Statistical Inference Project Part 2

Camille Tolentino

10/7/2020

Introduction

In this report we will analyze the ToothGrowth data in the R data sets package which shows the effect of vitamin C on tooth growth in guinea pigs. 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).

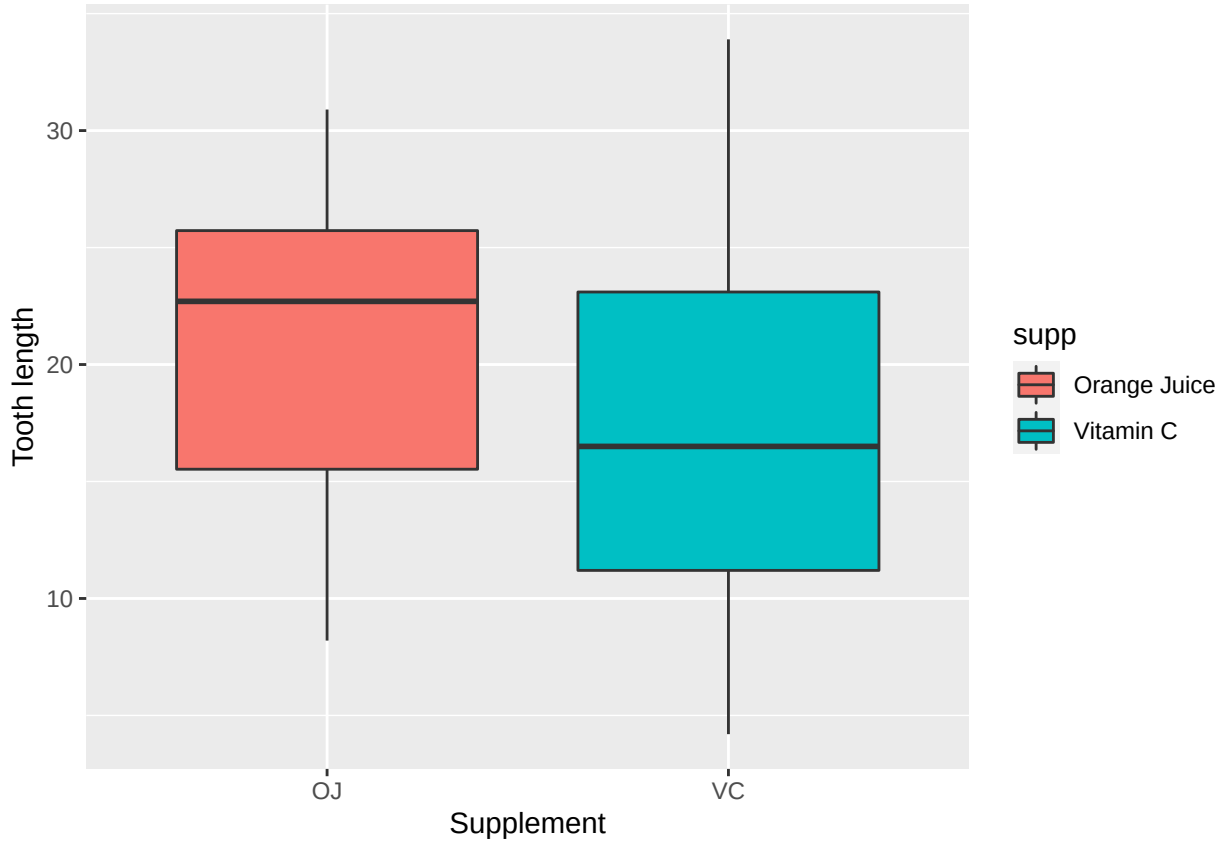
```
library(ggplot2)
data("ToothGrowth")
summary(ToothGrowth)
```

```
##       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
```

Does the supplement significantly affect tooth growth?

First, we look at the the effect of the two supplements on tooth length.

```
ggplot(aes(x=supp, y=len), data=ToothGrowth)+
  geom_boxplot(aes(fill=supp))+
  xlab("Supplement") +ylab("Tooth length") +
  scale_fill_discrete(labels = c("Orange Juice", "Vitamin C"))
```



Visually, orange juice and vitamin C are neck and neck based on tooth length. To support this, we can conduct a t-test. We consider our null hypothesis to be $H_0 : \mu_{OJ} = \mu_{VC}$ that is, there is no significant difference between the mean tooth length for guinea pigs given orange juice supplements (mean = 20.66) and vitamin C (mean = 16.96).

```
t.test(ToothGrowth$len[ToothGrowth$supp=="OJ"],
       ToothGrowth$len[ToothGrowth$supp=="VC"],
       paired = FALSE,var.equal = FALSE)
```

