# Analysis Project

Alper Halbutogullari, November 10, 2015

## Overview

In this project we will perform some basic exploratory analyses on the ToothGrowth data. The data is set of 60 observations, 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).

Our goals are to:
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.
4. state our conclusions.

## Load Data

```r
# load neccesary libraries
library(ggplot2)
library(datasets)

# The Effect of Vitamin C on Tooth Growth in Guinea Pigs
data(ToothGrowth)
toothGrowth <- ToothGrowth
toothGrowth$dose <- factor(toothGrowth$dose) # convert to factor
```

## Basic Summary of the data

```r
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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

```r
summary(toothGrowth)
```

```
##       len        supp      dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

```
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
```

```
table(toothGrowth$supp, toothGrowth$dose)
```

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

Please see Appendix for plots of the relations between the features and the results of the experiments.

As can be seen in those plots, there is a clear positive correlation between the tooth length and the dose levels of Vitamin C, for both delivery methods. Next we will analize the effect of the dosage and the supplement to support these claims with data.

## Analysis

Let's first split the data by supplement and dosage, and then build models from them.

Tooth-length vs. supplement:

```
ToothGrowth.supp_OJ <- subset(ToothGrowth, ToothGrowth$supp == "OJ")
fit_OJ <- lm(len ~ dose, data = ToothGrowth.supp_OJ)
summary(fit_OJ)
```

```
##
## Call:
## lm(formula = len ~ dose, data = ToothGrowth.supp_OJ)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7.2557 -3.7979 -0.0643  3.3521  7.9386
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   11.550      1.722   6.708 2.79e-07 ***
## dose           7.811      1.302   6.001 1.82e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.446 on 28 degrees of freedom
## Multiple R-squared:  0.5626, Adjusted R-squared:  0.547
## F-statistic: 36.01 on 1 and 28 DF,  p-value: 1.825e-06
```

```
ToothGrowth.supp_VC <- subset(ToothGrowth, ToothGrowth$supp == "VC")
fit_VC <- lm(len ~ dose, data = ToothGrowth.supp_VC)
summary(fit_VC)
```

```
##
## Call:
## lm(formula = len ~ dose, data = ToothGrowth.supp_VC)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -8.2264 -2.6029  0.0814  2.2288  7.4893
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.295      1.427   2.309   0.0285 *
## dose          11.716      1.079  10.860 1.51e-11 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.685 on 28 degrees of freedom
## Multiple R-squared:  0.8082, Adjusted R-squared:  0.8013
## F-statistic: 117.9 on 1 and 28 DF,  p-value: 1.509e-11
```

The R-square values are high and the p-values are low above, so models fit well, i.e. there is a correlation between the tooth increase and dosage increase.

