Statistical Inference Project - Part 2

Oye Akinnikawe

January 2, 2017

Synopsis

This is a project for the Cousera Statistical Inference Class. The project consists of two parts:

1. Simulation exercise to explore inference
2. Basic inferential data analysis using the ToothGrowth data in the R datasets package

Part 2 - Inferential Data Analysis

Overview

The goal of this report is to analyze the ToothGrowth data in the R dataset package. The following tasks should be completed

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.

Basic Summary of Data

first we will load the necessary libraries for the project.

```
library(ggplot2)
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 3.2.5
```

Loading the ToothGrwoth data and looking at the structure of the data.

```
data("ToothGrowth")

## display data structure
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 ...
```

Provide a quick 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
```

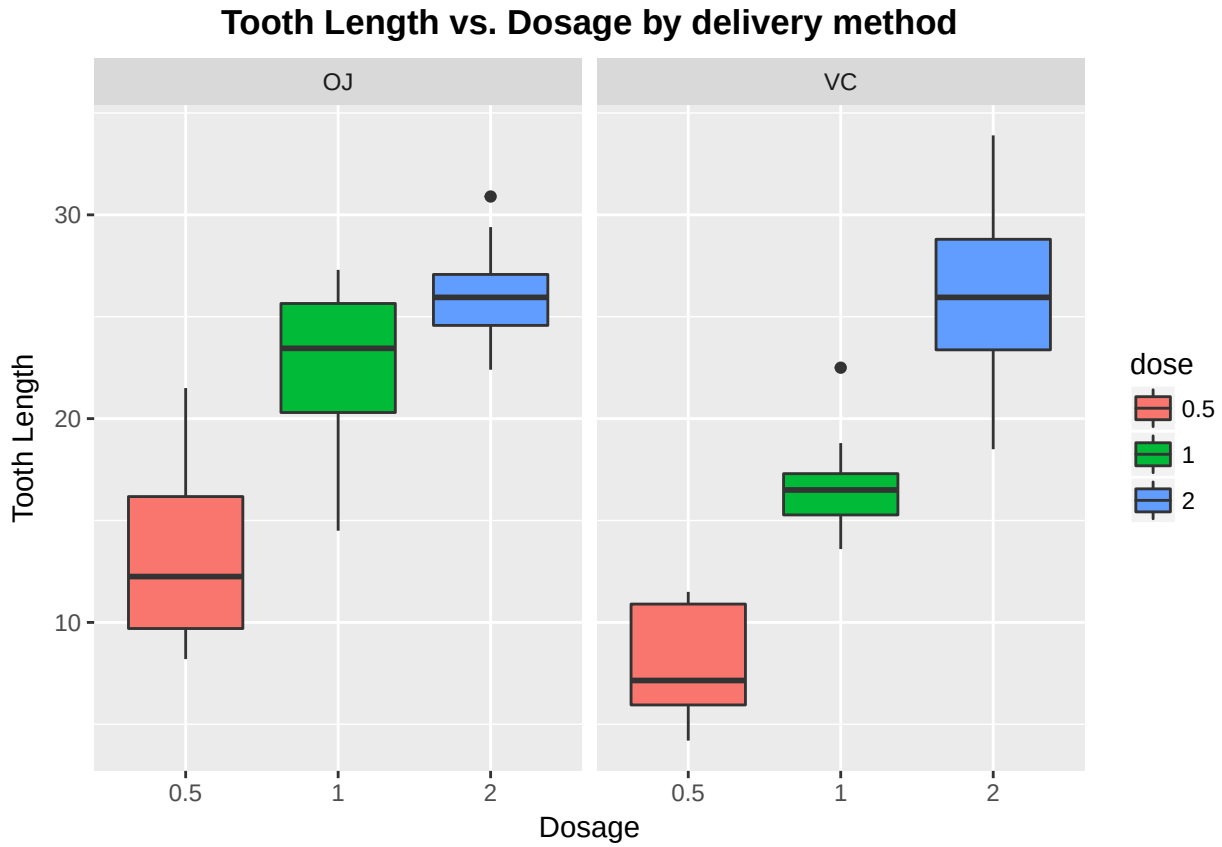
```
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

Exploratory Data Analysis

In this section we will create some plots to explore the data.

