Tooth Growth Analysis

Yadder Aceituno

January 24, 2018

Overview

In this analysis we will use the "Toot Growth" data, this data set registers the effect of vitamin C on pigs' tooth growth. Over this data we will discover some relations on the data's attributes. To confirm that relations we will use plots and t tests to get some information.

Loading Data and Summary

For this analysis I will use the "Tooth Growth" data. Let's load it and view a summary of the data.

```
data("ToothGrowth");
```

Now, let's look a summary of the data.

```
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
```

Now, let's look around the firsts rows of the 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
```

You can find more information about the data clicking on this link

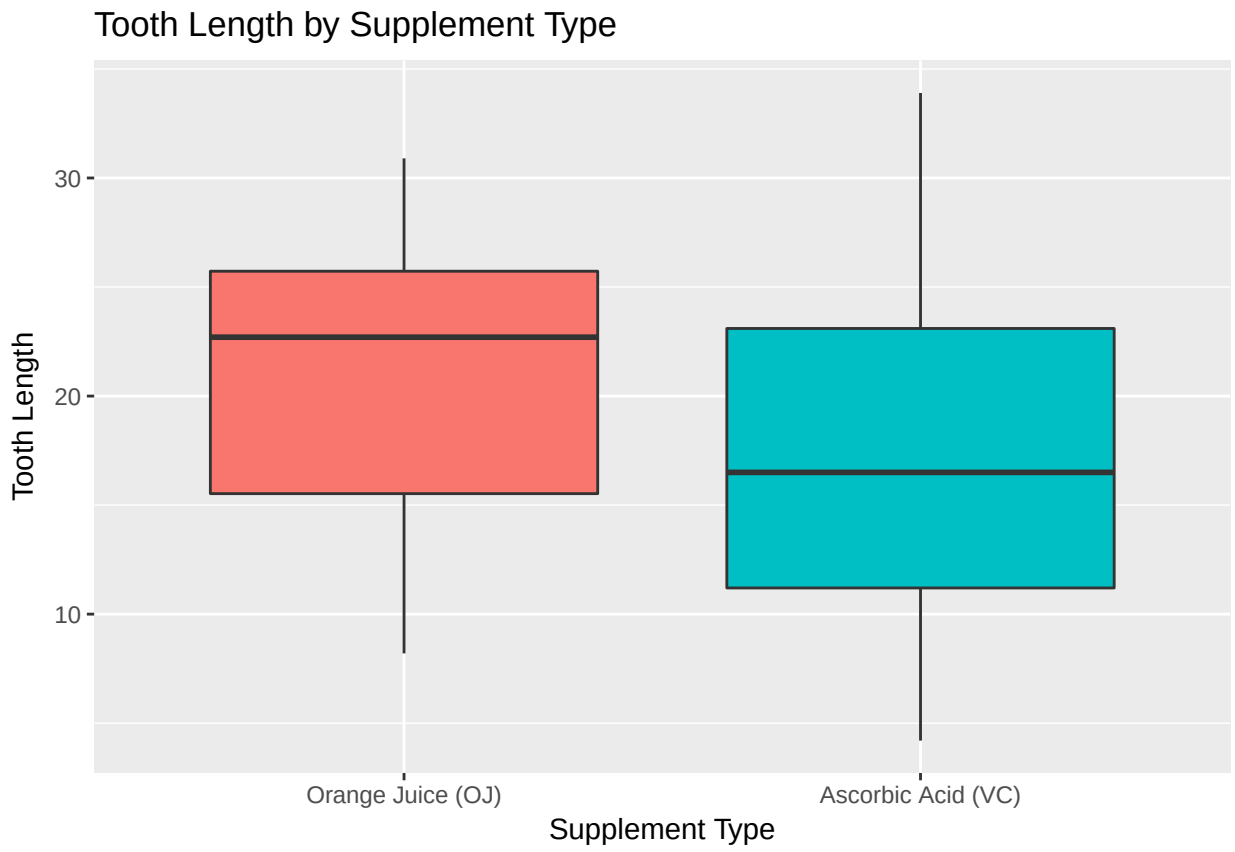
Comparing By Delivery Method

I will create a plot about length by delivery method to easily explore the data.

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose);

plot <- ggplot(aes(x=supp, y=len), data = ToothGrowth);
plot <- plot + geom_boxplot(aes(fill=supp)) +
  xlab("Supplement Type") +
  ylab("Tooth Length") +
  ggtitle("Tooth Length by Supplement Type") +
```

```
  # Removes the legends
  theme(legend.position="none") +
  # Labeling supplement type
  scale_x_discrete(labels=c("Orange Juice (OJ)","Ascorbic Acid (VC)"));
plot;
```



With that plot I can see that both samples looks like they are almost the same. So it could be possible that the supplement type (delivery method) doesn't have relation with tooth length . Now, in order to discover that assumption, I will use t test to compare the data by supplement(delivery method).

```
t.test(len~supp, data = ToothGrowth);
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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 in group OJ mean in group VC
##         20.66333         16.96333
```

Notice that the p-value (0.06) is greater than 0.05. So we can state that there isn't difference between both data sets (Orange Juice delivery and Ascorbic Acid) related to tooth growth.

