

GRADUATE STUDENT STAT 840 A1

Vsevolod Ladtchenko 20895137

Problem 10

(a)

```
generate_samples = function(n)
{
  # generate 100*n samples. then reshape them into matrix 100 x n
  disc = sample(x = c(0,1,7), n*100, replace = T, prob = c(0.25, 0.25, 0.5))
  exps = rexp(n*100, .2)
  disc = matrix(disc, n)
  exps = matrix(exps, n)

  # mean of each column (sample of size n) yielding vector of size 100
  disc_means = colMeans(disc)
  exps_means = colMeans(exps)

  # join results into data frame
  return(data.frame(disc_means, exps_means))
}
set.seed(1)
df = generate_samples(15)
```

(b)

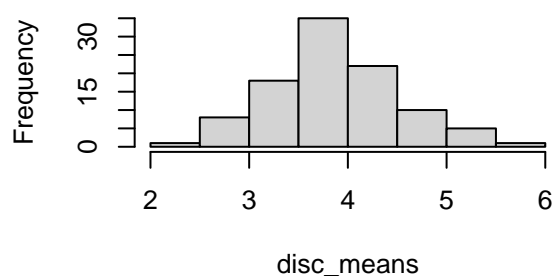
We will now look at whether the CLT provides a good approximation of the distribution of means, for $n = 15$. We will see that the shape depends on the underlying distribution (discrete vs. exponential), but they will both converge as n increases.

Histogram

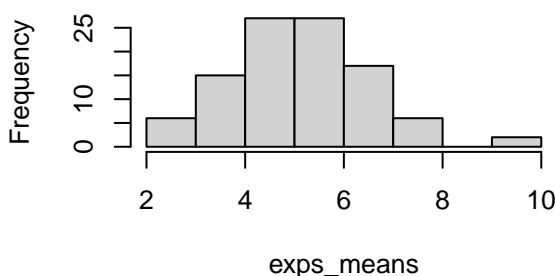
The histogram of discrete means appears more bell-shaped. The histogram of exponential means appears skewed, not very bell-shaped.

```
par(mfrow = c(2,2))
disc_means = df[1][,]
exps_means = df[2][,]
p1 = hist(disc_means)
p2 = hist(exps_means)
plot(p1, col=rgb(0,0,1,1/4), xlim=c(0,10))
plot(p2, col=rgb(1,0,0,1/4), xlim=c(0,10), add=T)
```

