Statistical Inference Project - Part 2

yxy

Saturday, January 17, 2015

Part 2 Basic Inferential Data Analysis

Overview

In this part of the project, the dataset "ToothGrowth" is analyzed. The data record the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two supplement types (orange juice/OJ or ascorbic acid/VC).

Basic Exploratory Analysis

Data were first loaded and a simple exploratory analysis was performed. We can see that the data have 60 observations and 3 variables: len, supp, and dose. supp is a factor with two supplement types: OJ and VC.

```
library(datasets)
data(ToothGrowth)
head(ToothGrowth)
```

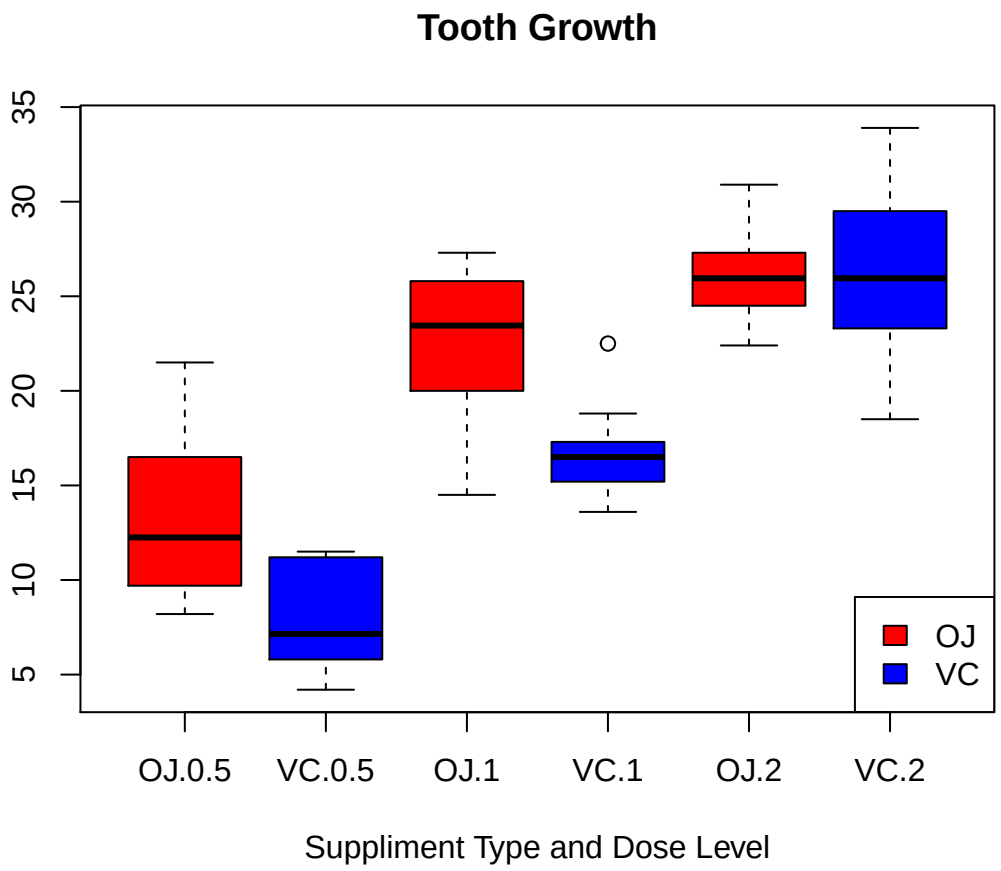
```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
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 ...
```

Fig. 3 shows that the tooth length increases when the dose level increases.

```
# Box Plots
boxplot(len ~ supp * dose, data = ToothGrowth,
        col = (c("red","blue")), main = "Tooth Growth",
        xlab="Suppliment Type and Dose Level")
legend('bottomright', c("OJ", "VC"),
       fill = c("red", "blue"))
```



Basic Summary of the Data

The summary statistics of the data are shown below:

```
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
```

Compare tooth growth by supp and dose

An independent t-test was used to compare the tooth growth when two different supplement types were administered. The results show that the 95% confidence interval is [-0.17, 7.57], which include the value 0. Also, the p-value = 0.06 is larger than 0.05. Thus, the null hypothesis that the means of the tooth growth for OJ and VC are not significantly different is accepted.

```r
# Compare tooth growth by supp
bysupp <- split(ToothGrowth$len, ToothGrowth$supp)
# OJ vs VC
t.test(bysupp$OJ, bysupp$VC)
```

```
##
##  Welch Two Sample t-test
##
## data:  bysupp$OJ and bysupp$VC
## 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 of x mean of y
##  20.66333  16.96333
```

To compare the tooth growth for the three different dosage levels, an ANOVA could be used. As ANOVA has not been covered in this course, three independent t-tests were used. For dosage 0 and dosage 0.5, their tooth growth are significantly different. This is supported by a p-value smaller than 0.05 and a 95% confidence interval that does not include 0.

```r
# Compare tooth growth by supp
bydose <- split(ToothGrowth$len, ToothGrowth$dose)
# Dosage 0.5 vs Dosage 1
t.test(bydose$'0.5', bydose$'1')
```

```
##
##  Welch Two Sample t-test
##
## data:  bydose$"0.5" and bydose$"1"
## 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 of x mean of y
##    10.605    19.735
```

For dosage 0 and dosage 2, their tooth growth are also significantly different. This is supported by the evidence that p-value < 0.05 and the 95% confidence interval does not include 0.

```r
# Dosage 0.5 vs Dosage 2
t.test(bydose$'0.5', bydose$'2')
```

```
##
##  Welch Two Sample t-test
##
## data:  bydose$"0.5" and bydose$"2"
## 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 of x mean of y
##    10.605    26.100
```

For dosage 1 and dosage 2, their tooth growth are also significantly different. This is supported by the evidence that p-value < 0.05 and the 95% confidence interval does not include 0.

```
# Dosage 0.5 vs Dosage 2
t.test(bydose$'0.5', bydose$'2')
```

```
##
##  Welch Two Sample t-test
##
## data:  bydose$"0.5" and bydose$"2"
## 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 of x mean of y
##    10.605    26.100
```

Conclusions

The first assumption made in this analysis is that this study is a randomized control trial where the guinea pigs were randomly allocated to one of the supplement type with different dosage level. Also, the data are normally distributted. Thus, an independent t-test was used to compare the tooth growth for different treatments. one or other of the different treatments under study. The results from the t-tests show that:
1. The mean tooth growth for the supplement type OJ and VC are not significantly different. 2. The mean tooth growth for the different dose levels are significantly different. Further analysis shows that the tooth length is correlated with the dose level.