Statistical inference

Saurabh Yadav

PART1-A Simulation Excersise

Simulating exponantial variables

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.1
```

```
n<-1000
col<-40
expovar<-rexp(n*col,0.2)
```

creating matrix for the exponantial variables

```
mat<-matrix(expovar,n,col)
```

taking the mean of the 40 exponantial variables

```
expomean<-apply(mat, 1, mean)
```

Comparisons

Comparison between sample mean and the theoretical mean

1.mean of the population

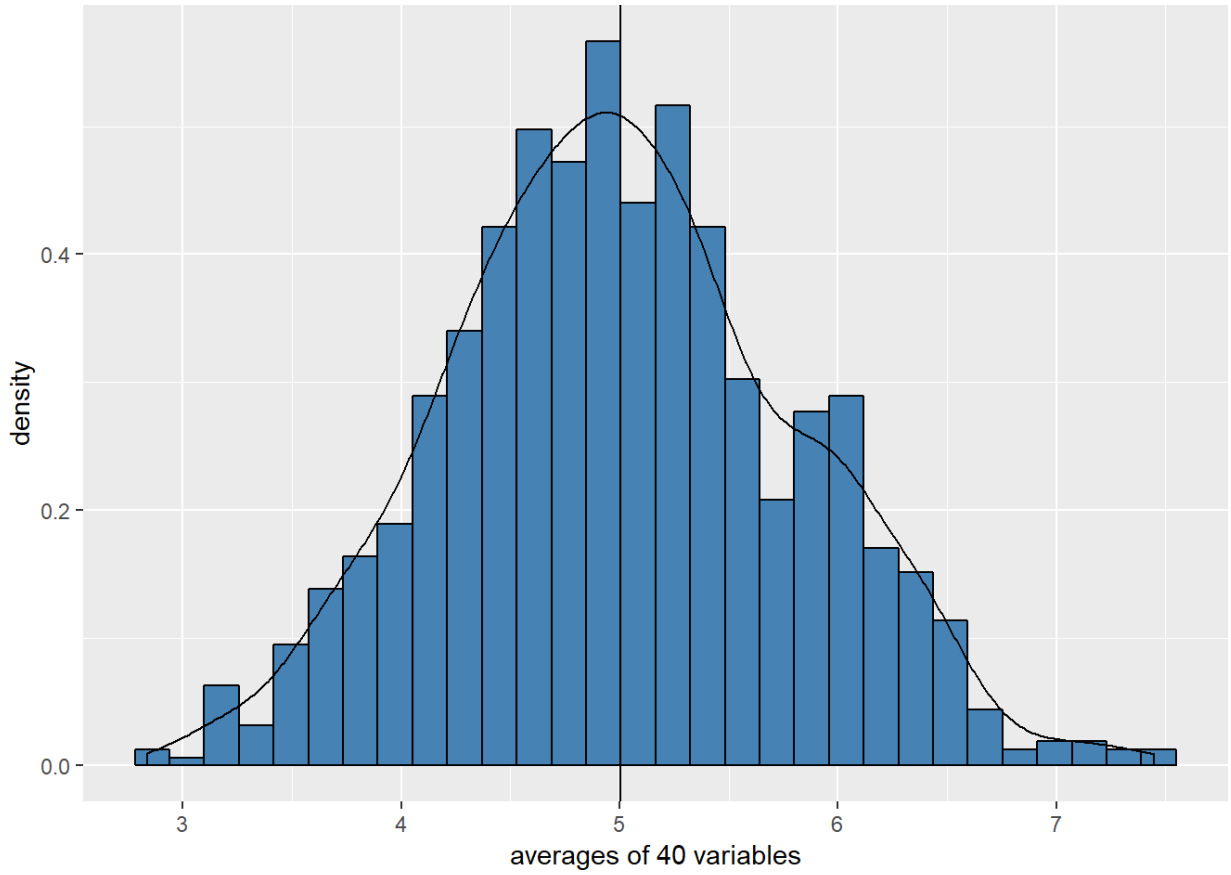
```
mean(expovar)
```

```
## [1] 5.00458
```

plotting the distribution of averages

```
ggplot(data.frame(expomean),aes(expomean))+geom_histogram(aes(y=..density..),col="black",fill="steelblue")+geom_density(col="black")+geom_vline(aes(xintercept=mean(expomean, na.rm=T)))+labs(x="averages of 40 variables")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



RESULT

1.The plot shows the distribution of averages is normally distributed and follows the Central Limit Theorm

2.The population mean and the mean of averages are approximate beacause the plot is centered around the population mean