Reading Monthly Mean and Yearly Mean Data from US Historical Climate Network: Using Excel to Calculate Basic Statistics for Data Near Your Home.

Learning objectives:
Use Excel To:

- Import US Historical Climate Network data
- Calculate mean, variance, standard deviation, maximum, minimum, and trends estimates for temperature data at a station near your town.
- Estimate the statistical error in trend estimates
- Calculate temperature anomalies
- Use a running mean filter to smooth data
- Graph temperature data and anomalies and trend estimates.

In what follows we assume that we are interested in using data from Vancouver, WA. You will want to choose data from a station close to your home and will have to adapt the instructions below accordingly.

The site
http://cdiac.esd.ornl.gov/epubs/ndp019/ushcn_r3.html
has the following US Historical Climatic Network (USHCN) data by state

- Monthly Mean Maximum Temperature (degrees F)
- Monthly Mean Average Temperature (degrees F)
- Monthly Mean Minimum Temperature (degrees F)
- Total Precipitation (inches)

This makes it easy to find data for a station near you home town.

The monthly mean temperature data is found:
http://cdiac.esd.ornl.gov/epubs/ndp019/statemean.html

Washington State data is found http://cdiac.esd.ornl.gov/r3d/ushcn/statemean.html#WA And then clicking on Vancouver Washington (my town) gives you a file of all state locations with the pointer at Vancouver. If you select (in this case) the rows for 1891 to 1994 and **Edit-copy** and then **Edit-paste** into Excel, you get all the data set in column A so you must use the **Data-Text to Columns command** to get the data into separate columns. More current data can be found but you must import and read the data files using a program written in something like Fortran which is beyond our scope here.

The format of the temperature data is as follows:

Station number, year, Jan, Feb, Mar, Apr, May, Jun, Jul, Aug, Sep, Oct, Nov, Dec, Winter (Dec, Jan, Feb), Spring (Mar, Apr, May), Summer (Jun, Jul, Aug), Fall (Sep, Oct, Nov), and the annual average temperature (Jan-Dec). All values are in degrees F and -99.99 indicates missing data.

Using Excel for basics statistics.

Since the first 4 years 1891-1894 having missing data for Vancouver we delete those rows to make the rest of this analysis easier. Do this for your city also, but try to find a USHCN location next to your city that has approximately 80 to 100 years of observations.

Insert at least one row at the top of your worksheet and appropriately label the top of each column, Station, year, Jan , Feb, ….. Annual.

For this description assume: 1) that there are 100 years of data 1895 through 1994 and 2) this data is in columns A through S and in rows 2 through 101; row 1 has the column labels.

In cell C102 (jan column) enter: **=average(c2:c101)** the Jan avg 1895-1994 temp.
In cell C103 enter: **=var (c2:c101)** The variance of the Jan temp
In cell C104 enter: **=stdev (c2:c101)** The standard deviation Jan of the temp
In cell C105 enter: **=max (c2:c101)** The maximum Jan temp for those years
In cell C106 enter: **=min (c2:c101)** The minimum Jan temp for those years

Select Cells C102 through S106 (5 rows and 17 columns) and **Edit-Fill-Right.** This calculates the mean, variance … for all months, all seasons, and annual average conditions. Label cells B102 to B106 . Select cell C2 to S106 and Use **Format-Cell-number** and set at 2 decimal places to format all values to two decimal place precession.

Questions.

From what city and state is your data?

For what years do you have data?

During what month is the temperature Maximum? Knowing that summer solstice is Around June 21, what is the approximate the in months between maximum solar input and maximum temperature?

During what month is the temperature minimum? Knowing that Winter solstice is December 21, what is the lag in months between minimum solar input and minimum temperature?

During what times of year (estimate dates) is the temperature typically close to the annual average temperature?

Which month has the greatest temperature variability?

Which month has the smallest temperature variability?

Does this make sense in terms of when you would expect the atmosphere to be most turbulent? Explain in a complete sentence or two and discuss the differences observed between these two months on the graph that you create below.

Graph: On the same set of axes plot the temperature for the largest variability month and the smallest variability month on the y-axis vs. Year on the x-axis. Remember, to select columns for graphing that are not adjacent you must hold down the Crtl key. Make the graph look nice with labels and formatting, and print it out.

Calculating anomalies.

Select all cells A1 through S106. **Edit-Copy** and then click on Cell A108 and **Edit Paste**. This should put a copy of everything in cells A108 to S213. The data itself is in cell A108 through S208. In cell
c109 enter: =c2-c$103. This take the temperature and subtracts the average from it. This is the anomaly. Often anomalies are calculate from a specific average period like 1950 to 1990. Here we use the full 1895 to 1994 average. The problem with this is that other stations will not have that many years of data so if we were to compare stations a shorter averaging period would be best. This is not an issue here.

Select cells C109 through S109 and use **Edit-Fill -Right** and then select C109 through S208 and use **Edit –Fill-Down.**

Questions.

Compare the statistics calculated for the anomalies with those calculated for the actual temperatures. For example:

What are the average values of the Anomalies?

How do the anomaly variance and standard deviations compare with the temperature variance and standard deviation?

How does the range (Max-Min) in anomalies compare with the range in Temperatures?

Temperature trends?

Select column T111; this is a new column to the right of the annual average and adjacent to the third row of anomaly values. (the actual cell will depend on how much data you have for your city).
Enter T111**: =average(S109:S113)** This is a 5-year average of the annual average temperatures. Copy this formula down to three rows from the bottom of the actual data. For the Vancouver case this is to cell T206. That is select T111 to T206 and **Edit-Fill-Down.**

Using Excel for basics statistics.

