Tooth Growth Analysis in Guinea Pigs

Chris Selig

January 7, 2017

## Part 2

# Overview

In the second part of this report, I will perform hypothesis testing on tooth growth rate in guinea pigs to show that for a dose level of 0.5mg/day or 1.0mg/da, the orange juice delivery method is superior to the ascorbic acid method.

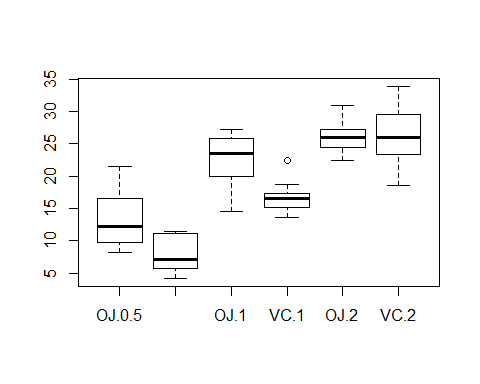
### Data

We will be exploring the response length of odontoblasts in guinea pigs to three different dose levels of vitamin C (0.5, 1, and 2 mg/day) and through two different delivery methods (orange juice, coded as "OJ" and ascorbic acid, coded as "VJ").

### Exploratory Data Analysis

Basic structure of the data is 60 observations in 3 variables, with a length (len), supp (delivery method), and dose (dose level of vitamin C).

## '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 ...

Furthermore, it appears that as you increase the dose level, the length of tooth growth also increases. 

Also, here are some summary statistics of the data.

## Source: local data frame [6 x 5]  
## Groups: supp [?]  
##   
## supp dose mean sd var  
## <fctr> <dbl> <dbl> <dbl> <dbl>  
## 1 OJ 0.5 13.23 4.46 19.89  
## 2 OJ 1.0 22.70 3.91 15.30  
## 3 OJ 2.0 26.06 2.66 7.05  
## 4 VC 0.5 7.98 2.75 7.54  
## 5 VC 1.0 16.77 2.52 6.33  
## 6 VC 2.0 26.14 4.80 23.02

Indeed, you can see that the means all increase when the dose level increases.

### Hypothesis Testing

First up, we will do a high level test to see if the means of each delivery method equal each other. So: Ho: muOJ = muVC  
Ha: muOJ != to muVC

## [1] -0.1670064 7.5670064  
## attr(,"conf.level")  
## [1] 0.95

## [1] 0.06039337

The difference in means is not significant (confidence interval includes 0), and the p-value > 0.05, so we fail to reject the null hypothesis. Surprisingly, it appears that the means could equal each other for the two delivery methods.

Lets explore a bit further. From our calculated means, it appears that the orange juice (OJ) delivery method is better for all doses except 2mg/day. Lets test to see if the difference is significant.

First up, 0.5mg/day. Our hypotheis to test are: Ho: muOJ0.5 = muVC0.5 and Ha: muOJ0.5 > muVC0.5

## [1] 1.770262 8.729738  
## attr(,"conf.level")  
## [1] 0.95

At a 95% confidence interval, we get a p-value = 0.0053037, which means we can reject the null hypothesis, and it appears that OJ is the better delivery method at the 0.5mg/day dose level.

Next, we will do the same for the 1mg/day dose level. #Ho muOJ1.0 = muVC1.0 and Ha muOJ1.0 > muVC1.0

At a 95% confidence interval, we get a p-value = 7.807261710^{-4}, which means we can reject the null hypothesis, and it appears that OJ is the better delivery method at the 1mg/day dose level.

And for the 2mg/day dose level, we find that the p-value = 0.9637098, which is greather than 5%, so we fail to reject the null hypothesis. So, we cannot say that OJ is the better delivery for 2mg/day

Finally, lets test to see, for the OJ delivery method, if an increase in dose is significant for tooth growth. First up, is checking if 1mg/day is greater than 0.5mg/day, so our hypothesis test is: Ho: mu0.5 = mu1.0 and Ha: mu1.0 > mu0.5

At a 95% confidence interval, we get a p-value = 1.26629710^{-7}, which means we can reject the null hypothesis, and it appears 1mg/day is greather than 0.5mg/day.

And, for: Ho: mu1.0 = mu2.0 and Ha: mu2.0 > mu1.0, we get a p-value of 1.810828510^{-5}, so we can reject the null hypothesis heare as well. 2mg/day seems better than 1.0mg/day, and by default, better than 0.5mg/day.

### Assumptions

To conduct the hypothesis tests, I used t-tests due to low sample size (n = 10) for each dose and delivery method. This assumes that given enough observations, the data would be approximately normal. Furthermore, to conduct the t-tests, I assumed the group variances were equal. Finally, I did not adjust for false positive errors because with so few hypothesis tests, I would only have 0.3 false positives, which I believe is an acceptable number for this analysis.

### Conclusion

We concluded that the orange juice delivery is the better delivery method, and that the 2mg/day is the prefered dosage.

