Statistical Inference Course Project Part 1

Louis

9/19/2020

Coursera Statistical Inference Course Project

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(knitr)
```

Part 1 : Finding the mean and standard deviation of an randomly generated set of numbers.

Loading the values

```
set.seed(13)
#Write down the values
n <- 40
sim_num <- 1000
lambda <- 0.2
#Generate the sample numbers
sim_sample <- replicate(sim_num, rexp(n,lambda))
#Get the means of the sample numbers
mean_sampl <- apply(sim_sample, 2, mean)
```

Finding the theoretical and sample means

```
theo_mean <- 1/lambda
mean_samp <- round(mean(mean_sampl),2)
print(paste("Theoretical mean is:", theo_mean))
```

```
## [1] "Theoretical mean is: 5"
```

```
print(paste("Sample Mean is:",round(mean(mean_sampl),2)))
```

```
## [1] "Sample Mean is: 4.97"
```

Finding the theoretical and sample variance

```
theo_var <- (1/lambda)^2 / n
sampl_var <- round(var(mean_sampl),3)
print(paste("Theoretical variance is:", theo_var))
```

```
## [1] "Theoretical variance is: 0.625"
```

```
print(paste("Sample variance is:", sampl_var))
```

```
## [1] "Sample variance is: 0.623"
```

Finding the theoretical and sample standard deviation

```
theo_sd <- round(1/(lambda * sqrt(n)),2)
sampl_sd <- round(sd(mean_sampl),2)
print(paste("Theoretical standard deviation is:", theo_sd))
```

```
## [1] "Theoretical standard deviation is: 0.79"
```

```
print(paste("Sample standard deviation is:", sampl_sd))
```

```
## [1] "Sample standard deviation is: 0.79"
```

Making a plot for the data

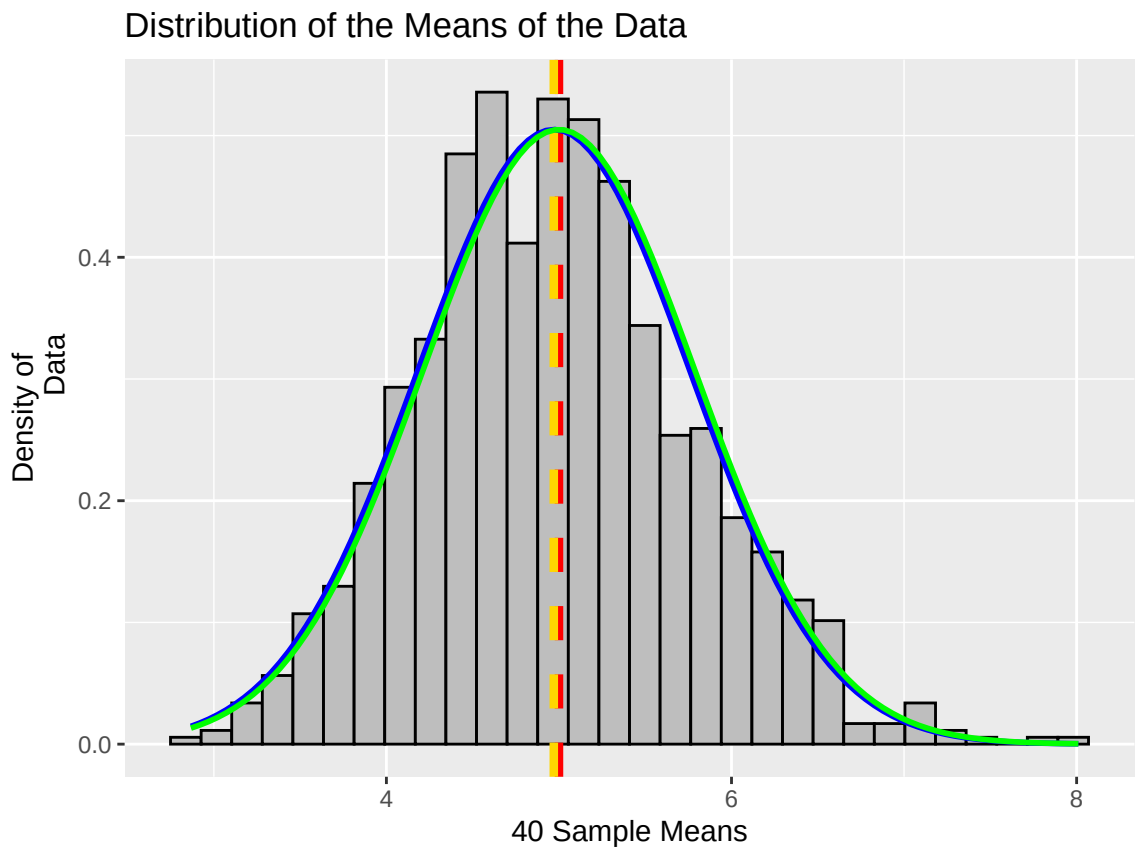
```
data_sim <- data.frame(mean_sampl)
data_plot <- ggplot(data = data_sim, aes(mean_sampl))

data_plot <- data_plot +
        geom_histogram(aes(y=..density..),colour="black", fill = "grey") +
```

```
        geom_vline(xintercept = theo_mean, color = "red", linetype = "dashed", size = 1.5) +
        geom_vline(xintercept = mean_samp, color = "gold", linetype = "dashed", size = 1.5)
data_plot <- data_plot +
        stat_function(fun = dnorm, args = list(mean = mean_samp, sd = sampl_sd), color = "blue",
                      size = 1.0)
data_plot <- data_plot +
        stat_function(fun = dnorm, args = list(mean = theo_mean, sd = theo_sd), colour = "green",
                      size = 1.0)
data_plot <- data_plot +
        labs(title = "Distribution of the Means of the Data", x = "40 Sample Means", y = "Density of
             Data")


data_plot
```

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



Red is the theoretical mean(5) and Gold is the sample mean(4.97)

Green is the theoretical standard deviation(0.79) and Blue is the sample standard deviation(0.79)