

## TIPS FOR DOING STATISTICS IN EXCEL

Before you begin, make sure that you have the DATA ANALYSIS pack running on your machine. It comes with Excel. Here's how to check if you have it, and what to do if you don't.

Go to TOOLS on the main menu bar. If Data Analysis appears on that pull-down menu, you are ready to go. If it does not, select ADD INS from the same TOOLS pull-down menu. When that window opens up, there should be the opportunity to select Analysis Tool Pack. Once you do that, Data Analysis should appear on the TOOL menu. If Analysis Tool Pack is not listed under ADD INS, then you must get out your Excel installation disk and add it to your program.

### Set Up of Raw Data Files

1. Each row of data is a set of scores for one individual.
2. Each column represents a different variable and should be clearly labeled with a header.

### Sorting Data

Excel can do an excellent job of sorting your data for you. You should begin by saving your workbook under a new name. That way, if you made any errors in sorting, you can go back to your original data set and start again.

- First, highlight all your data. You can do this by clicking the uppermost left-hand corner of the worksheet. The entire screen will go grey.
- Then, under DATA on the menu bar, select SORT. You can sort by three variables at a time. If you have a header row, make sure to click that button on the bottom of the SORT box. Then, select the headers of the columns you wish to sort. Presto, it's done!

NOTE: If you fail to select all of your data, you may end up sorting only some of the columns and messing up your data.

### Generating Descriptive Statistics

You can generate different statistics in Excel by using the formula box at the top of the spreadsheet.

When you use Excel, I would like you to type in the formulas directly. Here are the formulas we will use frequently on exams:

= AVERAGE ( firstcell: lastcell)	= MODE ( firstcell: lastcell)
= MEDIAN ( firstcell: lastcell)	= MAX ( firstcell: lastcell)
= STDEV ( firstcell: lastcell) - <i>sample</i>	= MIN ( firstcell: lastcell)
= STDEVP (firstcell: lastcell) - <i>population</i>	= VAR ( firstcell: lastcell)

Note that:

- Every formula is preceded by “=”. Don’t forget that!
- The (firstcell:lastcell) is your way of identifying the data values that Excel should use in your calculation. Select a column of data to work with, and then identify the first and last cells in that column that are to be included in the analysis. For example, to find the average of scores in column A, from row 2 to row 86, you would type:

= AVERAGE (A2:A86)

**How do Do It:**

1. Select a portion of your spreadsheet, clearly separate from your data, as a place to record your descriptive statistics.
2. In this spot, create a “table” for recording your statistics. If you are going to be calculating average, median and mean, type these names as labels in your table.
3. Now, to fill in the table:
  1. Click on the location in your table where you want to record “average.”
  2. Now take your cursor to the formula bar and type in your formula:  
 = AVERAGE (A2:A84)
  3. Hit the red “  ” to the left of the formula bar.
  4. The average of the data from A2:A84 should now appear in your table.

**Using COUNT Statements**

You can ask Excel to count various things on your spreadsheet. Here are some of the COUNT formulas you might find useful:

COUNTA(firstcell:lastcell)

This formula will count all of the cells in the range that have data in them.

COUNTIF(firstcell:lastcell, “SCORE”)

This formula will count the number of times that the value (SCORE) appears in the cells in your range.

You may wish to relate the number of cells with a certain value, to the number of cells that have data in them. For example, if you were taking a proportion. To do this, simply use the formula to divide two quantities:

= firstcell/lastcell

### **Repeating Formulas**

Once you have typed a formula into a cell, you can ask Excel to repeat that formula down the entire column. For example, if you have data that looks like this:

	A	B	C
1	100	50	=(B1/A1)
2	210	65	
3	185	48	

If you trying to calculate the ratio of B/A for each row, type the formula shown into cell C1. Then hit return. You will see the value (0.50 in this case). If you click on that cell, you'll see a small black box in the bottom right-hand corner of the box. Drag the box down as far as you need to, and it will automatically apply that formula to the entire column.

