Statistical Inference Course Project - Part 2

Chandrasekar Ganesan

September 1, 2017

Overview

Objective is to analyze the ToothGrowth data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

Tooth Growth Data Analysis

```
library(ggplot2)
data("ToothGrowth")
summary(ToothGrowth)
```

```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

```
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
```

```
unique(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
## 7  11.2   VC  0.5
## 9   5.2   VC  0.5
## 10  7.0   VC  0.5
## 11 16.5   VC  1.0
## 13 15.2   VC  1.0
```

```
## 14 17.3   VC  1.0
## 15 22.5   VC  1.0
## 17 13.6   VC  1.0
## 18 14.5   VC  1.0
## 19 18.8   VC  1.0
## 20 15.5   VC  1.0
## 21 23.6   VC  2.0
## 22 18.5   VC  2.0
## 23 33.9   VC  2.0
## 24 25.5   VC  2.0
## 25 26.4   VC  2.0
## 26 32.5   VC  2.0
## 27 26.7   VC  2.0
## 28 21.5   VC  2.0
## 29 23.3   VC  2.0
## 30 29.5   VC  2.0
## 31 15.2   OJ  0.5
## 32 21.5   OJ  0.5
## 33 17.6   OJ  0.5
## 34  9.7   OJ  0.5
## 35 14.5   OJ  0.5
## 36 10.0   OJ  0.5
## 37  8.2   OJ  0.5
## 38  9.4   OJ  0.5
## 39 16.5   OJ  0.5
## 41 19.7   OJ  1.0
## 42 23.3   OJ  1.0
## 43 23.6   OJ  1.0
## 44 26.4   OJ  1.0
## 45 20.0   OJ  1.0
## 46 25.2   OJ  1.0
## 47 25.8   OJ  1.0
## 48 21.2   OJ  1.0
## 49 14.5   OJ  1.0
## 50 27.3   OJ  1.0
## 51 25.5   OJ  2.0
## 52 26.4   OJ  2.0
## 53 22.4   OJ  2.0
## 54 24.5   OJ  2.0
## 55 24.8   OJ  2.0
## 56 30.9   OJ  2.0
## 58 27.3   OJ  2.0
## 59 29.4   OJ  2.0
## 60 23.0   OJ  2.0
```

```
unique(ToothGrowth$dose)
```

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

```
unique(ToothGrowth$len)
```

```
##  [1]  4.2 11.5  7.3  5.8  6.4 10.0 11.2  5.2  7.0 16.5 15.2 17.3 22.5 13.6
## [15] 14.5 18.8 15.5 23.6 18.5 33.9 25.5 26.4 32.5 26.7 21.5 23.3 29.5 17.6
## [29]  9.7  8.2  9.4 19.7 20.0 25.2 25.8 21.2 27.3 22.4 24.5 24.8 30.9 29.4
## [43] 23.0
```