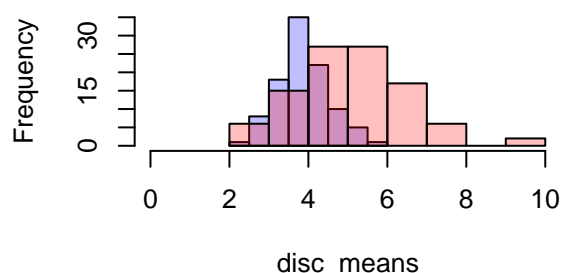
Histogram of disc_means



Histogram of exps_means



Histogram of disc_means



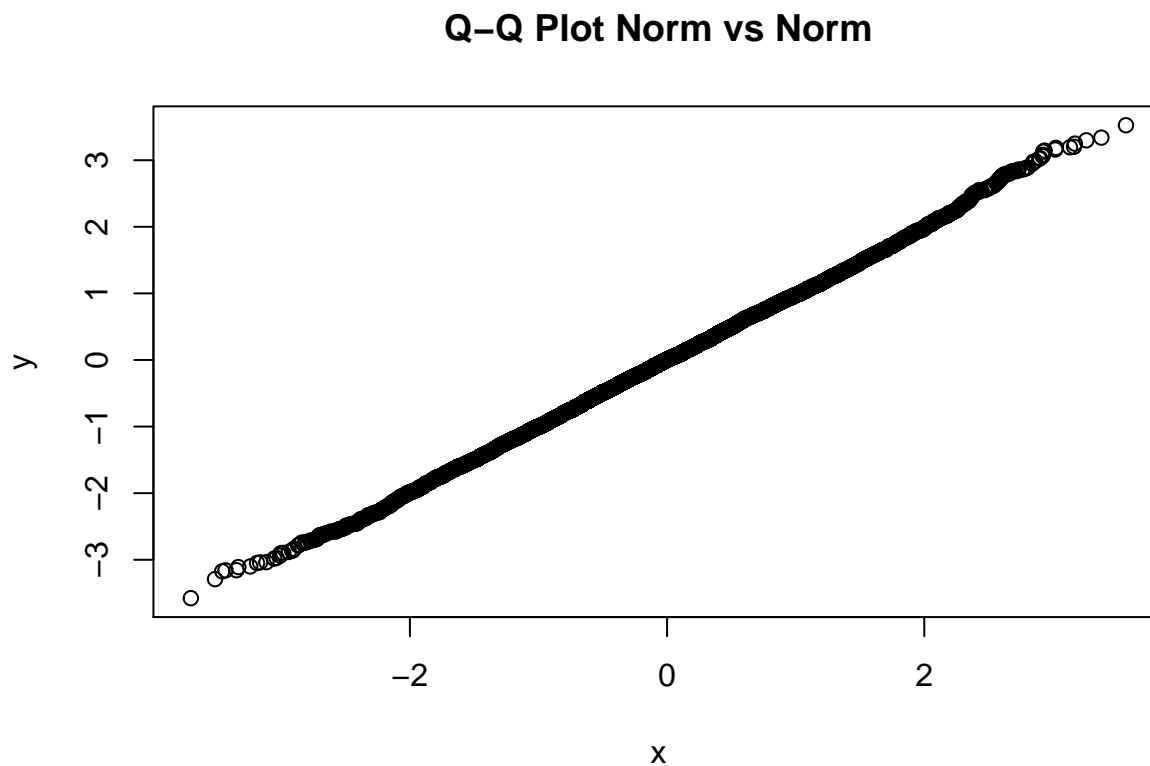
QQ Plot

Using a sample of 100 for the normal sample to compare against, I found that the QQ plot is too wild and they all look the same. Changing the number of normal samples to 10_000 allows a better visualization as more of the normal points are placed into the quantiles of our distribution of means, and more clearly shows the curvature instead of random zig zags.

The QQ plot for two normal samples x and y have an almost linear plot, as expected. Comparing against the distribution of the sample means, we see that they are far from a straight line meaning the distribution is

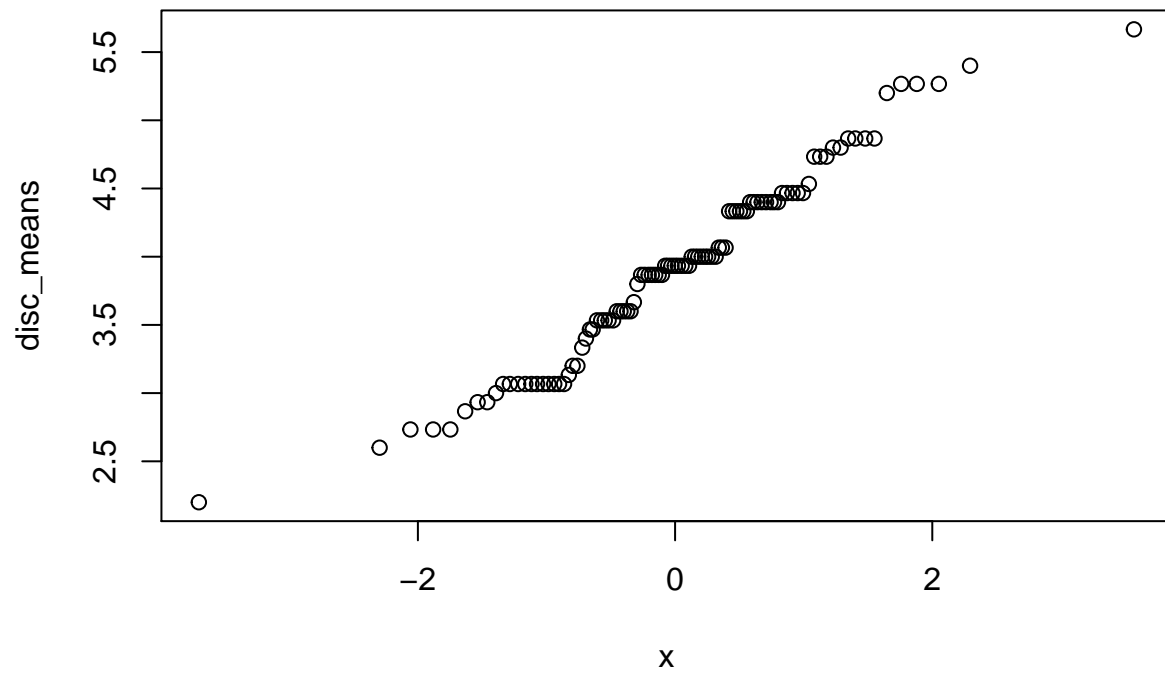
far from normal. The means from the discrete distribution do have a more linear QQ plot, but with farther outliers, suggesting a different tail. The exponential means show more curvature in the QQ plot.

```
#par(mfrow = c(2,2))  
x = rnorm(10000, 0, 1)  
y = rnorm(10000, 0, 1)  
p1 = qqplot(x, y, main = "Q-Q Plot Norm vs Norm")
```



