CHAPTER 2

Appendix

Examining Relationships with Excel, JMP, Minitab, SPSS, CrunchIt!, R, and a TI-83/-84 Calculator

Scatter Plots

To create a scatterplot in Excel, the data for the two variables should be placed in two adjacent columns, with the column associated with the Y variable to the right of the column associated with the X variable. If the data file does not have the X variable first, copy and paste the data column(s) into another area of the spreadsheet. Click and drag the mouse to highlight the cells of the two columns of data.

1. Click **Insert** followed by **Scatter** in the Charts group.
2. Choose the scatterplot option with no line connections.
3. If the gridlines are not desired, you can click on them and delete them by hitting the delete key or by right-clicking and selecting "Delete."
4. To add a grouping variable, watching the video referenced below is recommended. There is no automated procedure in Excel to do this.

Note: The layout of the scatterplot can also be manipulated by choosing among a variety of options offered within the Charts Layouts group found under the Design tab. To add the regression line, click **Add Chart Element ➔ Trend line ➔ Linear**.

For videos to help with these topics, see the Excel Video Technology Manuals on Scatterplots and Scatterplot by Groups.

1. Click **Analyze ➔ Fit Y by X**.
2. Select the Y variable and click "Y, Response."
3. Select the X variable and click "X, Factor."
4. Click **OK**.

To add groups,

1. Click **Graph ➔ Graph Builder**.
2. Click and drag the two variables to the X and Y axes.
3. Drag the grouping variable to the "Color" box.
4. Click **Done**.

For videos to help with these topics, see the JMP Video Technology Manuals on Scatterplots and Scatterplot by Groups.

1. Click **Graph → Scatterplot**.
2. Select "Simple" (the default) for the type of scatterplot, then click **OK**.
3. Click to highlight the name of the data column associated with the Y (response) variable, then select it into row 1 of the Y variables box. Repeat to select the X (explanatory) variable into row 1 of the X variables box.
4. Click **OK**.

To add groups, select the "With Groups" option instead of "Simple." Click and highlight the Y and X variables, then add the categorical grouping variable in the box below.

For videos to help with these topics, see the Minitab Video Technology Manuals on Scatterplots and Scatterplot by Groups.

1. Click **Graphs → Chart Builder**
2. Select **Scatter/Dot** and drag the first plot type (scatter without lines) to the graph area.
3. Select the response variable and drag it to the "Y Axis?" box.
4. Select the predictor variable and drag it to the "X Axis?" box.
5. Add any title or expanded axis labels you desire using the "Edit Properties of" area.
6. Click **OK**.

To add a categorical grouping variable, click **Groups/Points ID**. Place a check next to "Grouping/stacking variable." Select that variable and drag it into the box labeled "Set color."

For videos to help with these topics, see the SPSS Video Technology Manuals on Scatterplots and Scatterplot by Groups.

1. Click **Graphics → Scatterplot**.
2. Select the X variable for "X" and Y variable for "Y."
3. In the parameters section, you can specify additional options and labels.
4. Click **Calculate**.

To add a categorical grouping variable, enter it in the box labeled "Group by."

For videos to help with these topics, see the CrunchIt! Help Video on Scatterplots.

1. Press [STAT] and select "Edit" to enter the list editor.
2. In L1, enter X values.
3. In L2, enter Y values.
4. Press STAT PLOT ([2nd] [Y=]) and select the scatterplot ⁘.
5. Press [2nd] [1] = [L1] for Xlist and [2nd] [2] = [L2] for Ylist.
6. Press [ZOOM] [9].

To add a categorical grouping variable, you will need to have lists for each group. Define up to three scatterplots using different graphing symbols.

For videos to help with these topics, see the TI-83/-84 Video Technology Manuals on Scatterplots and Scatterplot by Groups.

