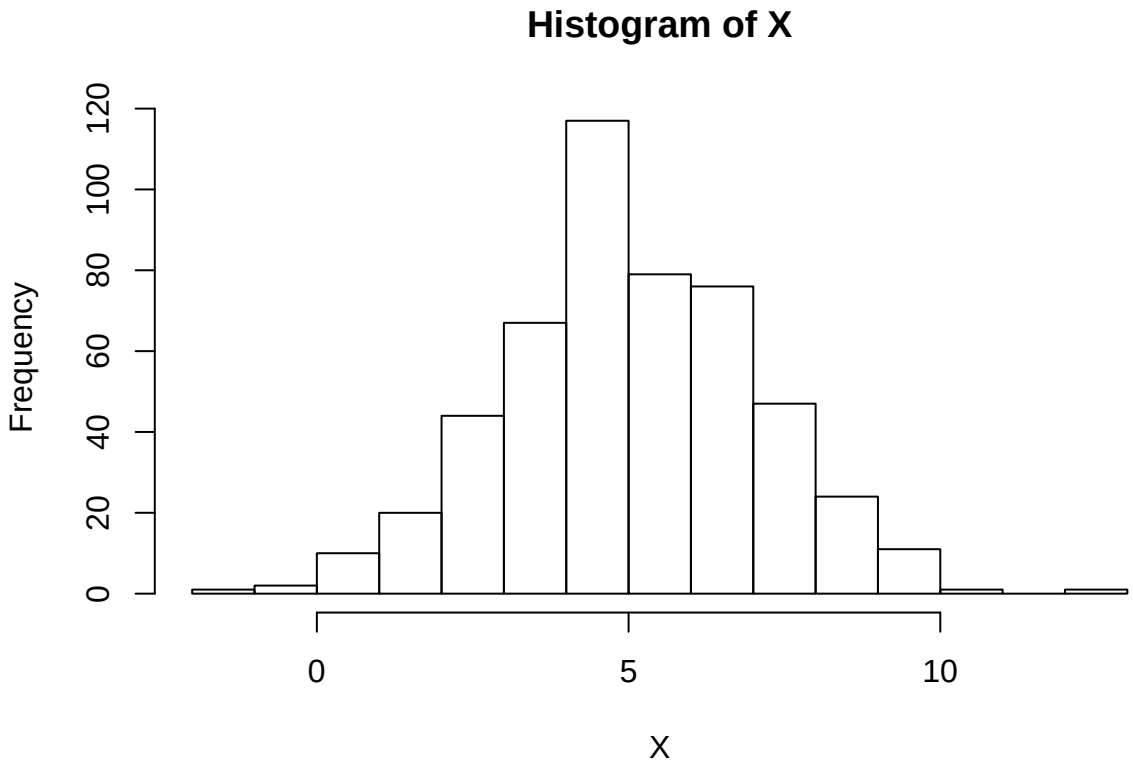
STAT3011 Fall2019 Week8 Lab

Draw A Sample and Calculate Confidence Interval

During today's lab, we will discuss more about the Confidence Interval through simulations. Simulation is an important way to learn statistics. First, let's look at how to generate random sample from a given population.

Suppose we want to draw a sample with size 500, from population follows normal distribution with $\mu = 5$ and $\sigma = 2$. We use function *rnorm()* to do so:

```
X = rnorm(n = 500,mean = 5,sd = 2)
hist(X)
```



Once we get the sample X, we could use the function *t.test()* to get Confidence Interval with given confidence level, for its true mean, μ. Here we get the Confidence Interval with confidence level 0.9:

```
t_test_result = t.test(X, conf.level = 0.9, alternative = "two.sided")
CI = t_test_result$conf.int
CI
```

```
## [1] 4.896136 5.194440
## attr(,"conf.level")
## [1] 0.9
```

We can see, from sample X, we get the 0.9 Confidence Interval is [4.896136, 5.194440]. We can determine if true $\mu = 5$ is in it or not:

```
(5 > CI[1]) & (5 < CI[2])
```

```
## [1] TRUE
```

Draw 10000 Samples and Calculate Confidence Intervals

Now let's draw more samples, say 10000 samples, calculate Confidence Intervals from each of them, and determine if true $\mu = 5$ is in it or not. We use a variable *result* to store the results for each time:

```
number_simulation = 10000
result = NULL ### a null variable, use to store results
for(i in 1:number_simulation)
{
  ### same as previous part
  X = rnorm(500,mean = 5,sd = 2)
  t_test_result = t.test(X, conf.level = 0.9, alternative = "two.sided")
  CI = t_test_result$conf.int

  ### concatenate new result after previous results
  result = c(result,(5 > CI[1]) & (5 < CI[2]))
}
```

Let's see how many times in total 10000 times, we have $\mu = 5$ in the confidence interval:

```
sum(result)
```

```
## [1] 8969
```

We can see, there are 8969 out of 10000 times, the confidence interval cover $\mu = 5$. So, the proportion would be $\frac{8969}{10000} = 0.8969 \approx 0.9$.

Repeat, For A Different Confidence Level

Let's repeat, but change the confidence level to 0.99:

```
number_simulation = 10000
result = NULL ### a null variable, use to store results
for(i in 1:number_simulation)
{
  ### same as previous part
  X = rnorm(500,mean = 5,sd = 2)
  t_test_result = t.test(X, conf.level = 0.99, alternative = "two.sided")
  CI = t_test_result$conf.int

  ### concatenate new result after previous results
  result = c(result,(5 > CI[1]) & (5 < CI[2]))
}
sum(result)
```

```
## [1] 9916
```

Now we can see, there are 9916 our of 10000 times, the confidence interval cover $\mu = 5$. So, the proportion would be $\frac{9916}{10000} = 9916 \approx 0.99$.