Homework #1

Due in class on Wednesday January 30, 2019. *Instructions for "theoretical" questions: Answer all of the following questions. The theoretical problems should be neatly numbered, written out, and solved. Please do not turn in messy work. Working in small groups is allowed, but it is important that you make an effort to master the material and hand in your own work;Identical solutions will be considered as a violation of the Student Honor Code.* **Note that you are also required to turn in the computational portion of this assignment. It is on Canvas, under the name:**

`A4570SP19HW01-computational.ipynb.`

ON THE FRONT OF YOUR HOMEWORK CLEARLY PRINT THE FOLLOWING:

- Your name.

- Lecture number (either **APPM 4570** or **APPM 5570** or **STAT 4000** or **STAT 5000**).

- Homework number.

- **Points will be deducted if these instructions are not followed.**

Remember that writing style, clarity and completeness of explanations is always important. Justify your answers. (Be sure to place your homework in the correct pile, either "undergraduate" or "graduate".)

Theoretical Questions

1. For each of the following research scenarios below, write down (i) a description of the population, (ii) the variable of interest, and (iii) a description of a typical sample of size four.

   (a) All distances that might result when Chip shoots a hockey puck on a 200 foot hockey rink.

   (b) Page lengths of books to be published in 2020.

   (c) All possible earthquake-strength measurements (Richter scale) that might be recorded in California during 2019.

   (d) All possible yields (in grams) from a certain chemical reaction carried out in a laboratory.

   (e) The lengths of Philip Lindsay's Tweets.

2. Say your research deals with social networks. Your first step is to study the properties of the Facebook network of college students at CU-Boulder campus. The next step is to compare your findings to the national college student Facebook network.

   (a) What are the populations you are concerned with?

   (b) What is the relationship between these populations?

   (c) What are some of the characteristics of the networks you might consider? Pick three as an example.

   (d) If you had infinite time and resources, would you be able to measure these characteristics for every member of these populations?

   (e) Say you don't have infinite time and resources – how would you go about estimating those population characteristics?

3. Let $x_1, ..., x_n$ be a variable measured for units in a sample with sample variance given by $s_x^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ where $\bar{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$ is the mean of the *sample*. Let μ denote the mean of the *population* from which the sample came.

   (a) For what value of c is the quadratic polynomial $p(c) = \sum_{i=1}^{n}(x_i - c)^2$ minimized? (Use techniques of Calculus I to solve this.)

(b) Using the result from part (a), which of the two quantities $\sum_{i=1}^{n}(x_i - \bar{x})^2$ and $\sum_{i=1}^{n}(x_i - \mu)^2$ will be smaller than the other (assuming that $\bar{x} \neq \mu$)?

(c) Let $y_i = x_i - \bar{x}$, for $i = 1, ..., n$. How do the values of s_y^2 and s_y compare to s_x^2 and s_x? Prove your result. (More on this in the Computational section of the homework.)

(d) Let $z_i = (x_i - \bar{x})/s_x$ for $i = 1, ..., n$. What are s_z^2 and s_z? Prove your result. (More on this in the Computational section of the homework.)

4. (**APPM 5570/STAT 5000 Students Only**) Read the following article: `https://www.nature.com/news/2004/040927/full/news040927-9.html`. Can you think of any reasons why the authors' conclusions about 100 meter sprint times might be wrong?