The Effect of Vitamin C on Guinea Pig Tooth Length by Supplement Source and Dosage

Submission by Connor Lenio. Email: cojamalo@gmail.com. Completion Date: Apr 17, 2017.

Overview

The following report analyzes the ToothGrowth data set to determine the statistical significance of the results of an experiment involving guinea pig odontoblast length versus supplement type and dose per day of vitamin C. The analysis includes multiple hypothesis testing using both parametric and bootstrapping techniques. The p-values are adjusted to control the false positive rate and the statistical power of each test is calculated to consider the false negative rate. The majority of the dose/supplement combinations cause a significant difference in the true mean odontoblast length in the population of guinea pigs. For most of the treatment groups, increasing vitamin C dose causes an increase (positive difference) in mean odontoblast length and, orange juice causes a greater increase in mean odontoblast length than ascorbic acid.

Load Packages

```
library(pander); library(ggplot2); library(gridExtra);library(tidyr); library(dplyr); library(datasets)
```

Load and Tidy the Data

First, the ToothGrowth data from the data sets package is loaded:

```
# Source the data
data(ToothGrowth)
# Rename the columns for clarity
names(ToothGrowth) <- c("tooth_length", "supp_type", "dose_mg_day")
# Factor the dose variable
ToothGrowth$dose_mg_day <- factor(ToothGrowth$dose_mg_day)
# Take a glimpse of the data
glimpse(ToothGrowth)
```

```
## Observations: 60
## Variables: 3
## $ tooth_length <dbl> 4.2, 11.5, 7.3, 5.8, 6.4, 10.0, 11.2, 11.2, 5.2, ...
## $ supp_type    <fctr> VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, VC, ...
## $ dose_mg_day  <fctr> 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5, 0.5...
```

The data has three variables and sixty observations. From the ToothGrowth documentation, each observation represents a single guinea pig, its odontoblast length, `tooth_length`, what type of supplement it received, `supp_type`, and the dose size of vitamin C received per day, `dose_mg_day`. There are two levels of the `supp_type` variable, "OJ" (orange juice) or "VC"(ascorbic acid), and three levels of the ordinal `dose_mg_day` variable, 0.5 mg/day, 1mg/day, and 2 mg/day. These variables were renamed from their originals for clarification.

The main problem in this inferential analysis is to compare tooth growth by supp and dose. Thus, `tooth_length` is the continuous response variable in this analysis and the `dose_mg_day` variable will be treated as a categorical explanatory variable in addition to the categorical explanatory `supp_type` variable.

The basic structure of the analysis is a categorical variable (C) explaining a quantitative variable (Q) or:

$$C \rightarrow Q$$

Both `dose_mg_day` and `supp_type` are combined into a single categorical variable, `dose_supp`, which yields six groups of ten gerbils. Therefore, the inferential analysis will concern:

$$`dose\_supp` \rightarrow `tooth\_length`$$

```
# Create the dose_supp character variable and factor this variable
ToothGrowth <- ToothGrowth %>% mutate(dose_supp = paste(dose_mg_day, supp_type))
ToothGrowth$dose_supp <- factor(ToothGrowth$dose_supp)
summary(ToothGrowth) # View the resulting data frame
```

