**GLOBAL WARMING PROJECT—data analysis**

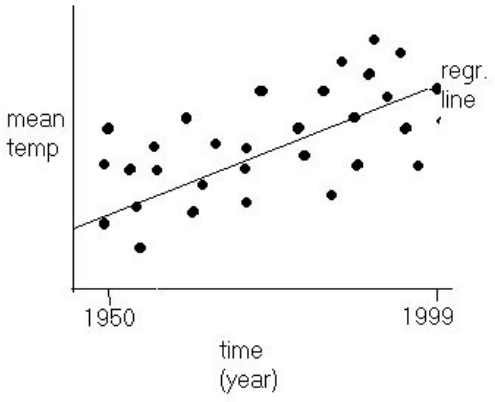
In this assignment, you are to choose three cities in the southwest, and evaluate whether or not global warming is occurring in this region. To accomplish this, Jim has provided you with data sets in Excel format which include the maximum and minimum temperatures for each month of the year from 1949 to 1999.

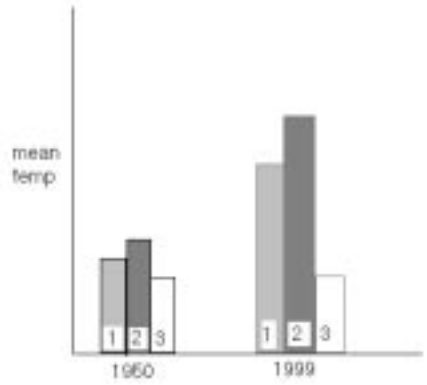
Before you begin, ask yourself which cities you will investigate, and why. For example, you may want to choose your towns based on a latitudinal distribution, you may want to choose high elevation towns, or Great Basin locations or a combination of any of the above. Justify your choice.

Now that you have chosen your cities, you can ask yourself any number of questions, but to get you started, here are some ideas: 1) has the temperature changed significantly over the past 50 years? 2) has there been a temperature change in the winters (or summers..) over the past 50 years? 3) if so, are the differences comparable between high and low elevation cities? Are miminum temperatures rising? 4) has the variation from mean temperature increased/decreased over the given time period? I'm sure you can think of many other interesting questions to evaluate. Understand that this is a question/hypothesis-driven paper: you are asking a question about the data set, and then analyzing the numbers to get an answer that you may or may not expect. Your conclusions about the issue will be based specifically on the results of your analysis.

What will your data look like? Here are some examples of what you can plot:

**EXAMPLE A**: **Regression analysis** will answer a question such as: has the yearly average minimum temperature for CityX changed over the past 50 years? Mathematically, you are asking whether the slope of the regression line is significantly different from 0. But you could analyze this another way… see Example B






**EXAMPLE B:** (above) Here you may wish to simply compare mean annual (or seasonal or whatever) temperatures in two different time periods. So, you would ask yourself: are temperatures in 1999 significantly different than in 1950? A simple statistical analysis here would involve **a T-test**, where you may wish to compare the temperatures of City1 in 1950, to those of 1999. A t-test is good for comparing two group (sample) means. Alternatively, the y-axis on the above graph could be the magnitude of variance in the temperature ie: the deviation from the mean. Recall, that one of the predictions of global climate change models is an increase in climate (temperature??) variation.

**The basic procedure for a t-test:**
First of all, the question you are asking is "are two sample means (groups of data) significantly different?" or more specifically "is the average winter temperature different in 1950-1960 than in 1989-1999 in Pocatello?" For our purposes, let's assume that the 'mean' is the same thing as the 'average', and the 'variance' (specifically the cumulative difference between observed temperatures and the mean temperature; see below for formula) between the two groups is equal.

1.Our hypothesis is that there is no difference in mean annual winter temperatures between those two time periods.
2. Calculate your $mean_1$ and $mean_2$ of your two data sets, and your variance. The **variance** is given as

$\mathbf{s^2 = \Sigma (X_i - X_{mean})^2 / n-1}$ where $X_I$ is your observed value and $X_{mean}$ is the mean, n is the sample size. For example, the temperature in november 1950 is 30, and the november mean from 1950 to 1960 is 40. Then $(30 - 40)^2$ is 100.

3. Use the following formula to calculate your T-statistic:

$$\mathbf{t_s = X_{mean1} - X_{mean2} / sqr.\ root\ (s^2_1 = s^2_2/n)}$$

4. Assume a P=0.05 level of probability. In other words, assume that there is a 5% chance that the data is significantly different due to chance. Or rather, be 95% confident, that the observed differences/results are not due to chance. Also, assume a two-tailed distribution for reasons I don't want to get into.
5. Assume $n_1 + n_2 - 2$ degrees of freedom.

6. Look up your critical t-statistic on the t-table (see other attachment). If the calculated statistic exceeds the critical statistic then the two groups of data (the means specifically) are significantly different. If not, assume that Mean1 is equal to Mean2 and they are no different.

Or very simply, calculate your T-Test in Excel. In the main menu, click 'insert', then 'function' then 'statistical', then T-test. Make sure your data is in columns.

**Regression Analysis**

The regression analysis deals with the investigation of the relationship between two (or more) variables, such as time and temperature. This can be easily done in most spreadsheet, graphing or statistics packages including Excel, but see me if you need formulae. In the Excel choose 'Insert' in the main menu, then 'function', then 'statistical', then RSQ. This will give you the 'r-square' value, which is a measure of how much of the variation in the data is explained by the regression line. So, the higher your $r^2$, the greater the proportion of variation that is explained by your model, or specifically, the regression line. If your $r^2$ value is low, then there may not be a trend, or, another analysis may better fit your data. Anything above 0.75 is considered pretty good. Most statistical packages will tell you via a P-value whether your slope is significantly different from 0 or not. Again, if you're working at a P=0.05 significance level, then you want a P-value lower than 0.05.

You can perform basic statistics using most spreadsheet and statistics packages such as Excel, SigmaStat, SAS, or InStat, and you are highly encouraged to work as a group on this! Not only will you all benefit, but it'll be easier to solve any problems that may arise during the statistical analysis. The following websites below give very detailed information on how to perform statistics in Excel.

Of course, you may consult any number of stats textbooks for clarification, but if you lack the time, check out some **useful statistics websites**:
http://www.statsoft.com/textbook/stathome.html ---excellent rudimentary intro to the topic with a cool search engine
http://panther.bsc.edu/~agannon/Bscstats.htm#t-test ---this is from a website for an upper division physiology course, which nicely describes t-tests, anova's and regression analysis, includes basic formulas, and gives a t-table values for a p of 0.05; but be careful! Their formula for variance is wrong! It should read $\Sigma (X_i - X_{mean})^2 /n-1$…..IMPORTANTLY this site gives you a step by step guide on performing basic statistical procedures in Excel, including graphing.

Lastly, feel free to stop by my lab in 229 South Biology if you need any help.

**A word on format, figures, analyses and style**

Follow the 5:3:15 rule as Jim indicated on the 5460 website! Papers significantly longer than 5 pages may be subject to a deduction in your grade. Don't say I didn't warn you! Please spellcheck and proofread! All figures should be completely legible including the data points, labels and legends. You will of course be evaluated on your analyses and interpretation, but not on the outcome of your data (ie: if you see a change or not). I simply want you to evaluate your results logically, and come to a reasonable conclusion about the trends or lack thereof.