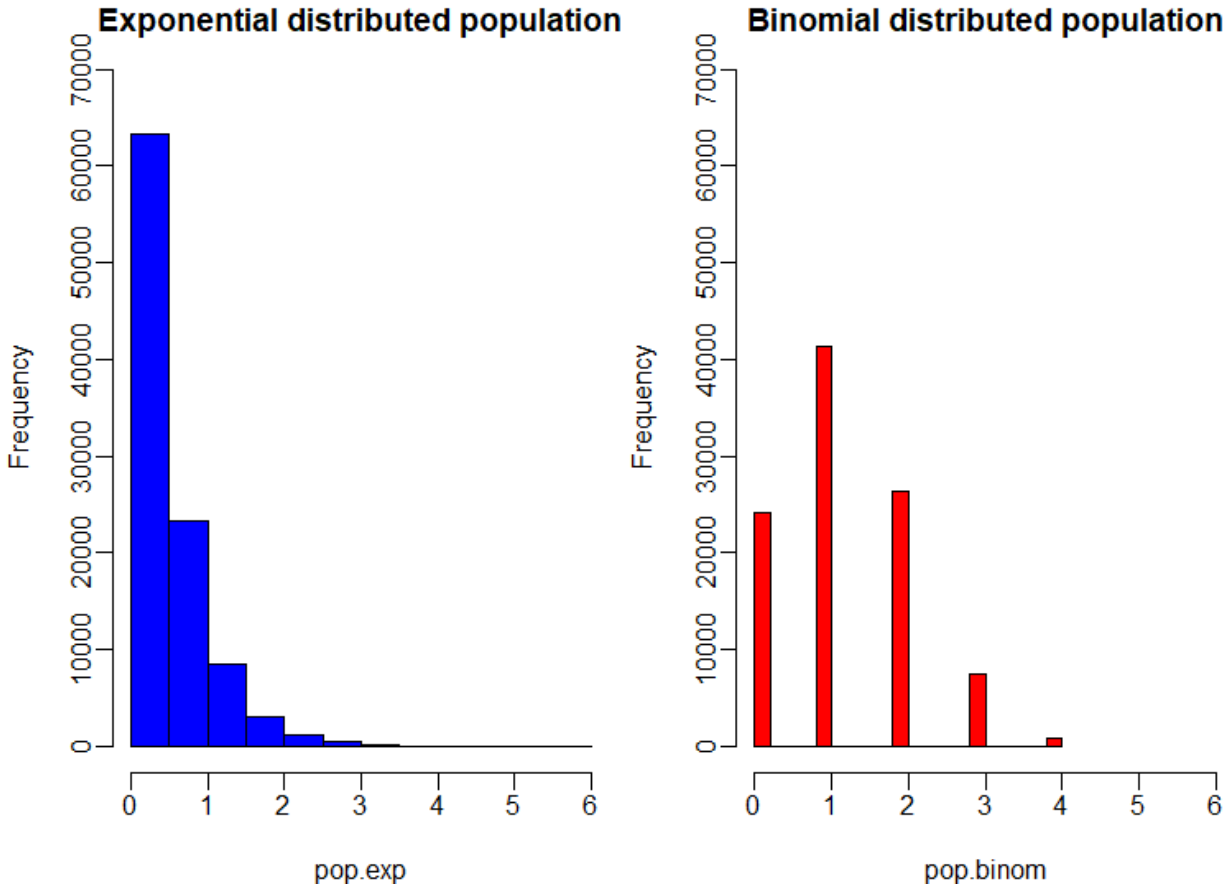
Name: Steve G. Mwangi
Date: November 17th, 2019
Assignment: TMATH 390 R Lab 7 Document.