Now for supplement:

```
ToothGrowth.dose_05 <- subset(ToothGrowth, ToothGrowth$dose == "0.5")
fit_dose_05 <- lm(len ~ supp -1, data = ToothGrowth.dose_05)
summary(fit_dose_05)
```

```
##
## Call:
## lm(formula = len ~ supp - 1, data = ToothGrowth.dose_05)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.030 -3.305 -0.830  3.220  8.270
##
## Coefficients:
##        Estimate Std. Error t value Pr(>|t|)
## suppOJ   13.230      1.171  11.296 1.33e-09 ***
## suppVC    7.980      1.171   6.814 2.22e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.704 on 18 degrees of freedom
## Multiple R-squared:  0.9063, Adjusted R-squared:  0.8959
## F-statistic: 87.02 on 2 and 18 DF,  p-value: 5.585e-10
```

```r
ToothGrowth.dose_1 <- subset(ToothGrowth, ToothGrowth$dose == "1")
fit_dose_1 <- lm(len ~ supp -1, data = ToothGrowth.dose_1)
summary(fit_dose_1)
```

```
##
## Call:
## lm(formula = len ~ supp - 1, data = ToothGrowth.dose_1)
##
## Residuals:
##     Min      1Q Median      3Q     Max
## -8.200  -1.745  0.130   2.147   5.730
##
## Coefficients:
##         Estimate Std. Error t value Pr(>|t|)
## suppOJ    22.70       1.04   21.83 2.11e-14 ***
## suppVC    16.77       1.04   16.13 3.81e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.288 on 18 degrees of freedom
## Multiple R-squared:  0.9762, Adjusted R-squared:  0.9735
## F-statistic: 368.4 on 2 and 18 DF,  p-value: 2.496e-15
```

```r
ToothGrowth.dose_2 <- subset(ToothGrowth, ToothGrowth$dose == "2")
fit_dose_2 <- lm(len ~ supp -1, data = ToothGrowth.dose_2)
summary(fit_dose_2)
```

```
##
## Call:
## lm(formula = len ~ supp - 1, data = ToothGrowth.dose_2)
##
## Residuals:
##     Min      1Q Median      3Q     Max
## -7.640  -2.615  -0.150  1.765   7.760
##
## Coefficients:
##         Estimate Std. Error t value Pr(>|t|)
## suppOJ   26.060      1.226   21.25 3.36e-14 ***
## suppVC   26.140      1.226   21.32 3.19e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.877 on 18 degrees of freedom
## Multiple R-squared:  0.9805, Adjusted R-squared:  0.9784
## F-statistic: 453.1 on 2 and 18 DF,  p-value: 4.031e-16
```

The R-square values are high and the slopes tell us are low above, so models fit well, i.e. there is a correlation between the tooth increase and supplements. The Orange Juice has a higher coeffient at lower levels, but the coefficients (hence the effect) are equal for dose=2.

Let's also check the supplement change with t-tests:

```
testResults = list()

tr <- t.test(len ~ dose, data=subset(ToothGrowth.supp_OJ, ToothGrowth.supp_OJ$dose != "1"), var.equal=TI
testResults <- rbind(testResults, list(test="len vs dose for supp=OJ", p_value=tr$p.value, Conf_int_low=

tr <- t.test(len ~ dose, data=subset(ToothGrowth.supp_VC, ToothGrowth.supp_VC$dose != "1"), var.equal=TI
testResults <- rbind(testResults, list(test="len vs dose for supp=VC", p_value=tr$p.value, Conf_int_low=

tr <- t.test(len ~ supp, data=ToothGrowth.dose_05, var.equal=TRUE)
testResults <- rbind(testResults, list(test="len vs supp at dose=0.5", p_value=tr$p.value, Conf_int_low=

tr <- t.test(len ~ supp, data=ToothGrowth.dose_1, var.equal=TRUE)
testResults <- rbind(testResults, list(test="len vs supp at dose=1", p_value=tr$p.value, Conf_int_low=ti

tr <- t.test(len ~ supp, data=ToothGrowth.dose_2, var.equal=TRUE)
testResults <- rbind(testResults, list(test="len vs supp at dose=2", p_value=tr$p.value, Conf_int_low=ti

testResults
```

```
##      test                     p_value       Conf_int_low Conf_int_high
## [1,] "len vs dose for supp=OJ" 3.401859e-07 -16.27822    -9.381777
## [2,] "len vs dose for supp=VC" 4.957286e-09 -21.83284    -14.48716
## [3,] "len vs supp at dose=0.5" 0.005303661  1.770262     8.729738
## [4,] "len vs supp at dose=1"   0.0007807262 2.840692     9.019308
## [5,] "len vs supp at dose=2"   0.9637098    -3.722999    3.562999
```

These results are all in line with the previous results.

## Conclusions

Assuming that the sample is a representative of the normally distributed population (*normality*) and the treatements have been applied randomly (*homogeneity of variance*) and exclusively (*independence*), we can make the following conclusions:

Both the supplement and the dosage have a clear positive effect on the length of teeth of guinea pigs. Supplement type also makes a change, with Orange Juice having a greater influence at lower dosages (less than 2 mg to be precise).

Tooth–length vs Dose, by Supplement



Tooth–length vs Supplement, by Dose

Tooth−length vs Supplement



Tooth−length vs Dose