Course Project Part II: Inferential Data Analysis

Matthew Dunne

August 19, 2017

Overview

This a basic inferential analysis on the Tooth Growth data that comes with R Studio. Here we will examine whether there is a difference in tooth growth by supplement and by dosage.

Summary of Data

This data consists of 60 records, each consisting of a numeric value of length in tooth growth, a type of supplement (either OJ or VC), and a dosage for that supplement (0.5, 1.0, or 2.0). There are 30 records for each supplement, which are further divided into 10 records for each dosage within that supplement.

Is there a difference in tooth growth between supplements?

To answer this question we set our Null Hypothesis: length in tooth growth using VC = length in tooth growth using OJ. Assuming we want the probability of rejecting under this Null Hypothesis to be 5%, we split it equally between the upper tail and the lower tail a straightforward t.test function in R tells us that the difference in means is:

```
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

or that the OJ group had a mean growth of about 3.7 more than the VC group.

But this does not necessarily mean we reject the null hypothesis. We only reject the null hypothesis if the observed difference in means is so extreme as to fall within the 5% probability of rejection.

There are several ways we can tell whether this observed difference is so extreme as to be rejected. The first is to look at its confidence interval. If that confidence interval includes 0, our analysis tells us that the difference could still be 0 and we would fail to reject the null hypothesis.

The lower bound for our confidence interval is:

```
## [1] -0.1710156
```

meaning the confidence interval includes 0. Therefore, we fail to reject the null hypothesis.

Is there a difference in growth by dosage?

Here we will split our records into two groups based on supplement type and within those groups do the same t-test as above based on dosage. Because the t-test function only allows us to compare two factors at once, we shall compare the relevant pairs: 0.5 vs. 1.0, 1.0 vs. 2.0, and 0.5 vs 2.0. This means we will run the t-test on six different supplement, dosage combinations.

In all cases, we see larger growth with the larger dose. For example, the mean increase in growth from am OJ dose of 1.0 vs 0.5 is:

```
dosediff<-t.test(len~dose, paired=FALSE, data=ojhalfv1)$estimate
dosediff<-dosediff[2]-dosediff[1]
names(dosediff)<-NULL
dosediff
```

```
## [1] 9.47
```

But what are the chances of getting differences in growth as extreme as those observed? Put another way, are the p-values above 5%, meaning we would fail to reject the null hypotheses?

The clear answer is no. In all cases and combinations, the p-values are less than 5%. Therefore, in all cases we reject the null hypothesis and say there is a difference in growth between dosages.

Conclusion

There is no difference in growth by supplement. There is a difference in growth by dosage.