Tooth Growth Analysis and Hypothesis Tests

Ash Chakraborty

Wednesday, June 17, 2015

PART 2: Analyzing Tooth Growth Data

We are going to load the ToothGrowth dataset that's part of the "datasets" package in R. We run a basic analysis on this dataset, in addition to verifying assumptions around confidence intervals to extract any correlations between supplement dosage and tooth length.

We see that the dataset is setup as follows:

```
##       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
```

Each observation in this dataset seems to record the length of the tooth when administered with a certain dosage of supplement OJ or VC. Doses are discrete levels of 0.5, 1.0, 1.5, and 2.0 units. The mean length of tooth growth when OJ is administered is 20.66 units, and when VC is administered the mean length is 16.96 units.

Now, let's get an idea of the relationship between length and dosage per supplement:

Scatter: Tooth Growth Length Vs. Dose
of Supplements OJ and VC
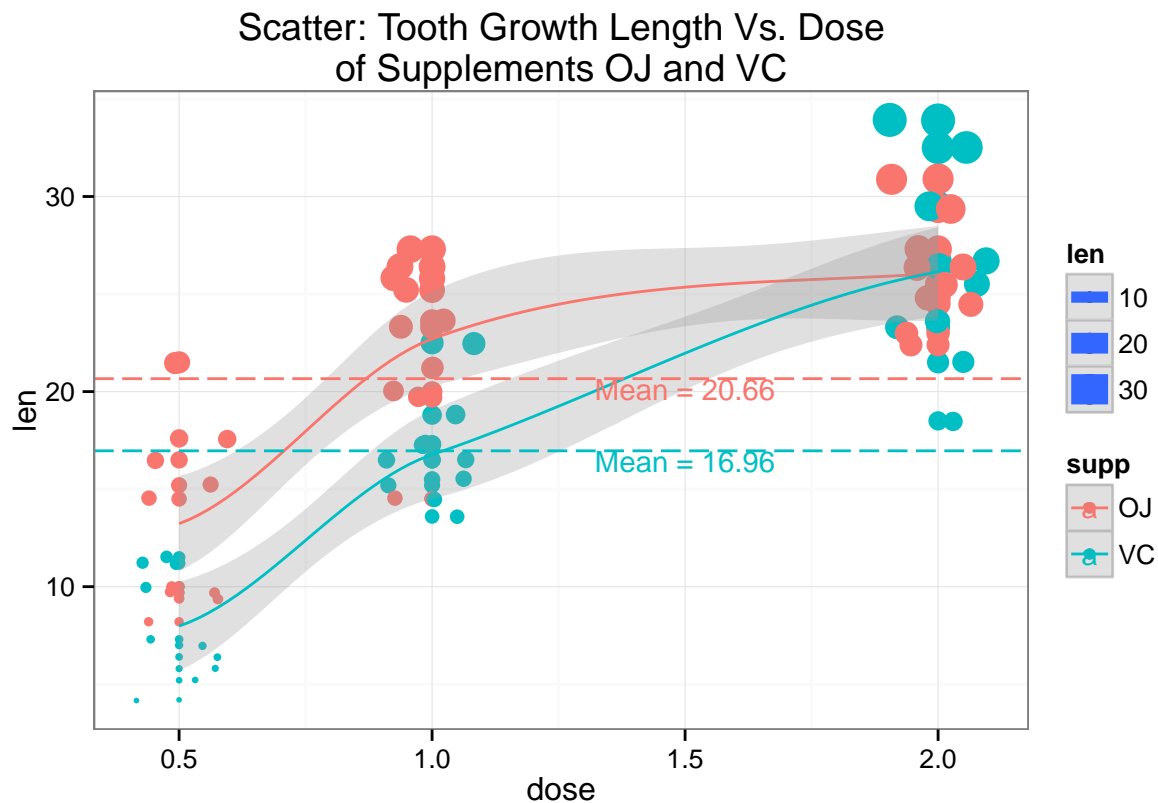
The plot *seems to* suggest the following:

- There is a correlation between length and dosage amount 0.5 and 1.0 for both supplements.
- There is **no data** for either supplement at dosage amount 1.5.
- Any effect on length at dosage amount of 2.0 for either supplement is **unclear**.
- Overall, the average sample tooth growth length is greater for "OJ" than it is for "VC".

Hypothesis Tests

Given the summary above, we setup the following tests to help determine the effect of the 2 supplements and their doses on tooth growth. In the first test (which serves as the template for the subsequent tests), we compare the effect on tooth growth between supplements OJ and VC at dose = 0.5. **We assume that the NULL hypothesis, $H_0 : \mu_{x_{OJ}} = \mu_{x_{VC}}$, is true**.

```
# Function to take in the dosage group and perform a t-test on supplement effect on length
doseTest <- function(d) {
        testdf <- NULL
        testdf <- subset(ToothGrowth, dose==d, select=c('len', 'supp'))
        # apply t.test to compare the mean lengths by supplement at the dosage group
        t.test(len ~ as.factor(supp), paired=FALSE, data=testdf)}
doseTest(0.5)


##
##  Welch Two Sample t-test
##
```

```
## data:  len by as.factor(supp)
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23            7.98
```

Here, we see that the 95% confidence interval is above zero, suggesting that the mean of supplement OJ at dose=0.5 is larger than the mean of VC at the same dose. The p value of 0.64% is very unlikely, thus causing us to **reject** the NULL hypothesis. **Under similar assumptions, we repeat the *t-test* to compare the 2 supplements in dose groups 1.0 and 2.0, respectively**:

```
doseTest(1)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by as.factor(supp)
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70           16.77
```

```
doseTest(2)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by as.factor(supp)
## t = -0.0461, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06           26.14
```

We note that for dose = 1.0, the p value of 0.1% is unlikely, thus causing us to **reject** the NULL hypothesis. However, for dose = 2.0, the p value of 96.39% is well within the 95th percentile of values that comply with the NULL hypothesis. In this case, then, the NULL hypothesis **stands**.

Assumptions and Conclusions

The following assumptions must be restated for the tests conducted above:

- The subjects are independent, i.e. the distribution in each group is not paired.

- Unequal variance between groups has been assumed.

The scatter plot and the tests performed on the dataset help us conclude the following:

- At dose=0.5, OJ results in higher tooth growth than VC
- At dose=1.0, OJ results in higher tooth growth than VC
- At dose=2.0, there is no discernible difference in tooth growth between the supplements
- Overall, it seems that OJ is more effective than VC at increasing tooth growth when administered at doses 0.5 or 1.0.

END OF REPORT

APPENDIX 1: GIT CODE

The entire markdown file can be found in this github repo.

APPENDIX 2: PLOT CODE

Data Scatter by Supplement Group and Dosage

```
g1 <- ggplot(data=ToothGrowth, aes(x=dose, y=len, group=supp, col=supp, size=len) )
g1 + geom_point()+
        geom_jitter(position=position_jitter(width=0.1))+
        geom_smooth(alpha=0.3, method="loess")+
        geom_hline(data=means.df, aes(yintercept=c(means[[1]], means[[2]]), col=supp), linetype="longda
        geom_text(data=means.df, aes(x=1.5, y=c(means[[1]], means[[2]]), col=supp,
                                    label=c(paste0("Mean = ", round(means[[1]], 2)),
                                            paste0("Mean = ", round(means[[2]], 2)))),
                                        size=4, vjust=1)+
        labs(title="Scatter: Tooth Growth Length Vs. Dose\nof Supplements OJ and VC") +
        theme_bw()
```