Statistical Inference and Tooth Growth

Eamon Corr

Monday, January 11, 2016

## ToothGrowth Data Analysis

##### Analysis of tooth growth by supplement and dosage using confidence intervals and hypothesis tests.

#### Loading & Data Processing

Load the dataset ToothGrowth from the R package datasets and check the first few rows.

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(datasets)  
data(ToothGrowth)  
head(ToothGrowth,3)

## len supp dose  
## 1 4.2 VC 0.5  
## 2 11.5 VC 0.5  
## 3 7.3 VC 0.5

Checking the dataset's dimensions, variables and charcteristics:

str(ToothGrowth)

## '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 ...

unique(ToothGrowth$supp)

## [1] VC OJ  
## Levels: OJ VC

unique(ToothGrowth$dose)

## [1] 0.5 1.0 2.0

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

This quick check reveals the variable (factor) 'supp' with two levels, 'OJ' and 'VC'. The dose variable has three unique values of 0.5, 1, and 2. The 'len' (tooth length) has a median of 19.25 and a range from 4.2 to 33.90.

The data is organised firstly by 'supp' and then dose with equal quantities for each. There's no NA's.

summary(ToothGrowth$len)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 4.20 13.08 19.25 18.81 25.28 33.90

any(is.na(ToothGrowth$len))

## [1] FALSE

table(ToothGrowth$dose)

##   
## 0.5 1 2   
## 20 20 20

table(ToothGrowth$supp)

##   
## OJ VC   
## 30 30

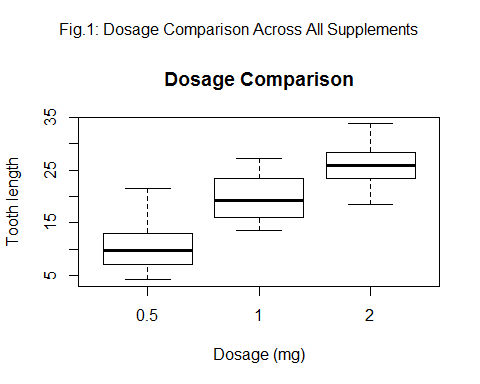
The dataset comes with no supporting information regarding the experiment/trial prerequisites, conditions, variable names and observations. Therefore we will make the following assumptions:

'supp' is short for supplement and the supplements include 'OJ', possibly Orange Juice and 'VC'. The dose (0.5, 1, 2) is a measure of vitamin C and the units are milligrams (mg).

#### Some Exploratory Data Analysis

Let's look at the dosage comparison across all supplements for tooth length (growth), means and standard deviations for all groups.

par(mfrow = c(1,1), oma= c(0,0,2,0))  
boxplot(len ~ dose, data=ToothGrowth, xlab='Dosage (mg)', ylab='Tooth length', main='Dosage Comparison')  
mtext('Fig.1: Dosage Comparison Across All Supplements', outer = TRUE)



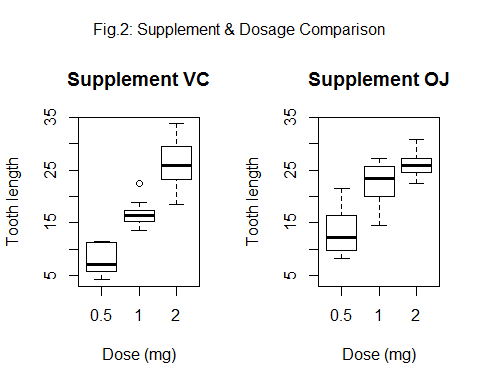
a <- group\_by(ToothGrowth, supp,dose)  
summarise(a, mean(len), round(sd(len),2))

## Source: local data frame [6 x 4]  
## Groups: supp [?]  
##   
## supp dose mean(len) round(sd(len), 2)  
## (fctr) (dbl) (dbl) (dbl)  
## 1 OJ 0.5 13.23 4.46  
## 2 OJ 1.0 22.70 3.91  
## 3 OJ 2.0 26.06 2.66  
## 4 VC 0.5 7.98 2.75  
## 5 VC 1.0 16.77 2.52  
## 6 VC 2.0 26.14 4.80

In general we can see a step increase in tooth growth as dosage increases across all measurements.