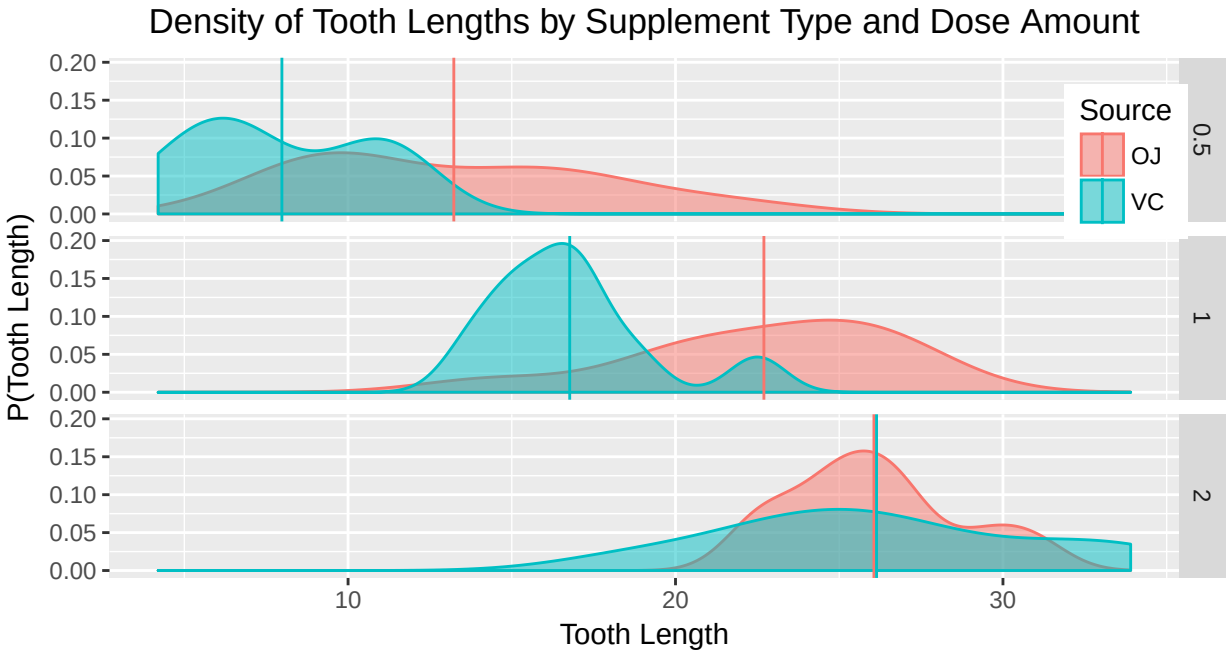
```
##   tooth_length   supp_type dose_mg_day  dose_supp
##  Min.   : 4.20   OJ:30     0.5:20      0.5 OJ:10
##  1st Qu.:13.07   VC:30     1  :20      0.5 VC:10
##  Median :19.25             2  :20      1 OJ  :10
##  Mean   :18.81                         1 VC  :10
##  3rd Qu.:25.27                         2 OJ  :10
##  Max.   :33.90                         2 VC  :10
```

With `dose_supp` included, the data is now ready for analysis.

Exploratory Data Analysis

The data is plotted to view the distribution of `tooth_length` by `supp_type` and `dose_mg_day`. This plot can help illustrate any associations in the data prior to determining statistical significance. The mean odontoblast length of each of the six groups is also plotted to compare odontoblast lengths.

```
# Plot the density curve for the sample populations split by dose amount and supplement type
vline.dat <- data.frame(dose_mg_day = levels(ToothGrowth$dose_mg_day),
                        v1 = c(mean(filter(ToothGrowth, dose_mg_day == "0.5" & supp_type == "OJ")$tooth
                        mean(filter(ToothGrowth, dose_mg_day == "1" & supp_type == "OJ")$tooth_length),
                        mean(filter(ToothGrowth, dose_mg_day == "2" & supp_type == "OJ")$tooth_length),
                        mean(filter(ToothGrowth, dose_mg_day == "0.5" & supp_type == "VC")$tooth_length
                        mean(filter(ToothGrowth, dose_mg_day == "1" & supp_type == "VC")$tooth_length),
                        mean(filter(ToothGrowth, dose_mg_day == "2" & supp_type == "VC")$tooth_length))
                                    colors = factor(c("OJ", "OJ","OJ","VC","VC","VC")))
ggplot(ToothGrowth, aes(tooth_length)) + geom_density(aes(fill = supp_type,color = supp_type),
                                                      alpha = 0.5) +
    facet_grid(dose_mg_day~.) +
    geom_vline(data = vline.dat, aes(xintercept = v1, color = colors)) +
    labs(title = "Density of Tooth Lengths by Supplement Type and Dose Amount",x="Tooth Length",
         y = "P(Tooth Length)",fill = "Source",  color = "Source") +
    theme(legend.position = c(0.95, 0.8), plot.title = element_text(hjust = 0.5))
```



A couple of interesting features stand out from this plot. First, the variance of each group is different with differing spreads of `tooth_length` for each group of ten guinea pigs. Moreover, some distributions are fairly normal in shape, while others appear multimodal. These differences in distribution are likely due to the small sample sizes of n = 10.

In terms of associations between groups, as `dose_mg_day` increases, so does the mean `tooth_length`. In addition, for the two smaller doses, the mean tooth growth of gerbils receiving OJ is greater than the gerbils receiving VC.

