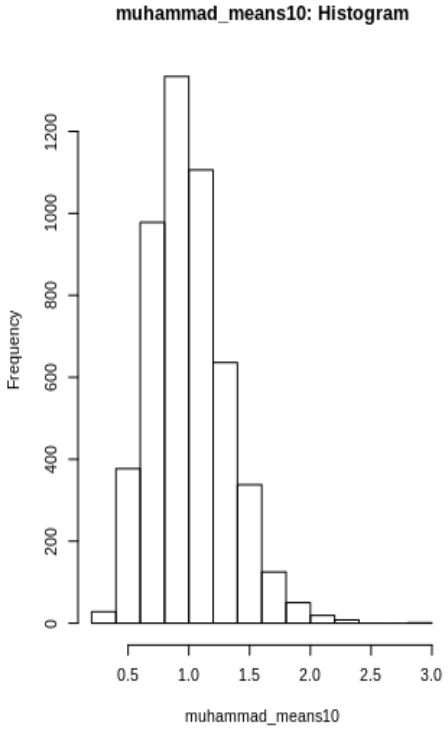
Tahir Muhammad
1002537613

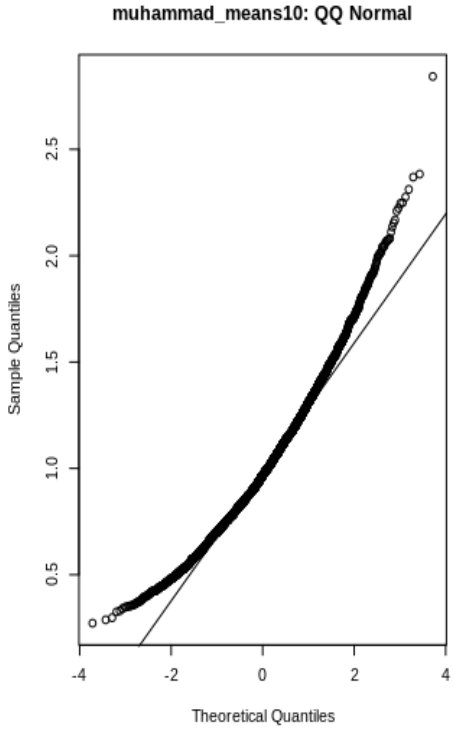
Question 1 Answer:

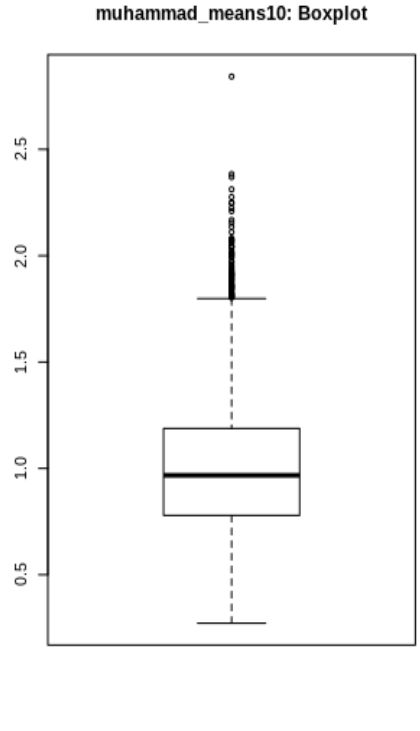
a) muhammad_means10 <- c(rep(NA,5000))

b) for (i in 1:5000){
muhammad_means10[i] <-mean(rgamma(10,1,1))
}

c) hist(muhammad_means10, main="muhammad_means10: Histogram")
qqnorm(muhammad_means10, main="muhammad_means10: QQ Normal")
boxplot(muhammad_means10, main="muhammad_means10: Boxplot")







