Statistics Course Assignment 2 Tooth Growth

JB

September 26, 2017

Overview

The following report describes a basic statistical analysis of the tooth growth data from the R library. 60 guinea pigs were split into two groups of 30. One group was given Vitamin C delivered by orange juice (OJ). The other group was given vitamin C delivered as ascorbic acid (VC). In each group, the guinea pigs were further divided into three groups given Vitamin C at three different dose rates, 0.5, 1 and 2 mg/day. This was done to evaluate the effect of Vitamin C dose rate and delivery method on tooth growth in the guinea pigs.

Exploratory Analysis

In this section some basic exploratory evaluation of the tooth growth data is done after it has been loaded.

Install the needed plotting libraries and load the data. Convert the dose column to a factor for plotting.

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.4.1
```

```
library(gridExtra)
```

```
## Warning: package 'gridExtra' was built under R version 3.4.1
```

```
tooth_data <- data.frame(ToothGrowth)
tooth_data$dose <- as.factor(tooth_data$dose)
```

Show the summary of the tooth growth data.

```
summary(tooth_data)
```

```
##       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
```

Data Analysis

The following section goes into more detail in analyzing the tooth growth data. Specifically a box and whisker plot is done to show the median, quartiles and range of the tooth lengths for different delivery methods and dose rates. Also, hypothesis testing is done to determine if there are statistically significant differences in tooth length across different delivery methods and dose rates.

Box Plot

The following figure shows a box and whisker plot showing tooth lengths across the three different dose rates for each of the two Vitamin C delivery methods: orange juice ("OJ") and ascorbic acid ("VC").

```
box_plot <- ggplot(tooth_data, aes(dose,len)) + geom_boxplot(color="black",
fill="orange")+ facet_grid(.~supp) +
labs(title="Relationship Between Tooth Length and Vitamin C Dose")+
labs(x=" Vitamin C Dose(mg/day)", y="Tooth length") + theme(text = element_text(size=16))
print(box_plot)
```

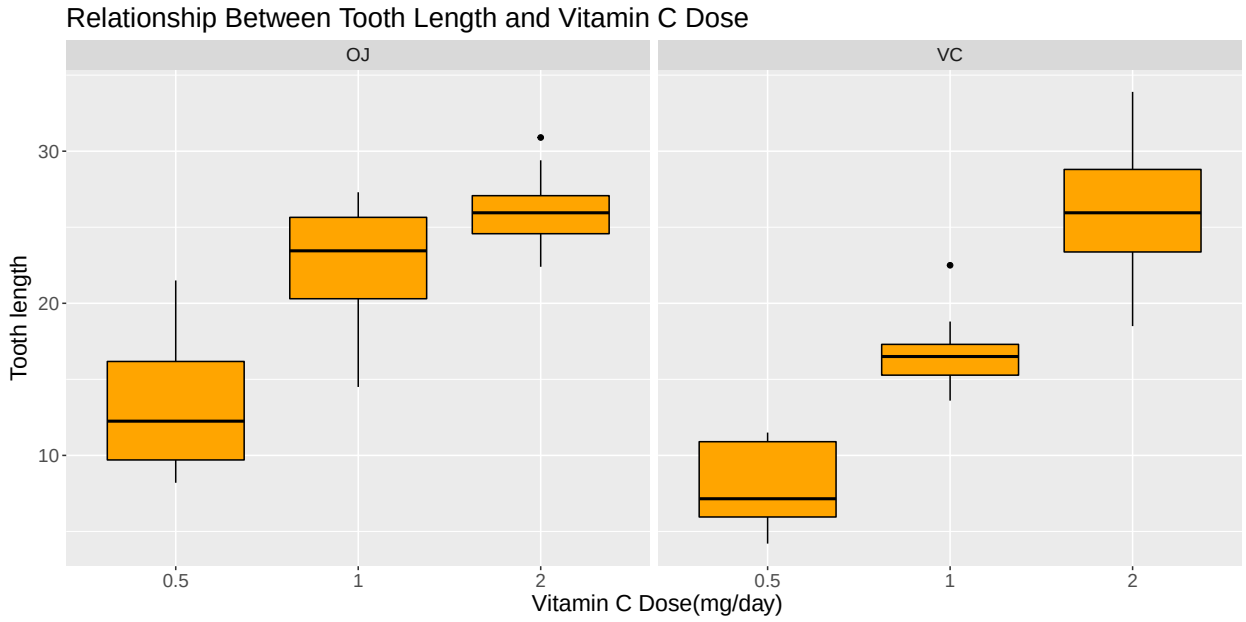

Figure 1: Box and whisker plot for tooth length for each guinea pig group assigned a specific delivery method and dose rate.

This figure shows that increasing the Vitamn C dose shows a trend of increasing tooth length for both delivery methods, but the spread in tooth length overall seems slightly larger with the ascorbic acid guinea pig group.

Hypothesis Testing

A two-sample, unpaired Student's t-test will be used to compare tooth length between the OJ and VC Vitamin C delivery groups. This test was chosen because two small sets of data will be compared that were collected independently from one another.This test will be done both one tailed and two tailed. For the two tailed test, the alternative hypothesis is that the mean tooth lengths between the OJ and VC delivery groups are different. For the one-tailed test, the alternative hypothesis is that the mean tooth length in the OJ group is larger than in the VC group. Only the one-tailed output will be shown since the two-tailed test will have a P-value of double the one-tail test result.

Two tailed test:

```
t.test(tooth_data$len~tooth_data$supp)
```

One-tailed test:

```
t.test(tooth_data$len~tooth_data$supp, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  tooth_data$len by tooth_data$supp
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
```

```
## 95 percent confidence interval:
##  0.4682687       Inf
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

It is also feasible to compare the tooth lengths versus dose rates using this test. One tailed tests will be used to see if the mean tooth length at higher dose rates is larger than the tooth length at lower dose rates.

Get the tooth lengths for each dose rate, for both delivery methods.

```
tooth_length_oj_dose1 <-tooth_data$len[tooth_data$dose==1 & tooth_data$supp =="OJ"]
tooth_length_oj_dose0p5 <-tooth_data$len[tooth_data$dose==0.5 & tooth_data$supp =="OJ"]
tooth_length_oj_dose2 <-tooth_data$len[tooth_data$dose==2 & tooth_data$supp =="OJ"]

tooth_length_vc_dose1 <-tooth_data$len[tooth_data$dose==1 & tooth_data$supp =="VC"]
tooth_length_vc_dose0p5 <-tooth_data$len[tooth_data$dose==0.5 & tooth_data$supp =="VC"]
tooth_length_vc_dose2 <-tooth_data$len[tooth_data$dose==2 & tooth_data$supp =="VC"]
```

Do the hypothesis tests comparing the tooth lengths between different dose rates, for each delivery method. Only the output of OJ delivery comparing dose levels 1 and 2 mg/day will be shown to save space.

```
t.test(tooth_length_oj_dose1,tooth_length_oj_dose0p5, alternative = "greater")
t.test(tooth_length_vc_dose1,tooth_length_vc_dose0p5, alternative = "greater")
t.test(tooth_length_vc_dose2,tooth_length_vc_dose1, alternative = "greater")
```

```
t.test(tooth_length_oj_dose2,tooth_length_oj_dose1, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  tooth_length_oj_dose2 and tooth_length_oj_dose1
## t = 2.2478, df = 15.842, p-value = 0.0196
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.7486236       Inf
## sample estimates:
## mean of x mean of y
##     26.06     22.70
```

Conclusions

For the delivery comparison, these results show that in the one-tailed case, the p-value is less than the significance level α=0.05, meaning that we can reject the null hypothesis and assume that the mean tooth length with OJ is larger than with VC delivery. However,the two-tailed test shows that the null hypothesis of there being no difference in mean tooth lengths between the two delivery methods cannot be rejected since the P-value is larger than α=0.05. The two-tailed test is a stricter test, and perhaps it shows that more data are required to determine if significant differences between mean tooth lengths are present for the two Vitamin C delivery methods.

For the dose rate comparisons, all P-values showed that increasing dose rate increased tooth length for a significance level of $\alpha = 0.05$. Even if two-tailed tests had been done, the P-values would be doubled but still less than $\alpha = 0.05$, so one could accept the alternative hypothesis that increasing dose rate increases tooth length, for this significance level.