Statistical Inference Course Project

1. Simulation Exercise

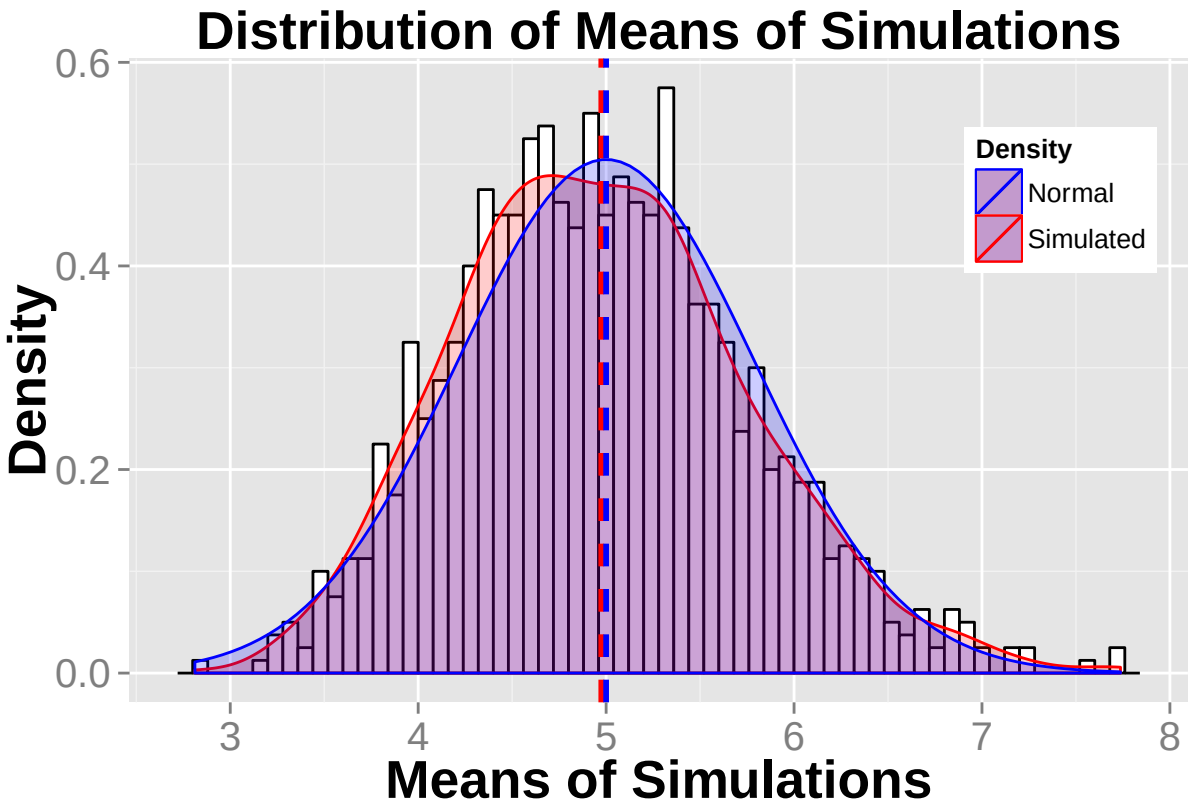
1. Show where the distribution is centered at and compare it to the theoretical center of the distribution

```
##      [,1] [,2]                                  [,3]
## [1,] "1"  "Theoretical Centre of distribution" "5"
## [2,] "1"  "Centre of Simulation Distribution"  "4.97423877125153"
```

2. Show how variable it is and compare it to the theoretical variance of the distribution.

```
##      [,1] [,2]                                   [,3]
## [1,] "2"  "Theoretical Variance of Distribution" "25"
## [2,] "2"  "Vairance of Simulation Distribution"  "23.3726136368744"
```

3. Show that the distribution is approximately normal. See image below: This shows the distribution of means of the 1000 sample Simulations is approximately normally distributed



