ECO239 Statistics I R HW 2

Due Date: January 11th 2019, at 13:00 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

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

Answer the following questions:

Using calc_streak, compute the streak lengths of sim_basket.

a. Describe the distribution of streak lengths. What is the typical streak length for this simulated independent shooter with a 45% shooting percentage? How long is the player's longest streak of baskets in 133 shots?

b. Change the shooting percentage to (YOUR CHOICE %) and repeat a.

c. How does Kobe Bryant's distribution of streak lengths compare to the distribution of streak lengths for the simulated shooter? Using this comparison, do you have evidence that the hot hand model fits Kobe's shooting patterns? Explain.

d. Plot the distribution of the streak lengths for Kobe and for the independent shooter separately.

e. Calculate the probability that the streak length is greater than 2 for Kobe, and for the independent shooter using R command.

f. Find the event which can be represented with binomial distribution in your daily life (no "flipping coins" or "throwing dice", please). Define the probability of "success" and the sample size by yourself, simulate, plot the distribution and calculate the probability of your choice.

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: Find your own data set containing at least 3 continuous variables.

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

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

i. Pick one variable (which is the closest to normal distribution), calculate the probability P(x0<x<x1) by setting up your own question

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