

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")

## Installing package into 'C:/Users/Ann/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'dplyr' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\Ann\AppData\Local\Temp\Rtmp2jK00f\downloaded_packages
install.packages("ggplot", repos = "http://cran.us.r-project.org")

## Installing package into 'C:/Users/Ann/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## Warning: package 'ggplot' is not available (for R version 3.3.2)
install.packages("gridExtra", repos = "http://cran.us.r-project.org")

## Installing package into 'C:/Users/Ann/Documents/R/win-library/3.3'
## (as 'lib' is unspecified)

## package 'gridExtra' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
## C:\Users\Ann\AppData\Local\Temp\Rtmp2jK00f\downloaded_packages
library(ggplot2)

## Warning: package 'ggplot2' was built under R version 3.3.3
library(grid)
library(gridExtra)

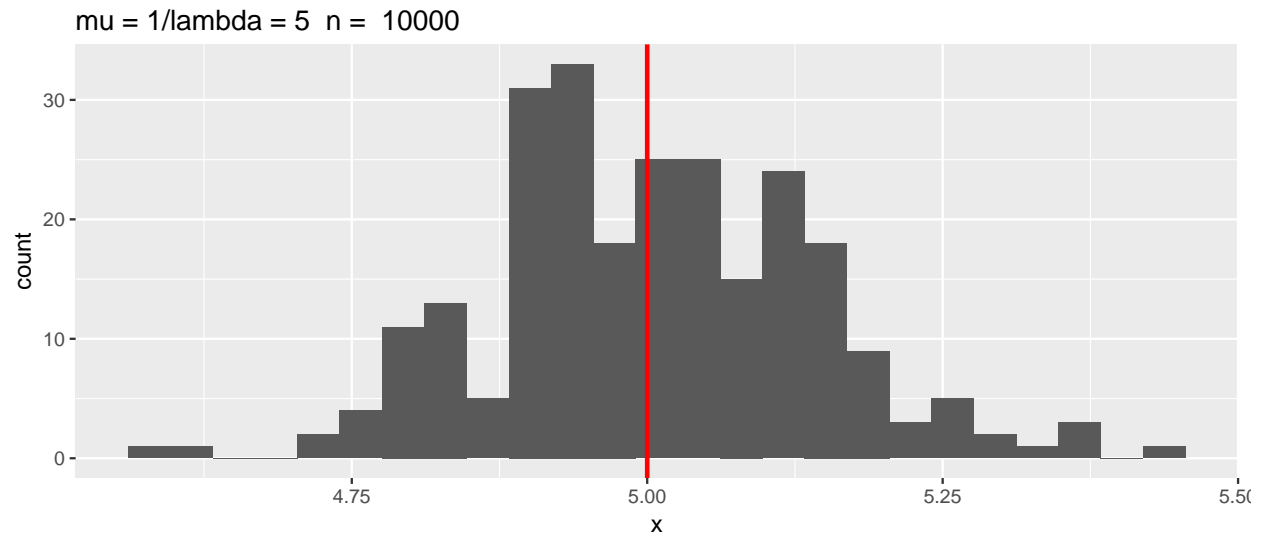
## Warning: package 'gridExtra' was built under R version 3.3.3
```

