

PEfthimion_Homework4

The Central Limit Theorem states that if you have a sufficient number of randomly selected, independent samples, the means of those samples will follow a normal distribution, even if the population we are sampling does not.

The first thing that we need in a bootstrap is a sample. Here, we create a sample with a normal distribution and find its mean.

```
x <- rnorm(50, 22, 5)
xbar <- mean(x)
```

Next we need to say how many simulations we will be doing. Here we say 1000. We also make sure that R knows that the number of simulations is a numeric value.

```
nsims <- 1000
bootnorm <- numeric(nsims)
```

Now, for each $nsim = 1000$ simulations, we take a sample from x . Then, we save the average of that sample to $bootnorm$. We will use these averages to prove if the Central Limit Theory is true for x .

```
for(i in 1:nsims){
  temp <- sample(x, 50, replace = TRUE)
  bootnorm[i] <- mean(temp)
}
```

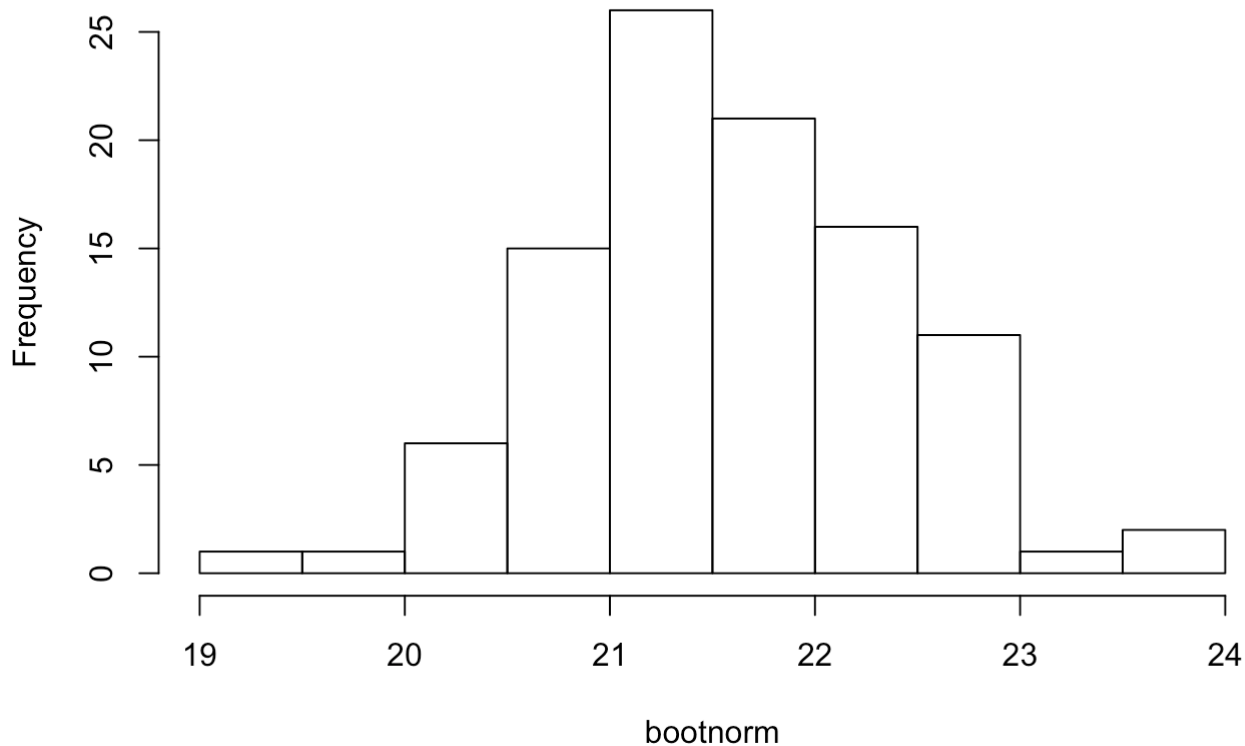
Now let's check to see if the Central Limit Theory applies.

```
summary(bootnorm)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    19.20   21.04   21.53   21.59   22.22   23.93
```

```
hist(bootnorm)
```

Histogram of bootnorm



And let's double check with the bias to see any difference between this sample and the theoretical mean

```
est.bias <- xbar - mean(bootnorm)
est.bias
```

```
## [1] 0.04303727
```

We can see here there is some, but very little bias. Thus we prove the Central Limit Theorem.

