

Statistical Inference Course Project

Author: Dale Hunscher

Creation Date: Tuesday, 24 April 2018

Part 2: Tooth Growth Data Set Analysis

Overview

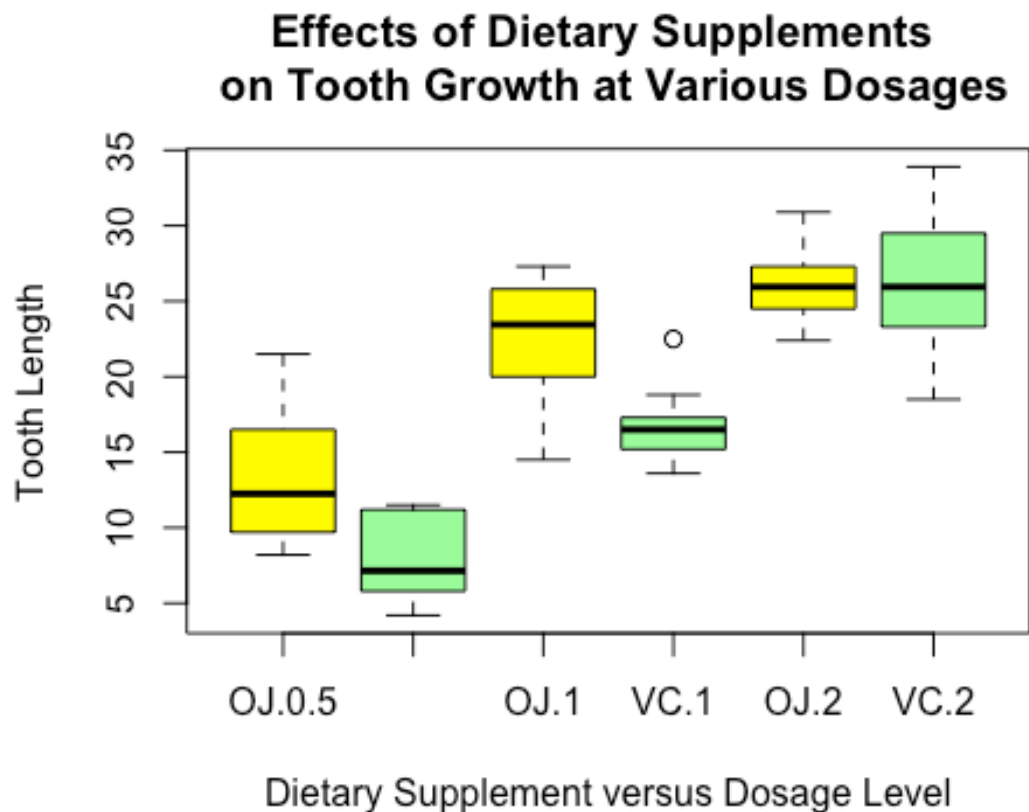
We will be investigating the ToothGrowth data set from the R datasets package. We'll start by getting familiar with the data set, using exploratory statistical techniques. Then we posit some hypotheses with respect to the effects of choice of dietary supplement and dosage level. Finally we'll summarize the conclusions we can draw from our investigation.

Exploratory Analysis

Let's look at a summary of the data set and see what we can learn from it.

##	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

We have 60 data points, 30 of which are associated with the dietary supplement factor "OJ" (ascorbic acid in orange juice), and the remaining 30 with "VC" (pure ascorbic acid). Each row contains a tooth length growth value (len column), a dietary supplement factor (supp column), and a dosage value (dose column).



From this chart, we see that orange juice (OJ) appears to be more effective at stimulating tooth growth at 0.5 and 1.0 mg/day. At the 2.0 dosage level, the effects are much closer.

Our goal is to provide information to support any decision-making with respect to supplement and dosage level. To do this, we will test some hypotheses against the data set.

Hypothesis Testing: Does Type of Supplement Matter?

We choose to use the Student's T-test statistic because the sample groups are small and the statistic is relatively robust with respect to the normalcy of the data. *However, we are NOT making an assumption that the variances are approximately equal.*

Our null hypothesis is that the effects of the two groups are the same at any given dosage level. Our alternative hypothesis is that the mean of the OJ group is significantly greater than the mean of the VC group with a confidence of 95%.

We will do three separate analyses to test the significance of the difference between the two treatments at each dosage level.

- Dosage = 0.5 mg/day

T-Test of the alternative hypothesis that OJ mean significantly greater than VC mean:

P-value is 0.0031793.

The P-value is less than 0.95, at 95% confidence level, so we can reject the null hypothesis. At a dose of 0.5 mg/day, ascorbic acid via orange juice is more effective than pure ascorbic acid alone.

- Dosage = 1.0 mg/day

T-Test of the alternative hypothesis that OJ mean significantly greater than VC mean.

P-value is 0.0005191879.

The P-value is less than 0.95, at 95% confidence level, so we can reject the null hypothesis. At a dose of 1.0 mg/day, ascorbic acid via orange juice is more effective than pure ascorbic acid alone.

- Dosage = 2.0 mg/day

T-Test of the alternative hypothesis that OJ mean significantly greater than VC mean:

The p-value is much greater than 0.05, so we can accept the null hypothesis, that the difference in means is insignificant. At a dose of 2.0 mg/day, ascorbic acid via orange juice is NOT more effective than pure ascorbic acid alone.

Hypothesis Testing: Does Dosage Matter?

Our analysis would be incomplete if we did not also look at the significance of the different dosage levels under each supplement. The length limitation imposed on this paper rules out including the results of such experiments. A separate Appendix document contains the results of these tests, and the code used in both parts of the project is also shown inline.

Conclusions Of Tooth Growth Analysis

- The higher the dosage the better in terms of stimulating tooth growth, regardless of supplement. For both supplements, the difference between each incrementally adjacent dosage pair (0.5-1.0 and 1.0-2.0 mg/day) is significant.
- At lower dosages (0.5 and 1.0 mg/day), the performance of orange juice is significantly better than pure ascorbic acid. The box plot showed that the difference is most pronounced at the 1.0 dosage level.
- At the highest dosage level tested, 2.0 mg/day, the performance of orange juice and pure ascorbic acid are roughly equivalent.

Appendix

The Rmd source code for my project documents can be found at https://github.com/DHunscher/Statistical_Inference/tree/master/project

To avoid tempting others to plagiarism, this repository will be removed from public view after course grades have been announced. The document ending in Appendix is an expanded version of both parts 1 and 2 combined, each part too large to be submitted for this course. This document exposes all code used.