**Assignment 1 – Surface Weather Stations & Univariate Analysis**

**Due Friday February 1st at 5:00 p.m. by email (twford@siu.edu) or in person**

Surface weather stations were the original source of weather and climate information, and are still regarded as the gold standard of climate data. However, because of issues with station moves, sensor changes, and missing or erroneous measurements, station data can be difficult to work with. The purpose of this exercise is to gain experience downloading, processing, and analyzing climate information from surface weather stations.

You will download daily weather information from one surface weather station, process that data, statistically analyze the data, and create figures to display your results. For this assignment we are going to focus on changes in individual variables over time, and therefore utilize univariate statistical methods. Once you download, process, and analyze the data, you will write and turn in a typed report that summarizes your results.

**Tasks:**

1. Use the Iowa Environmental Mesonet database to download daily summary data from an ASOS station in Chicago, Illinois.
   1. Go to this webpage: <https://mesonet.agron.iastate.edu/request/daily.phtml>, select “Illinois ASOS” from the dropdown menu, and click “Switch Network”
   2. Download all daily data from the Chicago O’Hare station from January 1, 1948 to December 31, 2018.
   3. Many variables are measured at this station, we are only concerned with maximum temperature, minimum temperature, maximum dewpoint temperature, and minimum dewpoint temperature. You can delete the remaining variables.
2. Use the computer program of your choice to process the data in the following ways:
   1. Erroneous measurements are more insidious than missing values, as they take the form of the variable measured instead of some moniker (like “NaN” or “none”) as missing values do. Therefore, before you do any processing, you must account for erroneous measurements. Typically this is done using thresholds beyond which measurements are not physically possible. In this case, replace any values you deem erroneous with some sort of moniker that indicates the value is missing (like “NaN” or “none”). Do NOT replace with another number like -999, as this will affect your statistics
   2. Now that you have accounted for erroneous measurements, use daily maximum and minimum temperature to compute a daily mean temperature. Remember that you can only calculate mean temperature for a day if that day’s maximum AND minimum temperature are not missing. Do the same for dewpoint temperature, using maximum and minimum dewpoint to calculate a daily mean. Make sure you retain maximum, minimum, and mean temperature and dewpoint data because we will run statistics on each
   3. Because they are used for operational purposes first and foremost, ASOS stations report variables in standard units (i.e., temperature in °Fahrenheit). But, we are scientists and will use scientific units; so convert temperature and dewpoint temperature to °Celsius.
   4. For our statistical analysis we will be calculating annual averages and extremes, but we can only use data from a calendar year if the year does not have too much missing data. How much is too much missing data is a somewhat-subjective decision. In this case, we will use 10% as a threshold for too many missing temperature and dewpoint observations. Find the number of missing observations for each variable as a percent of the total number of observations in each year. So, for example, if there are 29 maximum temperature observations missing in the year 2009 that would represent roughly 8% of total 2009 maximum temperature observations (29/365). If there any years in which a variable is missing more than 10% of observations, replace all of this variable’s measurements in that year with your missing moniker. Make sure to only make that variable missing; so it is possible there are too many maximum temperature observations missing, but maximum dewpoint is fine.
   5. Calculate the annual mean of each of your variables, so you will have one mean value for each year in the time series, 1948-2018.
   6. Along with looking at changes in the mean of each variable, we also want to look at changes in extremes (hot and cold, humid and dry). We will use simple percentile thresholds for identifying extremes, in this case the 90th and 10th percentiles. For each variable, find the number of days in each year (1948-2018) that exceed the 90th percentile. Then for each variable, find the number of days in each year that fall below the 10th percentile.
3. Finally, we are at the point where we can start analyzing the data. The question we are interested in answering for this assignment is “to what extent have temperature and dewpoint changed over the last 70 years in Chicago”. This question applies to both the mean of temperature and dewpoint variables, as well as the extremes. There are many statistical methods, including those we have covered in lecture that can help us answer this question. Your job is to select a univariate statistical method that is appropriate for the data and the research question, and apply the method to answer the question. Understand that you may need a different statistical test to answer the question for the mean of the variables than you do for the extremes.

**Report**

You will turn in a typed report that summarizes the work you did for this assignment. The report should be written in a professional format, summarizing your work while addressing the following questions. Figures, charts, and tables are not required, but are usually helpful in summarizing your work.

1. How many daily observations from each variable were missing or erroneous, and how many years (if any) did you have to delete as part of your quality control?
2. What univariate statistical methods did you select to answer the analysis question and why? It is important you provide a sound argument for the statistical method you chose, including that the assumptions of that method were not violated by the data.
3. How have variable means changed over the last 70 years in Chicago? How have the frequencies of extremes changed over the last 70 years?