# Assignment Two

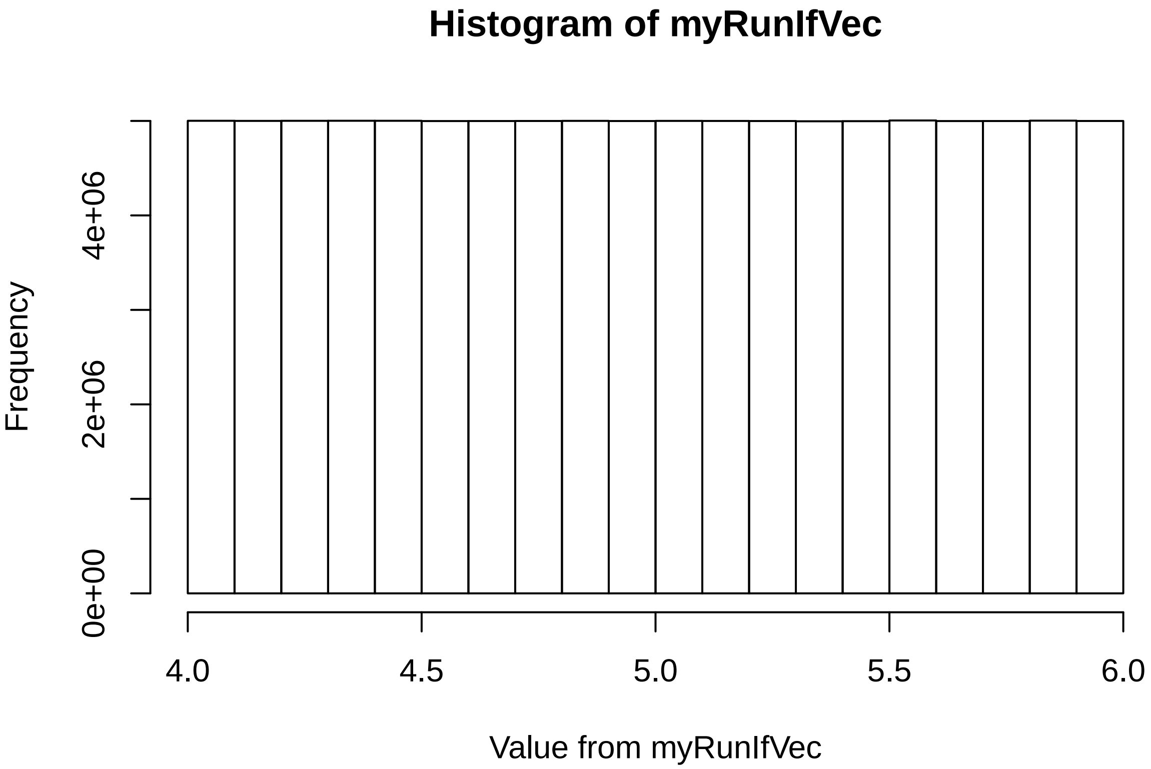
*Louise Lai*

*July 21, 2018*

## Question 1.1

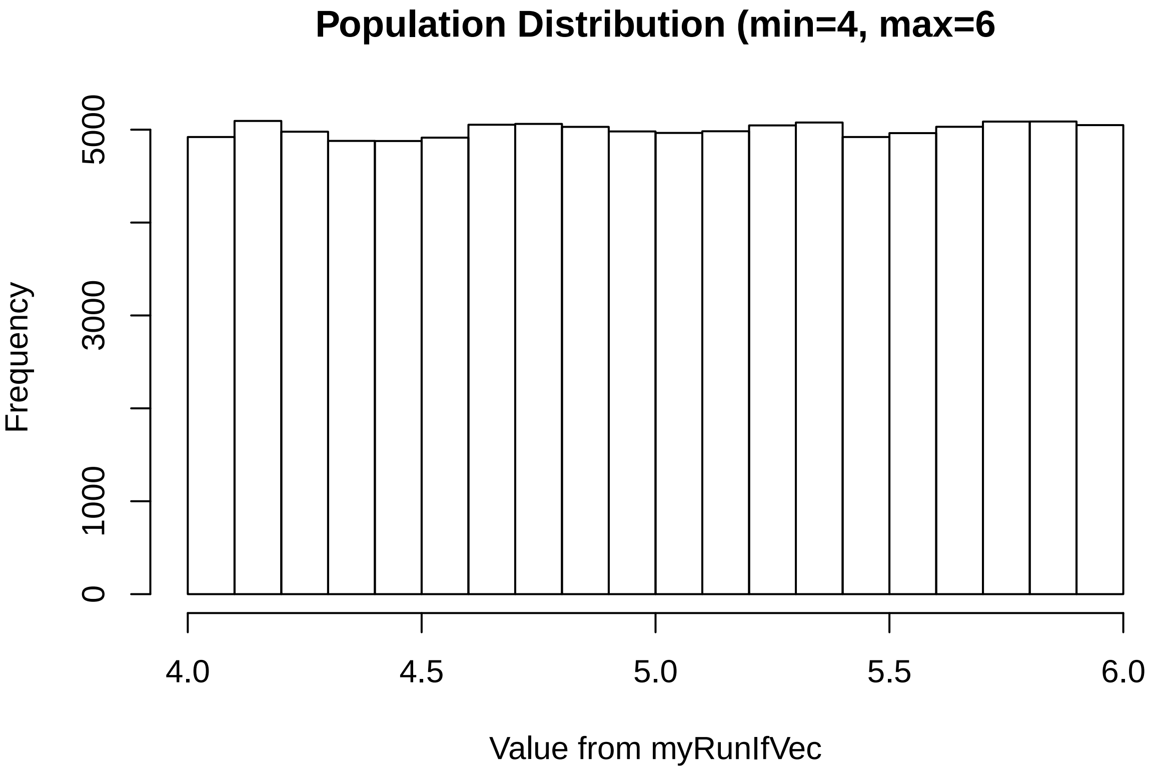
1.1.1. Create 10,000,000 random variates $\sim\mathcal{U}4, 6$ and store the result in a vector called `myRunIfVec`. Create a histogram.

```
myRunIfVec <- runif(100000000, min=4, max=6)
hist(myRunIfVec, xlab="Value from myRunIfVec")
```



1.1.2. Sample randomly 100,000 times from myRunIfVec and plot the sample histogram. Describe the shape of the sampling distribution and note if it is different from the population distribution.

```
hist(sample(myRunIfVec, 100000), xlab="Value from myRunIfVec", main="Population Distribution (min=4, ma
```



The sample distribution looks uniform with no peaks or troughs. It looks very similar to the population distribution.

