transfer the Time column into the **Response:** box, and then select and transfer the Type column into the **Model** box. Click the **Graphs** button and select the **Four in one** option. Click **OK**, and then click **OK** again.

Extended Activities

28. To sort the data from smallest to largest, enter the five data points into C1. From the menu bar, select **Data > Sort**. Select and transfer C1 into the **Sort Columns:** box and click OK.

Enter the numbers 1, 2, 3, 4, 5 into C2. From the menu bar, select **Calc > Calculator**. Select and transfer C3 into the **Store result in variable:** box, and then enter '(C2-.5)/5' into the **Expression:** box. Click **OK**.

From the menu bar, select **Calc > Probability Distributions > Normal**. Select the **Inverse Cumulative probability** option. Be sure that the value in the **Mean** box is 0, and the value in the **Standard deviation** box is 1. Select and transfer C3 into the **Input column:** box, and then select and transfer C4 into the **Optional storage:** box. Click **OK**.

29. Copy and paste the data from the `Normal` data set into the Minitab worksheet. From the menu bar, select **Stat > Descriptive Statistics > Normality Test**. Select and transfer C1 (or any appropriate column) into the **Variable:** box. Click **OK**.

30. a) From the menu bar, select **Stat > Descriptive Statistics > Normality Test**. Select and transfer the Resid column into the **Variable:** box. Click **OK**.

b-c) Create random samples of size 40 from N(0,1) distribution by going to menu bar and selecting **Calc > Random Data > Normal Distribution**. Enter '40' into the **Number of rows to generate:** box, and select and transfer C1-C9 (or any appropriate columns) into the **Store in columns:** box. Click **OK**.

31. a) Copy and paste the data from the `Emission` data set into the Minitab worksheet. From the menu bar, select **Graph > Individual Value Plots > With Groups > OK**. Select and transfer the Emission column into the **Graph Variables:** box, and select and transfer the Year column into the **Categorical variables:** box. Click **OK**.

To show graphs with properly sorted data, right click the Year column in the Minitab worksheet and select **Column > Value Order**. Select the **User-specified order** option. Be sure the years are in proper order in the **Define an order** box, and then click **OK**.

b) From the menu bar, select **Calc > Calculator**. Select and transfer LnEm into the **Store result in variable:** box, then select and transfer LN('Emission') into the **Expression: box**. Click **OK**.

c) From the menu bar, click **Stat > ANOVA > General Linear Models**. Select and transfer LnEm into the **Responses:** box, then select and transfer the Year column into the **Model:** box. Click the **Graphs** button and select the **Four in one** option. Click **OK**, then click **OK** again.

32-33. Follow the steps outlined in the previous activities.

34. Enter 2.2862 into C1.

To find P(t < 2.2862), go to the menu bar and select **Calc > Probability Distributions > t**. Select the **Cumulative Probability** option and enter 0 into the **Noncentrality parameter** box, then enter 38 into the **Degrees of freedom** box. Select and transfer C1 into the **Input column:** box and then select and transfer C2 into the **Optional storage:** box. Click **OK**.

The p-value is 2(1- P(t < 2.2862)).

From the menu bar, select **Calc > Calculator**. Select and transfer C3 to the **Store results in variable:** box, then enter 2*(1-C2) into the **Expression:** box. Click **OK**.

35. Follow the steps outlined in Activity 34.

Chapter 3
Multiple Regression: How Much is Your Car Worth?

Activities

1. Copy and paste the data from the Cars data set into a new Minitab worksheet.

From the menu bar, select **Graph > Scatterplot > With Regression > OK**. Select and transfer the Price column into the **Y variables** box, then select and transfer the Mileage column into the **X variables** box. Click **OK**.

2. From the menu bar, select **Stat > Regression > Regression**. Select and transfer the Price column into the **Response:** box, then select and transfer the Mileage column into the **Predictor:** box. Click **OK**.

3. To calculate all residuals, go to the menu bar and select **Stat > Regression > Regression**. Select and transfer the Price column into the **Response:** box, then select and transfer the Mileage column into the **Predictor:** box. Click the **Storage** button and check the **Residuals** box. Click **OK**, and then click **OK** again.

4. **a)** From the menu bar, select **Stat > Regression > Regression**. Select and transfer the Price column into the **Response:** box. One at a time, input each of the seven explanatory variables into the **Predictor:** box. Click **OK**.

b) From the menu bar, select **Stat > Regression > Regression**. Select and transfer the Price column into the **Response:** box. For each of the six remaining variables, input C1 and one of the other six explanatory variables into the **Predictor:** box. Click **OK**.

c) From the menu bar, select **Stat > Regression > Stepwise**. Select and transfer the Price column into the **Response:** box, then select and transfer all available explanatory variable columns (including Leather, Cruise, and Sound) into the **Predictor:** box. Click **OK**.

5. From the menu bar, select **Stat > Regression > Best Subsets**. Select and transfer the Price column into the **Response:** box, then select and transfer all available quantitative explanatory variables into the **Free Predictors:** box. Click **OK**.

7. From the menu bar, select **Stat > Regression > Regression**. Select and transfer the Price column into the **Response:** box, then select and transfer the appropriate values into the **Predictor:** box. Click the **Graphs** button and check the **Residuals versus fits** box, then select and transfer all explanatory variables in your model into the **Residuals versus the variables: box.** Click **OK**, and then click **OK** again. Note that "Fits" is another term for predicted or estimated retail price.

8. From the menu bar, select **Calc > Calculator**. Use the function menu to select Log ten of the Price column, LOGT(Price). Click **OK**.

Repeat, this time using the function menu to select the Square root of the Price column, SQRT(Price). Click **OK**.

