Statistical Inference Project 2

Niraj Patel

October 26, 2015

# Loading Initial Libraries and Dataset

# Load necessary libraries  
library(datasets)  
library(ggplot2)  
library(gridExtra)

## Loading required package: grid

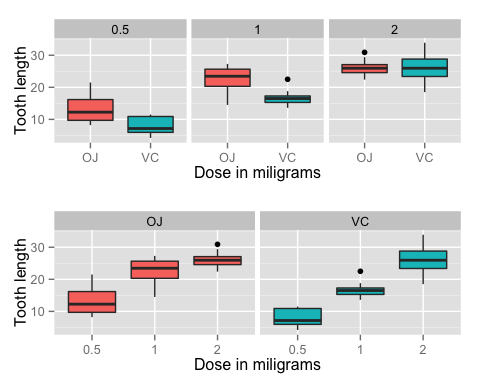
Data derived from ToothGrowth data sets are used. ToothGrowth describes the effect of Vitamin C on Tooth growth in Guinea pigs. The dosage of Vitamin C is administered using two different kinds of liquids, i.e. Ornage Juice (OJ) and Ascorbic Acid (VC).

data(ToothGrowth)  
TG <- ToothGrowth  
  
# Check the variables in supp in different columns  
str(TG)

## '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 ...

From the above, it can be seen that supp and dos are categorical variables.

# Visualise the data



To better understand the effect of each variable in isolation, a regression model needs to be constructed. For the purpose of this exercise, the method used will be the Linear Regression model.

# Linear Regression Model

# Study the correlation between variables , i.e. dose and supp type on len  
  
fit <- lm(len ~ dose + supp, data=TG)  
summary(fit)

##   
## Call:  
## lm(formula = len ~ dose + supp, data = TG)  
##   
## 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 \*\*\*  
## dose 9.7636 0.8768 11.135 6.31e-16 \*\*\*  
## suppVC -3.7000 1.0936 -3.383 0.0013 \*\*   
## ---  
## 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

From the value, we can see that 70% of the variance in the data is captured by the model. The intercept is 9.2725, meaning that without any Vitamin C supplied to the Guinea pigs, the average tooth length is 9.2725 units. The coefficient of dose is 9.7635714, meaning that regardless of type of Vitamin C, the length of teeth will increase by 9.76 units above the intercept value for every 1mg increase in dose of the Vitamin C.

The last coefficient is for the supplement type. Since the supplement type is a categorical variable, a dummy variable, suppVC, is created. The variable looks at the effect of administering Vitamin C via Ascorbic Acid while dosages are kept constant. The coefficient shows that the length of the tooth decreases by -3.7 when Ascorbic Acid is used as Vitamin C. This is clear in the graphs plotted above (row 1) as well for lower dosage values, however for the dosage of 2 mg and above, the trend is the opposite. To properly validate this data, a larger sample size would be required.

Keeping in mind that Vitamin C source is a categorical variable, we can also preliminarily deduce that length of the tooth will increase by -3.7 when Orange Juice is used as Vitamin C while keeping dosage constant.

# Confidence Interval and Conclusion

95% confidence intervals for two variables and the intercept are as follows.

confint(fit)

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

The confidence intervals mean that if we collect a different set of data and estimate parameters of the linear model many times, 95% of the time, the coefficient estimations will be in these ranges.

In conclusion, from the data above, it can be observed that both dosage and type of Vitamin C liquid used have an effect on the tooth length of the Guinea pigs. In coming to this conclusion, the following assumptions were taken.

1. The distribution of tooth length is normal
2. The guinea pigs in the sample are representative of the population of guinea pigs.