Central Limit Theorem

Statistical Inference Project Part I, Class 6 in data science series

Ann Crawford

```
# Clear our workspace
rm(list=ls())
 #dependencies
#install.packages("dplyr",repos = "http://cran.us.r-project.org")
#install.packages("ggplot", repos = "http://cran.us.r-project.org")
#install.packages("gridExtra",repos = "http://cran.us.r-project.org")
```

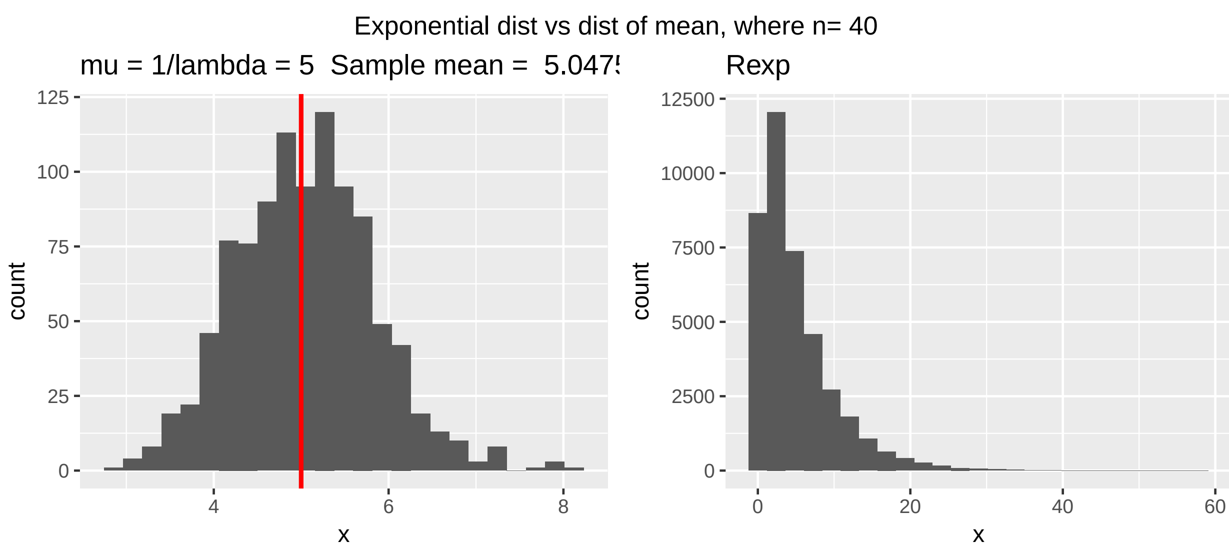
Central Limit Theorem

The Central Limit Theorem (CLT) states that the distribution of averages of independent and identically distributed (iid) variables becomes that of a **standard normal** as the sample size increases even if the original variables are not normally distrubted. This document investigates the exponetial distribution and the distribution on the average of 40 exponetials generated using R function rexp.

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\text{Estimate} - \text{Mean of estimate}}{\text{Std. Err. of estimate}}.$$

The sample mean and distribution compared to to the population

The mean of the sampling distribution of means approaches mu (population mean) as n increases. The graphs below also shows that although the population distribtion is not normal the distribution of sample means is normal with sample mean approaching populatio mean.



The sample variance vs population variance

The histogram shows a distribution with a shape similar to the normal curve. The variance of the sampling distribution of the sample means is less than population variance and approaches 0 as n goes to infinity. For n = 40, sample variance for exponentials = 0.625