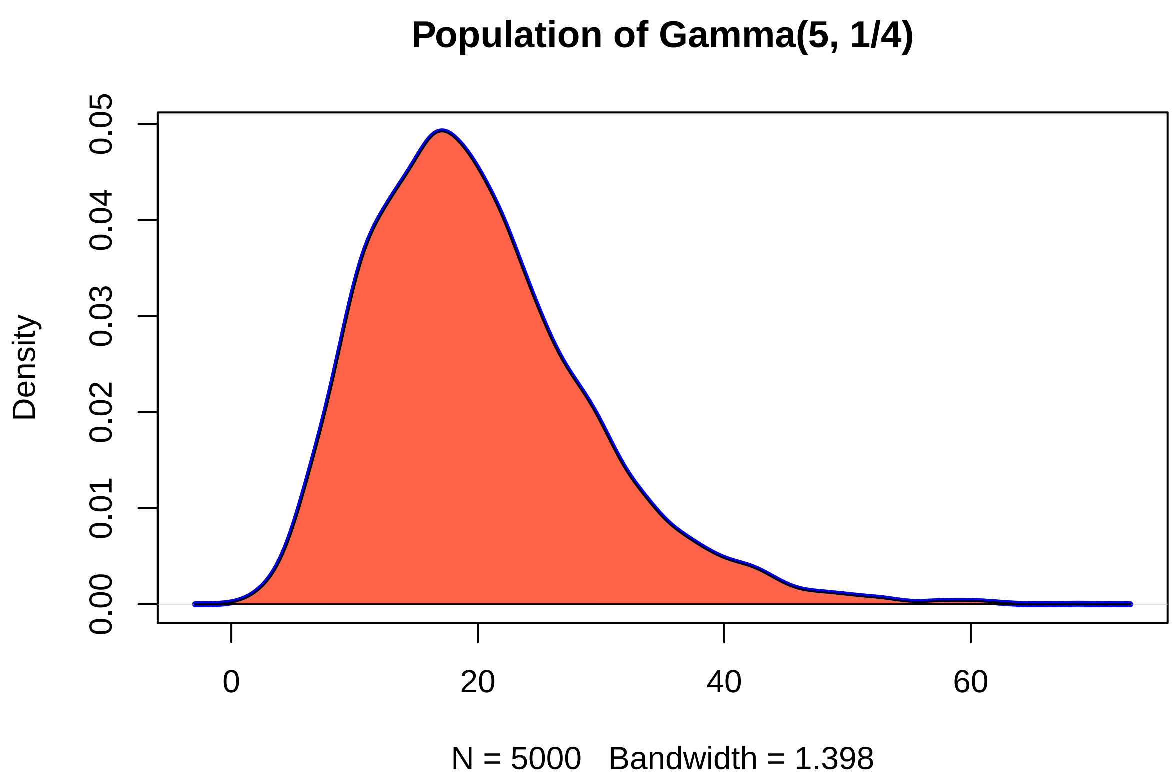
Advanced Statistical Inference: Assignment 2

ankit.tewari@estudiant.upc.edu

We're creating a population that has gamma distribution with paramters r=5 and lambda=1/4 Although it's unspecified about which parameter to take shape and which one to take scale, so we've tried to carry out the simulation by passing the values in such a way that the shape becomes 5 and rate becomes 1/4 Which in turn makes scale=1/rate i.e. scale=4

