

## Problem 5

Yu Du

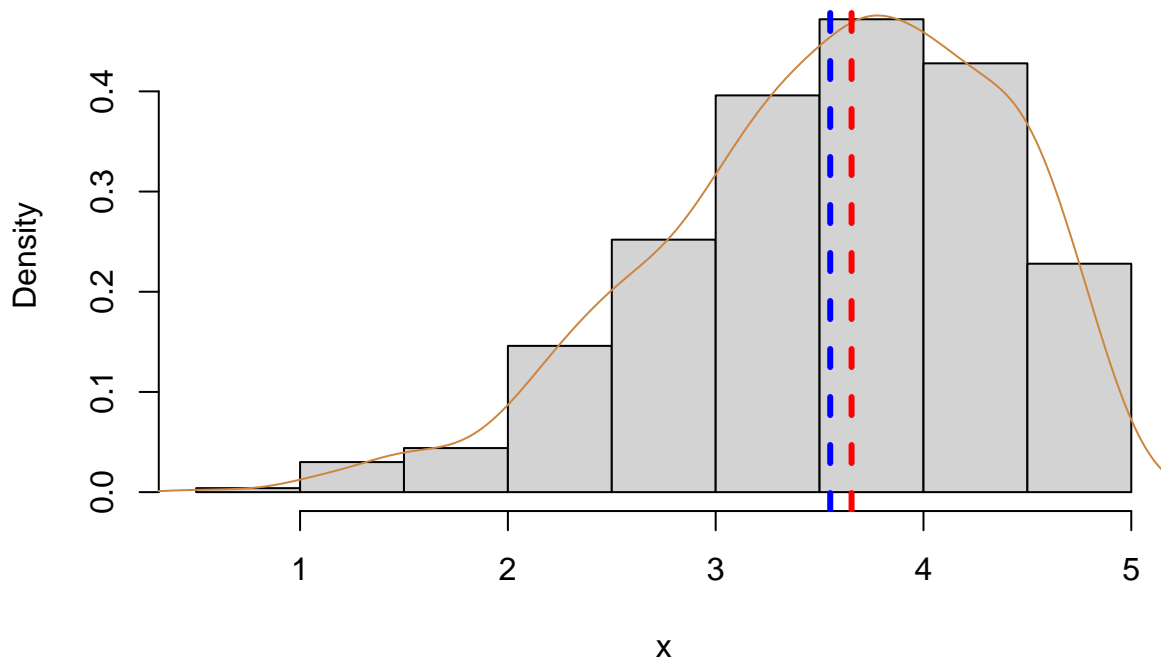
The answers for problem 1 to 4 are in another pdf file.

Problem 5: Select a skewed distribution from which to sample. Using R, demonstrate the convergence of the mean value of your samples to the Normal distribution. Assume you are making this demonstration to someone who has little or no statistical training. Your demonstration should take no more than 10 minutes. Along with your code, outline the commentary you would use in your demonstration.

First, using the `rbeta` function to generate 1000 random variables between 0 and 5. The distribution of these variables is left-skewed as below.

```
set.seed(123)
x=rbeta(1000,5,2)*5
hist(x, freq=FALSE)
lines(density(x), col='peru', lwd=1)
abline(v = c(mean(x),median(x)), col=c("blue", "red"), lty=c(2,2), lwd=c(3, 3))
```

**Histogram of x**



The mean, median and standard deviation of population.

```
round(mean(x),4)
```

```
## [1] 3.5512
```

```
round(median(x),4)
```

```
## [1] 3.6539
```

