Statistical Inference Project - Part 2

Andrew Mendonca

July 28, 2017

Overview

This report will analyze the ToothGrowth data in the R datasets package.

Analysis

1. Load the data and perform some exploratory data analyses

Load the following library.

```
library(ggplot2)
```

Load the data.

```
data("ToothGrowth")
```

View all of the dataset variables.

```
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

Convert the variable dose from numeric to factor

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
```

See the variable after conversion.

```
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

2. Provide a basic summary of the data.

Summarize all stats for every single variable

```
summary(ToothGrowth)
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

Display the first few rows of the ToothGrowth data.

```r
head(ToothGrowth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

3. Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose.

Compare tooth growth by supplement.

```r
t.test(len ~ supp, data = ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

The p-value is 0.06 and the confidence interval is 0. This means that we cannot ignore the null hypothesis that supplement types have no effect on tooth growth.

Compare tooth growth by dose based on the different pairs of dose values.

```r
ToothGrowth_subdata <- subset(ToothGrowth, ToothGrowth$dose %in% cbind(1.0, 0.5))
t.test(len ~ dose, data = ToothGrowth_subdata)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

```r
ToothGrowth_subdata <- subset(ToothGrowth, ToothGrowth$dose %in% cbind(0.5, 2.0))
t.test(len ~ dose, data = ToothGrowth_subdata)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

```
ToothGrowth_subdata <- subset(ToothGrowth, ToothGrowth$dose %in% cbind(1.0, 2.0))
t.test(len ~ dose, data = ToothGrowth_subdata)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```

The p-value of each test is 0 and the confidence interval of each test is also 0.

So, we can assume that the average tooth length increases with a higher dose, and thus, the null hypothesis is ignored.

Conclusions

We made a few assumptions that the sample is representative of the population and the distribution of the sample means follows the Central Limit Theorem.

Based on the tests above, we conclude that the supplement delivery method does not have any effect on tooth growth, but higher dosages result in increased tooth length.