Objective: Using simulated data to explore sampling distributions

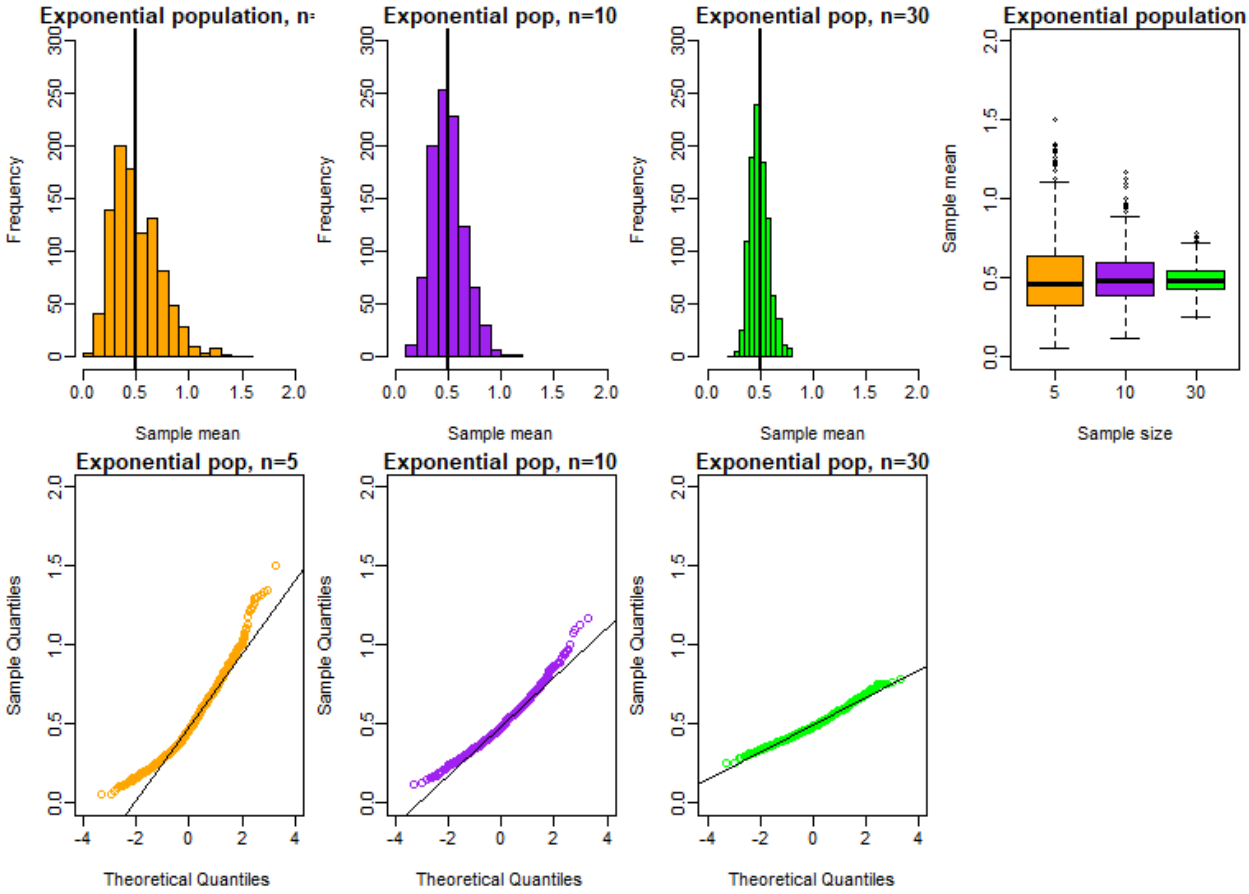
C1 (3). Submit the R script you used to complete this computer lab assignment.

Submitted

C2 (1) Provide the graph showing the distribution of each population.



C3 (2) Provide the graph generated by the code above for the distribution of the sample mean for the exponential distribution.