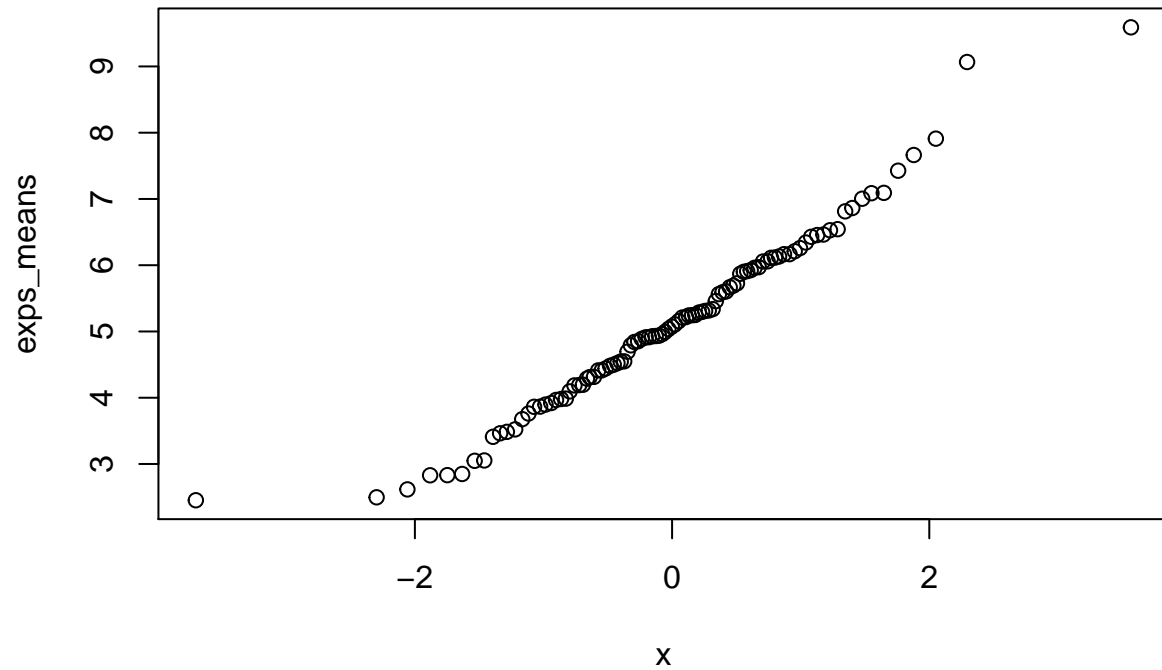
```
p2 = qqplot(x, disc_means, main = "Q-Q Plot Norm vs Discrete")
```

Q-Q Plot Norm vs Discrete



```
p3 = qqplot(x, exps_means, main = "Q-Q Plot Norm vs Exps")
```

Q-Q Plot Norm vs Exps

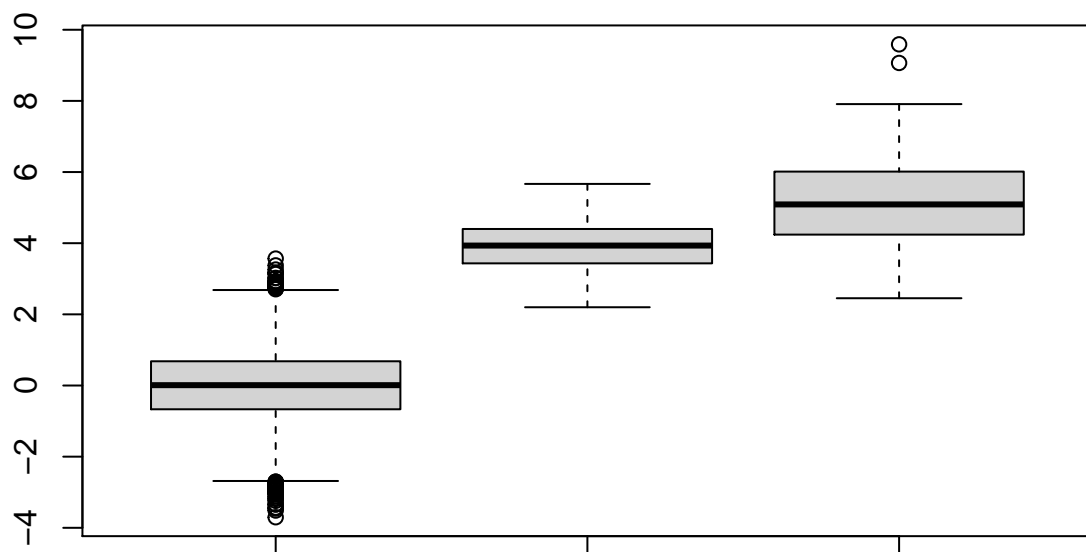


Boxplot

Looking at the box plot, we see the normal has a symmetric box plot and many outliers. Discrete also appears symmetric, with less outliers. The exponential is not symmetric.

```
boxplot(x, disc_means, exps_means, main='Normal vs Discrete vs Exp')
```

