The Impact of Supplement and Dose on Tooth Growth

Richard Ashley

Sunday, August 23, 2015

Project for the Statistical Inference Coursera Class

Overview

We now look at some standard Tooth Growth data included in R. We will be comparing the impact of `Supplements` and `Dose` of Supplements on `Tooth Growth`. We will explore the data and then look at the confidence intervals of Tooth Growth at each combination of factor and then drawing conclusions on the impact of these factors.

Data Exploration

1. Load the ToothGrowth data and perform some basic exploratory data analyses.

First, lets load the `Tooth Growth` data and then take a look at the first few rows of data.

```
data(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
```

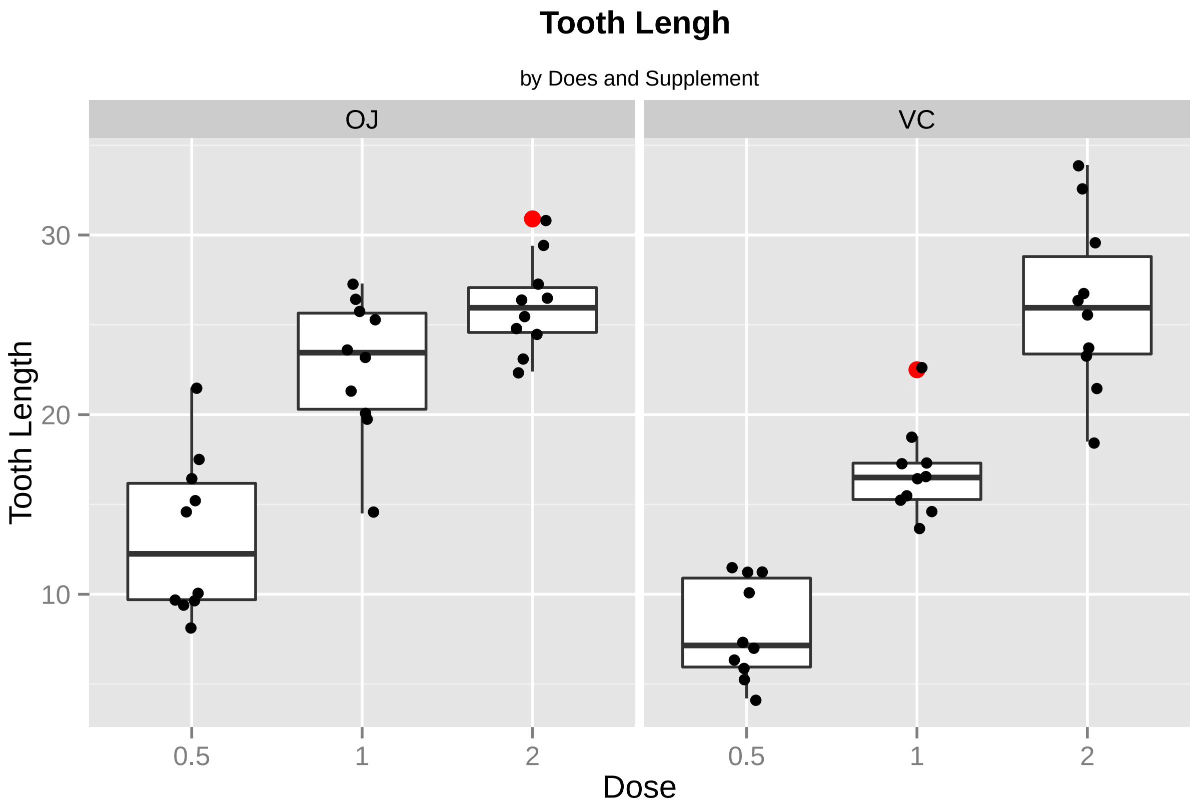
2. Provide a basic summary of the data.

If we now look at both that structure and a summary of the data we find 3 variables, `len` is the length of the tooth, `supp` is the Supplement used (2 levels), and `dose` is the dosage used (3 levels)

```
## '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 ...
```

```
##       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
```

Next, lets create a box-plot of the 2 factors `supp` and `dose` and look at the response of tooth length `len`.



Visually, we see a significant increase at higher dose as well stronger impact to tooth growth at lower doses for the 'OJ' supplement.

Some additional statistics on each combination. First lets look at the means by `supp` and `dose`.

```
with(ToothGrowth,aggregate(len, list(supp=supp, dose=dose), mean))
```

```
##   supp dose     x
## 1   OJ  0.5 13.23
## 2   VC  0.5  7.98
## 3   OJ    1 22.70
## 4   VC    1 16.77
## 5   OJ    2 26.06
## 6   VC    2 26.14
```

Now lets look at the stander deviation by `supp` and `dose`.

```
with(ToothGrowth,aggregate(len, list(supp=supp, dose=dose), sd))
```

```
##   supp dose        x
## 1   OJ  0.5 4.459709
## 2   VC  0.5 2.746634
## 3   OJ    1 3.910953
## 4   VC    1 2.515309
## 5   OJ    2 2.655058
## 6   VC    2 4.797731
```

Factor Analysis

We will not compare tooth growth for each of the factor levels and create confidence intervals in order to draw conclusions about the differences of the mean growth.

