Statistical Inference Course Project - Part 2

Amber Beasock

27 December 2015

Introduction

In the second part of the project, we analyze the `ToothGrowth` data in the R datasets package.

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

Load the data & some basic exploratory data analyses

```
# Visualizations will later be created using ggplot2
library(ggplot2)

# Load R datasets packages
library(datasets)

# Documentation on the Tooth Growth dataset
??ToothGrowth

# Visualizing the first few observations of the ToothGrowth dataset
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
```

The documentation on this R dataset provides us with a description of the data and some information on the variables.

The ToothGrowth data set contains 60 observations of the lenth of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

ToothGrowth Variables:

- **len** (numeric): Tooth Length
- **supp** (factor): Supplement type (VC or OJ)
- **dose** (numeric): Dose in miligrams

Summarize 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
```


Compare tooth growth by supp and dose

Determine if tooth length depends on the Vitamin C delivery method (OJ vs VC)

```
# The average length by supp
supp <- split(ToothGrowth$len, ToothGrowth$supp)
sapply(supp, mean)
```

```
##       OJ       VC
## 20.66333 16.96333
```

```
# Plot the average tooth length by supplement type
ggplot(aes(x=supp, y=len),data=ToothGrowth) +
    geom_boxplot(aes(fill=supp)) +
    xlab("Supplement Type") +
    ylab("Tooth Length")
```
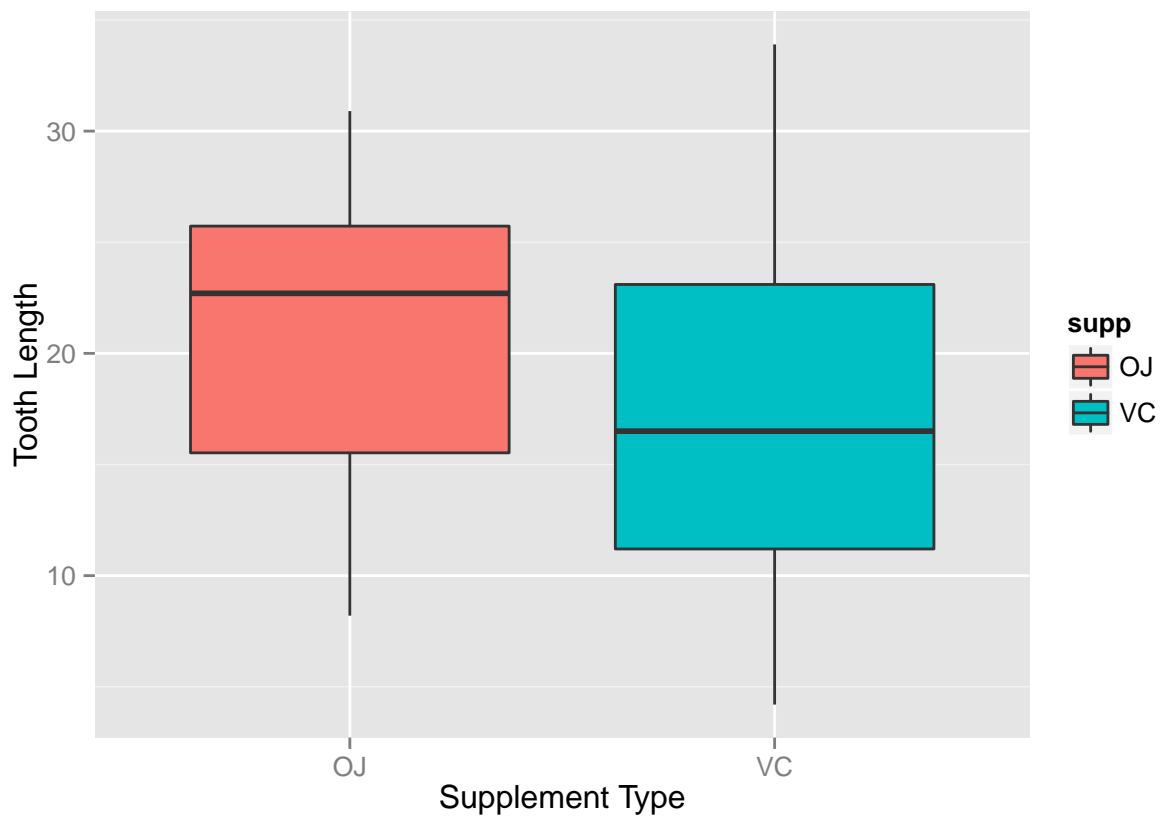
```
# Variance between supplement types
sapply(supp, var)
```

