Statistical Inference Course Project Part 1

1- Introduction

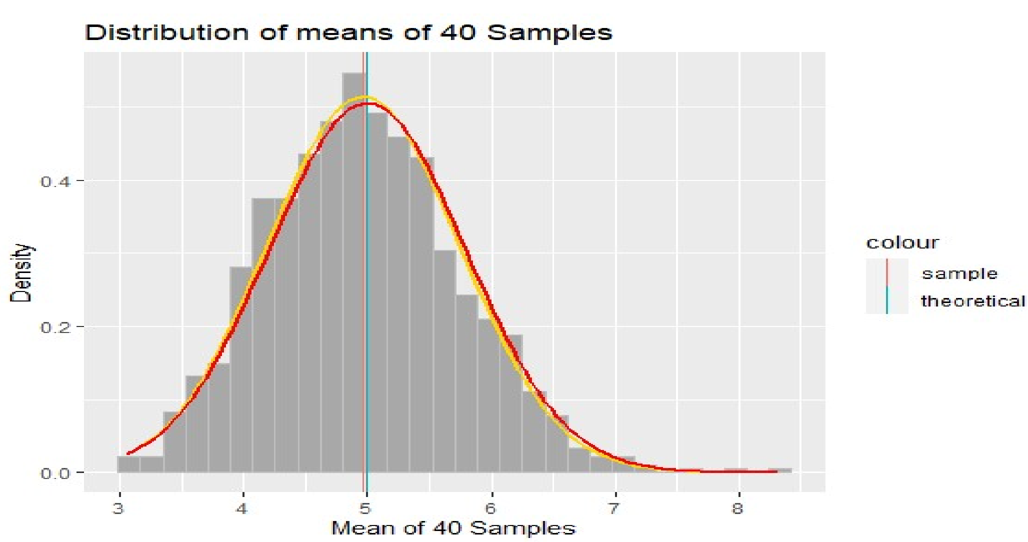
It is demonstrated that the distribution proves the Central Limit Theorem. This assignment will make calculation and plots and compare confidence intervals, and eventually proves that the distribution is approximately normal.

2- Project Mind Flow

1. Create Data sample for simulation
2. Show the sample mean and compare it to the theoretical mean of the distribution.
3. Show how variable the sample is (via variance) and compare it to the theoretical variance of the distribution.
4. Show that the distribution is approximately normal.

3- Conclusion & Output

1. The mean of the sample means is 4.9668844 and the theoretical mean is 5. The sample mean and the theoretical mean (expected mean) are very close.
2. The variance of the sample means is 0.6048268 and the theoretical variance of the distribution is 0.625. Both variance values are very close to each other. Meanwhile, the standard deviation of the sample means is 0.7777061 and the theoretical standard deviation of the distribution is 0.7905694. Both standard deviation values are very close to each other.
3. The density of the actual data is shown by the light blue bars. The theoretical mean and the sample mean are so close that they nearly overlap. The "red" line shows the normal curve formed by the the theoretical mean and standard deviation. The "gold" line shows the curve formed by the sample mean and standard deviation. As you can see from the graph, the distribution of means of 40 exponential distributions is close to the normal distribution with the expected theoretical values based on the given lambda.
4. The sample confidence interval is (4.726, 5.208) and the theoretical confidence level is (4.755, 5.245) . The confidence levels also match closely. Again, proving the distribution is approximately normal.



Statistical Inference Course Project Part 1

4- Coding

```
## Set Random data
set.seed(127)
lambda <- 0.2
n <- 40
sample_size <- 1000
simulated_sample <- replicate(sample_size, rexp(n, lambda))
means_exponentials <- apply(simulated_sample, 2, mean)
##  Meam Comparison
sample_mean <- mean(means_exponentials)
theo_mean <- 1 / lambda
sample_mean
theo_mean
mat.mean  <- (sample_mean / theo_mean)
mat.mean
##  Variance Comparison
sample_var <- var(means_exponentials)
theo_var  <- (1 / lambda)^2 / (n)
sample_var
theo_var
mat.Var  <- (sample_var / theo_var)
sample_sd <- sd(means_exponentials)
theo_sd  <- 1/(lambda * sqrt(n))
sample_sd
theo_sd
mat.sd  <- (sample_sd / theo_sd)
mat.sd
## Distribution _ Plotting
plotdata <- data.frame(means_exponentials)
m <- ggplot(plotdata, aes(x =means_exponentials))
m <- m + geom_histogram(aes(y=..density..), colour="grey",
                fill = "grey66")
m <- m + labs(title = "Distribution of means of 40 Samples", x = "Mean of 40
Samples", y = "Density")
m <- m + geom_vline(aes(xintercept = sample_mean, colour = "sample"))
m <- m + geom_vline(aes(xintercept = theo_mean, colour = "theoretical"))
```

```
m <- m + stat_function(fun = dnorm, args = list(mean = sample_mean, sd =
sample_sd), color = "gold1", size = 1.0)
m <- m + stat_function(fun = dnorm, args = list(mean = theo_mean, sd =
theo_sd), colour = "red", size = 1.0)
m
```