9-10. From the menu bar, select **Stat > Regression > Regression**. Input the appropriate response and explanatory variables into the **Response:** and **Predictor:** boxes, respectively. Click the **Graphs** button and check the **Residuals versus order** box. Click **OK**, and then click **OK** again.

11. From the menu bar, select **Stat > Regression > Regression**. Input the appropriate response and explanatory variables into the **Response:** and **Predictor:** boxes, respectively. Click the **Graphs** button and check the **Normal Plot of Residuals** box. Click **OK**, and then click **OK** again.

From the menu bar, select **Editor > Brush**, then move the cursor around and click on any outliers to identify them.

12. Follow the steps outlined for the previous activities.

13. From the menu bar, select **Stat > Regression > Regression**. Input the appropriate response and explanatory variables into the **Response:** and **Predictor:** boxes, respectively. Click the **Graphs** button and select the **Four in one** options. Click **OK**, and then click **OK** again.

14. From the menu bar, select **Stat > Regression > Regression**. Input the appropriate response and explanatory variables into the **Response:** and **Predictor:** boxes, respectively. Click **OK**.

16. From the menu bar, select **Graph > Scatterplot > Simple**. Select and transfer the Cyl column into the **Y variable** box, then select and transfer the Liter column into the **X variable** box. Click **OK**.

17. From the menu bar, select **Graph > Individual Value Plot > One Y > With Groups > OK**. Select and transfer the TPrice column into the **Graph variables:** box, then select and transfer the Make column into the **Categorical variables for grouping:** box. Click **OK**.

Repeat the process, transferring the Model, Trim, and Type columns in the **Categorical variables for grouping:** box.

18. From the menu bar, select **Calc > Make Indicator Variables**. Select and transfer the Make column into the **Indicator variables for:** box, and then enter any six unused columns into the **Store results in:** box. Name the columns, in order: Buick, Cadillac, Chevrolet, Pontiac, SAAB, Saturn.

19. From the menu bar, select **Stat > Regression > Regression**. Input the appropriate response and explanatory variables into the **Response:** and **Predictor:** boxes, respectively. Click **OK**.

20. From the menu bar, select **Calc > Make Indicator Variables**. Select and transfer the Type column into the **Indicator variables for:** box, and then enter any five unused columns into the **Store results in:** box. Name the columns, in order: Buick, Convertible, Coupe, Hatchback, Sedan, Wagon. Click **OK**.

From the menu bar, select **Stat > Regression > Regression**. Input the appropriate response and explanatory variables into the **Response:** and **Predictor:** boxes, respectively. Click **OK**.

21-22. Follow the steps outlined for the previous activities.

Extended Activities

23. Copy and paste the data from the `Cavalier` data set into a new Minitab worksheet.

To create a vector of values of y_hat values, go to the menu bar and select **Calc > Calculator**. Enter 'yhat' into the **Store result in variable:** box, then enter 15244 - 0.111*Mileage into the **Expression:** box. Click **OK**.

To create a vector of values of y_bar values, enter 12962 into the first row of a new column labeled ybar; copy and paste it into the 30 rows below.

From the menu bar, select **Calc > Calculator** and enter 'ybar-yhat)*(yhat-ybar)' into the **Expression:** box and store the results in a new column. From the menu bar, select **Calc > Column Statistics > OK** to sum the column.

30-35. Follow the steps outlined for the previous activities.

36. Copy and paste the data from the `4-8Cyl` data set into a new Minitab worksheet.

Create indicator variables for the Make column by going to the menu bar and selecting **Calc > Make Indicator Variables**. Select and transfer the Make column into the **Indicator variables for:** box and store the results in new columns C13-C15. Click **OK**.

From the menu bar, select **Stat > Regression > Regression**. Select and transfer the Price column into the **Response:** box, then select and transfer the Mileage, Cadillac, and SAAB columns into the **Predictor:** box. Click **OK**.

Create a column of yhat values. From the menu bar, select **Calc > Calculator**. Enter yhat into the **Response:** box, then select and transfer the Mileage column into the **Predictor:** box. Click **OK**.

Create a scatterplot that includes X(Mileage) versus Y(Price) and X (Mileage) and Y(yhat). From the menu bar, select **Graph > Scatterplot > Simple > OK**. Enter the following information into the specified columns:

Y variables:	**X variables:**
Price	Mileage
Yhat	Mileage

Click the **Multiple Graphs** button and select the **Overlaid on the same graph** option. Click **OK**, then click **OK** again.

37-43. Follow the steps outlined for the previous activities.

Chapter 4
Designing Factorial Experiments: Microwave Popcorn

Activities

4. Copy and paste the data from the `Popcorn` data set into a new Minitab worksheet.

From the menu bar, select **Stat > Basic Statistics > Display Descriptive Statistics**. Select and transfer the PopRate column into the **Variables:** box, then select and transfer the Brand column into the **By variables:** box. Click **OK.**

From the menu bar, select **Stat > Basic Statistics > Display Descriptive Statistics**. Select and transfer the PopRate column into the **Variables:** box, then select and transfer the Time column into the **By variables:** box. Click **OK.**

From the menu bar, select **Stat > Basic Statistics > Display Descriptive Statistics**. Select and transfer the PopRate column into the **Variables:** box. Click **OK.**

5. From the menu bar, select **Stat > Basic Statistics > Display Descriptive Statistics**. Select and transfer the PopRate column into the **Variables:** box, then select and transfer the Brand and Time columns into the **By variables:** box. Click **OK.**

13. From the menu bar, select **Stat > ANOVA > General Linear Models**. Select and transfer the PopRate column into the **Responses:** box, then select and transfer the Brand and Time columns (Brand*Time) into the **Model:** box. Click **OK.**

14. From the menu bar, select **Stat > ANOVA > General Linear Models**. Select and transfer the PopRate column into the **Responses:** box, then select and transfer the Brand and Time columns (Brand*Time) into the **Model:** box. Click the **Graphs** button and select the **Normal plot of residuals** box. Click **OK**, then click **OK** again**.**

16. From the menu bar, select **Graph > Individual Value Plot > With Groups > OK**. Select and transfer the PopRate column into the **Graph variables:** box, then select and transfer the Brand, Time, and Microwave columns into the **Categorical variables:** box. Click **OK**.

