Statistical inference course project

Shawvin

7/3/2019

Part 1: Simulation exercise

Overview

This Simulation exercise investigate the exponential distribution and compare it with Central Limit Theorem.The exponential distribution can be simulated in R as rexp(n,lambda). The mean and standard devitaion for the exponential distribution will be 1/lambda.

Simulations

We will set lambda=0.2 for the simulation.The sample will be averages of 40 exponentials for 1000 simulation.

```r
set.seed(1234)
lambda<-0.2
mns<-NULL

#1000 simulations are run, and each simulation will calculate the mean of 40 random exponential and gen
for(i in 1:1000)
  mns<-c(mns,mean(rexp(40,lambda)))

#Sample Mean will be the mean of these 1000 sample means
#Sample Variance will be the variance of these 1000 sample means
Sample_Mean<-mean(mns)
Sample_Var<-var(mns)

#Theoretical Mean and Theoretical Variance are 1/lambda and 1/(lambda^2*n) respectivly
Theoretical_Mean<-1/lambda
Theoretical_Var<-1/(lambda^2*40)
```
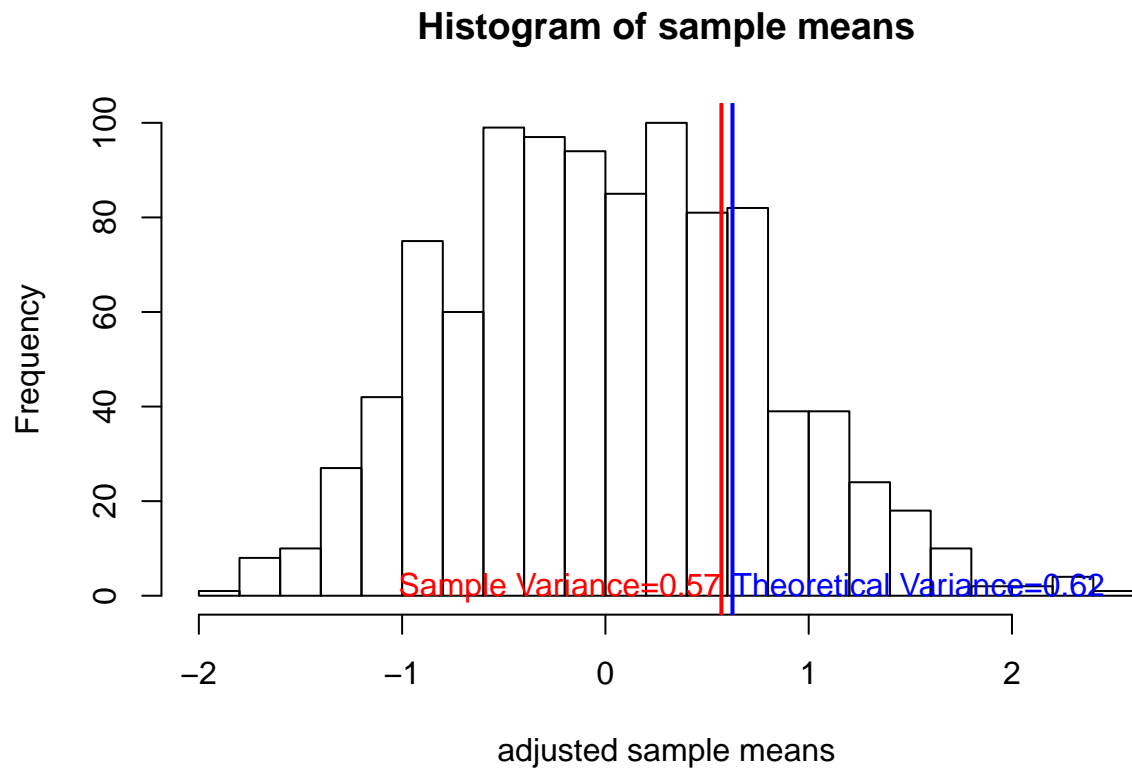
Histogram of sample means
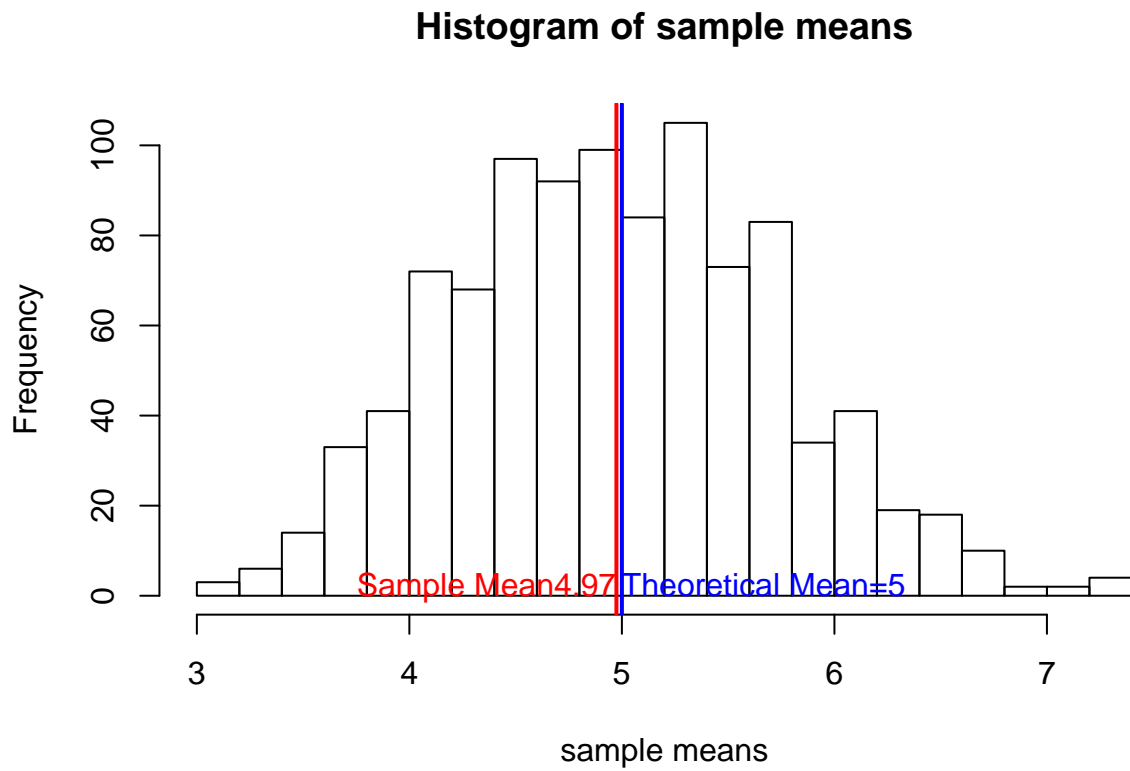
The histogram of 1000 sample means is plotted and the mean of these 1000 sample means is used as Sample Mean. In the plot, redline denotes the Sample Mean and the blue line denotes the Theoretical Mean. Their value are 4.9742388 and 5

