Statistical Inference Course Project: Part 2, Inferential Data Analysis

Amal Haq

June 20, 2015

```
library(ggplot2)
library(reshape2)
```

```
## Warning: package 'reshape2' was built under R version 3.1.2
```

The instructions ask us to use an existing dataset in R, called “ToothGrowth”. We are going to take a look at this data to see what has bee recorded, and use the datapoints to make some ‘inferences’ about tooth growth and how it is affected by different supplements and different doses of those supplements.

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

```
data(ToothGrowth)
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, 3)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
```

```
tail(ToothGrowth, 3)
```

```
##     len supp dose
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

```
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

`?ToothGrowth` tells us that we are looking at data for the effect of Vitamin C on Tooth Growth in Guinea Pigs. The data is stored in a data.frame. It has 60 rows and only 3 columns, which correspond to the tooth length, vitamin C supplements, and the actual dose of vitamin C. There are only two types (factor levels) of supplements, Orange Juice or ascorbic acud. Using the `unique(ToothGrowth$dose)` function, also shows that there are only three possible values for this variable, 0.5, 1.0 or 2.0.

Provide a basic summary of the data

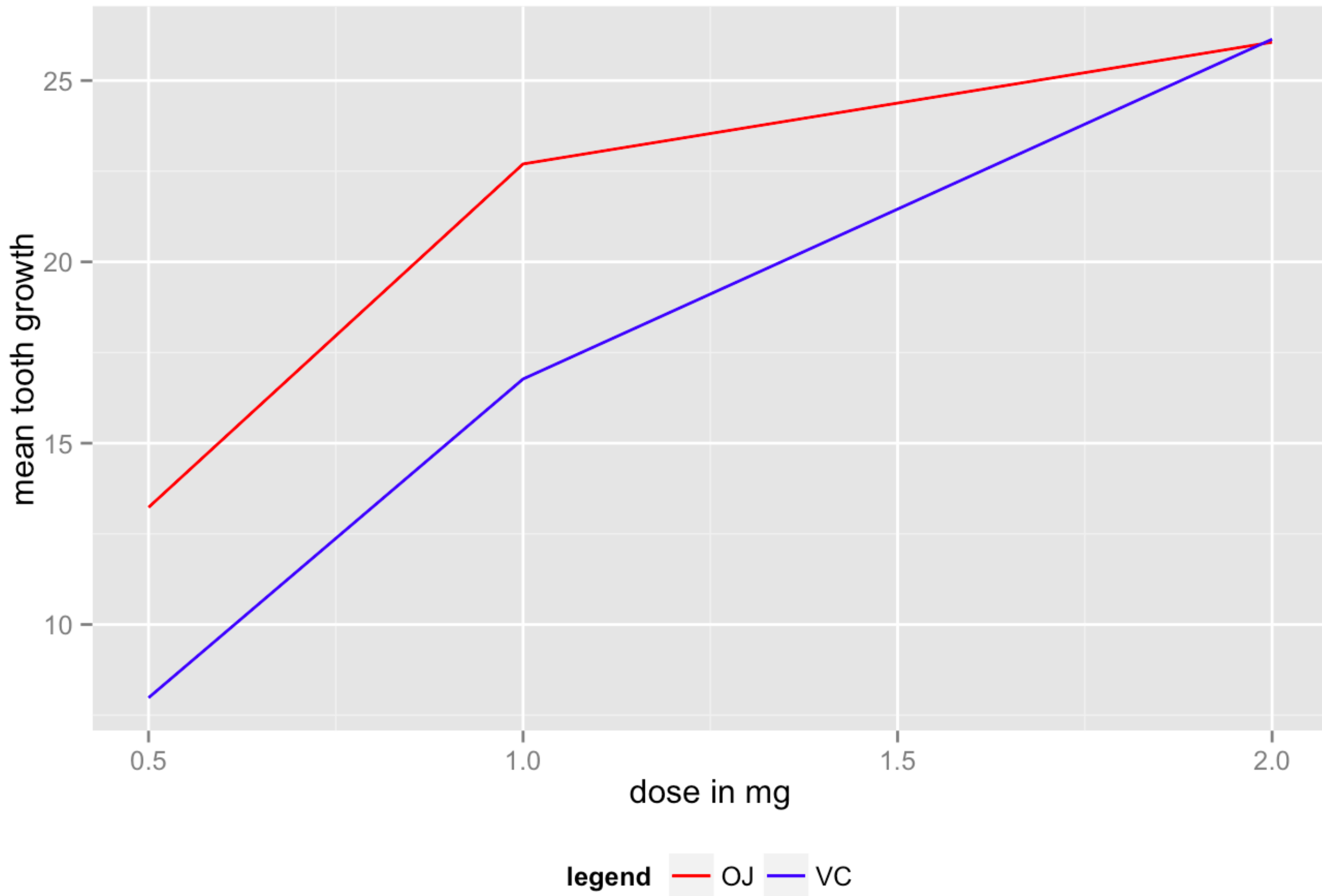
```
summary(ToothGrowth)
```

```
##       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  
```

