IST-687 Chapter Notes Template: After Completing Please Submit as a PDF.
Originality Assertion: By submitting this file you affirm that this writing is your own.

Name: Yicun Deng
Class: M003
Date: 9/15
Chapter Number: #10
Title of Chapter: Sample in a Jar

Lectures 4.1 - Sampling

We are looking deeper into the sampling this week.
Sample can only reveal part of the nature of the dataset because of the chaotic nature of the dataset
Use sample(x, size=, replace=TRUE) replace means put it back in and reroll
Replicate(number, code) to run the code multiple times
The more times you ran the code, the closer of the guess mean to the real mean(make sense)
The more time you run, the shape of the means of samples would tend to be more normal distribution
quantile(vector, prob=c()) used to see which number fall into which range

Lecture 4.2 - Roundtable on Sampling

The reason why we do sampling is that we might not be able to run the test with the entire dataset and it is more efficient and economy to do sampling.

Lecture 4.3 - Mystery Samples

Use the quantile and summary to compare mystery sample to the actual sample
sd(samples)/sqrt(numbers of samples) = standard or error
Comparison of the existing dataset and new samples can get you a basic idea of what happened to the new data.

R Coding 4.4 - Using sampling

intall.package("moments")
library(moments
skewness() to see whether the data in dataset seat in one side more or the other side more
turn replace to false would disable the possibility of repicking an object over and over.
means of means should be a normal distribution while the actual data's distribution might not.

Roundtable 4.5 - Data Science in the Real World - New Coke Machines

The ultimate goal of data science is to make people's life easier. For example, for people such as manager of Five Guys, they would be able to predict which drink is more popular and should get more through data science of the coke machine.

Exercise Review
Chapter10

```
testSet<- rnorm(20, mean=3090, sd=2)
standarderror <- sd(testSet)/sqrt(20)
```

IST-687 Chapter Notes Template: After Completing Please Submit as a PDF.
Originality Assertion: By submitting this file you affirm that this writing is your own.

quantile(testSet, prob = c(0.025,0.975))

Question for Class
So for homework3, I used the “dumb” way to finish the homework, does it increase or decrease my actual efficiency of running the program though.