**Statistical Inference Final Project**

**PART2: Basic Inferential Data Analysis**

TASK:

This report aims to analyze the Tooth Growth data in the R datasets package. Per the course project instructions, the following items should occur:

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 thereâs other approaches worth considering).
4. State your conclusions and the assumptions needed for your conclusions.

ANALYSIS:

First we will load the necessary libraries.

**library**(ggplot2)

Now we will load the data and provide a quick summary.

*# Load ToothGrowth data*

data("ToothGrowth")

*# Display a 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

*#Display the first few rows of data*

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

Next we will create some plots to explore the data.

*# Convert dose to a factor*

ToothGrowth$dose<-as.factor(ToothGrowth$dose)

*# Plot tooth length ('len') vs. the dose amount ('dose'), broken out by supplement delivery method ('supp')*

ggplot(aes(x=dose, y=len), data=ToothGrowth) + geom\_boxplot(aes(fill=dose)) + xlab("Dose Amount") + ylab("Tooth Length") + facet\_grid(~ supp) + ggtitle("Tooth Length vs. Dose Amount \nby Delivery Method") +

theme(plot.title = element\_text(lineheight=.8, face="bold"))

*# Plot tooth length ('len') vs. supplement delivery method ('supp') broken out by the dose amount ('dose')*

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"))

Now we will compare tooth growth by supplement using a t-test.

*# run t-test*

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

The p-value of this test was 0.06.  
Since the p-value is greater than 0.05 and the confidence interval of the test contains zero we can say that supplement types seems to have no impact on Tooth growth based on this test.

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

*# run t-test using dose amounts 0.5 and 1.0*

ToothGrowth\_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(1.0,0.5))

t.test(len~dose,data=ToothGrowth\_sub)

## 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

*# run t-test using dose amounts 0.5 and 2.0*

ToothGrowth\_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,2.0))

t.test(len~dose,data=ToothGrowth\_sub)

##

## 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

*# run t-test using dose amounts 1.0 and 2.0*

ToothGrowth\_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(1.0,2.0))

t.test(len~dose,data=ToothGrowth\_sub)

##

## 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

As can be seen, the p-value of each test was essentially zero and the confidence interval of each test does not cross over zero (0).

Based on this result we can assume that the average tooth length increases with an increasing dose, and therefore the null hypothesis can be rejected.

ASSUMPTIONS AND CONCLUSION:

Given the following assumptions:

1. The sample is representative of the population
2. The distribution of the sample means follows the Central Limit Theorem

**In reviewing our t-test analysis from above, we can conclude that supplement delivery method has no effect on tooth growth/length, however increased dosages do result in increased tooth length**.