1. The simplest command is

```
> plot(x,y)
```

2. Add labels and titles and specify a data set by adding other parameters (note: the + symbol used here is the R prompt for a continuation line):

```
> plot (x,y,main="title",xlab="X label",ylab="Y
+ label", data=dataset)
```

To add a grouping variable, modify the command to something like below (for 3 groups)

```
> plot (x,y,main="title",xlab="X label",ylab="Ylabel",
+ data=dataset,bg=c("red","green","blue")[unclass(gpvar)])
```

For videos to help with these topics, see the R Video Technology Manuals on Scatterplots and Scatterplot by Groups.

Correlation

1. Click **Data ➔ Data Analysis ➔ Correlation ➔ OK**.
2. Enter the cell range of the data on the two variables (placed in adjacent columns) into the "Input Range" box.
3. Click **OK**.

The Excel Video Technology Manual on Correlation offers more help and an example.

1. Click **Applets ➔ Demonstrate Regression**.
2. Click the radio button next to My Sample. Select the appropriate response ("Sheet 1" for a data table already loaded or "Open another data table" followed by electing a new file.
3. Click **OK**.
4. Select the Y variable and click "Y, Response."
5. Select the X variable and click "X, Factor."
6. Click **OK**.
7. The correlation will be displayed on the left, just above the sample size, and is labeled "Corr."

The JMP Technology Video Manuals on Correlation offers more help and an example.

1. Click **Stat ➔ Basic Statistics ➔ Correlation**.
2. Highlight the variable names and select them into the "Variables" box.
3. Click **OK**.

The Minitab Video Technology Manual on Correlation offers more help and an example.

1. Click **Analyze ➔ Correlate ➔ Bivariate**.
2. Select at least two variables and click the right arrow to enter variables.
3. Click **OK**.

The SPSS Technology Video Manual on Correlation offers more help and an example.

CRUNCHIT!

1. Click **Statistics ➔ Correlation**.
2. Check the boxes next to variables you wish to correlate (at least two).
3. Click **Calculate.**

The CrunchIt! Help Video on Correlation offers more help and an example.

TI-83/-84

1. Press [STAT] and select "Edit" to enter the list editor.
2. In L1, enter X values.
3. In L2, enter Y values.
4. Press [STAT] [▸] to Calc and select option 8: LinReg(a+bx). Press [ENTER].
5. Press [2nd] [1] for Xlist and [2nd] [2] for Ylist (or the lists with the variables of interest). FreqList should be set to 1, and press [ENTER].
6. The correlation coefficient is returned along with the linear regression line.

Note: You must first "turn on" the display of r and r^2. Do this by pressing [2nd] [0] (Catalog), [x^{-1}] (C in alpha mode), then arrow down to DiagnosticOn. Press [ENTER] twice.

The TI-83/-84 Technology Video Manual on Correlation offers more help and an example.

The correlation command is

```
> cor (x,y)
```

You can specify the data set name with an optional "data=dataset" parameter.

The R Technology Video Manuals on Correlation offers more help and an example.

Least-Squares Regression

EXCEL

1. Click **Data ➔ Data Analysis ➔ Regression ➔ OK**.
2. Enter the cell range of the response data into the "Input Y Range" box and the cell range of the explanatory data into the "Input X Range" box.
3. If you want the residual values, then place a check mark next to the "Residuals" option. A check next to "Residual plots" will produce a plot of the residuals against X.
4. Click **OK**.
5. If you wish to produce a scatterplot superimposed with the least-squares regression line, then do *not* select the "Line Fit Plots." Instead, create a scatterplot as described earlier in this appendix; click **Add Chart Element ➔ Trend Line ➔ Linear**.

The Excel Technology Video Manuals on Correlation, Fitted Line Plot, Linear Regression, Prediction, and Residual Plots offer more help and examples.

1. Click **Analyze ➔ Fit Y by X**.
2. Select the Y variable and click "Y, Response."
3. Select the X variable and click "X, Factor."
4. Click **OK**.
5. Click the Red Triangle ➔ **Fit Line**.

Note: To show diagnostic plots, click the red triangle next to "Linear Fit" and select "Plot Residuals." You can save residuals and predicted values from this menu.