### Function Definition

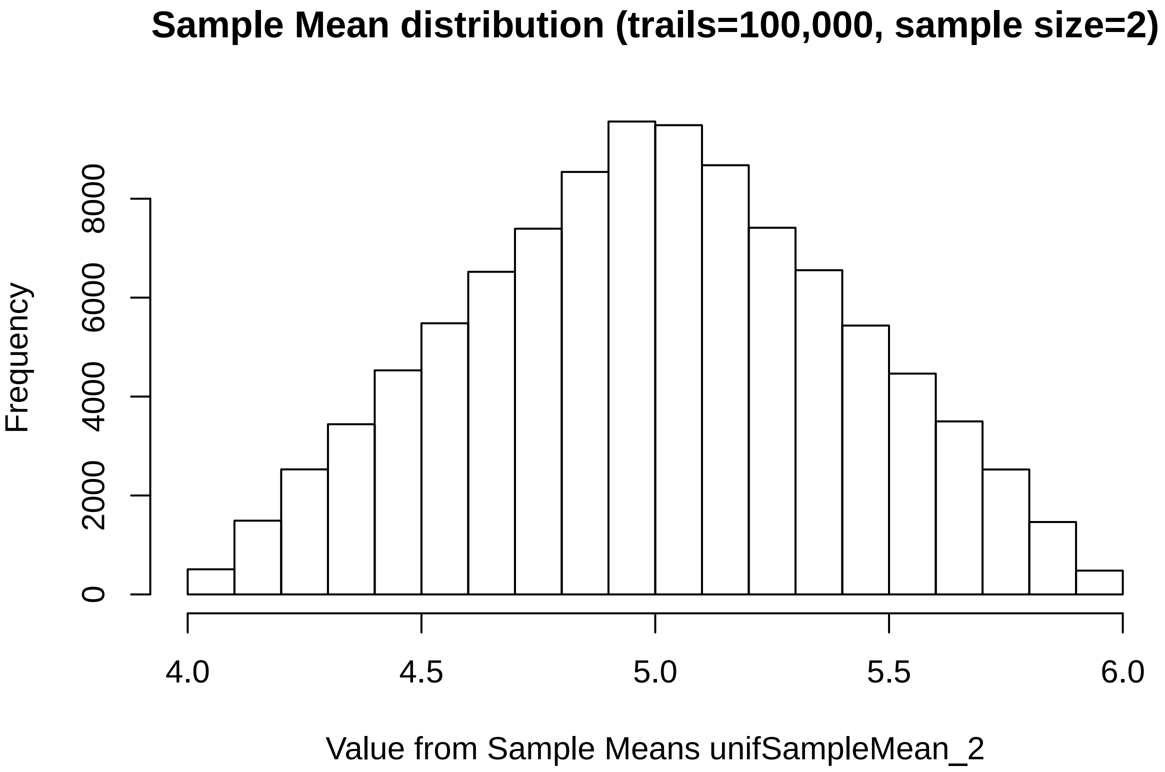
```
# Calculates sample means given a distribution, sample size and n
getNSampleMeans <- function(n, sampleSize, distribution){

  # 1) create sampleSize vectors
  vectors <- vector("list", sampleSize)
  for(i in 1:sampleSize){
    vectors[[i]] = sample(distribution, n) # e.g. vector is 100,000 ints
  }

  # 2) create n sample means
  sampleMeans <- c()
  for(j in 1:n){
    sum = 0
    for(k in 1:sampleSize){
      sum = sum + vectors[[k]][j] # step through COLUMN wise (the inner k changes faster)
    }
    sampleMeans[j] <- (sum/sampleSize) # get the average
  }
  return(sampleMeans)
}
```

1.1.3. Sample two random elements of myRunIfVec, take the mean of those two elements, and store the value in unifSampleMean_2. Repeat this step 100,000 times, so that you will have sample 200,000 elements from myRunIfVec and created 100,000 2-sample means in unifSampleMean_2. Plot a histogram of unifSampleMean_2, describe the shape of the sampling distribution of the mean, and note if it is different from the population distribution.

```
unifSampleMean_2 <- c()
unifSampleMean_2 <- getNSampleMeans(100000, 2, myRunIfVec)
hist(unifSampleMean_2, xlab="Value from Sample Means unifSampleMean_2", main="Sample Mean distribution
```

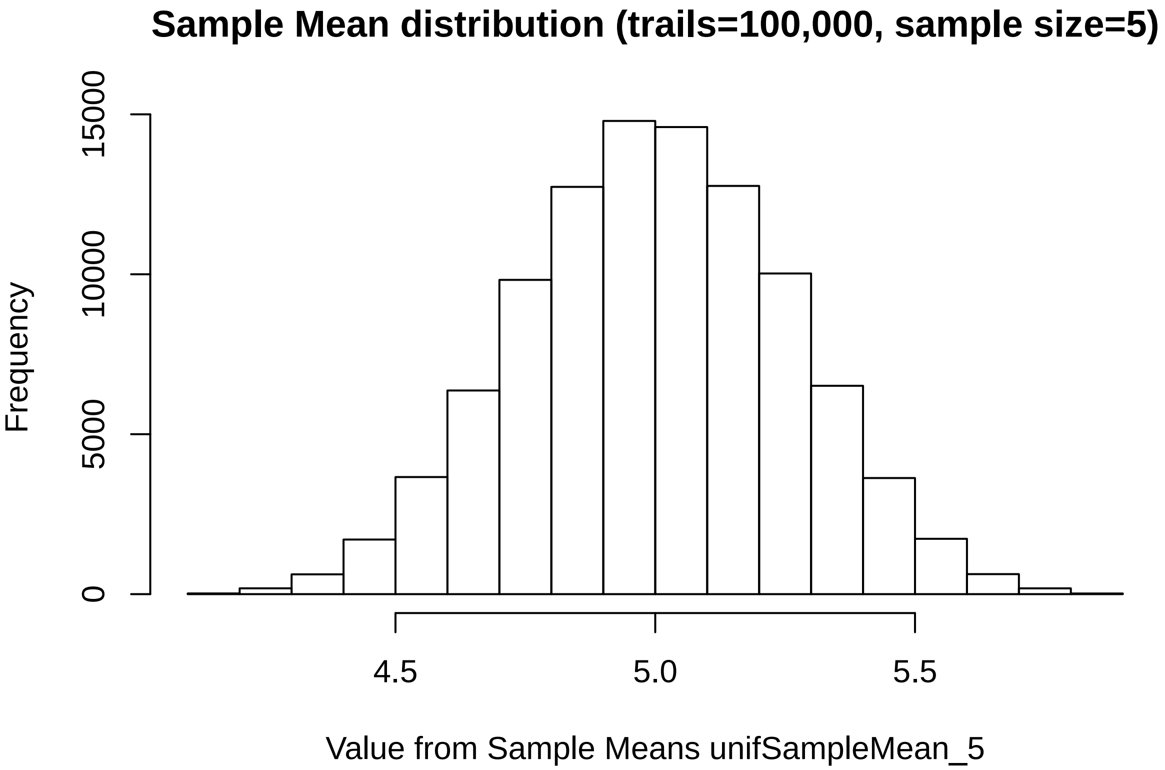


Wow!

The distribution appears like a triangle, with a the highest density centered around the mean of 5.0.

1.1.4. Repeat (1.1.3), but this time sample five random elements, take the mean, and store the value in `unifSampleMean_5`. Repeat this step 100,000 times. Plot a histogram of `unifSampleMean_5`, describe the shape of the sampling distribution of the mean, and note if it is different from the population distribution.

