1 One-Variable Statistics

The analysis toolpak can handle most of the one-variable statistics such as the mean and median.

1. Type a set of quantitative data into Excel. For example, use the fuel economy data given on page 31 in the text. (Use the city data or the highway data, but not both.)

2. Go to Tools→Data Analysis and select “Descriptive Statistics”.

3. Select the data to be the input range, select an empty cell to be the output range, and check the “Summary Statistics” box.

4. You’ll get a lot of information about the data, including many numbers that probably won’t mean anything to you at this point. The mean, median, mode, standard deviation, minimum, and maximum are all in there, and you can ignore the rest of the information it gives you.

5. Two important measures you do not get are the first and third quartiles. To get these, follow these instructions:
   
   (a) Select an empty cell, and type =quartile( into the cell. Include the open parentheses, and don’t hit the enter key yet.
   (b) Now, highlight the original set of data.
   (c) Finally, type a comma, then a “1” (for first quartile), close the parentheses “)” and hit “Enter” again. This will output the first quartile.
   (d) Select a different empty cell and repeat the procedure, replacing “1” with “3”, and you’ll have the third quartile.

6. Incidentally, the procedure for producing the quartiles can be repeated for any of the other one-variable statistics if you don’t want to produce the entire descriptive statistics output. (Excel uses =average( for mean rather than =mean(), though, so be careful there.)

2 Scatterplots

Scatterplots are pretty easy in Excel; you don’t even need the analysis toolpak.

1. Type in two sets of data that you think might share some relationship. For example, you could use both the city and highway fuel economy data on page 31.

2. Select the data you want to portray on a scatterplot.

3. Select Insert→Chart and select the “XY (Scatter)” option.

4. Click “Next” until you reach Step 3 in the chart creation.

5. At Step 3, you can enter a chart title and label the horizontal and vertical axes if you want.

6. More importantly, though, scatterplots look better without gridlines in them. Select the “Gridlines” tab and uncheck the “Major gridlines” box under “Value (Y) Axis”.
7. Continue clicking “Next” until you’ve got the chart.

For multiple categorical values in a scatterplot (as with Figure 2.2 on page 89, the example with heights and handspans of men vs. women, and Problem 2.21 on page 98)

1. First, you have to separate the data by category. You’ll need two columns (one for explanatory variable and one for response variable) for each category.

2. Proceed as before with the scatterplot creation, using the first category’s data as input, but stop at Step 2.

3. You’ll need to input the other categories’ data now. Click the “Series” tab, and then click the “Add” button.

4. Select the X and Y values for input in the next series.

5. Continue with this process until you have input all categories’ data.

6. Then proceed as before to finish the scatterplot.

3 Correlation

Excel calculates correlation very easily. “Correlation” is one of the options in the analysis toolpak. The only problem is that the output takes a bit to understand at first. If you do correlation between two sets of data, you’ll get a $2 \times 2$ table as output. The number you want for correlation is in the lower left-hand corner of the table, where “Column 1” down and “Column 2” across meet. The entry for “Column 1” both down and across is 1, as that’s the correlation between the data in Column 1 and itself. (Of course a data set will correlate perfectly with itself!) Likewise, the entry for “Column 2” both down and across is 1.

4 Linear Regression

If you just want to add the regression line to a scatterplot, click on the scatterplot. Then go to Chart→Add Trendline and select the “Linear” option. If you want the linear regression equation, then there’s a little more to do.

1. Make sure you do NOT have the scatterplot selected, go to Tools→Data Analysis, and select the “Regression” option.

2. Select the range of values of the RESPONSE variable to go in the “Input Y Range” box. (It seems a little backwards to me to have the response variable first, but that’s Excel.)

3. Then select the range of values of the EXPLANATORY variable to go in the “Input X Range” box.

4. Then check the “Output Range” option and select an empty cell for the output range.

5. Finally, click “OK”.

6. All the regression data should now be on the screen. There’s a lot of stuff here, only a little bit of which we actually need. For the regression line $\hat{y} = a + bx$, the value of $a$ is given by the first number in the row labeled “Intercept”, and the value of $b$ is given by the first number in the row labeled “X Variable”. (Incidentally, both the correlation $r$ and $r^2$ are also given here; they’re the two numbers “Multiple R” and “R Square,” respectively.)