Normal vs Discrete vs Exp

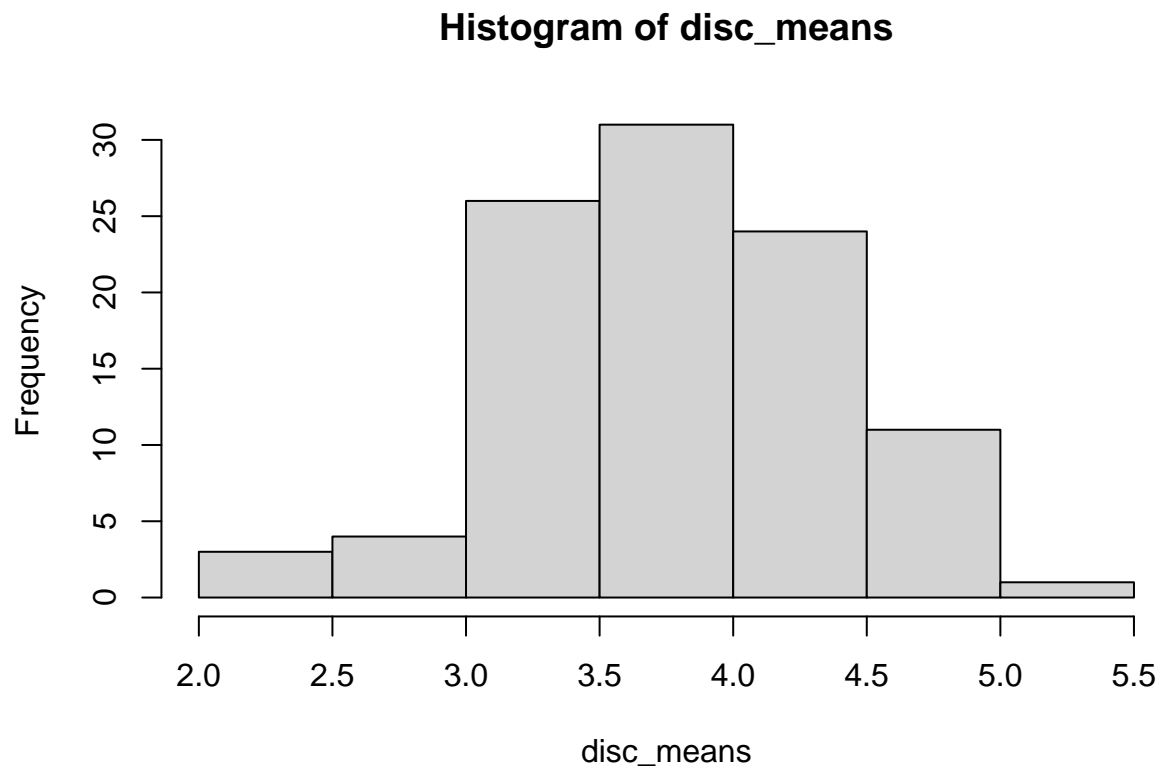


(c)

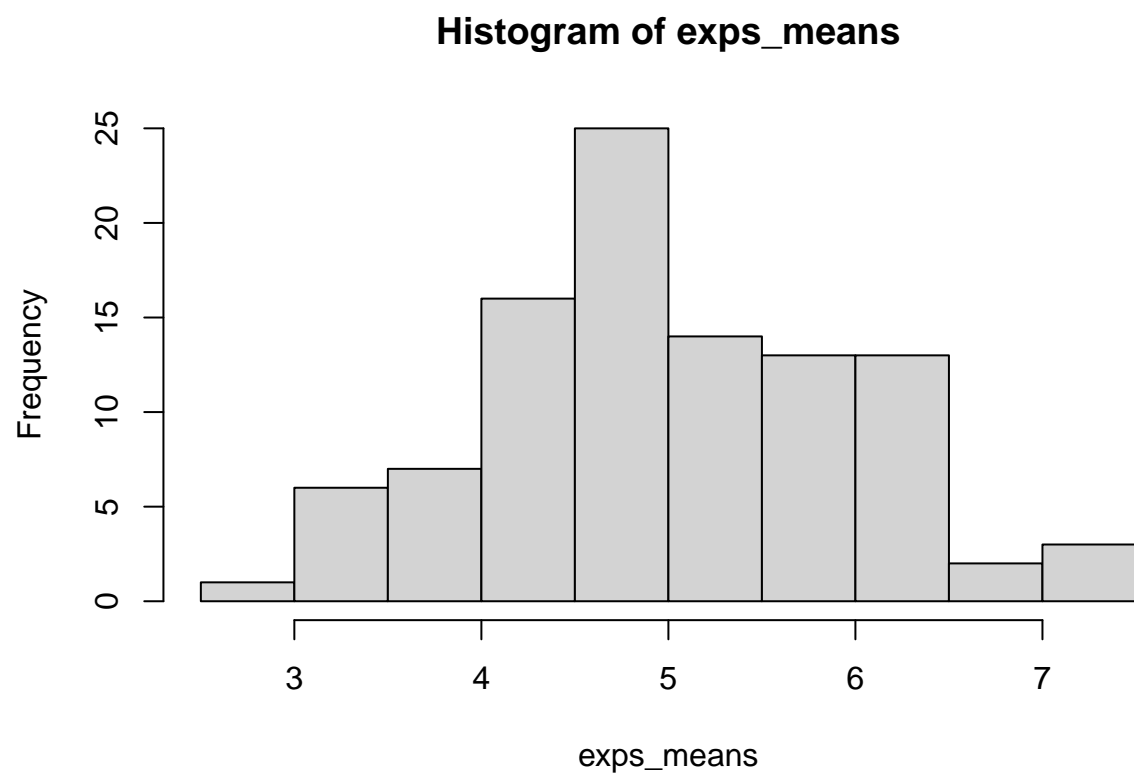
n = 30

Performing the same analyses, we start again with the histogram. The discrete mean distribution appears even more Gaussian. The exponential mean distribution is less skewed than before. The discrete QQ plot appears more linear. Similarly the exponential QQ plot has slightly less curvature. The boxplots are about the same as before.

```
df = generate_samples(30)
disc_means = df[1][,]
exps_means = df[2][,]
hist(disc_means)
```