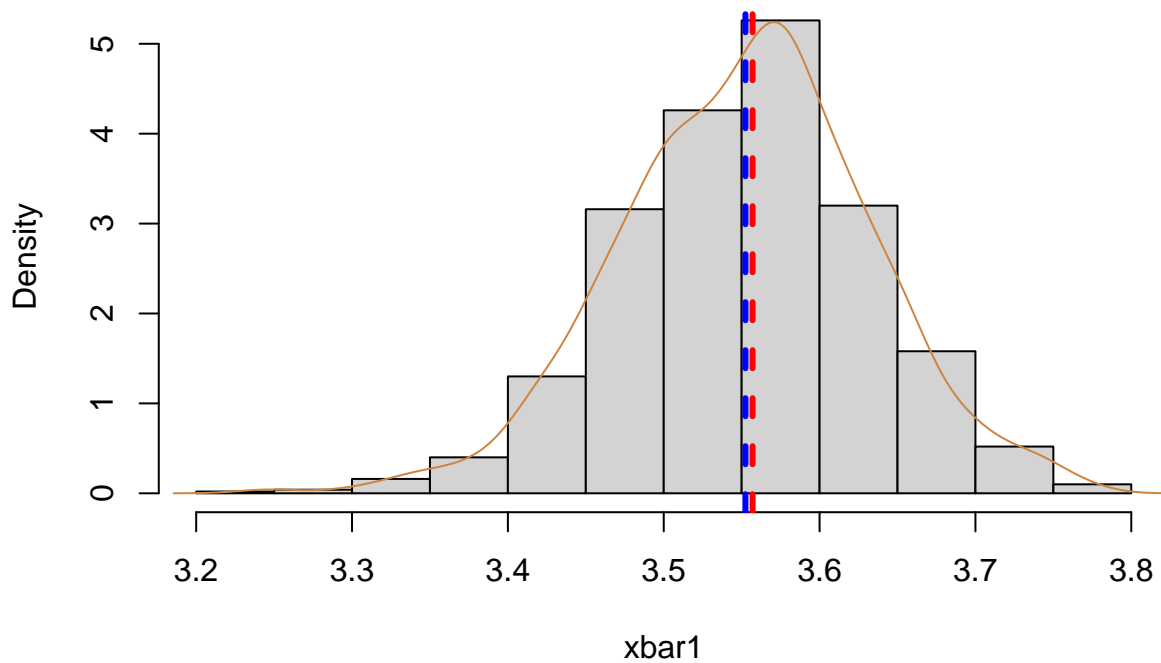
```
round(sd(x),4)
```

```
## [1] 0.8053
```

We use 4 sample sizes: 100, 150, 200, 500. As the size gets larger, the curve becomes closer to the normal distribution. The mean of sample means is closer to the population mean.

```
#samples=1000
set.seed(123)
n<-1000
samplesize1<-100
xbar1<-rep(NA,n)
for (i in 1:n){
  sample1<-sample(x,samplesize1,replace=TRUE)
  xbar1[i]<-mean(sample1)
}
hist(xbar1, freq=FALSE,main="Sample size=100")
lines(density(xbar1), col='peru', lwd=1)
abline(v = c(mean(xbar1),median(xbar1)), col=c("blue", "red"), lty=c(2,2), lwd=c(3, 3))
```

### Sample size=100



```
round(mean(xbar1),4)
```

```
## [1] 3.5524
```

```
round(median(xbar1),4)
```

```
## [1] 3.557
```

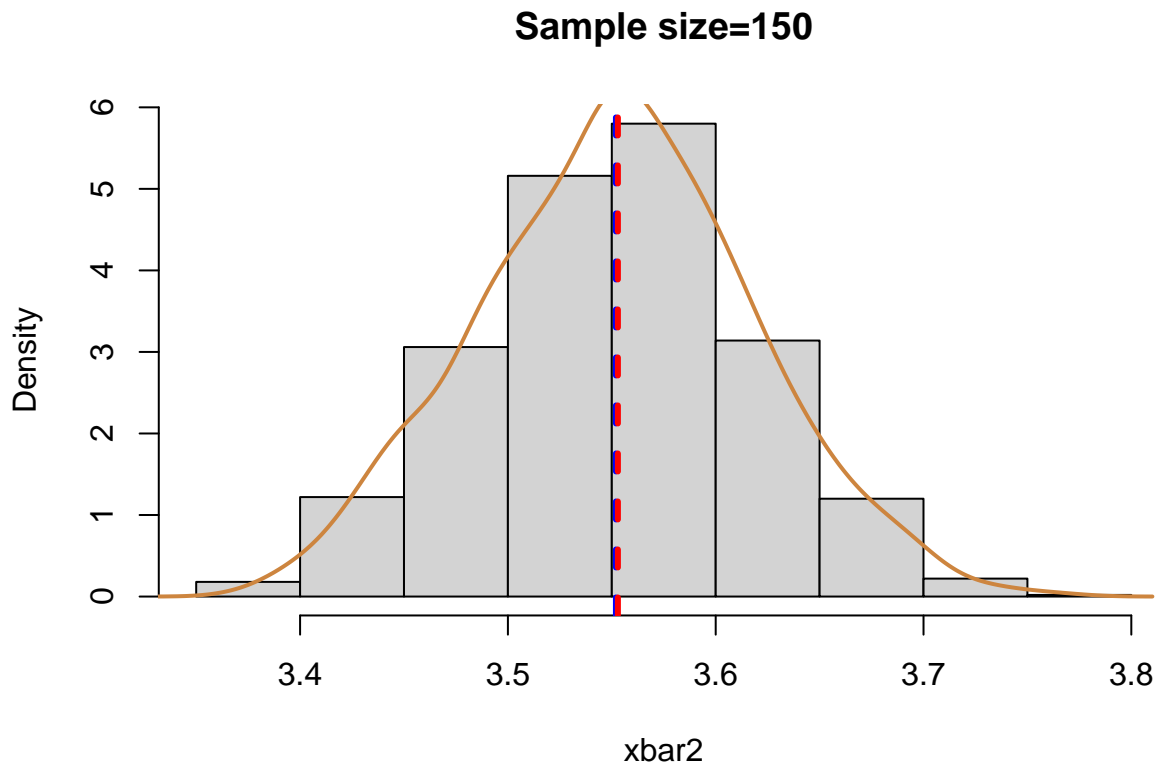
```

round(sd(xbar1),4)

## [1] 0.0806

set.seed(123)
samplesize2<-150
xbar2<-rep(NA,n)
for (i in 1:n){
  sample2<-sample(x,samplesize2,replace=TRUE)
  xbar2[i]<-mean(sample2)
}
hist(xbar2, freq=FALSE,main="Sample size=150")
lines(density(xbar2), col='peru', lwd=2)
abline(v = c(mean(xbar2),median(xbar2)), col=c("blue", "red"), lty=c(2,2), lwd=c(3, 3))

