Analysis of a Guinea Pig Tooth Growth Experiment

Charles Wylie - August 27, 2016

Statistical Inference Course Project 2

Synopsis

In this project we investigate an experiment on the effect of vitamin C on tooth growth in guinea pigs, reported in *Crampton (1947), The Journal of Nutrition* [1]. The data from this experiment is contained in the *ToothGrowth* dataset in the R *datasets* Library. The purpose of this exploratory data analysis is to assess whether the data meets the required assumptions for hypothesis testing with t-tests, confidence intervals, or questions such as "is the data normally distributed?" [6][7] We then use hypothesis tests to compare tooth growth by supplement and dose.

R Libraries and Data

```r
library(datasets)
library(pastecs)
data(ToothGrowth)
```

Summary of the ToothGrowth Dataset

The response is the length of odontoblasts (cells responsible for tooth growth) measured in microns, in 60 guinea pigs, each receiving one of three dose levels of Vitamin C (0.5, 1.0, and 2.0 mg/day) with one of two delivery methods (orange juice or an aqueous solution of ascorbic acid). No guinea pig received a dose of zero as they would acquire scurvy at that dose [4].

```r
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
str(ToothGrowth)
```

```
## '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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

Key Assumptions

We must assess whether the data meets the required assumptions for hypothesis testing with t-tests or confidence intervals. The Student's t-test assumes that the data represent independent samples with equal variances from a normal population distribution. Knowing the distribution of the test statistic under the null hypothesis allows for accurate calculation of p-values [5].

We assume that the 60 guinea pigs tested in this experiment represent a random sample of the same normally distributed population, therefore any division of the group will give samples with equal variances. Also, our test groups are of equal sample size: 30 OJ and 30 VC supplements, and 10 each for the three dose levels.

Our test data meets the assumptions for the Student's t-test, therefore we will use it to test our hypotheses.

Hypotheses

We can ask two questions of the ToothGrowth data: "Does the delivery method affect the response?", and "Does the dose level affect the response?" We can answer each question by testing a null hypothesis with respect to an alternate hypothesis.

Our null hypothesis for Question 1 is "the response to the two delivery methods is the same", and the alternate hypothesis is "the response to delivery methods is not the same." We can test this with a two-tailed t-test.

Similarly, the null hypothesis for Question 2 is "the response to a higher dose is the same as to a lower dose", and the alternate hypothesis is "the response to a higher dose is greater than to a lower dose." A one-tailed test can be used.

Hypothesis Test - Question 1

We test the null hypothesis "the response to the two delivery methods is the same."

```r
t.test(len ~ supp, paired = FALSE, var.equal = TRUE, conf.level = 0.975,
       data = ToothGrowth)
```

```
##
##  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
## 97.5 percent confidence interval:
##  -0.7453352  8.1453352
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

We see that the 97.5% confidence interval contains 0 (barely), so we accept the null hypothesis and reject the alternate. We can say that there is no significant difference in delivery methods.

