

Toothgrowth Analysis

Dominic Lloyd

October 24, 2015

Overview

Now we're going to analyze the ToothGrowth data in the R datasets package.

We will do the following in keeping with the brief:

- Load the ToothGrowth data and perform some basic exploratory data analyses
- Provide a basic summary of the data.
- 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)
- State your conclusions and the assumptions needed for your conclusions.

Some criteria that you will be evaluated on

- Did you perform an exploratory data analysis of at least a single plot or table highlighting basic features of the data?
- Did the student perform some relevant confidence intervals and/or tests?
- Were the results of the tests and/or intervals interpreted in the context of the problem correctly?
- Did the student describe the assumptions needed for their conclusions?

Exploratory Data Analysis

```
rm(list=ls())  
data("ToothGrowth")
```

Output of ?ToothGrowth in R

The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid).

The dataset shows The Effect of Vitamin C on Tooth Growth in Guinea Pigs

The dataset is a data frame with 60 observations on 3 variables.

[,1] len numeric Tooth length
[,2] supp factor Supplement type (VC or OJ).
[,3] dose numeric Dose in milligrams.

```
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 ...
```

```
summary(ToothGrowth)
```

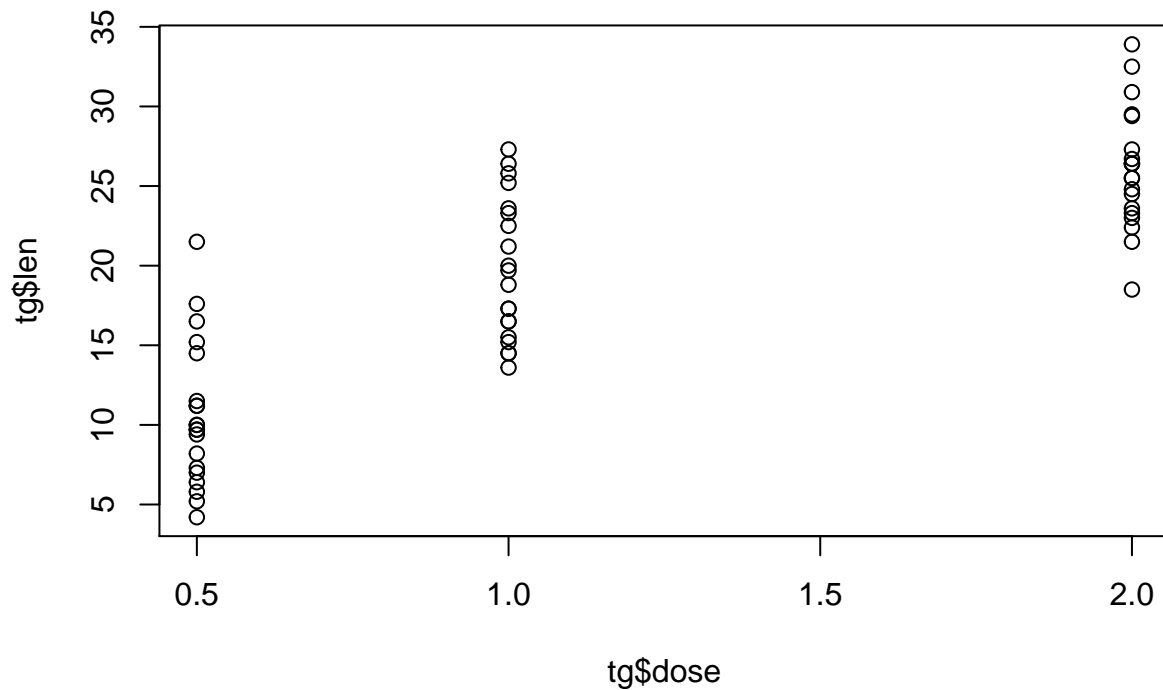
```
##      len      supp      dose
## Min.   : 4.20   OJ:30   Min.    :0.500
## 1st Qu.:13.07   VC:30   1st Qu.:0.500
## Median :19.25           Median :1.000
## Mean   :18.81           Mean   :1.167
## 3rd Qu.:25.27           3rd Qu.:2.000
## Max.   :33.90           Max.    :2.000
```

As the description tells us there are 6 independent sets of 10 guinea pigs.

The 6 independent sets are as follows, with each independent set containing 10 guinea pigs:

- Dose level 0.5 mg Vitamin C delivered as ascorbic acid
- Dose level 0.5 mg Vitamin C delivered as orange juice
- Dose level 1.0 mg Vitamin C delivered as ascorbic acid
- Dose level 1.0 mg Vitamin C delivered as orange juice
- Dose level 2.0 mg Vitamin C delivered as ascorbic acid
- Dose level 2.0 mg Vitamin C delivered as orange juice

```
tg <- ToothGrowth
plot(tg$dose, tg$len)
```



We have a relatively small set of data. The graph shows us that there appears to be a correlation between an increase in tooth length and the increase in dose of vitamin C. We will now do some further research to break the guinea pigs down into their individual groups so we can start to apply confidence levels to the range of upper and lower values for average tooth length given dose and method of delivery.