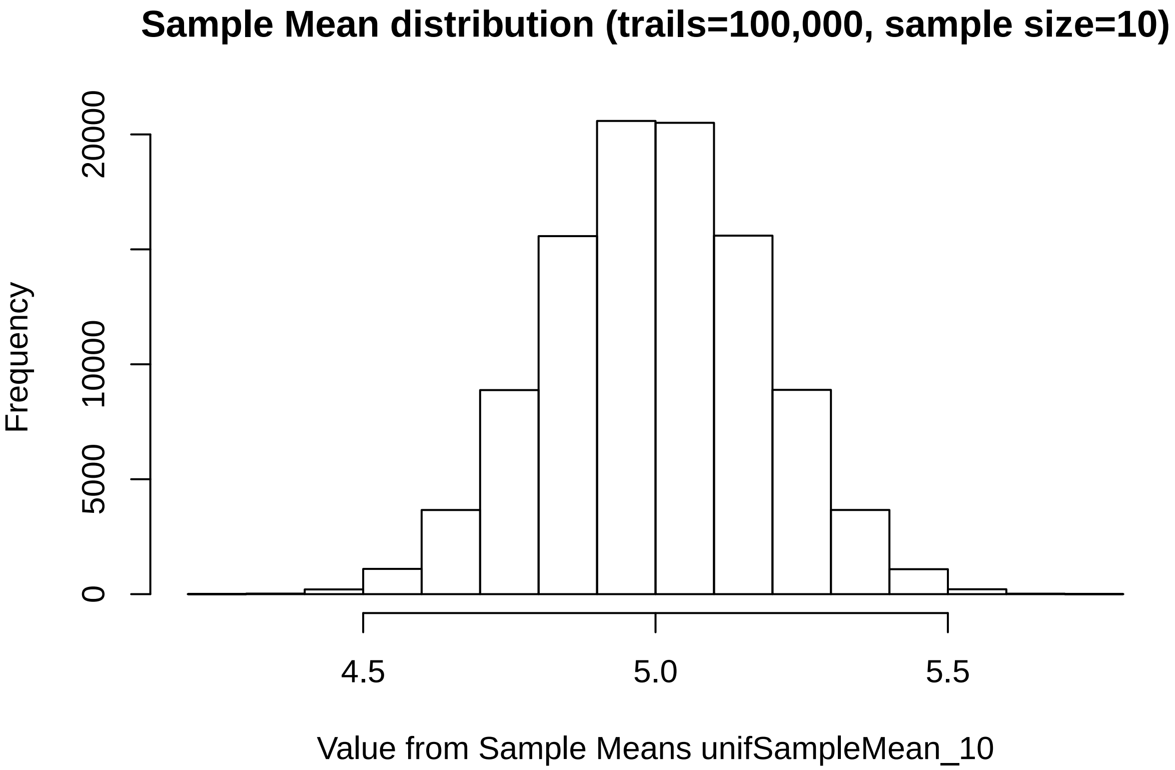
```
unifSampleMean_5 <- c()
unifSampleMean_5 <- getNSampleMeans(100000, 5, myRunIfVec)
hist(unifSampleMean_5, xlab="Value from Sample Means unifSampleMean_5", main="Sample Mean distribution
```



The graph looks 'tighter' around the mean, as the variance is lower due to the increase of sample size from 2 to 5. It is beginning to approximate the normal distribution.

1.1.5. Repeat (1.1.4), but this time sample ten random elements, take the mean, and store the value in `unifSampleMean_10`. Repeat this step 100,000 times. Plot a histogram of `unifSampleMean_10`, describe the shape of the sampling distribution of the mean, and note if it is different from the population distribution.

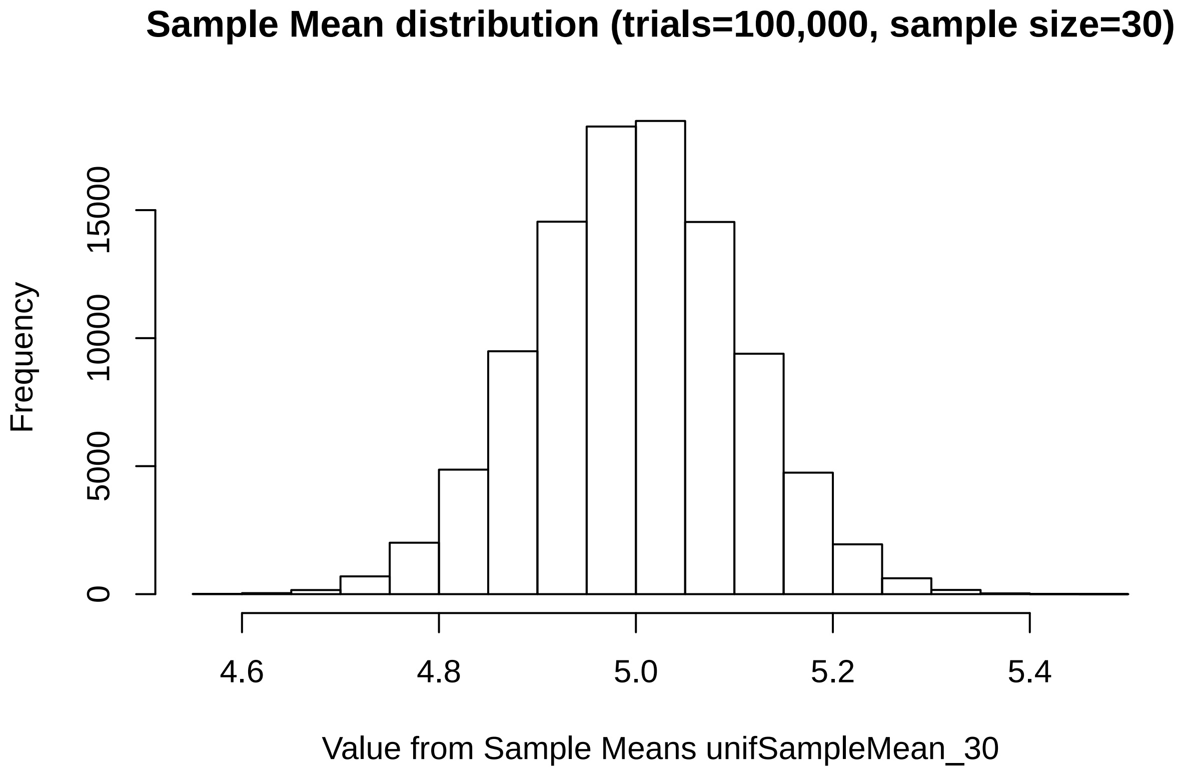
```r
unifSampleMean_10 <- c()
unifSampleMean_10 <- getNSampleMeans(100000, 10, myRunIfVec)
hist(unifSampleMean_10, xlab="Value from Sample Means unifSampleMean_10", main="Sample Mean distributio
```



It appears to be much tighter around the mean compared to the previous graphs in 1.1.4 and 1.1.3. There are far fewer samples with a sample means further out.

1.1.6. Repeat (1.1.5), but this time sample thirty random elements, take the mean, and store the value in `unifSampleMean_30`. Repeat this step 100,000 times. Plot a histogram of `unifSampleMean_30`, describe the shape of the sampling distribution of the mean, and note if it is different from the population distribution.

```
unifSampleMean_30 <- c()
unifSampleMean_30 <- getNSampleMeans(100000, 30, myRunIfVec)
hist(unifSampleMean_30, xlab="Value from Sample Means unifSampleMean_30", main="Sample Mean distributio
```



This graph doesn't look wildly different from the previous graph of 1.1.5. It still looks normally distributed and has a smaller variance than in 1.1.3 and 1.1.2.

