

# Statistical Inference Project, Part 2

*Tiffany Stoeke*

*September 19, 2015*

## Summary:

In the second half of the course project, we evaluate the data found in the Tooth Growth R data file and perform some basic exploratory data analysis to compare tooth growth by supplement and dose. This file describes the effect of Vitamin C on tooth growth in each of 10 guinea pigs at three dosage levels under two delivery methods. Our dosage levels are 0.5, 1, and 2 mg and our delivery methods are orange juice or ascorbic acid.

Please note that due to page restrictions all relevant t tests will be found in the appendix.

## Assumptions:

Our assumptions when working with this data are that the data sets are paired per the information in the help file of the data set (note: only 10 guinea pigs were used, with six different variations of the study). This assumption directly affects our t-test outcomes. We also assume that the underlying study was performed with relevant wash-out periods (if required), across equivalent time intervals for all observations, and that there are no other confounder variables that would affect tooth growth during the time of the study. We also assume that no guinea pigs were harmed in the making of this study!

## Exploratory Data Analysis:

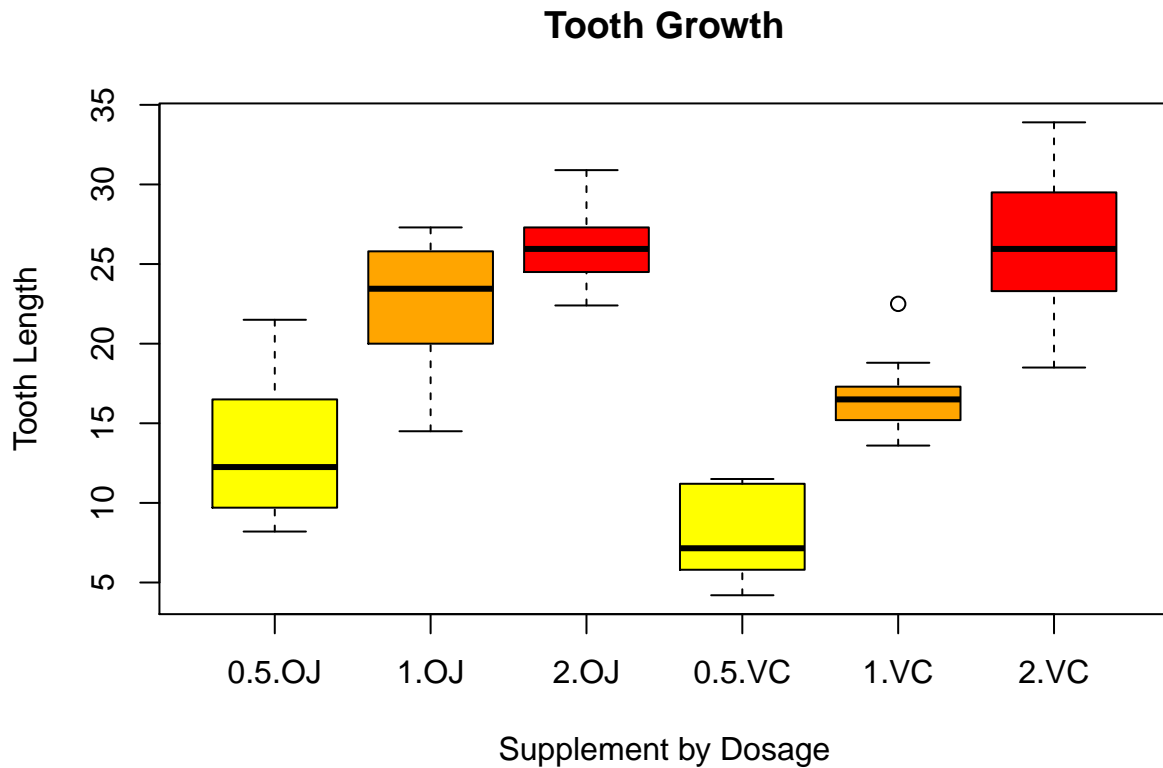
In order to begin our analysis, let's gather some summary information about the data:

```
library(ggplot2)
##Load the ToothGrowth Data and show a basic summary of the data.
data(ToothGrowth)
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
```

The below box plot shows a summary of the tooth growth data, with each delivery method broken out into dosage levels. Our initial review shows that dosage level plays a role in increasing tooth length across both delivery methods.

```
##Exploratory Data Analysis - predictor variables are dosage and supplement,
##outcome variable is length of tooth growth.
boxplot(len~dose+supp,data=ToothGrowth,col=(c("yellow","orange","red")),main="Tooth Growth",
        xlab="Supplement by Dosage",ylab="Tooth Length")
```



### Confidence Intervals, T Test Results and Conclusions:

Our null hypothesis is that the delivery method (orange juice vs. ascorbic acid) does not affect guinea pig tooth growth. The alternative hypothesis is that delivery method does indeed play a role in tooth growth. Please see appendix for the relevant confidence intervals and t test results referenced here.

