Statistical Inference, Part 2: Basic Inferential Data Analysis of Tooth Growth Data Set

Beth Skierski

3/31/2021

Overview

This report contains basic analysis of the "ToothGrowth" data set. The data is summarized and hypotheses are created about which factors have an effect on tooth growth. The hypotheses are tested and conclusions are summarized.

Loading Data & Initial Summary

First, the ToothGrowth data is loaded, and an understanding of the data set is established.

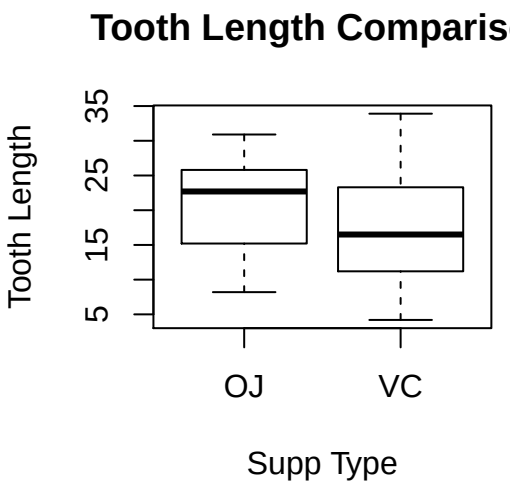
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

The data set is a data frame with 60 observations of 3 variables, len, supp, and dose. Length is the output, which is a numerical continuous variable. Supplement type is a factor variable with two levels, and dose is a numerical variable with three levels.

Confidence Interval Analysis

The first hypothesis to be investigated is: H0: length from OJ = length from VC; Ha: length from OJ is not = length from VC. To investigate, we can perform a t-test to look at length by supplement type.

```
##
##  Welch Two Sample t-test
##
## data:  tooth$len by tooth$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
```



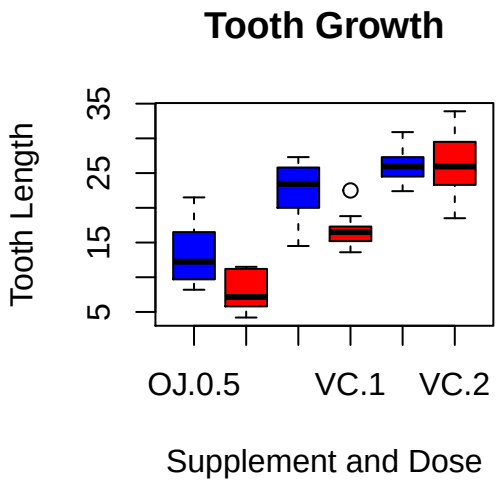
A p-value of 0.06 indicates that we cannot reject the null hypothesis there is no significant difference in length by supplement type. This is emphasized by overlap in the box plots.

The second hypothesis test to be investigated is: H0: length from dose groups are equal; Ha: length from dose groups are not equal. To investigate, we can perform a t-test to look at length by dose level.

```
##
##  Welch Two Sample t-test
##
## data:  tooth$len and tooth$dose
## t = 17.81, df = 59.798, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  15.66453 19.62881
## sample estimates:
## mean of x mean of y
## 18.813333  1.166667
```

A p-value less than 0.05 indicates that we can reject the null hypothesis; there is a significant difference in length for various dose amounts. This is emphasized by a lack of overlap in the box plots.

We can further create a box plot to compare tooth length for each supplement type by dose level.



This boxplot shows that within doses 0.5 and 2, there is no difference in tooth length between supplement types. In dose group 1, there is indeed a difference in length between supplement types, where OJ produces longer tooth growth.

Conclusions

This basic data analysis of tooth growth data shows that supplement type alone does not have an effect on tooth length in the overall population, but it does have an effect in the "1" dose group. The analysis shows that dose level consistenly has an effect on tooth length, where higher doses result in longer teeth.

Assumptions

In order to perform this analysis, we assume that the samples within each group are not related to one another or paired. We further assume that all test subjects were random, and thus that other influencing factors were minimized.

Appendix: Code for Report

```
knitr::opts_chunk$set(echo = TRUE)
tooth <- ToothGrowth
str(tooth)
t.test(tooth$len~tooth$supp)
boxplot(tooth$len~tooth$supp, main="Tooth Length Comparison",
   xlab="Supp Type", ylab="Tooth Length")
t.test(tooth$len, tooth$dose)
boxplot(len~dose, tooth, main = "Tooth Growth", xlab="Dose",
        ylab="Tooth Length")
boxplot(len~supp*dose, tooth, col=(c("blue","red")),
  main="Tooth Growth", xlab="Supplement and Dose",
  ylab="Tooth Length")
```