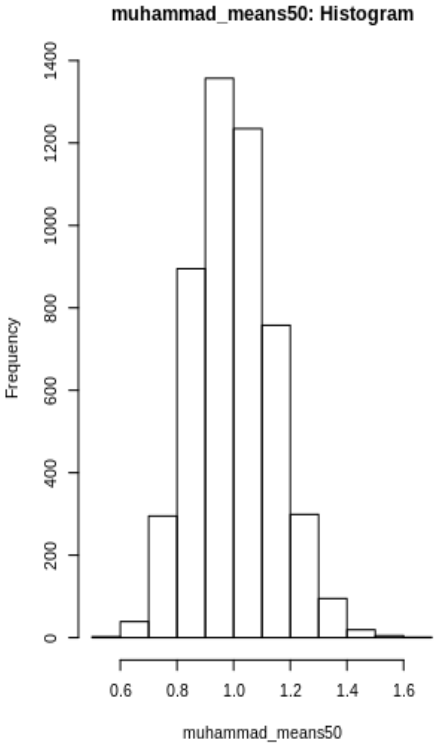
d) mean(muhammad_means10)
[1] 0.9982071
sd(muhammad_means10)
[1] 0.3128598

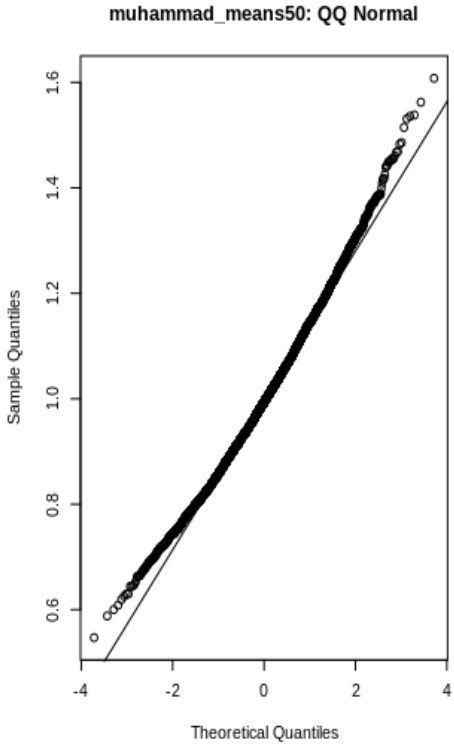
Question 2 Answer:

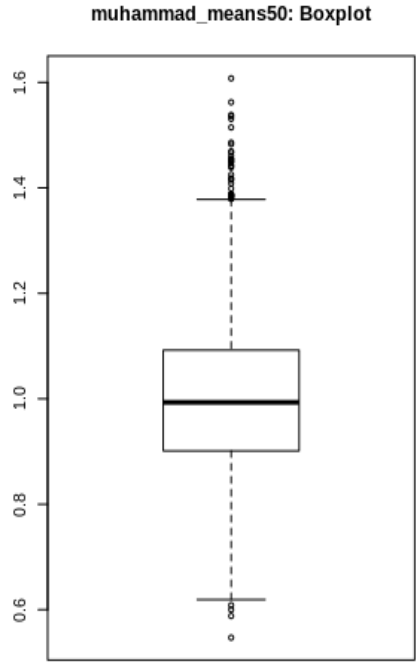
a) muhammad_means50 <- c(rep(NA,5000))

b) for (i in 1:5000){
muhammad_means50[i] <-mean(rgamma(50,1,1))
}

c) hist(muhammad_means50, main="muhammad_means10: Histogram")
qqnorm(muhammad_means50, main="muhammad_means10: QQ Normal")
boxplot(muhammad_means50, main="muhammad_means10: Boxplot")







d) `mean(muhammad_means50)`
`[1] 0.9992662`

`sd(muhammad_means50)`
`[1] 0.1415601`

Question 3 Answer:

- We can see that for Q1 the population seems positively skewed (from the histogram)
- It also contains alot of outliers in the upper region of the boxplot (for Q1)
- We also note that the QQ-plot demonstrates enough evidence to say this is not exactly a normal distribution (Most of the points are not on the line, and has many outliers on both ends
- However, in Q2, we observe:
    - A more “bell shaped” curve histogram. Not only is it more normal in its fit, but also tigher around the mean
    - QQ-plot is mostly on the 45degree line, except a few outliers at the ends
    - We can see that the Skew has vanished in Q2 (with more sample size)

- Based on these observations being demonstrated, we can conclude that the bigger the sample size, the more normally distributed your data is. And intuitively, we can see why this would happen as because when your sample size is large, the odds of “falling” to an extreme end has lower likihood, (i.e 25/25 or a 100/100) and usually falls more towards the expected value. This means that your sample mean is less likely to be far away from the true mean, and this is why it is preffered to use a larger sample size (your standard deviation is less, and estimation is more accurate)