Tooth Growth Analysis: Correlation with diet and Vitamin C

Winston

October 21, 2014

#### Note to reader

This .pdf may look at bit strange because just as I was finishing I upgraded my Mac OS to Yosemite. The LaTeX version seemed to stop working with RPubs. So I saved as a word file and the saved that as a .pdf, which I uploaded.

## Get the data and take a quick look at it

data(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 ...

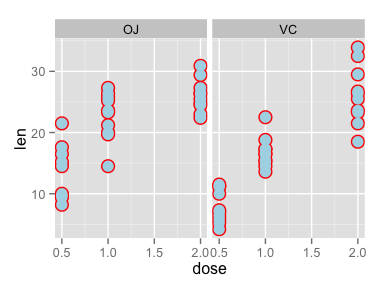
## Exploratory Analysis

Here is a plot of the data showing behavior trends in tooth growth factored by supplement.

require(ggplot2)

## Loading required package: ggplot2

ggplot(ToothGrowth, aes(x=dose, y = len))+geom\_point(size = 5, colour="red")+geom\_point(size = 4, colour="lightblue")+facet\_grid(. ~ supp)



A summary of the data:

## len supp dose   
## Min. : 4.2 OJ:30 Min. :0.50   
## 1st Qu.:13.1 VC:30 1st Qu.:0.50   
## Median :19.2 Median :1.00   
## Mean :18.8 Mean :1.17   
## 3rd Qu.:25.3 3rd Qu.:2.00   
## Max. :33.9 Max. :2.00

shows the data appear to be well behaved (10 unpaired observations at each condition) and do not require cleaning for this anlaysis.

## len supp dose  
## 1 4.2 VC 0.5  
## 2 11.5 VC 0.5  
## 3 7.3 VC 0.5  
## 4 5.8 VC 0.5  
## 5 6.4 VC 0.5

## Confidence Interval Analysis

The first step is to look at the statistical significance of the shift in the data with increasing dose. The conclusion after looking at all the pairs is that there is an upward trend (meaning higher doses correlate to statistically significant higher length) among all the data.

To keep the analysis part clean, first split up the data by Supplement and Dose

##Group the data into individual sets with somewhat descriptive names (Supplement and Dose)  
 ##Vitamin C set  
 VC05<-ToothGrowth[1:10,]  
 VC10<-ToothGrowth[11:20,]  
 VC20<-ToothGrowth[21:30,]  
 ##Ornage Juice Set  
 OJ05<-ToothGrowth[31:40,]  
 OJ10<-ToothGrowth[41:50,]  
 OJ20<-ToothGrowth[51:60,]

## Within Vitamin C Data

Some sample analysis

#### Dose0.5 to Dose1.0

a<- t.test(VC10$len, VC05$len, paried=FALSE)  
lcb<-a$conf.int[1]  
ucb<-a$conf.int[2]

The confidence interval difference between tooth length at these doses is 6.3143 to 11.2657 which does not contain 0.

#### Dose1.0 to Dose2.0

The confidence interval difference between tooth length at these doses is 5.6857 to 13.0543 which does not contain 0.

## Within Orange Juice

We can draw a similar conclusion that the correlation of increased does and tooth length is significnat statistically.

#### Dose0.5 to Dose1.0

a<- t.test(OJ10$len, OJ05$len, paried=FALSE)  
lcb<-a$conf.int[1]  
ucb<-a$conf.int[2]

The confidence interval difference between tooth length at these doses is 5.5244 to 13.4156 which does not contain 0.

#### Dose1.0 to Dose1.0

a<- t.test(OJ10$len, VC10$len, paried=FALSE)  
lcb<-a$conf.int[1]  
ucb<-a$conf.int[2]

The confidence interval difference between tooth length from the Orange Juice and Vitamin C at doses of 1.0 is 2.8021 to 9.0579 which does not contain 0.

#### Conclusion

In all the cases above the shifts in the mean, both with increasing dose and the comparison of Vitamin C to Orange Juice are significant.

## Hypothesis Testing

### Hypothesis to be tested: *Orange Juice is more effective (produces more tooth growth) than Vitamin C*

#### Test at dose of 1.0.

b<-t.test(OJ10$len, VC10$len, paired=FALSE, var.equal=TRUE)

The anlaysis shows that for OJ10 and VC10 using the Two Sample t-test the t-statistic is 18 and the confidence interval is 2.8407 and 9.0193.

### For var.equal=FALSE

Note that for the assumption var.equal=FALSE analysis shows the t-statistic is 15.3577 and the confidence interval is 2.8021 and 9.0579. Not significantly different than above.

#### Test at dose of 2.0

b<-t.test(OJ20$len, VC20$len, paired=FALSE, var.equal=TRUE)

The anlaysis shows that for OJ20 and VC20 using the Two Sample t-test the t-statistic is 18 and the confidence interval is -3.723 and 3.563.

## Assumptions

Here I assume that the variances of the distributions are equal between the Orange Juice and the Vitamin C populations. I tested this in one case and found it be not e important.

## Conclusion

The hypothesis that Orange Juice is more effective than Vitamin C in tooth length increases is **FALSE** since at higher doses the difference in the outcome includes zero.

However, at lower dose = 1.0, Orange Juice is more effective than vitamin C to a high degree of statistical signifance since the confidence interval does not contain zero.