Assignment for a Typical Statistics 101 course

Total Points: 40

Learning Goals

1. Students apply regression analysis and hypothesis testing to an example from climate change. (Psychomotor or Skills Domain)
2. Students master the ability to apply the appropriate technique (especially hypothesis testing) and articulate their reasons. (Cognitive Domain).
3. Students connect the mathematics to pressing real-world issues that did not exist just a decade or two ago (Affective Domain)

Learning Objective (practically speaking)

To get a visceral understanding of rising CO_2 levels by plotting the data yourself and performing hypothesis testing to show that the change is statistically significant at the 0.05 level.

Instructions

Read my article "Connecting Algebra to Real World Issues" located at
http://userhome.brooklyn.cuny.edu/skingan/Kingan_p.238-242.pdf

Read my instructional materials - a complete Statistical Literacy Textbook located at https://www.dropbox.com/sh/13isgj4fbwaavyh/H8jIzNvmiD

Chapter 3, 9 and 10 are particularly relevant to this assignment. It should take about two hours to review regression, z-test and t-test.

The accompanying Excel spreadsheet (on Google docs) gives the Mauna Loa temperature data from 1958. The first column is the year-month. The second is the number of days of data per month. There are missing days and some months it is not known how many days of data are available. The data has to be what we called "cleaned." If there are missing days concentrated either early or late in the month, the monthly mean is corrected to the middle of the month using the average seasonal cycle. -1 in the third column indicates no data on number of days. The third column gives the monthly mean CO_2 *mole fraction* determined from daily averages.
The mole fraction of CO_2 is the number of molecules of CO_2 in every one million molecules of dried air (water vapor removed). It is expressed as parts per million (ppm). Missing monthly means are denoted by -99.99. This is why we need another column called "interpolated." Compute for each month the average seasonal cycle in a 5-year window around each monthly value. In this way the seasonal cycle is allowed to change slowly over time. Then determine the "trend" value for each month by removing the seasonal cycle; this result is shown in the "trend" column. Trend values are linearly interpolated for missing months. The interpolated monthly mean is then the sum of the average seasonal cycle value and the trend value for the missing month. Interpolated values are computed in what sounds like a complicated way, but if we had access to the details it would be straightforward.

Part I: Learning the Techniques

1) **[5 Points]** Draw a time series graph for this data with year and month on the *x*-axis and interpolated mean CO_2 on the *y*-axis. Your graph should look like the Mauna Loa graph like Figure 1.

2) **[5 Points]** On the second sheet of the Excel file is the annual mean CO_2 levels obtained by taking the mean of all the monthly interpolated data. Notice that 1958 is missing since not all monthly mean is available. Draw a linear regression line and predict what the mean annual CO_2 level will be in 2015.

3) [5 Points] Consider the 53 years from 1959 to 2011 as the population and find the population mean and population standard deviation. Consider the last 21 years from 1991 to 2011 as a sample. Do a one-sample test of significance (z-test) to determine if this sample could have come from the population of 53 years. The null hypothesis is that the sample comes from the population and the alternative hypothesis is that it does not.

4) [5 Points] Let us consider as two samples the 32 years from 1959 – 1990 and the last 21 years from 1991 – 2011 both inclusive. Perform a two-sample test of significance (t-test -- since here we don't know the true population mean and standard deviation) to determine if these two samples could have come from the same population.

<u>Part II: Interpreting the answers</u>

1) [2 Points] In four or five lines explain how this assignment helped you connect statistics to the real-world?

2) [2 Points] Based on your graph in (1), what happened to CO_2 levels at Mauna Loa between 1958 and 2012?

3) [2 Points] Explain your prediction in (3) to a lay person in a paragraph.

4) [2 Points] Someone who doesn't understand the details might find "cleaning the data" suspicious. How would you explain that you didn't do anything wrong? (There is no incorrect answer here. Real-world data is never clean like textbook data. Give your best explanation. Just thinking about this will help you in your future efforts.)

5) [2 Points] What is the conclusion of your hypothesis testing? How will you explain it to a friend?

6) [2 Points] Is the z-test or t-test better for this data? (Or perhaps it doesn't matter?)

7) [8 Points] Apply the techniques in this assignment to another real-world problem. Write your real-world situation in a concise paragraph, provide the data, and solve the problem. (Keep it short and simple)

References:

1. Sandra Kingan (2013) Connecting algebra to real-world issues, in *Rethinking Mathematics*, 2nd Edition, Editors Eric Gutstein and Bob Peterson, Rethinking Schools, p. 238 - 242.

2. Pieter Tans, NOAA/ESRL (www.esrl.noaa.gov/gmd/ccgg) and Ralph Keeling, Scripps Institution of Oceanography (scrippsco2.ucsd.edu/).

3. Mauna Loa Carbon Dioxide level http://celebrating200years.noaa.gov/datasets/mauna/welcome.html#1958

Grading Rubric for Problems

1) There is one correct solution to the four problems in the assignment. In order to get full credit students must have the correct solution and all the in between steps must also be correct.

2) Partial credit will be generously given if the final answer is wrong, but the steps are more or less correct.

3) You may work in teams and help each other, but each student has to submit individual work and not just copy from a friend. It is easy to tell when a student just copies without understanding.

Grading Rubric for Interpretive Exercises

1) While there are clearly correct answers for the interpretative exercises, the verbal responses will be generously graded.

2) The breakdown for Problem 6 is 4 points for the problem and data and 4 points for solving it correctly.