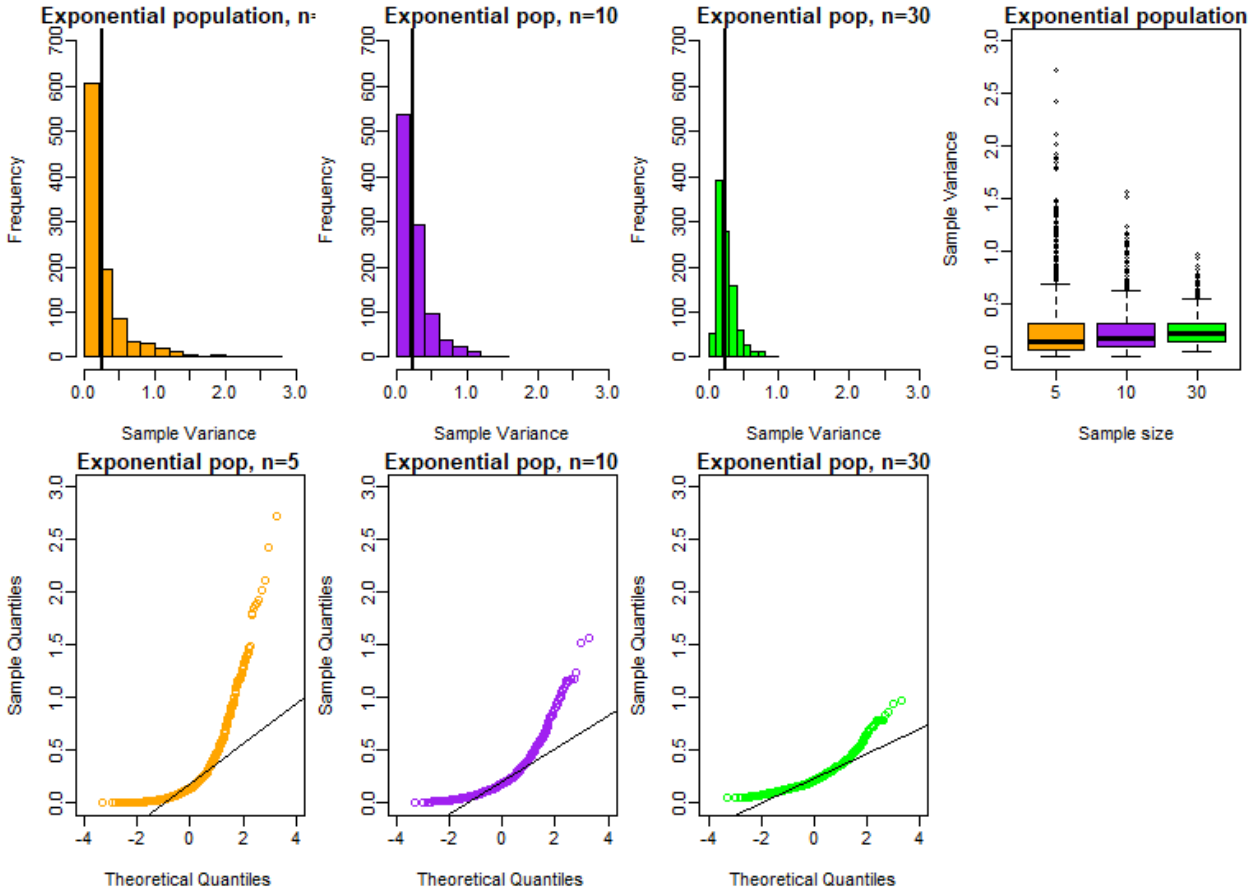
C4 (2) Describe how the sampling distribution of the sample mean for the exponential distribution changes with increasing sample size, including the shape, variability, and center of the distribution.

For the histograms, all the sample means for the samples of size 5, 10 and 30 are centered about the population exponential mean, of value 0.4991. The shape of the histograms is bell-shaped, similar to the normal population they are drawn from. Most variability is seen in the spread of the small sized sample, size = 5. At size 30, the spread is narrow, and centered about the mean.

All the qqplots show distributions that are right skewed. This makes sense since both the start and the end of the sample sizes are above the population qqline, hence they have values that are higher than those in the population, at the beginning and end of the Qq plots. Some values in the population are missing from the samples hence the population's qqline lying under the scatter plots at the beginning and at the end.

The boxplots confirm, especially in the n = 5 and 10 plots, that the data is right skewed, since the plots are longer on the upper side. All boxplots medians are about the same level, confirming the mimic the population's median. First boxplot, n =5 has greater variability as seen in the larger interquartile range compared to the rest.

C5 (2) Provide the graph generated by the code above for the distribution of the sample variance.