Looking at the data in more detail (Fig.2 below) we can see a clear increase in tooth length as dosage is increased for supplement VC. However, supplement OJ is a little less clearer. There's a lot more overlap between the three dosages particularly dosage 1 and 2. Supplement OJ seems to be more effective for lower doses but seems to be on par with VC at the higher level dosage. We'll do a little more analysis on supplement OJ when we investigate the confidence intervals later.

vc <- subset(ToothGrowth, ToothGrowth$supp == 'VC')  
oj <- subset(ToothGrowth, ToothGrowth$supp == 'OJ')  
#to scale the plots so that they are on the same range  
#we calcualte the range of the dataset for both  
rng <- range(vc$len, oj$len, na.rm=T)  
par(mfrow = c(1,2), oma= c(0,0,2,0))  
boxplot(len ~ dose, data=vc, ylim = rng, xlab='Dose (mg)', ylab='Tooth length', main='Supplement VC')  
boxplot(len ~ dose, data=oj, ylim = rng, xlab='Dose (mg)', ylab='Tooth length', main='Supplement OJ')  
mtext('Fig.2: Supplement & Dosage Comparison', outer = TRUE)



### Confidence Intervals and Hypothesis Testing to Compare Tooth Growth by Supplement & Dosage.

##### Assumptions

Before carrying out further analysis we will make the following assumptions. There's no evidence that the dataset consists of paired observations, i.e. no id variable to indicate remeasurement of the units. We will assume the variance is not equal. Looking at the boxplots in Fig.1 we can see variations in density across dosage levels across supplements and within supplements. Therefore we will assume the variance are not equal. All calculations will be carried out using the 95% confidence interval.

Note: If the data was treated as paired, intervals may be different and may not include zero. It's also likely we will get different intervals when considering variances to be equal or unequal.

##### Confidence Intervals & Hypothesis Testing.

Just a quick note on confidence intervals and hypothesis testing. 95% refers to the fact that if one were to repeatably get samples of size n, about 95% of the intervals would contain mu, the parameter we're trying to estimate. If an interval contains zero we cannot rule out zero as the possibility for the population difference between the two groups.

##### Determine confidence intervals and hypothesis test to compare supplements for tooth growth across both supplements VC and OJ in general.

Hypothesis Test: Ho: µ(oj) = µ(vc); H1: µ(oj) != µ(vc)

vc <- ToothGrowth$len[1:30]  
oj <- ToothGrowth$len[31:60]  
t.test(oj,vc, paired=FALSE, var.equal = FALSE)

##   
## Welch Two Sample t-test  
##   
## data: oj and vc  
## 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 of x mean of y   
## 20.66333 16.96333

P-value = 0.06063 Therefore do not reject the null hypothesis, i.e. there is no statistical difference between the two supplements (means of tooth growth) overall. The confidence interval also contains zero. We should note that the p-value is not large and from out exploratory data analysis we know there are differences (Fig.2) when we break the data down further and analysis dosage and tooth growth.

##### Determine confidence intervals and hypothesis test to compare supplements for tooth growth for dosage = 0.5mg

Hypothesis Test: Ho: µ(oj\_0.5) = µ(vc\_0.5) ; H1: µ(oj\_0.5) != µ(vc\_0.5)

vc\_0.5 <- ToothGrowth$len[1:10]  
oj\_0.5 <- ToothGrowth$len[31:40]  
t.test(oj\_0.5,vc\_0.5, paired=FALSE, var.equal = FALSE)

##   
## Welch Two Sample t-test  
##   
## data: oj\_0.5 and vc\_0.5  
## t = 3.1697, df = 14.969, p-value = 0.006359  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 1.719057 8.780943  
## sample estimates:  
## mean of x mean of y   
## 13.23 7.98

P-value = 0.006359 indicating there is a significant difference between the two supplements for a dosage of 0.5mg. The confidence interval does not contain zero. Therefore we can reject the null hypothesis and accept the alternative that the two supplements (means of tooth growth) are not equal.