There are 6 independent sets of different guinea pigs each of a small sample size of 10 pigs. Given the small sample size we will use T confidence intervals to work out a 95% confidence interval for the average length for a particular dose and delivery method.

Let's use R to break down the data into each of our test sets. We will refer to 0.5mg as low, 1.0mg as medium and 2.0mg as high dose. We will calculate all the variables required for our confidence test.

```
library("dplyr")
```

```
## Warning: package 'dplyr' was built under R version 3.2.1
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
##
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
##
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
## conventions in variable names
## ld = low dose
## md = medium dose
## hd = high dose
## aa = ascorbic acid
## oj = orange juice
## len = length

set_ld_aa <- filter(ToothGrowth, supp == 'VC', dose == 0.5)
set_md_aa <- filter(ToothGrowth, supp == 'VC', dose == 1.0)
set_hd_aa <- filter(ToothGrowth, supp == 'VC', dose == 2.0)

set_ld_oj <- filter(ToothGrowth, supp == 'OJ', dose == 0.5)
set_md_oj <- filter(ToothGrowth, supp == 'OJ', dose == 1.0)
set_hd_oj <- filter(ToothGrowth, supp == 'OJ', dose == 2.0)

## work out set means and standard deviations
mean_len_ld_aa <- mean(set_ld_aa$len)
mean_len_md_aa <- mean(set_md_aa$len)
mean_len_hd_aa <- mean(set_hd_aa$len)
mean_len_ld_oj <- mean(set_ld_oj$len)
mean_len_md_oj <- mean(set_md_oj$len)
mean_len_hd_oj <- mean(set_hd_oj$len)

sd_len_ld_aa <- sd(set_ld_aa$len)
sd_len_md_aa <- sd(set_md_aa$len)
sd_len_hd_aa <- sd(set_hd_aa$len)
sd_len_ld_oj <- sd(set_ld_oj$len)
sd_len_md_oj <- sd(set_md_oj$len)
sd_len_hd_oj <- sd(set_hd_oj$len)
```

Since we want to calculate a 95% confidence level for the upper and lower levels of our means in each of our sets of ten we have referred to our t-table cross-referencing 95% confidence and degrees of freedom 9 (sample size - 1) to give a t distribution factor of 2.262.

Now we can continue building our lower and upper confidence intervals using the established methods.

In this case we are using the method as detailed on this website: <http://www.statisticshowto.com/how-to-construct-a-confidence-interval-from-data-using-the-t-distribution/>

```
tfactor <- 2.262

## now we will divide each std dev by sqrt of sample size
## in keeping with the method described at the website this is step 5 of 8 so we
## will refer to it as step5
step5_ld_aa <- sd_len_ld_aa / sqrt(10)
step5_md_aa <- sd_len_md_aa / sqrt(10)
step5_hd_aa <- sd_len_hd_aa / sqrt(10)
step5_ld_oj <- sd_len_ld_oj / sqrt(10)
step5_md_oj <- sd_len_md_oj / sqrt(10)
step5_hd_oj <- sd_len_hd_oj / sqrt(10)

## step 6 is to multiply step 5 by the tfactor
step6_ld_aa <- step5_ld_aa * tfactor
step6_md_aa <- step5_md_aa * tfactor
```

```

step6_hd_aa <- step5_hd_aa * tfactor
step6_ld_oj <- step5_ld_oj * tfactor
step6_md_oj <- step5_md_oj * tfactor
step6_hd_oj <- step5_hd_oj * tfactor

```

```

## Now we can work out each of the lower and upper end of the ranges
## of each confidence interval

```

```

## lower end of range subtract step6 value from mean
## upper end of range add step6 value to mean
lower_ld_aa <- mean_len_ld_aa - step6_ld_aa
upper_ld_aa <- mean_len_ld_aa + step6_ld_aa

```

```

lower_md_aa <- mean_len_md_aa - step6_md_aa
upper_md_aa <- mean_len_md_aa + step6_md_aa

```

```

lower_hd_aa <- mean_len_hd_aa - step6_hd_aa
upper_hd_aa <- mean_len_hd_aa + step6_hd_aa

```

```

lower_ld_oj <- mean_len_ld_oj - step6_ld_oj
upper_ld_oj <- mean_len_ld_oj + step6_ld_oj

```

```

lower_md_oj <- mean_len_md_oj - step6_md_oj
upper_md_oj <- mean_len_md_oj + step6_md_oj

