Part 2: Basic Inferential Data Analysis Instructions

TEJASWI N RAO

26TH JULY 2018

Overview

In this part of the project we are going to do some basic inferential data analysis. In order to do that, we are going to analyze the ToothGrowth data in the R datasets package.

Load ToothGrow data

Load the ToothGrowth data and perform a quick summary to explore it.

```
library(ggplot2)
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
# Load ToothGrowth data
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
```

```
#Unique Values
unique(ToothGrowth$len)
```

```
##  [1]  4.2 11.5  7.3  5.8  6.4 10.0 11.2  5.2  7.0 16.5 15.2 17.3 22.5 13
.6
## [15] 14.5 18.8 15.5 23.6 18.5 33.9 25.5 26.4 32.5 26.7 21.5 23.3 29.5 17
.6
```

```
## [29]  9.7  8.2  9.4 19.7 20.0 25.2 25.8 21.2 27.3 22.4 24.5 24.8 30.9 29
.4

## [43] 23.0
```

```
unique(ToothGrowth$supp)
```

```
## [1] VC OJ

## Levels: OJ VC
```

```
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

Provide a basic summary of the data

Let's go to represent graphicaly the Tooth Growing versus Supplement Delivery Method,
taking into account the 3 different Dose Amount types

```
ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp))
+ xlab("Supplement Delivery") + ylab("Tooth Length") + facet_grid(~ dose) +
ggtitle("Tooth Length vs. Delivery Method \nby Dose Amount") +

    theme(plot.title = element_text(lineheight=.8, face="bold"))
```

Tooth Length vs. Delivery Method
by Dose Amount
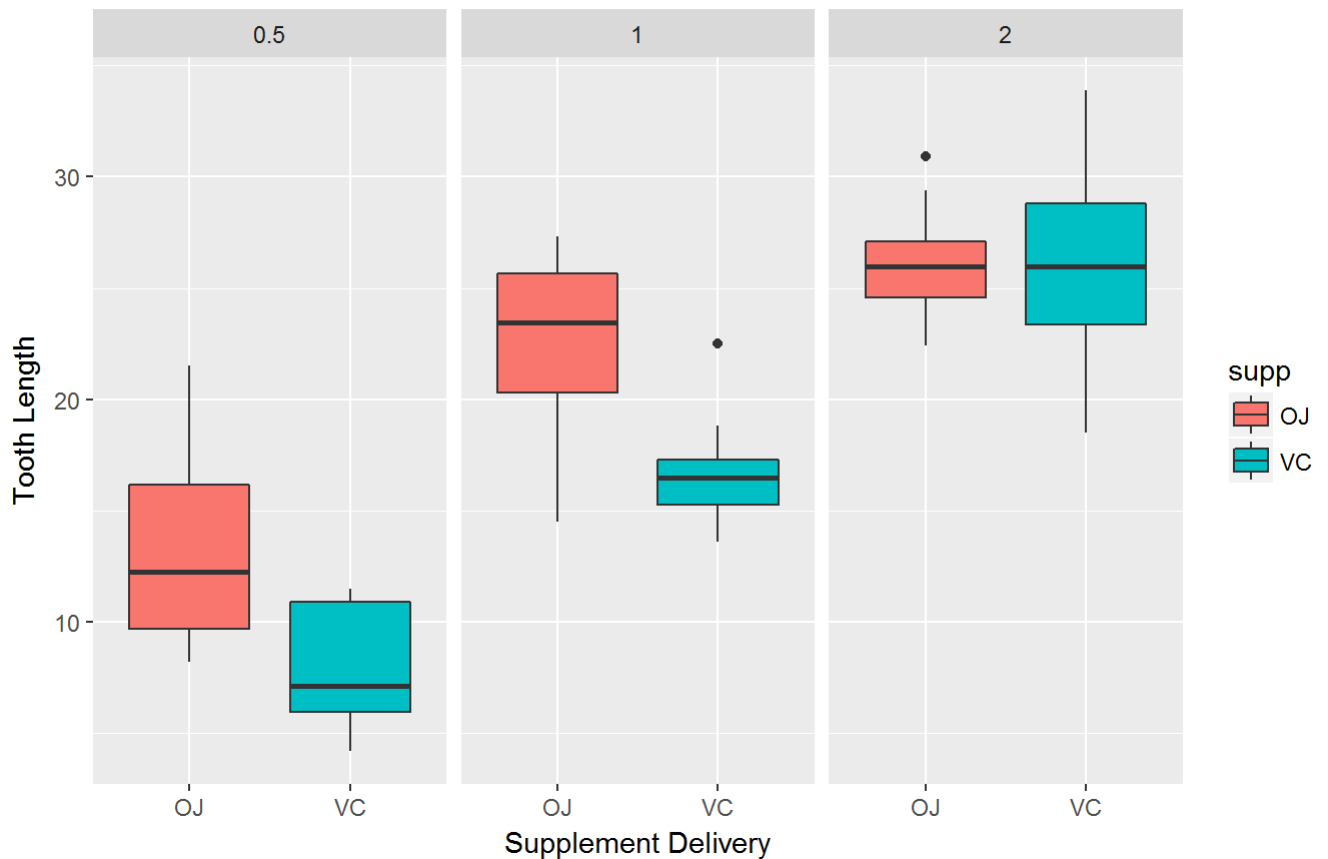
Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose

Now we are going to test if there is any correlation between Tooth Grow and Supplement

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

P-value is 0.06 (>0.05)
Confidence Interval is -0.17 to 7.57 (contains 0)
We can say that supplement types seems to have no impact on Tooth growth based on this test

Let's try another test, now comparing Tooth Grow with Dose Amount looking at the different pairs of dose values

```
# Dose amounts 0.5 and 1.0
t.test(len~dose,data=subset(ToothGrowth, ToothGrowth$dose %in% c(1.0,0.5)))
```

```
##
##   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
```

```
# Dose amounts 0.5 and 2.0
t.test(len~dose,data=subset(ToothGrowth, ToothGrowth$dose %in% c(2.0,0.5)))
```

```
##
##   Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   -18.15617 -12.83383
## sample estimates:
## mean in group 0.5    mean in group 2
```

```
##               10.605                 26.100
```

```
# Dose amounts 1.0 and 2.0
t.test(len~dose,data=subset(ToothGrowth, ToothGrowth$dose %in% c(2.0,1.0)))
```

```
##
##   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
```

P-values are all close to 0 ($<<0.05$)
Confidence Interval of each test doesn't cross over zero (0).
We can say that supplement types seems to have no impact on Tooth growth based on this test.
Based on this results, we can assume that the average tooth length increases with an increasing dose

Conclusions

Given the following assumptions:

- The sample is representative of the population

- The distribution of the sample means follows the Central Limit Theorem

We can conclude that:

- Supplement Delivery Method has no effect on tooth growth/length

- Increased Dose do result in increased tooth length