Statistical Inference Project 2: ToothGrowth Analysis

Andreas Hadimulyono

Overview

This analysis shall look into the ToothGrowth dataset provided within datasets package in R, perform exploratory data analysis, and try to compare the tooth growth by supp and dose.

Exploratory Analysis

```
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 ...
```

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

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

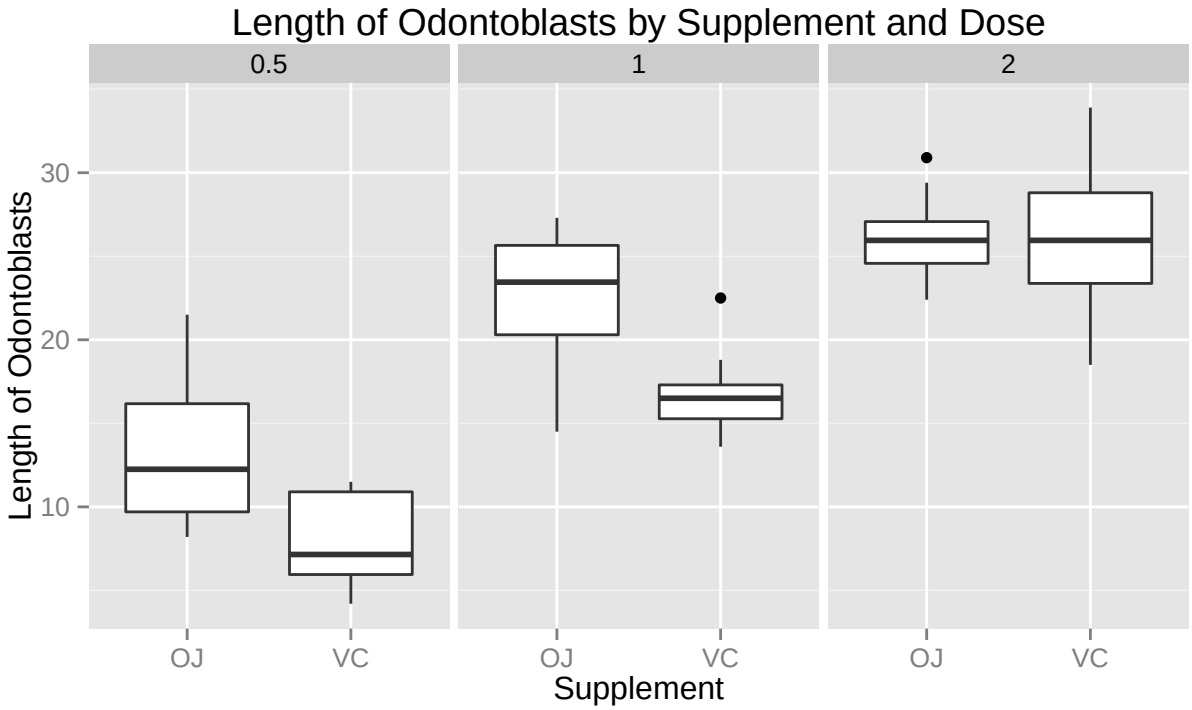
The ToothGrowth dataset contains 60 observations, and 3 variables. The dose variable is numeric, however looking at the data it seems it can be changed to Factor since the values are discrete.

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
analysisDataSet <- ToothGrowth
dcast(analysisDataSet, supp + dose ~ ., length)
```

```
## Using dose as value column: use value.var to override.
```

```
##   supp dose  .
## 1   OJ  0.5 10
## 2   OJ    1 10
## 3   OJ    2 10
## 4   VC  0.5 10
## 5   VC    1 10
## 6   VC    2 10
```

The table above describes the number of observations grouped by supp and dose. We have exactly 10 subjects for each combination of supp and dose. The analysis shall try to determine if there is any difference between two supplements, “OJ” and “VC”, at each dose level. Figure 1 shows the box plot of each group of observations.



Hypothesis

The following null hypotheses are proposed:

- There is no difference in effect to length of Odontoblasts between delivery method of OJ and VC at dose 0.5 milligram per day (H0_0.5: u = 0 and Ha_0.5: u != 0)
- There is no difference in effect to length of Odontoblasts between delivery method of OJ and VC at dose 1 milligram per day (H0_1 = 0: u = 0 and Ha_1: u != 0)
- There is no difference in effect to length of Odontoblasts between delivery method of OJ and VC at dose 2 milligram per day (H0_2 = 0 and Ha_2 != 0)

Assumptions

We are going to make some assumptions on the dataset:

- The underlying data are IID and Gaussian
- The distribution of the underlying data is roughly symmetric and mound shaped
- The variance between OJ and VC are assumed to be not equal
- The data is not paired

T Tests and Conclusions

As the dataset measures the difference in a measured variable among two different groups, the author decides to use T test to calculate the statistic. The codes used to perform the T test is described in the Appendix.

T Test for the group with 0.5mg Vitamin C per day

```
##
##  Welch Two Sample t-test
##
## data:  subset05[subset05$supp == "OJ", "len"] and subset05[subset05$supp == "VC", "len"]
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

From the T statistic above, we can reject the null hypothesis, which means at the dosage of 0.5mg/day, the method of delivery is likely to make a difference to the measured length of Odontoblasts.

T Test for the group with 1mg Vitamin C per day

```
##
##  Welch Two Sample t-test
##
## data:  subset1[subset1$supp == "OJ", "len"] and subset1[subset1$supp == "VC", "len"]
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

From the T statistic above, we can reject the null hypothesis, which means at the dosage of 1mg/day, the method of delivery is likely to make a difference to the measured length of Odontoblasts.

T Test for the group with 2mg Vitamin C per day

```
##
##  Welch Two Sample t-test
##
## data:  subset2[subset2$supp == "OJ", "len"] and subset2[subset2$supp == "VC", "len"]
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

From the T test above, we fail to reject the null hypothesis, which means at the dosage of 2mg/day, the method of delivery is NOT likely to make a difference to the measured length of Odontoblasts.

APPENDIX

T TEST CODES:

T test for 0.5mg/day

```
subset05 <- analysisDataSet[analysisDataSet$dose=="0.5",]
t.test(subset05[subset05$supp =="OJ","len"],
       subset05[subset05$supp =="VC","len"],
       var.equal = FALSE,
       paired=FALSE)
```

T test for 1mg/day

```
subset1 <- analysisDataSet[analysisDataSet$dose=="1",]
t.test(subset1[subset1$supp =="OJ","len"],
       subset1[subset1$supp =="VC","len"],
       var.equal = FALSE,
       paired=FALSE)
```

T test for 2mg/day

```
subset2 <- analysisDataSet[analysisDataSet$dose=="2",]
t.test(subset2[subset2$supp =="OJ","len"],
       subset2[subset2$supp =="VC","len"],
       var.equal = FALSE,
       paired=FALSE)
```