Graph: select B109toB208 (x-axis) and S109toS208 (y-axis) and use the chart wizard to plot the average annual temperature vs. year. Label your graph and format it to make it look nice. Now click on the data points of the graph so that the graph data is selected. The formula bar shows you what you've selected so if you make a mistake click somewhere in the graph and then try clicking on the data point again. Go to **Chart-Add Trendline-Linear** select **options tab** and check **Display equation of Chart** and **display R squared on chart.** After clicking okay the linear fit to the data and R-squared values are set on the chart. Often they are hidden within the data points but you can move them and format them to make them look nice. It's nice to use T instead of y and Year instead of x to make things more clear.

Note: A trend value of 0.004 for the slope would correspond to a 0.04 °F increase in temperature per decade and a 0.4°F increase in temperature per century.

Now add the 5 year running mean curve to your graph by using the **Chart-Add Data command.** When asked for the range simply select cells T111 to T206 in your worksheet with the mouse by dragging (the easy way)(or type in Sheet1!T111:T206 if you are working in the worksheet named Sheet1). After entering okay click on the new data points. Since the T column has fewer data points than the first line the year values are not correct. In the formula bar edit the B values from

=SERIES(,Sheet1!$B$**109**:$B$**208**,Sheet1!$T$111:$T$206,3)
to
=SERIES(,Sheet1!$B$**111**:$B$**206**,Sheet1!$T$111:$T$206,3) so everything matches.

Clean up your graph , add a legend,label it properly using **Chart-source Data -series**, and print it out.

Questions: What is the temperature trend of your station. Over the past 100 years the global average trend in annual surface temperature is about 0.9 °F/century. How does your city compare with the global average?

The R-squared value is and estimate of how much the linear fit contributes to the total pattern (variance) contained within your data . A value of 1.0 means your data is a perfect straight line with no other wiggles and a value of 0.00 mean that there is no linear component to your data at all.
What is the R-squared value for your fit? Comment on it's meaning for your data.

The same sort of information **plus a liitle** more can be obtained using the **Linest function**.

Select cells U108 to V112 Actually any convenience 5 rows and two columns will do. Click on the paste function wizard ***fx*** and find the **Linest** function (a statistical function).
Know Ys **S109:S208,** Known xs **B109:B208**, Constant: **True**
Stats; **True**

Now for the tricky part. Hold down the shift and Ctrl keys simultaneously and then press enter. If you need to edit this you must select all 10 cells again, do your editing, and then Shift-ctrl-enter. If you get stuck in this the escape key can get you out.

For The Vancouver data we get in the 10 cells:

	0.00869	-16.897		Slope and intercept
	0.00407	7.915956		Standard deviation of slope and intercept
	0.044437	1.174993		R-squared, standard error of y estimate
	4.557318	98		F value, degrees of freedom
	6.291871	135.2996		Regression and residual sum of squares,

A search of Linest in the Excel help menu can give you more detail and one should consult a basic statistic book for more information regarding theory.

The point here is that **Linest** can give us an uncertainty estimate for our calculated trend. The trend(slope) divided by the standard error is distributed as a t-statistic. For degrees of freedom between 80 and 100, +/- two standard errors correspond to an approximate 95% confidence interval that the trend is different from zero. Using this estimate we would report out trend as our trend is 0.87°F/century +/- 0.80°F/century. This is just statistically significant at roughly 95% confidence level.

Repeat this sort of estimate for your station's data and report your findings here.

Does the trend in the first half of the data differ from that of the last half. Once Linest is typed in and entered (with Crtl-Shift-enter) it can also be modified very easily to give trend estimate for any segment of data. To modify the calculation you must select all 10 cells again, do your editing, and then Shift-ctrl-enter.
Try this with your data
For our example changing:
=LINEST(S109:S208,B109:B208,TRUE,TRUE)
to
=LINEST(S109:S*164*,B109:B*164*,TRUE,TRUE) gives trend for 1895 to 1950
to
=LINEST(S*164*:S208,B*164*:B208,TRUE,TRUE) gives trend for 1950 to 1994
to
=LINEST(*T111*:*T206*,B*111*:B*206*,TRUE,TRUE) gives trend for smoothed (5-year running mean) temperature between 1897 and 1992.

Calculate the trends in annual mean temperature for the first half of your data and the second half of your data included uncertainty estimates and comment on what you find.

Calculate the trend in smoothed (5 –year running mean) annual mean temperature. Does the trend for the smooth data differ significantly form that of the unfiltered data.

Using Excel for basics statistics.

Statistical references on the Web.
http://www.statsoftinc.com/textbook/stbasic.html Basic Statistics On-line
http://www.physics.csbsju.edu/stats/Index.html On-line Statistical Calculators
http://thesaurus.maths.org/dictionary/map/word/2794 t-Table for statistical significance
When using the Linest function in Excel to estimate trends, the slope and the standar error for the slope estimate are given. These fit a t-distribution with n-2 degrees of freedom. For example is you have 102 years of data, and linest gives 2.0 and 0.50 for slope and standard error then the 95 % confidence limits will be 2.0 +/- 0.5(1.98). The value 1.98 comes from the t-Table 0.025 (two tailed) with degrees of freedom =100. The Excel Help on Linest gives more information.
http://www.sjsu.edu/faculty/gerstman/StatPrimer/regression.pdf a PDF file on linear regression and error estimates