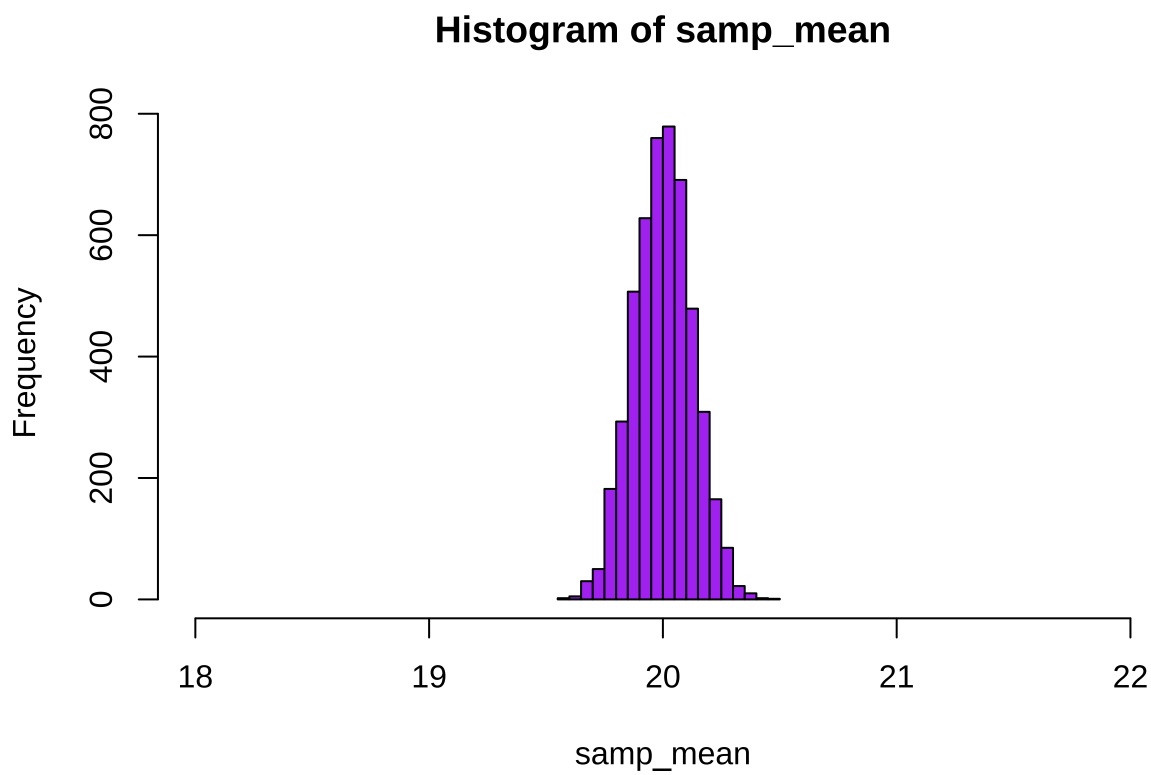
```r
set.seed(1234)
dist<-rgamma(5000, 5, 1/4)
mean_dist<-mean(dist)
sd_dist<-sd(dist)

plot(density(dist), col="blue", lwd=3, main = "Population of Gamma(5, 1/4)")
polygon(density(dist), col="tomato", border="black")
```



Now, we'll attempt to simulate the "Sampling Distribution of Means" based on the population that we've just created. This attempt is based on the fact that for a population distributed as Gamma(r,lambda), the sampling distribution of sample means of that population are distributed as Gamma(nr, n*lambda), where n is the population size under consideration. We'll first compute the important statistics for this sampling distribution and thereafter, we'll plot a histogram of the Sampling Distribution to visulize our results along with fitting a density function to it.

```r
set.seed(1234)
samp_mean<-rgamma(5000, 5000*5, 5000*(1/4))
```

```
samp_mu<-mean(samp_mean)
samp_sd<-sd(samp_mean)

hist(samp_mean, col = "purple", xlim = c(18,22), breaks = 20)
```



