Statistical inference course project | Part 2

Danna

May 28, 2017

## Overview

This report aims to analyze the ToothGrowth data in the R datasets package. Per the course project instructions, the following items should occur:

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/or hypothesis tests to compare tooth growth by supp and dose (only use the techniques from class, even if thereâs other approaches worth considering). 4.State your conclusions and the assumptions needed for your conclusions.

### 1. Load the ToothGrowth data and perform some basic exploratory data analyses

# Load ToothGrowth data  
library(datasets)  
x <- ToothGrowth  
# convert dose to factor  
x$dose <- as.factor(x$dose)

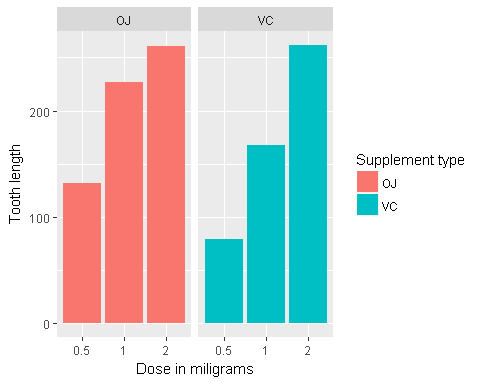
str(x)

## '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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...

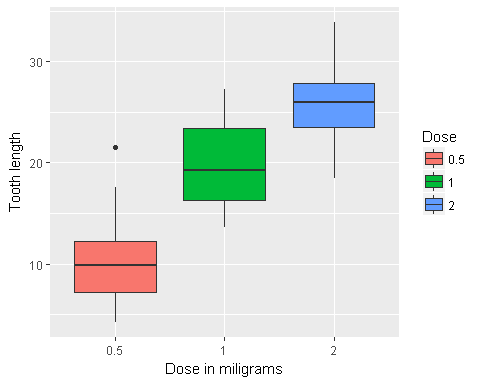
table(x$dose, x$supp)

##   
## OJ VC  
## 0.5 10 10  
## 1 10 10  
## 2 10 10

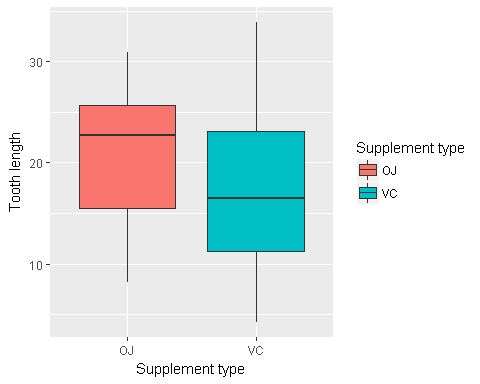
library(ggplot2)  
ggplot(data=x, aes(x=dose, y=len, fill=supp)) + geom\_bar(stat="identity",) + facet\_grid(. ~ supp) + xlab("Dose in miligrams") + ylab("Tooth length") + guides(fill=guide\_legend(title="Supplement type"))



ggplot(aes(x=dose, y=len), data=x) + geom\_boxplot(aes(fill=dose)) + xlab("Dose in miligrams") + ylab("Tooth length") + guides(fill=guide\_legend(title="Dose"))



ggplot(aes(x=supp, y=len), data=x) + geom\_boxplot(aes(fill=supp)) + xlab("Supplement type") + ylab("Tooth length") + guides(fill=guide\_legend(title="Supplement type"))



### 2. Provide a basic summary of the data.

# Display a summary of the data  
summary(x)

## len supp dose   
## Min. : 4.20 OJ:30 0.5:20   
## 1st Qu.:13.07 VC:30 1 :20   
## Median :19.25 2 :20   
## Mean :18.81   
## 3rd Qu.:25.27   
## Max. :33.90

### 3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there’s other approaches worth considering)

t.test(len ~ supp, data = x)

##   
## 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

Null hypothesis can not be rejected as confindence intervals contain zero and p-value is 0.06.

Supplement types seems to have no impact on Tooth growth.

# three groups as per dose level pairs  
x.doses\_0.5\_1.0 <- subset (x, dose %in% c(0.5, 1.0))   
x.doses\_0.5\_2.0 <- subset (x, dose %in% c(0.5, 2.0))   
x.doses\_1.0\_2.0 <- subset (x, dose %in% c(1.0, 2.0))   
  
# Check for dose levels (0.5, 1.0)  
t.test(len ~ dose, data = x.doses\_0.5\_1.0)

##   
## Welch Two Sample t-test  
##   
## data: len by dose  
## t = -6.4766, df = 37.986, p-value = 1.268e-07  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -11.983781 -6.276219  
## sample estimates:  
## mean in group 0.5 mean in group 1   
## 10.605 19.735

# Check for dose levels (0.5, 2.0)  
t.test(len ~ dose, data = x.doses\_0.5\_2.0)

##   
## Welch Two Sample t-test  
##   
## data: len by dose  
## t = -11.799, df = 36.883, p-value = 4.398e-14  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -18.15617 -12.83383  
## sample estimates:  
## mean in group 0.5 mean in group 2   
## 10.605 26.100

# Check for dose levels (1.0, 2.0)  
t.test(len ~ dose, data = x.doses\_1.0\_2.0)

##   
## Welch Two Sample t-test  
##   
## data: len by dose  
## t = -4.9005, df = 37.101, p-value = 1.906e-05  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -8.996481 -3.733519  
## sample estimates:  
## mean in group 1 mean in group 2   
## 19.735 26.100

The p-value is less than 0.05 and confindence intervals don’t contian 0. The average toot length increases with an inceasing dose. The null hypothesis can be rejected.

### 4. State your conclusions and the assumptions needed for your conclusions.

Given the following assumptions: 1.The sample is representative of the population 2.The distribution of the sample means follows the Central Limit Theorem 3.For the t-tests, the variances are assumed to be different for the two groups being compared. This assumption is less stronger than the case in which the variances are assumed to be equal.