Statistical Inference - Course Project 2

K H Chen

Feb 20, 2014

About This Project

This project is to analyze the ToothGrowth data in the R datasets package and perform some exploratory data analysis and perform hypothesis testing on the data. The response is the length of 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 delivery methods (orange juice or ascorbic acid).

1. Basic Exploratory Data Analysis

```
#Loading the Library Required #
library(ggplot2)

# Loading the dataset #
library(datasets)
data(ToothGrowth)

# Checking the dataset #
nrow(ToothGrowth)
```

```
## [1] 60
```

```
# Previewing Data #
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 ...
```

```
# Pre-processing of Data #
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
```

2. Summary of the Data

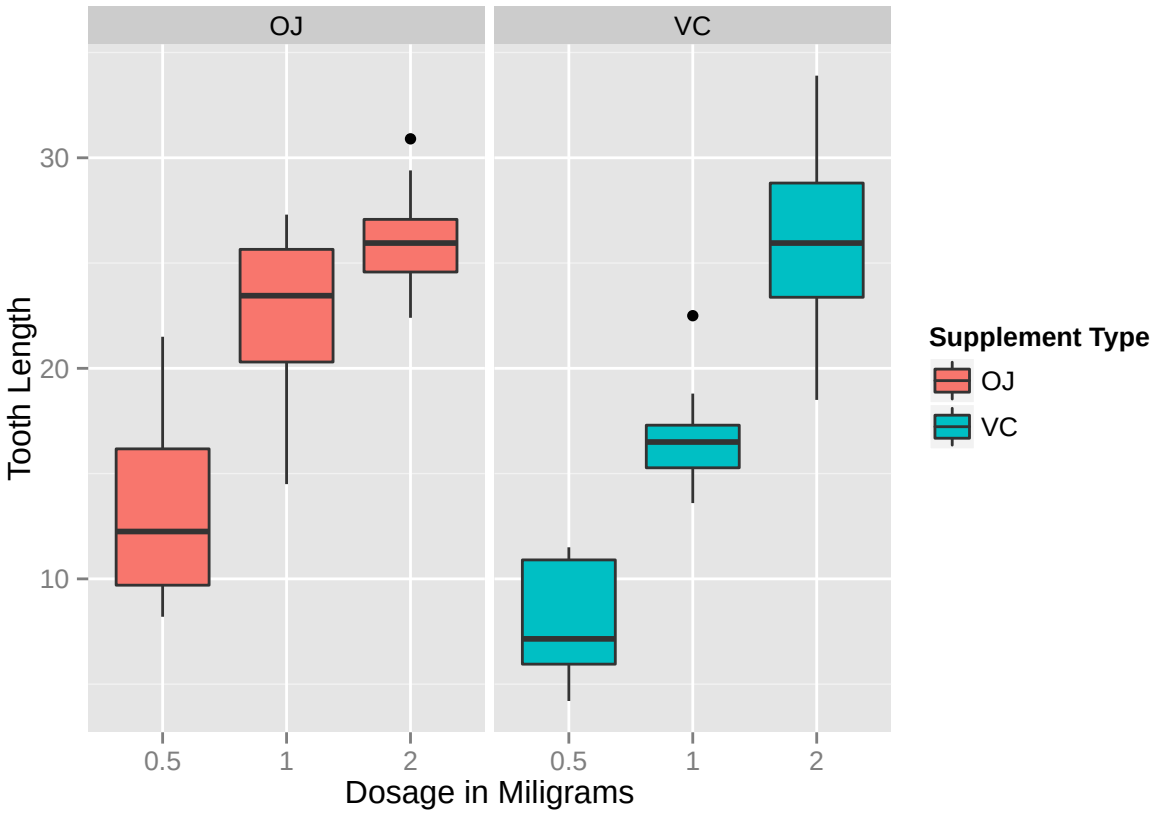
```
# Generating Statistics of the Data #
summary(ToothGrowth)
```

```
##       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
```

```
# Spliting the data #
table(ToothGrowth$dose, ToothGrowth$supp)
```

```
##
##       OJ VC
##   0.5 10 10
##   1   10 10
##   2   10 10
```

```
ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) +
    geom_boxplot() +
    facet_grid(. ~ supp) +
    xlab("Dosage in Miligrams") +
    ylab("Tooth Length") +
    guides(fill=guide_legend(title="Supplement Type"))
```



Based on the summary, there is a positive correlation between the tooth length and the dosage levels for both delivery method, as the length of the tooth increases when the dosage increases. It also showns taht OJ delivery method yields a greater length increate than the VC.

3. Confidence Intervals and Hypothesis Tests

Statistical inference

Hypothesis testing is used to compare whether the same dosage of each supplement yields the same tooth growth length. The null hypothesis test is as following:

$H_0 : \mu_{OJ|0.5} = \mu_{VC|0.5}$
$H_0 : \mu_{OJ|1.0} = \mu_{VC|1.0}$
$H_0 : \mu_{OJ|2.0} = \mu_{VC|2.0}$

```
# Slitting the Data by Dosages #
d0.5 <- subset(ToothGrowth, dose == 0.5)
d1.0 <- subset(ToothGrowth, dose == 1.0)
d2.0 <- subset(ToothGrowth, dose == 2.0)

# Conducting a t-test between Supplements #
test0.5 <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = d0.5)
test0.5$p.value; test0.5$conf[1]
```

```
## [1] 0.006358607
```

```
## [1] 1.719057
```

```r
test1.0 <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = d1.0)
test1.0$p.value; test1.0$conf[1]
```

```
## [1] 0.001038376
```

```
## [1] 2.802148
```

```r
test2.0 <- t.test(len ~ supp, paired = FALSE, var.equal = FALSE, data = d2.0)
test2.0$p.value; test2.0$conf[1]
```

```
## [1] 0.9638516
```

```
## [1] -3.79807
```

Observations

a. P Value of Dosages 0.5 and 1.0

Dosages 0.5 and 1.0 have p-values of 0.0063586 and 0.0010384, respectively. As is below 0.01, this shows there is a strong presumption against the null hypothesis. It is show there is a significant difference in mean values between the supplements.

b. Confidence Interval of Dosage 1.0 and 2.0

Dosage 1.0 and 2.0 have a confidence interval between 1.7190573 - 8.7809427 and 2.8021482 - 9.0578518, respectively.

c. P Value and Confidence Interval of Dosage 2.0

Dosage 2.0 has a very high p-value of 0.9638516 and a confidence interval below zero (-3.7980705 - 3.6380705). The p-value of Dosage shows there is no presumption against the null hypothesis. Therefore, there is no significance between the supplemen at the dosage at this amount.

4. Conclusion and Assumption

Based on the above analysis, supplements with lower dosages (0.5 and 1.0) are more effective on the tooth length growth. However, supplement with higher dosage (2.0) has shown the decrease of the effect on tooth growth.

Assumptions Applied

a. Test subjects are similar size, age and diet

b. Samples are unpaired, with unequal variances

c. No other factors other affecting tooth growth