**Statistical Inference Course Project**

Author: Dale Hunscher

Creation Date: Tuesday, 24 April 2018

## Part 1: Simulation Exercise

### Overview

As stated in the description of the project in Coursera, this part of the project is an exploration of the Central Limit theorem. We'll compare the mean and variance of a series of randomly generated sample distributions of size 40 drawn from the exponential distribution with their theorical values computed using the same lambda value (0.2). Then we will use histograms to do a visual comparison of the sample distribution with the normal paradigm.

### Explorations Part One: Comparing Sample and Theory

Compute the exponential sample data set, and then the sample and theoretical means and variances.

Comparing means: first the sample and theoretical means...

```
## [1] 5.032989
```

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

...and the sample and theoretical variances...

```
## [1] 0.6276361
```

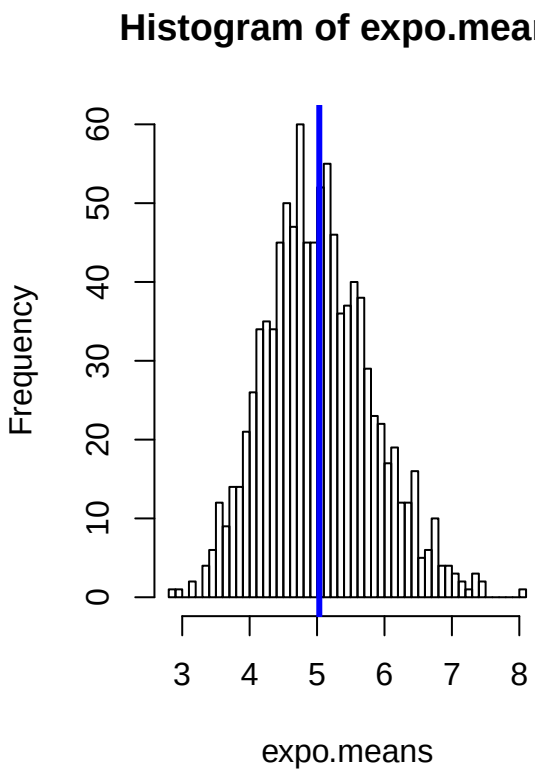
```
## [1] 0.625
```

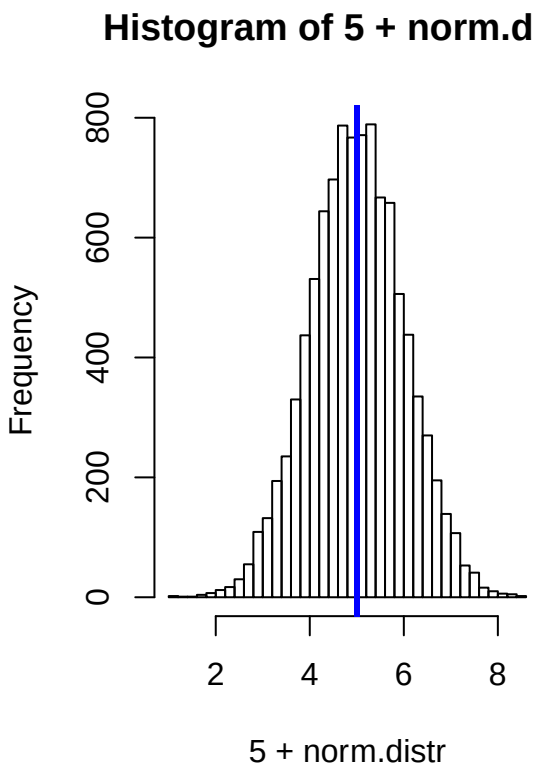
We can see that the mean and variance values computed from the sample data sets are very close to their theoretical equivalents.

### Explorations Part Two: A Visual Comparison

The first histogram below histogram shows the distribution of the sample means. The vertical bar marks the mean of the distribution of sample means.

For comparison purposes, we'll compute a random sample drawn from the normal distribution and show a second histogram. We'll up the sample "n" to 10,000 to ensure it's close to the paradigm for a normal distribution. We'll also add 5 (theoretical mean of the sample exponential distribution) so the histograms' scales are comparable.




We can see that the distribution of even a small set of means of random sample data sets drawn from the exponential distribution approximates the normal paradigm, and the means are very close to the theoretical value for the exponential distribution.

The tails on the sample distribution appear less evenly balanced than the normal distribution. According to the Law of Large Numbers, the distribution of sample means will become closer to the normal distribution as the size of each sample data set increases (for example, from 40 to 1000). The page limit for the project precludes a visual demonstration of this effect, which in any case is not in the project scope.

**Conclusions**

- The sample mean and variance values for are quite close to the theoretical mean and variance coomputed using the same lambda value. This is the expected result, because the sample mean and variance are unbiased estimators of their theoretical equivalents.
- A histogram showing the distribution of means of 1000 sample 40-value data sets drawn from the exponential distribution are visually similar to a histogram of the normal distribution.
- These conclusions demonstrate the Central Limit Theorem in action.

**Appendix**

The Rmd source code for my project documents can be found at https://github.com/DHunscher/Statistical_Inference/tree/master/project

To avoid tempting others to plagiarism, this repository will be removed from public view after course grades have been announced.

The document ending in Appendix is an expanded version of both parts 1 and 2 combined, each part too large to be submitted for this course. This document exposes all code used. Minor and mostly cosmetic code changes have been made to the separate project part documents since the two project parts were split, but the Appendix is substantively correct.