Identifying Impact of Vitamin C Treatments on Tooth Growth in Guinea Pigs

Author: Lowell Ricketts
Date: March 2016

Synopsis

Using hypothesis testing and associated confidence intervals, I find that delivering vitamin C supplements in the form of orange juice rather than ascorbic acid leads to greater length of odontoblasts (cells responsible for tooth growth) when the dosage levels of the supplement are less than 2 mg/day. At levels of 2 mg/day there was no significant difference in odontoblast length between the two groups.

Introduction

This project was completed during the Statistical Inference course offered by Coursera in the Data Science specialization tract. The course was taught by Brian Caffo from Johns Hopkins Bloomberg School of Public Health.

For this project I use the ToothGrowth dataset found in the R Datasets package. The data measures the response in 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 or ascorbic acid (a form of vitamin C.)

In the following analysis I begin with an exploratory look at the underlying data. Next, I conduct hypothesis tests to see if there is a statistically significant difference between mean tooth length for the two delivery methods.

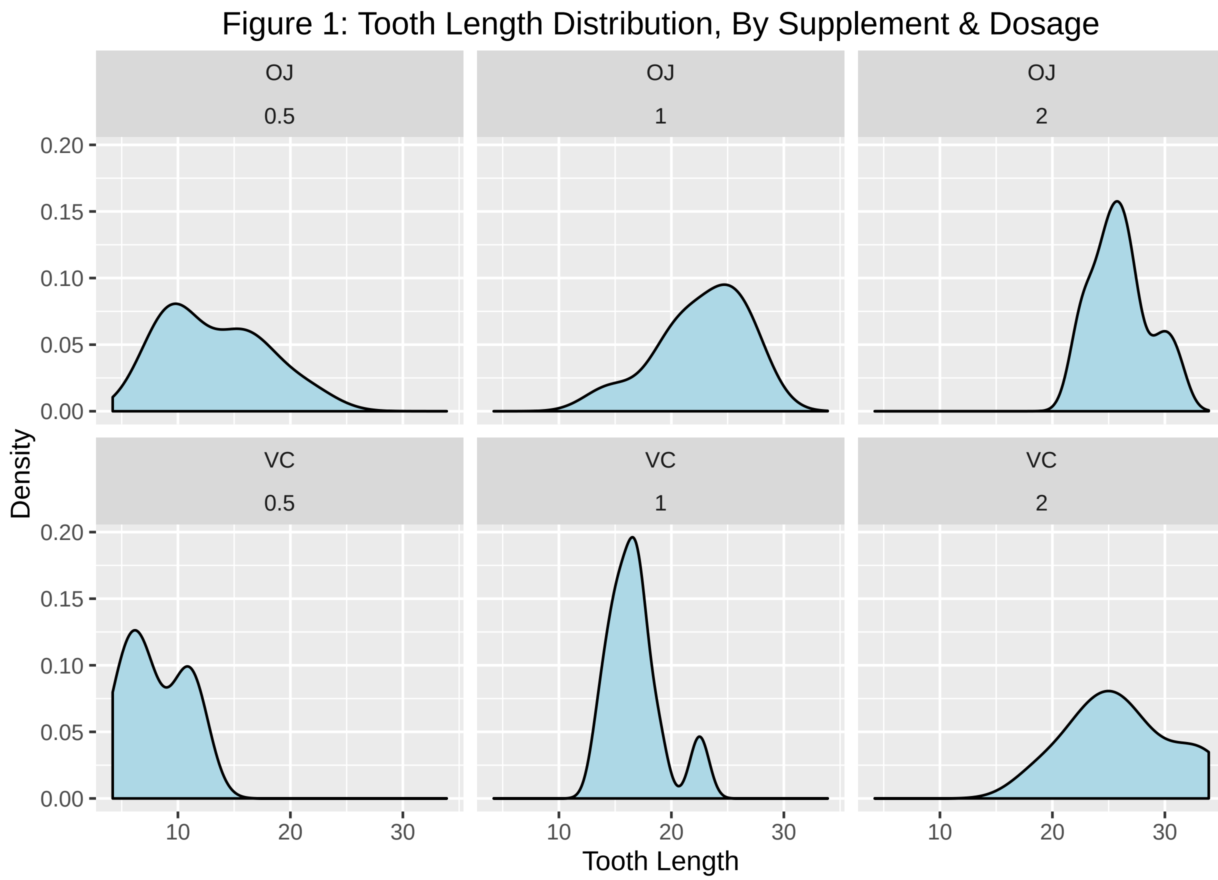
Please see the Appendix for all supporting R code.

