Statistical Inference Project Part 2

J.E. Panzik

5/27/2020

Overview

This report aims to analyze the ToothGrowth data in the R datasets package. Per the course project instructions, the following items should occur:

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 are other approaches worth considering).
4. State your conclusions and the assumptions needed for your conclusions.

The analysis below comes to the conclusion that the type of supplement given IS NOT statistically significant, but the dosage of supplement IS statistically significant in the tooth length of guinea pigs.

Data: Description and Exploratory Plots

The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice or ascorbic acid (a form of vitamin C and coded as VC).

```
#Load the datasets and open ToothGrowth data
library(datasets)
data(ToothGrowth)
```

The following code displays the structure of the data and how many incomplete entries exist.

```
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 ...
```

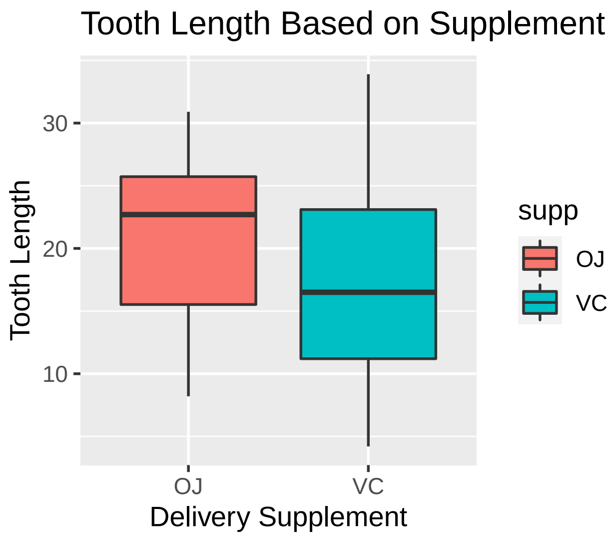
```
sum(!complete.cases(ToothGrowth))
```

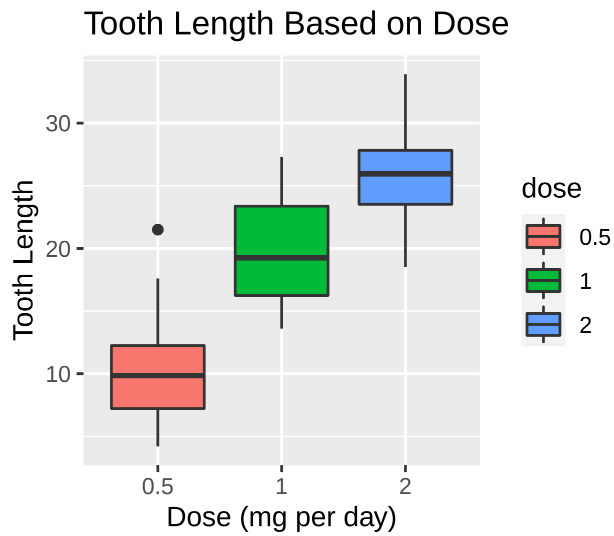
```
## [1] 0
```

The following plots show a quick visual behavior of tooth length of the subject guinea pigs related to the supplements given and the doses given.

```
library(ggplot2); library(gridExtra)
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
ToothGrowth$supp <- as.factor(ToothGrowth$supp)
```

```r
#First plot of tooth length vs supplement
p1 <- ggplot(ToothGrowth, aes(x=supp, y=len, fill=supp)) +
      geom_boxplot() +  ggtitle("Tooth Length Based on Supplement") +
      xlab("Delivery Supplement") + ylab("Tooth Length")
#Create second plot of tooth length vs dose
p2 <- ggplot(ToothGrowth, aes(x=dose, y=len,fill=dose)) +
      geom_boxplot() +  ggtitle("Tooth Length Based on Dose") +
      xlab("Dose (mg per day)") + ylab("Tooth Length")
#Combine both plots
grid.arrange(p1, p2, nrow=1)
```





Hypothesis Testing: Impact of Supplement

For all hypothesis testing, data is analyzed using a t test and only the t statistic, p value, and confidence interval are displayed.
Null Hypothesis: The type of supplement **does not** impact tooth growth.
Alternate Hypothesis: The type of supplement **does** impact tooth growth.

```r
tSupp <- t.test(len ~ supp, data = ToothGrowth)$statistic
pSupp <- t.test(len ~ supp, data = ToothGrowth)$p.value
confSupp <- t.test(len ~ supp, data = ToothGrowth)$conf.int
```

Test Statistic: 1.9152683
Confidence Interval: -0.1710156, 7.5710156
p Value: 0.0606345

Hypothesis Testing: Impact of Dose

Null Hypothesis: The dose amount **does not** impact tooth growth.
Alternate Hypothesis: The dose amount **does** impact tooth growth.
To test the significance between each dose, the dataset is broken up into subsets based on the two dosages being compared.