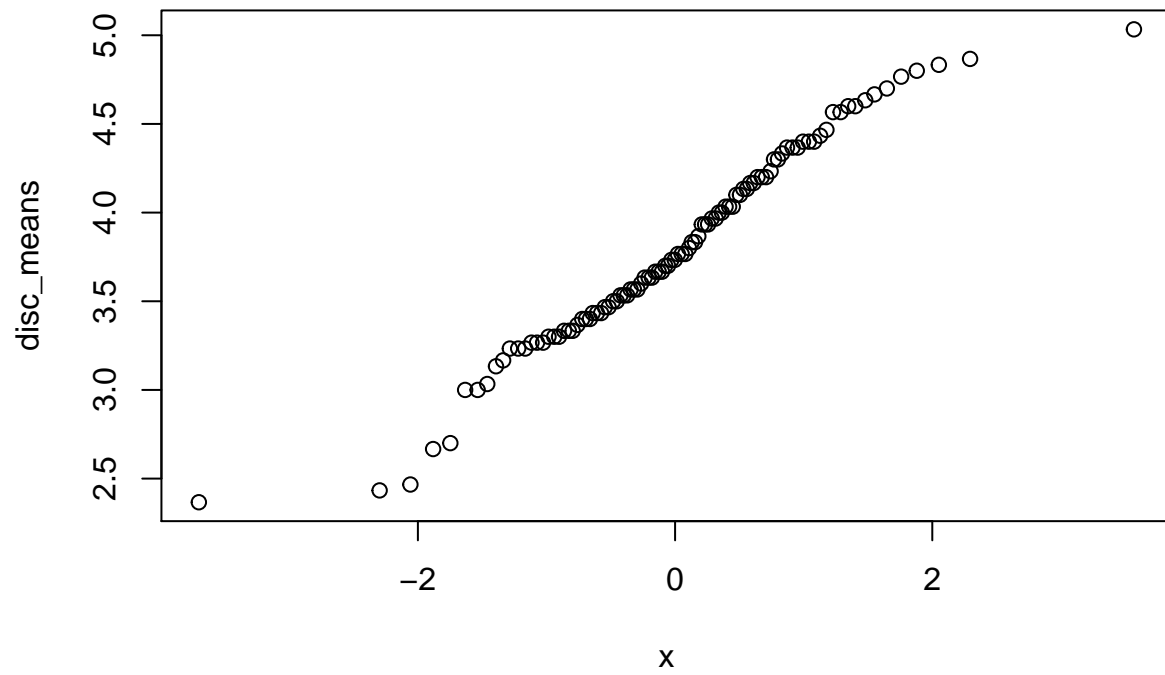


```
hist(exps_means)
```



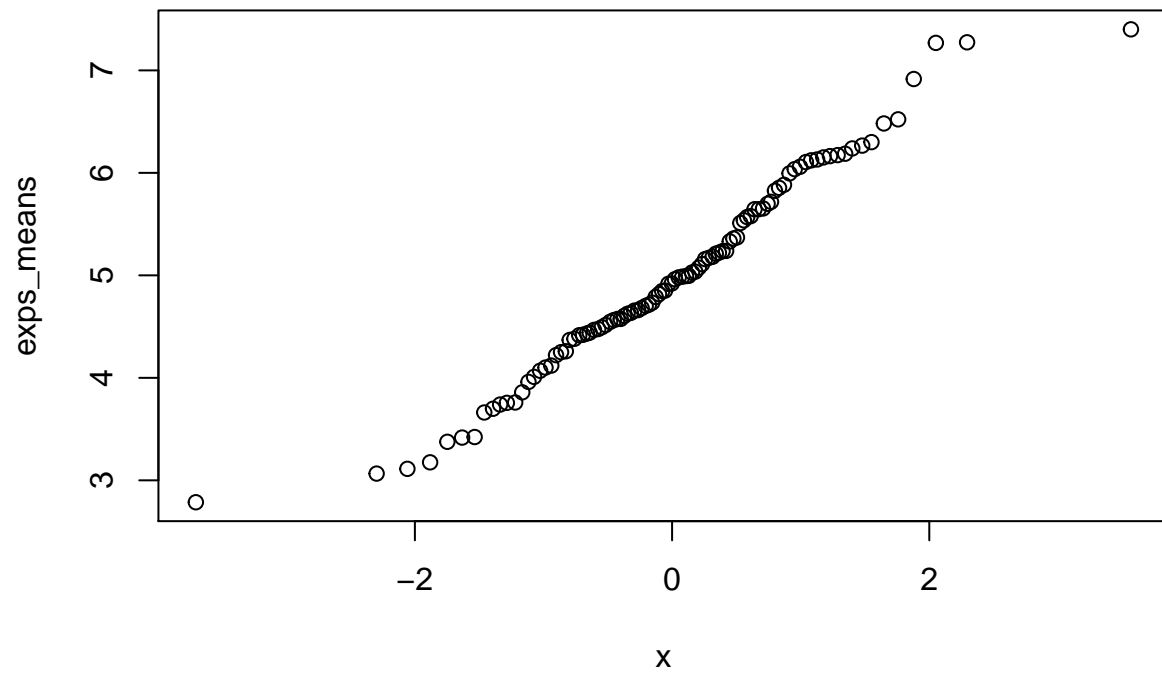
```
qqplot(x, disc_means, main = "Q-Q Plot Norm vs Discrete")
```


