# Impact of Vitamin C Delivery Method on Tooth Growth

*Jim Callahan*

*October 1, 2015*

**Overview**

The data in this project is an R supplied data set, "Tooth Growth". The official description in the R version 3.2.2 help file is:

> The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

**Data**

The data in this project is an R supplied data set, "Tooth Growth" which can be accessed from the "datasets" package using the data command:

```r
# Access the "datasets" package using the R library command
# and then the data set itself using the R data command.
library(datasets)
data(ToothGrowth)

# Exploratory Data Analysis (EDA)
# Describe the structure of the data set, the first few rows
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 ...
```

```r
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
```

```r
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
```

```r
# Note the help file for the R data set can be accessed by typing
# help(ToothGrowth)
# at the R command line.
```

The **R** help file also has sample code to produce a graph.

```r
# Graph Code from R help file
require(graphics)
coplot(len ~ dose | supp, data = ToothGrowth, panel = panel.smooth,
       xlab = "ToothGrowth data: length vs dose, given type of supplement")
```

Given : supp



ToothGrowth data: length vs dose, given type of supplement
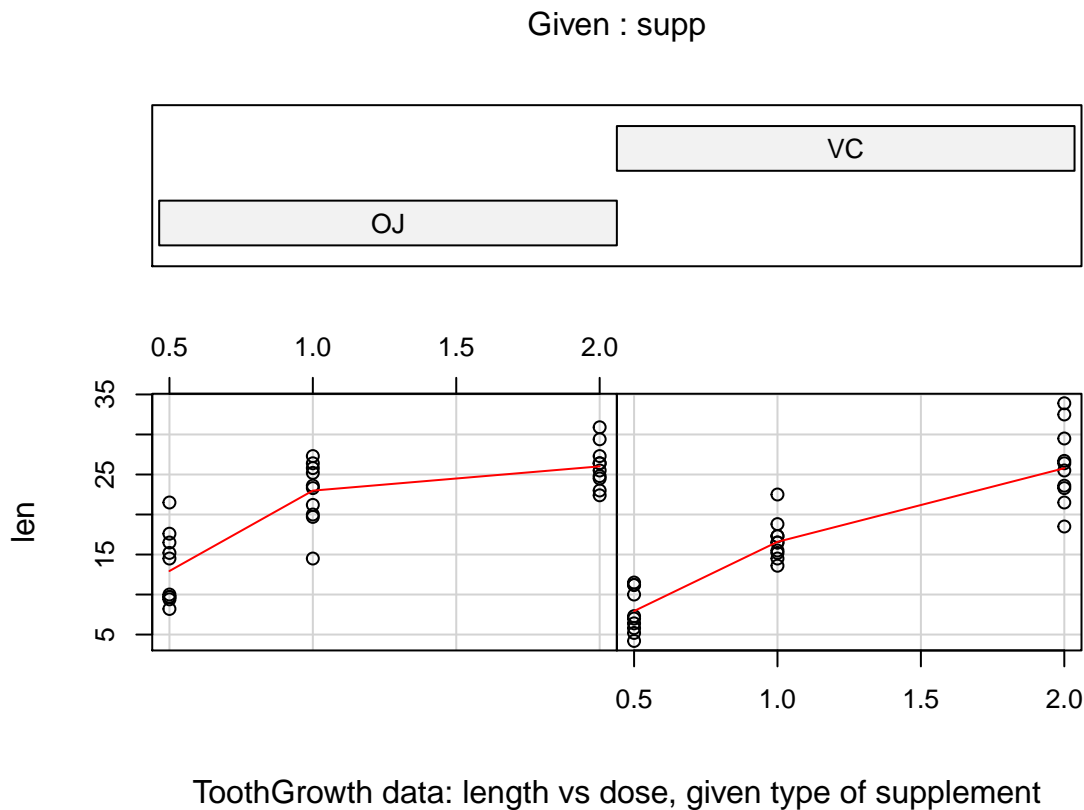
The obvious thing to do here is to run a regression with:

```
dose = f(supp, data)
```

or an ANOVA, but the instructions are explicit: > Only use the techniques from class, even if there's > other approaches worth considering.

