Study the ToothGrowth Data in R

Gurmeet Singh

November 16, 2015

Overview

The report investigates the dataset ToothGrowth in R and uses confidence intervals and hypothesis tests to compare affect on tooth growth by dose and type of supplement. The documentation of the ToothGrowth dataset has this to say:

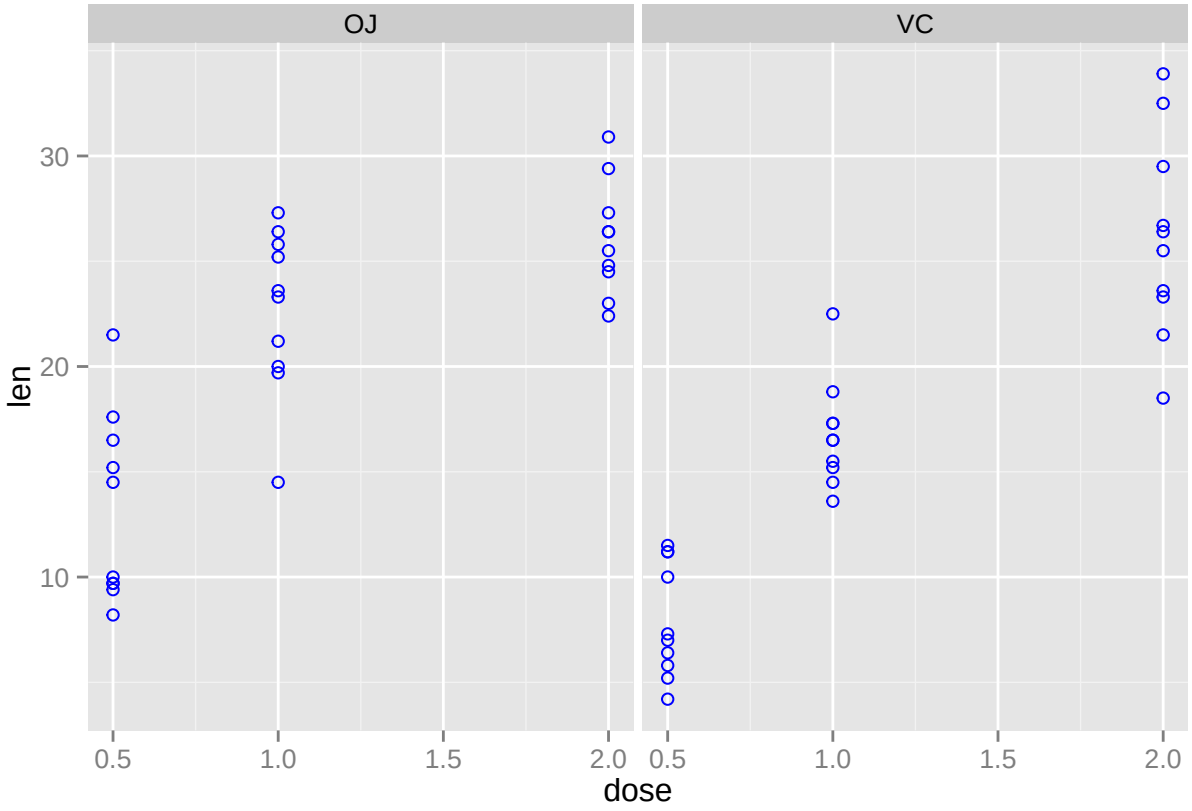
```
The Effect of Vitamin C on Tooth Growth in Guinea Pigs

Description: The response is the length of odontoblasts (cells responsible for tooth growth) in 60
guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by
one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC).

Format: 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/day
```

Data Exploration

Let's plot the data using ggplot2:



A graph is plotted showing the effect of vitamin C supplements on the growth of cells responsible for tooth growth. Without proceeding to any numerical analysis, conclusions can be drawn from it:

1. Both(VC,OJ) supplements affect the growth of cells. There is a positive correlation between dosage of each and the growth of cells.
2. At lower doses (0.5,1.0 mg/day), OJ seems to be more effective in increasing growth.
3. At 2.0 mg/day dosage level, it is unclear whether OJ or VC is more effective but clearly VC data has higher variance compared to the OJ data.