Sample Variance versus Theoretical Variance

Histogram of sample means

The histogram of 1000 sample means is plotted.Sample Variance is the variance of these 1000 sample means.For comparing the variance more esaily, the Sample Mean of these 1000 average is shifted to 0 by minusing Sample Mean from each sample mean.

In the plot, redline denotes the sample variance and the blue line denotes the theoretical variance. Their value are 0.5706551 and 0.625

Distribution

Histogram of normalized sample means
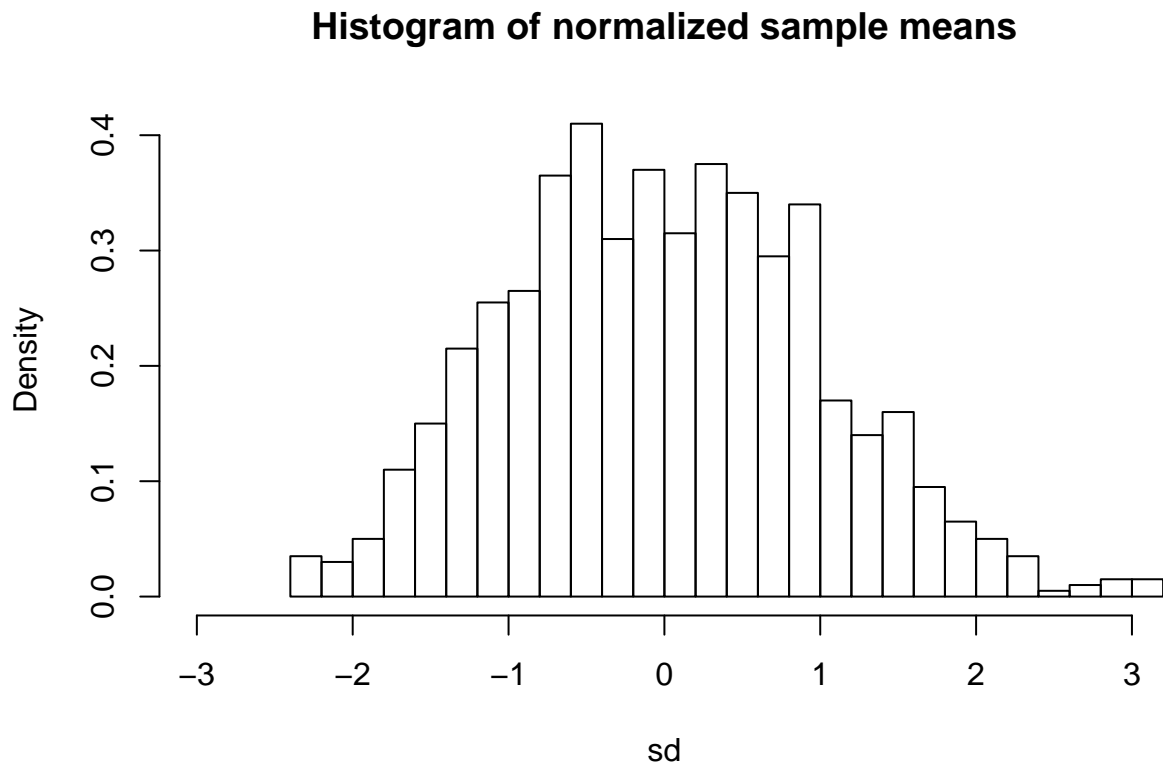

```
##    sd cum probability of Sample Mean distribution
## 1 -3                                        0.000
## 2 -2                                        0.013
## 3 -1                                        0.169
## 4  0                                        0.513
## 5  1                                        0.848
## 6  2                                        0.974
## 7  3                                        0.997
##   cum probability of normal distribution
## 1                             0.001349898
## 2                             0.022750132
## 3                             0.158655254
## 4                             0.500000000
## 5                             0.841344746
## 6                             0.977249868
## 7                             0.998650102
```

These 1000 sample means are normlized and plotted.The probabily is compared to normal distribution.

Part 2: Basic Inferential Data Analysis Instructions

basic exploratory data analyses

```
data(ToothGrowth)
dim(ToothGrowth)
```

```
## [1] 60  3
```

```
head(ToothGrowth)
```

```
##     len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
table(ToothGrowth$supp,ToothGrowth$dose)
```

```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

We load the ToothGrowth data. This dataframe has 60 rows and 3 columns. The first column is the length of tooth growth. The second column is the type of supplement used and the third column is the dose.

We tabulate the supplement and dose and find that there are two supplement and three different dose used.

basic summary of the data

```
suppressMessages(library(dplyr))
library(ggplot2)
plotdata<- ToothGrowth %>% group_by(supp,dose) %>% summarise(mean=mean(len),se=sd(len)/sqrt(length(len))
plotdata
```

```
## # A tibble: 6 x 4
## # Groups:   supp [2]
##   supp   dose  mean    se
##   <fct> <dbl> <dbl> <dbl>
## 1 OJ      0.5 13.2  1.41
## 2 OJ      1   22.7  1.24
## 3 OJ      2   26.1  0.840
## 4 VC      0.5  7.98 0.869
## 5 VC      1   16.8  0.795
## 6 VC      2   26.1  1.52
```

