ToothGrowth Analysis - Confidence Intervals

Jason Giacomo

November 20, 2015

Synopsis

This report will analyze the ToothGrowth data set in R. The goal is to look at the data and perform statistics based on confidence intervals to answer questions about whether there is a difference in the type of supplement given and whether the dose has an effect on tooth growth.

Tooth Growth Data

Begin by loading the packages we'll use to complete the analysis in this report.

```
library(datasets)  # should be loaded by default, but just in case
library(dplyr)  # for manipulation of data
library(ggplot2)  # for plotting of data
```

Let's take a quick look at the dataset.

```
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: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
summary(ToothGrowth)
```

```
##       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
```

From this quick look we see that there are three variables: len (length of tooth), supp (supplement type), and dose (amount given). These are true variables and it appears each of the 60 rows contains one observation so the data are tidy. Also, the summary indicates there are no missing values. It appears that length is a measure of the length of a tooth. It also appears that only two supplements were given in the study: OJ or orange juice and VC or vitamin C. Now we'll take a look at the doses given for each supplement.

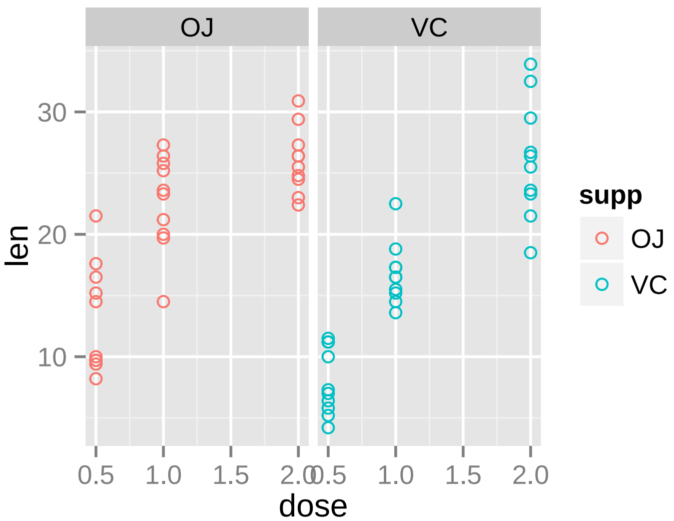
```
table(ToothGrowth$supp, ToothGrowth$dose)
```

```
##     
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

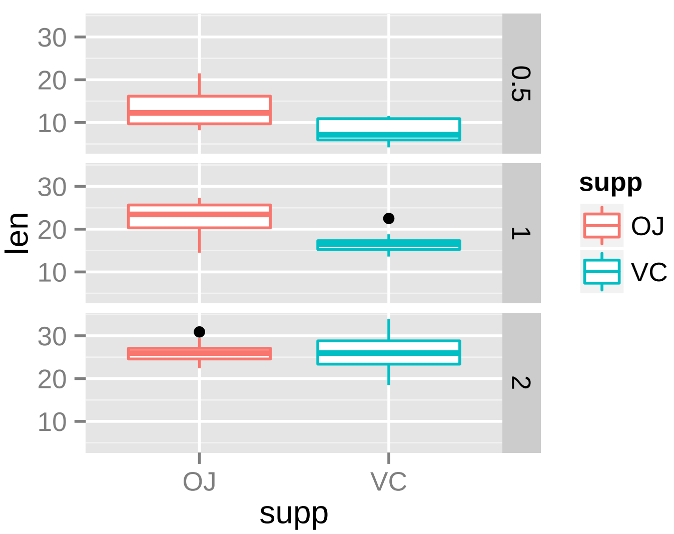
Here it is easily seen that each supplement was given in three doses (0.5, 1, 2) and that each combination of supplement and dose was given to 10 subjects. With only the information given it is likely that each of the 60 subjects is unique (no subject was tested with one dose and supplement and then tested again with a different combination). This means we do not have any paired data.

To further understand our data let's plot the data for each supplement and then compare the data between supplements at each dose level.

```
p1 <- ggplot(ToothGrowth, aes(x=dose, y=len, color=supp))
print(p1 + geom_point(shape=1) + facet_grid(. ~ supp))
```




```
p2 <- ggplot(ToothGrowth, aes(x=supp, y=len, color=supp))
print(p2 + geom_boxplot() + facet_grid(dose ~ .))
```



From the plots it appears that dose has an effect on tooth growth and also that orange juice supplements induce more tooth growth than vitamin C at the lower doses but maybe not at the higher dose. Our eyes can be deceiving however, and it's best that we perform a statistical analysis to check these observations.

Comparing Vitamin C with Orange Juice Supplements

From the boxplots we plotted it appears that there may be an advantage in tooth growth in subjects given orange juice compared to the vitamin C group except perhaps in the highest dose. We'll split the data by dose and then compare both supplements. First, we split the data.

