Inferential Statistics - Tooth Growth Analysis

Joe Pechacek

July 22, 2015

Overview

The scope of this analysis is to determine if there are clear differences in the tooth growth of guinea pigs based on the amount and type of Vitamin C administered. The study measured the tooth growth for 10 guinea pigs after receiving doses of 0.5mg, 1.0mg and 2.0mg of Vitamin C either through Orange juice or Ascorbic acid.

Basic Data Exploration

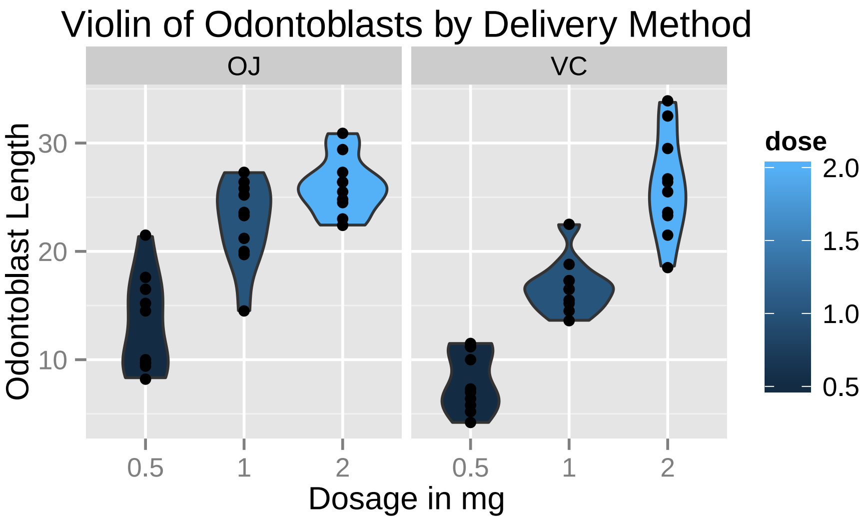
Each of the 10 guinea pigs received all 6 combinations of dose/method. This resulted in a data set with 60 observations. Calling `str()` on the data set provides a quick look at the data we will be working with.

```
## '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 ...
```

The *len* variable is the length of odontoblasts in each subject. There are also *supp* and *dose* variables that indicate the type of supplement and dosage for each observation. Sample values are shown in the output from `str()` above.
A further view of the data can be seen through a comparison of the mean tooth growth by supplement and dose as shown below. From this table we can see a potential increase in tooth growth based on dosage. There also appears to be a relationship where the Orange juice supplement results in greater growth overall.

```
## Source: local data frame [6 x 3]
## Groups: dose
##
##   dose supp  mean
## 1  0.5   OJ 13.23
## 2  0.5   VC  7.98
## 3  1.0   OJ 22.70
## 4  1.0   VC 16.77
## 5  2.0   OJ 26.06
## 6  2.0   VC 26.14
```

The relationship between the supplement type and dosage can be more clearly seen in the chart on the following page. This Violin Plot shows the relationship between the dosages for each supplement. Note that the shape of the Violin plot indicates the concentration of observations, wide areas have more observations while narrow areas have fewer observations. As an example, the 2.0mg dose for supplement Ascorbic acid (VC) is tall and narrow which is an indicator that the data has a large variance and few values are clustered together. Compare this to the 1.0mg dose of Ascorbic acid where we can see it is more concentrated - i.e. wider near the bottom.



The key observations to focus on with this plot is that the Ascorbic acid (VC) appears to have a stronger relationship to dosage while Orange juice appears to result in greater growth overall.
So the question becomes, can we say that the dosage of Vitamin C is a factor in tooth growth in guinea pigs? This will be addressed in the next section.

Testing the Hypothesis

The previous graph indicates that there could be a positive relationship between the dosage of Vitamin C and tooth growth, but can we be confident that this is true? To provide support for this observation, we can make use of Hypothesis testing and the P-value for various paired observations.
In order to run these comparisons, I split the data set into 6 smaller sets with the observations for a specific supplement and dosage in each. Specifics on the code used can be found in the Appendix.

The first Hypothesis test will check if the supplement type makes a difference with the same dosage. The Null Hypothesis would be "Is the mean growth the same between supplements for the same dosage." The P-values for each comparison between Ascorbic acid and Orange Juice are as listed below.

```
## $p.value.small.dose
## [1] 0.01547205
##
## $p.value.medium.dose
## [1] 0.008229248
##
## $p.value.large.dose
## [1] 0.9669567
```

These results indicate that for the small (0.5mg) and medium (1.0mg) dosages we would reject the Null Hypothesis and say that the supplement type does make a difference as the P-value is less than 0.05 for a 95% confindence. For the large (2.0mg) dosage we cannot reject the Null Hypothesis as the P-value is very large. Based on the fact that the large dose P-value is so large, we should accept the Null Hypothesis that supplement type does not support a relationship to tooth growth at larger doses.

