

# Guinea Pigs Tooth Growth Data Analysis

*Sam Hewamanage*

*August 4, 2014*

## Problem

Now in the second portion of the class, we're going to analyze the ToothGrowth data in the R datasets package.

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 hypothesis tests to compare tooth growth by supp and dose. (Use the techniques from class even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

## Analysis

### Observations:

- No missing values.
- Two independent variables, dose and supplement are factors with 3 and 2 different treatment levels, totalling 6 different treatment conditions.
- One continuous dependent variables, len.

```
rm(list=ls())
set.seed(1234)
data(ToothGrowth)
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 ...
```

```
summary(ToothGrowth)
```

```
##      len      supp      dose
## Min.   : 4.2   OJ:30   Min.    :0.50
## 1st Qu.:13.1   VC:30   1st Qu.:0.50
## Median :19.2           Median :1.00
## Mean   :18.8           Mean   :1.17
## 3rd Qu.:25.3           3rd Qu.:2.00
## Max.   :33.9           Max.    :2.00
```

```
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
```

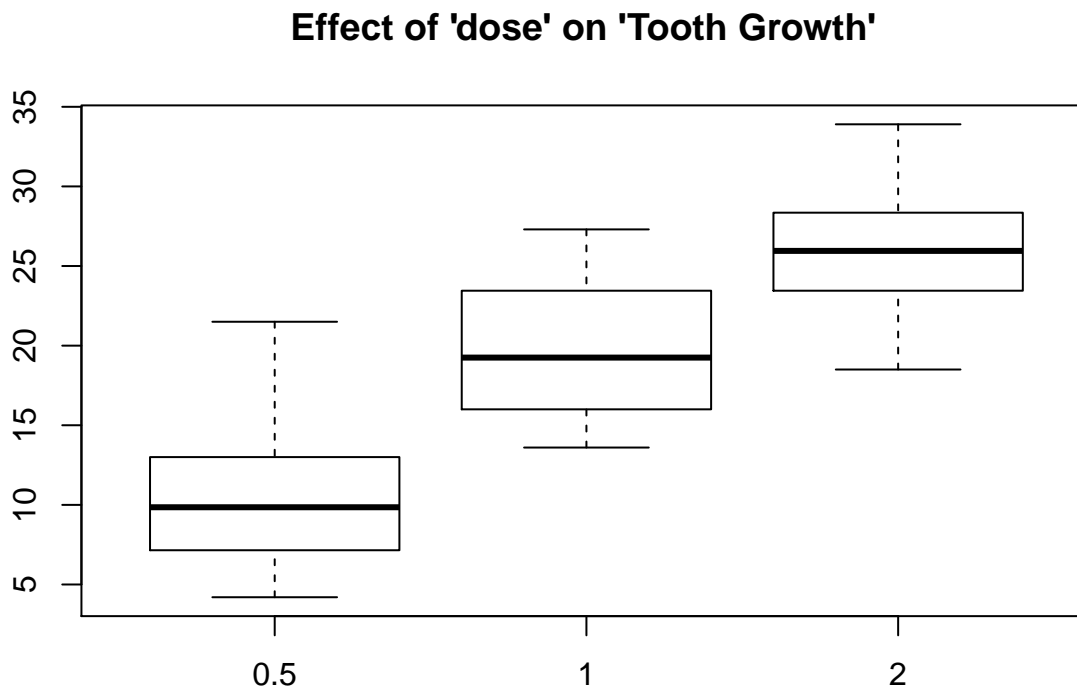
### Observations:

There is dependence of tooth length (len) on dose and supplement type. The width of len distribution gets smaller with the dose (approximately).