17. From the menu bar, select **Stat > Basic Statistics > Display Descriptive Statistics**. Select and transfer the PopRate column into the **Variables:** box, then select and transfer the Brand, Time, and Microwave columns into the **By variables:** box. Click the **Statistics** button and verify that the **Standard deviation** box is checked. Click **OK**, and then click **OK** again**.**

18. From the menu bar, select **Stat > ANOVA > General Linear Models**. Select and transfer the PopRate column into the **Responses:** box, then select and transfer the Brand, Time , and Microwave columns (Brand*Time, Microwave*Brand, Microwave*Time) into the **Model:** box. Click the **Graphs** button and select the **Normal plot of residuals** box. Click **OK**, and then click **OK** again**.**

Extended Activities

25. Copy and paste the data from the `PaperTowels` data set into a new Minitab worksheet.

From the menu bar, select **Stat > ANOVA > Main Effects Plot**. Select and transfer the Strength column into the **Responses:** box, then select and transfer the Brand and Water columns into the **Factors:** box. Click **OK**.

28. From the menu bar, select **Stat > ANOVA > Interaction Plot**. Select and transfer the Strength column into the **Responses:** box, then select and transfer the Brand and Water columns into the **Factors:** box. Click **OK**.

34. From the menu bar, select **Stat > Basic Statistics > Display Descriptive Statistics**. Select and transfer the Strength column into the **Variables:** box, then select and transfer the Brand and Water columns into the **By variables:** box. Click the **Statistics** button and verify that the **Standard deviation** box is checked. Click **OK**, then click **OK** again**.**

From the menu bar, select **Calc > Calculator**. In the **Store result in variable:** box enter LnStrength. In the **Expression:** box, use the function menu to select Natural Log of the Strength column, LOGE(Strength), or select the Square Root of the Strength column, SQRT(Strength). Click **OK**.

35. From the menu bar, select **Graph > Boxplots > With Groups > OK**. Select and transfer the SQRTStrength column into the **Graph variables:** box, then select and transfer the Brand and Water columns into the **Categorical variables:** box. Click **OK**.

36. From the menu bar, select **Stat > ANOVA > General Linear Models**. Select and transfer the SQRTStrength column into the **Responses:** box, then select and transfer the Brand and Water columns into the **Model:** box. Click the **Graphs** button and select the **Four in one** option. Click **OK**, then click **OK** again**.**

Chapter 5
Block, Split-Plot and Repeated Measure Designs: What Influences Memory?

Activities

1. Copy and paste the data from the `Memory` data set into a new Minitab worksheet.

b) From the menu bar, select **Stat > ANOVA > General Linear Models**. Select and transfer the Score column into the **Responses:** box, then select and transfer the columns Worldlist, Distracter, and Wordlist*Distracter into the **Model:** box. Enter the interaction term as shown. Minitab uses the * to identify interactions between terms. Click **OK**.

c) From the menu bar, select **Stat > ANOVA > General Linear Models**. Select and transfer the Score column into the **Responses:** box, then select and transfer the columns Worldlist, Distracter, and Wordlist*Distracter into the **Model:** box. Click the **Graphs** button and check the **Normal plot of residuals** box. Click **OK**, then click **OK** again.

From the menu bar, select **Graph > Boxplots > One Y With Groups > OK**. Select and transfer the Score column into the **Graph variables:** box, then select and transfer the Wordlist and Distracter columns into the **Categorical variables:** box. Click **OK**.

d) From the menu bar, select **Stat > ANOVA > General Linear Models**. Select and transfer the Score column into the **Responses:** box, then select and transfer the columns Worldlist, Distracter, and Wordlist*Distracter into the **Model:** box.

Click the **Factor Plots** button. Check the **Main Effect Plot** box, then select and transfer the Worldist and Distracter columns into the **Factors:** box. Check the **Interaction Plot** box, then select and transfer the Worldist and Distracter columns into the **Factors:** box. Click **OK**, and then click **OK** again.

2. From the menu bar, select **Stat > ANOVA > General Linear Models**. Select and transfer the Score column into the **Responses:** box, then select and transfer the columns Student, Worldlist, Distracter, and Wordlist*Distracter into the **Model:** box. Select and transfer the Student column into the **Random factors:** box.

Click the **Factor Plots** button. Check the **Main Effect Plot** box, then select and transfer the Worldist, Student, and Distracter columns into the **Factors:** box. Check the **Interaction Plot** box, then select and transfer the Worldist and Distracter columns into the **Factors:** box. Click **OK**.

Click the **Graphs** button and check the **Normal plot of residuals** box. Click **OK**, and then click **OK** again.

To place all main effect plots on the same Y-axis, right click the Main Effects Plot output, then click the **Panel** option. On the **Arrangement** tab, click **Custom**. Enter 1 into the **Rows:** box, and then enter 3 into the **Columns:** box. Click **OK**.

3-7. Minitab is not required to complete these activities.