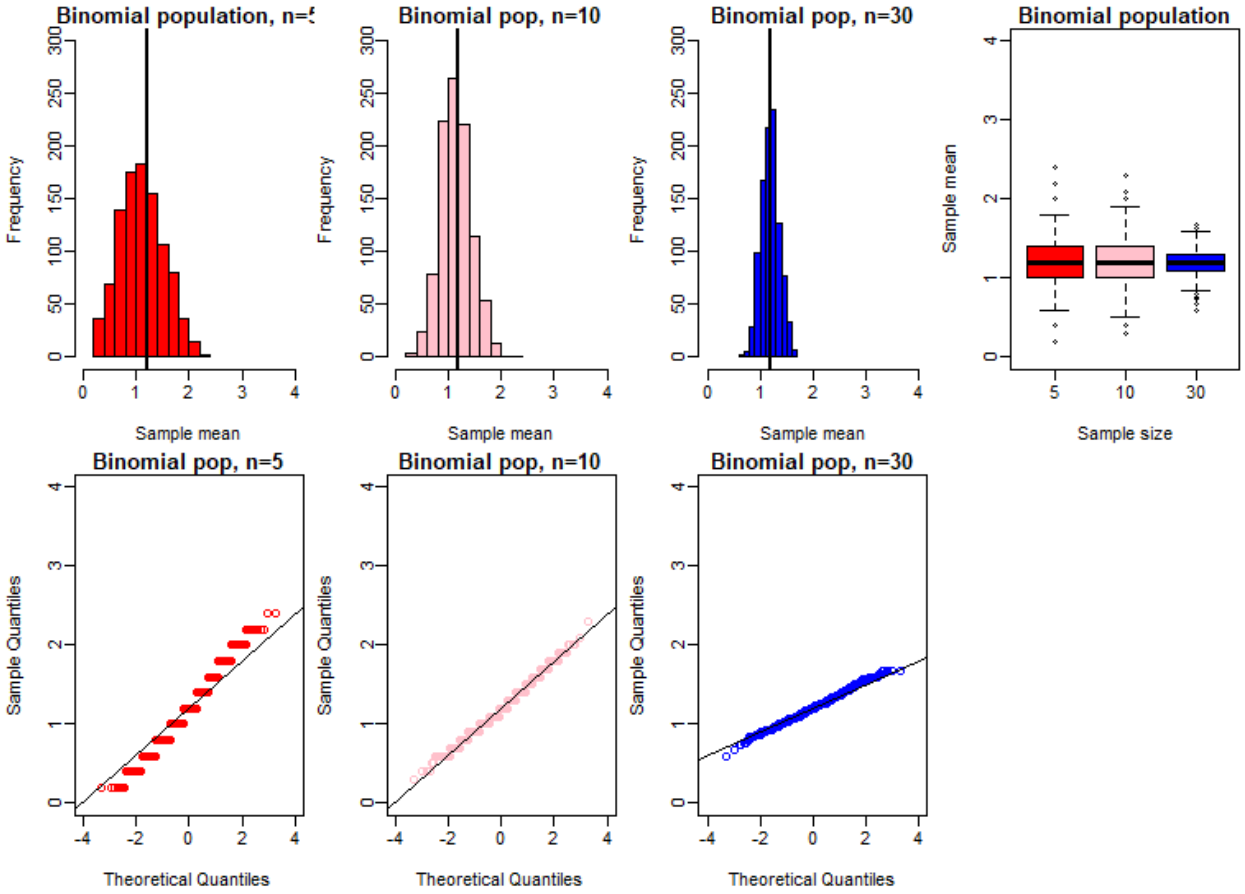


C6 (2) Describe how the sampling distribution of the sample sample variance changes with increasing sample size, including the shape, variability, and center of the distribution.

All histograms are unimodal, and the variance of the population lies about where the peak is, showing that their small sized distributions mimic the underlying population's variance. This is shown also by the boxplots, with their medians about the where the population's variance is, at 0.25.

The boxplots and the scatter plots confirm what the histograms show, that the data is right skewed. The variance of the sample of size 5 has highest number of outliers than the other two(as seen in the boxplots and scatter plots), confirming variability is higher in smaller sample sizes.

C7 (2) Provide the graph generated by the code above for the distribution of the sample mean when sampling from a binomial distribution. Set the axis limits to (0,4)