## Question 1.2

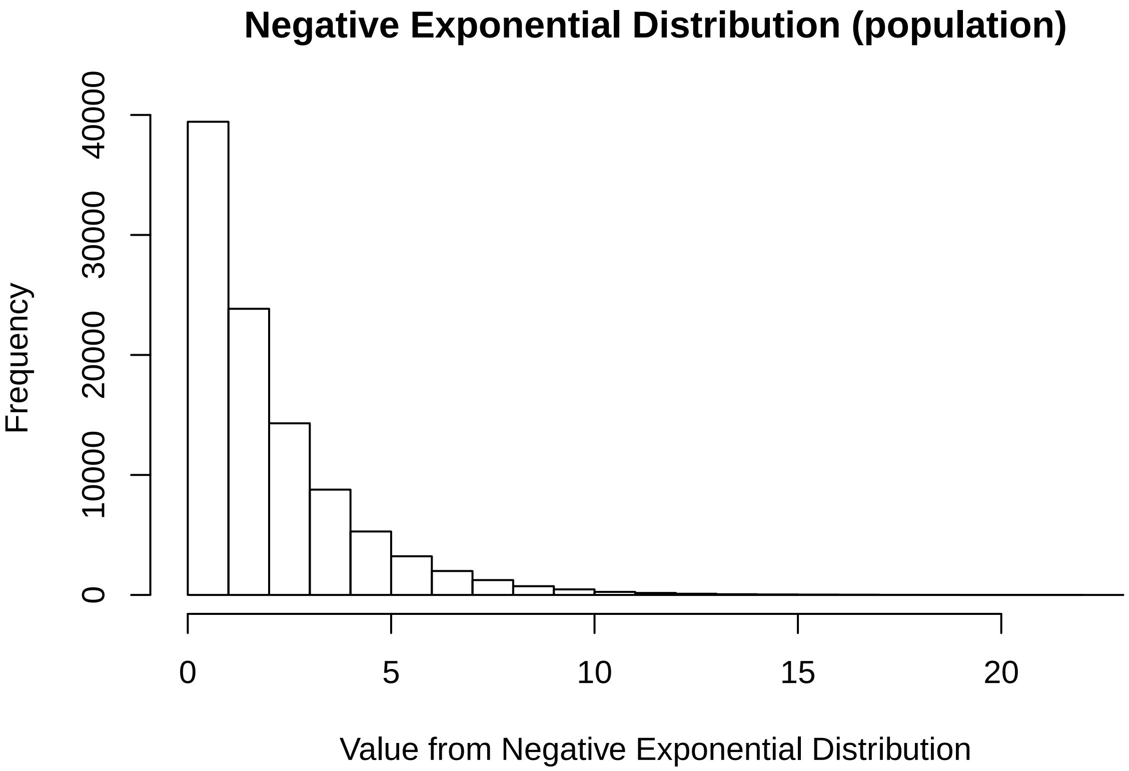
Repeat **all** steps of Question #1, but this time initializing the process with a sample of 10,000,000 random variates from a negative exponential distribution with $\lambda = 0.5$. Comment on how the results of this exercise differ from those of the previous question.

```
negativeExponentialDistribution <- rexp(10000000, rate=0.5)
```

### .2

Sample 100,000 from the population distribution and plot.

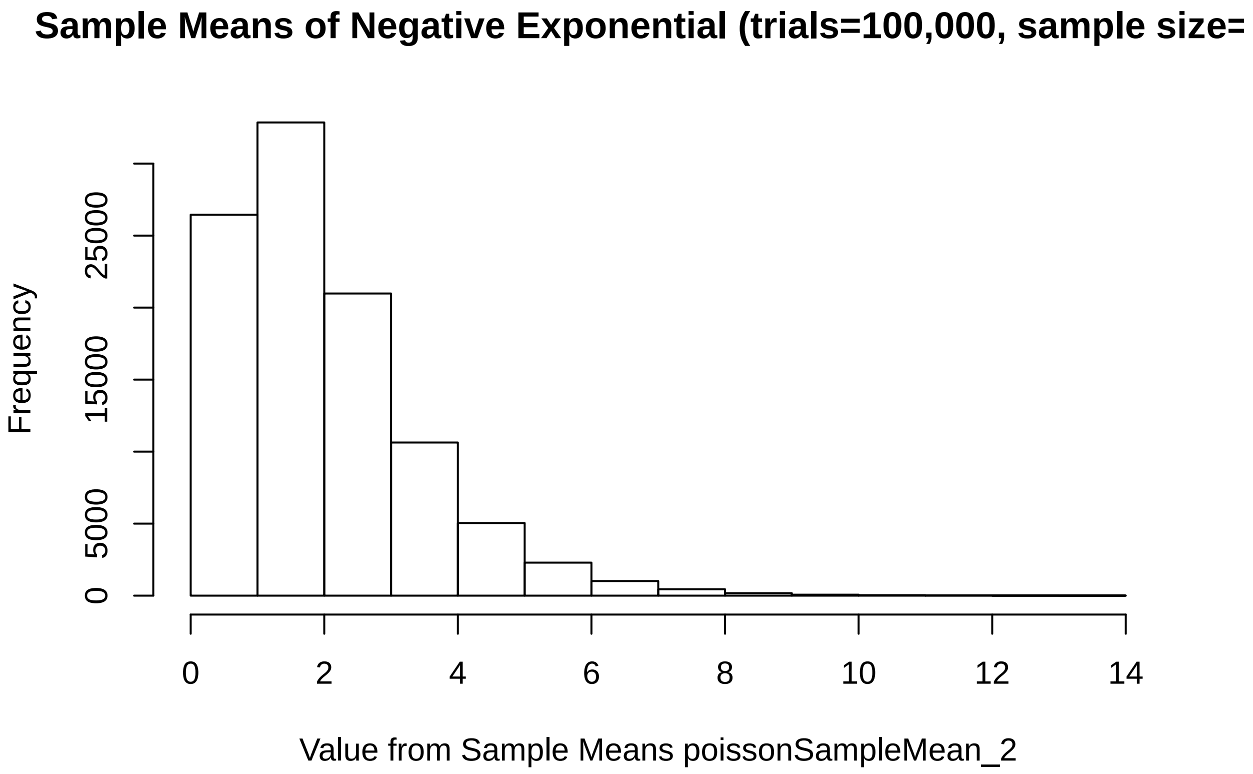
```
hist(sample(negativeExponentialDistribution, 100000), xlab="Value from Negative Exponential Distributio
```



### .3

Trials = 100,000, Sample size = 2

```
poissonSampleMean_2 <- c()
poissonSampleMean_2 <- getNSampleMeans(100000, 2, negativeExponentialDistribution)
hist(poissonSampleMean_2, xlab="Value from Sample Means poissonSampleMean_2", main="Sample Means of Nega
```

