PeerAssignment

Tanika

June 11, 2018

Statistical Inference Assignment

Introduction :

The project consists of two parts:

1. A simulation Exercise.

2. Basic Inferential Data Analysis.

You will create a report to answer the questions. Given the nature of the series, ideally you'll use knitr to create the reports and convert to a pdf. (I will post a very simple introduction to knitr). However, feel free to use whatever software that you would like to create your pdf.

Each pdf report should be no more than 3 pages with 3 pages of supporting appendix material if needed (code, figures, etcetera).

Part 1: Simulation Exercise Instructions.

In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponentials. Note that you will need to do a thousand simulations.

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponentials. You should

1.Show the sample mean and compare it to the theoretical mean of the distribution.

2.Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

3.Show that the distribution is approximately normal.

In point 3, focus on the difference between the distribution of a

```r
install.packages("knitr",repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/tanik/Documents/R/win-library/3.4'
## (as 'lib' is unspecified)
```

```
## package 'knitr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\tanik\AppData\Local\Temp\RtmpSGs9sB\downloaded_packages
```

```r
install.packages("ggplot2",repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/tanik/Documents/R/win-library/3.4'
## (as 'lib' is unspecified)
```

```
## package 'ggplot2' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\tanik\AppData\Local\Temp\RtmpSGs9sB\downloaded_packages
```

```r
library(knitr)
library(ggplot2)
```

set the seed value

```r
set.seed(100)
```

set the lambda

```r
lambda <- 0.2
```

Number of samples

```r
n <- 40
```

Number of simulations

```r
simulation <- 1000
```

Calculate the simulations

```r
simulations_exp <- replicate(simulation, rexp(n,lambda))
```

Means of the simulations

```
simulations_means <- apply(simulations_exp,2,mean)
```

theoretical mean

```
theory_mean <- 1/lambda
theory_mean
```

```
## [1] 5
```

calculate SD

```
theory_SD <- (1/lambda)/sqrt(n)
theory_SD
```

```
## [1] 0.7905694
```

Variance

```
theory_var <- theory_SD ^ 2
theory_var
```

```
## [1] 0.625
```

1.Show the sample mean and compare it to the theoretical mean of the distribution.

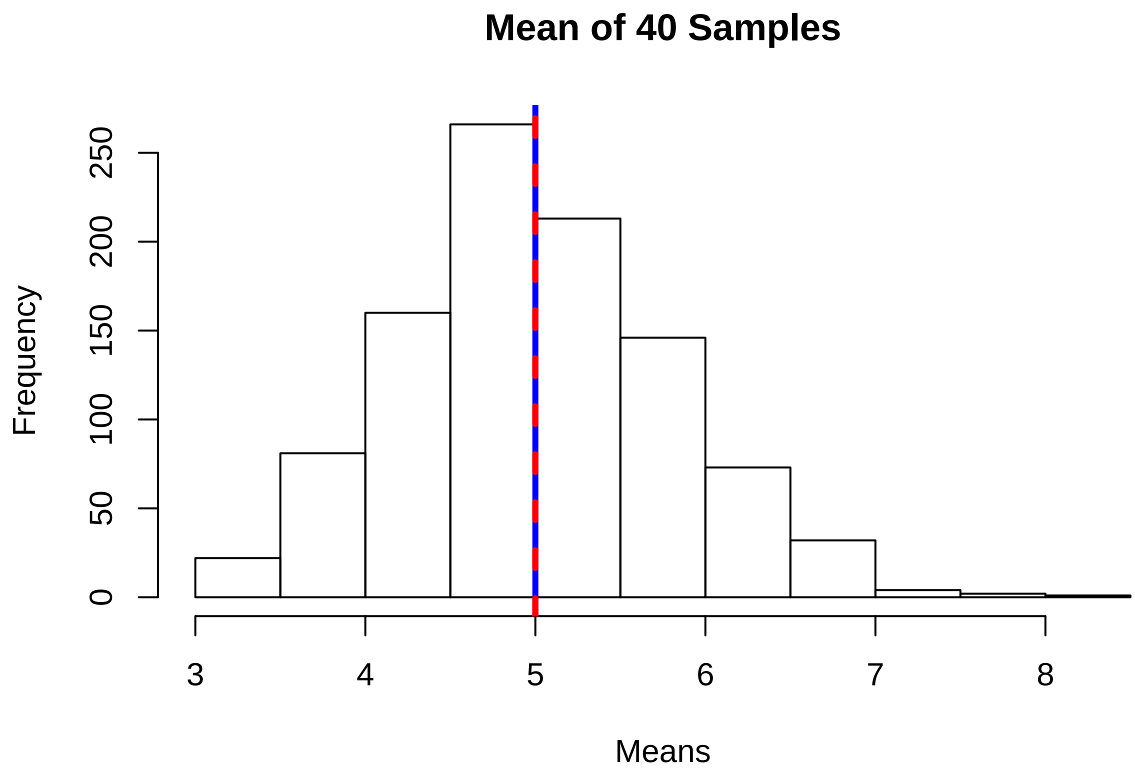
```
mean_simulation <- mean(simulations_means)
mean_simulation
```

```
## [1] 4.999702
```

Hence, the Theoretical Mean = 5 & Simulation Mean = 4.99 , which is much closer.

Visualization

```
hist(simulations_means,xlab = "Means", main = "Mean of 40 Samples")
abline(v= theory_mean, col = "blue",lwd =3,lty=1)
abline(v= mean_simulation, col = "red", lwd =3,lty=2)
```



2.Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.

Calculate the Variance of the Distributions:

First Calculate the Standard Deviation

```
simulation_SD <- sd(simulations_means)
simulation_SD
```

```
## [1] 0.8020251
```

variance

```
simulation_var <- simulation_SD^2
simulation_var
```

```
## [1] 0.6432442
```

Theoretical Variance : 0.625

Simulation Variance : 0.6303972

3.Show that the distribution is approximately normal.

```
meanData <- data.frame(simulations_means)
g<- ggplot(data = meanData, aes(x=simulations_means))
g<- g+ geom_histogram(aes(y= ..density..), fill = I("blue"),binwidth = 0.2,color=I("red4"))
g<- g+ stat_function(fun = dnorm,args = list(mean = theory_mean,sd = theory_SD), color = I("red1"),lwd =

g
```