Statistical Inference Week 4 Assignment Part 1

Kaitlyn McGrew

1/16/2018

This project is the final assignment for the course statistical inference part of the data science specialization offered by coursera. It involves two parts a simulation exercise and basic inferential data analysis.

Part 1: Simulation Exercise

as described in the prompt: “In this project you will investigate the exponential distribution in R and compare it with the Central Limit Theorem. The exponential distribution can be simulated in R with rexp(n, lambda) where lambda is the rate parameter. The mean of exponential distribution is 1/lambda and the standard deviation is also 1/lambda. Set lambda = 0.2 for all of the simulations. You will investigate the distribution of averages of 40 exponential. Note that you will need to do a thousand simulations.”

1.Show the sample mean and compare it to the theoretical mean of the distribution. First I will make a matrix of 40 exponential observed through a 1000 simulations. Then take the mean of those observations. The sample mean is `mean2`

```
library(ggplot2)
set.seed(12) ## to make this reproducable
n <- 40
lambda <- 0.2
simulation <- replicate(1000, rexp(n, .2))
mean_simulation <- apply(simulation, 2, mean)
qplot(mean_simulation, geom = "histogram",
      main = "Mean Simulation",
      xlab = "Mean observations")
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```