Computer Project2

Patrick Wong

11/07/2018

MATH 324 Computer Project 2

Algorithm 1: Generate 500 Sample Means from Sample Size n and for a Particular Distribution

```
a <- 0
b <- 5
sample_size <- 500
numList <- runif(sample_size, min = a, max = b)

number_trials <- 15
p <- 0.2
bino <- rbinom(sample_size, size = number_trials, prob = p)

lambda <- 5
exp <- rexp(sample_size, rate = lambda)

mu <- 2
pos <- rpois(sample_size, mu)

###Algorithm)
sampleMeans = 500
storedData5 = rep(NA, sampleMeans)
storedData50 = rep(NA, sampleMeans)

for (i in 1:sampleMeans){
  arrayVal5 = runif(5, min = 0, max = 5)
  arrayVal50 = runif(50, min = 0, max = 5)

  average5 <- mean(arrayVal5)
  average50 <- mean(arrayVal50)

  storedData5[i] <- average5
  storedData50[i] <- average50
}
```

Question 1)

```
paste0("Mean for Sample Size 5: ", mean(storedData5))
```

```
## [1] "Mean for Sample Size 5: 2.54257085117279"
```

```
paste0("Mean for Sample Size 50: ", mean(storedData50))
```

```
## [1] "Mean for Sample Size 50: 2.502100544318"
```

```r
theoreticalMean <-  (0.5*(0+5))
paste0("Theoretical value for the uniform sample means: ", theoreticalMean)
```

```
## [1] "Theoretical value for the uniform sample means: 2.5"
```

Question 2)

```r
paste0("Variance for Sample Size 5: ", var(storedData5))
```

```
## [1] "Variance for Sample Size 5: 0.457397547302951"
```

```r
paste0("Variance for Sample Size 50: ", var(storedData50))
```

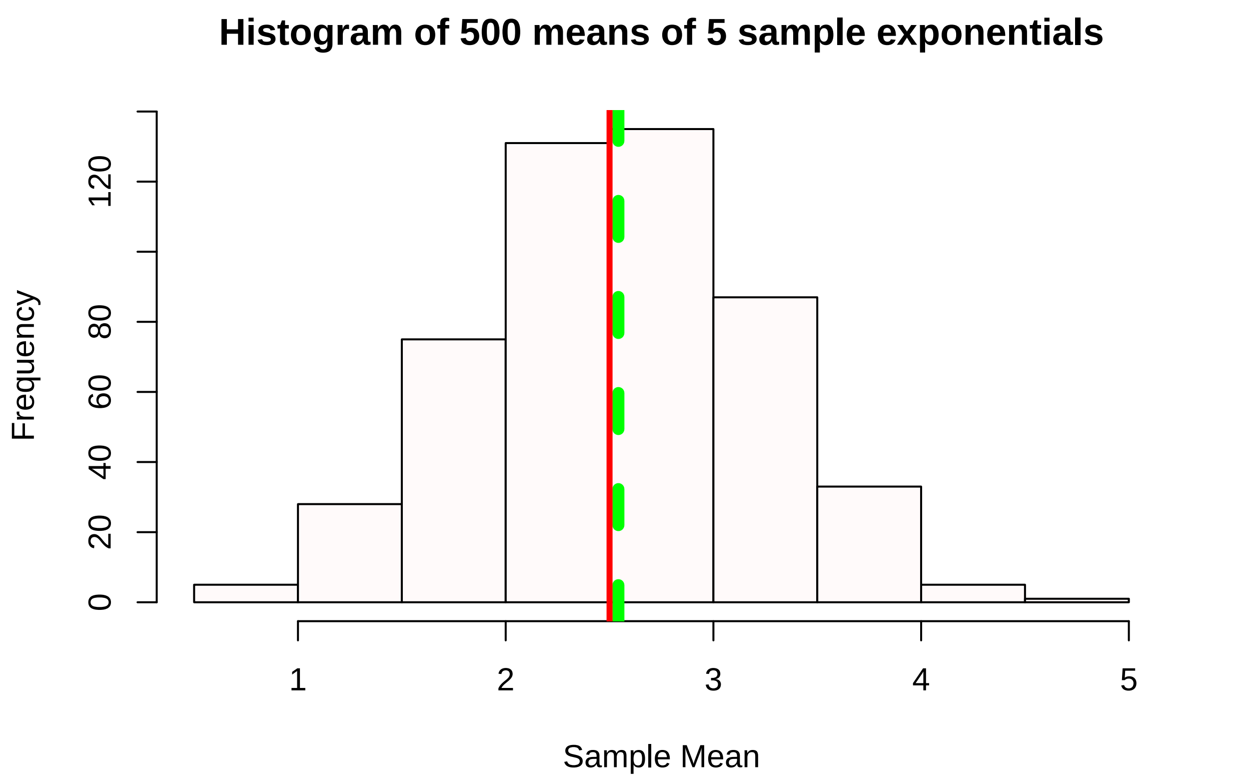
```
## [1] "Variance for Sample Size 50: 0.038176375816325"
```

```r
theoreticalVar  <- ((1/12)*(5-0)^2)
paste0("Theoretical value for the sample variance: ", theoreticalVar )
```

```
## [1] "Theoretical value for the sample variance: 2.08333333333333"
```