Assumptions for Inferential Analysis

The next step is to evaluate the data for the necessary conditions for a valid inferential analysis. A two-sample t-test will be used to determine statistical significance for the mean odontoblast length between the six different testing groups. Moreover, the observations are unpaired and the variances between the groups are unequal. The six groups of guinea pig will be compared to each other in fifteen hypothesis tests for differences between means for all of the possible combinations of doses and supplement types.

The first condition needed for a two-sample t-test is the data must represent random samples, or, specifically, independent, identically distributed (IID) variables. While the documentation for the data set does not give details on the experiment that generated the data, for the purpose of statistical inference, it is assumed that sixty guinea pigs were randomly selected for the experiment and that they were randomly assigned to the various testing groups with any possible confounding variables for odontoblast length controlled by the experimental design. In addition, the hypothetical population of all guinea pigs is at least 10-20 times larger than the sample size. These assumptions ensure that the guinea pigs represent IID samples.

The second condition needed for a two-sample t-test is the distribution of the response variable in both sample populations must be normal, or the samples should be large enough such that the Central Limit Theory is valid when applied to the sample (a common rule of thumb is that a sample > 30 is considered large enough). A table of the data arranged by each of the six groups can be used to check the size of the samples as well as to check the sample means and variances:

```
supp_dose_table <- ToothGrowth %>%
            tbl_df %>%
            group_by(dose_supp) %>%
```

```
            summarize(n = n(), Mean = round(mean(tooth_length),1),
                      Variance = round(var(tooth_length), 1))
pandoc.table(supp_dose_table)
```

dose_supp	n	Mean	Variance
0.5 OJ	10	13.2	19.9
0.5 VC	10	8.0	7.5
1 OJ	10	22.7	15.3
1 VC	10	16.8	6.3
2 OJ	10	26.1	7.0
2 VC	10	26.1	23.0

One may assume that the variables under consideration do vary normally under in a population of guinea pigs and simply proceed with the analysis. However, each sample has large differences in variance and the sample sizes are all less than 30. The chosen parametric two-sample t-test accounts for unequal variances and small sample size, but the tests rely on the assumption that the population distribution is normal when the sample size is small. Thus, this analysis considers a lack assumed normality of the population distribution by using bootstrapped t-distributions to double check the parametric t-test results. These extra calculations are important if any of the sample p-values are close to the 0.05 significance level or if it turns out the variable actually are not normally distributed in the population of guinea pigs.

The exploratory analysis did identify some positive associations between the explanatory and response variables, however, since no prior knowledge about vitamin C's interaction with guinea pig odontoblast length is assumed, the two sample t-tests will be two-sided. Thus, each hypothesis will all take the following form for the null and alternative hypotheses:

$$H_0 = \mu_1 - \mu_2 = 0$$
$$H_a = \mu_1 - \mu_2 \neq 0$$

Results of the Analysis

First, a general function, `two_samp_ht` is written that will return the p-value, confidence interval, and difference in means for the parametric t-test as well as the bootstrapped t-test p-value for a given combination of sample groups:

```
set.seed(123) # Set a seed for reproducibility of any simulated results

# two_Samp_ht function begins
two_samp_ht <- function(x, y, ..., pooled_xy, boot_reps = 50000) {
test <- t.test(x, y) # run the parametric two sample t.test

# Compute bootstrapped p-value for unequal variance
## Adjust both samples to the same mean to simulate null hypothesis
xt <- x - mean(x) + mean(pooled_xy)
yt <- y - mean(y) + mean(pooled_xy)

## This loop creates a bootstrap t distribution specifically for this null hypothesis
## by resampling from the mean-adjusted samples
boot.t <- c(1:boot_reps)
for (i in 1:boot_reps){
    sample.x <- sample(xt,replace=TRUE)
    sample.y <- sample(yt,replace=TRUE)
    boot.t[i] <- t.test(sample.x,sample.y)$statistic
```

```
}
# Compute and return all needed data
p.sample <- test$p.value
p.boot <- sum((abs(boot.t) >= abs(test$statistic))) / boot_reps
diff_means <- signif(-1*diff(test$estimate)[[1]], digits = 5)

return(data.frame(p_h0_sample = p.sample, conf_low = test$conf.int[1],conf_hi = test$conf.int[2],
                  p_h0_boot = p.boot, diff_means = diff_means))
     }
```

Using this function will simplify the code for the multiple hypothesis tests.