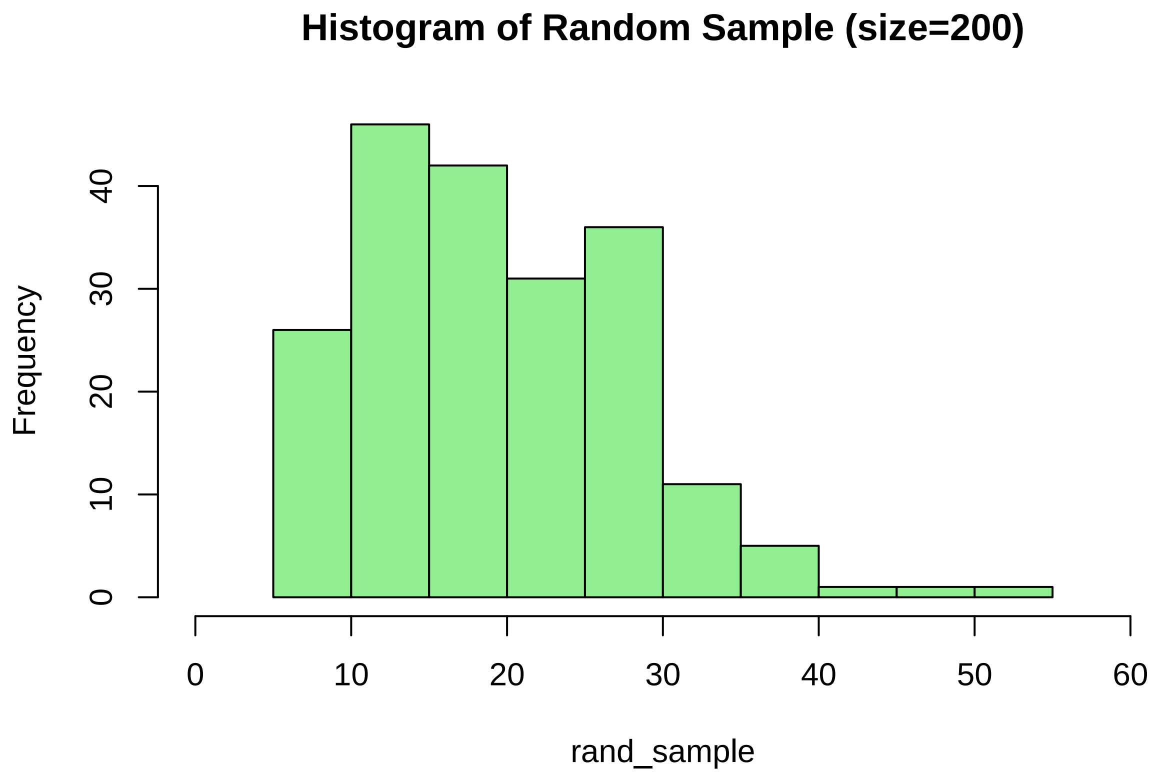
Below, we have desribed the sampling distribution of sample means using some summary statistics.

```
print(summary(samp_mean))
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   19.57   19.92   20.00   20.00   20.08   20.45
```

Now, we will be creating a random sample of size 200 from the population. This sample will be the base for simulating our bootstrap distribution.

```
rand_sample<-sample(dist, 200, replace = F)

mean_rand_sample<-mean(rand_sample)
sd_rand_sample<-sd(rand_sample)
hist(rand_sample, col = "lightgreen", main = "Histogram of Random Sample (size=200)", axes=T,
     xlim = c(0,60))
```



