Analyzing ToothGrowth data

BK

Monday, October 19, 2015

Objective

To compare the effectiveness of two different supplements on tooth growth.

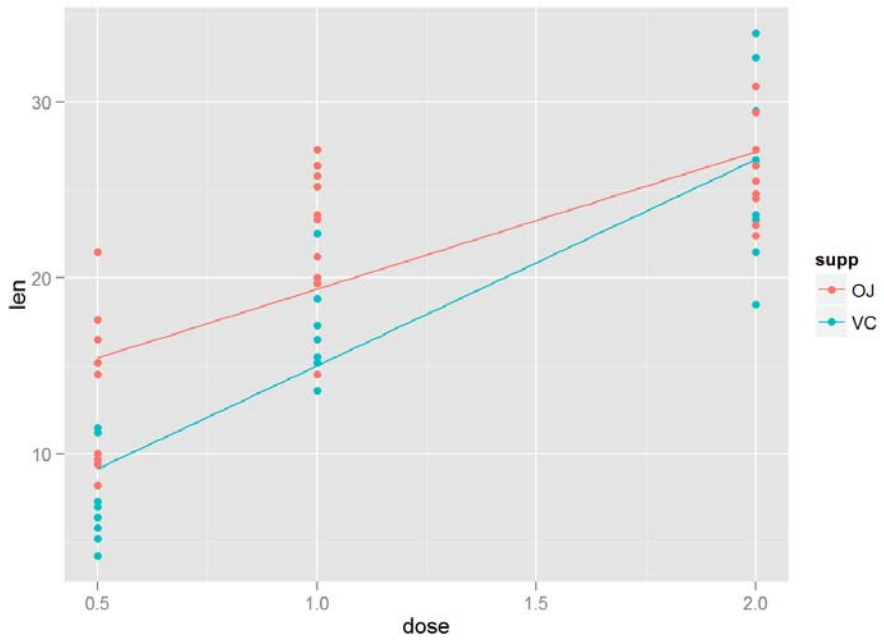
Exploratory analysis

We first learn that the data is structured in long form to show length, supplement and dosage.
The number of rows in ToothGrowth is 60 and columns is 3. The head of the data looks like this:

```
##    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
```

We then plot the data to tease out insights that require further numerical analysis.

```
## Warning: package 'ggplot2' was built under R version 3.1.3
```



Insight gained: That Orange Juice (OJ) is a more effective supplement than ascorbic acid (VC) at lower dosage levels and that the outcomes at the highest dosage are too close to call.

Reshaping data

The data needs to be recast from its given long table form in order to compare Supplement+Dosage combinations. The recast data appears below:

```
##    VC_0.5 VC_1.0 VC_2.0 OJ_0.5 OJ_1.0 OJ_2.0
## 1     4.2   16.5   23.6   15.2   19.7   25.5
## 2    11.5   16.5   18.5   21.5   23.3   26.4
## 3     7.3   15.2   33.9   17.6   23.6   22.4
## 4     5.8   17.3   25.5    9.7   26.4   24.5
## 5     6.4   22.5   26.4   14.5   20.0   24.8
## 6    10.0   17.3   32.5   10.0   25.2   30.9
## 7    11.2   13.6   26.7    8.2   25.8   26.4
## 8    11.2   14.5   21.5    9.4   21.2   27.3
## 9     5.2   18.8   23.3   16.5   14.5   29.4
## 10    7.0   15.5   29.5    9.7   27.3   23.0
```

Determining the difference in efficacy of the supplements, at different dosages

The exploratory analysis with the ggplot above showed Orange Juice to be more effective at low doses. We determine the confidence intervals below.

At 0.5 dosage:

Finding: Orange Juice (OJ) is more effective for tooth growth than ascorbic acid (VC). The confidence interval does not cross/ include zero.

```
## [1] 1.263458 9.236542
## attr(,"conf.level")
## [1] 0.95
```

At 1.0 dosage:

Finding: Orange Juice (OJ) is more effective for tooth growth than ascorbic acid (VC). The confidence interval does not cross/ include zero.

```
## [1] 1.951911 9.908089
## attr(,"conf.level")
## [1] 0.95
```

At 2.0 dosage:

Finding: There is no statistically significant difference between the effectiveness of Orange Juice (OJ) over ascorbic acid (VC) for tooth growth. The confidence interval does cross/ include zero.

```
## [1] -4.328976  4.168976
## attr(,"conf.level")
## [1] 0.95
```

Assumptions

1. The help file provided with the data, does not indicate that the observations were paired in any manner. So it is assumed that data is not paired.
2. Similarly, no information on variance of these unpaired populations is provided. So, the we make the conservative assumption that the variances are unequal.

Appendix

All the code used for this analysis is shown below.

```r
data(ToothGrowth)
head(ToothGrowth)
summary(ToothGrowth)
nrow(ToothGrowth)

library(ggplot2)
g <- ggplot(ToothGrowth, aes(x=dose, y=len, group = supp, colour = supp))
g <- g + geom_point() + geom_smooth(method = "lm", se = F)
g



vctg <- ToothGrowth[ToothGrowth$supp=="VC",]
ojtg <- ToothGrowth[ToothGrowth$supp=="OJ",]

vctgW <- as.data.frame(with(vctg,split(len,dose)))
names(vctgW) <- c('VC_0.5','VC_1.0','VC_2.0')

ojtgW <- as.data.frame(with(ojtg,split(len,dose)))
names(ojtgW) <- c('OJ_0.5','OJ_1.0','OJ_2.0')

TGW <- cbind(vctgW,ojtgW)
TGW

t.test(TGW["OJ_0.5"] - TGW["VC_0.5"], paired = F, var.equal=F)$conf

t.test(TGW["OJ_1.0"] - TGW["VC_1.0"], paired = F, var.equal=F)$conf

t.test(TGW["OJ_2.0"] - TGW["VC_2.0"], paired = F, var.equal=F)$conf
```