Q-Q Plot Norm vs Discrete



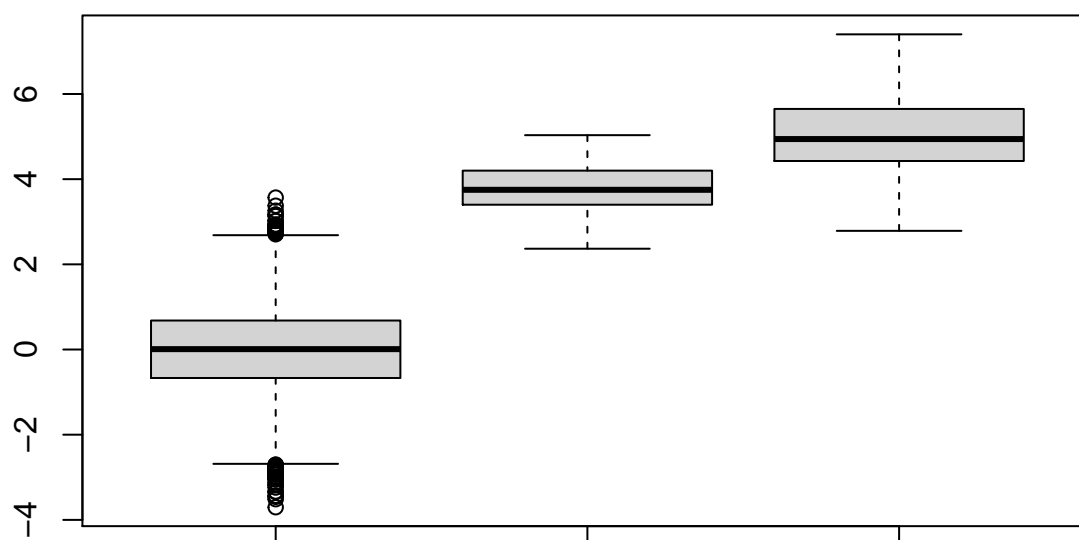
```
qqplot(x, exps_means, main = "Q-Q Plot Norm vs Exps")
```

Q-Q Plot Norm vs Exps



```
boxplot(x, disc_means, exps_means, main='Normal vs Discrete vs Exp')
```