```



```

round(mean(xbar2),4)

## [1] 3.552

round(median(xbar2),4)

## [1] 3.5529

round(sd(xbar2),4)

## [1] 0.0661

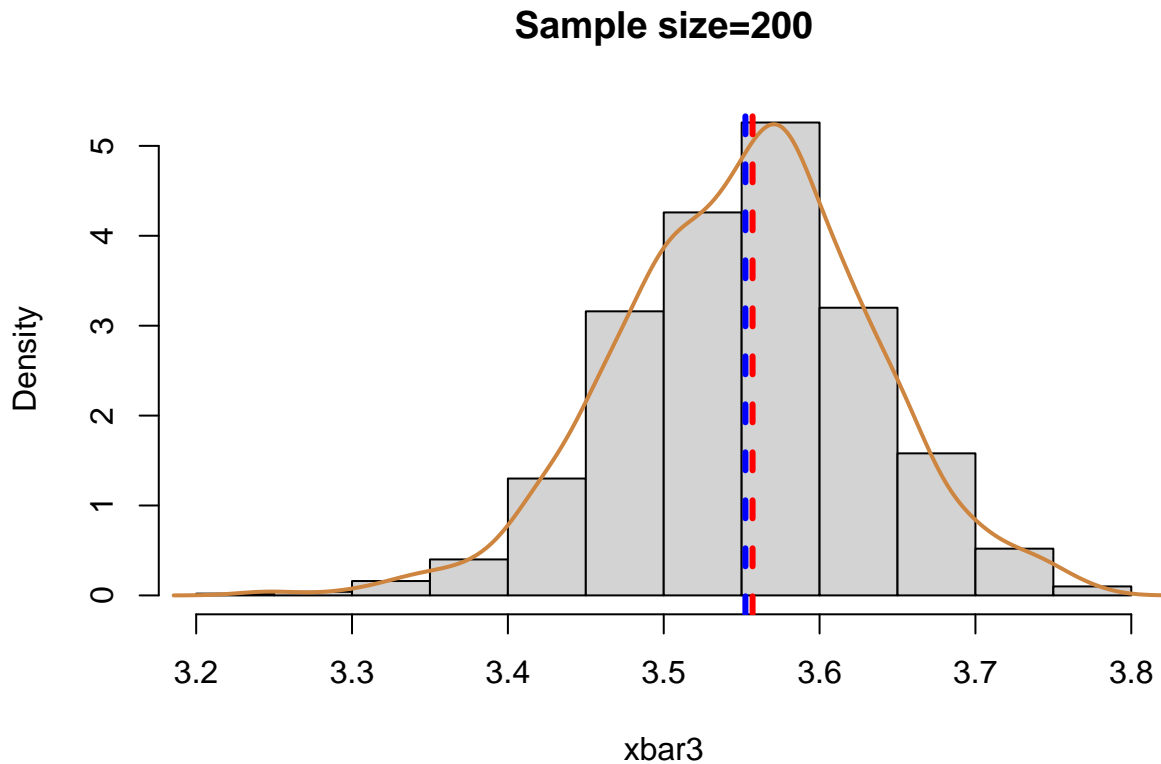
set.seed(123)
samplesize3<-200
xbar3<-rep(NA,n)
for (i in 1:n){
  sample3<-sample(x,samplesize1,replace=TRUE)
  xbar3[i]<-mean(sample3)
}