These means and medians don’t really tell us much. Given that there are only three dose levels, and two supplement types, it will be a lot more informative to see average tooth growth by dose and by supplement. *See Code Chunk 1 in Appendix*.

```
##    dose    OJ    VC
## 1   0.5 13.23  7.98
## 2     1  22.7 16.77
## 3     2 26.06 26.14
## 4 total 20.66 16.96
```

Initially, it looks like the total mean tooth growth from 30 observation of OJ is more than the total mean tooth growth from 30 observation of ascorbic acid. However, when we look at a plot of mean tooth growth by dose of the supplement, we see something more (*See Code Chunk 2 in Appendix*):



So, it looks like, at smaller doses of vitamin C, OJ seems to result in more mean tooth growth than ascorbic acid. However, at larger doses, the mean tooth growth is about the same regardless of whether vitamic C is administered through OJ or ascorbic acid.

Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose.

So far we have done some basic exploratory data analysis. Let's see if we can back our initial analysis with some confidence intervals and hypothesis tests.

First Test

Let's compare the type of supplement (OJ vs ascorbic acid) to determine whether the way in which vitamin C is administered has an impact on tooth growth.

The null hypothesis *H0* is that there is no impact. The alternative hypothesis *H1* is that there is an impact, and total mean tooth growth from OJ and ascorbic acid are different.

Because the guinea pigs that receive ascorbic acid versus those that receive OJ are different and receive different treatment, we will consider these independent groups and thus will not perform a paired t-test.

```
t.test(ToothGrowth$len[ToothGrowth$supp=="OJ"], ToothGrowth$len[ToothGrowth$supp=="VC"], paired=FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$supp == "OJ"] and ToothGrowth$len[ToothGrowth$supp ==
"VC"]
## 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
```

The means do look different; we had seen this in our exploratory data analysis as well. However, to reject H0, we should see a p-value of 0.05 or less, and our 95% confidence interval should not contain 0. **Our p-value = 0.06 and confidence interval = (-0.171, 7.571). Therefore, we do not have enough evidence to reject the null hypothesis, and we cannot say with confidence that the type of supplement affects tooth growth.**

Second Test

Let's compare dosage to determine whether the amount of vitamin C administered has an impact on tooth growth.

Let's check 0.5mg versus 1.0mg. H0 is that mean tooth growth from 0.5mg and 1mg of vitamin C is the same (difference in the mean is 0). H1 is that the mean tooth growth from a 0.5mg dose and 1.0mg dose of vitamin C is different.

Because we are looking at four sets of guinea pigs, and each set receives a different dose and supplement type, we will again consider these as independent groups and thus will not perform a paired t-test (*See Code Chunk 3 in Appendix*).

```
## [1] 1.268301e-07
```

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

The *p-value (1.268e-07)* of this test is very small and the *confidence interval (6.276, 11.984)* does not include 0. **Therefore, we have enough evidence to reject the null hypothesis; increasing the dose of vitamin C from 0.5 mg to 1.0 mg does impact mean tooth growth, regardless of the supplement.**

Similarly, comparing 1.0mg dose to 2.0mg dose by grabbing the relevant p-value and confidence interbval (*See Code Chunk 4 in Appendix*):

```
## [1] 1.90643e-05
```

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

Therefore, we have enough evidence to reject the null hypothesis; increasing the dose of vitamin C from 1.0 mg to 2.0 mg does impact mean tooth growth, regardless of the supplement.

Third Test

Let's do one last test. This is the more nuanced test but it will complement the first two tests rather well. Let's check supplement type and dosage together. Are certain combinations of dose and supplement types better than others?

The H0 is that a given dose of vitamin C (e.g. 0.5 mg) administered through one type of supplement (e.g. OJ) does not impact tooth growth any differently than that same dose from the other type of supplement e.g. ascorbic acid). The H1 is that it does have an impact.

Because we are looking at guinea pigs that receive the same amount of vitamin C dose, we will consider this data paired, and perform paired t-tests (*See Code Chunk 5 in Appendix*).

