Sample.Rmd

Comparison of an Exponential Distributions and the Central Limit Theorem

Overview

This exercise will look at the difference between a sample mean and it's related population mean and how the Central Limit Theorem works. The Central Limit Theorem states that the larger a sample size, the closer the sample's mean and variance will reflect the true population mean and variance.

Simulations

Two sets of simulations will be covered:

- Create a sample of 40 draws, **Draw40** and compare the mean and standard deviation
- Create 1000 iterations of 40 draws, **Draw1000**, and compare the observed means and standard deviations

One of the **GIVENS** in this assignment is to use a **lambda value of 0.2**.

```
lambda <- 0.2   # this is a GIVEN in the assignment
```

This implies that the theoretical **true mean and true standard deviation** of the population is:

```
true_mean <- 1/lambda
true_std <- 1/lambda
```

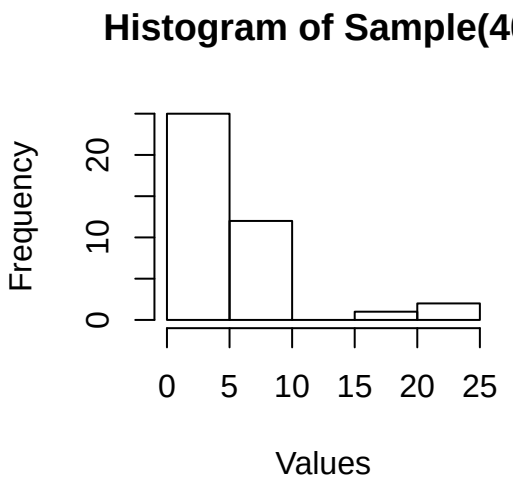
NOTE: We will use `set.seed(2)` before every sequence to ensure reproducibility.

Simulation of 40 Draws

If we simulate 40 draws, what are the mean and standard deviation of this sample?

```
set.seed(2)
sample_data <- rexp(40, rate=lambda)
sample_mean <- mean(sample_data)
sample_std <- sd(sample_data)
```

```
hist(sample_data, main="Histogram of Sample(40)", xlab="Values")
```



The **true mean is 5**, how does this compare with the sample mean?

The **sample_mean** is 5.199 and the standard deviation is 5.275. To see if this is meaningful we calculate the confidence interval for the sample size of 40 given the true mean:

```
n <- 40
true_mean + c(-1,1) * qt(0.95, n-1) * true_std/sqrt(n)
```

```
## [1] 3.667989 6.332011
```

We can see that the observed mean 5.199 falls within this confidence interval, meaning that 95% of the time we will observe a sample's mean value within this interval as an estimation of the **true mean**.

Simulation of 1000 Draws of 40

Note that as the size of the sample increases, the confidence interval decreases as there is less variation areound the theoretical mean

Create a matrix of 100 draws of 40

```
set.seed(2)
Draw1000 <- matrix(rexp(40000, rate=lambda), 1000,40)

Draw1000_mean <- apply(Draw1000, 1, mean)
Draw1000_std <- apply(Draw1000, 1, sd)

mean(Draw1000_mean)
```

```
## [1] 5.016356
```

```
mean(Draw1000_std)
```

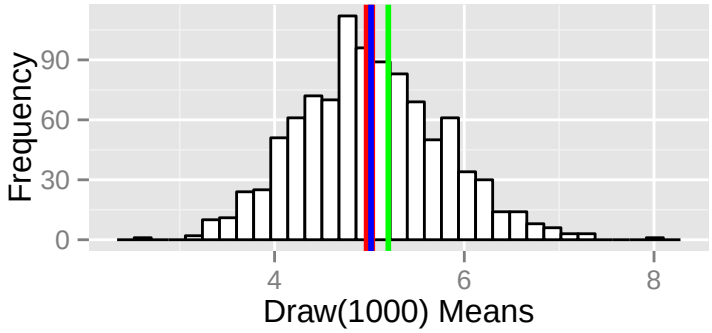
```
## [1] 4.898651
```

```r
n <- 40000
mean(Draw1000_mean) + c(-1,1) * qt(0.95, n-1) * mean(Draw1000_std)/sqrt(n)
```

```
## [1] 4.976067 5.056645
```

Also note that by increasing the number of observations, the confidence interval has shrunk.

```
## Loading required package: ggplot2
```



Appendix

```
hist(Draw1000, breaks=40)
```