Project 1 Part 1

Kaylee Walsh

October 23, 2014

About

This is the project for the statistical inference class. In it, you will use simulation to explore inference and do some simple inferential data analysis. The project consists of two parts:

1. Simulation exercises.
2. Basic inferential data analysis.

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 also 1/lambda. Set lambda = 0.2 for all of the simulations. In this simulation, you will investigate the distribution of averages of 40 exponential(0.2)s. Note that you will need to do a thousand or so simulated averages of 40 exponentials.

Part 1

Simulation

For this simulation we wish to replicate 1000 times an exponential distribution of 40 observations centered about 5. To guarantee reproducibility, we set the seed for randomization. We then can create a matrix of data with 1000 rows and 40 columns with each column representing an observation and each row represents a simulation.

```
lambda <- 0.2
pmu <- 1/lambda
psd <- 1/lambda

n <- 40
nsims <- 1000

set.seed(2)
sim.matrix <- matrix(rexp(n*nsims, lambda), nsims)
```

Inferential Data Analysis

Illustrate via simulation and associated explanatory text the properties of the distribution of the mean of 40 exponential(0.2)s. You should

1. Show where the distribution is centered at and compare it to the theoretical center of the distribution.
2. Show how variable it is and compare it to the theoretical variance of the distribution.
3. Show that the distribution is approximately normal.
4. Evaluate the coverage of the confidence interval for 1/lambda: X±1.96 S/sqrt(n).

Center of Distribution

The center of the distribution will lie with the mean which we can find by taking the mean for each simulation and then taking the mean of all simulations.

```
xmu <- mean(apply(sim.matrix, 1, mean))
xmu
```

```
## [1] 5.016
```

We see that the sample mean is 0.33 percent different from the theoretical mean of 5.

Variance

After finding the mean for each simulation, we wish to find the standard deviation of these means. This is noted as the standard error.

```
pse <- 5/sqrt(n) # need to divide sd by sqrt(n) to account for sample size
pse
```

```
## [1] 0.7906
```

```
xsd <- sd(apply(sim.matrix, 1, mean)) # the standard deviation of the means
xsd
```

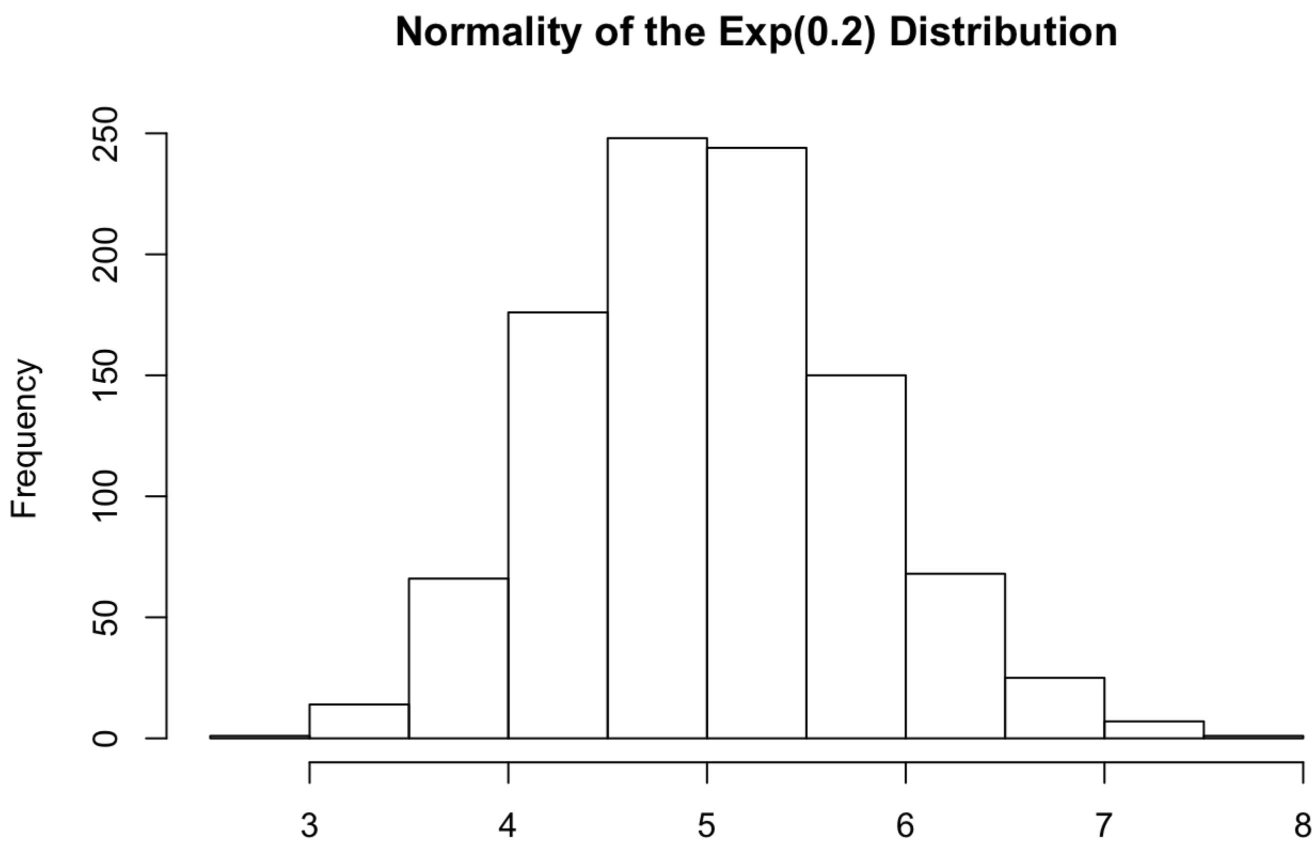
```
## [1] 0.76
```

From the theoretical standard deviation, the standard error is 3.87 percent different.

Normality

Plotting the distribution of means we see this is indeed roughly normal.

```
hist(apply(sim.matrix, 1, mean), main = "Normality of the Exp(0.2) Distribution", xlab = "")
```



Confidence Interval

For the confidence interval we see it is fairly close to the population confidence interval.

```
xmu + c(-1, 1) * 1.96 * xsd
```

```
## [1] 3.527 6.506
```

```
pmu + c(-1, 1) * 1.96 * psd/sqrt(n) # population CI
```

```
## [1] 3.45 6.55
```