Analysing the Tooth Growth dataset

Reza

April 17, 2016

Synopsis

Vitamin C can affect the tooth growth. In this report we compare the effect of different doses of vitamin C, delivered in two methods on the tooth growth of guinea pigs. The sample are 60 guinea pigs. Each animal received 0.5, 1 or 2 mg/day vitamin C in the form of orange juice or ascorbic acid.

The study is done on the ToothGrowth dataset in R. The dataset contains 60 rows and 3 columns. Columns are "len", "supp" and dose which indicate the length of odontoblasts (cells responsible for tooth growth), supplement method ("OG" for orange juice and "VC" for ascorbic acid) and the daily dose, respectively.

Summary of the data

The aim would be comparing the average effect of different doses of different delivery methods. So lets look at how they look like. Fist we load the data and summarise it based on supp and dose.

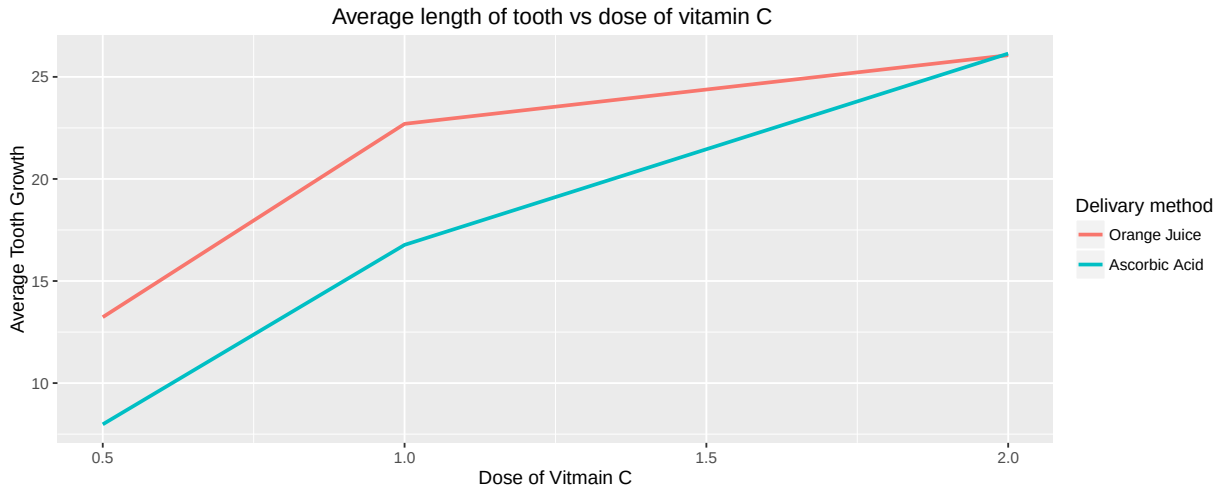
```
library(plyr)
data("ToothGrowth")
total <- ddply(ToothGrowth, .(supp, dose), summarise,
               Average_Growth = mean(len))
```

Lets look at the summarised table.

```
##   supp dose Average_Growth
## 1   OJ  0.5          13.23
## 2   OJ  1.0          22.70
## 3   OJ  2.0          26.06
## 4   VC  0.5           7.98
## 5   VC  1.0          16.77
## 6   VC  2.0          26.14
```

It seems that the orange juice method shows better effect in 0.5 and 1 mg/day dose but ascorbic acid method fills the gap in 2 mg/day dose. This is shown in the diagram below.

```
library(ggplot2)
p <- ggplot(total, aes(dose, Average_Growth, color = supp)) +
       geom_line(stat = "identity", size = 1)
p + labs( title= "Average length of tooth vs dose of vitamin C ", x = "Dose of Vitmain C", y = "Average
```



Comparing average tooth growth

In this section we will test our observations in the last section. Exactly we are going to test the null hypothesis

- H_0: The average tooth growth in a similar dosage by both delivery methos are equal