Observation	OJ	VC
shape	left skewed	more symmetric
correlation	strong	weak
spread	narrow	much broader

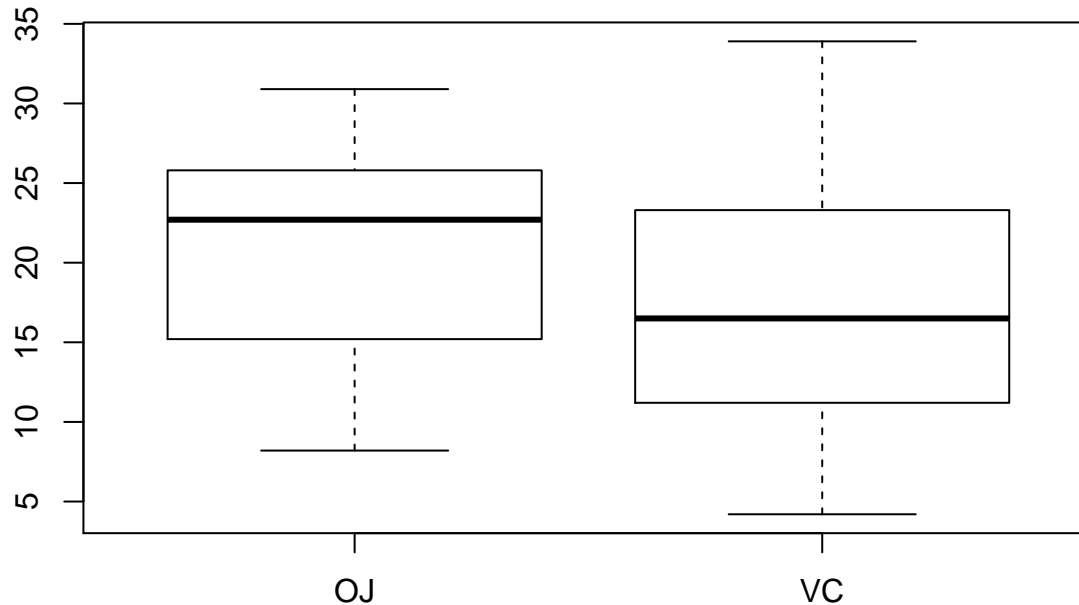
```
library(ggplot2)
p <- ggplot(data=ToothGrowth)

boxplot(len ~ dose, data=ToothGrowth)
title("Effect of 'dose' on 'Tooth Growth'")
```



```
boxplot(len ~ supp, data=ToothGrowth)
title("Effect of 'supplement type' on 'Tooth Growth'")
```

## Effect of 'suppliment type' on 'Tooth Growth'



Compare the additive model with interaction model and a model with both.

With the factors the additive explains ~70% of the variance. The interaction model itself show ~72% correction which is slightly higher than the additive model.

### Additive Model (1)

```
lm.form.1 = formula('len ~ dose + supp')
lm.form.1

## len ~ dose + supp

#oj.df <- subset(ToothGrowth, supp=='OJ')
#lm.model.1 <- lm(lm.form.1, data=oj.df)
lm.model.1 <- lm(lm.form.1, data=ToothGrowth)
summary(lm.model.1)

##
## Call:
## lm(formula = lm.form.1, 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)    9.273     1.282     7.23 1.3e-09 ***
```

```
## dose          9.764      0.877   11.14  6.3e-16 ***
## suppVC        -3.700      1.094   -3.38  0.0013 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.24 on 57 degrees of freedom
## Multiple R-squared:  0.704, Adjusted R-squared:  0.693
## F-statistic: 67.7 on 2 and 57 DF, p-value: 8.72e-16
```

```
confint(lm.model.1)
```

```
##              2.5 % 97.5 %
## (Intercept)  6.705 11.84
## dose         8.008 11.52
## suppVC       -5.890 -1.51
```

## Interaction Model (2)

```
lm.form.2 = formula('len ~ dose * supp')
lm.form.2
```

```
## len ~ dose * supp
```

```
lm.model.2 <- lm(lm.form.2, data=ToothGrowth)
summary(lm.model.2)
```

```
##
## Call:
## lm(formula = lm.form.2, data = ToothGrowth)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.23  -2.85   0.05   2.29   7.94
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    11.55      1.58     7.30 1.1e-09 ***
## dose           7.81      1.20     6.53 2.0e-08 ***
## suppVC        -8.26      2.24    -3.69 0.00051 ***
## dose:suppVC     3.90      1.69     2.31 0.02463 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.08 on 56 degrees of freedom
## Multiple R-squared:  0.73, Adjusted R-squared:  0.715
## F-statistic: 50.4 on 3 and 56 DF, p-value: 6.52e-16
```

```
confint(lm.model.2)
```

```
##           2.5 % 97.5 %
## (Intercept)  8.3821 14.718
## dose        5.4167 10.206
## suppVC      -12.7351 -3.775
## dose:suppVC  0.5176  7.291
```