4. Evaluate the coverage of the confidence interval
Using Equation

$$X_{sample} \pm 1.96 * \frac{S}{\sqrt{(n)}}$$

Confidence Interval

```
## [1] 4.675 5.274
```

Confidence Interval Coverage

```
## [1] 0.04947
```

2. Basic Inferential Data Analysis

1. Load the Tooth Growth Dataset & Basic Exploratory Analysis

Types of Supplement

```
## [1] VC OJ
## Levels: OJ VC
```

Doses Used in Trial

```
## [1] 0.5 1.0 2.0
```

Number of Each Supp

```
## [1] 30
```

```
## [1] 30
```

Number of Each Dose

```
## [1] 20
```

```
## [1] 20
```

```
## [1] 20
```

2. Basic Summary of Data

```
##       len       supp         dose
##  Min.   : 4.2   OJ:30   Min.   :0.50
##  1st Qu.:13.1   VC:30   1st Qu.:0.50
##  Median :19.2           Median :1.00
##  Mean   :18.8           Mean   :1.17
##  3rd Qu.:25.3           3rd Qu.:2.00
##  Max.   :33.9           Max.   :2.00
```

Supplements and Doses Summary Table

```
##    supp dose meanLength sdLength
## 1:   VC  0.5       7.98    2.747
## 2:   VC  1.0      16.77    2.515
## 3:   VC  2.0      26.14    4.798
## 4:   OJ  0.5      13.23    4.460
## 5:   OJ  1.0      22.70    3.911
## 6:   OJ  2.0      26.06    2.655
```

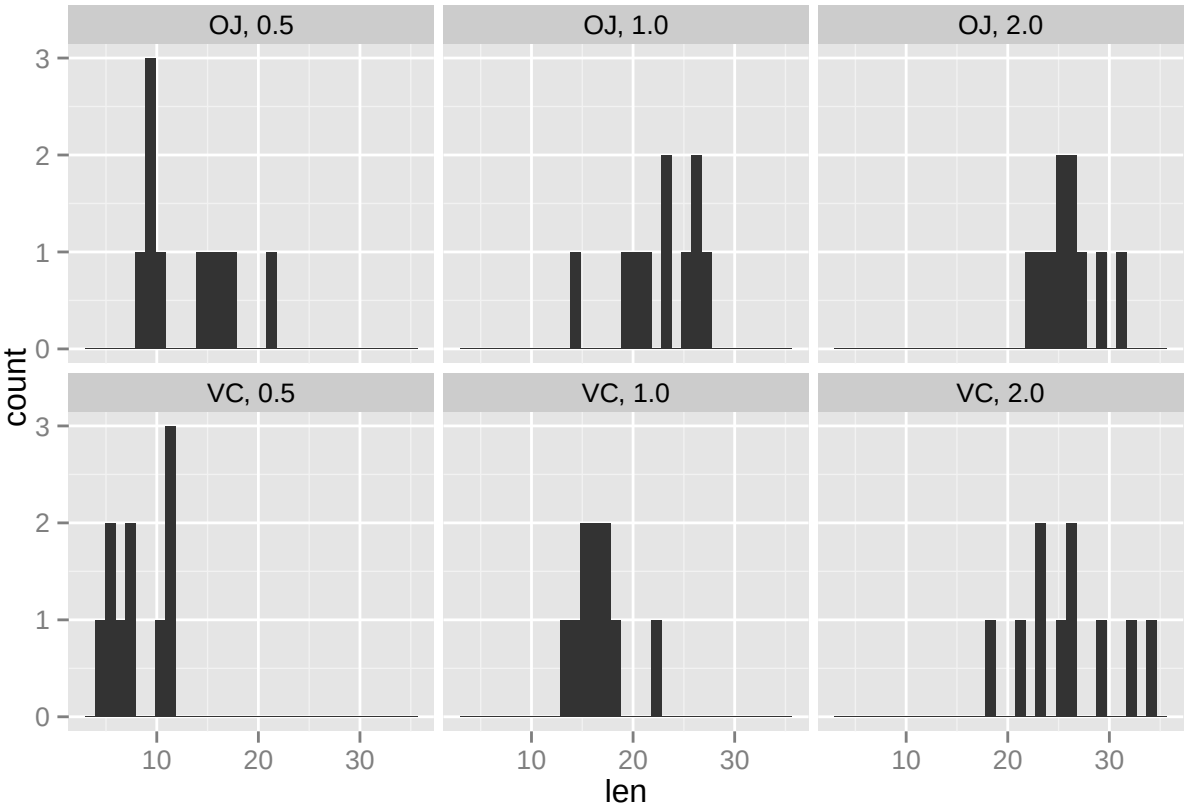
Supplements Summary Table

```
##    supp meanLength sdLength
## 1:   VC      16.96    8.266
## 2:   OJ      20.66    6.606
```