```
unique(ToothGrowth$supp)
```

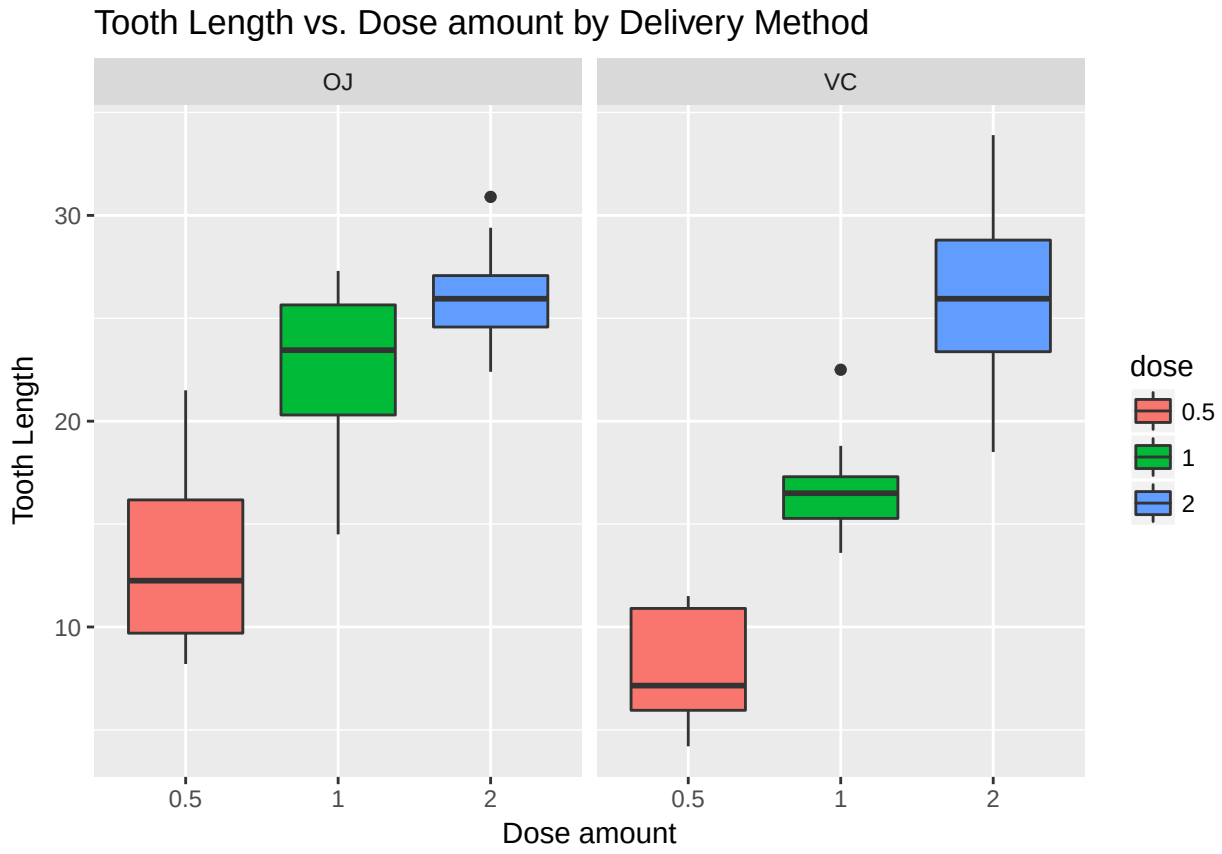
```
## [1] VC OJ
## Levels: OJ VC
```

Convert dose as factor variable

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

Plot the data

```
g = ggplot(aes(x = dose, y = len), data = ToothGrowth)
g = g + geom_boxplot(aes(fill=dose))
g = g + xlab("Dose amount") + ylab("Tooth Length")
g = g + facet_grid(~ supp)
g = g + ggtitle("Tooth Length vs. Dose amount by Delivery Method")
g
```




```
g = ggplot(aes(x = supp, y = len), data = ToothGrowth)
g = g + geom_boxplot(aes(fill=supp))
g = g + xlab("Delivery Method") + ylab("Tooth Length")
g = g + facet_grid(~ dose)
g = g + ggtitle("Tooth Length vs. Delivery Method by Dose amount")
g
```

Compare tooth growth by supplement using t-test

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

```
##
##  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 of the test is **0.06** (greater than 0.05) The confidence interval of the test **contains zero**. We can conclude that supplement types seems to have no impact to tooth growth based on this test.

CASE A: Compare tooth growth by dose (0.5 and 1.0)

```
ToothGrowth_dosage <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,1))
test <- t.test(len ~ dose, data=ToothGrowth_dosage)
test
```

```
##
##  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
```

Case A Analysis The p-value of the test is almost **0.00** (less than 0.05) The confidence interval of the test **does not cross zero**.

CASE B: Compare tooth growth by dose (1.0 and 2.0)

```
ToothGrowth_dosage <- subset(ToothGrowth, ToothGrowth$dose %in% c(1,2))
test <- t.test(len ~ dose, data=ToothGrowth_dosage)
test
```

```
##
##  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
```

Case B Analysis The p-value of the test is almost **0.00** (less than 0.05) The confidence interval of the test **does not cross zero**.

CASE C: Compare tooth growth by dose (0.5 and 2.0)

```
ToothGrowth_dosage <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,2))
test <- t.test(len ~ dose, data=ToothGrowth_dosage)
test
```

```
##
##  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
```

Case C Analysis The p-value of the test is almost **0.00** (less than 0.05) The confidence interval of the test **does not cross zero**.

Conclusion

Refering to analysis from Case A, B & C, we can find that the p-value of each test is close to zero and each confidence interval does not cross over zero. Based on this we can assume that the average tooth length

proportional to dose and null hypothesis can be rejected.

Overall, we can also conclude that the supplement delivery method has no effect on tooth growth, but an increased dosages do result in increased tooth length.