Dosage: 0.5 mg per day vs. 1 mg per day

```r
ToothGrowth1 <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,1))
tDose1 <- t.test(len ~ dose, data = ToothGrowth1)$statistic
pDose1 <- t.test(len ~ dose, data = ToothGrowth1)$p.value
confDose1 <- t.test(len ~ dose, data = ToothGrowth1)$conf.int
```

Test Statistic: -6.4766477
Confidence Interval: -11.9837813, -6.2762187
p Value: 1.2683007×10^{-7}

Dosage: 0.5 mg per day vs. 2 mg per day

```r
ToothGrowth2 <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,2))
tDose2 <- t.test(len ~ dose, data = ToothGrowth2)$statistic
pDose2 <- t.test(len ~ dose, data = ToothGrowth2)$p.value
confDose2 <- t.test(len ~ dose, data = ToothGrowth2)$conf.int
```

Test Statistic: -11.799046
Confidence Interval: -18.1561665, -12.8338335
p Value: 4.397525×10^{-14}

Dosage: 1 mg per day vs. 2 mg per day

```r
ToothGrowth3 <- subset(ToothGrowth, ToothGrowth$dose %in% c(1,2))
tDose3 <- t.test(len ~ dose, data = ToothGrowth3)$statistic
pDose3 <- t.test(len ~ dose, data = ToothGrowth3)$p.value
confDose3 <- t.test(len ~ dose, data = ToothGrowth3)$conf.int
```

Test Statistic: -4.9004843
Confidence Interval: -8.9964805, -3.7335195
p Value: 1.9064295×10^{-5}

Conclusions

Supplements
Based on the t test applied to variation of tooth length between supplements the type of supplement *does not* have a statistically significant difference. Because the p value: 0.0606345>0.5, and the confidence interval: (-0.1710156, 7.5710156) contains 0, we cannot reject the null hypothesis.

Dosage
Based on the t test applied to variation of tooth length between dosages 0.5 and 1, the dosage amount *does* have a statistically significant difference. Because the p value: 1.2683007×10^{-7}<0.5, and the confidence interval: (-11.9837813, -6.2762187) does not contain 0, we reject the null hypothesis.

Based on the t test applied to variation of tooth length between dosages 0.5 and 2, the dosage amount *does* have a statistically significant difference. Because the p value: 4.397525×10^{-14}<0.5, and the confidence interval: (-18.1561665, -12.8338335) does not contain 0, we reject the null hypothesis.

Based on the t test applied to variation of tooth length between dosages 1 and 2, the dosage amount *does* have a statistically significant difference. Because the p value: 1.9064295×10^{-5}<0.5, and the confidence interval: (-8.9964805, -3.7335195) does not contain 0, we reject the null hypothesis.

Supplemental Material

Though the analysis above only compared all the doses of supplements to each other. There may be a different relation between each supplement of the same dose. The analysis does not explore this further, but displays the posibility.

```
#Create plot for tooth length vs dose separated by supplement
p3 <- ggplot(ToothGrowth, aes(x=dose, y=len, fill=dose)) +
      geom_boxplot() +  ggtitle("Tooth Length Based on Dose by Supplement") +
      xlab("Dose (mg per day)") + ylab("Tooth Length") +facet_grid(~supp)
#Create plot for tooth length vs supplement separated by dosage
p4 <- ggplot(ToothGrowth, aes(x=supp, y=len, fill=supp)) +
      geom_boxplot() +  ggtitle("Tooth Length Based on Supplement by Dose") +
      xlab("Delivery Supplement") + ylab("Tooth Length") +facet_grid(~dose)
#Show plots as 1 column grid
grid.arrange(p3, p4, ncol=1)
```