An Analysis of the ToothGrowth Data Set: Comparing the Means and Intervals for Significance

S Duffy

March 9, 2016

Overview

in 1946, E.W. Crampton was exploring the potency of vitamin C in various food sources for the Canadian government. This was to ensure the most effective delivery of this vitamin to the armed forces during times of war. A portion of the resulting data is stored in the R dataset 'ToothGrowth'. This study will test the effectiveness of two sources of vitamin C, orange juice and a vitamn C pill in the form of ascorbic acid.

Data Exploration

In this section we will explore the ToothGrowth data set from the R library. First, loading the data and then some summarization.

The data set comes from research published in the paper "The Growth of the Odontoblast of the Incisor Teeth as a Criterion of Vitamin C Intake of the Guinea Pig." By E. W. Crampton as found in The Journal of Nutrition. Excellent reading for those with insomnia.

First load the data set and take a peak at it.

```
##load the datasets library and give the data set a short alias
library(datasets)
tg <- ToothGrowth

##dimensions of the dataset
a <- dim(tg)
```

Our data set ToothGrowth has dimensions of 60 rows by 3 columns. The structure of the data is thus:

```
##structure of the dataset
str(tg)

## '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 beginning and end of the dataset looks like this.

```
head(tg)
```

```
##     len supp dose
## 1   4.2   VC  0.5
## 2  11.5   VC  0.5
## 3   7.3   VC  0.5
## 4   5.8   VC  0.5
## 5   6.4   VC  0.5
## 6  10.0   VC  0.5
```

```
tail(tg)
```

```
##      len supp dose
## 55 24.8   OJ    2
## 56 30.9   OJ    2
## 57 26.4   OJ    2
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

Let it be noted that len is the Length of the tooth at measurement, supp is the Supplement used where VC = Vitamin C and OJ = Orange Juice, and dose is, as expected, the dose of the supplement provided in milligrams.

The following plots the data with trendlines to get a sense of what it looks like

```
library(ggplot2)
tgplot <-ggplot(tg, aes(x=dose, y=len, colour=supp)) +
          geom_point(alpha=.6) +
          geom_smooth(alpha=.2, size=1) +
          ggtitle("Tooth Growth")

tgplot
```
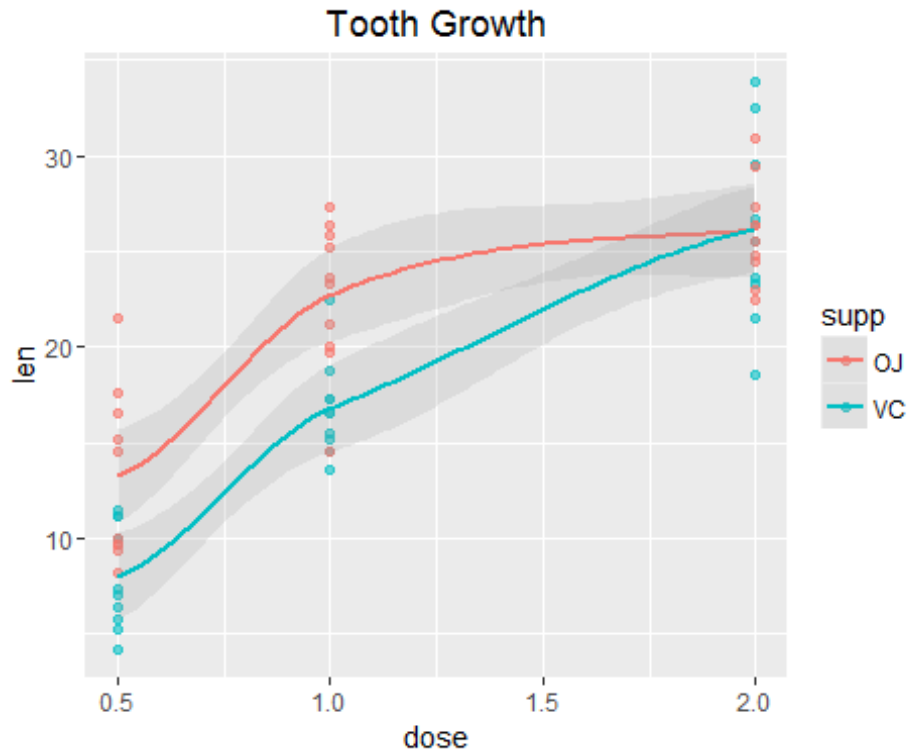
Tooth Growth

