Statistical Inference: Peer Graded Assigment - Part 2

Wayne Heller

April 22, 2017

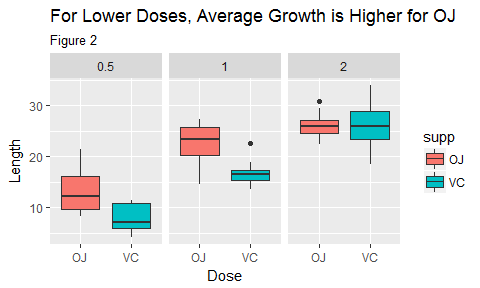
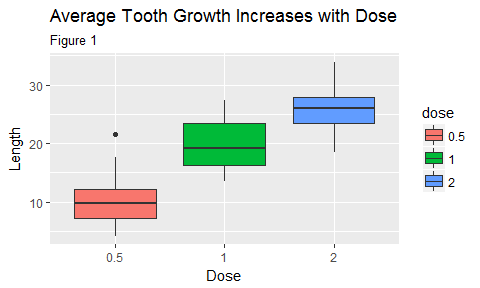
# OVERVIEW

*Part 2 Basic Inferential Data Analysis:* This part of the analysis uses the ToothGrowth data in the R datasets library and compares tooth growth across dosage and delivery methods. The specific instructions of this assignemnt are to "Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose"

## Description of the data

The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, orange juice coded as OJ or ascorbic acid a form of vitamin C and coded as VC.

# EXPLORATORY ANALYSIS



## Summary of Exploratory Analysis

1. The data indicate there maybe a strong relationship between an increase in tooth length and dose. Figure 1 on Left
2. The data indicate there maybe a difference in effectiveness between Orange Juice and Vitamin C supplements. Figure 2 on Right

# HYPOTHESIS TESTING

### Hypothesis Test 1: Over all dosage levels, with 95% confidence is Orange Juice a more effective delivery method than Vitamin C supplements

**Assumptions:** Tooth growth is normally distributed

**Approach:** Use Permutation testing to find percentage of results that produce a more extreme difference in averages than was measured

**Ho:** There is no significant difference in tooth growth between Orange Juice and Vitamin C

**Ha:** Average tooth growth is greater for Orange Juice

y <- ToothGrowth$len  
group <- ToothGrowth$supp  
testStat <- function(l, g) mean(l[g == "OJ"]) - mean(l[g == "VC"]) # test statistic difference of means between groups  
observedStat <- testStat(y, group) # test statistic for real observations  
permutations <- sapply(1:10000, function(i) testStat(y, sample(group))) # test statistics for permutations  
observedStat

## [1] 3.7

mean(permutations > observedStat) # percentage of permutations that have a more extreme result than was observed

## [1] 0.0282

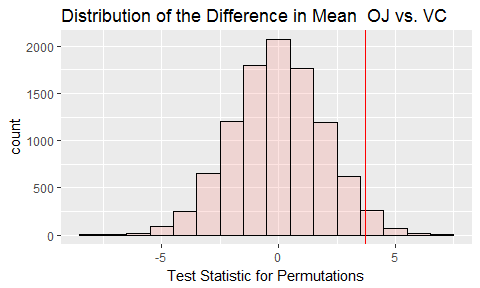


Figure 3: Permutations Analysis

##### Conclusion for Hypothesis 1

Based on the very low probability of finding a more extreme example, we should **REJECT the null hypothesis Ho:** There is no significant difference in tooth growth between Orange Juice and Vitamin C

### Hypothesis Test 2: At low dose (0.5), with 95% confidence is Orange Juice a more effective delivery method than Vitamin C supplements

**Assumptions:** Tooth growth is normally distributed

**Ho:** There is no difference in average growth

**Ha:** Average tooth growth is greater for Orange Juice

lenOJ <- ToothGrowth$len[ToothGrowth$dose == 0.5 & ToothGrowth$supp ==   
 "OJ"]  
lenVC <- ToothGrowth$len[ToothGrowth$dose == 0.5 & ToothGrowth$supp ==   
 "VC"]  
t.test(lenOJ, lenVC, alternative = "greater", paired = FALSE, var.equal = TRUE)$p.value

## [1] 0.002651831

##### Conclusion for Hypothesis 2

Based on the Student's T Test, we should **REJECT the null Hypothesis Ho:** There is no difference in average growth

### 3. At high dose (2.0), with 95% confidence is Orange Juice a more effective delivery method than Vitamin C supplements

**Assumptions:** Tooth growth is normally distributed

**Ho:** There is no difference in average growth

**Ha:** Average tooth growth is greater for Orange Juice

lenOJ <- ToothGrowth$len[ToothGrowth$dose == 2 & ToothGrowth$supp ==   
 "OJ"]  
lenVC <- ToothGrowth$len[ToothGrowth$dose == 2 & ToothGrowth$supp ==   
 "VC"]  
t.test(lenOJ, lenVC, alternative = "greater", paired = FALSE, var.equal = TRUE)$p.value

## [1] 0.5181451

##### Conclusion for Hypothesis 3

Based on the Student's T Test, we should **ACCEPT the null Hypothesis Ho:** There is no difference in average growth

# APPENDIX

### Code for Figure 1

createFig1

## function() {  
## g <- ggplot(data=ToothGrowth, aes(y=len, x=dose, fill=dose)) + geom\_boxplot()  
## g <- g + labs(title = "Average Tooth Growth Increases with Dose", x="Dose", y="Length", subtitle="Figure 1")  
## print(g)  
## }

### Code for Figure 2

createFig2

## function() {  
## g <- ggplot(data=ToothGrowth, aes(y=len, x=supp, fill=supp)) + geom\_boxplot()  
## g <- g + facet\_grid(. ~ dose)  
## g <- g + labs(title = "For Lower Doses, Average Growth is Higher for OJ", x="Dose", y="Length", subtitle="Figure 2")  
## print(g)  
## }

### Code for Figure 3

createFig3

## function() {  
## dat <- data.frame(x=permutations)  
## g <- ggplot(dat, aes(x = x, fill="red"))   
## g <- g + geom\_histogram(alpha = .20, binwidth=1, color = "black")   
## g <- g + geom\_vline(xintercept=observedStat, color='red') + theme(legend.position="none")  
## g <- g + labs(title="Distribution of the Difference in Mean OJ vs. VC", x="Test Statistic for Permutations")  
## print(g)  
## }