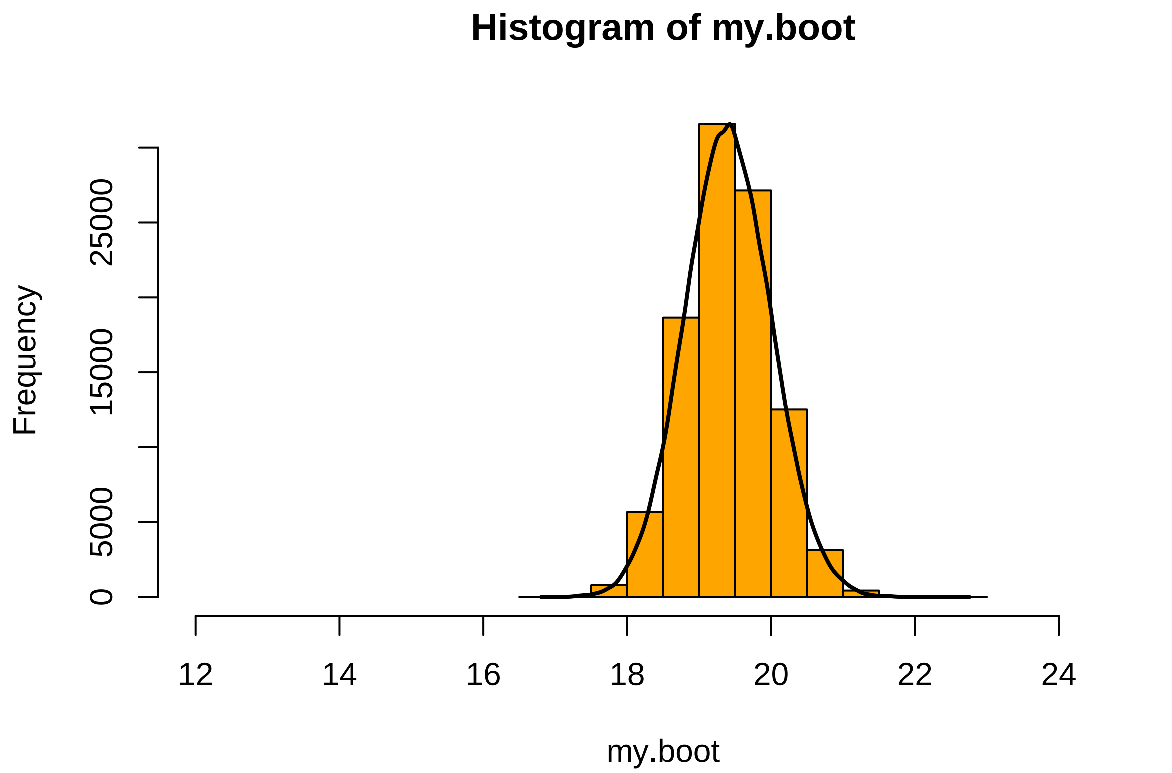
We have exhibited the summary statistics for this random sample in the table below-

```
summary(rand_sample)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   5.025  12.714  18.897  19.403  25.346  51.816
```

Bootstrap Distribution We will use the method of resampling i.e. sampling with replacement for creating a bootstrap distribution of our means.

```
N<-10^5
my.boot<-numeric(N)

for(i in 1:N){
  x<-sample(rand_sample, 200, replace = T)
  my.boot[i]<-mean(x)
}
mean_boot<-mean(my.boot)
sd_boot<-sd(my.boot)

hist(my.boot, col = "Orange", xlim = c(12,25))
par(new=T)
plot(density(my.boot), main = "", xlab = "", ylab = "", lwd=2, axes=F, xlim=c(12,25))
```