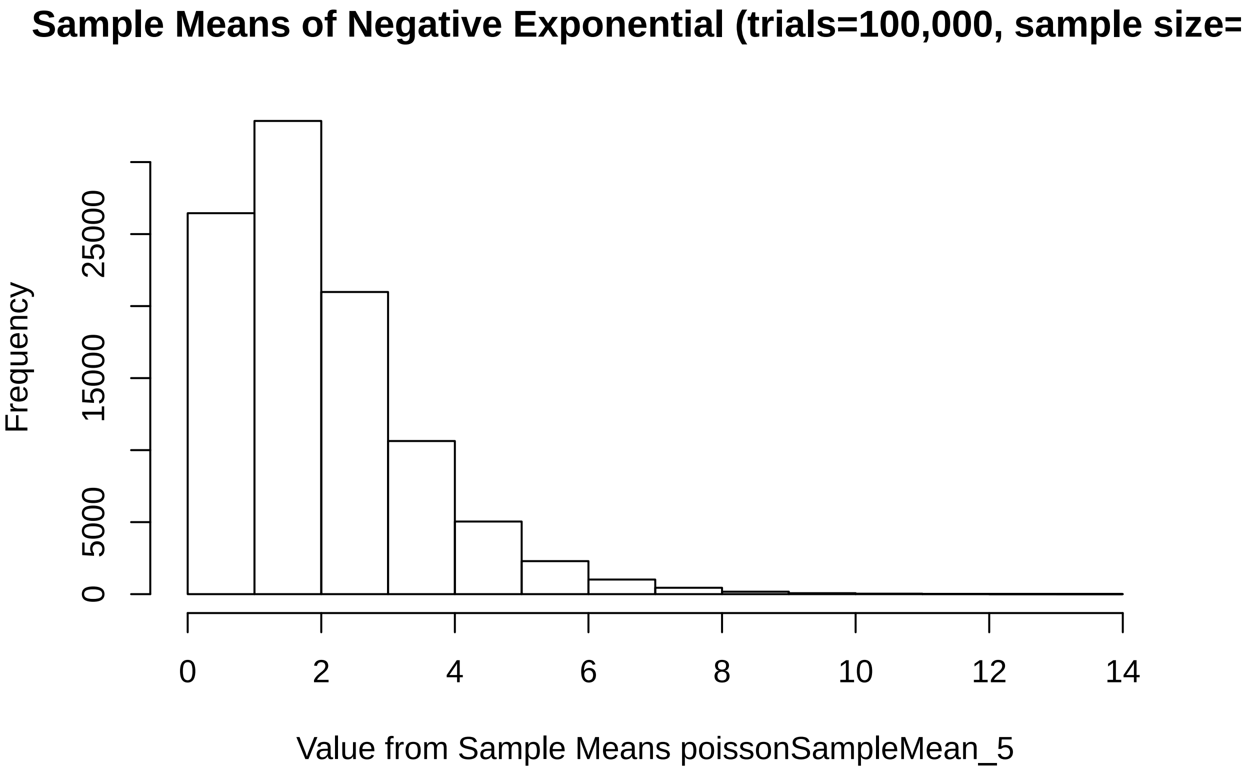


This distribution looks similar to the population distribution - it doesn't look like the normal distribution.

### .4

Trials = 100,000, Sample size = 5

```
poissonSampleMean_5 <- c()
poissonSampleMean_5 <- getNSampleMeans(100000, 5, negativeExponentialDistribution)
hist(poissonSampleMean_2, xlab="Value from Sample Means poissonSampleMean_5", main="Sample Means of Nega
```

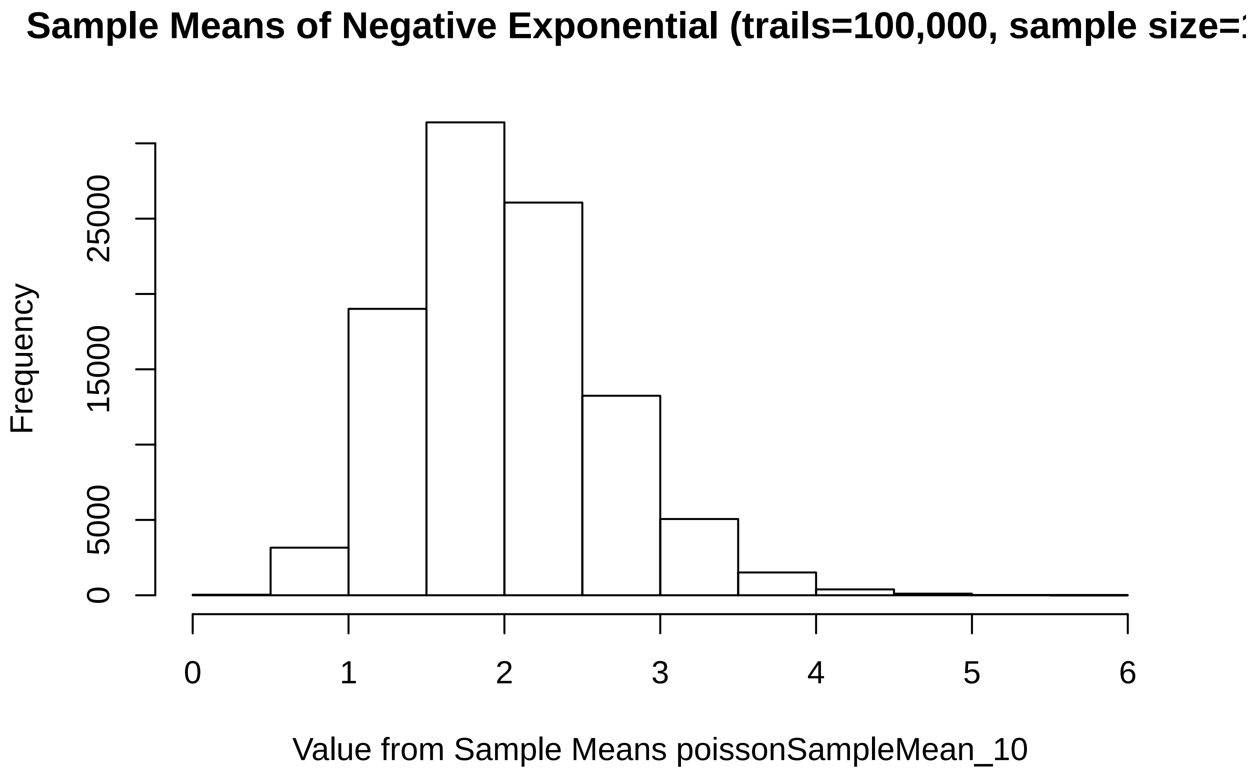


The graph still doesn't look like the normal distribution.