##### Determine Confidence intervals and hypothesis test to compare supplements for tooth growth for dosage = 1mg.

Hypothesis Test: Ho: µ(oj\_1.0) = µ(vc\_1.0) ; H1: µ(oj\_1.0) != µ(vc\_1.0)

vc\_1.0 <- ToothGrowth$len[11:20]  
oj\_1.0 <- ToothGrowth$len[41:50]  
t.test(oj\_1.0,vc\_1.0, paired=FALSE, var.equal = FALSE)

##   
## Welch Two Sample t-test  
##   
## data: oj\_1.0 and vc\_1.0  
## t = 4.0328, df = 15.358, p-value = 0.001038  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 2.802148 9.057852  
## sample estimates:  
## mean of x mean of y   
## 22.70 16.77

P-value = 0.001038 indicating there is a significant difference between the two supplements for a dosage of 1mg. The confidence interval does not contain zero. Therefore we can reject the null hypothesis and accept the alternative that the two supplements (means of tooth growth) are not equal.

##### Determine Confidence intervals and hypothesis test to compare supplements for tooth growth for dosage = 2mg.

Hypothesis Test: Ho: µ(oj\_2.0) = µ(vc\_2.0) : H1: µ(oj\_2.0) != µ(vc\_2.0)

vc\_2.0 <- ToothGrowth$len[21:30]  
oj\_2.0 <- ToothGrowth$len[51:60]  
t.test(oj\_2.0,vc\_2.0, paired=FALSE, var.equal = FALSE)

##   
## Welch Two Sample t-test  
##   
## data: oj\_2.0 and vc\_2.0  
## 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 of x mean of y   
## 26.06 26.14

P-value = 0.9639 indicating there is not a significant difference between the two supplements for a dosage of 2mg. The confidence interval does contain zero. Therefore we cannot reject the null hypothesis and the two supplements (means of tooth growth) are equal.

##### Determine Confidence intervals and hypothesis test for supplement OJ for tooth growth for dosage 0.5mg and 1mg.

During the exploratory data analysis we discussed overlap between the three dosages particularly dosage 1 and 2 for supplement OJ. We will analyse this further using confidence intervals and hypothesis testing.

Hypothesis Test: Ho: µ(oj\_0.5) = µ(oj\_1.0) ; H1: µ(oj\_0.5) != µ(vc\_1.0)

t.test(oj\_1.0,oj\_0.5, paired=FALSE, var.equal = FALSE)

##   
## Welch Two Sample t-test  
##   
## data: oj\_1.0 and oj\_0.5  
## t = 5.0486, df = 17.698, p-value = 8.785e-05  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 5.524366 13.415634  
## sample estimates:  
## mean of x mean of y   
## 22.70 13.23

P-value = 8.785e-05 confirms our believe there is a significant difference between the two doses for supplement OJ. The confidence interval does not contain zero. Therefore we can reject the null hypothesis and confirm there is a difference between the 0.5mg and the higher 1mg dosage.

##### Comparing tooth growth for dosage 1mg and 2mg for the OJ supplement.

Hypothesis Test: Ho: µ(oj\_1) = µ(oj\_2) ; H1: µ(oj\_1) != µ(vc\_2)

t.test(oj\_1.0,oj\_2.0, paired=FALSE, var.equal = FALSE)

##   
## Welch Two Sample t-test  
##   
## data: oj\_1.0 and oj\_2.0  
## t = -2.2478, df = 15.842, p-value = 0.0392  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -6.5314425 -0.1885575  
## sample estimates:  
## mean of x mean of y   
## 22.70 26.06

P-value = 0.0392 indicating there is a significant difference between the 1gm and 2gm doses for supplement OJ. The confidence interval does not contain zero. Therefore we can reject the null hypothesis and confirm there is a difference between the 1mg and the higher 2mg dosage.

### Conclusions

From this analysis We can conclude that in general, increases in vitamin C dosages for supplements OJ and VC, will have a positive effect on the growth of tooth length.

Dosages of 0.5mg and 1mg will result in larger tooth growth for supplement OJ than for supplement VC.

However, a dosage of 2mg will have a similar impact on tooth growth for both supplements.