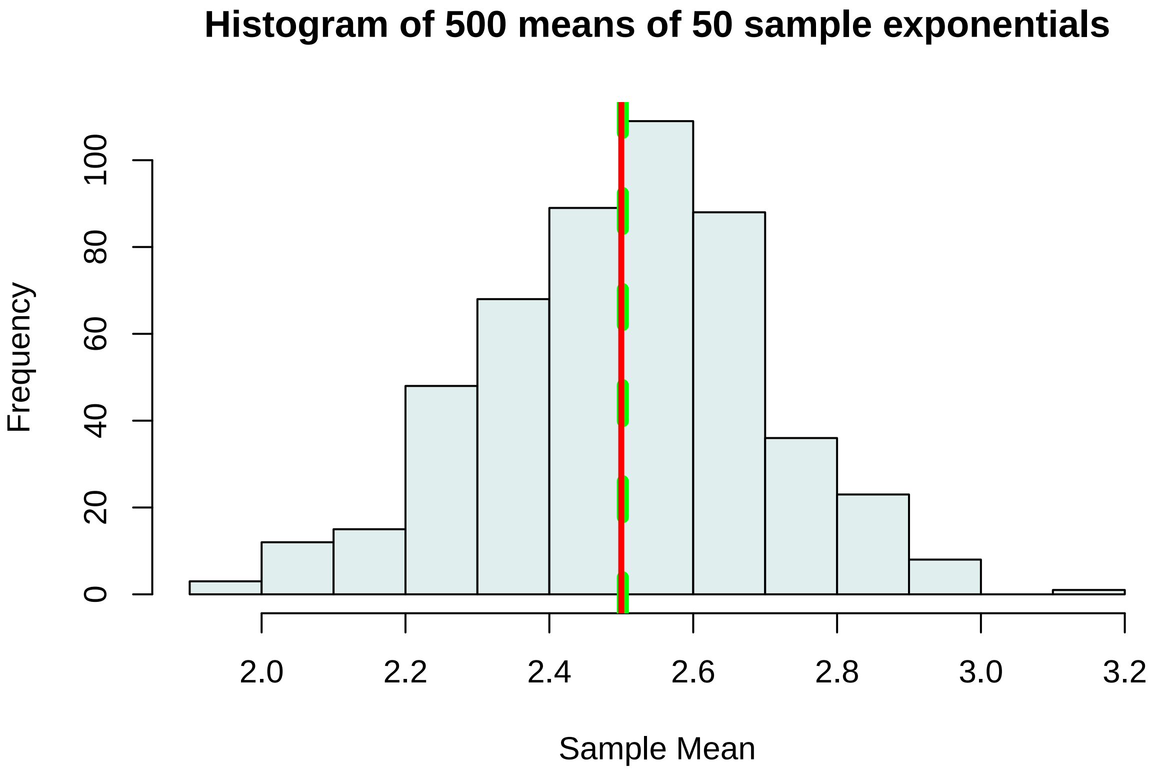
Question 3)

```r
hist(storedData5, col="snow", main="Histogram of 500 means of 5 sample exponentials"
     , xlab="Sample Mean", ylab="Frequency")
abline(v=mean(storedData5), col="green", lwd=6, lty=2)
abline(v=theoreticalMean, col="red", lwd=3)
```