0.5 mg from OJ vs. 0.5 mg from Ascorbic Acid:

```
## [1] 0.01547205
```

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

1.0 mg from OJ vs. 1.0 mg from Ascorbic Acid:

```
## [1] 0.008229248
```

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

2.0 mg from OJ vs. 2.0mg from Ascorbic Acid:

```
## [1] 0.9669567
```

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

How interesting! The data are statistically significant for the combination of lower doses of vitamin C and OJ. However, at higher doses of vitamin C, this difference is diminished. This is what we had seen in our exploratory data analysis, and we have now proven this hypothesis via demonstration with p-values and confidence intervals.

State your conclusions and the assumptions needed for your conclusions.

CONCLUSIONS:

Our inferential data analysis has shown that tooth growth is increased if the dose in mg of vitamin C is increased. We can also conclude that generally, the type of supplement has no effect on tooth growth, however, at low doses of 0.5mg and 1.0mg, OJ is a better supplement than ascorbic acid for vitamin C delivery and results in greater tooth growth in guinea pigs.

ASSUMPTIONS:

It appears that each guinea pigs is randomly assigned a dose amount and a supplement type. Therefore, in the first and second tests, we have assumed that the data being compared are *independent and not paired*. Because these are independent, unpaired groups that are treated under different conditions (different supplements and different dose amounts), we considered them to be representative of populations of guinea pigs that are different from each other. Therefore, for the first and second tests, *variances were assumed to be unequal*.

In the third test, we had to make some judgment calls. We make the assumption that the data are *independent but paired*. The test is performed on two sets of guinea pigs that receive different supplements (independent) but the exact same amount of vitamin C dose in mg (paired). The guinea pigs are considered to be representative of the entire population of guinea pigs, and so the *variances were assumed to be equal*.

APPENDIX FOR R-CODE

Code Chunk 1

```
TGmeans<- round(dcast(ToothGrowth, dose ~ supp, value.var= "len", fun.aggregate=mean),2)
totalmeans<-as.data.frame(cbind(c("total"), round(mean(TGmeans$OJ),2), round(mean(TGmeans$V
C),2)))
colnames(totalmeans)<- c("dose", "OJ", "VC")
TGdf<- rbind.data.frame(TGmeans, totalmeans)
print(TGdf)
```

Code Chunk 2

```
g<- ggplot(TGmeans, aes(x=dose)) +
        geom_line(aes(y=TGmeans$OJ, color="OJ")) +
        geom_line(aes(y=TGmeans$VC, color = "VC")) +
        labs(x="dose in mg", y="mean tooth growth") +
        theme(legend.position="bottom") +
        scale_color_manual(name ="legend", values = c(OJ="red", VC="blue"))
```

Code Chunk 3

```
t.test(ToothGrowth$len[ToothGrowth$dose==1.0], ToothGrowth$len[ToothGrowth$dose==0.5], paire
d=FALSE)$p.value
t.test(ToothGrowth$len[ToothGrowth$dose==1.0], ToothGrowth$len[ToothGrowth$dose==0.5], paire
d=FALSE)$conf.int
```

Code Chunk 4

```
t.test(ToothGrowth$len[ToothGrowth$dose==2.0], ToothGrowth$len[ToothGrowth$dose==1.0], paire
d=FALSE)$p.value
t.test(ToothGrowth$len[ToothGrowth$dose==2.0], ToothGrowth$len[ToothGrowth$dose==1.0], paire
d=FALSE)$conf.int
```

Code Chunk 5

```
OJ0.5<- ToothGrowth$len[ToothGrowth$dose== 0.5 & ToothGrowth$supp=="OJ"]
VC0.5<- ToothGrowth$len[ToothGrowth$dose== 0.5 & ToothGrowth$supp=="VC"]
t.test(OJ0.5, VC0.5, paired=TRUE, var.equal=TRUE)$p.value
t.test(OJ0.5, VC0.5, paired=TRUE, var.equal=TRUE)$conf.int

OJ1.0<- ToothGrowth$len[ToothGrowth$dose== 1.0 & ToothGrowth$supp=="OJ"]
VC1.0<- ToothGrowth$len[ToothGrowth$dose== 1.0 & ToothGrowth$supp=="VC"]
t.test(OJ1.0, VC1.0, paired=TRUE,  var.equal=TRUE)$p.value
t.test(OJ1.0, VC1.0, paired=TRUE,  var.equal=TRUE)$conf.int

OJ2.0<- ToothGrowth$len[ToothGrowth$dose== 2.0 & ToothGrowth$supp=="OJ"]
VC2.0<- ToothGrowth$len[ToothGrowth$dose== 2.0 & ToothGrowth$supp=="VC"]
t.test(VC2.0, OJ2.0, paired=TRUE,  var.equal=TRUE)$p.value
t.test(VC2.0, OJ2.0, paired=TRUE,  var.equal=TRUE)$conf.int
```