```
doseHalf <- ToothGrowth %>% filter(dose==0.5)
doseOne <- ToothGrowth %>% filter(dose==1)
doseTwo <- ToothGrowth %>% filter(dose==2)
```

Next we form our hypothesis. We will use the same hypothesis across all doses. Null hypothesis: vitamin C and orange juice have the same effect on tooth growth.

$$H_o : \mu_{VC} = \mu_{OJ}$$

Alternative hypothesis: vitamin C increased or decreased tooth length over orange juice.

$$H_a : \mu_{VC} \neq \mu_{OJ}$$

Also, since the treatment of the subjects and the measurments were likely the same for each supplement and dose, we will make the assumption that the variance is the same for each supplement.

To test our hypothesis we will calculate the 95% confidence interval of the difference between the vitamin C and orange juice supplements.

```
# dose = 0.5
doseHalf.len <- split(doseHalf$len, doseHalf$supp)
doseHalf.conf <- t.test(doseHalf.len$VC, doseHalf.len$OJ,
                        alternative = "two.sided", var.equal = TRUE,
                        conf.level = 0.95)$conf[c(1,2)]

# dose = 1
doseOne.len <- split(doseOne$len, doseOne$supp)
doseOne.conf <- t.test(doseOne.len$VC, doseOne.len$OJ,
                       alternative = "two.sided", var.equal = TRUE,
                       conf.level = 0.95)$conf[c(1,2)]

# dose = 2
doseTwo.len <- split(doseTwo$len, doseTwo$supp)
doseTwo.conf <- t.test(doseTwo.len$VC, doseTwo.len$OJ,
                       alternative = "two.sided", var.equal = TRUE,
                       conf.level = 0.95)$conf[c(1,2)]
```

Dose	Confidence Interval
0.5	-8.73, -1.77
1	-9.02, -2.84
2	-3.56, 3.72

From these results it is clear that a difference of 0 is not contained within the confidence intervals for doses of 0.5 and 1 so we reject the null hypothesis and state that vitamin C does have a different effect than orange juice. However, the confidence intervals are both less than zero so, we must conclude that the vitamin C results in less tooth growth compared to orange juice.

When a dose of 2 is used the confidence interval does contain 0. Therefore, we find in favor of the null hypothesis and conclude that we're 95% confident that vitamin C and orange juice have the same effect on tooth growth at a dose of 2.

The Effect of Dose on Tooth Growth

Now that we have compared the supplements to each other and determined that vitamin C is as effective as orange juice at a dose of 2, we should determine if a dose of 2 has an increased effect over a dose of 1. The higher the dose the higher the cost presumably so it would be of benefit to ensure that the higher dose of Vitamin C has an increased effect over a dose of 1. We set our hypotheses as follows:

$$H_o : \mu_{dose2} = \mu_{dose1}$$

$$H_a : \mu_{dose2} > \mu_{dose1}$$

The t-test will be performed the same as before only this time we are interested in testing if dose 2 is greater than dose 1 so we will be using a single tailed test.

```
suppVC.dose2 <- doseTwo %>% filter(supp=="VC") %>% select(len)
suppVC.dose1 <- doseOne %>% filter(supp=="VC") %>% select(len)

t.test(suppVC.dose2, suppVC.dose1,
       alternative = "greater", var.equal = TRUE,
       conf.level = 0.95)$conf[c(1,2)]
```

```
## [1] 6.399483      Inf
```

The confidence interval is greater than zero so we can say a vitamin C dose of 2 increases the effect on tooth growth over a dose of 1 with 95% confidence.

Conclusions

The statistics calculated from the tooth growth data indicate that the vitamin C supplement is only as effective as orange juice at a dose of 2 (at 95% confidence). Also, we are 95% confident that a vitamin C dose of 2 has a greater effect on tooth growth than does a dose of 1.

These results were calculated based on the assumptions that all of the subjects are unique, so there was no paired data. Also, that the distribution of lenghts all have equal variance. This assumption was made because there is no indication to the contrary that the measurements were all completed in the same manner and study subjects were picked at random.