Homework 8

Due on Gradescope 11/22/2016 at 4:00PM (Before Lecture)

1. Lost order: Let $\boldsymbol{x}$ be a list of three numbers $\{1, 3, 5\}$. Let $\boldsymbol{y}$ be a list of another three numbers $\{2, 4, 6\}$, but suppose that you lost the order of these three numbers, so you're not sure whether $\boldsymbol{y}$ is $\{4, 2, 6\}$, or $\{6, 4, 2\}$, etc.
   (a) What is the arrangement of numbers in $\boldsymbol{y}$ that has maximum correlation with $\boldsymbol{x}$?
   (b) What is the arrangement of numbers in $\boldsymbol{y}$ that has minimum correlation with $\boldsymbol{x}$?

2. Variance of a sum: We have two lists of numbers $\boldsymbol{x}$ and $\boldsymbol{y}$ each with n elements. $\boldsymbol{x}$ has average $\bar{x}$ and standard deviation s_x; $\boldsymbol{y}$ has average $\bar{y}$ and standard deviation s_y. $\boldsymbol{x}$ and $\boldsymbol{y}$ have covariance $Cov(\boldsymbol{x}, \boldsymbol{y})$.

   We construct a new list $\boldsymbol{z}$ with n elements, whose entries are determined by $z_i = x_i + y_i$.

   (a) Write the standard deviation of z in terms of $\bar{x}$, $\bar{y}$, s_x, s_y, and $Cov(\boldsymbol{x}, \boldsymbol{y})$. (Hint: Use the computational forms of the variance and covariance).
   (b) How does correlation between the lists $\boldsymbol{x}$ and $\boldsymbol{y}$ affect the variance of the list $\boldsymbol{z}$?

3. Freebie from polling notebook: In the polling example we discussed, we had two lists with length N composed of 1's and 0's: $\boldsymbol{c}$ of candidate preferences for each person in the population, and $\boldsymbol{t}$ of voter turnout status. We said that the vote share for the candidate of interest was given by:

   $$\mu_v = \frac{\sum_{i=1}^{N} t_i c_i}{\sum_{i=1}^{N} c_i},$$

   but because it is difficult to measure t_i for each person, we wanted to figure out whether we could instead measure population candidate prefernece instead:

   $$\mu_c = \frac{1}{N} \sum_{i=1}^{N} c_i.$$

   Prove that if t_i and c_i have correlation 0, then $\mu_c = \mu_v$.