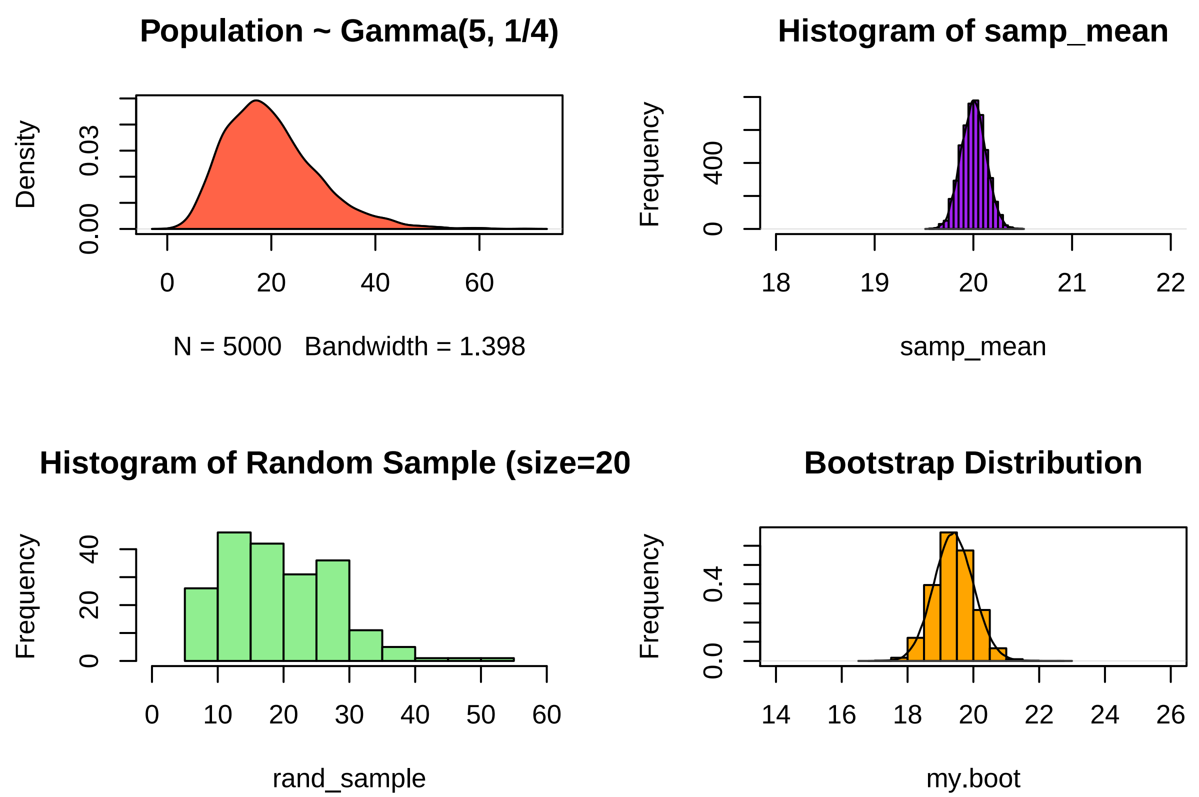
Comparison table In this comparison table, we can see the results associated with our different distributions which can serve the purpose of numerical comparison. In the next section, we have made an attempt for graphical comparison.

```
r1<-c("Population", mean_dist, sd(dist))
r2<-c("Sampling Distribution of Mean", mean_value=samp_mu, sd_value=samp_sd)
r3<-c("Sample (size=200)", mean_rand_sample, sd_rand_sample)
r4<-c("Bootstrap Distribution", mean_boot, sd_boot)

result<- as.data.frame(rbind(r1, r2, r3, r4))
names(result)<-c("", "Mean", "Standard Deviation")
rownames(result) <- c()

result
```

```
##                                              Mean Standard Deviation
## 1                    Population 20.0110700360581   8.89725334298631
## 2 Sampling Distribution of Mean 20.0016304192706  0.124507939779036
## 3             Sample (size=200) 19.4026149986386   8.52175059735526
## 4        Bootstrap Distribution 19.4048023356967  0.601708386963664
```

```
par(mfrow=c(2,2))


plot(density(dist), main = "Population ~ Gamma(5, 1/4)")
polygon(density(dist), col="tomato", border="black")
```

