Infrential Data Analysis

Utkarsh Chauhan

8/28/2020

Overview

In this analysis, we will analyze the ToothGrowth data set by comparing tooth length by supplement and dose. We will compare confidence intervals in order to make conclusions about the tooth growth.

Loading and Exploring Data

```
library(ggplot2)
data(ToothGrowth)
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 ...
```

```
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
```

```
g <- ggplot(data = ToothGrowth, aes(x=factor(dose),y=len))
label <- c("Orange Juice", "Vitamin C")
names(label) <- c("OJ", "VC")
g + geom_boxplot(aes(fill=supp),show.legend=FALSE) +
    facet_grid(.~supp, labeller = labeller(supp=label)) +
    labs(title = 'Tooth length by dose', x='Dose (mg/day)', y='Tooth length')
```


Tooth length by dose
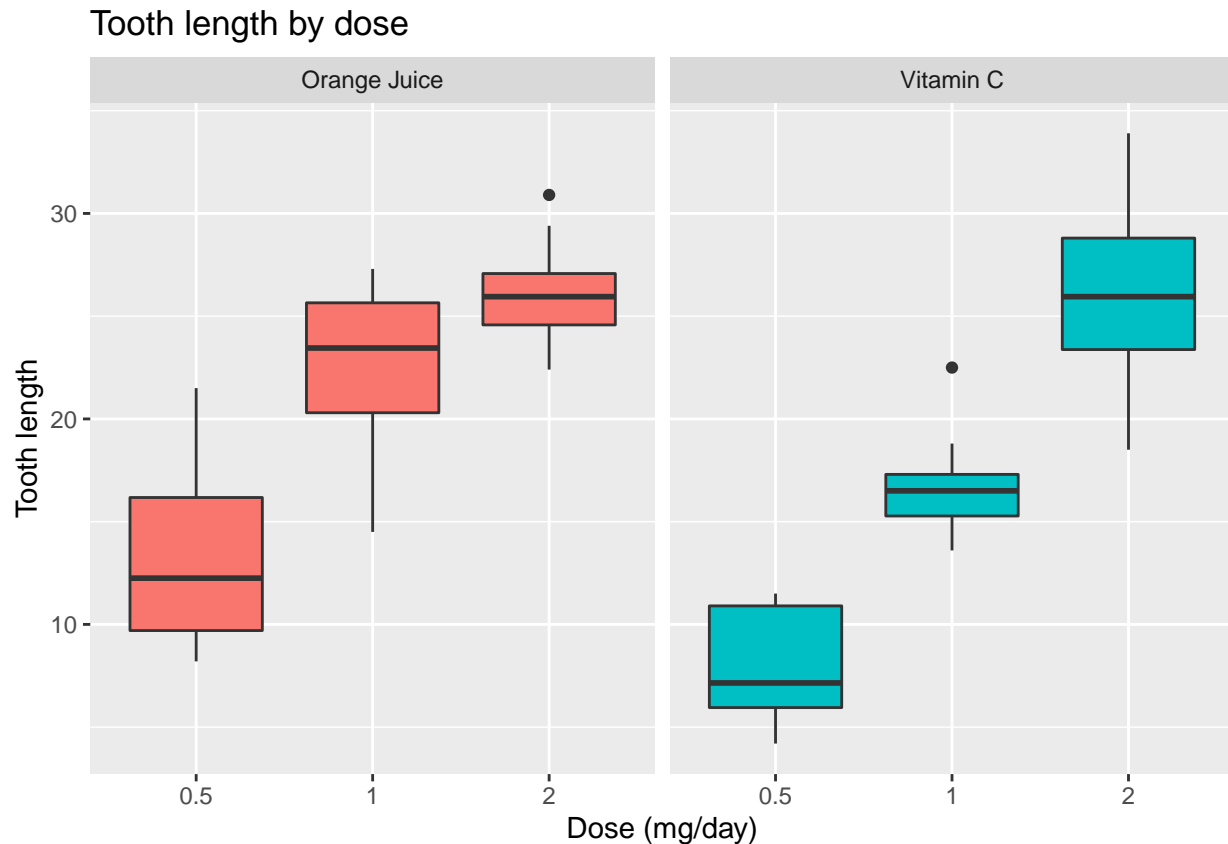
It can be seen that as dose increases, tooth growth also increases for each supplement. Also, orange juice seems to provide better results for less quantity of dose.

Hypothesis Testing

Lets look at tooth growth by supplement.

H0: Tooth growth not correlated with supplements
Ha: Tooth growth correlated with supplements

```
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
```

2

```
## 95 percent confidence interval:
##   -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##          20.66333          16.96333
```

Here the confidence interval includes 0 and the p-value is greater than threshold of 0.05 (alpha). Therefore the null hypothesis can not be rejected.

Now we will compare growth for each dose.

- For does=0.5

```
dose1 <- subset(ToothGrowth,dose==.5)
t.test(len~supp, data = dose1)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

The confidence interval does not contain 0 and p-value is much below 0.05 showing orange juice *provides better growth* than vitamin-C with a dose of 0.5 mg/day.

- For does=1

```
dose2 <- subset(ToothGrowth,dose==1)
t.test(len~supp, data = dose2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

The confidence interval does not contain 0 and p-value is below 0.05 showing orange juice *provides better growth* than vitamin-C with 1 mg/day dose.

- For does=2

```
dose3 <- subset(ToothGrowth,dose==2)
t.test(len~supp, data = dose3)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

The confidence interval contains 0 and p-value is above 0.05 showing orange juice and vitamin-C *provide same growth* at dose of 2 mg/day.

Conclusion and Assumptions

We can see that Orange juice shows more growth than Vitamin-C for small doses(0.5 and 1 mg/day) whereas there is no such trend for higher dose(2 mg/day). Overall we conclude that tooth growth depends strongly on dose but not on supplement.

Assumptions :

- sample is a good approximation of the population.

- distribution of the means is normal and follows CLT.

- no other unmeasured factors affect tooth growth.