Now lets try this again with a much larger sample size and see how much more normalized the sample becomes.

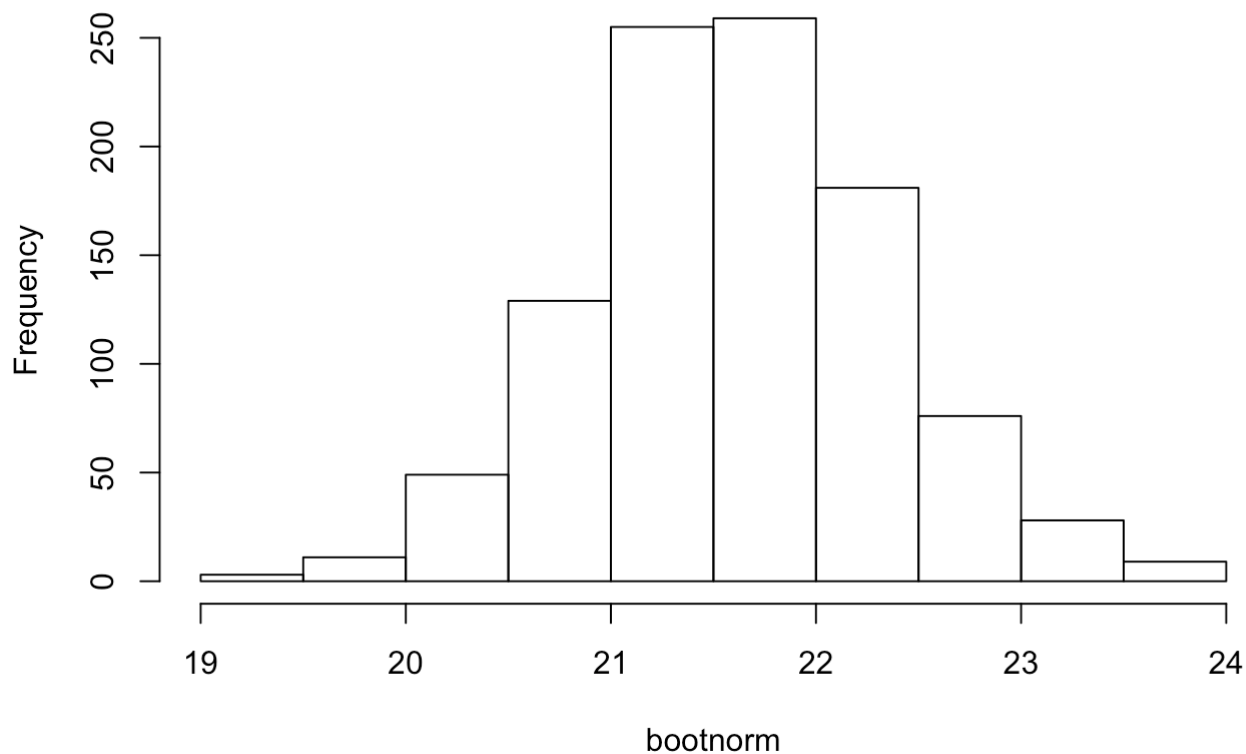
Other than changing the sample size, so nsims from 100 to 1000.

```
nsims <- 1000
bootnorm <- numeric(nsims)
for(i in 1:nsims){
  temp <- sample(x, 50, replace = TRUE)
  bootnorm[i] <- mean(temp)
}
summary(bootnorm)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  19.41   21.12   21.59   21.62   22.12   23.86
```

```
hist(bootnorm)
```

Histogram of bootnorm



```
est.bias <- xbar - mean(bootnorm)
est.bias
```

```
## [1] 0.01065105
```

As we can see here, the histogram looks a lot more normal for a sample for size 1000 than it did for 100. This is consistent with the Central Limit Theorem.

