

Weather Lab Procedure and Writeup Specs

On the lab computer you're using, there should be two Excel spreadsheets on the Desktop. One is called AlbanySince1795.xls and has average monthly temperatures in Albany, NY, for most months since 1795. The data for each year are on a single row with columns for the twelve months and a column for the average temperature for the year. The other spreadsheet is called Albany30yrs.xls and gives the monthly averages for 1976 to 2005 in one long column.

These data are based on actual observations. Note that some of the early years are missing, and a few years are missing data for one or more months. In the latter case "99.99" is entered. A brief historical digression: Simeon DeWitt made the 18th century observations; he was also one of the founders of Ithaca, NY and the Surveyor General of New York for many years, responsible for some of the first maps of the region. In addition, he had a role in promoting the construction of the Erie Canal. The data was downloaded from the United States Historical Climatology Network (USHCN); the URL is

<http://www.ncdc.noaa.gov/oa/climate/research/ushcn/ushcn.html>

It took some additional work to extract the Albany data, and weed out extraneous information, in order to produce these Excel worksheets.

Your goal should be to analyze the data in as many different ways as you can think of, including calculating numbers and drawing charts. Contrary to the cliché, it is not true that "the data speak for themselves." You will have to speak (or rather write) for them.

Respond thoughtfully to the following questions. Whenever possible, try to support claims both with evidence from a graph and from calculations of a numerical statistic.

1. How hot was the hottest month ever recorded in Albany? Which month in which year? (replace any "99.99" with a blank to avoid having it show up as the maximum) Similarly, how cold was the coldest month, and when did it happen? This data does not include daily temperatures, but what would you guess for temperature on the hottest day ever recorded in Albany, and on the coldest day?

2. Pick two months, in two different seasons, and make line charts of the temperatures. Is there a clearly visible upward or downward trend? Confirm this by having Excel draw a regression line, and noting the size (and units of measurement) of the slope. What are the minimum and maximum temperatures observed in each month? What are the first and third quartiles? Half of the temperatures will lie between these two numbers. If you ONLY look at the most recent 40 years, what proportion of the temperatures lie above the third quartile which you just measured for ALL the years? If that proportion is a lot bigger than 25%, it would be evidence for a warming trend.

The standard deviation of the temperatures is another way to measure how "spread out" the numbers are about the center. When you look at the chart for one of your two months, do temperatures appear to fluctuate more in one section of the graph compared to another (of the same duration)? Try to back this up by calculating standard deviations for temperatures in the two time periods.

Finally, make histograms for each of your two months, following the Basic Instructions for Statistics handout. They will be a little "bumpy", but are they basically bell-shaped?

Are there any gaps in the histogram, so that a few observations at the left or right ends are separated from the rest of the observations? Also mark the median and the first and third quartiles on the x -axis of the histogram (remember that a number below the x -axis labels the first tick mark to the right of that number; do your best to estimate when a number lies between two tick marks). 25% of the total area of the bars should lie to the left of the first quartile, 50% of the total area should lie to the left of the median, and 75% of the area should lie to the left of the third quartile.

3. These are more speculative questions: the measuring station in Albany has not always been in exactly the same location, though since 1938 it has been located at the airport. Is this “a problem” (why or why not)? If so, how serious is it? Are there ways you could think of to compensate for the need to move a weather station from one location to another?

4. Using the single-column worksheet, make a chart of month-by-month average temperatures in Albany over the past 30 years. Is there a cyclical pattern? Is there an upward trend? Are there any “unusual” months which show up on the chart? If so, when and what makes them “unusual”? For extra credit, create a function involving the sine function which does a reasonable job of approximating the values on your chart, and demonstrate this by plotting the data and the graph of your function in the same figure.

5. Write down a question about Albany weather that you would be interested in having an answer to, being specific both about what sort of data would be needed, and how you would need to analyze it.

EC1. For extra credit, you can figure out a way to plot a “5 year moving average” for one of your months you used in #2. For example, this means that the height of the plot for March 2003 should be the average of the temperatures in March 1999, March 2000, March 2001, March 2002, and March 2003. This chart should be compared to the chart you made in #2.

EC2. For extra credit, make a chart of the annual average temperatures and compare it to the two monthly charts especially with respect to variability. You should use at least two different approaches to measuring variability. Is there some general principle about average measurements, for which this is a special case?

Hints: relevant Excel commands include:

AVERAGE, CORREL, INTERCEPT, MAX, MIN, MEDIAN, QUARTILE, STDEV, SLOPE

To figure out how to use these functions, in the Insert menu, select Function, then select the Statistical category. Scroll to the command you want, and a quick description appears below.

See also the ExcelForStat handout, and the StatConcepts handout.