Multiple Hypothesis Test:

The following code runs the fifteen individual parametric and bootstrap hypothesis tests and aggregates the resulting data to a single table. The power and Tukey adjusted p-values for the samples are also calculated for each test. Tukey adjusted p-values are special adjusted p-values for multiple testing that take into consideration the variance of the data (see appendix A for a copy of the code used for the function `p.tukey`). The adjusted p-values are used to lower the false discovery rate inherent with running so many hypothesis tests at the same time.

```
# This function conducts a two-sample hypothesis test and bootstrap test using two_samp_ht
# and also calculates the power of the test
multi_ht_ind <- function(dose1, dose2, supp1, supp2) {
        x <- filter(ToothGrowth, dose_mg_day == dose1 & supp_type == supp1)
        y <- filter(ToothGrowth, dose_mg_day == dose2 & supp_type == supp2)
        xy <- filter(ToothGrowth, (dose_mg_day == dose1 & supp_type == supp1) |
                        (dose_mg_day == dose2 & supp_type == supp2))
        pvalues <- two_samp_ht(x$tooth_length, y$tooth_length, alternative = "two.sided", mu = 0,
                               paired = FALSE, var.equal = FALSE, boot_reps = 50000,
                               pooled_xy = xy$tooth_length)
        name <- paste0(dose1, "mg/day", supp1, "-", dose2, "mg/day", supp2,"=0")
        power <- power.t.test(n = nrow(x), delta = mean(x$tooth_length) - mean(y$tooth_length),
                              sd = sqrt((var(x$tooth_length)+var(y$tooth_length))/2),
                              sig.level = 0.05, alternative = "two.sided")$power
        data <- data.frame(name, pvalues[,1],pvalues[,2],pvalues[,3],pvalues[,4], pvalues[,5],power)
        names(data) <- c("h0", "p_h0_sample", "conf_low", "conf_hi","p_h0_boot", "diff_means","power")
        return(data)
}
# This function aggregates each individual test into a single data frame
multi_ht_all <-function(dose1, dose2, dose3, supp1, supp2) {
        data <- rbind(multi_ht_ind(dose1, dose1, supp1, supp2),
                      multi_ht_ind(dose1, dose2, supp1, supp2),
                      multi_ht_ind(dose1, dose3, supp1, supp2),
                      multi_ht_ind(dose1, dose2, supp2, supp1),
                      multi_ht_ind(dose2, dose2, supp1, supp2),
                      multi_ht_ind(dose2, dose3, supp1, supp2),
                      multi_ht_ind(dose1, dose3, supp2, supp1),
                      multi_ht_ind(dose2, dose3, supp2, supp1),
                      multi_ht_ind(dose3, dose3, supp1, supp2),
                      multi_ht_ind(dose1, dose2, supp1, supp1),
                      multi_ht_ind(dose1, dose3, supp1, supp1),
                      multi_ht_ind(dose2, dose3, supp1, supp1),
                      multi_ht_ind(dose1, dose2, supp2, supp2),
```

```
                        multi_ht_ind(dose1, dose3, supp2, supp2),
                        multi_ht_ind(dose2, dose3, supp2, supp2))
        return(data)
}
# All hypothesis tests are run, arranged by p-value, and then the Tukey p-values are calculated
multi_ht_results <- multi_ht_all("2", "1", "0.5", "VC","OJ") %>%
            arrange(desc(p_h0_sample)) %>%
            left_join(p.tukey(tooth_length, dose_supp, ToothGrowth)) %>%
            select(h0,p_h0_sample,p_h0_boot,adj_p_sample, power, conf_low, diff_means, conf_hi)
pandoc.table(multi_ht_results, split.table = Inf)
```