Exploratory Analysis

Summary statistics for tooth length are reported in the table below by supplement type and dosage. Standard deviations suggest that tooth length is more widely dispersed for guinea pigs receiving OJ than VC at lower dosages, while VC has the highest variability at the highest dose. When comparing the two supplements by dosage, mean tooth length is greater for OJ across low dosages and about equal at the highest dose.

```
## Source: local data frame [6 x 6]
## Groups: supp [?]
##
##     supp  dose  mean       sd   min   max
##   (fctr) (dbl) (dbl)    (dbl) (dbl) (dbl)
## 1     OJ   0.5 13.23 4.459709   8.2  21.5
## 2     OJ   1.0 22.70 3.910953  14.5  27.3
## 3     OJ   2.0 26.06 2.655058  22.4  30.9
## 4     VC   0.5  7.98 2.746634   4.2  11.5
## 5     VC   1.0 16.77 2.515309  13.6  22.5
## 6     VC   2.0 26.14 4.797731  18.5  33.9
```

Next, Figure 1 shows the distribution of tooth length by supplement and dosage. The distributions look gaussian with a high variance. Given the limited sample size I assume that the t-distribution is best suited to draw statistical inferences.



Hypothesis Testing

For the t-tests I assume that the observations are not paired given that all 60 guinea pigs were only sampled once. I don't assume that the groups have equal variance, the results of the following t-tests vary little based on that assumption. I define my type I error rate as 5 percent and use 95 percent confidence intervals for the estimated difference in group means.

The first test is whether the mean tooth length of the group given orange juice differs from the mean tooth length of the group given ascorbic acid. For this initial t-test I use all observations, regardless of dosage.

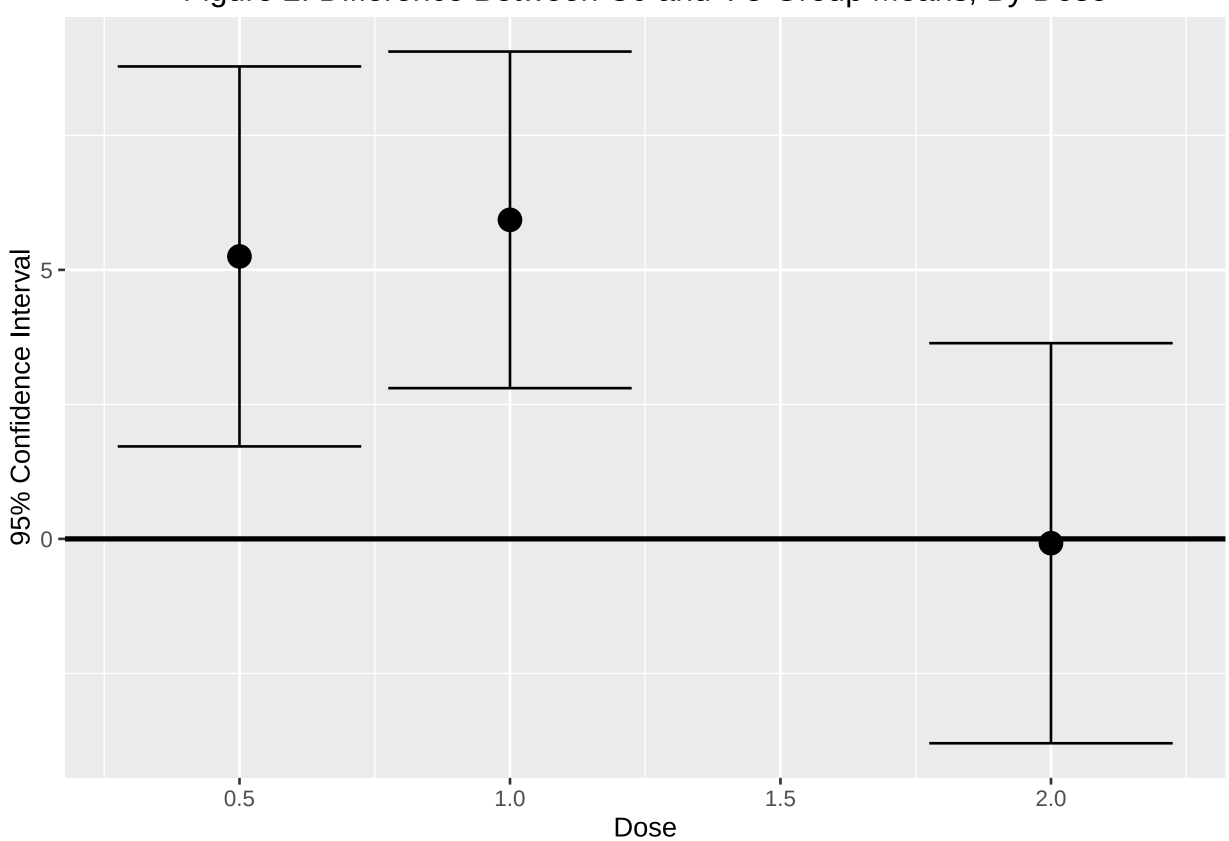
```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len by ToothGrowth$supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

The null hypothesis, that the difference in means is equal to zero, cannot be rejected at my designated significance level given that the p-value is greater than 0.05. This is also reflected in the confidence interval, zero is present on the lower end of the range. Certainly, if the confidence interval was narrowed by using a 10 percent type I error rate then we could say that OJ is more effective at effecting tooth growth. However, its important to stick to our significance thresholds rather than change them to tell a different story.

Next, I conduct a t-test for each of the three dosage levels to see if the delivery methods have different effectiveness based on dosage. Figure 2 shows the 95 percent confidence intervals for the difference in means across dosage levels. These confidence intervals were calculated using the same t-tests, see the Appendix for the exact R code used for the tests. The estimated difference in means from is shown in the center of the intervals.

The difference between the group means is positive and statistically significant for both the 0.5 and 1 mg/day dosage levels. Thus, we can conclude that orange juice leads to greater tooth growth when dosage levels are lower than 2 mg/day. The confidence interval for 2 mg/day shows that there is no significant difference in tooth growth between groups. In terms of our null hypothesis' of a zero difference in group means, we reject the null for 0.5 and 1 mg/day dosages and cannot reject the null at the 2 mg/day dosage.

Figure 2: Difference Between OJ and VC Group Means, By Dose



Conclusion

Orange juice is a more effective way to increase tooth length in guinea pigs when treatment dosages are lower than 2 mg/day. At dosages of 2 mg/day neither method (orange juice or ascorbic acid) was significantly more effective than the other.

Appendix