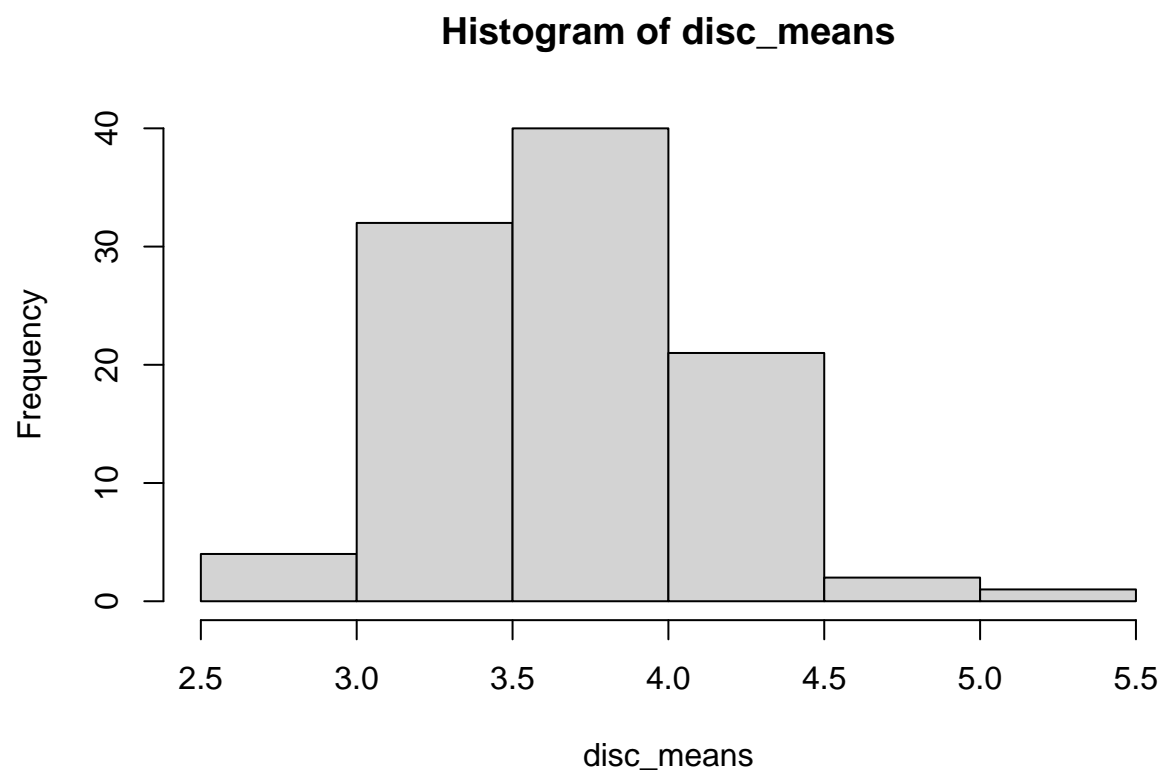
Normal vs Discrete vs Exp



n = 50

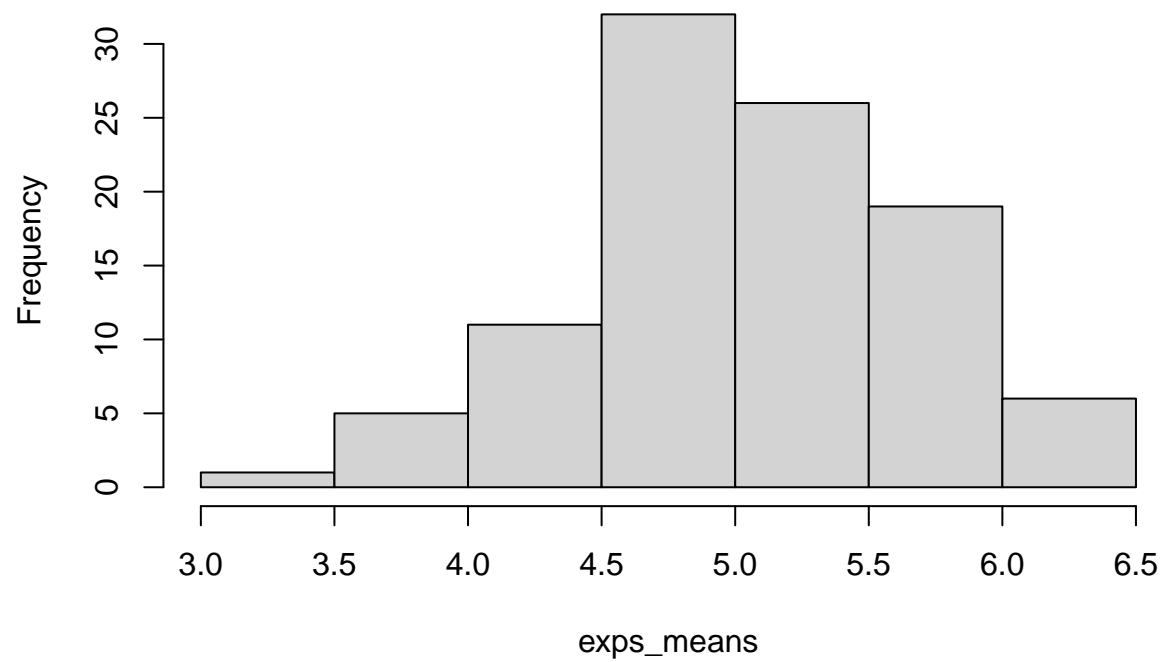
Both the discrete and exponential histograms are even more bell-shaped than for $n=30$. The QQ plots are becoming more linear. The boxplots are slowly becoming more symmetric.

```
df = generate_samples(50)
disc_means = df[1][,]
exps_means = df[2][,]
hist(disc_means)
```



