Statistical Inference Project Part 2: Inferential Data Analysis

Andrew Chellinsky

August 21, 2016

Introduction

In the second part of this course project, we have four tasks:

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

Each of these tasks is in its own section throughout this document.

Load and Explore Data

First, load the data frame:

```
data("ToothGrowth")
```

Now, perform some simple exploration of the data frame:

```
dim(ToothGrowth)
```

```
## [1] 60  3
```

```
head(ToothGrowth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

```
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 ...
```

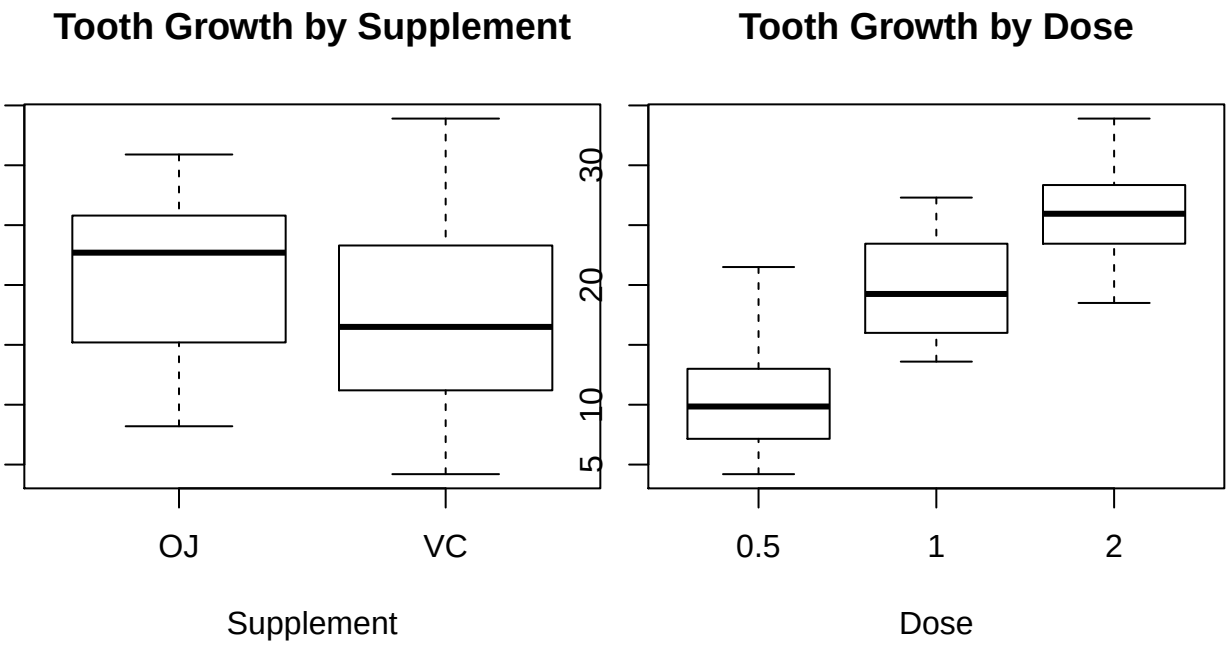
Looks like dose may be better as a factor. Let's clean that up:

```
tg2 <- ToothGrowth
tg2$dose <- as.factor(tg2$dose)
```

Let's make a couple simple plots of dose and supp compared to len.

```r
par(mfrow = c(1, 2), pin = c(3, 2))

plot(tg2$supp, tg2$len, type = "s", main = "Tooth Growth by Supplement", xlab = "Supplement", ylab = "L
plot(tg2$dose, tg2$len, type = "s", main = "Tooth Growth by Dose", xlab = "Dose", ylab = "Length")
```



Data Summary

To summarize the data briefly:

```r
summary(tg2)
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

In other words, the data frame consists of 60 measures. They are split equally between the supp and dose variables–30 for each supp and 20 for each dose. Searching R help reveals that supp refers to supplements, either orange juice (OJ) or ascorbic acid (VC). Dose refers to the dose of the supplement in milligrams per day.

Comparison of Tooth Growth by Supp and Dose

To compare tooth growth by supplement and dose, let's make a couple null hypotheses. First for supplements, the null hypothesis is that the mean tooth growth is the same for both OJ and VC supplement types. Second for dose, the null hypothesis is that the mean tooth growth is the same for all three doses: 0.5, 1, and 2.

Using a t-test on the first null hypothesis (for supplements) yields a result that shows the means are not significantly different at the 95% confidence interval. However, it is significant at the 94% confidence interval.

```r
OJ <- tg2[tg2$supp == "OJ", ]
VC <- tg2[tg2$supp == "VC", ]
t.test(OJ$len, VC$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ$len and VC$len
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

Similarly, using a t-test on the second null hypothesis (for dose) yields a result that shows each mean is significantly different. In other words, all dosages gave significantly different results at a 95% confidence interval.

```r
D5 <- tg2[tg2$dose == 0.5, ]
D1 <- tg2[tg2$dose == 1, ]
D2 <- tg2[tg2$dose == 2, ]
t.test(D5$len, D1$len)$conf.int
```

```
## [1] -11.983781  -6.276219
## attr(,"conf.level")
## [1] 0.95
```

```r
t.test(D5$len, D2$len)$conf.int
```

```
## [1] -18.15617 -12.83383
## attr(,"conf.level")
## [1] 0.95
```

```r
t.test(D1$len, D2$len)$conf.int
```

```
## [1] -8.996481 -3.733519
## attr(,"conf.level")
## [1] 0.95
```

Conclusions and Assumptions

Conclusions

To conclude, each level of dosage has a significant effect on tooth growth length. The type of supplement, on the other hand, does not have a significant effect on tooth growth length.

Assumptions

This analysis assumes normally distributed data and iid measurements. These assumptions appear to be safe based on the data, but without access to more information on the sampling procedures, it is left as an assumption.

Additionally, this analysis assumes the effect would be between doses or supplements, not an interaction of the two. Therefore, this analysis did not look at the combined effect of supplements and doses for significance.