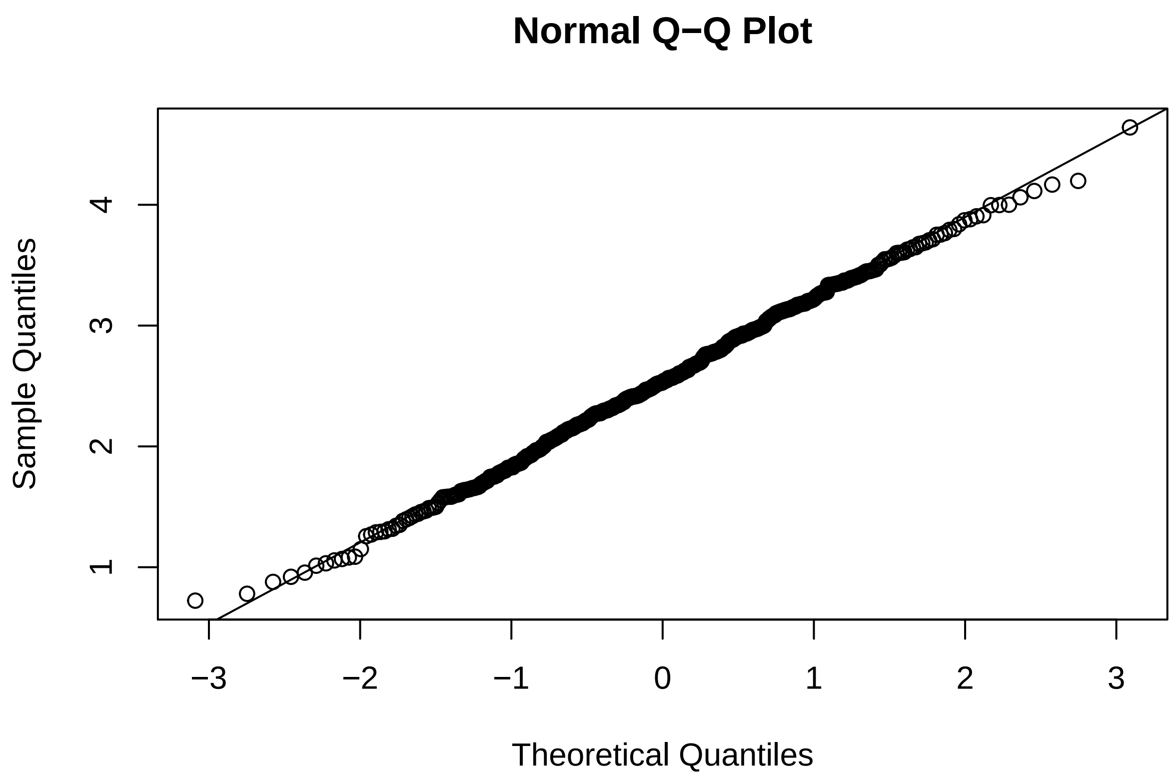

```r
hist(storedData50, col="azure2", main="Histogram of 500 means of 50 sample exponentials"
     , xlab="Sample Mean", ylab="Frequency")
```

```r
abline(v=mean(storedData50), col="green", lwd=6, lty=2)
abline(v=theoreticalMean, col="red", lwd=3)
```