 The plot shows two distinct trendlines for each of the supplements, with orange juice at higher levels of tooth growth for lower dosages. But is the difference significant?

Key Assumptions

This study assumes that orange juice and vitamin C are equally efficient at enabling tooth growth in guinea pigs. This will be the basis for hypothesis tests. The study will be conducted by supplement type and by dosage and will assume equal variances

Hypothesis Tests

The null hypothesis for this test will be that each type of supplement contributes an equal amount of growth at each dosage level. Thus, there will be three tests run, one for each dosage level (0.5, 1.0, 2.0) and for each the null hypothesis will state that there is no difference in the average tooth growth between orange juice and vitamin C. These will be two-sided tests assuming equal variance.

Test 1

Dosage level 0.5, H_o: mu_a = mu_b, alpha = 0.05

The following code will subset the data and conduct the tests using a two-sided Student's t test.

```
tghalf <- subset(tg,dose == .5)
test1 <- t.test(subset(tghalf,supp == 'VC')[,1],subset(tghalf,supp ==
'OJ')[,1],alt="two.sided",var.equal = TRUE)
```

```
pvt1 <- test1[[3]]
cit1 <- as.data.frame(test1[[4]])
```

Test 2

Dosage level 1.0, H_o: mu_a = mu_b, alpha = 0.05

```
tgone <- subset(tg,dose == 1.0)
test2 <- t.test(subset(tgone,supp == 'VC')[,1],subset(tgone,supp ==
'OJ')[,1],alt="two.sided",var.equal = TRUE)
pvt2 <- test2[[3]]
cit2 <- as.data.frame(test2[[4]])
```

Test 3

Dosage level 2.0, H_o: mu_a = mu_b, alpha = 0.05

```
tgtwo <- subset(tg,dose == 2.0)
test3 <- t.test(subset(tgtwo,supp == 'VC')[,1],subset(tgtwo,supp ==
'OJ')[,1], alt="two.sided",var.equal = TRUE)
pvt3 <- test3[[3]]
cit3 <- as.data.frame(test3[[4]])
```

Results

Test 1

In the first test, with a dosage equal to 0.5mg, the test returns a p-value of 0.0053037 and a confidence interval of -8.7297383 to -1.7702617. Because the p-value is less than our alpha of 0.05 and the hypothesized difference in means of 0 does not lie within the confidence interval, the null hypothesis is rejected for Test 1.

Test 2

In the second test, with a dosage equal to 1.0mg, the test returns a p-value of 7.807261710^{-4} and a confidence interval of -9.0193081 to -2.8406919. Because the p-value is less than our alpha of 0.05 and the hypothesized difference in means of 0 does not lie within the confidence interval, the null hypothesis is rejected for Test 2.

Test 3

In the third test, with a dosage equal to 2.0mg, the test returns a p-value of 0.9637098 and a confidence interval of -3.5629985 to 3.7229985. Because the p-value is greater than our alpha of 0.05 and the hypothesized difference in means of 0 lies within the confidence interval, the null hypothesis is not rejected for Test 3.

Conclusion

Considering the given data, only when the dosage of vitamin C and orange juice reaches two milligrams is it likely that the results of tooth growth in guinea pigs are equal to each

other. For dosages less than two milligrams (i.e., 0.5 and 1.0) there is a significant difference in the mean values that indicates the two supplements do not provide equal growth rate for guinea pig teeth. From the test results, and given the earlier view of the data, one can assume (somewhat unscientifically) that orange juice provides better tooth growth at lower dosages.