## Part 2 Appendix

### False Positive Rate

Checking for the rate of false positives. Didn't correct for a false positive rate, due to the limited number of hypothesis tests.  
For the amount of tests conducted, only 0.3 tests could be false positive, which I view as an acceptable level.

numhypotheistests <- 6  
levelsignificance <- 0.05  
numhypotheistests \* levelsignificance

## [1] 0.3

### Entire Code

library(dplyr)  
  
#Load Data  
toothgrowthdata <- datasets::ToothGrowth  
  
#Group the data by supp and then dose.  
groupeddata <- toothgrowthdata %>% group\_by(supp, dose)  
  
#Basic Structure of the data  
dim(toothgrowthdata)

## [1] 60 3

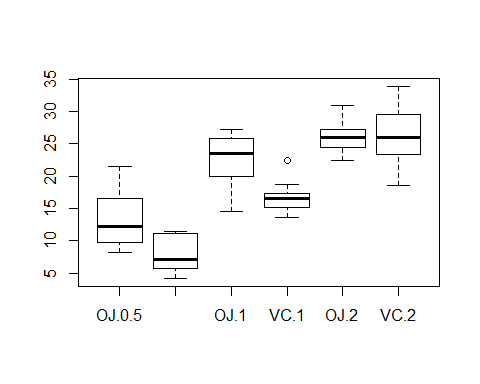
str(toothgrowthdata)

## '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 ...

summary(toothgrowthdata)

## 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

#Exploratory Plot  
boxplot(len ~ supp + dose, data = ToothGrowth)



#Confirm What the PLot Shows, Calulate Mean, Sd, Var  
summarizedgroupdata <- summarise(groupeddata, mean = round(mean(len),2), sd = round(sd(len),2), var = round(var(len),2))  
summarizedgroupdata

## Source: local data frame [6 x 5]  
## Groups: supp [?]  
##   
## supp dose mean sd var  
## <fctr> <dbl> <dbl> <dbl> <dbl>  
## 1 OJ 0.5 13.23 4.46 19.89  
## 2 OJ 1.0 22.70 3.91 15.30  
## 3 OJ 2.0 26.06 2.66 7.05  
## 4 VC 0.5 7.98 2.75 7.54  
## 5 VC 1.0 16.77 2.52 6.33  
## 6 VC 2.0 26.14 4.80 23.02

#Test if OJ is better than VC for each dose level  
  
##Statistical Tests  
#Ho = muOJ = muVC   
#Ha = muOJ != to muVC  
t1 <- t.test(len ~ supp, paired = FALSE, var.equal = TRUE, data = toothgrowthdata)  
t1$conf.int

## [1] -0.1670064 7.5670064  
## attr(,"conf.level")  
## [1] 0.95

t1$p.value

## [1] 0.06039337

#Since we didn't find that overall there was a significant differenct between delivery methods, I will check to each dose individually.  
#Subset data for the following hypothesis tests  
dose0.5 <- subset(toothgrowthdata, dose == 0.5)  
dose1.0 <- subset(toothgrowthdata, dose == 1.0)  
dose2.0 <- subset(toothgrowthdata, dose == 2.0)  
  
################################Check to see, for each dose level, if there is a difference between delivery methods  
#Ho muOJ0.5 = muVC0.5  
#Ha muOJ0.5 > muVC0.5  
t2 <- t.test(dose0.5[dose0.5$supp=="OJ",1], dose0.5[dose0.5$supp=="VC",1], lower.tail = FALSE, paired = FALSE,   
 var.equal = TRUE, data = dose0.5)  
t2$conf.int

## [1] 1.770262 8.729738  
## attr(,"conf.level")  
## [1] 0.95

t2$p.value

## [1] 0.005303661

#Check to see, for each dose level, if there is a difference between delivery methods  
#Ho muOJ1.0 = muVC1.0  
#Ha muOJ1.0 > muVC1.0  
t3 <- t.test(dose1.0[dose1.0$supp=="OJ",1], dose1.0[dose1.0$supp=="VC",1], lower.tail = FALSE, paired = FALSE,   
 var.equal = TRUE, data = dose1.0)  
t3$conf.int

## [1] 2.840692 9.019308  
## attr(,"conf.level")  
## [1] 0.95

t3$p.value

## [1] 0.0007807262

#Check to see, for each dose level, if there is a difference between delivery methods  
#Ho muOJ2.0 = muVC2.0  
#Ha muOJ2.0 > muVC2.0  
t4 <- t.test(dose2.0[dose2.0$supp=="OJ",1], dose2.0[dose2.0$supp=="VC",1], lower.tail = FALSE, paired = FALSE,   
 var.equal = TRUE, data = dose2.0)  
t4$conf.int

## [1] -3.722999 3.562999  
## attr(,"conf.level")  
## [1] 0.95

t4$p.value

## [1] 0.9637098

###########################Now, lets test to see if the increase is doses is significant for the OJ delivery method  
#Ho = mu0.5 = mu1.0  
#Ha = mu1.0 > mu0.5   
t5 <- t.test(toothgrowthdata[toothgrowthdata$dose==1.0,1], toothgrowthdata[toothgrowthdata$dose==0.5,1], lower.tail = FALSE, paired = FALSE,   
 var.equal = TRUE, data = toothgrowthdata)  
t5$conf.int

## [1] 6.276252 11.983748  
## attr(,"conf.level")  
## [1] 0.95

t5$p.value

## [1] 1.266297e-07

#Ho = mu1.0 = mu2.0  
#Ha = mu2.0 > mu0.5   
t6 <- t.test(toothgrowthdata[toothgrowthdata$dose==2.0,1], toothgrowthdata[toothgrowthdata$dose==1.0,1], lower.tail = FALSE, paired = FALSE,   
 var.equal = TRUE, data = toothgrowthdata)  
t6$conf.int

## [1] 3.735613 8.994387  
## attr(,"conf.level")  
## [1] 0.95

t6$p.value

## [1] 1.810829e-05