From the co-plot we can see that OJ appears to be clearly superior at the ".5 mg" and "1.0 mg" dose levels, while at the "2.0 mg" dose level the means appear to be closer, but even at 2.0 the Vitamin C observations seem to be more spread out suggesting a wider variance.

The questions then are: 1. How large are the differences in means? 2. Are the differences of means statistically significant?

```r
# Create a new variable "dosefactor" coercing dose to a factor
ToothGrowth$dosefactor <- as.factor(ToothGrowth$dose)
# Create a table showing how many observations we have at each level
# of supp and dosefactor
    ToothCount <- table(ToothGrowth$supp, ToothGrowth$dosefactor)
    ToothCount
```

```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

Although we could compute a mean for all of the "VC" observations and another mean for all of the "OJ" observations and compare the two, the procedure is suspect since we know there were different treatment levels and from the graph alone we can see there was likely some impact from the treatment levels so these are not "identically distributed" as we would want for an IID assumption.

So, we really need to compute six means, one for each combination of the level. And then perform three tests, one for each level of dose comparing "OJ" and "VC".

```r
LengthOJat05 <- ToothGrowth$len[ToothGrowth$supp == "OJ" & ToothGrowth$dosefactor == "0.5"]
LengthOJat10 <- ToothGrowth$len[ToothGrowth$supp == "OJ" & ToothGrowth$dosefactor == "1"]
LengthOJat20 <- ToothGrowth$len[ToothGrowth$supp == "OJ" & ToothGrowth$dosefactor == "2"]

LengthVCat05 <- ToothGrowth$len[ToothGrowth$supp == "VC" & ToothGrowth$dosefactor == "0.5"]
LengthVCat10 <- ToothGrowth$len[ToothGrowth$supp == "VC" & ToothGrowth$dosefactor == "1"]
LengthVCat20 <- ToothGrowth$len[ToothGrowth$supp == "VC" & ToothGrowth$dosefactor == "2"]

t.test(LengthOJat05, LengthVCat05)
```

```
##
##  Welch Two Sample t-test
##
## data:  LengthOJat05 and LengthVCat05
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   1.719057 8.780943
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

```r
t.test(LengthOJat10, LengthVCat10)
```

```
##
##  Welch Two Sample t-test
##
## data:  LengthOJat10 and LengthVCat10
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   2.802148 9.057852
```

```
## sample estimates:
## mean of x mean of y
##      22.70      16.77
```

```
t.test(LengthOJat20, LengthVCat20)
```

```
##
##  Welch Two Sample t-test
##
## data:  LengthOJat20 and LengthVCat20
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean of x mean of y
##      26.06      26.14
```

The t-tests confirm there is a statistically significant difference in the means at the "0.5 mg" and "1.0 mg" levels, but the the significance disappears at the "2.0 mg" level. We can build a table to summarize the results.

```
tdiff05 <- t.test(LengthOJat05, LengthVCat05)$statistic
tdiff10 <- t.test(LengthOJat10, LengthVCat10)$statistic
tdiff20 <- t.test(LengthOJat20, LengthVCat20)$statistic

pdiff05 <- t.test(LengthOJat05, LengthVCat05)$p.value
pdiff10 <- t.test(LengthOJat10, LengthVCat10)$p.value
pdiff20 <- t.test(LengthOJat20, LengthVCat20)$p.value

ttest  <- c(tdiff05, tdiff10, tdiff20)
pvalue <- c(pdiff05, pdiff10, pdiff20)
interpretation <- ifelse(pvalue <= .05, "significant", "not significant")

results <- cbind(ttest, pvalue, interpretation)
rownames(results) <- c("0.5 mg", "1.0 mg", "2.0 mg")
results
```

```
##        ttest                 pvalue                 interpretation
## 0.5 mg "3.16973278367081"    "0.00635860676409681"  "significant"
## 1.0 mg "4.03276963371935"    "0.00103837587229988"  "significant"
## 2.0 mg "-0.0461361049092349" "0.963851588723373"    "not significant"
```

We have chosen to claim "significance" for pvalues less than .05 ("pvalue <= .05") or an alpha level of 5%.

**ASSUMPTIONS**

This is assuming that the variables are IID (Independent and Identially Distributed) and follow a normal ("Gaussian") distribution that can be approximated by a Student's T distribution.