Alternatively, you can use **Applets → Demonstrate Regression** as detailed in the above section on correlation, but this option will not allow you to examine residuals.

The JMP Technology Video Manual on Linear Regression offers more help and examples.

1. Click **Stat → Regression → Regression → Fit Regression Model**.
2. Click in the response (Y) variable data column into the "Response" box.
3. Click in the explanatory (X) variable data column into the "Continuous Predictors" box.
4. If you want the residual values, click the "Storage" button and place a check mark next to the "Residuals" option; select the "Fits" option if you want the predicted values. Click **OK** to close this window.
5. Residual Plots can be created using the **Graphs** option in the regression menu box. The "4 in 1" plot does them all.
6. Click **OK**. Note: If you have asked for residuals plots, you will see those before the actual regression output.

If you wish to produce a scatterplot superimposed with the least-squares regression line, do the following pull-down sequence:

1. Click **Stat → Regression → Fitted Line Plot**.
2. Click in the response variable data column into the "Response (Y)" box and the explanatory variable data column into the "Predictor (X)" box.
3. Click **OK**.

The Minitab Technology Video Manuals: Correlation, Fitted Line Plot, Linear Regression, Prediction, and Residual Plots offer more help and examples.

1. Click **Analyze → Regression → Linear**.
2. Enter outcome variable in the "Dependent" box.
3. Enter predictor variable in "Independent(s)" box.
4. To save residuals, click **Save** and check the box next to the type of residual you wish to save (typically, Unstandardized). You can also save predicted values here.
5. You can click **Plots** and ask for a histogram and Normal probability plot of the residuals if desired.
6. Click **OK**.

The SPSS Technology Video Manuals on Correlation, Fitted Line Plot, Linear Regression, Prediction, and Residual Plots offer more help and examples.

1. Click **Statistics → Regression → Simple Linear**.
2. Enter outcome variable in "Dependent Variable."
3. Enter predictor variable in "Independent Variable."
4. Click **Calculate**.

For residual plots or a fitted line plot, change the option in the "Display" drop-down.

The CrunchIt! Help Video on Simple Linear Regression offers more help and examples.

TI-83/-84

To calculate the regression line, follow the instructions given in the Correlation section above.

For a fitted line plot, in the "Store RegEQ" line: VARS ▶ (to Y-Vars), ENTER (for Function) and ENTER (for Y1).

Press ZOOM 9 to display the scatterplot again.

For residuals plots, find the list of residuals (RESID) in the 2nd, STAT = LIST menu.

The TI-83/-84 Technology Video Manuals on Correlation, Fitted Line Plot, Linear Regression, Prediction, and Residual Plots offer more help and examples.

1. The simplest regression command is

```
> lm(y~x)
```

2. To replicate output and get a complete table of results, use a command like

```
> model<-lm(y~x)
> summary(model)
```

3. To create diagnostic (residuals) plots, access them as resid(model) as in

```
> hist(resid(model))
```

The R Technology Video Manuals on Correlation, Fitted Line Plot, Linear Regression, Prediction, and Residual Plots offer more help and examples.

Relationships in Categorical Data

When considering data input options, you will need to pay attention to what is desired/required for the different software. Excel wants data entered so they look like a two-way table as shown below (layout A).

District	Unsold	Sold
A	8889	24009
B	1329	32658

Most other software wants data entered in a different format, as shown below for the same data (layout B).

District	Result	Count
A	Sold	24009
A	Unsold	8889
B	Sold	32658
B	Unsold	1329

If you have marginal totals, instructions on producing a bar chart to display them can be found in the Bar Charts section of the Chapter 1 Tech Appendix. Below, we detail creating clustered or stacked bar charts to display conditional distributions.

EXCEL

1. Enter the data as a two-way table (layout A) and use the cursor to highlight the entire table.
2. Click **Insert**.
3. From the Charts section, click the down arrow by **Recommended Charts**.
4. Select the chart option you desire from the available options.