Compare the two models

```
anova(lm.model.1, lm.model.2)
```

```
## Analysis of Variance Table
##
## Model 1: len ~ dose + supp
## Model 2: len ~ dose * supp
##   Res.Df  RSS Df Sum of Sq    F Pr(>F)
## 1      57 1023
## 2      56  934  1      88.9 5.33  0.025 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The interaction term is significant and hence this model does better than previous model in explaining the variance.

## Conclusions

Both dependent variables are significant in describing the dependent variable, len. The interaction model seems to be a slightly better than additive model of the same variables as it seems to describes the variance a tad more. Both models' p-values are significant and hence we reject null hypothesis in favor of the alternate, 'the variable are correlated'. Futher more there is indication that the interaction between dependent variables is significant as evident by the higher adjusted R-squared in the interaction model.

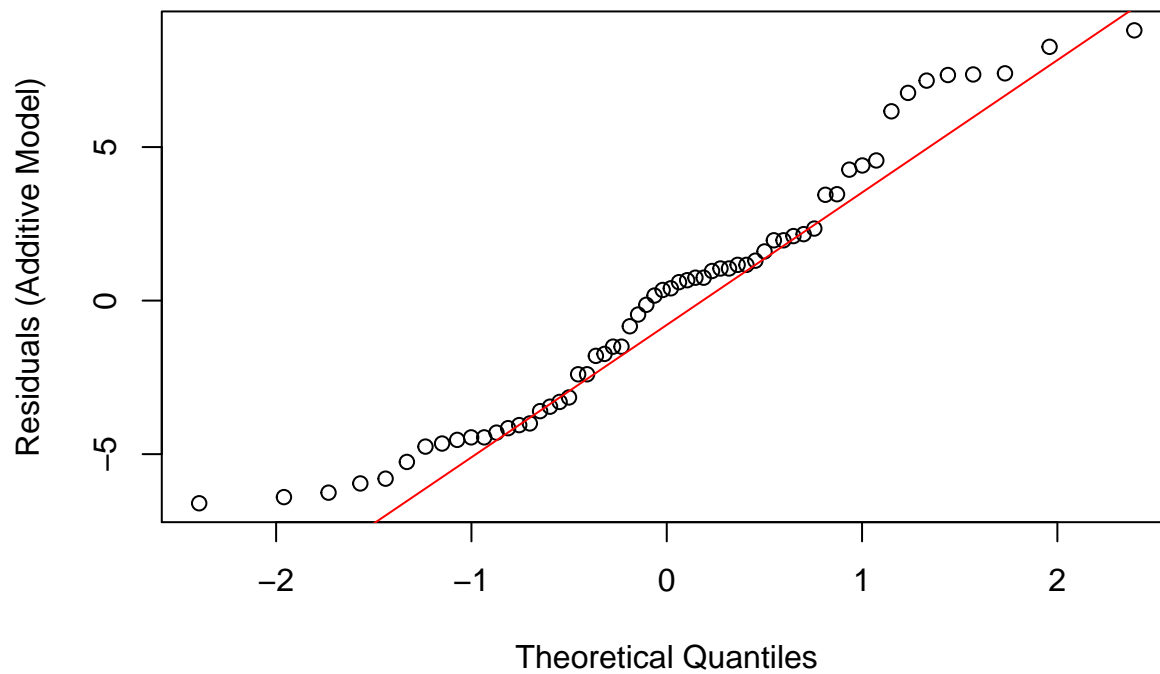
In conclusion, there is evidence that that dosage of vitamin C and the intake form have positive effect on the Guinea pigs' tooth growth.

## Assumptions:

1. The errors/residuals are normally distributed and iid. Looking the Q-Q plots below we can see the condition is approximately holds.

```
qqnorm(resid(lm.model.1), ylab="Residuals (Additive Model)")
qqline(resid(lm.model.1), col=2)
```

**Normal Q–Q Plot**



```
qqnorm(resid(lm.model.2), ylab="Residuals (Interaction Model)")  
qqline(resid(lm.model.1), col=2)
```

**Normal Q–Q Plot**