Comparing By Doses

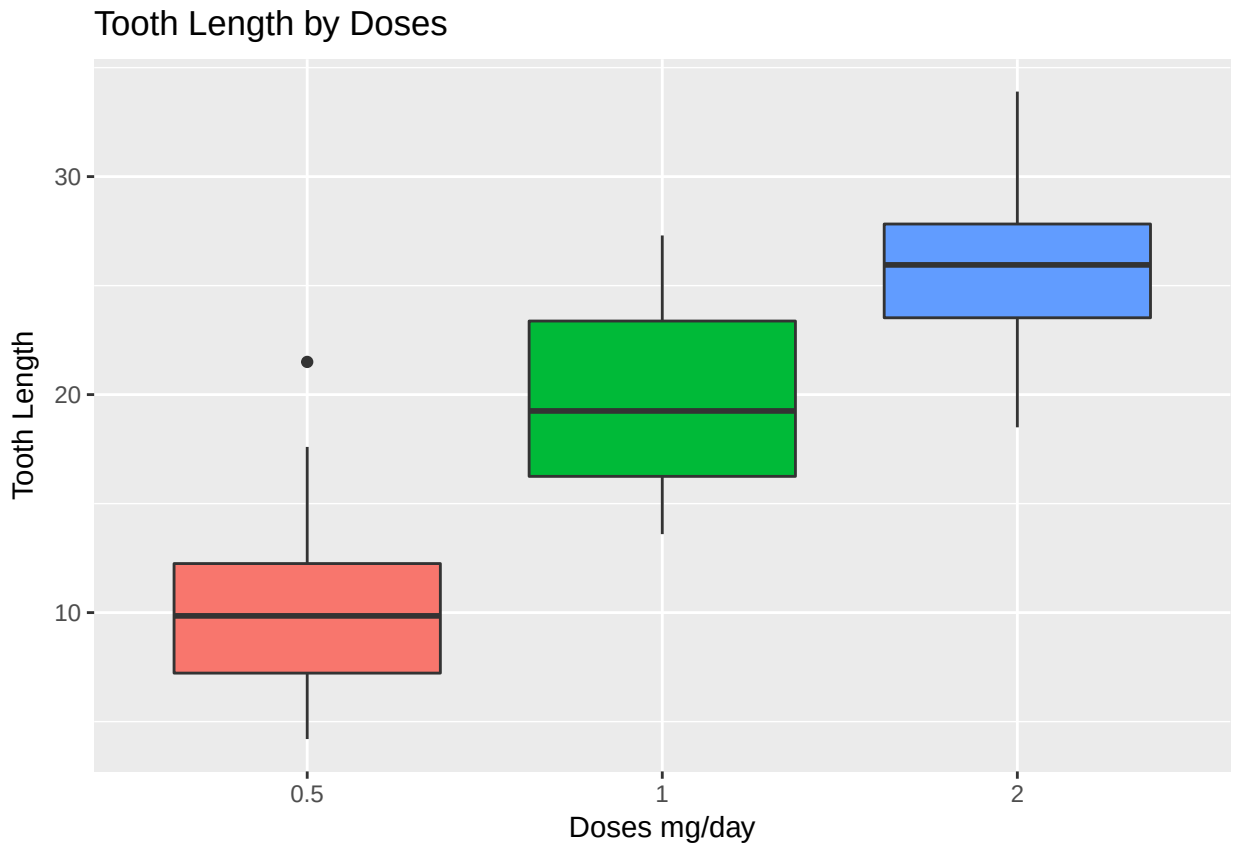
First, let's look how many doses we have in our data set.

```
unique(ToothGrowth$dose);
```

```
## [1] 0.5 1   2
## Levels: 0.5 1 2
```

Now, I already know that we have 3 different doses I will plot a boxplot by doses.

```
plot <- ggplot(aes(x=dose, y=len), data = ToothGrowth);
plot <- plot + geom_boxplot(aes(fill=dose)) +
  xlab("Doses mg/day") +
  ylab("Tooth Length") +
  ggtitle("Tooth Length by Doses") +
  # Removes the legends
  theme(legend.position="none");
plot;
```



In the graphic we can see that there are differences between the datasets. So, let's do the t test over the two firsts doses (0.5 and 1.0).

```
t.test(len~dose, data = subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,1.0)));
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
```

```
## 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 in group 0.5   mean in group 1
##            10.605            19.735
```

The p-value in this test is almost 0, that value tell us that those data set (0.5 and 1.0) are different.

Let's look the another ones, in this case applying the t test over the doses 1.0 and 2.0.

```
t.test(len~dose, data = subset(ToothGrowth, ToothGrowth$dose %in% c(1.0,2.0)));
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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 in group 1 mean in group 2
##          19.735          26.100
```

Again as we saw in the last test, the p-value is almost 0. This test also show that both data sets are different.

Now, let's test the last data sets, we can deduce that we will have the same result.

```
t.test(len~dose, data = subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,2.0)));
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## 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 in group 0.5   mean in group 2
##            10.605            26.100
```

As before, the p-value is almost 0. So, this test also indicates that both data sets are different.

Conclusion

Analyzing, and comparing the ToothGrowth data sets we can conclude that:

- There are not relations between tooth length and the delivery method.
- Exist relations between the tooth growth and the dose given to the pigs.
- With the plot, we can discover that if we raise the amount of dose, the tooth growth will increase.