CHAPTER 3 Appendix

Scatter Plots and Correlation

This chapter introduces the idea of examining relationships between two quantitative variables. In particular, we are interested in first graphing the data to determine whether a relationship between the predictor (x) and response (y) variables exists.

If a relationship exists, is it linear? If so, we can calculate the correlation coefficient, r, to describe the strength and direction of the linear relationship.

Scatterplots

Excel

To create a scatterplot in Excel, place the data for the two variables in two adjacent columns, with the column associated with the y variable being in the column 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.

1. Use the cursor to drag and highlight the range of data, including the column names.

2. **Insert ➔ Scatter** in the Charts group.

3. Select the scatterplot option with no line connections .
4. If the gridlines are not desired, you can delete them by first clicking on them, then either pressing the delete key or right-clicking and selecting Delete.
5. Add a meaningful title by double-clicking in the placeholder title (typically the name of the response variable). Use the cursor to highlight it, and type over with the desired text.
6. Add axis labels to describe the variables and units by clicking **+**, then place a check by **Axis titles**. Click the right arrow to add placeholder labels. Drag the cursor to highlight these labels and type over them.
7. To add a grouping variable, watching the video mentioned at the end of this section is recommended. There is no automated procedure in Excel to do this.

Note: You can also manipulate the layout of the scatterplot 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*.

jmp
Statistical Discovery.™ From SAS.

1. **Analyze ➔ Fit Y by X**
2. Select the y variable into **Y, Response**
3. Select the x variable into **X, Factor**
4. **OK**

To add groups and title, use the **Graph ➔ Graph Builder** option. Click and drag the two variables to the x and y axes; drag the grouping variable to the Color box.

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

Minitab

1. **Graph ➔ Scatterplot**
2. Select "Simple" for the type of scatterplot.
3. **OK**
4. 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.
5. Add a title and axis labels with **Labels**.
6. **Continue ➔ OK**

To add groups, select **With Groups** option instead of Simple. **Select** the y and x variables, and then **Select** and add the grouping variable into the "Categorical Variables for grouping" box.

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

SPSS

1. **Graph ➔ Chart Builder ➔ Scatter/Dot**
2. Select **Simple Scatter** on the left of the top row.
3. Drag the response variable to the **Y Axis** box.
4. Drag the predictor variable to the **X Axis** box.
5. **Titles/Footnotes ➔ Title 1**
6. Enter a meaningful title for the plot into the **Content** box at right.
7. If needed, give axes more meaningful labels by clicking **X-axis 1** and **Y-axis 1** and entering those labels into the **Content** box at right.
8. **Apply ➔ Close**
9. **OK**

To add a categorical grouping variable, select **Grouped Scatter** (the second plot on the top row) instead of Simple Scatter. Drag that variable to the **Set color** box.

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

CRUNCHIT!

1. **Graphics ➔ Scatterplot**
2. Select the x variable for "X" and the y variable for "Y" with the drop-downs.
3. In the parameters section, you can specify additional options, titles, and axis labels.
4. **Calculate**

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

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

TI-83/-84

1. Press $\boxed{\text{STAT}}$ $\boxed{\text{ENTER}}$ to enter the list editor.
2. In L1, enter the x values.
3. In L2, enter the y values.
4. Press STAT PLOT ($\boxed{\text{2nd}}$ $\boxed{\text{Y=}}$), and select the scatterplot ($\boxed{\cdots}$).
5. Press $\boxed{\text{2nd}}$ $\boxed{1}$ = [L1] for Xlist, and $\boxed{\text{2nd}}$ $\boxed{2}$ = [L2] for Ylist.
6. Press ZoomStat ($\boxed{\text{ZOOM}}$ $\boxed{9}$). If necessary, adjust the window using $\boxed{\text{WINDOW}}$.

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 Video Technology Manuals on *Scatterplots* and *Scatterplot by Groups*.

The simplest command is

```
> plot(mydat$xvar,mydat$yvar)
```

Add labels and titles by adding other parameters:

```
> plot(mydat$xvar,mydat$yvar,main="title",xlab="X
label",ylab="Y label")
```

To add a grouping variable, modify the command to something like

```
> plot(mydat$xvar,mydat$yvar,main="title",xlab="X
label",ylab="Ylabel", pch=mydat$catvar)
```

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

Correlation

Excel

1. **Data ➔ Data Analysis**
2. Select **Correlation** in the menu box.
3. **OK**
4. Use the cursor to drag and select the cell range for the two variables (placed in adjacent columns) into the **Input Range** box.
5. **OK**

The Excel Video Technology Manuals: *Correlation* offers more help and an example.

jmp

1. **Analyze ➔ Fit Y by X**
2. Select the *y* variable and click **Y, Response**.
3. Select the *x* variable and click **X, Factor**.
4. **OK**
5. Click the Red Triangle ➔ **Density Ellipse ➔ 0.95**.
6. Expand the bivariate Normal ellipse P = 0.950 section.

The JMP Technology Manuals: *Correlation* offers more help and an example.

Minitab

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

The Minitab Video Technology Manuals: *Correlation* offers more help and an example.

SPSS

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

The SPSS Video Technology Manuals: *Correlation* offers more help and an example.

CRUNCH IT!

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

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

TI-83/-84

1. Press $\boxed{\text{STAT}}$ $\boxed{\text{ENTER}}$ to enter the list editor.
2. In L1, enter the *x* values.
3. In L2, enter the *y* values.
4. Press $\boxed{\blacktriangleright}$ to CALC and select option 8: **LinReg(a+bx)**. Press $\boxed{\text{ENTER}}$.
5. Press $\boxed{\text{2nd}}\boxed{1}\boxed{,}$ $\boxed{\text{2nd}}\boxed{2}$ (for L1,L2, or the lists with the variables of interest) after LinReg(a+bx) in the command window, and press $\boxed{\text{ENTER}}$. The comma is above the $\boxed{7}$.
6. The correlation coefficient is returned along with the linear regression line.

Note: You must first "turn on" the diagnostics to see *r* and r^2. Do this by pressing $\boxed{\text{2nd}}\boxed{0}$ (Catalog), $\boxed{x^{-1}}$ (D in alpha mode), then arrow down to DiagnosticOn. Press $\boxed{\text{ENTER}}$ twice. This needs to be done only once—unless you are *really* slow at changing batteries and the calculator resets to factory default.

The TI Video Technology Manuals: *Correlation* offers more help and an example.

The correlation command is

```
> cor (mydat$xvar,mydata$yvar)
```

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