The next comparison is the relationship between dosages for a given supplement and we will need to run two Hypothesis tests - one for each supplement. Looking at Orange juice first, the Null Hypothesis would be "Is the mean growth the same between dosages for the same supplement." The resulting P-values are shown on the following page.

```
## $p.value.sm.med.oj
## [1] 0.00243514
##
## $p.value.sm.lrg.oj
## [1] 3.724102e-05
##
## $p.value.med.lrg.oj
## [1] 0.08383912
```

As can be seen in the output above, the P-value for the small-medium and small-large comparison are much less than 0.05 and support rejecting the Null Hypothesis. The P-value for the medium-large comparison is on the edge. For a 95% confidence level, we could not reject the Null Hypothesis as the P-value is just slightly greater than 0.05. If we were to lower the confidence level to 90%, then we could reject the Null Hypothesis in all cases and state that the Vitamin C dosage through Orange juice does impact tooth growth.

Next we will look at the Ascorbic acid supplement. Following the same procedure and Null Hypothesis as used above, the following results indicate that we can reject the Null Hypothesis in all cases since the P-value is less than 0.05. This would indicate that there is a relationship between dosage and tooth growth for the Ascorbic acid supplement.

```
## $p.value.sm.med.vc
## [1] 0.0001715165
##
## $p.value.sm.lrg.vc
## [1] 4.264256e-06
##
## $p.value.med.lrg.vc
## [1] 0.0004647951
```

Conclusions

Based on the results of the three Hypothesis tests performed, I would conclude that there does appear to be a relationship between tooth growth and Vitamin C for dosages between 0.5mg and 1.0mg. As the dosage increases past 1.0mg there appears to be a reduction in effectiveness. The confidence levels shown in the P-values indicate that it is possible for the mean tooth growth at dosages between 1.0mg and 2.0mg to be the same at a 95% confidence level and therefore we cannot state that the admistration of Vitamin C alone is a factor in tooth growth at these levels. Ascorbic acid does appear to have a higher correlation to tooth growth, but given the fact that the supplement types resulted in differing degrees of tooth growth between the 0.5mg and 2.0mg levels and the lack of support that the supplement type makes a difference at the 2.0mg dosage, I would be reluctant to state there is a positive correlation to tooth growth using the supplement Ascorbic acid at higher dosages. Though not the purpose of this analysis, it would seem conceivable from the data that the supplement Ascorbic acid may be a less effective source of Vitamin C and therefore takes greater dosages to reach the same results as seen with Orange juice. Additional study and analysis would be needed to confirm or deny this theory which again is not the purpose of this analysis to delve into.

Appendix

Code segments to load data and subset into 6 groups. Note that the ToothGrowth data set is part of the standard R sample datasets included with base R:

```
data <- ToothGrowth

data.sm.vc <- data[data$dose == 0.5 & data$supp == "VC", 1]
data.med.vc <- data[data$dose == 1.0 & data$supp == "VC", 1]
data.lrg.vc <- data[data$dose == 2.0 & data$supp == "VC", 1]

data.sm.oj <- data[data$dose == 0.5 & data$supp == "OJ", 1]
data.med.oj <- data[data$dose == 1.0 & data$supp == "OJ", 1]
data.lrg.oj <- data[data$dose == 2.0 & data$supp == "OJ", 1]
```

Code segments to run P-value comparisons:

```
by_dose <- NULL
by_dose$p.value.small.dose <-
     t.test(data.sm.vc, data.sm.oj, paired = TRUE)$p.value
by_dose$p.value.medium.dose <-
     t.test(data.med.vc, data.med.oj, paired = TRUE)$p.value
by_dose$p.value.large.dose <-
     t.test(data.lrg.vc, data.lrg.oj, paired = TRUE)$p.value

by_supp_vc <- NULL
by_supp_vc$p.value.sm.med.vc <-
     t.test(data.sm.vc, data.med.vc, paired = TRUE)$p.value
by_supp_vc$p.value.sm.lrg.vc <-
     t.test(data.sm.vc, data.lrg.vc, paired = TRUE)$p.value
by_supp_vc$p.value.med.lrg.vc <-
     t.test(data.med.vc, data.lrg.vc, paired = TRUE)$p.value

by_supp_oj <- NULL
by_supp_oj$p.value.sm.med.oj <-
     t.test(data.sm.oj, data.med.oj, paired = TRUE)$p.value
by_supp_oj$p.value.sm.lrg.oj <-
     t.test(data.sm.oj, data.lrg.oj, paired = TRUE)$p.value
by_supp_oj$p.value.med.lrg.oj <-
     t.test(data.med.oj, data.lrg.oj, paired = TRUE)$p.value
```