Data Science for Psych & Neuro 1/29/2018
Spring 2018

Homework 2: Data Types & Distributions

Complete this homework as a jupyter notebook titled “LASTNAME_Homework2.ipynb” posted on your GitHub account. *Answer all text questions using text cells and show all code & plots in code cells*. Turn in your assignment by emailing a link to the notebook to timothyv@andrew.cmu.edu by no later than 3pm on Thursday Feb. 8th, 2018.

Conceptual: Short answer questions. Be concise.
1. How does knowing data type (e.g., qualitative vs. quantitative) provide insights into the type of distribution a random variable comes from?

2. What exactly does it mean to say that a data point is a "random variable is normal with a mean of 100 and standard deviation of 15"?

Applied: Show your code & plots

The central limit theorem states that when independent random variables are added together, they sum to a normal distribution even if the original variables themselves are not normally distributed. For your homework test this assumption.

3. Using the *runif* function, run three experiments by simulating the outcomes of rolling a single, six-sided die. Show the distribution of each experiment. Show how the simulated means compare to the expected mean of a fair roll.
    a) Exp 1: 10 independent throws
    b) Exp 2: 1,000 independent throws
    c) Exp 3: 10,000 independent throws

4. Instead of rolling one die, run a set of experiments reporting the outcomes of rolling multiple dice at the same time.
    a) Exp 1: 10,000 throws, 1 die
    b) Exp 2: 10,000 throws, 2 dice
    c) Exp 3: 10,000 throws, 3 dice
    d) Exp 4; 10,000 throws, 6 dice.

Show the distribution of results for each experiment. Which of these experiments produces a distribution most similar to a normal normal distribution? Justify your conclusion using Q-Q plots.