Statistical Inference Course Project Part 1

yhuai

2018/4/21

Overview

This is the Coursera Johns Hopkins University Data Science Specialization – Statistical Inference Course Project

The project consists of two parts:

1.A simulation exercise. 2.Basic inferential data analysis.

Part 1: Simulation Exercise Instructions

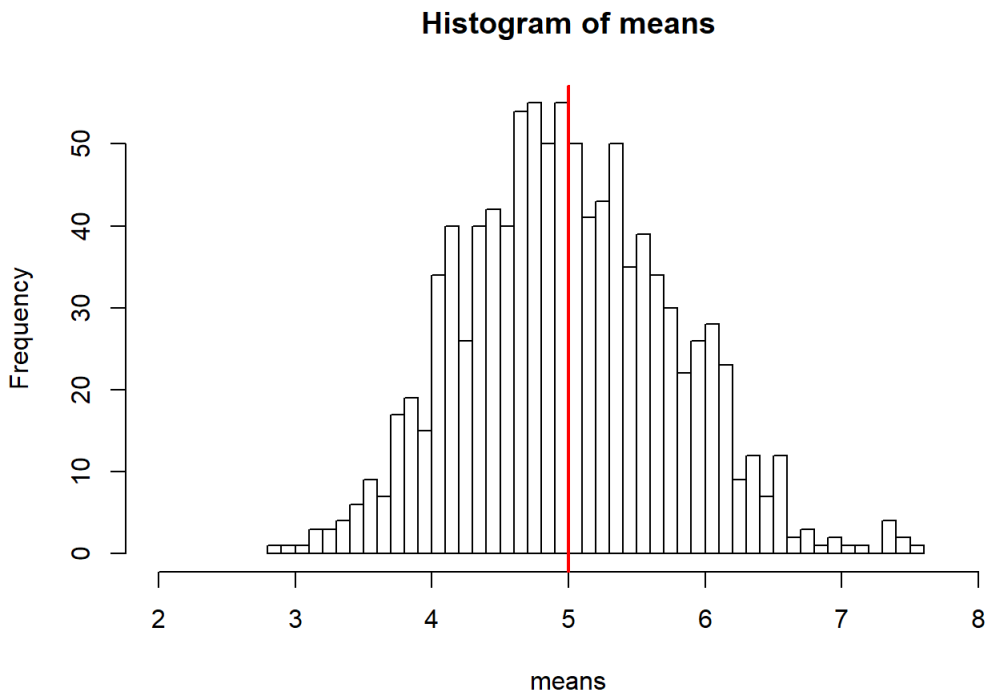
In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation 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.

```
set.seed(19930319)
lambda <- 0.2; n <- 40;
means = NULL
for (i in 1:1000) means = c(means, mean(rexp(n,lambda)))
hist(means,breaks=50,xlim=c(2,8))

mean(means)
```

```
## [1] 4.999559
```

```
abline(v=mean(means),col="red",lwd="2")
```



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

```
a1 <- lambda^-1
theoretical_mean <- a1
b1 <- mean(means)
sample_mean <- b1
table(theoretical_mean,sample_mean)
```

```
##                 sample_mean
## theoretical_mean 4.99955904380919
##                5               1
```

Show how variable it is and compare it to the theoretical variance of the distribution.

```
a2 <- (1/lambda)^2/n
theoretical_var <- a2
b2 <- var(means)
sample_var <- b2
table(theoretical_var,sample_var)
```

```
##                 sample_var
## theoretical_var 0.595911588567069
##           0.625               1
```

Based on Central Limit Theorem, our number of smaples is big(1000), the simulated exponential distribution is close to normal distribution.

```
x <- seq(2,8,length=2*n)
y <- dnorm(x,mean(means),sd(means))
plot(x,y,type = "l",col="red")
```