Pt2_BasicInfDataAnalysis

Alex Fleming

1/29/2018

Basic Analysis of the Tooth Growth Dataset

by Alex Fleming - student in the JHU Data Science Specialization

Overview

The basic dataset used is from the base R documentation.

Dataset: The Effect of Vitamin C on Tooth Growth in Guinea Pigs. The response is 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 and coded as VC).

The three variables: - len: numeric tooth length change (one observation per subject) - supp: factor of the type of supplement (30 of each of 2 types) - dose: numeric dose in mg/day (20 of each of 3 dose levels)

This analysis will be broken down into 3 hypothesis driven stages: 1) Can we reject the null hypothesis that vitamin C has NO impact on tooth length? We will test this using a students t driven hypothesis test. 2) Is there a significant difference in the tooth change based on type of supplement? We will attempt to divide two groups and prove they have significantly different distributions. 3) Does the level of dose have an impact on the magnitude of tooth length change? We will attempt to judge which of the doses cause changes tooth lengh outside the original confidence interval.

Part 1: Rejecting the Null

If we being with a null hypothesis that vitamin C has no impact on growth, then we would expect the mean of our length change to be zero. We will use the 95% students t-test confidence interval.

```
knitr::opts_chunk$set(echo = TRUE)
data("ToothGrowth")
n <- nrow(ToothGrowth)
mu0 <- 0
mu <- mean(ToothGrowth$len)
s <- sd(ToothGrowth$len)
test <- t.test(ToothGrowth$len)
print(test)
```

```
##
##   One Sample t-test
##
## data:  ToothGrowth$len
## t = 19.051, df = 59, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  16.83731 20.78936
## sample estimates:
## mean of x
##  18.81333
```

Since this t statistic is large and the 95% confidence interval does not contain zero, we can reject the null hypothesis. The vitamin C clearly has some statistically significant impact.

Part 2: Supplement Population Effect

We now will attempt to test the difference between the two supplement outputs to determine if there is a significant difference between orange juice or a vitamin C pill as a delivery mechanism,

```
test2 <- t.test(len ~ supp, paired=FALSE, var.equal=TRUE, data=ToothGrowth)
print(test2)
```

```
##
##   Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1670064  7.5670064
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

In this case, since the 95% CI does contain zero, and the p-value is greater than 0.05, we cannot reject the null hypothesis that there is no difference between supplement delivery methods.

Part 3: Dosage Impact

We can start by attempting to visualize the data across the dose groups.