### **Generating Frequency Tables**

You can also ask Excel to count scores for you so that you can create a frequency distribution. Again, you'll use the formula bar to do this. Use this formula:

= COUNTIF (firstcell:lastcell, "SCORE")

Here, the term "SCORE" is the score that you are interested in counting. For example, say you want to count how many times the score 39 shows up in cells A2:A84. You would enter the following formula in the formula box:

= COUNTIF (A2:A84, "39")

If you want to create a grouped frequency distribution, realize that you can also use the following symbols in your formula:

= COUNTIF (A2:A84, “ < 39”)	to count all scores less than 39
= COUNTIF (A2:A84, “ < = 39”)	to count all scores less than or equal to 39
= COUNTIF (A2:A84, “ > 39”)	to count all scores greater than 39
= COUNTIF (A2:A84, “ > = 39”)	to count all scores greater than or equal to 39

I won't tell you exactly how to use this to your advantage – that's for you to figure out.

### **How to Do It:**

1. As before, create a labeled table to record your frequency information.
2. Click on the cell where you want to enter your first frequency count.
3. Move the cursor to the formula bar. Type in the formula you are using.
4. Hit the red “  ” to the left of the formula bar.
5. The frequency count will now show up in your table.

### **Generating Graphs**

Excel can produce a number of different kinds of graphs for you. Once you have counted up the number of scores you have in given categories, you can ask Excel to create a histogram or bar graph showing this information.

1. Create a data table on the Excel spreadsheet. Name the columns and the rows. Enter the count data into the cells.
2. Then select the graph wizard from the menu, or, if it is not showing, go to INSERT, CHART...
3. Select the type of graph you want. Hit the NEXT button and Excel will walk you through the rest.
4. It is easiest to enter titles and make changes to the graph now rather than trying to edit it later. Some particularly nice features are (1) changing the scale of the axes if needed, (2) adding labels to the axes, and (3) putting a title on the graph.

### **Correlation**

Before you begin, be sure that the variables you want to correlate are all in adjacent columns. This is necessary.

1. Select TOOLS, Data Analysis, Correlation...
2. Now, highlight the adjacent columns (two or more) that you want to correlate. It's a good idea to highlight the label row as well. This information will show up in the INPUT window.
3. Click on the Labels box to indicate that you've selected the labels.
4. When you hit go, you will get a correlation matrix.

### Scatterplot

You can get Excel to print a scatterplot for two variables you are correlating. The variable in the left column will be placed on the x-axis and the variable in the right column will be placed on the y-axis.

1. Select INSERT, Chart, XY (scatter)...click the next button
2. Now, highlight the two adjacent columns of data that you want to graph.
3. You will have an opportunity to add titles, etc, before you finish the graph.

### T-Test

You can use Excel to run a one-sample or two-sample t-test. Here's how:

#### One-Sample

1. Enter the data into a single column.
2. Then, generate formulas to find the following necessary pieces for the test:
  - a. sample mean.
  - b. sample standard deviation.
  - c. sample size.
  - d. standard error.
3. Put these pieces into the formula for a single-sample t-test and generate the t value.
4. Check the t-value against your textbook to determine if the result is significant or not.

#### Two-Sample

1. Put your data into two adjacent columns, one for each sample. It is good to include a header.
2. Select TOOLS, Data Analysis, and then t-test...(select which kind you want)
  - Paired two-sample for means** (dependent t-test)
  - Two sample assuming equal variances** (independent t-test)
  - Two sample assuming unequal variances** (independent t-test)

**NOTE: to determine whether to use the equal or unequal variance test, do an F-max test.**  
You can do this in Excel, or by hand following your class notes.

3. Highlight the two adjacent columns that you want to test with a t-test, or type in the names of these ranges into the boxes provided. Check the box for labels if you have included the headers.
4. Enter the hypothesized mean difference (usually zero).
5. When you hit go, you will get the results of your t-test.

### **ANOVA**

You can use Excel to run a single-factor ANOVA. Here's how:

1. Put your data into adjacent columns.
2. Select TOOLS, Data Analysis, and then Anova: Single factor.
3. Highlight the three or more adjacent columns that you want to test, or type in the names of these ranges into the boxes provided. Check the box for labels if you have included the headers.
4. Enter the hypothesized mean difference (usually zero).
5. When you hit go, you will get the results of your ANOVA.