Now we will illustrate the central limit theorem with the exponential distribution instead of with a normal distribution. This would be possible because even if the data itself is not normally distributed, the means of the data will be. The bootstrap method uses the means of each sample taken so we will be able to use the exponential distribution to illustrate the central limit theorem.

Here is the one change in the code we will make for the exponential distribution. Since we are not doing a normal distribution we change `x` using `rnorm` into `rexp`. Also, like we did for the normal distribution, we will first be using 100 simulations with this example and 1,000 in the next.

```
x <- rexp(50)
xbar <- mean(x)
nsims <- 100
```

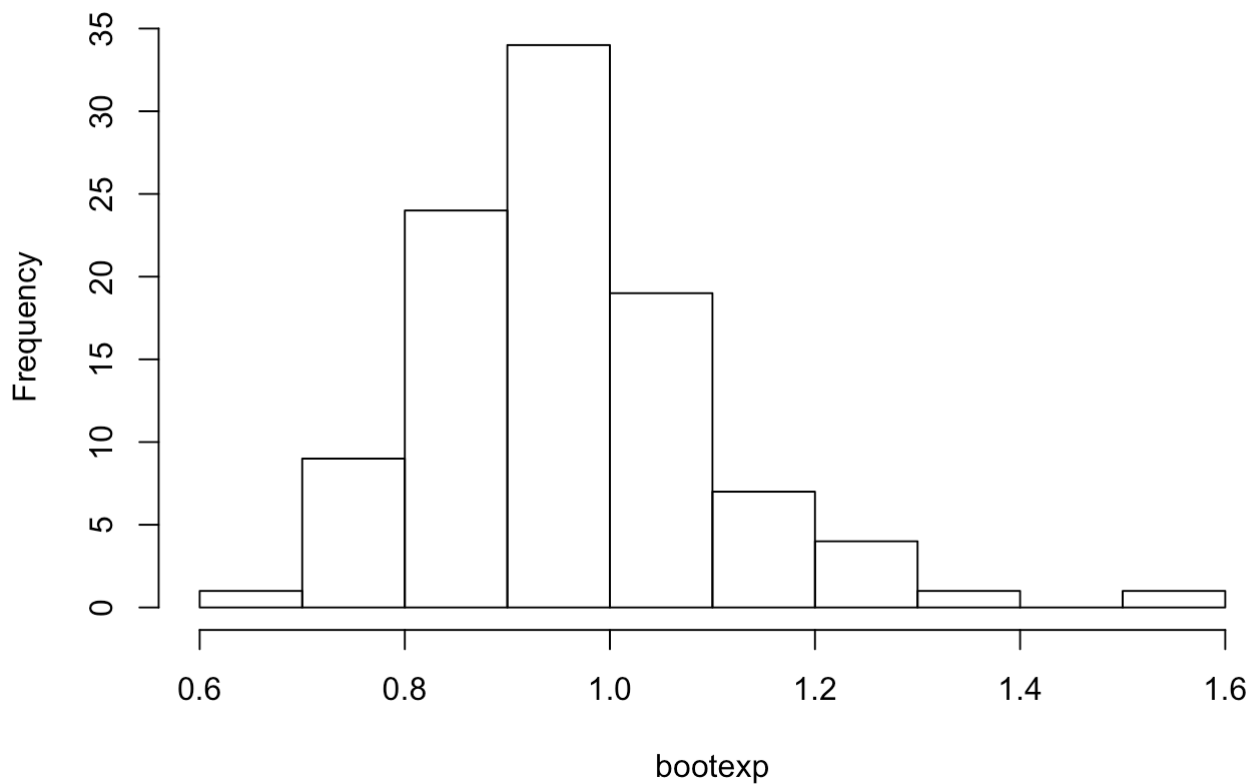
The rest of the code is the same as it was for the normal distribution as we explained above.

```
bootexp <- numeric(nsims)
for(i in 1:nsims){
  temp <- sample(x, 50, replace = TRUE)
  bootexp[i] <- mean(temp)
}
summary(bootexp)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.6947  0.8646  0.9339  0.9572  1.0270  1.5180
```

```
hist(bootexp)
```