Doses Summary Table

```
##    dose meanLength sdLength
## 1:  0.5      10.61    4.500
## 2:  1.0      19.73    4.415
## 3:  2.0      26.10    3.774
```

Graphical Summary of Supplements and Doses



3. Use Confidence Intervals and Hypothesis Tests to compare tooth growth by supplement and dose

a) Confidence Intervals

Comparing by Supp ### What is 95% CI for the Mean length using VC?

```
## [1] 14 20
```

What is 95% CI for the Mean length using OJ?

```
## [1] 14 20
```

Independent t confidence interval for OJ - VC

```
## [1] -0.17  7.57
```

Most of the time the OJ out performs the VC with 95% confidence

Comparing by Dose

Independent t confidence interval comparing doses 0.5 vs 1.0 milligrams

```
## [1]  6.28 11.98
```

Dose of 1.0 always out performs does of 0.5 in stimulating tooth growth with 95% confidence

Independent t confidence interval comparing doses 0.5 vs 2.0 milligrams

```
## [1] 12.84 18.15
```

Dose of 2.0 always out performs does of 0.5 in stimulating tooth growth with 95% confidence

Independent t confidence interval comparing doses 2.0 vs 1.0 milligrams

```
## [1] 3.74 8.99
```

Dose of 2.0 always out performs does of 1.0 in stimulating tooth growth with 95% confidence

b) Hypothesis Test

Comparing by Supp

```
##
##  Welch Two Sample t-test
##
## data:  dt[supp == "OJ"]$len and dt[supp == "VC"]$len
## t = 1.915, df = 55.31, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  -0.171  7.571
## sample estimates:
## mean of x mean of y
##     20.66     16.96

##     mean    sd
## 1: 18.81 7.649

## [1] "Reject H0"
```

Comparing by Dose

```
##
##  Welch Two Sample t-test
##
## data:  dt[dose == 0.5]$len and dt[dose == 1]$len
## t = -6.477, df = 37.99, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.984  -6.276
## sample estimates:
## mean of x mean of y
##     10.61     19.73

##
##  Welch Two Sample t-test
##
## data:  dt[dose == 0.5]$len and dt[dose == 2]$len
## t = -11.8, df = 36.88, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.16 -12.83
## sample estimates:
## mean of x mean of y
##     10.61     26.10

##
##  Welch Two Sample t-test
##
## data:  dt[dose == 1]$len and dt[dose == 2]$len
## t = -4.901, df = 37.1, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996 -3.734
## sample estimates:
## mean of x mean of y
##     19.73     26.10
```

4. Conclusions and Assumptions

Conclusions

1. Some Conculsions

Assumptions

1. Use t interval as not sure if data is normally distributed
2. Assume unequal variances for t distribution confidence interval
3. Central limit theorm for Z test
4. n must be large enough to be statistically signifcant
5. If n is small then Gossett's T test is used, n is small for each set of tests so use t test
6. Assuming a constant variance between groups of Guinea Pigs receiving difference amounts of treatment and different supplements