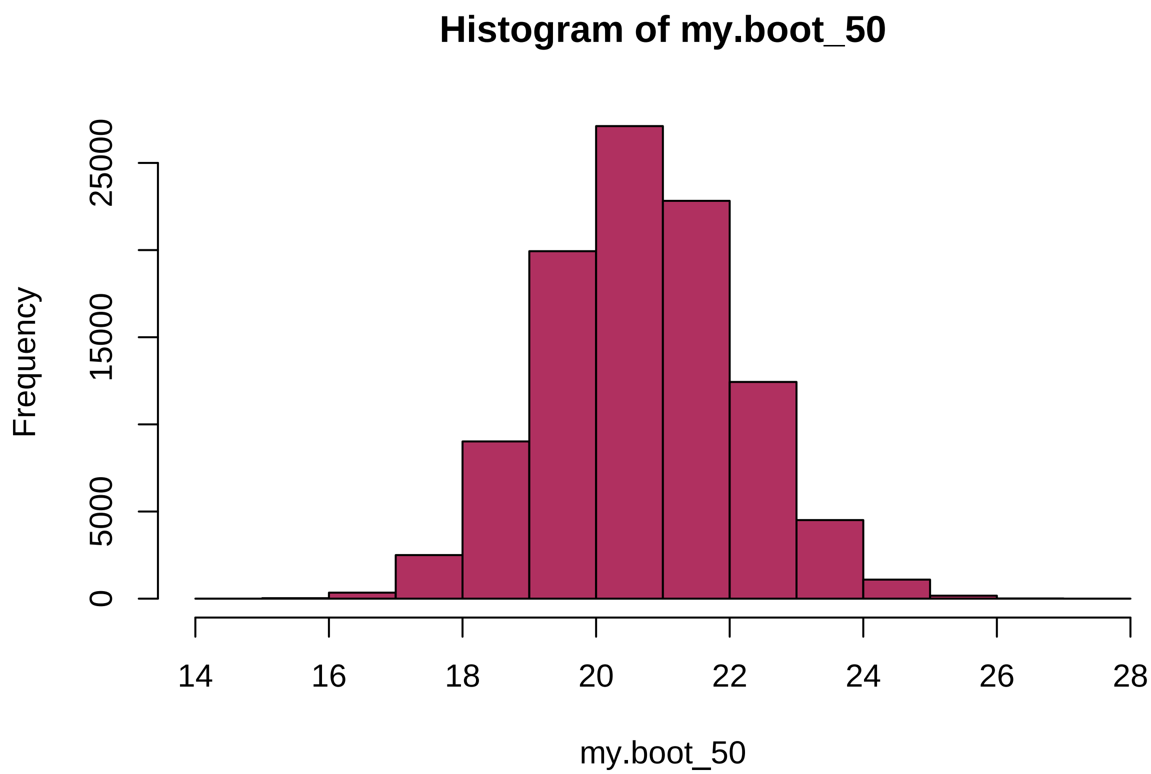
```
hist(samp_mean, col = "purple", xlim = c(18,22), breaks = 20)
par(new=T)
plot(density(samp_mean), main = "", xlab = "", ylab = "", lwd=1, xlim=c(18,22), axes=F)

hist(rand_sample, col = "lightgreen", main = "Histogram of Random Sample (size=200)",
     axes=T, xlim = c(0,60))

hist(my.boot, col = "Orange", axes = F, main = "Bootstrap Distribution", xlim=c(14,26),
     breaks = 20)
par(new=T)
plot(density(my.boot), main = "", xlab = "", ylab = "", lwd=1, xlim=c(14,26))
```



Now we'll attempt to simulate the bootstrap distribution for a sample size of 50. The code below exhibits this process.

```
rand_sample_50<-sample(dist, 50, replace = F)

mean_rand_sample_50<-mean(rand_sample_50)
sd_rand_sample_50<-sd(rand_sample_50)


N<-10^5
my.boot_50<-numeric(N)

for(i in 1:N){
```