### .5

Trials = 100,000, Sample size = 10

```
poissonSampleMean_10 <- c()
poissonSampleMean_10 <- getNSampleMeans(100000, 10, negativeExponentialDistribution)
hist(poissonSampleMean_10, xlab="Value from Sample Means poissonSampleMean_10", main="Sample Means of Ne
```

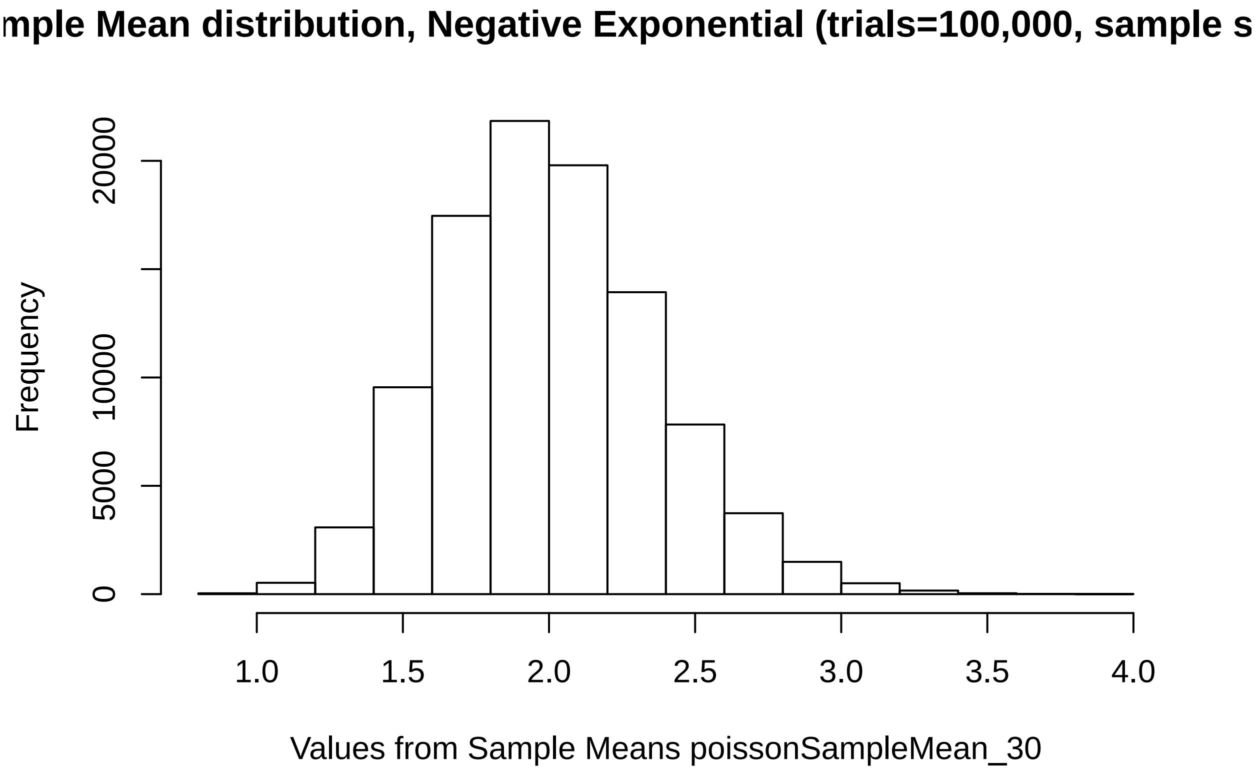


It's beginning to look more normal!

### .6

Trials = 100,000, Sample size = 30

```
poissonSampleMean_30 <- c()
poissonSampleMean_30 <- getNSampleMeans(100000, 30, negativeExponentialDistribution)
hist(poissonSampleMean_30, xlab="Values from Sample Means poissonSampleMean_30", main="Sample Mean dist
```

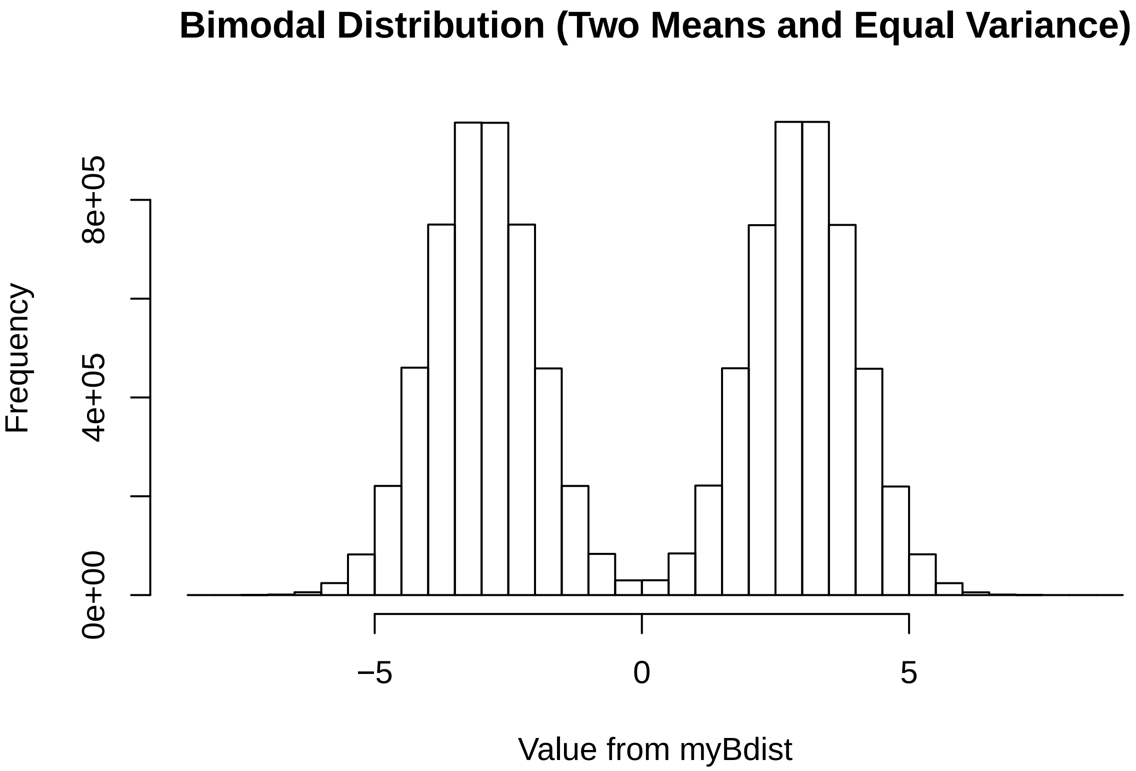


It looks shockingly like the normal distribution.