8. From the menu bar, select **Stat > ANOVA > General Linear Models**. Select and transfer the Score column into the **Responses:** box, then enter the following: Major, Student(Major), Wordlist, Distracter, and Wordlist*Distracter, into the **Model:** box. Select and transfer the Student column into the **Random factors:** box.

Click the **Factor Plots** button. Check the **Main Effect Plot** box, then select and transfer the Worldist, Major, and Distracter columns into the **Factors:** box. Click **OK**.

Click the **Graphs** button and check the **Normal plot of residuals** box. Click **OK**, and then click **OK** again.

9-10. Minitab is not required to complete these activities.

11. From the menu bar, select **Stat > ANOVA > General Linear Models**. Select and transfer the Score column into the **Responses:** box, then select and transfer the columns Major, Student2 ,Wordlist, Distracter, and Wordlist*Distracter into the **Model:** box. (Remove "Student" from the "Random Factors" box).Click **OK**.

12. From the menu bar, select **Stat > ANOVA > General Linear Models**. Select and transfer the Score column into the **Responses:** box, then select and transfer the columns Major, Student2(Major), Wordlist, Distracter, and Wordlist*Distracter into the **Model:** box. Click **OK**.

13-14. Minitab is not required to complete these activities.

15. From the menu bar, select **Stat > ANOVA > General Linear Models**. Select and transfer the Score column into the **Responses:** box, then select and transfer the columns Major, Student(Major), Wordlist, Distracter, Major*Wordlist , Major*Distracter and Wordlist*Distracter into the **Model:** box. Click **OK**.

16-18. Minitab is not required to complete these activities.

19. From the menu bar, select **Stat > ANOVA > General Linear Models**. Select and transfer the Score column into the **Responses:** box, then select and transfer the columns Student(Major), Wordlist, Distracter, Major*Wordlist, Major*Distracter and Wordlist*Distracter into the **Model:** box. Select and transfer the Student column into the **Random factors:** box. Click the **Graphs** button and select the **Four in one** option. Click **OK**, and then click **OK** again.

20. From the menu bar, select **Graph >Boxplots > With Groups >OK**. Select and transfer the Score column into the **Graph variables:** box, then select and transfer the Major, Wordlist, and Distracter columns into the **Categorical variables:** box. Click **OK**.

21. To calculate each student average, go to the menu bar and select **Stat > Basic Statistics > Store Descriptive Statistics**. Select and transfer the Score column into the **Variables:** box, then select and transfer the Major column into the **By Variables:** box. Click the **Statistics** button and verify that the mean is selected. Click **OK**, and then click **OK** again.

From the menu bar, select **Graph > Individual Value Plots > With Groups > OK**. Enter StudentAvg (called 'Mean' in Minitab) into the **Graph variables:** box, then select and transfer the Major column (often called 'ByVar1' in Minitab) into the **Categorical variables:** box. Click **OK**.

Extended Activities

25. Copy and paste the data from the `Flower` data set into a new Minitab worksheet.

From the menu bar, select **Stat > ANOVA > General Linear Models**. Select and transfer the Days column into the **Responses:** box, then select and transfer the Store and Water columns into the **Model:** box. Click the **Graphs** button and select the **Four in one** option. Click **OK**.

31. Copy and paste the data from the `Popcorn` data set into a new Minitab worksheet.

From the menu bar, select **Stat > ANOVA > General Linear Models**. Select and transfer the % Popped column into the **Responses:** box, then select and transfer the columns Brand, Box(Brand), Temp, and Brand*Temp into the **Model:** box. Click the **Graphs** button and select the **Four in one** option. Click **OK**.

42. Copy and paste the data from the `Handwash` data set into a new Minitab worksheet.

From the menu bar, select **Stat > ANOVA > General Linear Models**. Select and transfer the 1nAfter column into the **Responses:** box, then select and transfer the Cleanser column into the **Model:** box. Click the **Covariates** button, then select and transfer the 1nBefore column into the **Covariates:** box. Click **OK**, and then click **OK** again.

Chapter 6
Categorical Data Analysis: Is a Tumor Malignant or Benign?

Activities

3. Copy and paste the `Table 6.1` data set into a new Minitab worksheet

From the menu bar, select **Graph > Bar Chart**. Select **Values from a table** option from the **Bars represent:** dropdown menu, select **Two-way Table Stack**, and click **OK**.

Select and transfer the Malignant and Benign columns into the **Graph variables:** box, and then enter Shape (C1) into the **Row Labels:** box. Select the **Rows are outermost categories and columns are innermost** option, and make sure the **Stack the innermost category values** box is checked. Click the **Chart options** button and check the **Show Y as a Percent** box and select the **Within categories at level 1 (outermost)** option. Click **OK**, and then click **OK** again.

Note: to show the actual percentages on the graph itself, from within the **Bar Chart – Values from a table, Two-way table, Stack** window click the **Labels** button, select the **Data labels** tab, and select the **Use y-value labels** option. Click **OK**, and then click **OK** again.

6. Open a new Minitab worksheet and enter '1' into the first 24 rows of C1 (each representing a malignant cell) and '0' into the next 13 rows of C1 (each representing a benign cell). Label the column 'Malignancy.'

Click anywhere in the Session window, and from the menu bar, select **Editor > Enable Commands**. The Minitab prompt, `MTB >`, will appear in the Session window.

From the menu bar, select **Calc > Random Data > Sample from Columns**. Sample 21 rows from column C1 (Malignancy). Select and transfer C3 into the **Store samples in:** box. Make sure the **Sample with Replacement** box is not checked. Click **OK**. Notice that column C3 now has 21 rows of data and the session window now shows a small amount of computer code used to run this function. Entering the line of code `MTB > Sample 21 C1 C3` will execute the same function.

