Statistical Inference Project - Part 2

Tzachi Ezra Torf-Fulton

July 23, 2015

Overview

In this project we will analyze the ToothGrowth data in the R datasets package. We will compare the impact on the tooth length of guinea pigs by using 2 different supplement types, orange juice (OJ) or ascorbic acid (VC), at each of three dose levels of Vitamin C (0.5, 1, and 2 mg)

Basic summary of the ToothGrowth data

Loading and preprocessing the data

```
# Load ToothGrowth data
library(datasets)
library(ggplot2)
data(ToothGrowth)

# Change the dose attribute to be a factor. Thiw will help with the graphs
# coloring.
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
```

Data Summary

The data consists of 6 sets of 10 guinea pigs teeth samples. Each set relates to on of three dose levels of Vitamin C (0.5, 1, and 2 mg), and one of two delivery methods (orange juice or ascorbic acid). Here are the raw statistics:

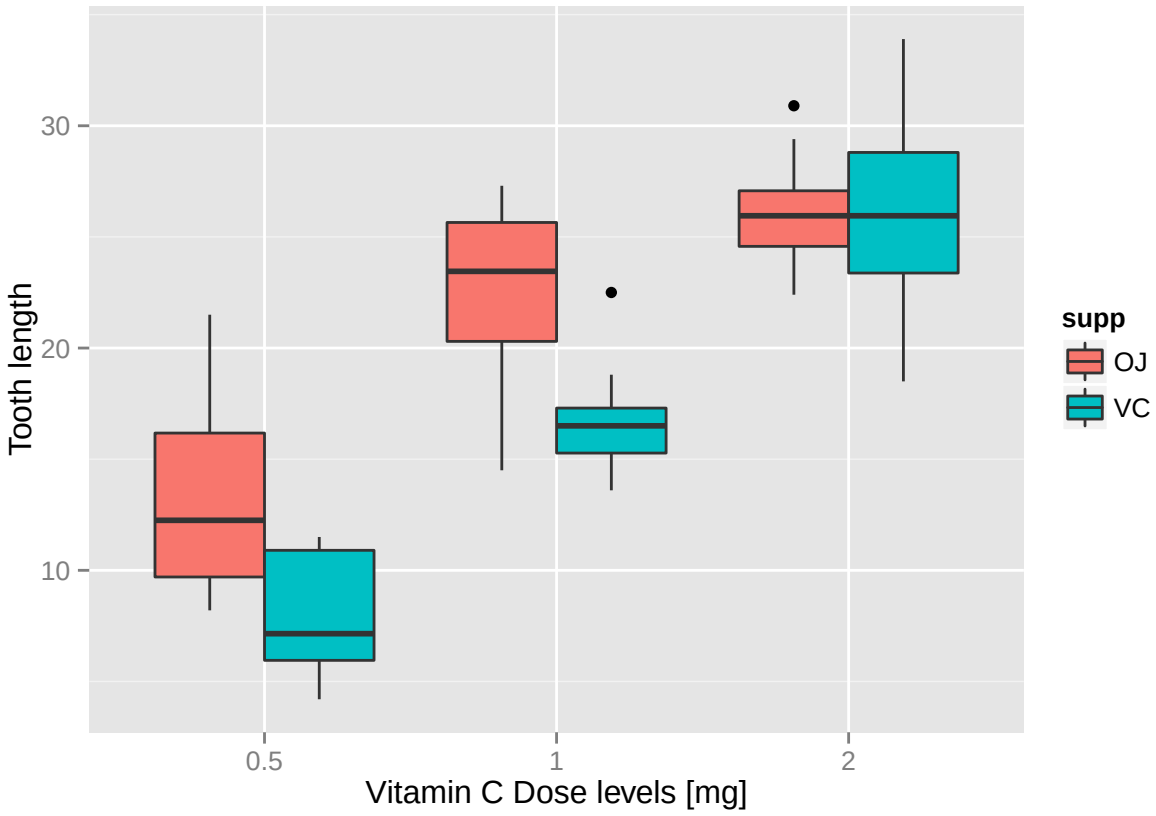
```
summary(ToothGrowth)
```

```
##       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
```

Box plot

A good way to visualize the data is by using box plots:

```
p <- ggplot(ToothGrowth, aes(factor(supp), x = dose, y=len, fill=supp))
p + geom_boxplot() + labs(x="Vitamin C Dose levels [mg]", y="Tooth length")
```



Compare tooth growth by supplement and dose

To decide if OJ and VC have the same effect on teeth length, we will coduct 3 tests comparing the different supplement for each dosage. Our null hypothesis (H0) is that OJ and VC has the same effect, and therefore the mean of the difference between the samples be around 0. We will check this hypothesis against each one of the 3 tests.

Since the ToothGrowth doesn't state it, we will assume that there is no pairing betweem the sample groups. We also don't assume equal variance.

```
# Split data by dosage and supplement
tables <- split(ToothGrowth,list(ToothGrowth$dose,ToothGrowth$supp))
```

0.5[mg] dosage

```
t.test(tables$'0.5.VC'$len - tables$'0.5.OJ'$len, paired=F, var.equal=F)
```

```
##
##  One Sample t-test
##
## data:  tables$"0.5.VC"$len - tables$"0.5.OJ"$len
## t = -2.9791, df = 9, p-value = 0.01547
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -9.236542 -1.263458
```

```
## sample estimates:
## mean of x
##     -5.25
```

1[mg] dosage

```
t.test(tables$'1.VC'$len - tables$'1.OJ'$len, paired=F, var.equal=F)
```

```
##
##  One Sample t-test
##
## data:  tables$"1.VC"$len - tables$"1.OJ"$len
## t = -3.3721, df = 9, p-value = 0.008229
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -9.908089 -1.951911
## sample estimates:
## mean of x
##     -5.93
```

2[mg] dosage

```
t.test(tables$'2.VC'$len - tables$'2.OJ'$len, paired=F, var.equal=F)
```

```
##
##  One Sample t-test
##
## data:  tables$"2.VC"$len - tables$"2.OJ"$len
## t = 0.0426, df = 9, p-value = 0.967
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -4.168976  4.328976
## sample estimates:
## mean of x
##      0.08
```

Conclusion

0.5[mg] and 1[mg]

It is possible to see that we get very small p-values, and the 95% confidence interval does not include zero, meaning that we can reject H0, and cannot assume the same effect using OJ and VC. Also, the fact that we get negative mean suggests that OJ has a better effect on teeth growth.

2[mg]

We can see a high p-value, a 95% confidence interval that contain zero, and a mean which is close to 0, therefore we cannnot reject H0 and can assume that for that dosage the effect of using OJ or VC is similar.