Homework #2

Statistical Methods with R, Spring 2019

Due in class on Wednesday February 6, 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:**

`A4570SP19HW02-computational.ipynb.`

ON THE FRONT OF YOUR HOMEWORK CLEARLY PRINT THE FOLLOWING:

- Your full name.
- Your 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. In class, we learned about three measures of center: the (arithmetic) mean, the median, and the mode. This question will explore some other measures of center.

   (a) The *trimmed mean* is the arithmetic mean after the lowest $x\%$ and highest $x\%$ have been removed.

   i. Find the 25% trimmed mean of the set $\{1, 2, 3, 4, 5, 6, 7, 8\}$.

   ii. Describe one advantage and one disadvantage to using the trimmed mean.

   iii. Is there another name for the trimmed mean when $x = 50$?

   (b) The *geometric mean* of $x_1, ..., x_n$ (all $x_i > 0$) is defined as

   $$\bar{x}_g = \left(\prod_{i=1}^{n} x_i\right)^{1/n} = (x_1 \times x_2 \times ... \times x_n)^{1/n}.$$

   i. Suppose that you bought a classic car on January 1, 2015 for \$20,000. In the first year, it's value increased to \$22,000. In the second year, it's value increased to \$23,500. In the third year, it increased to \$28,000. Define the *growth rate* in a given year to be $x = 1 + c$ where c is the fractional increase in price in that year, *i.e.* in the first year $x_1 = 1 + c_1 = \$22,000/20,000$. What is the (geometric) average growth rate, $\overline{x}_g$, in the value of your car over the course of three years?

   ii. Why is the geometric mean better than the arithmetic mean in this case?

2. (a) Let a and b be constants and let $y_i = ax_i + b$ for $i = 1, ..., n$. Find $\bar{y}$ in terms of $\bar{x}$. What is the relationship between s_x^2 and s_y^2?

   (b) A sample of temperatures for initiating a certain chemical reaction yielded a sample average of 87.3 degrees Celsius and a sample standard deviation of 1.04. What are the sample average and standard deviation measured in Fahrenheit? (*Recall that* $F = 9/5C + 32$.)

3. Let $\bar{x}_n$ and s_n^2 denote the sample mean and variance for the sample $\{x_1, ..., x_n\}$ and let $\bar{x}_{n+1}$ and s_{n+1}^2 denote the sample mean and variance when an additional, new observation x_{n+1} is added to the sample.

   (a) Show that $\bar{x}_{n+1} = \bar{x}_n + \dfrac{x_{n+1} - \bar{x}_n}{n+1}$. (You will verify this in the computational portion of the homework.)

(b) Why might these results be useful?

(c) In part (a), what happens as $n \to \infty$? 

4. (**APPM 5570/STAT 5000 Students Only**) The *harmonic mean* is defined as $\bar{x}_h = \dfrac{n}{\sum_{i=1}^{n} \frac{1}{x_i}}$. It is often appropriate when looking for the average of rates.

   (a) Show that, when $n = 2$, $\bar{x}_h = \dfrac{2x_1x_2}{x_1 + x_2}$.

   (b) Suppose you drive your classic car from your house to a friend's house and back. On the way there, you drive 50 miles per hour; on the way back, you drive 65 miles per hour. Find $\bar{x}_h$ and $\bar{x}$.

   (c) Let d be the distance from your house to your friend's house (measured in miles), x be your speed on the way there and y be your speed on the way back (both measured in miles per hour). Show that your average speed on the trip $\left(\dfrac{\text{total dist. traveled}}{\text{total time taken}}\right)$, is best described by the harmonic mean. (*Hint:* You may have to use part (a)!