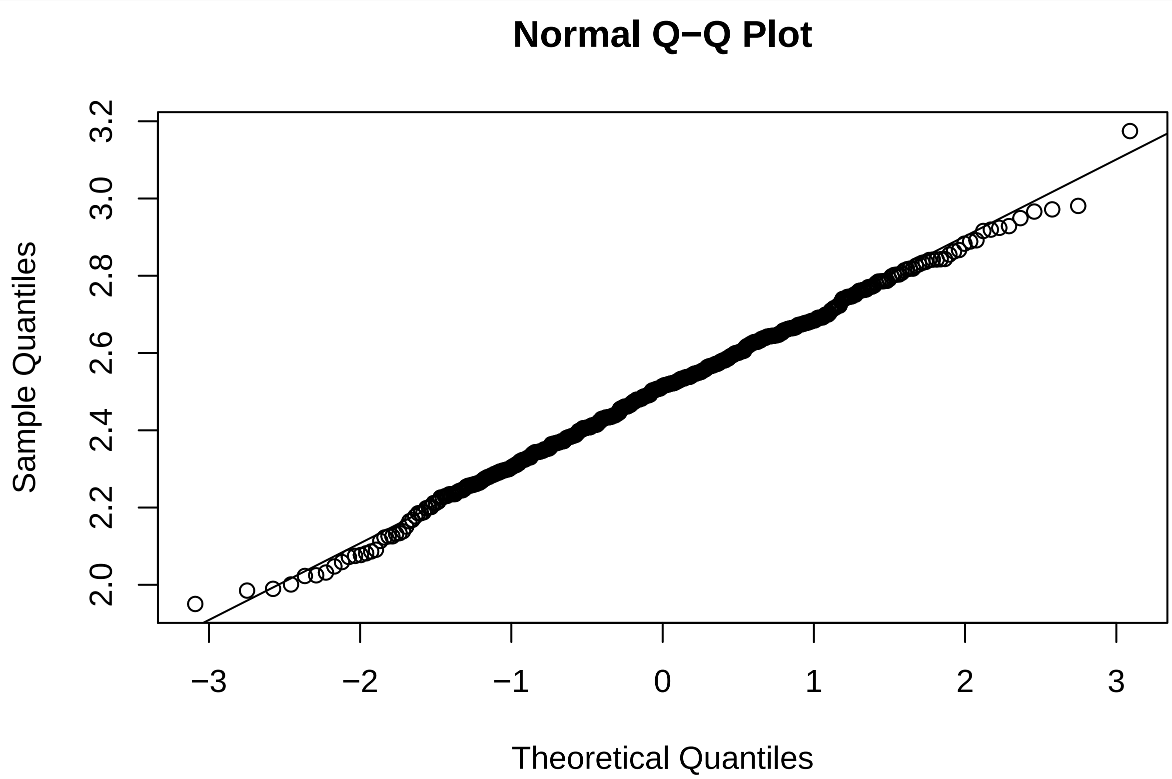


Question 4)

