Lab 2 – Probability

This week I provided an excel workbook with the data already loaded in sheets. The data is also posted in the class files as text files if you wish to use it in a different form or whatever. Download the workbook from courseworks and get started! Since there will be a lot of graphs in this lab you may want to include them in the worksheet they are associated with instead of as a separate sheet like I suggested last week.

Project 1: Probability

We will use the Central Park temperature in January to find the probability of temperatures below zero as discussed in class. This is the same dataset as we used last week, but I have deleted the months other than January for simplicity.

Categorize the temperature as being above or below zero. To do this you can use the excel function called "if". The syntax is IF(logical_test, value_if_true, value_if_false). The logical test is some statement using logical operators like <, >, =, >= etc. In our case we want to know if data in our temperature column is less than zero so we will type "=IF(C2<0,0,1)" in cell D2. Drag the formula down so that it applies to the entire time series (56 years). Notice that we have set the value of column D to zero for all days less than zero and 1 for all days greater than zero.

We want to estimate the probability of above zero and below zero, so we should count up how many days are in each category. We can do this by summing column D. This gives us the total number of days above zero. To get the days below zero you just need to subtract the number of above freezing days from the total, but these are not the probability values, because we know that probabilities lie on the interval between zero and one. To calculate probability we need to divide the number of occurrences of the event we are interested in by the total number of years in our sample.

For the next twenty years, estimate the probability distribution of the years with temperature above zero using binomal theory we talked about in lecture. Recall that in general, when each trial has probability *p* of success, the probability of exactly *s* successes in *n* trials is:

$$P[s] = \binom{n}{s} p^s (1-p)^{n-s}$$

Where the binomial coefficient is defined by:

$$\binom{n}{s} = \frac{n!}{s!(n-s)!}$$

To calculate the probability distribution for 0-20 years above zero we need to calculate the probability of having 0 years above zero, then 1 year above zero, 2 years above zero, etc., assume that the probability for the next twenty years is the same as the past 56 years. When you have a list of probabilities corresponding to the number of years above zero make graph to visually show the distribution. I recommend a column graph.

Project 2: Assessing the effect of increasing CO_2 in a Climate Model

We will be using monthly mean surface temperature for NY from two climate model runs, one without CO_2 increase, one with 1% per year CO_2 increase.

Compute mean, max, min, and standard deviation from each dataset for each month like we did last week. Make a plot of the mean temperature, maximum temperature, minimum temp, standard deviation, as a function of month for each of them. I suggest making one plot for each parameter (i.e. mean) with both cases (i.e. control vs. experiment) displayed on each plot. Observe the difference between the control and the increasing CO_2 run to answer the following questions:

- What is the general difference between the two cases?
- Which month shows the largest differences?
- Is the seasonal cycle the same or different in the control versus CO_2?
- What is the biggest problem of global warming based on the data you have (for example, maximum temperature is too high, minimum temperature too low, etc.)?
- What can you say about the variance of the two cases?

Make a plot for each month of the monthly mean temperature as a function of years for both the control and the CO_2 data. What do you see in these plots? Is there a temperature trend in the control run? CO_2 run? If so, what is your estimate of the trend? **Note**: You don't need to include all 16 plots that you just made in your write-up, only those you feel help illustrate your argument.

Next we will make a probability distribution for January and July for the two cases. We have to lead Excel every step along the way for this calculation, so we will need to do diagnostics. Excel needs us to tell it the upper limit of each bin for the histogram, so we need to pick both the range and bin size. First calculate the max and min of our range of data, then pick a reasonable bin size corresponding to, say, 8 bins or so. You can play with exactly how big by doing this multiple times and seeing how the shape changes (it does change!). Type a series of numbers that correspond to the upper limit of each bin's range (ie if my first bin is all numbers less than 10 I would type 10 as my first number, and if the second bin is numbers between 10 and 11 I would type 11 as the second number). Ok, now in a column near by highlight a number of rows corresponding to how many bins you have (i.e. if you have 8 bins, highlight 8 rows). We will use Excel's "frequency" function to calculate the frequency of occurrences in each bin range *for all bins at once*. The function syntax is "frequency(data_range, bin_limits)" and **must be executed by pressing Command-Enter** (control-shift-enter on PC). Whew!

Now we have a list of frequencies for a set of bins, but we want probabilities! Lets calculate a probability density by dividing each frequency value by the total number of samples (the number of years in our data set). Now we can plot the probabilities in a "column" graph. Make sure to choose the bin limits as your x (category) values. If you use the same bins for both the Control and Increased CO_2 data sets you can plot them on the column graph together by adding a second series. Ask me how if you need help.

Answer the following:

- What is the probability of having January temperature above 6^{o}C for each case?
- What is the probability of having July temperature above 25^{o}C?

Project 4. Monthly mean precipitation for NY from the control and the CO2 run.

Follow the same step in 2, and answer the question for precipitation in the same way as those for temperature.

- For probability distribution, what difference is there between temperature and precipitation?