Tooth Growth Analysis

Jordan DeHerrera

November 22, 2015

# Synopsis

The goal of the following procedures is to analyze the ToothGrowth data in R. The analysis is organized into the following sections: \* Exploratory Data Analysis \* Confidence Intervals and Hypothesis Testing

## Exploratory Data Analysis

### Necessary Libraries

# Use ggthemes, grid, and gridExtra for styling  
library(ggthemes)

## Loading required package: ggplot2

library(grid)  
library(gridExtra)  
  
# Use ggplot2 library for exploratory graphs  
library(ggplot2)

### Analysis

Let's look at what the data looks like with the summary function:

# Summary of ToothGrowth data  
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

Based on the summary information, we can infer the following about the data: \* There are two supplement types (OJ and VC) \* There is a dosage level with variability. Let's find out how many unique values exist within the attribute:

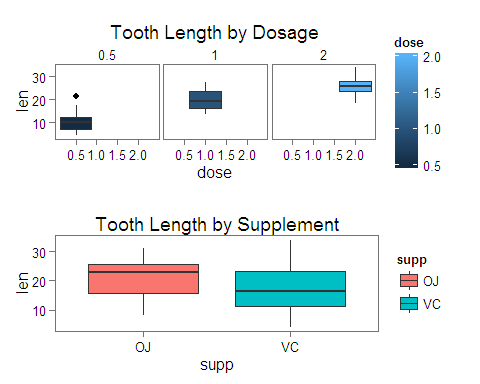
# Unique values of dose  
unique(ToothGrowth$dose)

## [1] 0.5 1.0 2.0

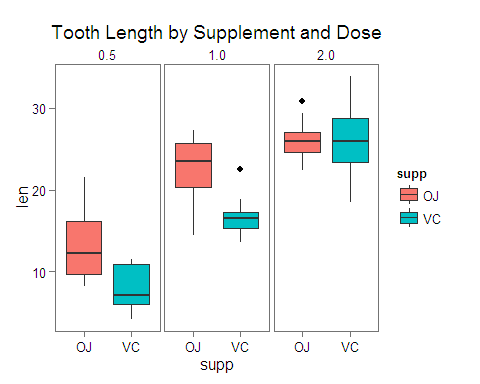
* Three unique dosage levels (0.5, 1.0, and 2.0)

Let's see how the population of data fits into these six different categories (3 dosages across 2 supplement types) using boxplot graphs with ggplot2:

# Create a boxplot of supplement types and doses  
plot.supplementDose <- (ggplot(aes(x = supp, y = len), data = ToothGrowth)   
 + geom\_boxplot(aes(fill=supp))  
 + facet\_wrap(~ dose)  
+ theme\_few()   
+ scale\_colour\_few()  
+ ggtitle("Tooth Length by Supplement and Dose"))  
  
# Create a boxplot of supplement types  
plot.supplement <- (ggplot(aes(x = supp, y = len), data = ToothGrowth)   
 + geom\_boxplot(aes(fill=supp))  
 + theme\_few()   
 + scale\_colour\_few()  
 + ggtitle("Tooth Length by Supplement"))  
  
# Create a boxplot of doses  
plot.dose <- (ggplot(aes(x = dose, y = len), data = ToothGrowth)   
 + geom\_boxplot(aes(fill=dose))  
 + facet\_grid(. ~ dose)   
 + theme\_few()   
 + scale\_colour\_few()  
 + ggtitle("Tooth Length by Dosage"))  
  
# Grid of all plots  
grid.arrange(plot.dose, plot.supplement, ncol=1)



plot.supplementDose



## Confidence Intervals and Hypothesis Testing

We'll use Welch's t-test to test the null hypothesis that two distinct populations have equal means. If the null hypothesis is rejected, then that means that there is some indication that the populations being examined have different sample means. In our specific analysis, that means that whatever attribute we're examining, could result in different tooth growth rates.

### Data Preparation

Let's start by separating the data into distinct comparative values. The Welch t-test requires data to be separated into 2 levels, so the supplement type is ready for analysis, but we will have to prepare the data into separate population groups based on unique pairings of the different dose types.

# Separate the data into separate groups based on dose  
popDose1 = subset(ToothGrowth, dose %in% c(0.5, 1.0))  
popDose2 = subset(ToothGrowth, dose %in% c(0.5, 2.0))  
popDose3 = subset(ToothGrowth, dose %in% c(0.1, 2.0))

### Analysis

Now, we can perform the Welch t-test on each set of data to determine when we should reject the null hypothesis:

#### Supplement Type

#### Dose Level

### Assumptions