h0	p_h0_sample	p_h0_boot	adj_p_sample	power	conf_low	diff_means	conf_hi
2mg/dayVC-2mg/dayOJ=0	0.9639	0.9635	1	0.02767	-3.638	0.08	3.798
2mg/dayVC-1mg/dayOJ=0	0.09653	0.101	0.2936	0.3836	-0.6843	3.44	7.564
1mg/dayVC-0.5mg/dayOJ=0	0.04601	0.04868	0.264	0.5434	0.07189	3.54	7.008
2mg/dayOJ-1mg/dayOJ=0	0.0392	0.0447	0.3187	0.5663	0.1886	3.36	6.531
0.5mg/dayVC-0.5mg/dayOJ=0	0.006359	0.00988	0.02425	0.8501	-8.781	-5.25	-1.719
1mg/dayVC-1mg/dayOJ=0	0.001038	0.00408	0.007393	0.9678	-9.058	-5.93	-2.802
2mg/dayVC-1mg/dayVC=0	9.156e-05	1e-04	5.774e-06	0.9993	5.686	9.37	13.05
1mg/dayOJ-0.5mg/dayOJ=0	8.785e-05	0.00044	4.612e-06	0.9975	5.524	9.47	13.42
2mg/dayVC-0.5mg/dayOJ=0	7.196e-06	2e-05	1.77e-09	1	8.556	12.91	17.26
2mg/dayOJ-0.5mg/dayOJ=0	1.324e-06	2e-05	2.125e-09	1	9.325	12.83	16.34
1mg/dayVC-0.5mg/dayVC=0	6.811e-07	0	2.101e-05	1	6.314	8.79	11.27
2mg/dayOJ-1mg/dayVC=0	2.361e-07	0	6.908e-06	1	6.86	9.29	11.72
2mg/dayVC-0.5mg/dayVC=0	4.682e-08	0	4.822e-13	1	14.42	18.16	21.9
1mg/dayOJ-0.5mg/dayVC=0	3.655e-08	0	2.986e-11	1	11.52	14.72	17.92
2mg/dayOJ-0.5mg/dayVC=0	1.362e-11	0	4.855e-13	1	15.54	18.08	20.62

This table gives all the necessary information to draw inferential conclusions for the data. It is worth noting that there are no cases where the bootstrap p-value contradicts the parametric t-test p-values.

Conclusions

Referencing the adjusted p values, there is evidence at the 5% significance level that the majority (11 of 15) of the different treatments (combinations of OJ/VC and dosage levels) cause a significant difference in the true mean odontoblast length in the population of guinea pigs. The null hypotheses of no difference between

the mean odontoblast length were rejected for these studies. The confidence intervals give the estimate for the true mean difference for each test using the two-sample t-test. For example, 1 mg/day of vitamin C from VC causes an average odontoblast length difference in guinea pigs between -2.8 to -9.1 units versus those guinea pigs that receive 1 mg/day of vitamin C from OJ. Thus, the majority of the associations in the mean odontoblast length identified in the exploratory analysis are statistically significant. For most of the test treatments, increasing vitamin C dose causes an increase in odontoblast length and, OJ causes a greater increase in odontoblast length than VC.

However, there are four treatment combinations that could not be determined to show a significant difference in mean odontoblast length. The highest adjusted p-value was one and the null hypothesis, 2mg/dayVC-2mg/dayOJ=0, failed to be rejected. Thus, at a dose of two mg per day, it is possible that OJ versus VC did not cause a significant difference in odontoblast length. However, the power, or, the likelihood that this particular hypothesis test could detect an effect, was only 0.03, which is extremely low (if desiring a power of 0.80 or higher). Thus, a much larger sample size is needed to provide evidence to convincingly rule out the possibility of an effect of `supp_type` at this dose level.

The other three treatments that could not be determined to show a significant difference in mean odontoblast length all involved the comparison of different doses to each other. Interestingly, they all involved two doses that are near to each other such as 0.5 mg and 1 mg per day. The null hypotheses, 2mg/dayVC-1mg/dayOJ=0 and 1mg/dayVC-0.5mg/dayOJ=0 failed to be rejected, and may provide evidence that these sets of different doses of different supplement types actually have the same mean effect on odontoblast length (which could support the claim that OJ causes a greater increase in odontoblast length than VC). However, the power of these tests is low, so this conclusion is not acceptable from these samples alone. The final hypothesis, 2mg/dayOJ-1mg/dayOJ=0, compares two different doses of the same supplement. This result may provide evidence that, for OJ, increasing the dose from 1 to 2 mg does not cause a significant difference in mean odontoblast length. Again, larger sample sizes are needed to increase the power of these tests to clearly rule the possibilities of false negatives.

Appendix A: The function `p.tukey` Which Retrieves Tukey Adjusted P-values

```
p.tukey <- function(response, explanatory, data, conf.level = 0.95) {
    data$response <- eval(substitute(response), data)
    data$explanatory <- eval(substitute(explanatory), data)
    fit <- aov(response ~ explanatory, data)
    tukey <- TukeyHSD(fit, conf.level = conf.level)
    pvalues <- tukey[[1]][,c(1,4)] %>% tbl_df %>% arrange(desc(`p adj`))
    pvalues$diff <- signif(pvalues$diff, digits = 5)
    names(pvalues) <- c("diff_means", "adj_p_sample")
    return(pvalues)
}
```