```
plot(ToothGrowth$len ~ ToothGrowth$dose, main="Length Difference by Dose Level", xlab="D
ose Level in mg/day", ylab="Change in Length", pch=18)
```

Length Difference by Dose Level
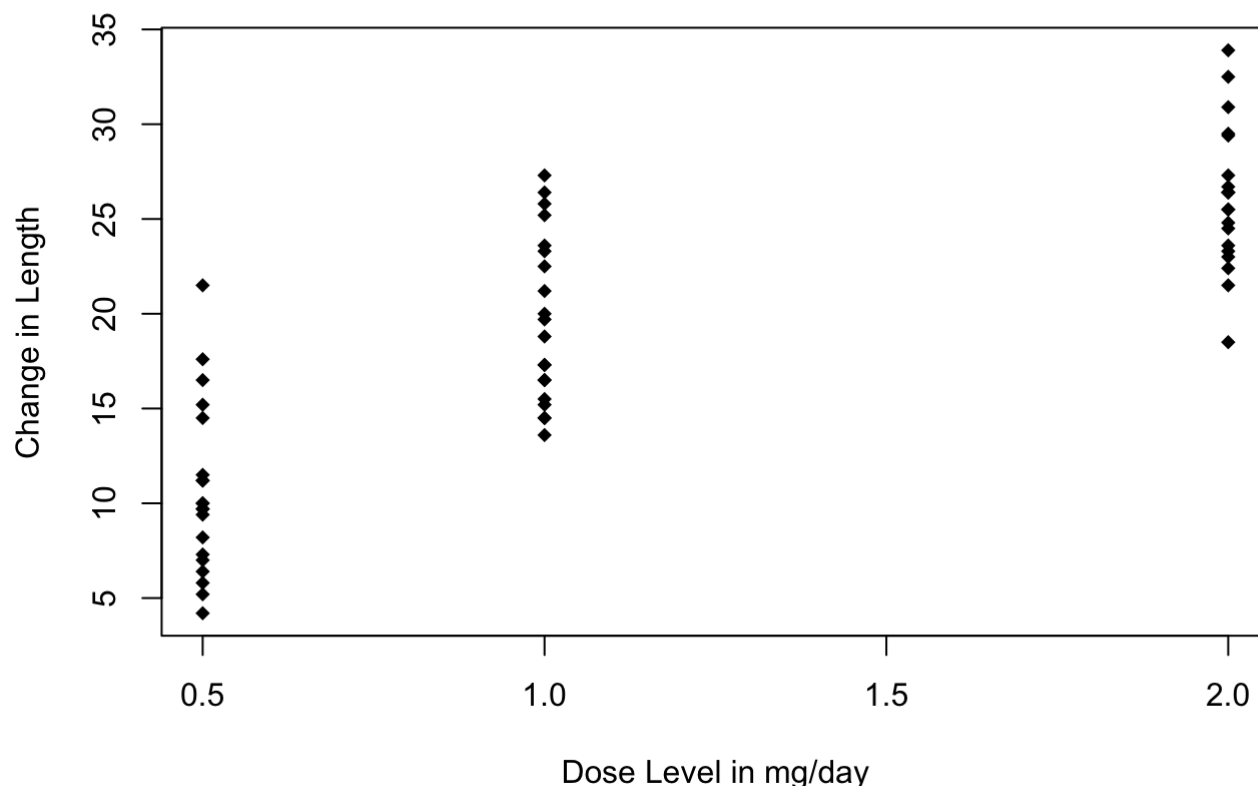
We see, qualitatively, from this plot that it appears there is a higher impact for a higher dose. One way to demonstrate this would be to see if there is an overlap in the 95% confidence intervals between the lowest dose level (0.5mg/day) and the highest dose level per day (2mg/day)

```
# Divide the dataset into low and high dose sets
ToothGrowthlow <- subset(ToothGrowth, ToothGrowth$dose == 0.5)
ToothGrowthhigh <- subset(ToothGrowth, ToothGrowth$dose == 2)
testlow <- t.test(ToothGrowthlow$len)
testhigh <- t.test(ToothGrowthhigh$len)
print(testhigh$conf.int)
```

```
## [1] 24.33364 27.86636
## attr(,"conf.level")
## [1] 0.95
```

```
print(testlow$conf.int)
```

```
## [1]  8.499046 12.710954
## attr(,"conf.level")
## [1] 0.95
```

Since there is no overlap in the confidence intervals for the high and low doses, there is a significant difference in the impact of the dose on the tooth length. Other methods such as regression or more detailed subsetting could be used to demonstrate the magnitude of this impact, but that is outside of our mandate.

Summary Output and Conclusions

The tooth growth data is a relatively small sample of the vitamin C impact on a specific animal species, and we do not know a great deal about the conditions of the data gathering. We can draw a few basic conclusions from a preliminary statistical analysis:

- There is a statistically significant increase in tooth size due to vitamin C doses administered.
- The method of delivery of the vitamin C does not appear to have a significant difference.
- The dosage level does have a significant impact on the size of growth, with larger doses resulting in significantly larger length impact.

These statements are applicable to this population and data only, and would be more impactful if they could be tested in a larger study with more situations and background data.

Contact: alex.c.fleming@gmail.com (mailto:alex.c.fleming@gmail.com)

Data Source

C. I. Bliss (1952) The Statistics of Bioassay. Academic Press.

References

McNeil, D. R. (1977) Interactive Data Analysis. New York: Wiley. Crampton, E. W. (1947) The growth of the odontoblast of the incisor teeth as a criterion of vitamin C intake of the guinea pig. The Journal of Nutrition 33(5): 491–504. http://jn.nutrition.org/content/33/5/491.full.pdf (http://jn.nutrition.org/content/33/5/491.full.pdf)