Histogram of bootexp



```
est.bias <- xbar - mean(bootnorm)
est.bias
```

```
## [1] -20.64774
```

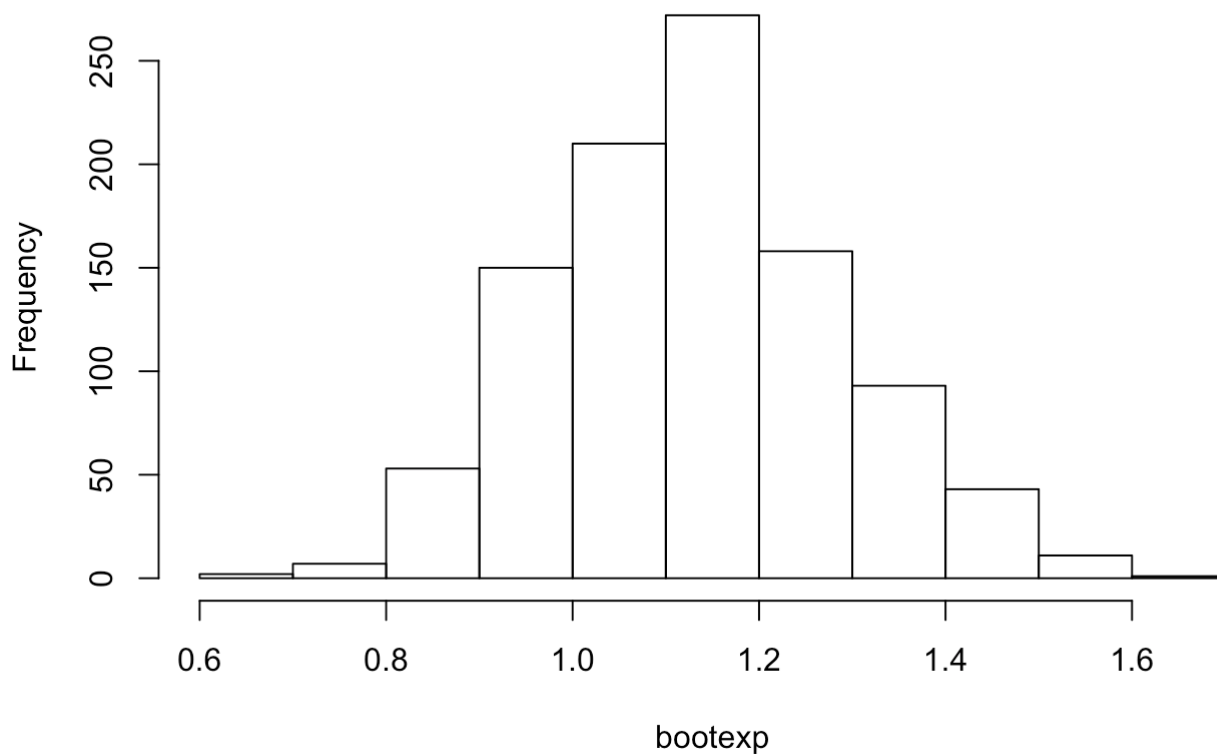
Now, we will do the same code again with the only change being doing 1000 simulations. According to the central limit theorem, this will make the data look more normal.

```
x <- rexp(50)
xbar <- mean(x)
nsims <- 1000
bootexp <- numeric(nsims)
for(i in 1:nsims){
  temp <- sample(x, 50, replace = TRUE)
  bootexp[i] <- mean(temp)
}
summary(bootexp)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.669   1.024   1.127   1.132   1.230   1.696
```

```
hist(bootexp)
```

Histogram of bootexp



This data looks much more normal with 1,000 simulations than it did when we only did 100. This is further proof of the central limit theorem.

```
est.bias <- xbar - mean(bootnorm)
est.bias
```

```
## [1] -20.48054
```