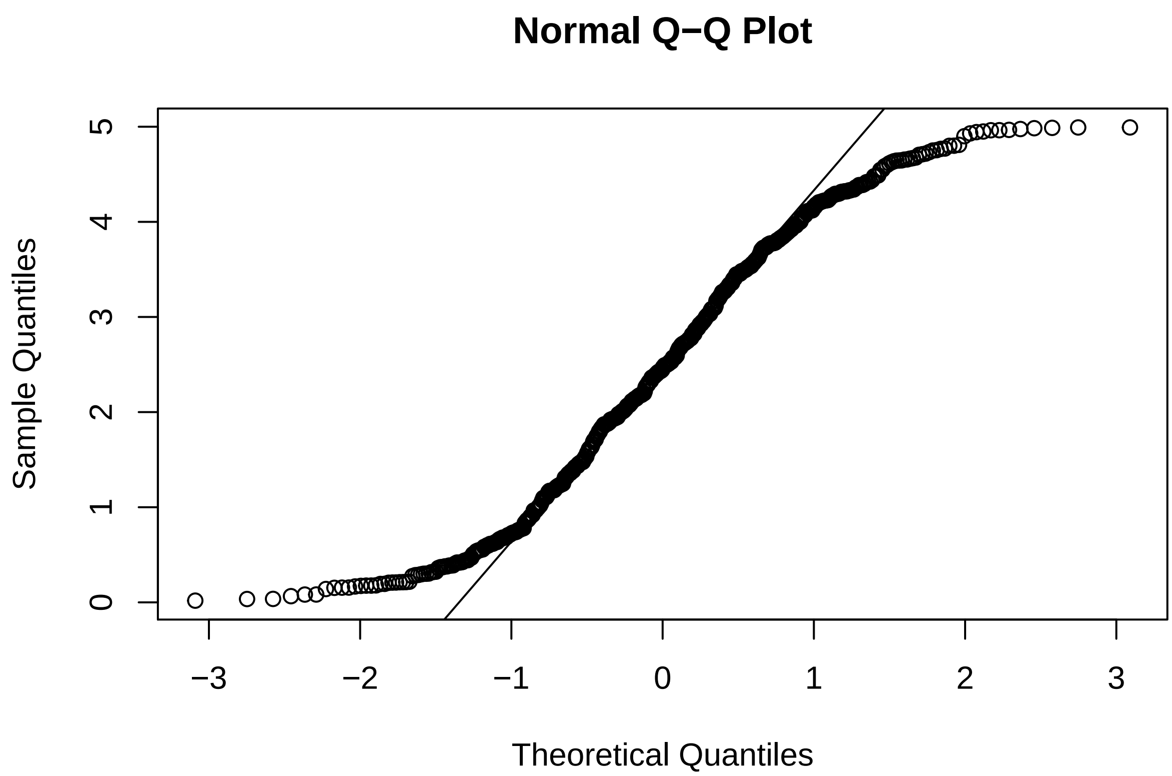
```r
x <- storedData5
y <- storedData50
a <- numList
b <- bino
c <- exp
d <- pos
qqnorm(x)
qqline(x)
```



