

Coursera Data Science Project: Statistical Inference (Part 1)

Rob Rolleston

August 23, 2015

Introduction (Part-1)

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

1. A simulation exercise.
2. Basic inferential data analysis.

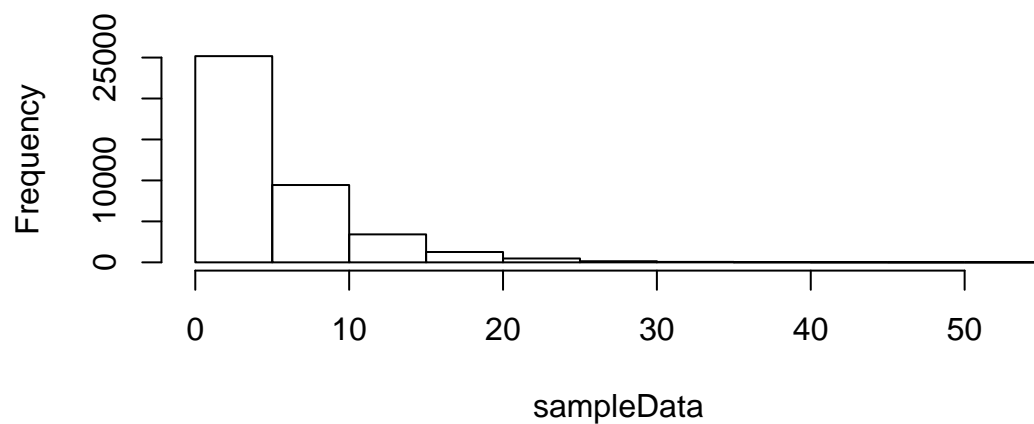
Simulation Exercise

tbd

Simulations

```
nosim <- 1000
n <- 40
lambda <- 0.2
populationMean <- 1./lambda
populationSD <- 1./lambda
sampleData <- matrix(rexp(nosim * n, rate=lambda), nosim)
hist(sampleData)
```

Histogram of sampleData

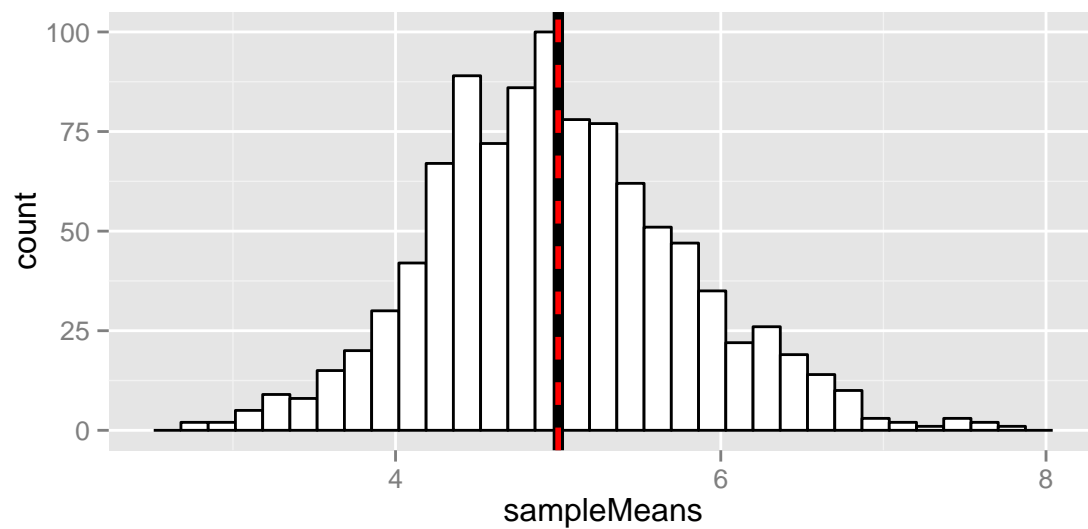


```
sampleData_tbl <- gather(tbl_df(data.frame(sampleData)))
#ggplot(sampleData_tbl, aes(x=value)) + geom_histogram()
#ggplot(sampleData_tbl, aes(x=value)) + geom_density()
```

Sample Mean versus Theoretical Mean

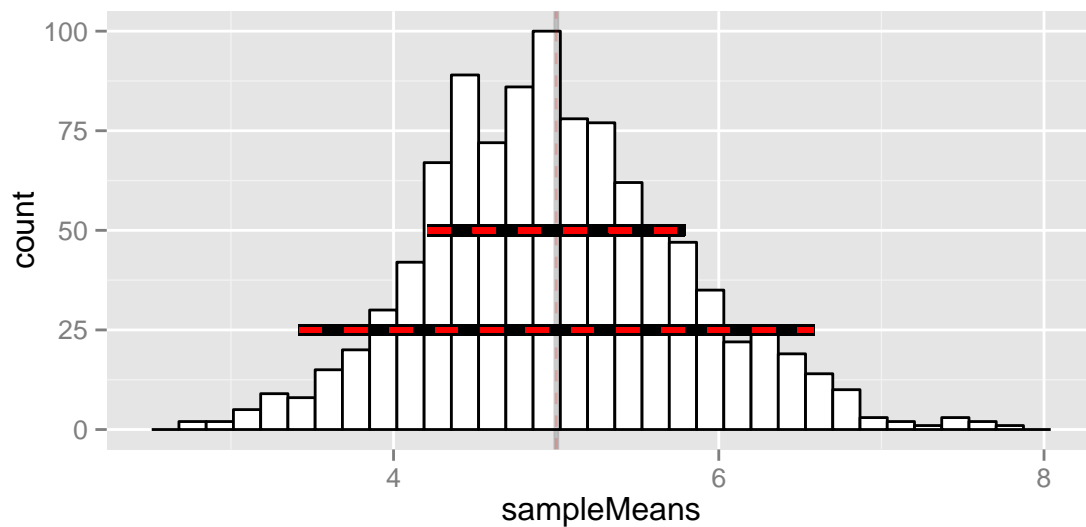
```
sampleMeans <- apply(sampleData, 1, mean)
sampleMean <- mean(sampleMeans)
sampleSD <- sd(sampleMeans)

inferredSD <- populationSD/sqrt(n)
```



Sample Variance versus Theoretical Variance

Compare the inferred variation of the means (inferredSD, via s/\sqrt{n}) with the actual value



Distribution

tbd

```
#x <- seq(0,50, length=100)
#hx <- dexp(x, rate=lambda)
```