From the menu bar, select **Calc > Column Statistics**. Select the **Sum** option, then select and transfer C3 into the **Input variable:** box. Click **OK**. The number of "1's" observed in C3 represents the number of concave nuclei that are malignant.

8. Open the program Notepad (**Programs > Accessories > Notepad**) and enter the following three lines:

```
Sample 21 C1 C3.
Let C5(k1) = sum(C3)
Let k1=k1+1
```

Click **File > Save as**. Name the file Malignant.mtb, and select **All File Types** from the dropdown menu before clicking **Save**.

In the Minitab Session window, enter the command `MTB > let k1=1`. From the menu bar, select **File > Other files > Run an exec**. In the **Number of times to execute** box, enter 1, then click on the **Select File** button. Locate the Malignant.mtb file on your computer and click **Open**. Verify that C3 contains 21 observations and C5 is the total number of malignant concave nuclei (total number of 1's in C3), then repeat the process 9999 more times.

In the Session window, enter the commands `MTB > let C6=(C5>=17)` and `MTB > sum C6`.

From the menu bar, select **Graph > Histogram > Simple > OK**. Select and transfer C5 into the **Graph variables:** box and click **OK**. In the Session window, enter the commands `MTB > let c6=(c5>=17)` and `MTB > sum C6`.

9. From the menu bar select **Calc > Probability Distributions > Hypergeometric**. Select the **Probability** option; enter 37 into the **Population size (N)** box , enter 24 into the **Successes in population (M)** box, and enter 21 into the **Sample size (n)** box. Use Counts as the **Input column**. Click **OK**. This will generate a column labeled Counts representing all possible outcomes starting with 0 and ending with 21.

10. Label the hypergeometric probabilities from the Activity 9 as "Prob."

From the menu bar, select **Graph > Histogram > Simple > OK**. Select and transfer the Counts column into the **Graph variables:** box. Click the **Data Options** button and then click the **Frequency** tab. Enter Prob*10000 into the **Frequency variable(s):** box and click **OK**. Click the **Scale** button and then click the **Y Scale** tab and select the **Percent** option. Click **OK**, and then click **OK** again.

11. In Minitab, the **Cumulative probability** tab can be used to find $P(X \geq 17)$, but note that since these are discrete outcomes, $P(X \geq 17) = 1 - P(X \leq 16)$.

19. Using the data from Table 6.1, from the menu bar select **Stat > Tables > Chi-Square Test**. Select and transfer the Benign and Malignant columns into the **Columns containing the table:** box. Click **OK**.

Extended Activities

34. Copy and paste the `Table 6.7` data set into a new Minitab worksheet.

From the menu bar select **Stat > Tables > Chi-SquareGoodness-of-Fit Test**. Select and transfer the Observed column into the **Observed counts:** box and select the **Test equal proportions** option. Click **OK**.

Chapter 7
Logistic Regression: The Space Shuttle Challenger

Activities

2. Copy and paste the `Shuttle` data into Minitab.

From the menu bar, select **Graph > Scatterplot > Simple > OK**. Select and transfer the Successful Launch column into the **Y variables:** box, and then select and transfer the Temperature column into the **X variables:** box. Click **OK**.

3. From the menu bar, select **Stat > Regression > Fitted line plot**. Select and transfer the Successful Launch column into the **Response (Y):** box, and then select and transfer the Temperature column into the **Predictor (X):** box. Click **OK**.

5. To generate X values, go to the menu bar and select **Calc > Make Patterned Data > Simple Set of Numbers**. Enter X into the **Store patterned data in:** box, enter 0 into the **From first value:** box, enter 30 into the **To last value:** box, enter 0.1 into the **In steps of:** box, and verify that 1 is entered into both the **Number of times to list each value:** and the **Number of times to list each sequence:** boxes.

To calculate probabilities (i.e. using b0=-10 and b1=.5), go to the menu bar and select **Calc > Calculator**. Enter Y1 into the **Store result in variable:** box, and enter EXP(-10+0.5*'X')/(1+EXP(-10+0.5*'X')) into the **Expression:** box. Repeat calculating Y2, Y3, etc. for each b0 and b1 value.

From the menu bar, select **Graph > Scatterplot > Simple > OK**. Input each Y1, Y2, Y3….. variable vs. X. Click the **Multiple Graphs** button and select the **Overlaid on the same graph** option. Click **OK**, and then click **OK** again.

6. From the menu bar, select **Stat > Regression > Binary Logistic Regression**. Select and transfer the Successful Launch column into the **Response:** box, and then select and transfer the Temperature column into the **Model:** box. Click the **Options** button and make sure that the **Logit** option is selected on the **Link Function** menu. Click **OK**. Click the **Storage** button and check the **Event probability** button. Click **OK**, and then click **OK** again.

In Minitab, the EPRO1 column now gives the calculated probabilities, $\hat{p}$, for each xi.

7. From the menu bar, select **Calc > Calculator**. Enter Y1 into the **Store result in variable:** box, and enter EXP(b0+b1*'X')/(1+EXP(b0+b1*'X')) [where b0 and b1 are MLE estimates from Activity 6] into the **Expression:** box. Click **OK**.

11. To generate X values, go to the menu bar and select **Calc > Make Patterned Data > Simple Set of Numbers**. Enter X into the **Store patterned data in:** box, enter 20 into the **From first value:** box, enter 90 into the **To last value:** box, enter 0.1 into the **In steps of:** box, and verify that 1 is entered into both the **Number of times to list each value:** and the **Number of times to list each sequence:** boxes.