3. Use confidence intervals to compare tooth growth by supp and dose.

To calculate the Confidence Intervals and P-value for each combination, we will first break up the data into a wide format and then use a `t.test` to calculate a confidence interval of the difference between each factor combination. Since the standard deviations that we looked at above did not look to be equal, I will go with the default of `var.equal = FALSE`. Below is a the CI between supplement factor `OJ` and `VC` at a dose of 2.0.

```
attach(ToothGrowth)

### Break up data wide to format and rearange
groups<-as.data.frame(split(len,list(supp,dose)))
groups <- groups %>% select(OJ.0.5, OJ.1, OJ.2, VC.0.5, VC.1, VC.2)

### Calculate P-Value and CI
paste0("P-value = ",round(t.test(groups$OJ.2,groups$VC.2)$p.value, 5))
```

```
## [1] "P-value = 0.96385"
```

```
paste0("Lower CI = ",round(t.test(groups$OJ.2,groups$VC.2)$conf.int[1], 2))
```

```
## [1] "Lower CI = -3.8"
```

```
paste0("Upper CI = ",round(t.test(groups$OJ.2,groups$VC.2)$conf.int[2], 2))
```

```
## [1] "Upper CI = 3.64"
```

As you can see, supplement factor `OJ` and `VC` at dose of 2.0, the the upper and lower CI (-3.8 , 3.64) includes 0 so we cannot conclude the two samples are different. The P-value of 0.96 also indicates this.

We will repeat this and produce a table of the CI and P-values for all combinations of `supp` and `dose`. The names of the comparisons are `supp.dose - supp.dose`

	P-value	L-Conf-int	U-Conf-int
OJ.0.5-OJ.1	0.00009	-13.42	-5.52
OJ.0.5-OJ.2	0.00000	-16.34	-9.32
OJ.0.5-VC.0.5	0.00636	1.72	8.78
OJ.0.5-VC.1	0.04601	-7.01	-0.07
OJ.0.5-VC.2	0.00001	-17.26	-8.56
OJ.1-OJ.2	0.03920	-6.53	-0.19
OJ.1-VC.0.5	0.00000	11.52	17.92
OJ.1-VC.1	0.00104	2.80	9.06
OJ.1-VC.2	0.09653	-7.56	0.68
OJ.2-VC.0.5	0.00000	15.54	20.62
OJ.2-VC.1	0.00000	6.86	11.72
OJ.2-VC.2	0.96385	-3.80	3.64
VC.0.5-VC.1	0.00000	-11.27	-6.31
VC.0.5-VC.2	0.00000	-21.90	-14.42
VC.1-VC.2	0.00009	-13.05	-5.69

4. State your conclusions and the assumptions needed for your conclusions.

As we can see from the CI and P-values in the table above, increasing `dose` is almost always significant independent of the supplement `supp`. In addition, the supplement `supp` is significant at lower `dose` levels but not at the higher `dose` levels and in this case OJ otperforms VC.

It was assumed that the variances were not equal and that the data was NOT from a paired test. In addition, the confidence level was left at the default at 0.95. Since we are dealing with small sample sizes, a t-test was used. As such, I am also taking advantage of looking at the average of 10 samples each and that this statistic will be approximately normally distributed.

Code Chunks

```
library(knitr)
library(markdown)
library(ggplot2)
library(tidyr)
library(dplyr)

### Load the data and look at some summary info
data(ToothGrowth)
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
str(ToothGrowth)

summary(ToothGrowth)

p <- ggplot(ToothGrowth, aes(factor(dose), len)) +
    geom_boxplot(outlier.colour = "red", outlier.size = 3) +
    geom_jitter(position = position_jitter(width = .1))
p <- p + ggtitle(expression(atop(bold("Tooth Lengh "),
                                 scriptstyle("by Does and Supplement")))) +
    theme(plot.title = element_text(size = 20)) +
    theme(legend.position = c(0.85, 0.85)) +
    labs(x = "Dose", y = "Tooth Length") +
    theme(plot.title = element_text(size = 12))

p + facet_grid(. ~ supp)

with(ToothGrowth,aggregate(len, list(supp=supp, dose=dose), mean))
with(ToothGrowth,aggregate(len, list(supp=supp, dose=dose), sd))


### Look at the CI and P-values
attach(ToothGrowth)

groups<-as.data.frame(split(len,list(supp,dose)))
groups <- groups %>% select(OJ.0.5, OJ.1, OJ.2, VC.0.5, VC.1, VC.2)

paste0("P-value = ",round(t.test(groups$OJ.2,groups$VC.2)$p.value, 5))
paste0("Lower CI = ",round(t.test(groups$OJ.2,groups$VC.2)$conf.int[1], 2))
paste0("Upper CI = ",round(t.test(groups$OJ.2,groups$VC.2)$conf.int[2], 2))

rnames<-vector()
```

```
count<-0
for ( f in 1:5 ) {
    for ( t in (f+1):6 ) {
        count<-count+1
        rnames[count]<-paste(as.character(names(groups)[f]),
                             as.character(names(groups)[t]),sep="-")
    }
}

test<-matrix(data=NA,nrow=length(rnames),ncol=3,byrow=TRUE,
             dimnames=list(rnames,c("P-value","L-Conf-int", "U-Conf-int")))
count<-0
for ( f in 1:5 ) {
    for ( t in (f+1):6 ) {
        count<-count+1;
        test[count,1] <- round(t.test(groups[,f],groups[,t])$p.value, 5)
        test[count,2] <- round(t.test(groups[,f],groups[,t])$conf.int[1], 2)
        test[count,3] <- round(t.test(groups[,f],groups[,t])$conf.int[2], 2)
        }
}
test
```