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.

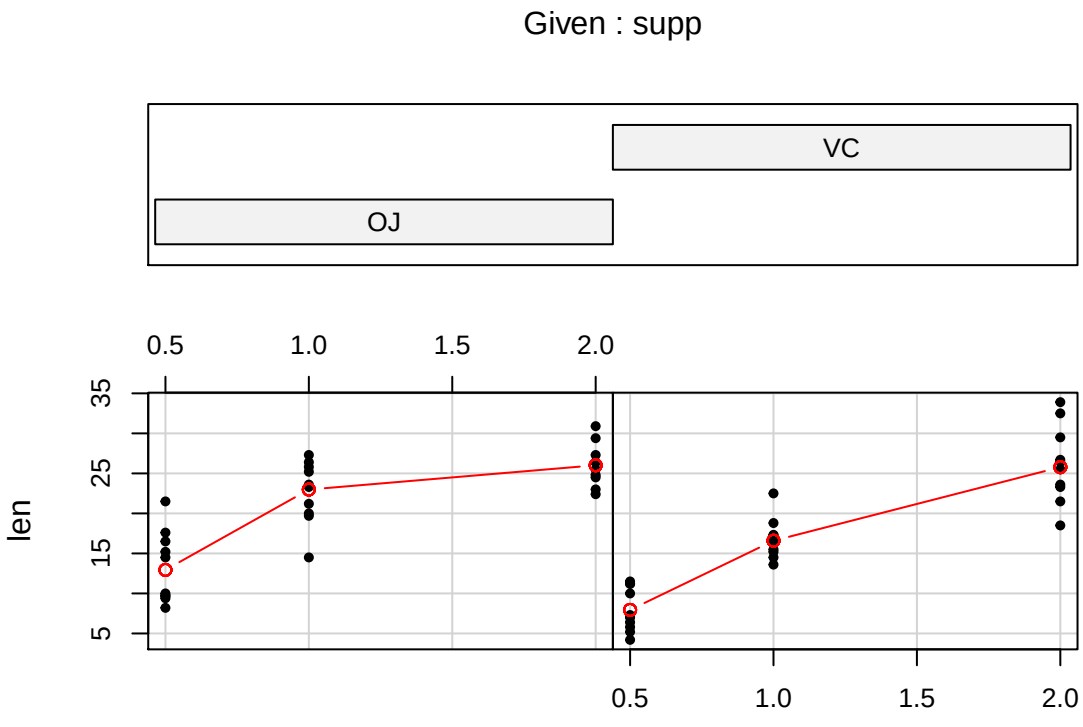
The documentation of the ToothGrowth dataset has this to say:

```
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



ToothGrowth data: length vs dose, given type of supplement

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

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:

```
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. 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.

Now let's compare the effect of OJ and VC on tooth growth, again using output of this T-test:

```
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 higher than the other. In fact, the confidence level is rather low. This is in agreement with what we saw earlier in the exploratory graph.

Conclusions

1. Increasing the supplement dosage has staistically significant positive effect on the odontoblast cell lengths.
2. At lower dosages, OJ is more effective, while at higher dosage of 2.0mg/day, VC and OJ have statistically similar affect 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 code used for the report can be found in the appendix and the R Markdown file itself can be found on GitHub

Code to generate an example of an exponential distribution.

```
data(ToothGrowth)

coplot(len ~ dose | supp,
       data = ToothGrowth,
       panel = panel.smooth,
       xlab = "ToothGrowth data: length vs dose, given type of supplement",
       type="b",
       pch=20)

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]
```

Code for runnning the simulation and displaying results.

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

```
t.test(dVC2p0,dVC0p5, var.equal = FALSE)
conf.int <- t.test(dVC2p0,dVC0p5, var.equal = FALSE)$conf.int
```