DATA 605 - Final Project

Zach Alexander

5/17/2020

Instructions

Your final is due by the end of the last week of class. You should post your solutions to your GitHub account or RPubs. You are also expected to make a short presentation via YouTube and post that recording to the board. This project will show off your ability to understand the elements of the class.

Problem 1

Using R, generate a random variable X that has 10,000 random uniform numbers from 1 to N, where N can be any number of your choosing greater than or equal to 6. Then generate a random variable Y that has 10,000 random normal numbers with a mean of $\mu = \sigma = \frac{(N+1)}{2}$

Following the instructions from above, here is my R syntax to generate a random variable X:

```
set.seed(123)

N <- 10
X <- runif(10000, 1, N)
Y <- rnorm(10000, (N+1)/2, (N+1)/2)
```

From the above syntax, you can see that I've created a variable X with 10,000 random uniform numbers from 1 to N (N=10). Additionally, I've created a random variable Y that has 10,000 random normal numbers with a μ and σ of $\frac{N+1}{2}$.

Probability: *Calculate as a minimum the below probabilities a through c. Assume the small letter x is estimated as the median of the X variable, and the small letter y is estimated as the 1st quartile of the Y variable. Interpret the meaning of all probabilities.*

To work through this, I'll first have to calculate x (the median) and y (the first quartile). Additionally, I've saved X and Y in a dataframe can stored the total number of rows (10,000) in a variable:

```
x <- quantile(X, 0.50)
y <- quantile(Y, 0.25)
df <- data.frame(X = X, Y = Y)
total_rows <- nrow(df)
```

a) $P(X > x \mid X > y)$

We can use the following equation to find the probability:

$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$

Where, $A = P(X > x)$ and $B = P(X > y)$

```r
P_AandB <- nrow(df %>% filter(X > x, X > y)) / total_rows
P_B <- nrow(df %>% filter(X > y)) / total_rows

P_AgivenB <- P_AandB / P_B
P_AgivenB
```

```
## [1] 0.5509642
```