```

```

}
hist(xbar3, freq=FALSE,main="Sample size=200")
lines(density(xbar3), col='peru', lwd=2)
abline(v = c(mean(xbar3),median(xbar3)), col=c("blue", "red"), lty=c(2,2), lwd=c(3, 3))

```



```
round(mean(xbar3),4)
```

```
## [1] 3.5524
```

```
round(median(xbar3),4)
```

```
## [1] 3.557
```

```
round(sd(xbar3),4)
```

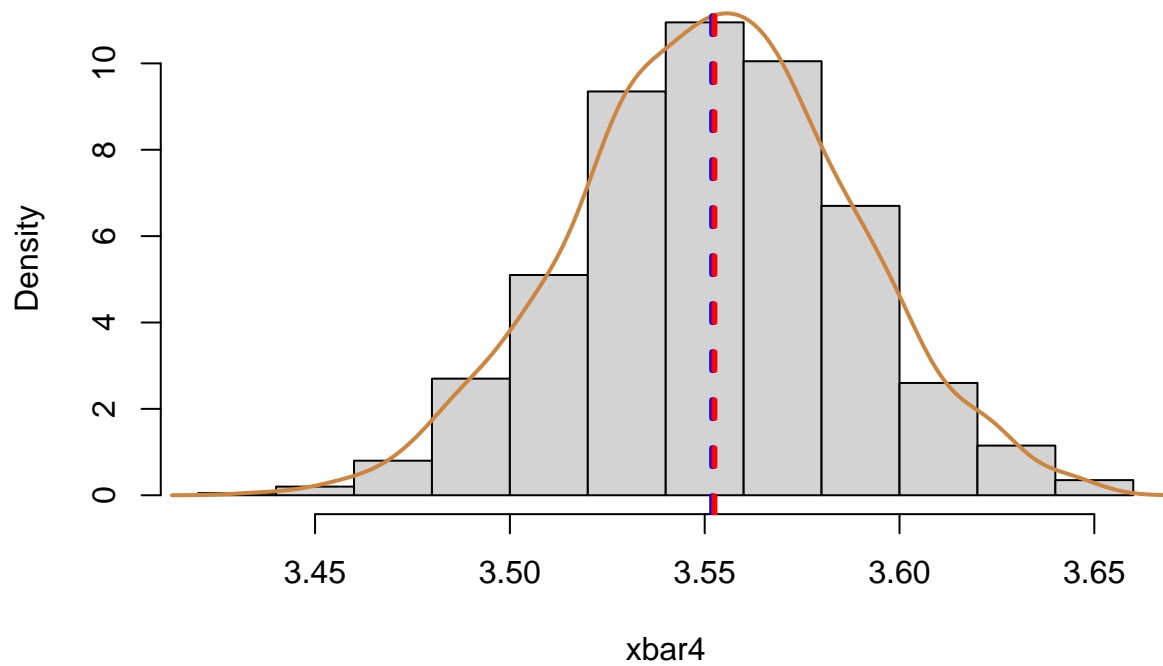
```
## [1] 0.0806
```

```

set.seed(123)
samplesize4<-500
xbar4<-rep(NA,n)
for (i in 1:n){
  sample4<-sample(x,samplesize4,replace=TRUE)
  xbar4[i]<-mean(sample4)
}
hist(xbar4, freq=FALSE,main="Sample size=500")
lines(density(xbar4), col='peru', lwd=2)
abline(v = c(mean(xbar4),median(xbar4)), col=c("blue", "red"), lty=c(2,2), lwd=c(3, 3))

```

**Sample size=500**



```
round(mean(xbar4),4)
```

```
## [1] 3.5519
```

```
round(median(xbar4),4)
```

```
## [1] 3.5525
```

```
round(sd(xbar4),4)
```

```
## [1] 0.0352
```