Statistical Analysis

Test 1

First, let's look at whether vitamin C in higher dosages causes increase in growth (compared to a lower dose) of tooth cells by using output of the T-test. Here, the null hypothesis H0: There's no effect of higher dosage.

```
t.test(dVC2p0,dVC0p5, var.equal = FALSE)
```

Here,

dVC2p0 = odontoblast cell lengths of guinea pigs given higher dose (2.0 mg/day) of vitamin C.

dVC0p5 = odontoblast cell lenghts of guinea pigs given lower dose (0.5 mg/day) of vitamin C.

```
##
##  Welch Two Sample t-test
##
## data:  dVC2p0 and dVC0p5
## t = 10.388, df = 14.327, p-value = 4.682e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  14.41849 21.90151
## sample estimates:
## mean of x mean of y
##     26.14      7.98
```

The confidence interval {14.418488, 21.901512} is clearly above zero and shows that the mean cell length is much greater with vitamin C in dosage of 2.0 mg/day compared to 0.5mg/day. The p value (rounded to 3 digits) of 0 of shows that the hypothesis H0 has very low probability of being correct, so it is rejected.

I've used `var.equal = FALSE` since there's nothing to suggest that the variance of the different data are equal since they may have been affected by the supplements differently.

Test 2

Now let's compare the effect of OJ and VC on tooth growth, again using output of this T-test. Here, the null hypothesis H0: There's no effect of using vitamin C as a supplement instead of Orange Juice.

```
t.test(dOJ2p0,dVC2p0, var.equal = FALSE)
```

Where dOJ2p0 and dVC2p0 are respectively the cell lengths of guinea pigs given 2.0 mg/day of OJ and vitamin C respectively.

The T-test output is:

```
##
##  Welch Two Sample t-test
##
## data:  dOJ2p0 and dVC2p0
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

The confidence interval {-3.7980705, 3.6380705} contains zero and is nearly symmetric around it. This shows that there's less than 95% confidence that mean cell lengths of either set are significantly different. The p value (rounded to 3 digits) of 0.964 of shows that the hypothesis H0 has high probability of being correct, so it is accepted, meaning that the effect of vitamin C is no more than that of orange juice.

Test 3

Similarly, we can make vitamin C vs Orange Juice comparison at dosage level of 0.5. The null hypothesis H0: mean of cell lengths with vitamin C dose of 0.5 mg/day is no larger than mean of cell lengths with orange juice.

The confidence interval here is {1.7190573, 8.7809427} and does not contain zero. This shows that there's more than 95% confidence that mean cell lengths of either set are higher than the other. The p value (rounded to 3 digits) of 0.006 shows that there's a low probability of null hypothesis being true and therefore vitamin C is more effective at increasing the cell legths than orange juice at dosage level 0f 0.5 mg/day.

All this is in agreement with what we saw earlier in the exploratory graph.

Conclusions

1. Increasing the vitamin C supplement dosage has statistically significant positive effect on the odontoblast cell lengths.
2. At lower dosages, orange juice is more effective, while at higher dosage of 2.0mg/day, vitamin C and orange juice have statistically similar effect on odontoblast cell lengths.

Appendix

The content of this report is authored in RStudio using R Markdown format and converted to PDF format using the **knitr** package. The R Markdown file itself can be found on GitHub

Plotting code

```
data(ToothGrowth)
library(ggplot2)

dVC0p5 <- ToothGrowth$len[ToothGrowth$supp == "VC" & ToothGrowth$dose == 0.5]
dVC1p0 <- ToothGrowth$len[ToothGrowth$supp == "VC" & ToothGrowth$dose == 1.0]
dVC2p0 <- ToothGrowth$len[ToothGrowth$supp == "VC" & ToothGrowth$dose == 2.0]
dOJ0p5 <- ToothGrowth$len[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 0.5]
dOJ1p0 <- ToothGrowth$len[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 1.0]
dOJ2p0 <- ToothGrowth$len[ToothGrowth$supp == "OJ" & ToothGrowth$dose == 2.0]

ggplot(data=ToothGrowth, aes(dose,len),colours=supp) + geom_point(col=4,pch=21) + facet_grid(. ~ supp)
```

Code for the T-tests and results' variables.

```
t.test(dOJ2p0,dVC2p0, var.equal = FALSE)
conf.int.vc.oj <- t.test(dOJ2p0,dVC2p0, var.equal = FALSE)$conf.int
p.vc.oj <- round(t.test(dOJ2p0,dVC2p0, var.equal = FALSE)$p.value,3)
```

```
conf.int.vc.oj.0p5 <- t.test(dOJ0p5,dVC0p5, var.equal = FALSE)$conf.int
p.vc.oj.0p5 <- round(t.test(dOJ0p5,dVC0p5, var.equal = FALSE)$p.value,3)
```

```
t.test(dVC2p0,dVC0p5, var.equal = FALSE)
conf.int.vc.vc <- t.test(dVC2p0,dVC0p5, var.equal = FALSE)$conf.int
p.vc.vc <- round(t.test(dVC2p0,dVC0p5, var.equal = FALSE)$p.value,3)
```