

## Lab 1 – Beginning Descriptive Statistics

If you are experienced with Excel you will not need to follow the directions in this lab too closely, but you still need to answer the questions!

### **Step 1:** Load the data sets into Excel.

You will need to get Excel to convert the text to columns, which it sometimes does all on its own. Most likely, Excel will pop up a box that contains the options for converting text to columns, but if that doesn't happen you can manually open the box from the data menu by choosing "text to columns". In this dialog box choose "delimited width" and click "finish". To remind us exactly what we are looking at lets add a row (Insert-"row") or two at the top as header information. The data is arranged by month, so columns D-O represent months January – December. For ease you should probably copy each dataset for the lab onto a sheet in the same Excel workbook to keep everything together. You can always add more sheets (Insert-"sheet") to the workbook if you need more space.

### **Step 2:** Compute the monthly means

Choose an empty row below the data and label the row in the first column "Average" or "Mean" or something like that. Select the box for this row in the January column and type your formula. You probably remember how to calculate the mean of a group of numbers, but to get accustomed to writing formulas in Excel try computing the mean both by writing a formula using only the sum command and one using the built in function called "average". Excel likes parenthesis, so you can always add them for clarification if you are unsure how Excel will calculate something. Ask me how if you have never written a formula in Excel. After typing the formula in one column you can drag the square in the lower left hand corner to apply the formula across a row or column. Notice that the cell references will also change, so that if you drag across a row your formula in the new column will use data in the same column for its calculation.

1. Examine the monthly average temperatures at Central Park and LaGuardia Airport for 1948 to 2003, both in (or very near) New York City. Which month averages the warmest at Central Park? at LaGuardia? Which month averages the coldest at Central Park? at LaGuardia?

### **Step 3:** Estimate the spread of data

How can we tell how much year-to-year variation in average temperature there is for a given month at Central Park or LaGuardia?

2. Invent two or three methods to measure this variation and describe them. Using your methods, find out whether there is more variation in January than in July at Central Park, and see if the same result applies to LaGuardia. How do your methods compare with one another?

### **Step 4:** Look at the data

3. When Central Park has a below average or above-average January temperature, does LaGuardia also tend to do likewise? What about for July?

4. Between Central Park and LaGuardia, which station tends to be colder in January? Which station tends to be colder in July? What explanation might you offer from the answers to those two questions, taken together?

### **Step 5:** Graph

5. Make a graph of the average temperature in Central Park for January for 1948 to 2003. Make a second graph for the same, except for July. To do this, highlight the y values (the “January” column in this case) including the label row (so 57 rows) and click the create graph button or select Insert – “chart”. In the dialog box that pops up choose “XY (scatter)” and then the picture with dots and smoothed lines. In the next window click the tab labeled “series” and then click the arrow box in the row labeled “X values”. This should minimize the chart wizard and you can drag your cursor over the x values (the years column, including the label) to select them. When you have selected the x values, press “enter” and the window will pop back up. Click “next”. Now you can type in an X label or modify the Y label for your graph. When you are done click “next”. In the last window you can choose to put the graph in the sheet with the data or as its own sheet in the workbook. I like to make it a separate sheet with a distinguishable name such as “Central Park Jan Temp”.

**Step 6: Mean Temperatures**

Calculate the annual mean temperature for each year from 1948 to 2003 in an empty column to the right of the data. We will do it using equal weighting for all months even though they have slightly different numbers of days.

6. Make a graph of the annual mean temperature.

**Step 7: Anomalies**

In climate studies we are usually most interested in variations about some mean value or mean pattern. Deviations from the prescribed mean are called anomalies and we usually calculate them by setting the mean to zero. To do this we will write a function that takes each monthly temperature value and subtracts the mean of our entire time period. Put your new variables in empty columns to the right of the data. Calculate the departure from normal (anomaly) for each year's January with respect to the average. Then do the same for every other month, and derive an anomaly data set.

7. In looking at the results of the whole data set, is there a tendency for gradual warming from the later 1940s to the present?

8. Is there a tendency for anomalies to last from one month to the following month within a given year? (That is, if March 1993 was warmer than average, would April 1993 also tend to be above average?)