Tooth Growth and Supplements

Kai Wakerman Powell

1 April 2016

Summary

The following is a very basic exploration of two supplements, orange juice and vitamin C, on tooth growth. Basic exploration of the data and plotting of the results make it clear that one type of supplement, orange juice, seems to have more impact on too growth length. Hypothesis testing of this against a Null Hypothesis that there is no difference confirm this to be true.

However further similar testing, while taking into account the dose levels of each of these supplements only confirm the advantage of orange juice for the 0.5 and 1 levels. It appears that the two supplements converge in the sample space (0.5-2.0). Further work probably involves regressions, which are not covered in the Statistical Inference course.

Setup

Basic setup and details for reproducibility.

```
setwd("~/Desktop/Coursera/SI/SIAss")
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.2.4
```

```
library(datasets)
set.seed(0987654321)
data("ToothGrowth")
```

Basic exploration

this shows some basic exploration of the data.

```
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 ...
```

```
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
```

```
summary(ToothGrowth)
```

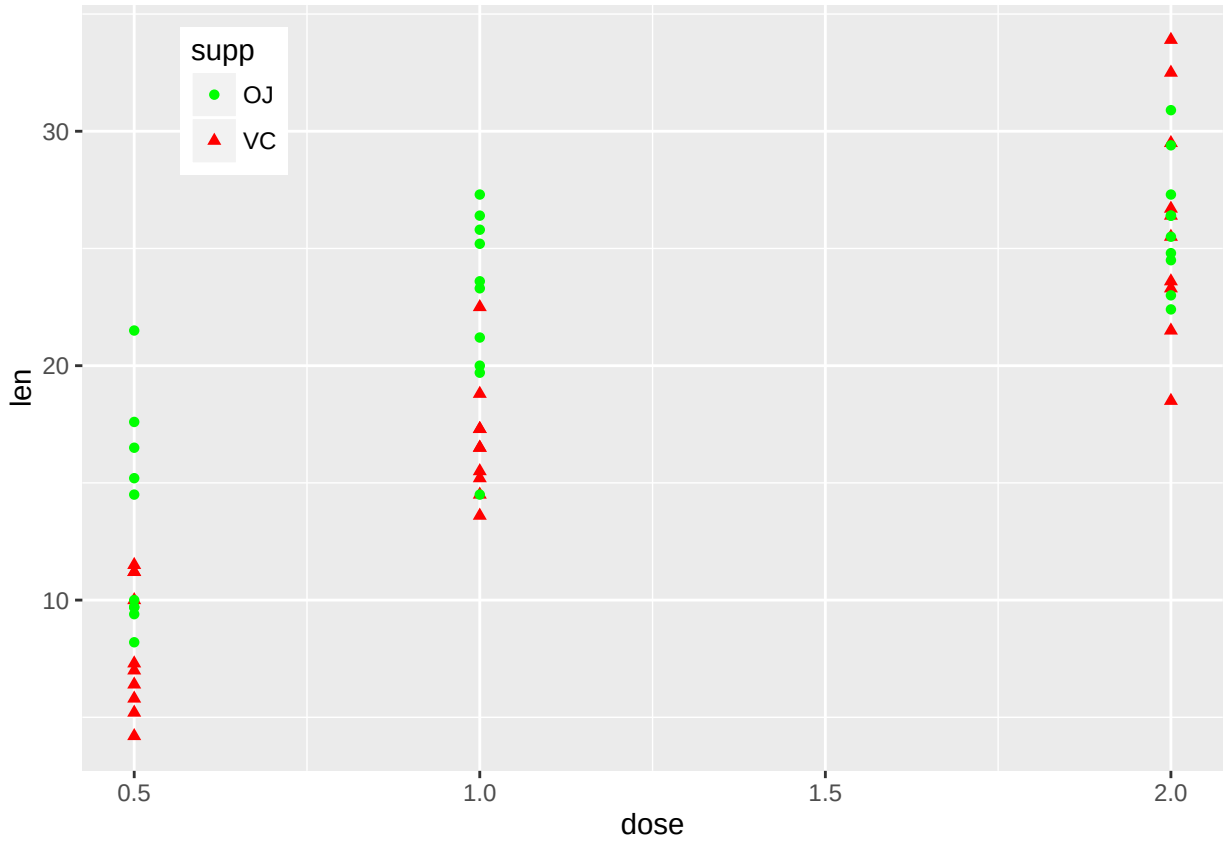
```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

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

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

However it is much clearer to plot the data than to read summaries.

```
plot <- ggplot(ToothGrowth,aes(dose,len,shape=supp,color=supp))+geom_point()+scale_color_manual(values=c
plot
```



Basic comparison

The plot appears to suggest there is a difference between orange juice (labelled OJ) and vitamin C (labelled VC). First step is to compare the raw sample means:

```
OJTooth<-ToothGrowth$len[ToothGrowth$supp=="OJ"]
VCTooth<-ToothGrowth$len[ToothGrowth$supp=="VC"]

meanvc<-mean(VCTooth)
meanvc
```

```
## [1] 16.96333
```

```
meanoj<-mean(OJTooth)
meanoj
```

```
## [1] 20.66333
```

```
meanoj-meanvc
```

```
## [1] 3.7
```

To confirm this, we can run a hypothesis test of OJ is greater than VC against a Null Hypothesis of no difference:

```
t.test(OJTooth,VCTooth,alt="greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  OJTooth and VCTooth
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687       Inf
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

This confirms that OJ has a greater impact on tooth growth length (assuming, oddly enough, that tooth growth length was measured after as a result of the supplements, not the other way around).

Dose-wise comparisons

However the two supplements only cover one of the two causal variables. We also want to examine whether the advantage of OJ over VC is maintained over varying dose levels:

```
oj05<-ToothGrowth$len[ToothGrowth$supp=="OJ"&ToothGrowth$dose==0.5]
oj10<-ToothGrowth$len[ToothGrowth$supp=="OJ"&ToothGrowth$dose==1]
oj20<-ToothGrowth$len[ToothGrowth$supp=="OJ"&ToothGrowth$dose==2]
vc05<-ToothGrowth$len[ToothGrowth$supp=="VC"&ToothGrowth$dose==0.5]
vc10<-ToothGrowth$len[ToothGrowth$supp=="VC"&ToothGrowth$dose==1]
vc20<-ToothGrowth$len[ToothGrowth$supp=="VC"&ToothGrowth$dose==2]

t.test(oj05,vc05,alt="greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  oj05 and vc05
## t = 3.1697, df = 14.969, p-value = 0.003179
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  2.34604     Inf
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

```
t.test(oj10,vc10,alt="greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  oj10 and vc10
## t = 4.0328, df = 15.358, p-value = 0.0005192
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  3.356158      Inf
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

```
t.test(oj20,vc20,alt="greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  oj20 and vc20
## t = -0.046136, df = 14.04, p-value = 0.5181
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  -3.1335     Inf
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

This shows that we can reject the Null Hypothesis for 0.5 and 1 dose levels, and that OJ appears to have an advantage at these levels. However we cannot reject the Null Hypothesis for a dose level of 2. This suggests that the two supplements converge on the upper dose levels within the sample space.

Assumptions

1. Samples are representative of the popultation
2. Data is unpaired
3. Variances are the same
4. Distributions are roughly normal/t

(to save space I didn't investigate all these)