Problem Set 2

ECON 480 — Fall 2022

Due by Class Wednesday September 28

Theory and Concepts

1. In your own words, explain the difference between endogeneity and exogeneity.

2.

a. In your own words, explain what (sample) standard deviation *means.*

b. In your own words, explain how (sample) standard deviation *is calculated.* You may also write the formula, but it is not necessary.

Problems

For the remaining questions, you may use `R` to *verify*, but please calculate all sample statistics by hand and show all work.

3. Suppose you have a very small class of four students that all take a quiz. Their scores are reported as follows:

$$\{83, 92, 72, 81\}$$

   a. Calculate the median.
   b. Calculate the sample mean, $\bar{x}$.
   c. Calculate the sample standard deviation, s.
   d. Make or sketch a rough histogram of this data, with the size of each bin being 10 (i.e. 70's, 80's, 90's, 100's). You can draw this by hand or use `R`.

If you are using ggplot, you want to use + geom_histogram(breaks = seq(start,end,by)) and add another layer + scale_x_continuous(breaks=seq(start,end,by)) . The first layer creates bins in the histogram, and the second layer creates ticks on the x axis; both by creating a sequence starting at some starting value, some ending value, by a certain interval (e.g. by 2, or by 10).

Is this distribution roughly symmetric or skewed? What would we expect about the mean and the median?

   e. Suppose instead the person who got the 72 did not show up that day to class, and got a 0 instead. Recalculate the mean and median. What happened and why?

4. Suppose the probabilities of a visitor to Amazon's website buying 0, 1, or 2 books are 0.2, 0.4, and 0.4 respectively.

a. Calculate the *expected number* of books a visitor will purchase.
b. Calculate the *standard deviation* of book purchases.
c. **Bonus**: try doing this in R by making an initial dataframe of the data, and then making new columns to the "table" like we did in class.

5. Scores on the SAT (out of 1600) are approximately normally distributed with a mean of 500 and standard deviation of 100.

a. What is the probability of getting a score between a 400 and a 600?
b. What is the probability of getting a score between a 300 and a 700?
c. What is the probability of getting *at least* a 700?
d. What is the probability of getting *at most* a 700?
e. What is the probability of getting exactly a 500?

6. Redo problem 5 by using the `pnorm()` command in `R`.

Hint: This function has four arguments:

1. the value of the random variable
2. the mean of the distribution
3. the sd of the distribution
4. `lower.tail` `TRUE` or `FALSE`.