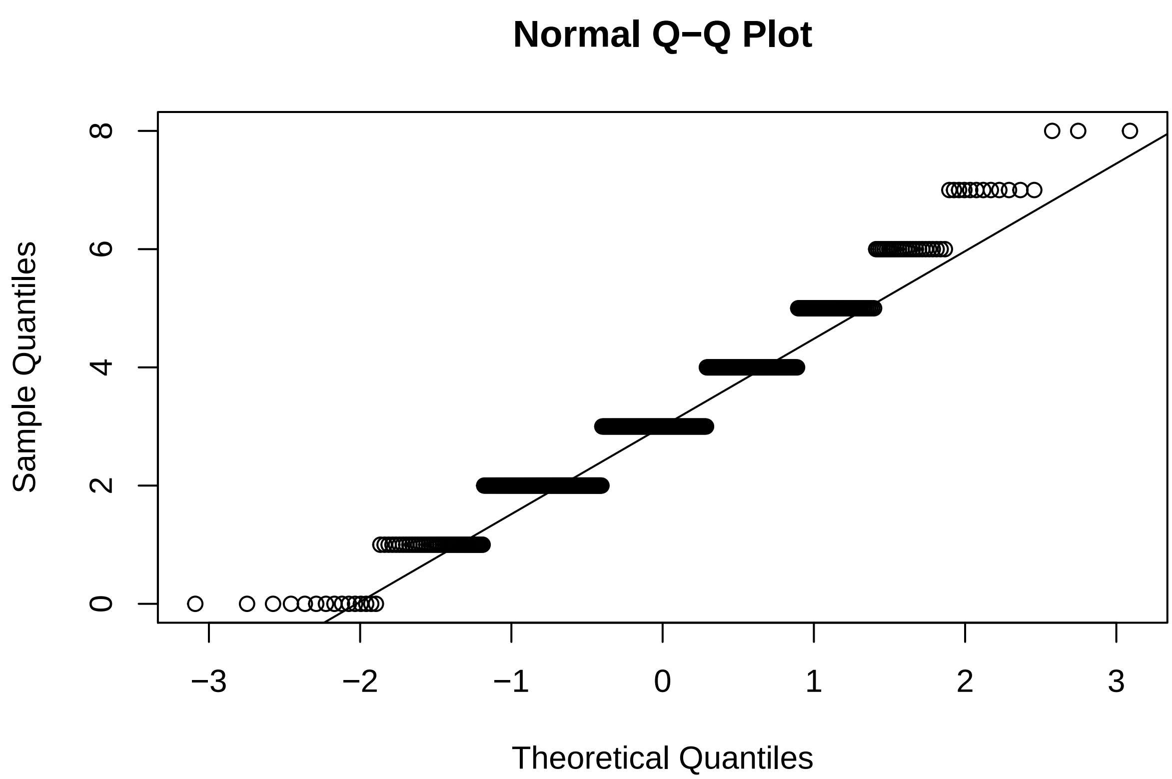

```r
qqnorm(y)
qqline(y)
```



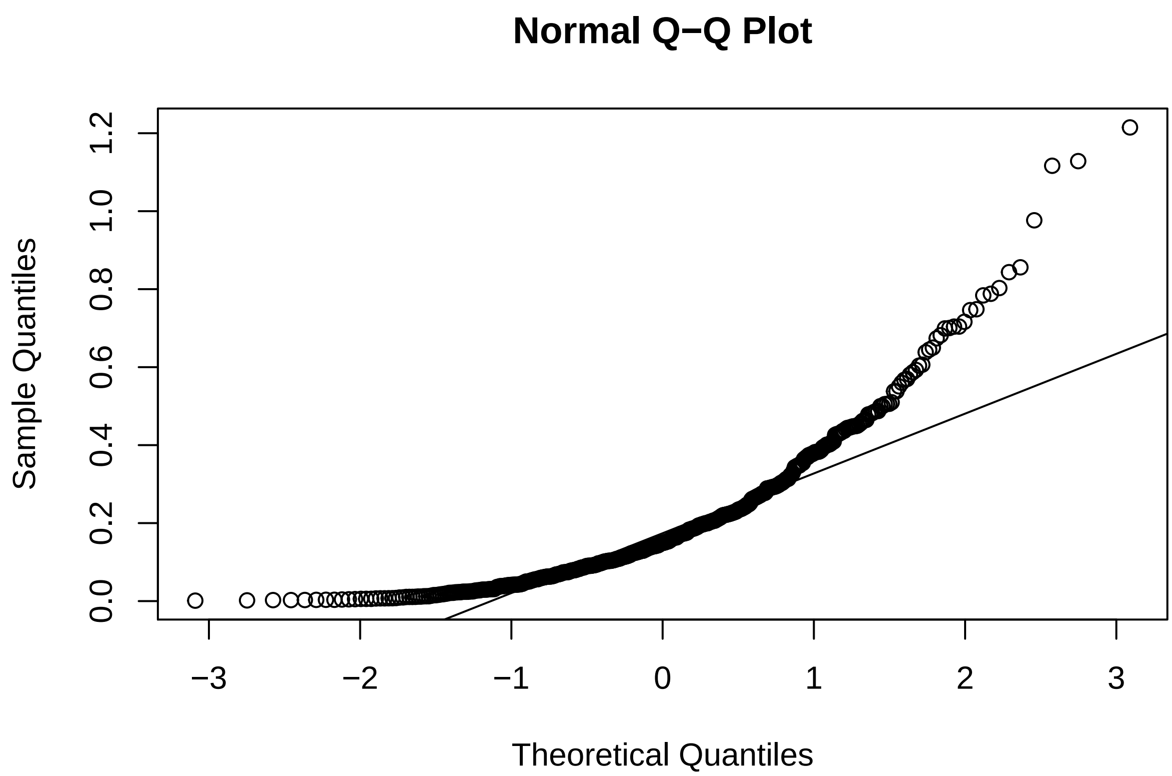

```r
qqnorm(a)
qqline(a)
```



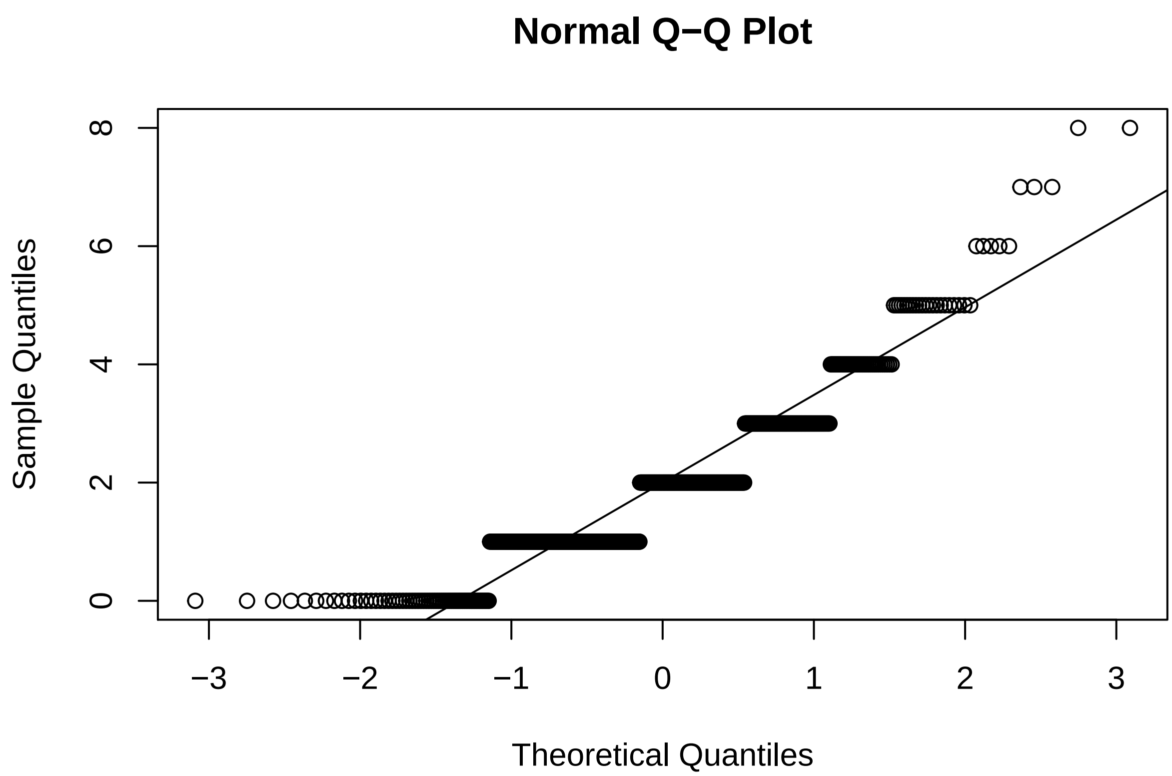

```
qqnorm(b)
qqline(b)
```




```
qqnorm(c)
qqline(c)
```




```r
qqnorm(d)
qqline(d)
```



Question 5) Summarize these findings for each of a-d, and use the central limit theorem to explain your findings.

It tells us that when the sample sizes get larger, the distribution of means will be calculated through repeating sampling until it reaches normality. It also tells us that the same population will be approximately equal to the mean of the population.

Algorithm 2: Use 40,000 Bootstrapped Means to Estimate Variance of the Mean

```
m2 <- 50
p2 <- 0.2
n = 5
X <- rbinom(m2,prob = p2, size = m2)
bMeans <- 40000
xMeanArray <- array(dim = bMeans)
for(i in 1:bMeans){
  # sample(<your_vector_of_50_binomial_random_numbers>, sample = 50, replace = TRUE)
  xMeanArray[i] <- mean(sample(X, n, replace = TRUE))
}
v <- var(xMeanArray)
```

A)

```
paste0("Bootstrap Variance: " , v)
```

```
## [1] "Bootstrap Variance: 1.81018094642366"
```

B)

```
#theoretical = np(1-p)/n
t <- m2*p2*(1-p2)/m2
paste0("Theoretical variance: ", t)
```

```
## [1] "Theoretical variance: 0.16"
```