Central Limit Theorem

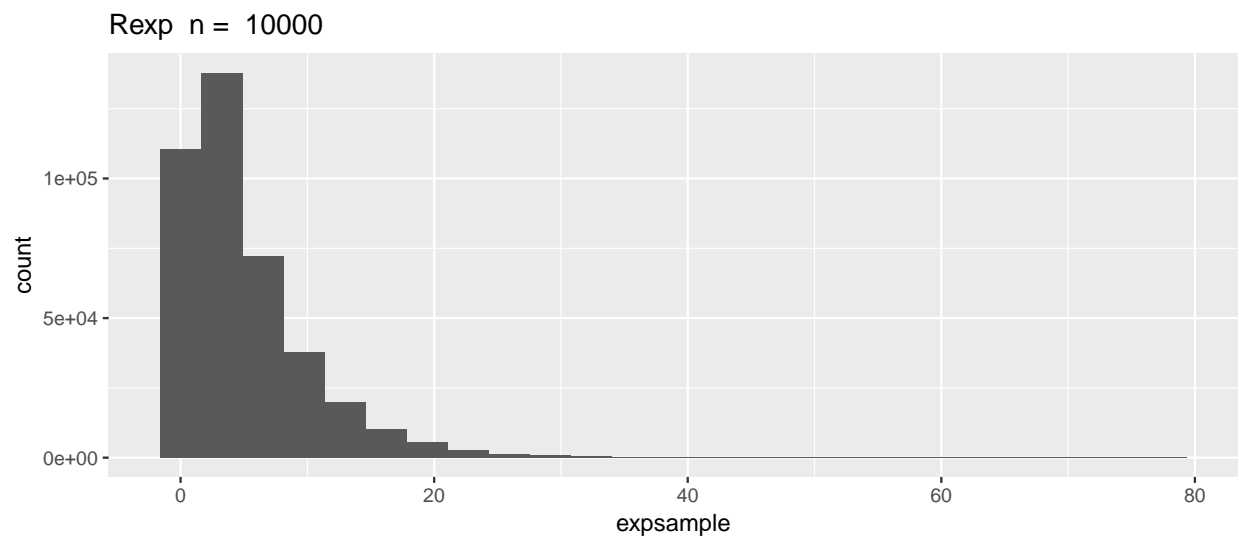
The Central Limit Theorem (CLT) states that the distribution of averages of independent and identically distributed (iid) variables becomes that of a **standard normal** as the sample size increases even if the original variables are not normally distributed. This document investigates the exponential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

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

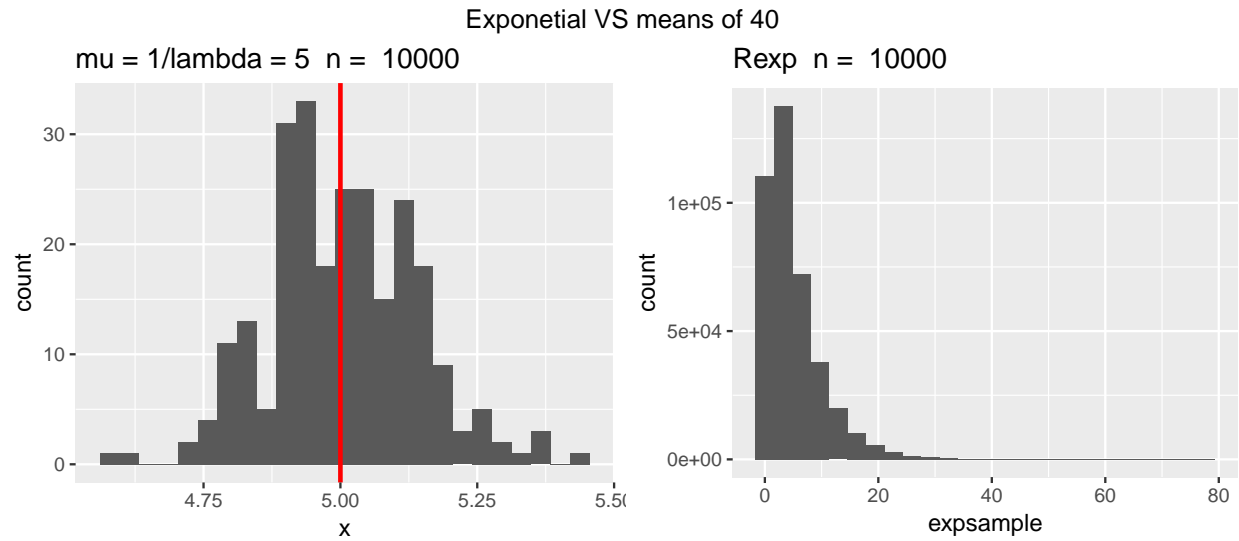
The sample mean compared to the theoretical mean



Don't know how to automatically pick scale for object of type data.frame. Defaulting to continuous.



Don't know how to automatically pick scale for object of type data.frame. Defaulting to continuous.



The sample variance vs theoritical variance

Show that the distribution of the sample means is normal

The histogram shows a distribution with a shape similar to the normal curve. The density proportions shown in the table below are approximately normal.