To calculate**,** go to the menu bar and select **Calc > Calculator**. Enter Y1 into the **Store result in variable:** box, and enter EXP(b0+b1*'X')/(1+EXP(b0+b1*'X')) into the **Expression:** box. Repeat for each b0 and b1 value**.**

From the menu bar, select **Graph > Scatterplot > Simple > OK**. Input each Y1, Y2, Y3….. variable vs. X. Click the **Multiple Graphs** button and select the **Overlaid on the same graph** option. Click **OK**, and then click **OK** again.

13. To change the X variable, go to the menu bar and select **Data > Code > Change Data type**. Select and transfer the Successful Launch column into the **Code data from columns:** box, then select and transfer to Damage column into the **Store coded data into columns:** box. Enter 1 and 0 into the **Original values:** box. Click **OK**.

From the menu bar, select **Stat > Regression > Binary Logistic Regression**. Select and transfer the Damage column into the **Response:** box, and then select and transfer the Temperature column into the **Model:** box. Click the **Options** button and make sure that the **Logit** option is selected on the **Link Function** menu. Click **OK**, and then click **OK** again.

14. From the menu bar, select **Stat > Regression > Binary Logistic Regression**. Select and transfer the Successful Launch column into the **Response:** box, and then select and transfer the Temperature column into the **Model:** box. Click the **Options** button and make sure that the **Logit** option is selected on the **Link Function** menu. Click **OK**, and then click **OK** again.

Extended Activities

15. Copy and paste the `Cancer2` data into Minitab.

From the menu bar, select **Stat > Regression > Binary Logistic Regression**. Select and transfer the Malignant? column into the **Response:** box, and then select and transfer the Radius and Concave columns into the **Model:** box. Click the **Options** button and make sure that the **Logit** option is selected on the **Link Function** menu. Click **OK**, and then click **OK** again.

16. From the menu bar, select **Stat > Regression > Binary Logistic Regression**. Select and transfer the Malignant? column into the **Response:** box, and then select and transfer the Radius column into the **Model:** box. Click the **Options** button and make sure that the **Logit** option is selected on the **Link Function** menu. Click **OK**, and then click **OK** again.

26. From the menu bar, select **Stat > Regression > Binary Logistic Regression**. Select and transfer the Malignant? column into the **Number of events:** box, enter Total into the **Number of Trials:** box, and then select and transfer the Radius column into the **Model:** box. Click the **Options** button and make sure that the **Logit** option is selected on the **Link Function** menu. Click **OK**, and then click **OK** again.

27-28. Repeat the steps give for Activity 26, this time clicking the **Storage** button and checking the **Pearson residuals** and **Deviance residuals** boxes before clicking **OK**, and then clicking **OK** again.

From the menu bar, select **Stat > Basic Statistics > Normality Test**. Select and transfer the residual values into the **Variable** box and click **OK**.

31. From the menu bar, select **Stat > Regression > Binary Logistic Regression**. Select and transfer the Malignant? column into the **Response:** box, and then select and transfer the Radius column into the **Model:** box. Click **OK**

32. From the menu bar, select **Stat > Regression > Binary Logistic Regression**. Select and transfer the Malignant? column into the **Response:** box, and then select and transfer the Radius column into the **Model:** box. Click the **Options** button and enter a value into the **Number of groups:** box. Click **OK**, and then click **OK** again.

33. From the menu bar, select **Stat > Regression > Binary Logistic Regression**. Select and transfer the Malignant? column into the **Response:** box, and then select and transfer the Radius column into the **Model:** box. Click the **Storage** button and check the **Pearson residuals** and **Deviance residuals** boxes, then click **OK**. Click the **Graphs** button and select the appropriate graphs. Click **OK**, and then click **OK** again.

34. To generate pi values, go to the menu bar and select **Calc > Make Patterned Data > Simple Set of Numbers**. Enter pi into the **Store patterned data in:** box, enter 0.1 into the **From first value:** box, enter 0.99 into the **To last value:** box, enter 0.1 into the **In steps of:** box, and verify that 1 is entered into both the **Number of times to list each value:** and the **Number of times to list each sequence:** boxes. Click **OK**.

From the menu bar, select **Calc > Calculator**. Enter **loglik** into the **Store result in variable:** box, and enter 5*log(pi)+7*log(1-pi) into the **Expression:** box. Click **OK**.

From the menu bar, select **Graph > Scatterplot > Simple**. Select and transfer the Loglik column into the **Y variable:** box, and select and transfer the pi column into the **X variable:** box. Click **OK**.

Chapter 8
Poisson Log-Linear Regression: Detecting Cancer Clusters

Minitab is not capable performing Poisson regression.

Chapter 9
Survival Analysis: Melting Chocolate Chips

Activities

19. Copy and paste the `MeltingChipsJS` data into Minitab. Note that in the Censor column, 0 represents a censored time and 1 represents a complete time.

From the menu bar, select **Stat > Reliability/Survival > Distribution Analysis (Right Censoring) > Nonparametric Distribution Analysis**. Select and transfer the Time column into the **Variables:** box, then select and transfer the Type column into the **By variable:** box.

Click the **Censor** button, then select and transfer the Censor column into the **Use censoring columns:** box. Click **OK**.

Click the **Graph** button and check the **Survival Plot** box. Click **OK**, and then click **OK** again.

25. Follow the steps outlined for Activity 19.

29. Use only the milk chocolate chip data with two columns Time and Censor from `MeltingChipsJS` data set.

From the menu bar and select **Stat > Reliability/Survival > Distribution Analysis (Right Censoring) > Nonparametric Distribution Analysis**. Select and transfer the Time column into the **Variables:** box.

Click the **Censor** button, then select and transfer the Censor column into the **Use censoring columns:** box. Click **OK**.

Click the **Graph** button and check the **Survival Plot** and **Display confidence intervals** boxes. Click **OK**.

