Statistical Inference Course Project

Tonya MacDonald

Purpose

Investigate the exponential distribution in R and compare it with the Central Limit Theorem.

Libraries

```
library("data.table")
library("ggplot2")
```

Setup

```
set.seed(1983)
# set lambda to 0.2, sample size to 40, 1000 simulations
l <- 0.2
n <- 40
s <- 1000

# calculate means
means <- apply(replicate(s, rexp(n, l)), 2, mean)
```

1 - Sample Mean versus Theoretical Mean

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

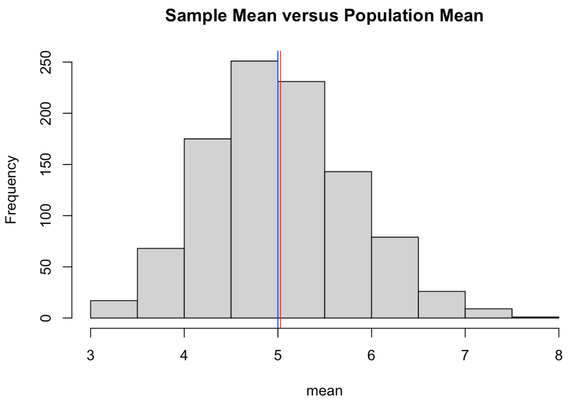
```
# sample mean
sample_mean <- mean(means)
sample_mean
```

```
## [1] 5.027677
```

```
# theoretical mean
theory_mean <- 1/l
theory_mean
```

```
## [1] 5
```

```
# histogram
hist(means, xlab = "mean", main="Sample Mean versus Population Mean")
abline(v = sample_mean, col = "red", )
abline(v = theory_mean, col = "blue")
```



2 - Sample Variance versus Theoretical Variance

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

```
# standard deviation of distribution
sd_dist <- sd(means)
sd_dist
```

```
## [1] 0.7857763
```

```
# standard deviation
sd_theory <- (1/l)/sqrt(n)
sd_theory
```

```
## [1] 0.7905694
```

```
# variance of distribution
v_dist <- sd_dist^2
v_dist
```

```
## [1] 0.6174444
```

```
# variance
v_theory <- ((1/l)*(1/sqrt(n)))^2
v_theory
```

```
## [1] 0.625
```

3 - Distribution

Show that the distribution is approximately normal.

```
# compare the distribution to a normal distribution

# histogram for simulated means
hist(means,breaks=n,prob=T,col="pink",xlab = "means",main="Distribution",ylab="density")

# overlay line for normal dist
x <- seq(min(means), max(means), length=100)
y <- dnorm(x, mean=1/l, sd=(1/l/sqrt(n)))
lines(x, y, pch=22, col="magenta", lty=5)
```