```
##        OJ        VC
## 43.63344 68.32723
```

T-test for supplement type

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

Determine if tooth length depends on dosage

```r
# The average length by dose
dose <- split(ToothGrowth$len, ToothGrowth$dose)
sapply(dose, mean)
```

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

```r
# Convert dose from numeric to factor
ToothGrowth$dose <- as.factor(ToothGrowth$dose)

# Plot the average tooth length by dose amount
ggplot(aes(x=dose, y=len),data=ToothGrowth) +
    geom_boxplot(aes(fill=dose)) +
    xlab("Dose in miligrams") +
    ylab("Tooth Length")
```
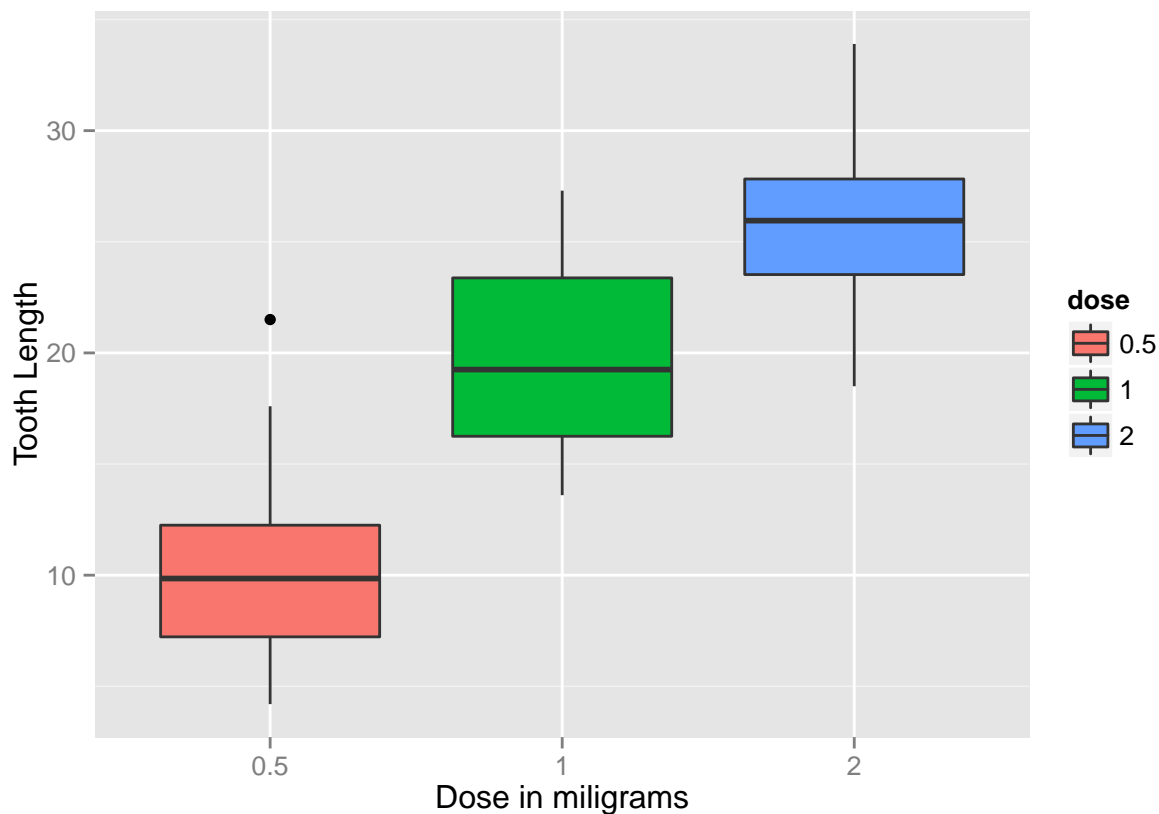

T-test for dosage of 0.5 & 2 mg

```r
t.test(ToothGrowth$len[ToothGrowth$dose==2], ToothGrowth$len[ToothGrowth$dose==0.5], paired = FALSE, va
```

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 2] and ToothGrowth$len[ToothGrowth$dose == 0.5]
```

```
## t = 11.799, df = 38, p-value = 2.838e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  12.83648 18.15352
## sample estimates:
## mean of x mean of y
##     26.100    10.605
```