```

```

lower_hd_oj <- mean_len_hd_oj - step6_hd_oj
upper_hd_oj <- mean_len_hd_oj + step6_hd_oj

```

```

## add our data to a data frame to plot the results for confidence intervals.
library(ggplot2)

```

```

## Warning: package 'ggplot2' was built under R version 3.2.2

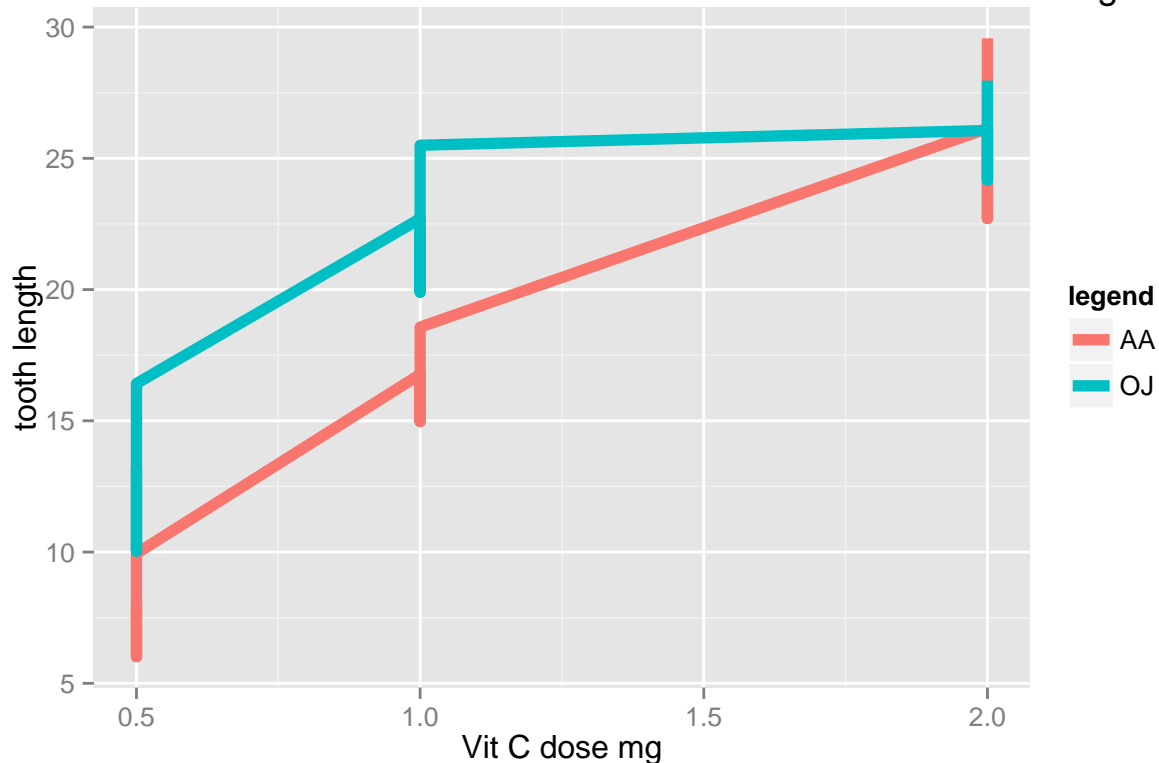
```

```

d <- data.frame(y = c(mean_len_ld_aa, mean_len_md_aa, mean_len_hd_aa, mean_len_ld_oj, mean_len_md_oj, mean_len_hd_oj),
x = c("aa", "oj"),
g <- ggplot(d, aes(x=x,y=y))
g <- g + geom_line(size = 2, aes(color = legend))
g <- g + ggtitle("T Conf intervals for vit c dose/method vs Tooth Growth in Guinea Pigs")
g <- g + xlab("Vit C dose mg")
g <- g + ylab("tooth length")
g

```

T Conf intervals for vit c dose/method vs Tooth Growth in Guinea Pigs



Conclusion and assumptions

We are assuming that doses of vitamin c are generally beneficial to tooth growth in guinea pigs. There is no data in this set for guinea pigs that have not been given supplements, so we can only work with the supplement data.

We are assuming that vitamin c does not have adverse health impacts on guinea pigs.

Our conclusion is:

- in low doses (0.5mg and 1.0mg) orange juice seems more effective at increasing tooth growth than ascorbic acid.
- from 1.0mg to 2.0mg there is only a small increase in the effectiveness of orange juice, showing a diminishing increase in tooth size.
- as the doses increase (2.0mg) we can see that the effectiveness of orange juice and ascorbic acid converge.
- as the doses increase across the range of doses, the 95% confidence intervals for orange juice appear to be narrowing and the confidence intervals for ascorbic acid appear to be expanding. It is not possible to say what this might mean given the small sample size, so more study might be needed with this dose and delivery method.