Hypothesis Test - Question 2

We test whether the response is the same for the highest and lowest dose.

```r
ToothGrowth2 <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5,2.0))
t.test(len ~ dose, paired = FALSE, var.equal = TRUE, conf.level = 0.95,
       data = ToothGrowth2)
```

```
##
##  Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 38, p-value = 2.838e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15352 -12.83648
```

```
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

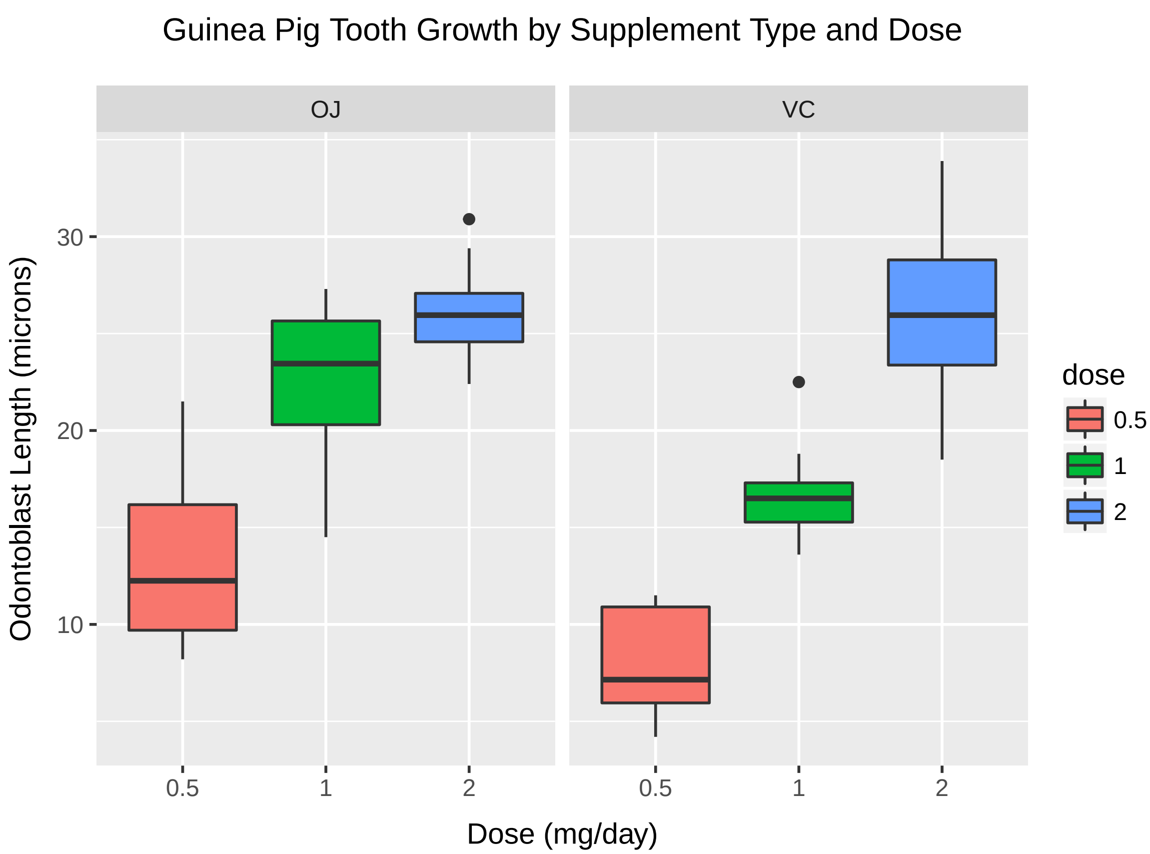
The hypothesis test for Question 2 shows that the 95% confidence interval does not contain 0, therefore we reject the null hypothesis and say that the response to a higher dose is greater than to a lower dose.

Results

We see from the box plot below that the response to the orange juice supplement is somewhat greater than to the ascorbic acid solution at lower doses, but at the highest dose the response is about the same. Overall, the two supplements have about the same response level, as the hypothesis test for Question 1 showed.

We see that the response to the dose level is high for both supplements, with orange juice giving a higher response at lower doses. The R code for Figure 1 is shown in the appendix.

Figure 1



Confounding Factors

The orange juice supplement may contain additional nutrients that contribute to odontoblast growth that are not found in the ascorbic acid supplement.

Conclusion

We have evaluated the data and tested two hypotheses using the given data. We have shown that the data does meet the requirements for hypothesis testing. We have also shown that there is a positive response to dose level, but no significant response to delivery method.

Appendix

References

1. *Crampton, E. W.* (1947) The Growth of the Odontoblasts of the Incisor Tooth as a Criterion of the Vitamin C Intake of the Guinea Pig: Five Figures. The Journal of Nutrition, Volume 33, Issue 5, 491-504.

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

3. *McNeil, D. R.* (1977) Interactive Data Analysis. New York: Wiley.

4. https://bugs.r-project.org/bugzilla3/show_bug.cgi?id=15953

5. http://www.mathworks.com/help/stats/hypothesis-test-assumptions.html?s_tid=gn_loc_drop

6. https://github.com/lgreski/datasciencectacontent/blob/master/markdown/edaInToothGrowthAnalysis.md

7. https://www.coursera.org/learn/statistical-inference/peer/3k8j5/statistical-inference-course-project

8. https://www.coursera.org/learn/statistical-inference/discussions/weeks/4

R code for Figure 1

```r
library(ggplot2)

fig1 <- ggplot(aes(x = dose, y = len), data = ToothGrowth) +
        geom_boxplot(aes(fill = dose)) +
        xlab("Dose (mg/day)") +
        ylab("Odontoblast Length (microns)") +
        facet_grid(~ supp) +
        ggtitle("Guinea Pig Tooth Growth by Supplement Type and Dose") +
        theme(plot.title = element_text(margin = margin(b = 0.6, unit = 'cm'))) +
        theme(axis.title.x = element_text(size = 12, margin = margin(10,0,0,0))) +
        theme(axis.title.y = element_text(size = 12, margin = margin(0,10,0,0))) +
        theme(plot.margin = unit(c(0, 0.5, 0.5, 0.5), 'cm')) # top, right, bottom, left

print(fig1)
```