against alternative hyphothesis

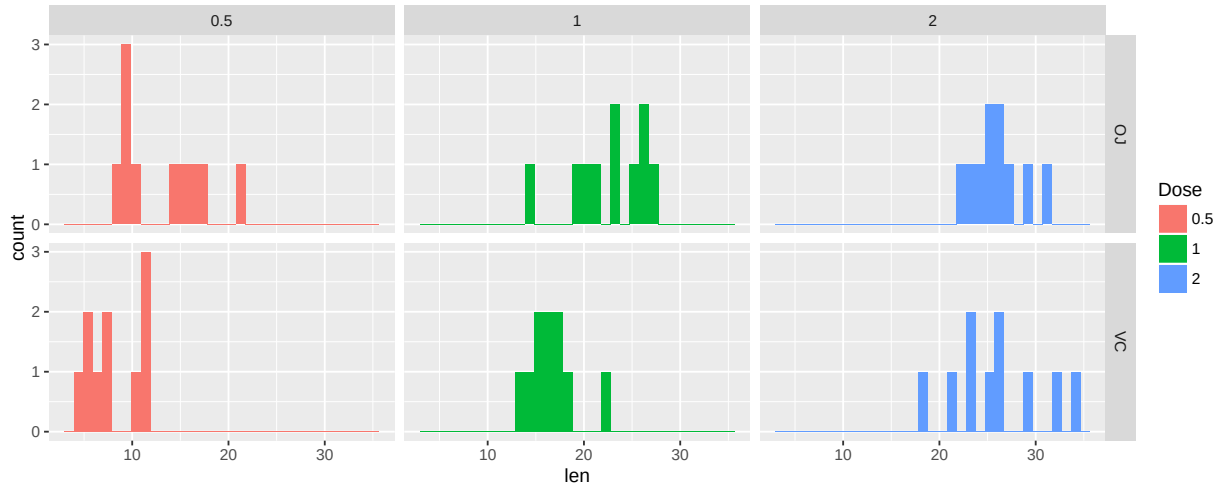
- H_1: The average tooth growth in a similar dosage by both delivery methos are not equal

This will be done with two methods. First t-confidence interval and then permutation tests.

t-confidence interval

Note that the number of samples are small (10 for each case) so t-test would be the parametric test candidate. Also looking at he histogram of len

```
qplot(len, data = ToothGrowth, geom = "histogram",facets = supp~dose, fill = as.factor(dose)) + scale_f
```



The distribution is acceptingly symmetric so we allow ourselves to use the t-test function in R. Here is the code for runnig the tests.

```
ttest <- function(d){
       t.test(ToothGrowth$len[ToothGrowth$supp == "OJ" & ToothGrowth$dose == d],
              ToothGrowth$len[ToothGrowth$supp == "VC" & ToothGrowth$dose == d], paired = F)
}
t <- lapply(c(.5,1,2), ttest)
```

For the dose of 0.5 mg/day, the average tooth growth by orange juice method is 13.23 and by ascorbic acid 7.98. Also the confidence interval is $(1.7190573, 8.7809427)$ which does not contain 0. Therefore we reject the null hypothesis and conclude that with dose of 0.5 mg/day the orange juice methos shows more effect.

For the dose of 1 mg/day the average tooth growth by orange juice method is 22.7 and by ascorbic acid 16.77. The confidence interval is $(2.8021482, 9.0578518)$ which again does not contain 0. Therefore we also reject the null hypothesis and conclude that the orange juice method in this dose had more effect too.

But when the dose is 2 mg/day, the story is different. The average tooth growth using orange juice and ascorbic acid are respectively 26.06 and 26.14 quite close. The confidence interval is $(-3.7980705, 3.6380705)$ containing 0. Therefore we fail to reject the null hypothesis and the effect of two methods in the 2 mg/day dose are probably similar.

Permutation tests

Let's assume no distribution for our data and run a non parametric permutation test.

On every dose we will change the grouping by delivery method (supp) ten thousand times. Then subtract the means of tooth growth assuming the new grouping, computing permutation statistics. If the cases where the permutation statistics are greater than the observed statistic (difference in the average len considering original grouping) is less than 2.5% or greater than 97.5%, then H_0 would be rejected by a 95% confidence interval.

The following is the code for running the test and showing the output

```
#function to compute permutation statistics based on new groupings
testStat <- function(w, g) mean(w[g == "OJ" ]) - mean(w[g == "VC"])
#computing observed statistic
observedstat <- sapply(c(.5,1,2), function(t) testStat(ToothGrowth$len[ToothGrowth$dose == t],
                        ToothGrowth$supp[ToothGrowth$dose == t]))
#computing permutation statistics
permutations <- lapply(c(.5,1,2), function(t) sapply(1:10000, function(i)
       testStat(ToothGrowth$len[ToothGrowth$dose == t],
                sample(ToothGrowth$supp[ToothGrowth$dose == t]))))
#pringting output into a table
data.frame(Dose = c(.5,1,2),
           "Observed Statistic" = observedstat,
           "Permutaion graeter than Observed%" = sapply(1:3, function(i) mean(permutations[[i]] > obser
```

```
##   Dose Observed.Statistic Permutaion.graeter.than.Observed.
## 1  0.5              5.25                            0.0027
## 2  1.0              5.93                            0.0006
## 3  2.0             -0.08                            0.5165
```

So the results of the permutation test confirms the result of the t-test.