Look at the tooth length by dosage per supplement type

```
# Plot the average tooth length by dose per supplemental type
ggplot(aes(x=dose, y=len),data=ToothGrowth) +
    geom_boxplot(aes(fill=supp)) +
    facet_grid(. ~ supp) +
    xlab("Dose in miligrams") +
    ylab("Tooth Length")
```
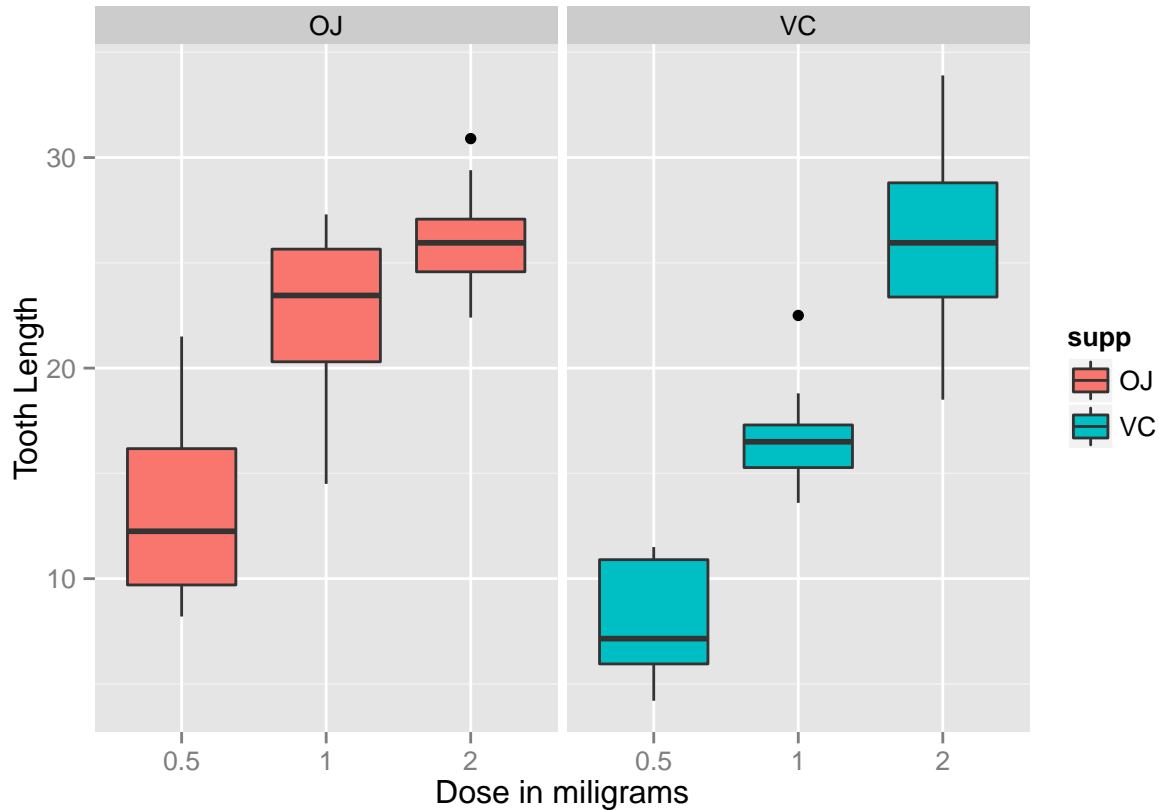

Conclusions

The first boxplot shows the impact on tooth length by supplement type, and shows OJ has more of a positive impact than VC. We can look at this further by performing a t-test. From the first t-test analysis above, the p-value is `0.06063`. This is close to the significance level of 5%, but not enough for us to reject the null

hyptothesis. Therefore, we conclude that there is no significant difference between the impact of OJ and VC on tooth growth.

The second boxplot shows that increasing dosage has a positive impact on tooth length. A t-test is performed to look at if increasing dosage from 0.5 to 2 mg has a impact on tooth growth. The p-value is 0. We can say increasing dosage has a positive effect on tooth growth.

The final boxplot shows that increase in dosage for both delivery methods has a postive correlation with tooth growth.

In conclusion, the delivery method seems to have no effect on tooth growth. However, increasing the dosage in either delivery method has a positive effect on tooth growth.