ECO239 Statistics I R HW 2

Due Date: January 10th 2020, at 12:15 sharp. NO LATE HOMEWORK ACCEPTED.

Submit a hard copy of the code/analysis/comments together with honor code and contribution sheet.

- Complete all tasks using R (R-console, R-studio, R-commander…your choice!)
- You can always conduct extra analyses you are interested in. (You may earn extra points!)

Task 1: Binomial Distribution

Reference: https://towardsdatascience.com/an-intuitive-real-life-example-of-a-binomial-distribution-and-how-to-simulate-it-in-r-d72367fbc0fa

http://htmlpreview.github.io/?https://github.com/andrewpbray/oiLabs-base-R/blob/master/probability/probability.html

a. Find real-life/real-world example for which Binomial Distribution is useful. Describe the situation including the definition of "Success" and probability of Success (P). (You may use one of the variables from your survey for 1st homework, but not obligatory. No "flipping a coin" or "throwing a die", please. Be creative!)

b. For the example described in part a, research and identify the probability of Success. If you use your survey data, you can calculate P directly. Explain in words.

c. Use *rbinom* function to simulate, (Set the number of trials as 10,000 (n). Decide "size" and use P from part b.) and plot the distribution. Comment on the distributional shape.

d. By calculating mean and standard deviation of the distribution derived for part c, decide what events are "unlikely" or "rare event". Use 2 standard deviation as the threshold.

e. Generate **three probability questions** regarding the binomial distribution considered for part 2. Use necessary R function to calculate the probability. (e.g. What is the probability that the number of heads obtained from 100 coin flips of biased coin (p=0.8) is between 85 and 90?).

Task 2: Normal Distribution

Refer to: http://htmlpreview.github.io/?https://github.com/andrewpbray/oiLabs-base-R/blob/master/normal_distribution/normal_distribution.html

Data Set: The data you've collected for 1st R homework. Use continuous variable for this task.

f. Generate Histograms for each variable and observe the distributional shape. Comment.

g. By using Q-Q plot, confirm if each variable is normally distributed. Comment on the distribution (normal, right/left skewed…).

h. Pick one variable (which is the closest to normal distribution),
calculate the probability P(x0<x<x1) by setting up **TWO** questions of your own.

i. Take natural log (ln) of each variable and repeat g and h procedures. Comment.