Statiscal Inference oroject part 1

Nguyen Son Linh

5/29/2020

Overview

In this part, we will explore the exponential distribution and compare it with the Central Limit Theorem (CLT) by simulation

Setup

Packages needed

```
library(ggplot2)
```

Part 1: Simulation

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

Per the instructions, the exponential distribution can be simulated in R with the function rexp(n, lambda) where lambda is the rate parameter. The simulation can be repeated multiple times using the repetition function.

The theoretical mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda.

For the 1000 simulations, lambda is assumed to be 0.2 and the sample size n is 40.

The seed is set to ensure reproducibility.

```
set.seed(21)
lambda <- 0.2
expo_1000 <- as.data.frame(replicate(1000, mean(rexp(40, lambda))))
names(expo_1000) <- c("sample_mean")

mean(expo_1000$sample_mean)
```

```
## [1] 4.98144
```

As seen, the sample mean is close to the theoretical mean

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

```
var(expo_1000$sample_mean)
```

```
## [1] 0.5904182
```

Likewise for the variance

1.3 Show that the distribution is approximately normal.

We plot the sample means distribution

```
ggplot(data = expo_1000, aes(x=sample_mean)) +
    geom_histogram(aes(y = ..density..),colour="black",fill="lightblue")+
    stat_function(fun=dnorm,args=list( mean=mean(expo_1000$sample_mean), sd=sqrt(var(expo_1000$sample_m
    ggtitle("Histogram of the Simulation Samples Means where n = 1000") +
    scale_x_continuous("Sample means")+
    ylab("Density")
```

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