QUIZ 5 rubric

a) Write code to take a sample of size 1000 from Unif(2,4) and make a qqplot for this sample, but with the quantiles of Unif(0,1) on the x-axis.

x = runif(1000,2,4) # 1 point

n = length(x)

X = seq(.5/n, 1-.5/n, length=n) # seq(0,1,length=n) is sufficient, as well. # 0.5 point for creating the probabilities

Q = qunif(X, 0,1) # 1 point, 0.5 partial for wrong arguments

plot(Q, sort(x))

b) Write code to superimpose on the above qqplot a line whose intercept and slope capture the overall pattern of the qqplot. Explore only integer values of slope and intercept.

abline(2,2) # 1 point for the numbers (0.5 each)

c) What is the relationship between the intercept and slope found above with the parameters of the uniform distribution, Unif(a,b), from which the sample is taken? Give your answer in the form of "intercept = ..." and "slope = ..." where the "..." are expressions/functions of the parameters a and b. Hint: you have had a hw related to this, but I cannot tell you which hw.

# intercept = a

# slope = (b-a)

# The main purpose of this question is to make a connection between the

# theoretical calculation done in hw\_lect9\_4 and the simulation done here.

# 1 point (0.5 each)

d) Write code to take a sample of size 1000 from an exponential distribution with parameter 2, i.e., from Exp(2), and make a qqplot for this sample, but with the quantiles of Exp(1) on the x-axis.

Additionally, on this qqplot superimpose a line with the interecept and slope you reported in part b. Remember to check the posted soln, later.

x = rexp(1000,2) # 1 point

n = length(x)

X = seq(.5/n, 1-.5/n, length = n)

Q = qexp(X, 1) # 1 point for using qexp( , 1)

plot(Q, sort(x))

abline(0,1/2) # 1 point for the numbers (0.5 each)

e) Write code for computing the correlation between cor(xi,yi) for i = 1, 2, 3, 4. Hint: You don't need a for-loop for this.

cor(x1,y1) # 0.8164205

cor(x2,y2) # 0.8162365

cor(x3,y3) # 0.8162867

cor(x4,y4) # 0.8165214

# 1 point for using cor() (-0.5 if the arguments are wrong). 1 point even if student doesn’t get to the actual data, as long as it’s in the correct form.

f) Note that the correlation values are very similar. Write code for exploring the relationship between xi and yi, i = 1, 2, 3, 4. For i = 2, 3, 4, accompany your code with an explantion of why the correlation is the same as that between x1 and y1. Use an appropriate tool for exploration of relationships.

# The appropriate tool for exploring relationships is the scatterplot.

plot(x1,y1)

plot(x2,y2) # The relationship is a sufficiently nonlinear to reduce r from 1 to 0.816

plot(x3,y3) # The outlier is just in the "right" place to reduce r from 1 to 0.816

plot(x4,y4) # The outlier is just in the "right" place to increase r from 0 to 0.816

# 1 point for using plot, then for each part 0.5 for correct explanation. (in total 2.5 points)

g) Write code to

- read in the data,

- make a scatterplot appropriate for predicting brain weight from head size,

- superimpose the OLS line on the scatterplot.

dat = read.table("http://sites.stat.washington.edu/marzban/390/spring21/brainhead\_dat.txt", header=T) # 0.5 point for importing the dataset with “header = T”

x = dat[,3]

y = dat[,4]

plot(x,y) # "hard" variable on y-axis, "easy" variable on x-axis. # 1 point for the correct scatterplot (-0.5 if x and y are swapped)

lm.1 = lm(y ~ x) # 1 point for using lm() correctly

abline(lm.1) # 0.5 point

h) Suppose your head size is 3760 cm^3. What is your brain weight?

lm.1 # y = 325.5734 + 0.2634 \* x

25.5734 + 0.2634 \* 3760 # 1015.957 grams # 1 point for correctly reporting this

# total: 15 points