Due to small sample size, Gosset's t distribution using  $n-1$  degrees of freedom was used. Per our t test of delivery methods when dosage is not taken into account, our t statistic is 3.3 which is higher than the expected qt value of 2.04523 at 29 degrees of freedom. The resulting p-value is also well below 5%, therefore we reject our null hypothesis - delivery method clearly affects tooth growth.

We also ran additional t tests to determine if the delivery method affects tooth growth at all three dosage levels. This will determine whether we reject the null hypothesis at all three dosage levels. Per our results (see appendix), delivery method plays an important role in dosage levels of 0.5 and 1 mg (t tests and p values indicate to reject the null hypothesis for both of the dosages per qt values at 9 degrees of freedom), but cannot be shown to affect tooth growth at the highest dosage level, 2mg. The null hypothesis is not rejected at the 2mg dosage level since the number 0 is included in our 95% confidence interval.

## Appendix

```
##Subset the data in order to perform hypothesis testing by supplement and by dosage
##separately so that relevant t-tests can be performed.
suppVC<-subset(ToothGrowth,supp=="VC")
suppOJ<-subset(ToothGrowth,supp=="OJ")
dose1<-subset(ToothGrowth,dose==1)
dose2<-subset(ToothGrowth, dose==2)
dosehalf<-subset(ToothGrowth,dose==0.5)

##Obtain confidence Intervals on the original data.
fit <- lm(len ~ ., data=ToothGrowth)
summary(fit)
```

```
##
## Call:
## lm(formula = len ~ ., data = ToothGrowth)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -6.600 -3.700  0.373  2.116  8.800
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.2725     1.2824   7.231 1.31e-09 ***
## suppVC        -3.7000     1.0936  -3.383  0.0013 **
## dose           9.7636     0.8768  11.135 6.31e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.236 on 57 degrees of freedom
## Multiple R-squared:  0.7038, Adjusted R-squared:  0.6934
## F-statistic: 67.72 on 2 and 57 DF,  p-value: 8.716e-16
```

```
confint(fit)
```

```
##              2.5 %    97.5 %
## (Intercept)  6.704608 11.840392
## suppVC       -5.889905 -1.510095
## dose         8.007741 11.519402
```

```
##Run t-tests on our null hypothesis.
##Null hypothesis (H_o): Supplement type doesn't matter.
t.test(len ~ supp, data=ToothGrowth, paired = TRUE)
```

```
##
## Paired t-test
##
## data:  len by supp
## t = 3.3026, df = 29, p-value = 0.00255
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
## 1.408659 5.991341
## sample estimates:
## mean of the differences
## 3.7
```

```
qt(.975,29)
```

```
## [1] 2.04523
```

```
##Additional tests to see if dosage level changes the outcome of tooth growth per supplement.
##Our null hypothesis for all three tests is that the dosage level per supplement
##doesn't affect the outcome. We now have 9 degrees of freedom after the subsetting.
qt(.975,9)
```

```
## [1] 2.262157
```

```
##t.test at 0.5 mg dosage level:
t.test(len ~ supp, data=dosehalf, paired = TRUE)
```

```
##
## Paired t-test
##
## data: len by supp
## t = 2.9791, df = 9, p-value = 0.01547
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 1.263458 9.236542
## sample estimates:
## mean of the differences
## 5.25
```

```
##t.test at 1 mg dosage level:
t.test(len ~ supp, data=dose1, paired = TRUE)
```

```
##
## Paired t-test
##
## data: len by supp
## t = 3.3721, df = 9, p-value = 0.008229
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 1.951911 9.908089
## sample estimates:
## mean of the differences
## 5.93
```

```
##t.test at 2 mg dosage level:
t.test(len ~ supp, data=dose2, paired = TRUE)
```

```
##
## Paired t-test
##
## data: len by supp
## t = -0.042592, df = 9, p-value = 0.967
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -4.328976  4.168976
## sample estimates:
## mean of the differences
## -0.08
```