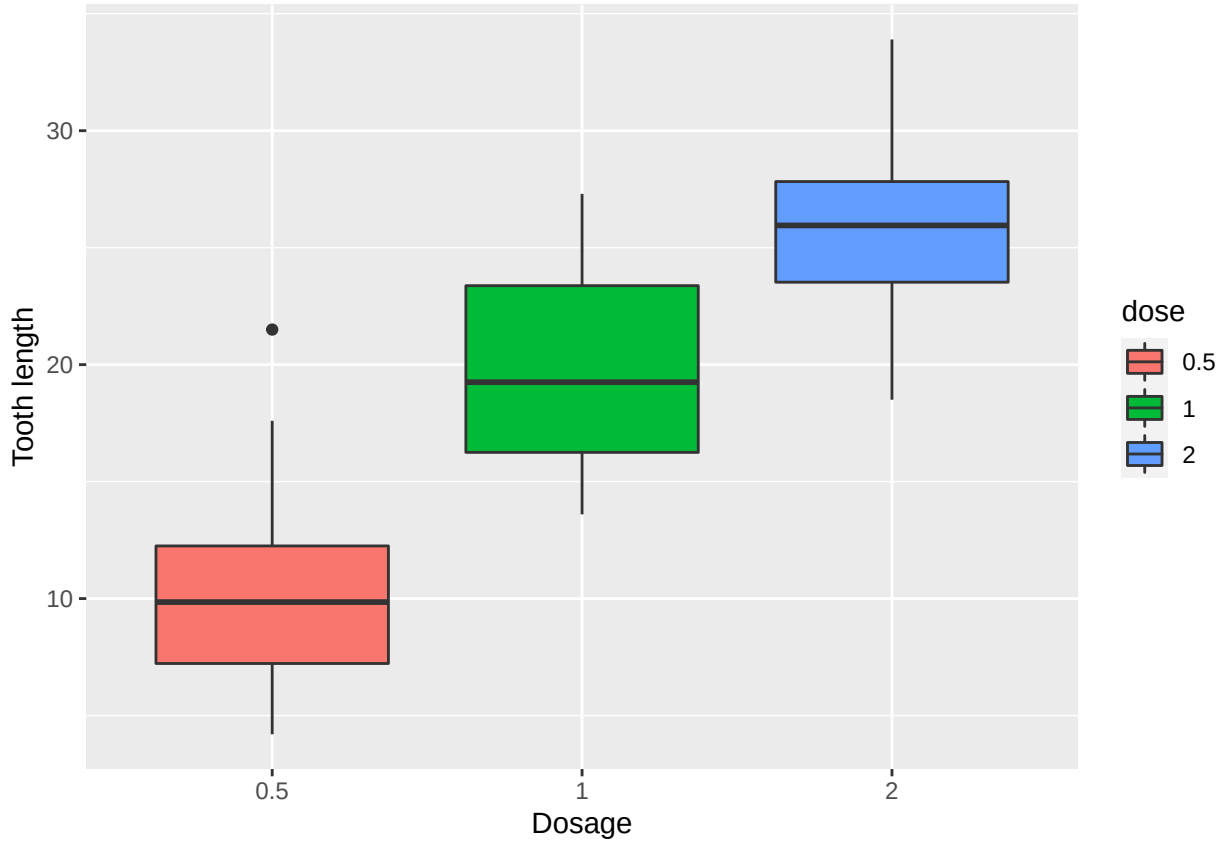
```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$supp == "OJ"] and ToothGrowth$len[ToothGrowth$supp == "VC"]
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

With $p - value = 0.06063 > 0.05$ and the interval $(-0.17, 7.57)$ containing 0 we can conclude that at 95 confidence level, we cannot reject the null hypothesis. This means, the difference between the change in length from orange juice and vitamin C is not significant.

Does dosage significantly affect tooth growth?

Similar to what we have done with the supplements, we first look at the box plot for each dosage. But first, we change the class of dosage from numeric to factor for us to separate each observation.

```
ToothGrowth$dose <-as.factor(ToothGrowth$dose)
ggplot(aes(x=dose, y=len), data=ToothGrowth)+
  geom_boxplot(aes(fill=dose))+
  xlab("Dosage") +ylab("Tooth length")
```



Looking at the means, we can see that the three dosages have spread out values.

```
sapply(split(ToothGrowth$len,ToothGrowth$dose),mean)
```

```
##    0.5      1      2
## 10.605 19.735 26.100
```

We can then check this observation using a t-test between dosages. For our null hypothesis, we can choose $H_0 : \mu_1 = \mu_2$.

```
t.test(ToothGrowth$len[ToothGrowth$dose=="1"],
       ToothGrowth$len[ToothGrowth$dose=="2"],
       paired = FALSE,var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == "1"] and ToothGrowth$len[ToothGrowth$dose == "2"]
## t = -4.9005, df = 37.101, p-value = 1.906e-05
```

```
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

With $p - value \approx 0 < 0.05$ and the interval $(-9, -3.73)$ lying strictly on one side of 0 we can conclude that at 95 confidence level, we can reject the null hypothesis. This means, the difference between the change in length from the increase in dosage is significant.

Conclusion

Supplement type does not significantly contribute to tooth length while dosage change significantly affects tooth length.