Statistical Inference - Course Project Part 2

Luke Moss

Sunday, April 26, 2015

Overview

This project will analyze the ToothGrowth data in the R datasets package and address the following:

1. Load the ToothGrowth da2. ta and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose. (Use the techniques from class even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

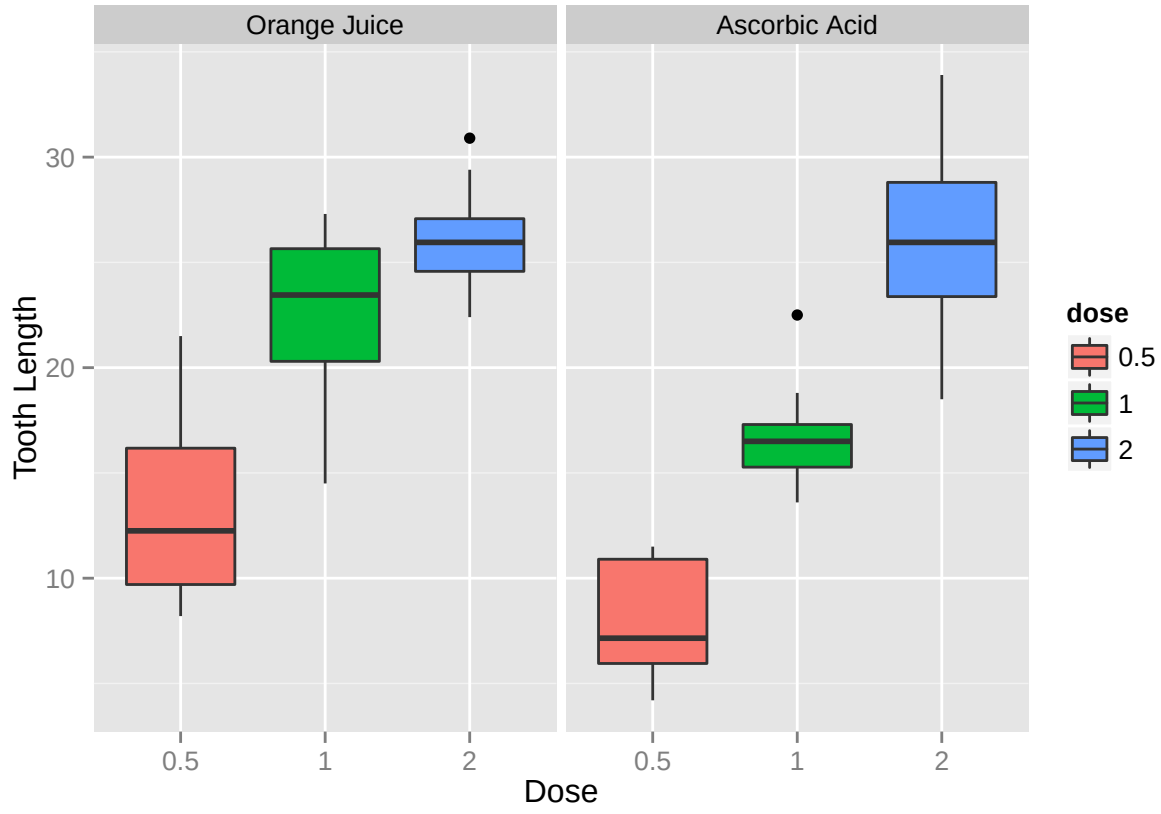
Load the ToothGrowth da2. ta and perform some basic exploratory data analyses

The ToothGrowth data is built into R and can be loaded by executing . The documentation about this data set can be viewed by looking at the help page (C:/Program Files/R/R-3.1.2/library/datasets/help/ToothGrowth).

From reading the documentation, this data represents the affect of Vitamin C on tooth growth in guinea pigs. The data consists of 60 observations (6 groups of 10 animals), divided into two groups based on the delivery method of the Vitamin C (orange juice or ascorbic acid) at each of 3 dosage levels of Vitamin C (0.5, 1 or 2 mg).

To form a quick exploratory analysis, first transform the data and then generate a boxplot to visualize the data and it's relationships.

```
data <- ToothGrowth
data$dose <- factor(data$dose)
levels(data$supp) <- c("Orange Juice", "Ascorbic Acid")
ggplot(data, aes(x = dose, y = len)) +
    geom_boxplot(aes(fill = dose)) +
    facet_wrap(~ supp) +
    ylab("Tooth Length") +
    xlab("Dose")
```



This quick analysis illustrates that:

1. Vitamin C dosage does seem to have an affect on tooth length.
2. Delivery method does seem to have an affect on tooth length. Around 2 mg, Ascorbic Acid seems to have more of an affect than Vitamin C.

Provide a basic summary of the data.

Below is a quick summary of the data:

```
summary(data)
```

```
##       len                  supp     dose
##  Min.   : 4.20   Orange Juice :30   0.5:20
##  1st Qu.:13.07   Ascorbic Acid:30   1  :20
##  Median :19.25                      2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose.

Hypothesis 1 - Mean tooth length is the same for different dosages

The first hypothesis is that the mean tooth length is the same for two different dosages. I will use a Welch 2 sample t-test to do this analysis with data from both delivery methods.

0.5 mg and 1 mg dosages:

```
data_05_1 <- subset(data, dose %in% c(0.5,1))
t.test(len ~ dose, paired=F, var.equal=F, data=data_05_1)
```

```
##
##  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 result is negative confidence intervals and a very small p-value. This indicates that the means are not equal and our hypothesis must be rejected. To ensure this is correct, do a second comparison for more conclusive evidence.

1 mg and 2 mg dosages:

```
data_1_2 <- subset(data, dose %in% c(1,2))
t.test(len ~ dose, paired=F, var.equal=F, data=data_1_2)
```

```
##
##  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
```

This further solidifies that the hypothesis has to be rejected. It also proves our observation about Vitamin C having an affect on tooth length (the higher the dosage, the longer the tooth length).

Hypothesis 2 - Abscorbic acid improves growth more than orange juice

The second hypothesis is that abscorbic acid improves tooth growth more than orange juice.

```r
data <- ToothGrowth
t.test(data$len[data$supp == "VC"], data$len[data$supp == "OJ"], paired = TRUE)
```

```
##
##  Paired t-test
##
## data:  data$len[data$supp == "VC"] and data$len[data$supp == "OJ"]
## t = -3.3026, df = 29, p-value = 0.00255
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -5.991341 -1.408659
## sample estimates:
## mean of the differences
##                    -3.7
```

The result is a low p value which would indicate that the hypothesis has to be rejected.

Conclusions

1. Vitamin C dosage does have an affect on tooth length. The higher the dosage, the longer the tooth.
2. Overall, the delivery method does not have an influence when looking at the entire dataset including all dosages. From looking at the summary boxplot, it would seem to indicate that delivery method makes a measurable difference at the 1 mg dosage, but the overall dataset indicates it doesn't make a difference.