## Question 1.3

1.3.1 Create a **single** vector with 5,000,000 random variates from a $\sim \mathcal{N}\{-3, 1\}$, 5,000,000 random variates from a $\sim \mathcal{N}\{3, 1\}$ and store these values in the vector `myBdist`. Create a histogram and describe the distribution. ### Function Definition

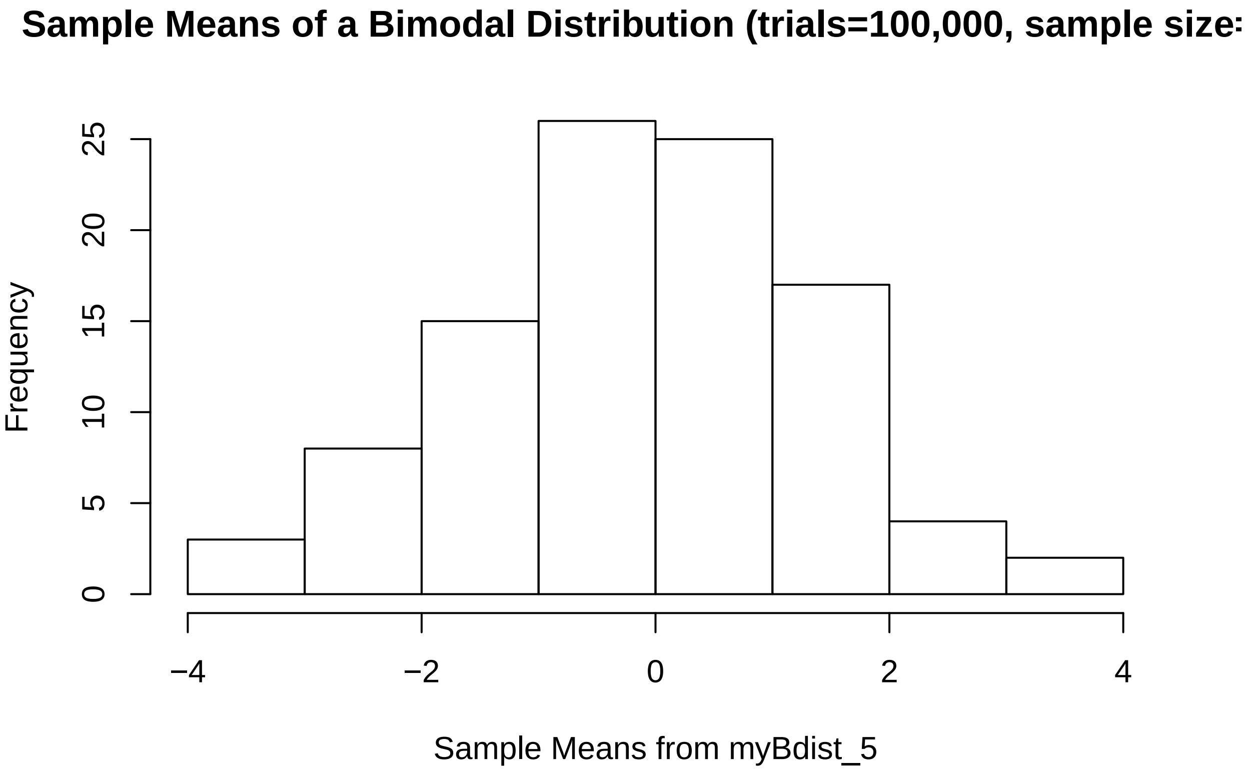
```
a <- rnorm(5000000, mean=-3, sd=1)
b <- rnorm(5000000, mean=3, sd=1)
myBdist <- c(a,b)
hist(myBdist, xlab="Value from myBdist", main="Bimodal Distribution (Two Means and Equal Variance)")
```



The distribution has two different means (-3 and 3) and equal variance. It has two distinct peaks, unlike the standard normal distribution, which typically only has one peak.

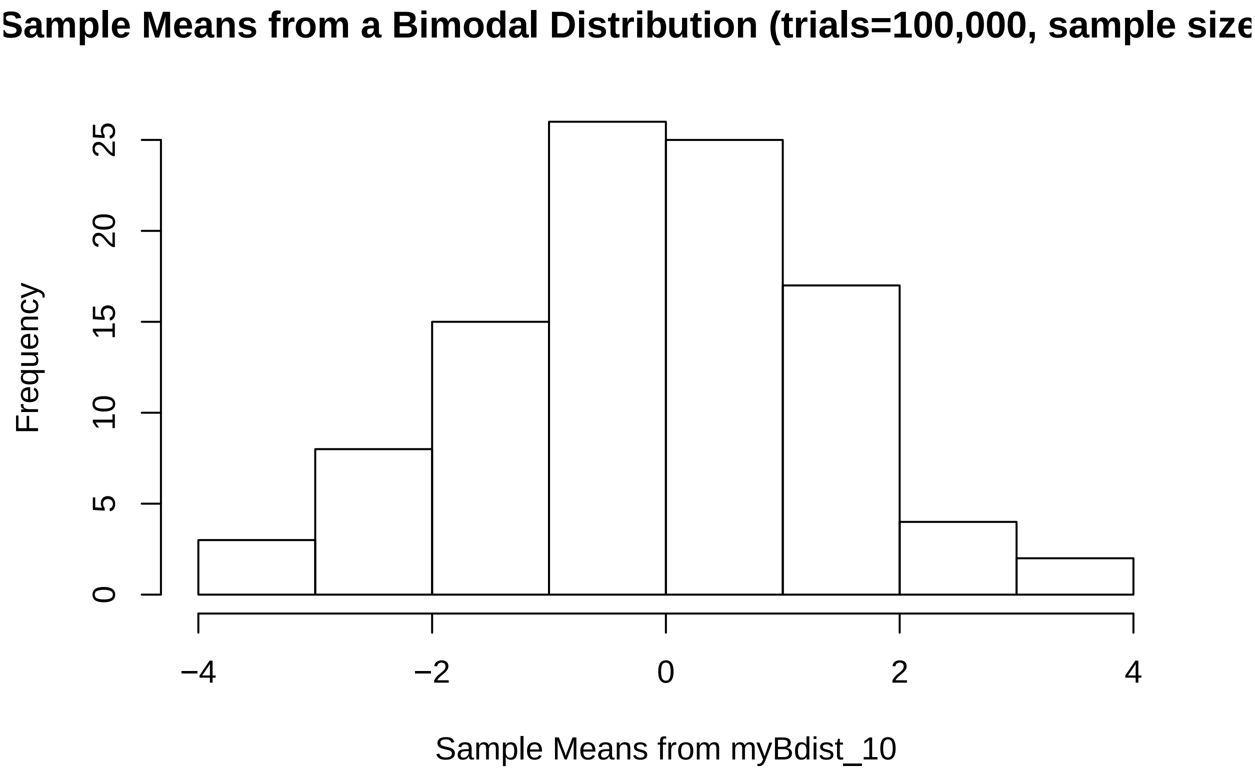
1.3.2 Sample five random elements of `myBdist`, take the mean of those five elements, and store the value in `myBdist_5`. Repeat this step 100,000 times, so that you will have sample 200,000 elements from `myBdist` and created 100,000 5-sample means in `myBdist_5`. Plot a histogram of `myBdist_5`, describe the shape of the sampling distribution of the mean, and note if it is different from the population distribution.

```
myBdist_5 <- getNSampleMeans(100, 5, myBdist)
hist(myBdist_5, xlab="Sample Means from myBdist_5", main="Sample Means of a Bimodal Distribution (trials
```