Click the **Storage** button and check the **Confidence limits for survival probabilities** box. Click **OK**.

Click the **Estimate** button and check the **Kaplan Meier** box. Enter 95% into the **Confidence level** box and then select Two-sided for the **Confidence intervals** option. Click **OK**.

31. Copy and paste the entire `MeltingChipsJS` data into Minitab. Note that in the Censor column, 0 represents a censored time and 1 represents a complete time.

From the menu bar, select **Stat > Reliability/Survival > Distribution Analysis (Right Censoring) > Nonparametric Distribution Analysis**. Select and transfer the Time column into the **Variables:** box, then select and transfer the Type column into the **By variable:** box.

Click the **Censor** button, then select and transfer the Censor column into the **Use censoring columns:** box. Click **OK**.

Click the **Graph** button and check the **Survival Plot** and **Display confidence intervals** boxes. Click **OK**.

Click the **Storage** button and check the **Confidence limits for survival probabilities** box. Click **OK**.

Click the **Estimate** button and check the **Kaplan Meier** box. Enter 95% into the **Confidence level** box and then select Two-sided for the **Confidence intervals** option. Click **OK**. Note that if you are comparing groups, by default the Kaplan-Meier curves will be plotted on the same graph; however, you have the option to plot them on separate graphs.

36. Follow the steps outlined for Activity 31.

Extended Activities

43-49. While it is possible to plot cumulative hazard functions (using the **Graph** option) the format is somewhat different than what is used in this chapter. Thus we suggest using R or other software package for these activities.

Chapter 10
Principal Component Analysis: Stock Market Values

Activities

1. Copy and paste the `2006Stocks` data into Minitab.

From the menu bar, select **Graph > Time Series Plot > Simple > OK**. Select and transfer the Dow and S&P columns into the **Series:** box. Click the **Multiple graphs** button and select the **Overlaid on the same graph** option. Click **OK**, and then click **OK** again.

2. From the menu bar select **Calc > Standardize**. Select and transfer the Dow column into the **Input column(s):** box. Select the **Subtract mean and divide by std. dev.** option, and enter Z1 into the **Store results in:** box. Click **OK**. Repeat this process for the S&P 500 column, storing the standardized data in Z2.

From the menu bar, select **Graph > Time Series Plot > Simple > OK**. Select and transfer the Z1 and Z2 columns into the **Series:** box. Click the **Multiple graphs** button and select the **Overlaid on the same graph** option. Click **OK**, and then click **OK** again.

3. From the menu bar select **Calc >Calculator**. Enter PC1 into the **Store result in variable:** box. In the **Expression:** box, enter Z1*(.707) + Z2*(.707). Click **OK**.

From the menu bar, select **Graph > Time Series Plot > Simple > OK**. Select and transfer the Z1, Z2, and PC1 columns into the **Series:** box. Click the **Multiple graphs** button and select the **Overlaid on the same graph** option. Click **OK**, and then click **OK** again.

4. Follow the **Calc > Standardize** instructions for Activity 2 to create column Z3.

Follow the **Calc > Calculator** instructions for Activity 3 to create column C1.

From the menu bar, select **Graph > Time Series Plot > Simple > OK**. Select and transfer the Z1, Z2, Z3, and C1 columns into the **Series:** box. Click the **Multiple graphs** button and select the **Overlaid on the same graph** option. Click **OK**, and then click **OK** again.

Repeat the process for C2, C3, and C4.

5. From the menu bar, select **Stat > Basic Statistics > Correlation**. Select and transfer the ZI and Z2 columns into the **Variables:** box. Check the **Store matrix (display nothing)** box. Click **OK**. This will automatically store the correlation matrix as Corr1.

To view this matrix it is best to use the Session editor. Click anywhere on the Session window, then go to the menu bar and select **Editor > Enable commands**. In the Session window, enter command `MTB > print Corr1` to view the correlation matrix.

6. Follow the instructions for Activity 5.

9. From the menu bar, select **Calc > Matrices > Eigen Analysis**. Enter Corr1 into the **Analyze matrix:** box. Have the eigenvalues output into a column called EValues, and and have the matrix of eigenvectors output into a column called EVectors. Click **OK**. Note the Corr1 was calculated in Activity 5.

Click on the Session window. From the menu bar, select **Editor > Enable commands**. In the Session window, enter the commands `MTB > print Evectors` and `MTB > print Evalues`.

10. To create PC1 using v1, go to the menu bar and select **Calc > Calculator**. Select and transfer column PC1 into the **Store result in variable:** box, then enter Z1*(.707) + Z2*(.707) into the **Expression:** box. Click **OK**.

Repeat the process using v2 to create a column called PC2.

a) From the menu bar, select **Stat > Basic Statistics > Correlation**. Select and transfer columns PC1 and PC2 into the **Variables:** box and click **OK**.

b) From the menu bar, select **Stat > Basic Statistics > Display Descriptive Statistics**. Select and transfer columns PC1 and PC2 into the **Variables:** box and click **OK**.

11. From the menu bar, select **Graph > Time Series Plot > Simple > OK**. Select and transfer the Z1, Z2, PC1, and PC2 columns into the **Series:** box. Click the **Multiple graphs** button and select the **Overlaid on the same graph** option. Click **OK**, and then click **OK** again.

12. Note that in Minitab, the eigenvectors are output as PC1, PC2, PC3 and PC4, but in the text we refer to them as v1, v2, v3 and v4. The principal components are called scores.

From the menu bar, select **Stat > Multivariate > Principal Components**. Select and transfer the columns Z1, Z2, and Z3 into the **Variables:** box. Click the **Graphs** button and select the **Scree plot** option. Click **OK**. Click the **Storage** button, then enter PC1, PC2, and PC3 (or select any three unused columns) into the **Scores:** box. Click **OK**, and then click **OK** again.

