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:

```r
# 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)
```
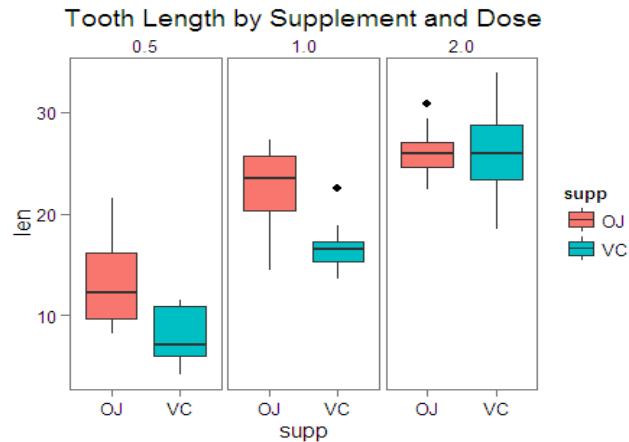

Tooth Length by Dosage

Tooth Length by Supplement

2
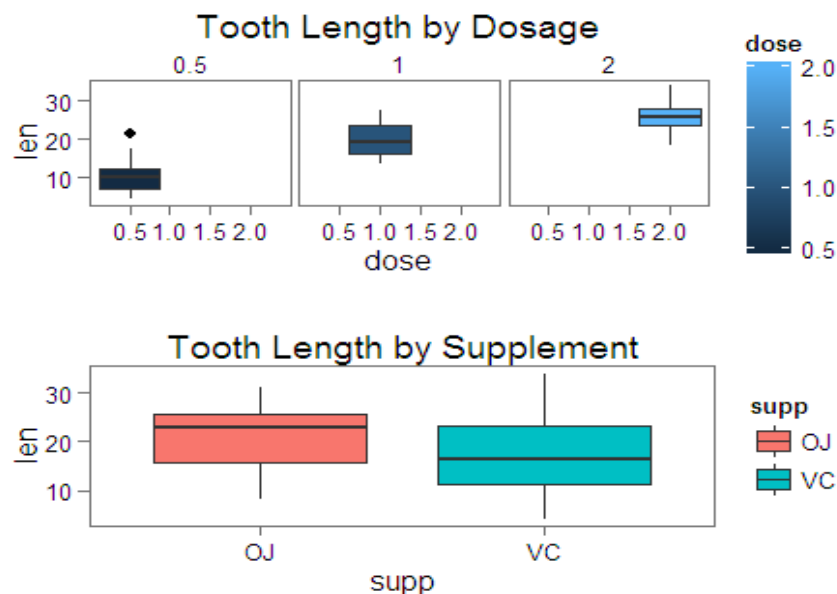
```
# Plot of Supplement and Dose
plot.supplementDose
```

Tooth Length by Supplement and Dose

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

```
t.test(len ~ supp, paired = F, var.equal = F, data = ToothGrowth)

##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
```

```
## mean in group OJ mean in group VC
##          20.66333          16.96333
```

As shown, we would not reject the null hypothesis. This indicates that the supplement type does not have an impact on tooth growth rate as we would expect based on the exploratory data analysis done.

Dose Level

```
t.test(len ~ supp, paired = F, var.equal = F, data = popDose1)

##
##   Welch Two Sample t-test
##
## data:  len by supp
## t = 3.0503, df = 36.553, p-value = 0.004239
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   1.875234 9.304766
## sample estimates:
## mean in group OJ mean in group VC
##           17.965           12.375

t.test(len ~ supp, paired = F, var.equal = F, data = popDose2)

##
##   Welch Two Sample t-test
##
## data:  len by supp
## t = 0.92163, df = 35.105, p-value = 0.363
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   -3.10849  8.27849
## sample estimates:
## mean in group OJ mean in group VC
##           19.645           17.060

t.test(len ~ supp, paired = F, var.equal = F, data = popDose3)

##
##   Welch Two Sample t-test
##
## data:  len by supp
## 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 in group OJ mean in group VC
##           26.06           26.14
```

We would reject the null hypothesis in each of the tests above this text. This indicates that the dosage has an impact on tooth growth rate with higher doses leading to longer tooth lengths as indicated by the difference in means of each of the sample populations. This is consistent with the expected results based on the exploratory analysis.

Assumptions

Some key assumptions in the analysis include the following:

- The populations given different doses and supplement types are normally distributed
- The populations are independent. That is, 60 guinea pigs were used so that each combination of dose level and delivery method were not affected by other methods
- Double blind research methods were used to prevent errors in recording observations due to unconscious bias from the observers/researchers