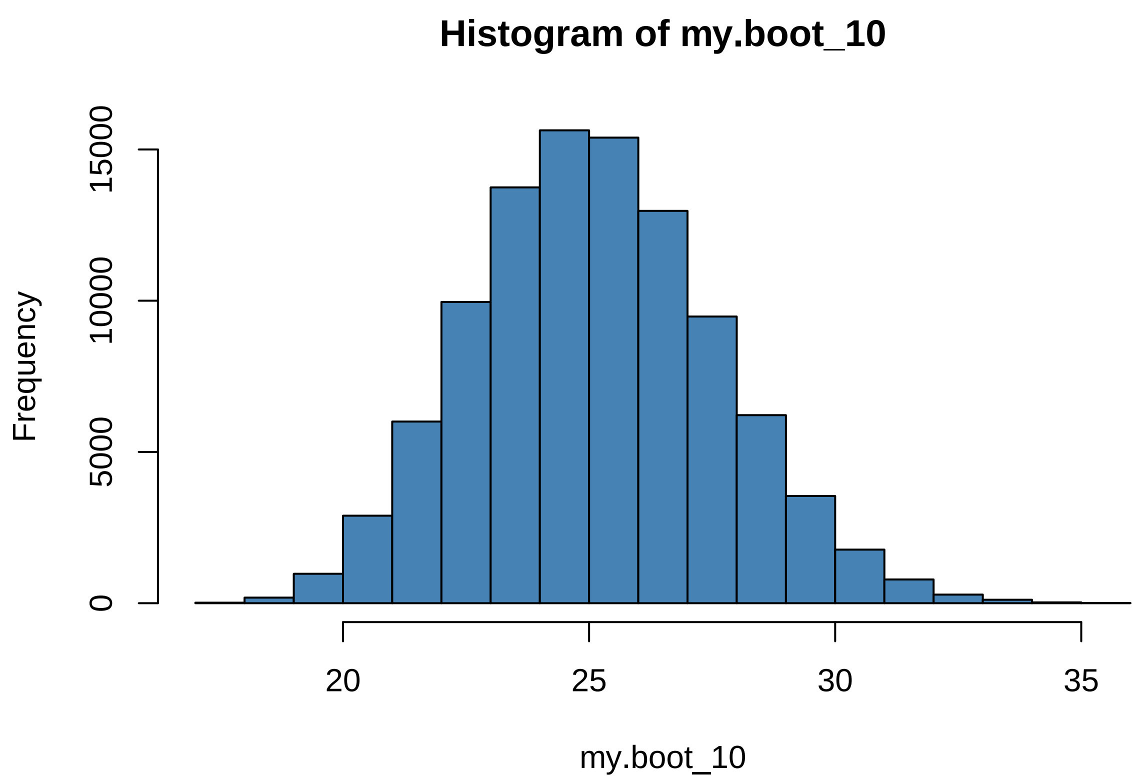
```
  x<-sample(rand_sample_50, 50, replace = T)
  my.boot_50[i]<-mean(x)
}
mean_boot_50<-mean(my.boot_50)
sd_boot_50<-sd(my.boot_50)

hist(my.boot_50, col = "maroon")
```



Next, we'll attempt to simulate the bootstrap distribution for a sample size of 10. The code below exhibits this process.

```
rand_sample_10<-sample(dist, 10, replace = F)

mean_rand_sample_10<-mean(rand_sample_10)
sd_rand_sample_10<-sd(rand_sample_10)


N<-10^5
my.boot_10<-numeric(N)

for(i in 1:N){
  x<-sample(rand_sample_10, 10, replace = T)
  my.boot_10[i]<-mean(x)
}
mean_boot_10<-mean(my.boot_10)
sd_boot_10<-sd(my.boot_10)
```

```r
hist(my.boot_10, col = "steel blue")
```



Finally, let us compare the three bootstrap distributions on the basis of the quantiles. We've compared all the three bootstrap distributions based on sample size of 200, 50 and 10 respectively.

```r
quantile(my.boot,c(0.1,0.9))
```

```
##      10%      90%
## 18.63795 20.17918
```

```r
quantile(my.boot_50,c(0.1,0.9))
```

```
##      10%      90%
## 18.85880 22.55807
```

```r
quantile(my.boot_10,c(0.1,0.9))
```

```
##      10%      90%
## 21.99089 28.38350
```