```r
library(datasets)
library(dplyr)
library(ggplot2)

# Load ToothGrowth data: The Effect of Vitamin C on Tooth Growth in Guinea Pigs
# Variable List:
# len = Tooth length
# supp = Supplement type (VC = ascorbic acid, OJ = orange juice)
# dose = Dose in milligrams/day
data(ToothGrowth)
head(ToothGrowth)

# Calculate summary statistics for tooth length, by supplement and dosage
summarize(group_by(ToothGrowth, supp, dose), mean = mean(len), sd = sd(len),
          min = min(len), max = max(len))

# Plot the distribution of tooth length, by supplement and dosage
# It appears that OJ has greater tooth lengths across lower dosages, and similar
# lengths at the highest dosage
ggplot(ToothGrowth, aes(len)) +
        geom_density(fill = "lightblue") +
        facet_wrap(~supp + dose) +
        ggtitle("Figure 1: Tooth Length Distribution, By Supplement & Dosage") +
        labs(x = "Tooth Length", y = "Density")

# Test difference in supplement type, was orange juice more effective than
# ascorbic acid? H0 = There was no difference in mean tooth length between VC
# and OJ.
mu_oj <- mean(ToothGrowth$len[ToothGrowth$supp == "OJ"])
mu_vc <- mean(ToothGrowth$len[ToothGrowth$supp == "VC"])
sd_oj <- sd(ToothGrowth$len[ToothGrowth$supp == "OJ"])
sd_vc <- sd(ToothGrowth$len[ToothGrowth$supp == "VC"])
n_oj <- length(ToothGrowth$len[ToothGrowth$supp == "OJ"])
n_vc <- length(ToothGrowth$len[ToothGrowth$supp == "VC"])

# Pooled variance estimate
sp <- sqrt( ((29 * (sd_oj ^ 2)) + 29 * (sd_vc ^ 2)) /
        (n_oj + n_vc -2))

# 95 % confidence interval for the difference in supplement group means
(mu_oj - mu_vc) + c(-1,1) * qt(.975, 58) * sp * (1/30 + 1/30) ^ .5

# Using a 90% confidence interval
(mu_oj - mu_vc) + c(-1,1) * qt(.95, 58) * sp * (1/30 + 1/30) ^ .5

t.test(ToothGrowth$len ~ ToothGrowth$supp, paired = FALSE)

# Does the difference between OJ and VC change by dose?
# Low Dose: Difference in mean highly significiant
lowdose <- filter(ToothGrowth, dose == 0.5)
t.test(lowdose$len ~ lowdose$supp, paired = FALSE)
```

```r
low <- t.test(lowdose$len ~ lowdose$supp, paired = FALSE)
low
c1 <- low$conf.int[1:2]
mu1 <- low$estimate[1] - low$estimate[2]

# Moderate Dose: Difference in mean highly significiant
meddose <- filter(ToothGrowth, dose == 1)
t.test(meddose$len ~ meddose$supp, paired = FALSE)
med <- t.test(meddose$len ~ meddose$supp, paired = FALSE)
med
ci2 <- med$conf.int[1:2]
mu2 <- med$estimate[1] - med$estimate[2]

# High Dose: Null is not rejected, CI is very wide
highdose <- filter(ToothGrowth, dose == 2)
t.test(highdose$len ~ highdose$supp, paired = FALSE)
high <- t.test(highdose$len ~ highdose$supp, paired = FALSE)
high
ci3 <- high$conf.int[1:2]
mu3 <- high$estimate[1] - high$estimate[2]

ci_intervals <- cbind(rbind(ci1, ci2, ci3),
                           dose = c(0.5, 1, 2),
                        mean = c(mu1, mu2, mu3))
ci_intervals <- data.frame(ci_intervals)
ci_intervals <- rename(ci_intervals, lower = V1, higher = V2)

g <- ggplot(ci_intervals, aes(x = dose, y = mean)) +
        geom_point(size = 4) +
        geom_errorbar(aes(ymax = higher, ymin = lower)) +
        geom_hline(yintercept = 0, size = 1) +
        labs(x = "Dose", y = "95% Confidence Interval") +
        ggtitle("Figure 2: Difference Between OJ and VC Group Means, By Dose")
g
```