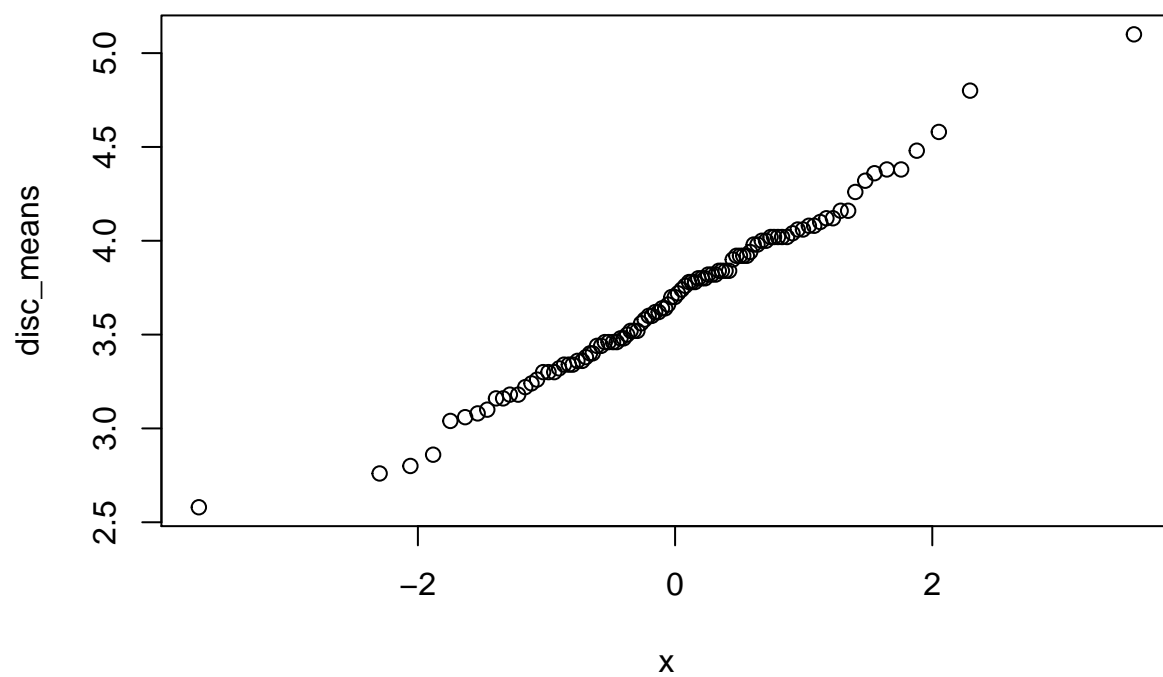
```
hist(exps_means)
```

Histogram of exps_means



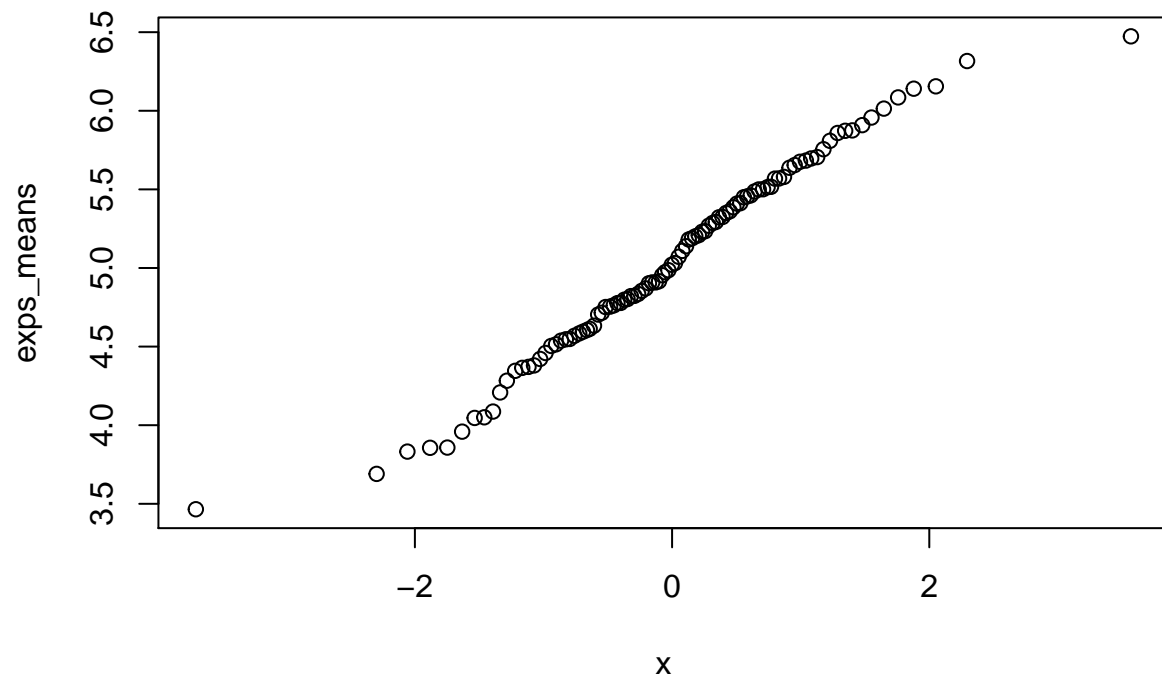
```
qqplot(x, disc_means, main = "Q-Q Plot Norm vs Discrete")
```

Q-Q Plot Norm vs Discrete



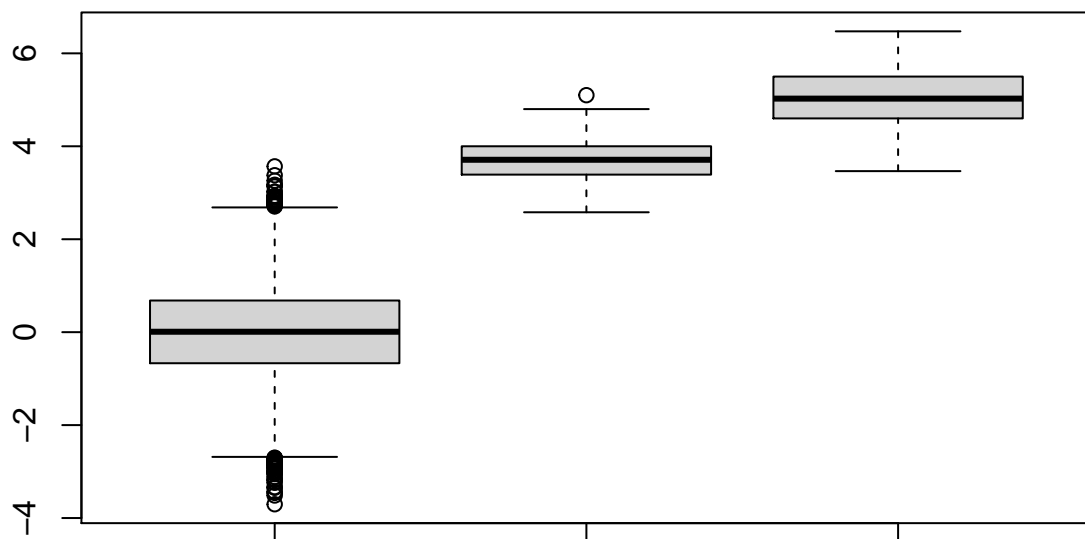
```
qqplot(x, exps_means, main = "Q-Q Plot Norm vs Exps")
```

Q-Q Plot Norm vs Exps



```
boxplot(x, disc_means, exps_means, main='Normal vs Discrete vs Exp')
```

Normal vs Discrete vs Exp



(d)

The conclusion is that as n increases, the distribution of the mean becomes more normal. The histograms become more bell shaped, although at different speeds depending on the underlying distribution. The QQ plots become more linear. the boxplots become more symmetric.