To create a clustered bar graph (or stacked bar graph) with the Y axis in counts, do the following:

1. Your data should be entered as in layout B shown above.
2. Click **Graph ➔ Graph Builder**.
3. Select the conditioning variable and drag it to the X axis.
4. Select the bar graph  from the chart styles row.
5. Select the "response" variable and drag it to the Overlay box.
6. Select and drag the count variable to the Freq box.

At this point, you have a side-by-side (clustered) bar graph with the Y axis in counts.

7. If you want a stacked bar chart, use the drop-down for "Bar Style" to select **Stacked**.

To duplicate the mosaic plot shown in Figure 2-23, do the following.

1. Click **Analyze ➔ Fit Y by X**.
2. Click the conditioning variable into the "X, Factor" box and the response category variable into the "Y, response" box.
3. Click to enter the column of counts into the "Freq" box.
4. Click **OK**.

MINITAB

1. Your data should be entered as shown in Layout B.
2. Click **Graph ➔ Bar Chart**.
3. Use the drop-down to change "Bars represent" to **Values from a table**.
4. Under "One column of values" select either **Cluster** or **Stack**. Click **OK**.
5. Highlight and **Select** the count variable column into the "Graph variables" box.
6. Highlight and **Select** the two categorical variables into the "Categorical variables for grouping" box. Make sure the variable you are using for conditioning is specified first.
7. Use **Labels** to give the bar graph a meaningful title. Click **Continue**.
8. Click **Chart Options**. Click to check the box next to "Show Y as Percent." Click to change the radio button to "Within categories at level 1."
9. Click **OK ➔ OK**.

SPSS

1. Your data should be entered as shown for layout B.
2. Click **Graphs ➔ Chart Builder ➔ Bar**.
3. Select and drag either the clustered bar or stacked bar icon into the graph area.
4. Select and drag the conditioning variable to the "X-axis?" box.
5. Select and drag the second categorical variable to "Cluster on X: set color" or "Stack: set color" box. Note: The option name depends on the type of graph selected.
6. Select and drag the count variable to the "Count" box on the Y axis.
7. Give the bar graph a meaningful title using **Titles/Footnotes**.
8. Click **OK**.

9. If sample sizes in the groups are different and you are creating a stacked bar chart, you should adjust from counts to percents. Double-click in the graph for the Graph Editor. On the menu ribbon, click **Scale to 100%**  at the far right.
10. Click the **X** to close the graph editor.

CrunchIt! cannot create bar graphs for conditional distributions.

TI-83/-84

TI calculators cannot create bar graphs for conditional distributions.

There may be easier ways to accomplish this with packages, but the following procedure uses only base R commands.

Enter the data in order as in layout B.

Here we look at the results of a study of complications from different types of gastric surgery intended for weight loss. "NLT" means non–life threatening.

```
> surgery <- c('banding','banding','banding',
+ 'bypass','bypass','bypass',
+ 'gastrectomy','gastrectomy ','gastrectomy ')
> comp<-c('NLT','none','serious','NLT',
+ 'none','serious','NLT','none','serious')
> count<-c(81,606,31,5253,8110,604,46,325,19)
> gastric=data.frame(surgery,comp,count)
> levels(gastric$surgery)
> levels(gastric$comp)
> data=count
> data=matrix(data,ncol=3,byrow=T)
> colnames(data)=levels(gastric$surgery)
> rownames(data)=levels(gastric$comp)
```

At this point, you have created a matrix data structure. To see the matrix, simply enter the command "data." The results are shown below.

```
> data
```

	banding	bypass	gastrectomy
NLT	81	606	31
none	5253	8110	604
serious	46	325	19

To build the bar graph, we build using the rows of this table to condition the percents.

```
> prop=prop.table(data,margin=2)
> par(mar=c(5.1, 4.1, 4.1, 7.1), xpd=TRUE)
> barplot(prop,
+ col=heat.colors(length(rownames(prop))),
+ width=2)
> legend("topright",inset=c(-0.25,0),
+ fill=heat.colors(length(rownames(prop))),
+ legend=rownames(data))
```

If you add the parameter beside = T to the barplot command, you will have a clustered bar graph.