```
g<-ggplot(plotdata,aes(factor(dose),mean,group=supp,color=supp))
g+geom_line(lwd=1)+geom_errorbar(aes(ymin=mean-se,ymax=mean+se),width=0.1)+labs(x="Dose",y="Length")+gg
```
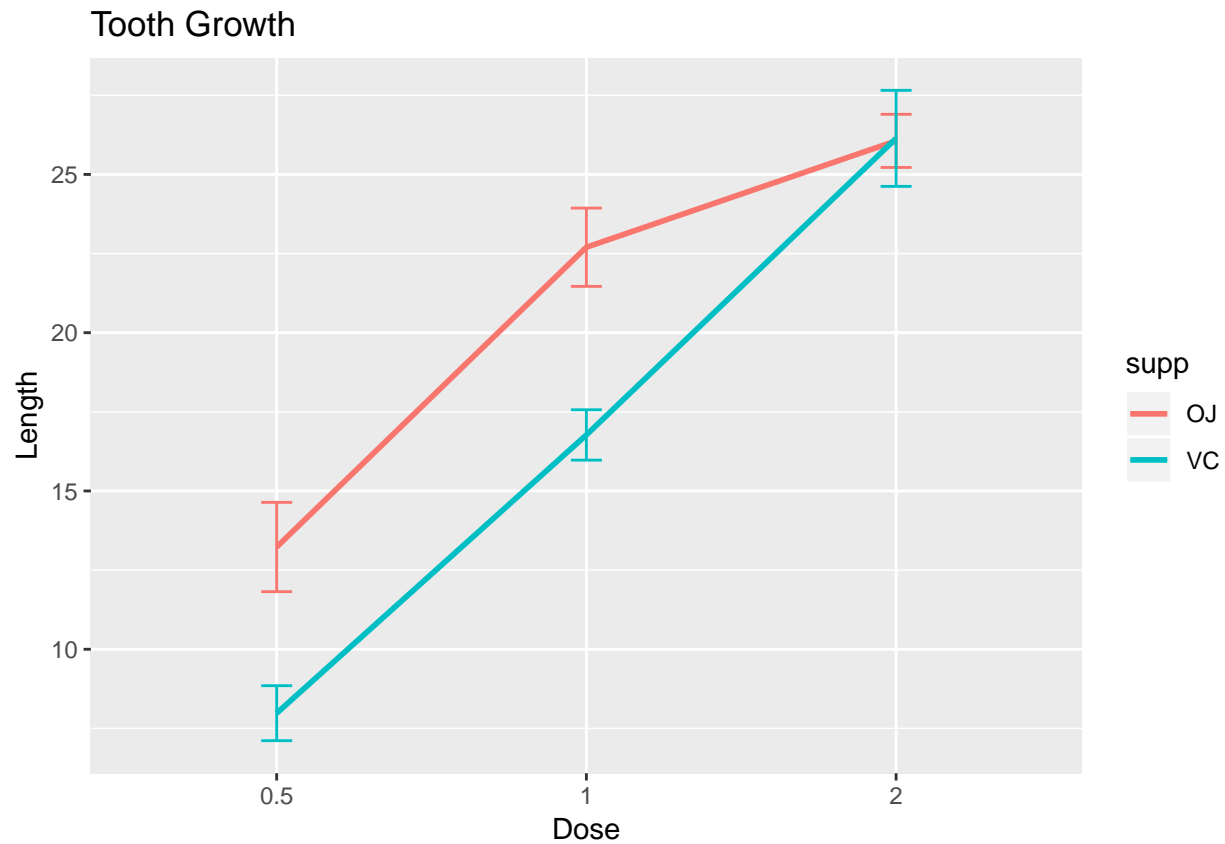
Tooth Growth

By the type of supp and the dose, we separate the data into 6 groups and plot the mean and sem for each group. From the plotdata, we find that the length of tooth growth increases as the dose of supp increases. The OJ supplement has better improvement than VC supplement in tooth growth at lower dose (0.5 and 1). The effect of two supplements are comparable at higher dose (2).

Hypothesis testing

```
OJ_0.5<-ToothGrowth[ToothGrowth$supp=="OJ"&ToothGrowth$dose==0.5,]$len
OJ_1<-ToothGrowth[ToothGrowth$supp=="OJ"&ToothGrowth$dose==1,]$len
OJ_2<-ToothGrowth[ToothGrowth$supp=="OJ"&ToothGrowth$dose==2,]$len

VC_0.5<-ToothGrowth[ToothGrowth$supp=="VC"&ToothGrowth$dose==0.5,]$len
VC_1<-ToothGrowth[ToothGrowth$supp=="VC"&ToothGrowth$dose==1,]$len
VC_2<-ToothGrowth[ToothGrowth$supp=="VC"&ToothGrowth$dose==2,]$len

#IN OJ supplement groups,compare the effect of different dose
t.test(OJ_0.5,OJ_1)$p.value
```

```
## [1] 8.784919e-05
```

```r
t.test(OJ_1,OJ_2)$p.value
```

```
## [1] 0.03919514
```

```r
#IN VC supplement groups,compare the effect of different dose
t.test(VC_0.5,VC_1)$p.value
```

```
## [1] 6.811018e-07
```

```r
t.test(VC_1,VC_2)$p.value
```

```
## [1] 9.155603e-05
```

```r
#Compare OJ supplement and VC supplement at the same dose
t.test(OJ_0.5,VC_0.5)$p.value
```

```
## [1] 0.006358607
```

```r
t.test(OJ_1,VC_1)$p.value
```

```
## [1] 0.001038376
```

```r
t.test(OJ_2,VC_2)$p.value
```

```
## [1] 0.9638516
```

Conclusion

For both supplement, the teeth growth increases significantly as the dose increase. The effect of OJ is significantly better than that of VC at the dose of 0.5 and 1 but the effect are comparable between these two supplement at the dose of 2.