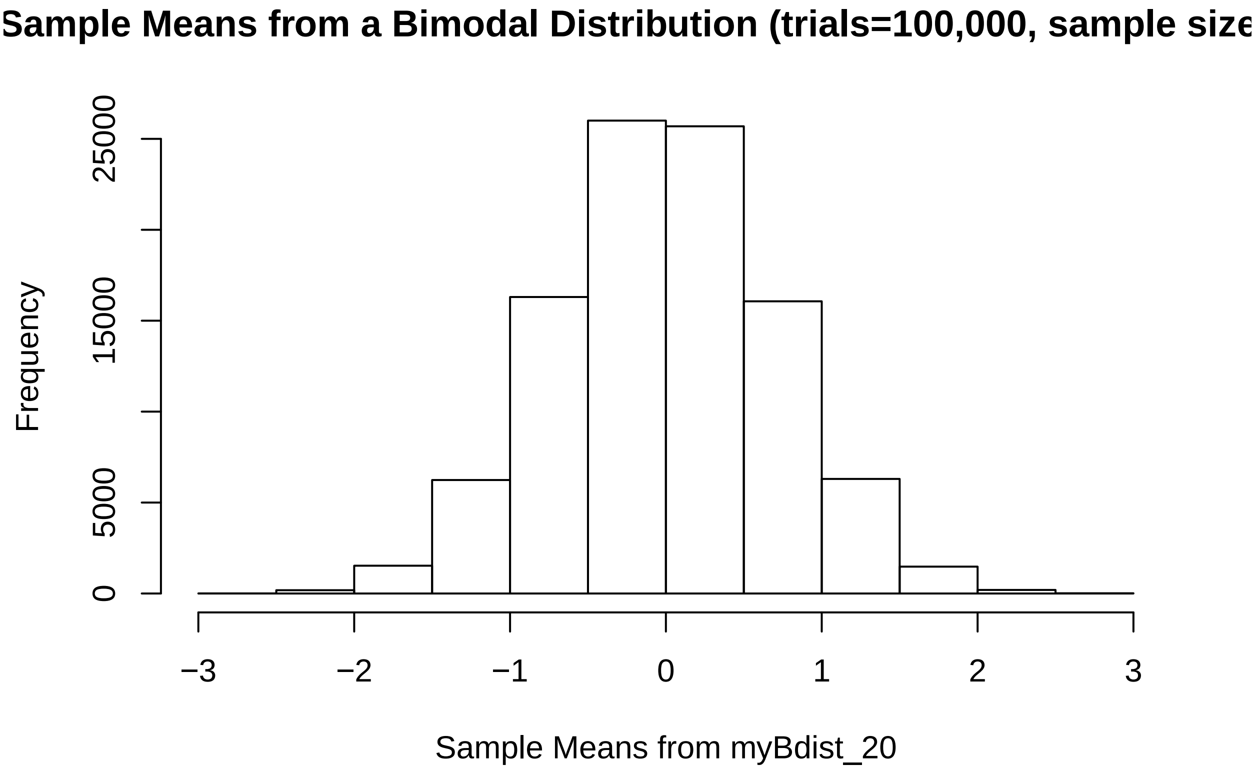


1.3.3 Repeat 1.3.2 with sample means of 10, 20 and thirty, creating histograms of each as you go along.

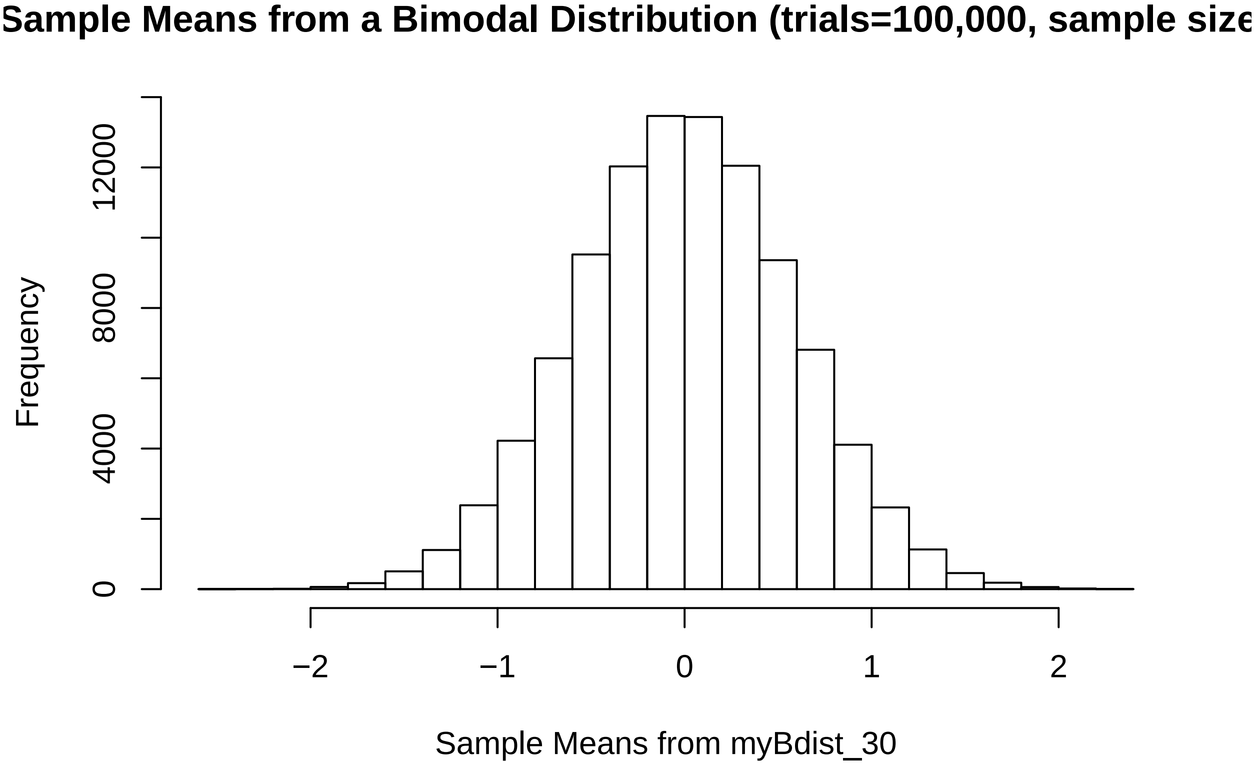
```
# Sample Size = 10
myBdist_10 <- getNSampleMeans(100000, 10, myBdist)
hist(myBdist_5, xlab="Sample Means from myBdist_10", main="Sample Means from a Bimodal Distribution (tr
```



```
# Sample size = 20
myBdist_20 <- getNSampleMeans(100000, 20, myBdist)
hist(myBdist_20, xlab="Sample Means from myBdist_20", main="Sample Means from a Bimodal Distribution (t
```



```
# Sample size = 30
myBdist_30 <- getNSampleMeans(100000, 30, myBdist)
hist(myBdist_30, xlab="Sample Means from myBdist_30", main="Sample Means from a Bimodal Distribution (t
```



1.3.4 Write a short summary of what you have observed, and relate it to the theory you have learned in MSAN 504. What is this behavior called?

We observe that sample means approximate the normal distribution and become more tightly distributed around the true population mean as sample size increases, and that bimodal distributions can really screw up our approximations of the true mean.

These observations relate to the key concepts of The Central Limit Theorem and the violation of the normality assumption. First, the Central Limit Theorem states that, regardless of the underlying distribution, sample distribution of sample means starts to become normally distributed as the sample size increases. We clearly observe this phenomena in this report.

However, if the underlying is bimodal, the sample means generated can be wildly wrong. Hence, it is important to have an idea of the underlying population distribution and remember the assumption of normailty when using the Central Limit Theorem.