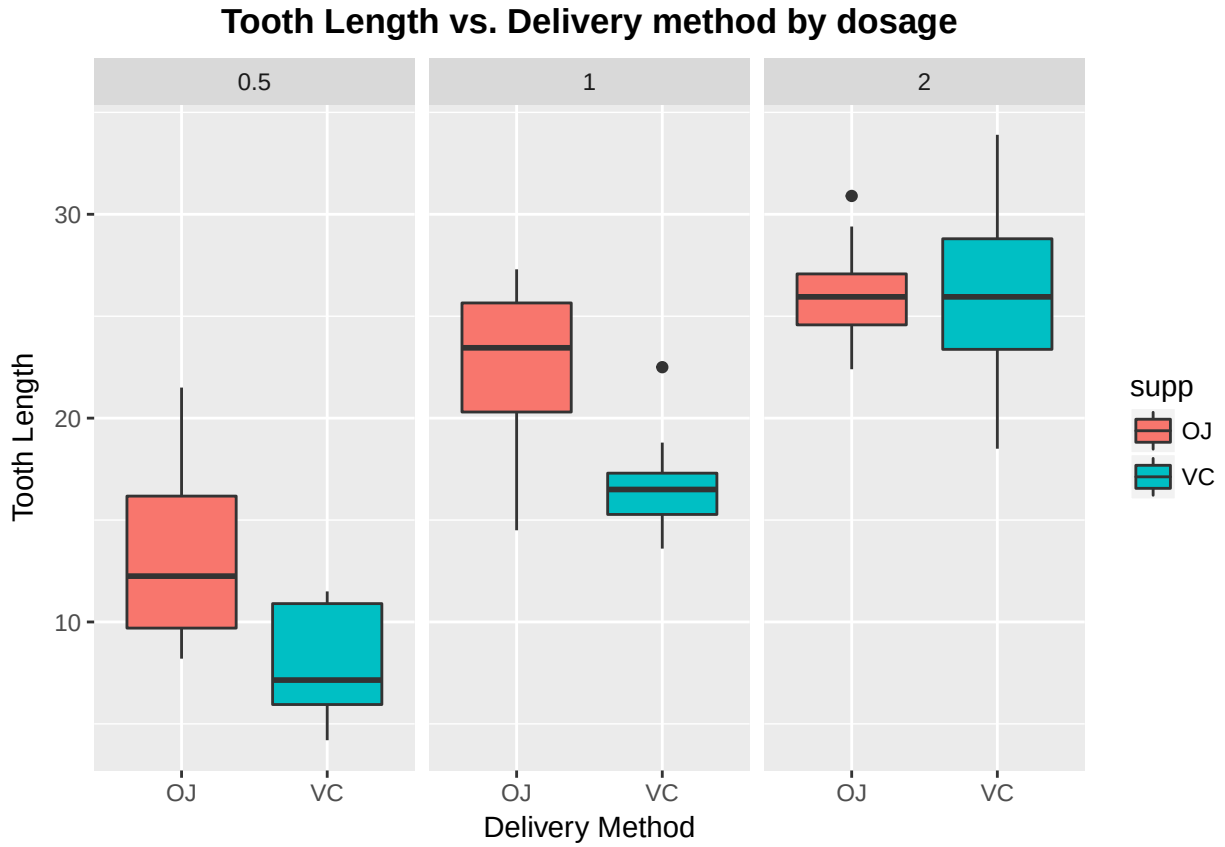
Plot tooth length, *len* vs. dose amount, *dose* broken out by supplement delivery method, *supp*

```
# convert dose to a factor
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
g <- ggplot(aes(x=dose, y=len), data = ToothGrowth)
g <- g + geom_boxplot(aes(fill=dose)) + xlab("Dosage") + ylab("Tooth Length")
g <- g + facet_grid(~ supp) + ggtitle("Tooth Length vs. Dosage by delivery method")
g <- g + theme(plot.title = element_text(hjust = 0.5, face = "bold"))
g
```



Plot tooth length vs. supplement delivery amount broken out by the dose amount.

```
g <- ggplot(aes(x=supp, y=len), data = ToothGrowth)
g <- g + geom_boxplot(aes(fill=supp)) + xlab("Delivery Method") + ylab("Tooth Length")
g <- g + facet_grid(~ dose) + ggtitle("Tooth Length vs. Delivery method by dosage")
g <- g + theme(plot.title = element_text(hjust = 0.5, face = "bold"))
g
```



From the plots we can see that the higher the dose, the longer the Tooth Length. For both delivery method. It seems that a given dose via the orange juice (OJ) delivery method is more efficient compared to the via vitamin C (VC). However, they both seem to have the same median for a dose of 2.

Using confidence intervals to compare tooth growth by supp and dose

```
# Convert ToothGrowth into a molten data frame
longTooth <- melt(ToothGrowth, measure.vars = c("supp", "dose"))
```

```
## Warning: attributes are not identical across measure variables; they will
## be dropped
```

```
head(longTooth)
```

```
##    len variable value
## 1  4.2     supp    VC
## 2 11.5     supp    VC
## 3  7.3     supp    VC
## 4  5.8     supp    VC
## 5  6.4     supp    VC
## 6 10.0     supp    VC
```

Perform t test on supplement

```
g1 <- longTooth$len[1:30] ; g2 <- longTooth$len[31:60]
t.test(g2, g1, paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  g2 and g1
## 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
```

The p-value is 0.06, and the confidence interval contains zero. This indicates that we cannot reject the null hypothesis that the supplement delivery methods have no effect on tooth length.

Now we will compare tooth growth by dose, looking at the different pairs of dose values.

first, we run t-test using dose amounts of 0.5 and 1

```
t.test(longTooth$len[longTooth$value==1], longTooth$len[longTooth$value==0.5], paired = FALSE, var.equal
```

```
##
##  Two Sample t-test
##
## data:  longTooth$len[longTooth$value == 1] and longTooth$len[longTooth$value == 0.5]
## t = 6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.276252 11.983748
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

Next, we run t-test using dose amounts of 1 and 2

```
t.test(longTooth$len[longTooth$value==2], longTooth$len[longTooth$value==1], paired = FALSE, var.equal =
```

```
##
##  Two Sample t-test
##
## data:  longTooth$len[longTooth$value == 2] and longTooth$len[longTooth$value == 1]
## t = 4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.735613 8.994387
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

The p-values of both tests were very low (close to 0) and the confidence intervals did not contain zero. Therefore, we will reject the null hypothesis and accept the alternative hypothesis. This means that there is an increase in tooth length with increased dose amount.

Conclusions

From the t-test analysis, we can conclude that using a 95% confidence interval the supplement delivery method has no effect on tooth length. However, increasing the dosage results in increased tooth length.