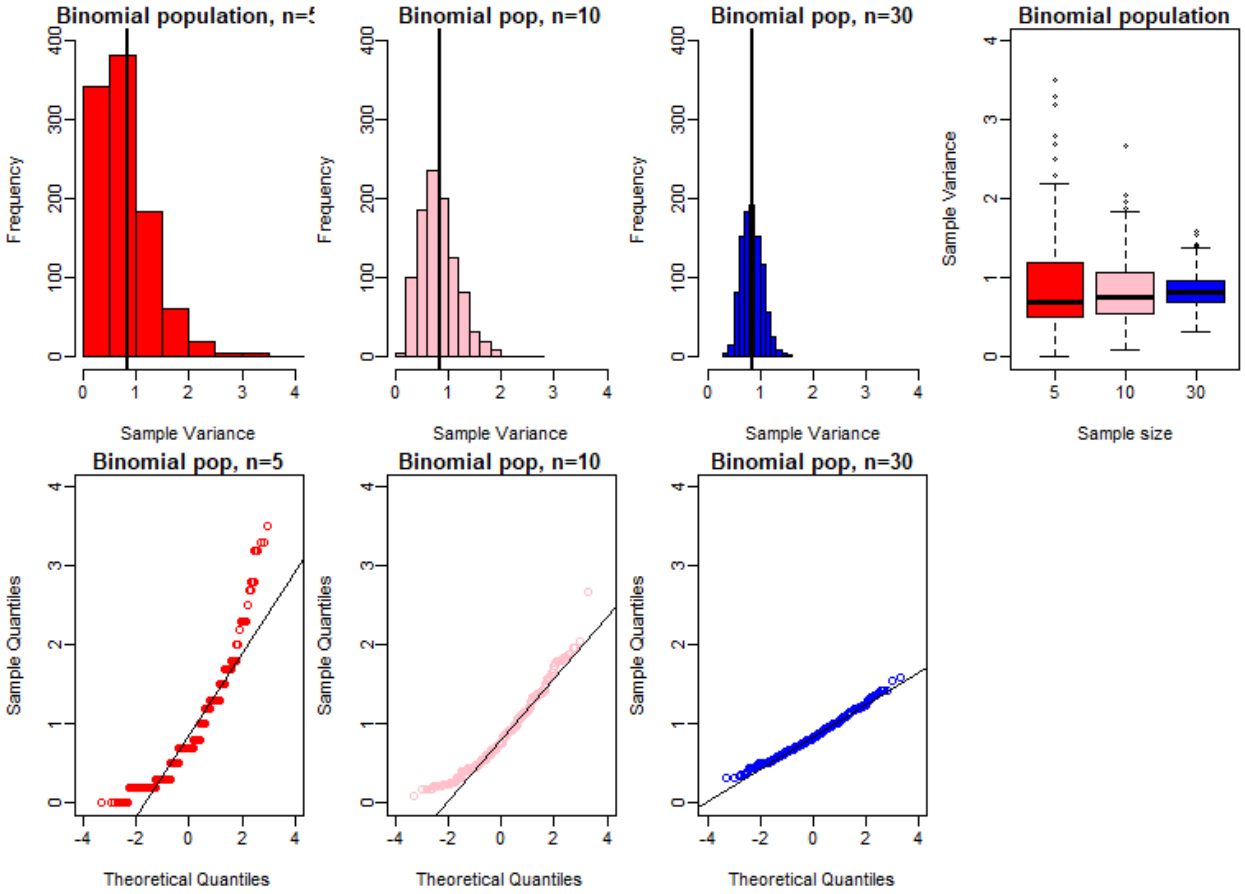


C8 (2) Describe how the sampling distribution of the sample mean (when sampling from an population that follows a binomial distribution) changes with increasing sample size, including the shape, variability, and center of the distribution.

All histograms are unimodal, with greater spread seen in sample with size n = 5 and 10, with bases that range in width from 0 to about 2 units. Base of histogram with n =30 has about a base that is 1 unit wide. All three boxplots have outliers at both ends, but n = 5 is longest compared to n = 10 and n = 30. At boxplot n = 30, has less variability given that it's interquartile range and overall height is less than the other two's IQR and height.

All scatterplots have positive association with values from the population as seen by how most are beside or on the qqline. Strongest linear relationship is seen in n = 10 and n = 30, where most values lie on qqline, showing an almost normal distribution. AT n=5, the qqplot is leptokurtic with fatter tails, also seen from the histogram.

C9 (2) Provide the graph generated by the code above for the distribution of the sample variance when sampling from a binomial distribution. Set the axis limits to (0,4)



C10 (2) Describe how the sampling distribution of the sample variance (when sampling from an population that follows a binomial distribution) changes with increasing sample size, including the shape, variability, and center of the distribution.

All histograms and scatterplots show the distribution of the binomial variance is right skewed. This distribution confirms what we have seen, that at n = 30, the distribution of the variance (and even the mean) is close to that of the underlying population compared to n= 5 and n=10. Shortest boxplot is n = 30, showing that its variance has less variability than at n = 5 and n= 10. At n = 10 and 30, the scatterplots follow a normal distribution closely compared to n = 5. At n = 30, most points on the scatterplot align with qqline of the binomial population's variance. Hence size n =30 has a close positive relationship with the population distribution than the rest of the data.