13. From the menu bar select **Graph > 3DScatterplot > Simple > OK**. Select and transfer the columns Z1, Z2, and Z3into any of the **X**, **Y**, and **Z variable** boxes. Click **OK.** Use the arrows on the **3D Graph Tools** to rotate the graph.

The process of drawing eigenvectors with Minitab requires multiple steps:

i. Add the points corresponding to the eigenvectors to your scatterplot (add the eigenvectors to the bottom of the Z1Z2 Z3 columns).

In other words, write
-.0540 in the 252th row of Z1,
-0.582 in the 252th row of Z2, and
-0.608 in the 252th row of Z3

Place the second eigenvector in the 253rd row of the Z1, Z2, and Z3 columns and place (0,0,0) on the 254th row of the Z1 Z2 Z3 columns.

ii. Create a new column called ‘CODE’ so that the “eigenvector points" are distinct from all the other data.

In the CODE column, write the word "original" in the first 251 rows (use the click and drag function to copy the word), then write "First" on the 252th row, "Second" on the 253rd row and "Origin" on the 254th row.

iii. Draw a three dimensional graph. From the menu bar, select **Graph > 3DScatterplot > With Groups > OK**. Select and transfer the columns Z1, Z2, and Z3 into any of the **X**, **Y**, and **Z variable** boxes. Select and transfer the CODE column into the **Categorical variable:** box. Click **OK**.

Use the arrows on the 3D Graph Tools to rotate the graph enough to convince yourself that the first eigenvector is in the direction of the most variation. Continue to rotate the graph to convince yourself that the second eigenvector is perpendicular to the PC1 vector. Submit this rotated graph on an attached page. This graph does not need to be perfectly rotated.

14. From the menu bar, select **Graph > Time Series Plot > Simple > OK**. Select and transfer the Z1, Z2, Z3, PC1, PC2, and PC3 columns into the **Series:** box. Click the **Multiple graphs** button and select the **Overlaid on the same graph** option. Click **OK**, and then click **OK** again.

Extended Activities

16. From the menu bar, select **Stat > Multivariate > Principal Components**. Select and transfer the X1, X2, and X3 columns into **the Variables:** box and select the **Covariance matrix** option. Click **OK**.

17. From the menu bar, select **Graph > Time Series Plot > Simple > OK**. Select and transfer the columns X1, X2, X3, and PC1 into the **Series:** box. Click the **Multiple graphs** button and select the **Overlaid on the same graph** option. Click **OK**, and then click **OK** again.

18. From the menu bar, select **Calc >Calculator**. Enter NewNasdaq into the **Store result in variable:** box. In the **Expression:** box, enter 5000*Nasdaq. Click **OK**.

Follow the previously provided instructions to perform a PCA.

19. Copy and paste the `Veriscolor` data into the first four columns (C1-C4) of a new Minitab worksheet. Note that in Minitab, the eigenvectors are output as PC1, PC2, PC3 and PC4, but in the text we refer to them as v1, v2, v3 and v4. The principal components are called scores.

From the menu bar, select **Stat > Multivariate > Principal Components**. Select and transfer the columns C1, C2, C3, and C4 into the **Variables:** box. Click the **Graphs** button and select the **Scree plot** and **Correlation matrix** options. Click **OK**, and then click **OK** again.

24. Copy and paste the `Cars` data into a new Minitab worksheet. Follow the previously provided instructions to calculate PC1.

From the menu bar, select **Stat > Regression > Regression**. Select and transfer the LnPrice column into the **Response:** box, and select and transfer the appropriate columns into the **Predictors:** box. Click the **Graphs** button and check the **Four in one** box. Click **OK**, and then click **OK** again.

Chapter 11
Bayesian Data Analysis: What Colors Come in Your M&M's® Candy Bag?

Activities

2. Copy and paste the `MMs` data into Minitab. Or, input your own data into a new Minitab worksheet.

From the menu bar, select **Graph > Scatterplot > With connect line > OK**. Select and transfer the Proportion column into the **Y variable:** box, then select and transfer the Total column into the **X variable:** box. Click **OK**.

Double click the vertical axis of the scatterplot. In the **Scale** tab, uncheck the **Minimum:** and **Maximum:** boxes, which have been automatically set at 0 and 1, respectively. Click **OK**.

Right click the vertical axis of the scatterplot, then click **Add > Reference lines**. In the **Show reference lines for Y positions:** box, enter 0.5{you can add additional lines if you like}. Click **OK**.

Extended Activities

34, 36, 38. Software that better conducts simulations, such as R, should be used to perform these activities.

41. Copy and paste the `MMs` data into Minitab. Or, input your own data into a new Minitab worksheet.

From the menu bar, select **Calc > Probability Distributions > Beta**. Select the **Inverse cumulative probability** option. In the **First shape parameter:** box, enter 24; in the **Second shape parameter:** box, enter 33; in the **Input constant:** box, enter 0.025. Click **OK**.

Repeat the process, this time entering 0.975 into the **Input constant:** box. Click **OK**.

45. The posterior distribution for pi is Beta (24, 33). Thus $p(pi < 0.5 \mid x) = 0.885597$. Note that this appears to be consistent with Figure 11.6

From the menu bar, select **Calc > Probability Distributions > Beta**. Select the **Cumulative probability** option. In the **First shape parameter:** box, enter 24; in the **Second shape parameter:** box, enter 33; in the **Input constant:** box, enter 0. 5. Click **OK**.

46. To find the 95% credible intervals, follow the instructions provided for Activity 41.