Class 6 - Course Project: Analysis of Tooth Growth

Nicholas Ng

Monday, January 26, 2015

Overview

This report was written as part of the course project requirements of class 6 of the Data Science specialisation, Statistical Inference. The aim of this report is to analyse the ToothGrowth dataset for efficacy of supplement types and amounts on guinea pig odontoblasts (teeth) length.

Exploratory Analysis

According to the vignette on the dataset, there are 60 observations on 3 variables, so to guide one on the next steps of the analysis, a summary table and quick exploratory scatterplot of the data is created:

```
data(ToothGrowth)
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
```

```
suppressMessages(require("ggplot2"))
overplot <- ggplot(ToothGrowth, aes(x = dose, y = len)) +
            geom_point() +
            facet_grid(. ~ supp)
```

As observed in the exploratory plot above, the length of teeth appear to increase with the dose of supplement, regardless of supplement used. Therefore, the following sections will more formally study if:

1. A higher dose of supplements does promote tooth growth, and
2. Which supplement is more effective.

Given that the number of observations is low, the t-test is used here for hypothesis testing.

Promotion of Tooth Growth via Higher Dosage

As observed in the previous plot, a higher dose of supplement is associated with a increased tooth growth. Therefore, the data is subsetted based on supplement used and tested based on whether the length of teeth in higher doses differs from the the lower doses.

Subsetting:

```
OJdat<-ToothGrowth[ToothGrowth$supp == "OJ",]
VCdat<-ToothGrowth[ToothGrowth$supp == "VC",]
```

Testing: (null hypothesis: len = 0 between doses)

```
suppressMessages(require("xtable"))
oja <- t.test(OJdat$len[OJdat$dose == 1], OJdat$len[OJdat$dose == 0.5],
              paired = F, var.equal = F)$conf
ojb <- t.test(OJdat$len[OJdat$dose == 2], OJdat$len[OJdat$dose == 1],
              paired = F, var.equal = F)$conf
vca <- t.test(VCdat$len[VCdat$dose == 1], VCdat$len[VCdat$dose == 0.5],
              paired = F, var.equal = F)$conf
vcb <- t.test(VCdat$len[VCdat$dose == 2], VCdat$len[VCdat$dose == 1],
              paired = F, var.equal = F)$conf
dosetable <- data.frame(null = c("OJ 0.5 = OJ 1", "OJ 1 = OJ 2",
                                 "VC 0.5 = VC 1", "VC 1 = VC 2"),
                        lower_interval = c(oja[1], ojb[1], vca[1], vcb[1]),
                        upper_interval = c(oja[2], ojb[2], vca[2], vcb[2]))
print(xtable(dosetable), comment = F, include.rownames = F)
```

null	lower_interval	upper_interval
OJ 0.5 = OJ 1	5.52	13.42
OJ 1 = OJ 2	0.19	6.53
VC 0.5 = VC 1	6.31	11.27
VC 1 = VC 2	5.69	13.05

As seen from the table above, none of the intervals contain 0. Hence one may conclude that one may reject the null hypothesis that higher doses do not cause tooth growth, at the 95% confidence level.

Efficacy of Supplement in Promotion of Tooth Growth

To determine which supplement is more effective in promoting tooth growth, the distribution of the length of teeth is tested by supplement and dose to observe if the lengths differ significantly from each group.

```
suppa <- t.test(OJdat$len[OJdat$dose == 0.5], VCdat$len[VCdat$dose == 0.5],
                paired = FALSE, var.equal = FALSE)$conf
suppb <- t.test(OJdat$len[OJdat$dose == 1], VCdat$len[VCdat$dose == 1],
                paired = FALSE, var.equal = FALSE)$conf
suppc <- t.test(OJdat$len[OJdat$dose == 2], VCdat$len[VCdat$dose == 2],
                paired = FALSE, var.equal = FALSE)$conf
supptable <- data.frame(null = c("OJ 0.5 = VC 0.5", "OJ 1 = VC 1",
                                 "OJ 2 = VC 2"),
                        lower_interval = c(suppa[1], suppb[1], suppc[1]),
                        upper_interval = c(suppa[2], suppb[2], suppc[2]))
print(xtable(supptable), comment = F, include.rownames = F)
```

As seen in the table table, at low doses (first 2 rows), 0 is not included in the interval, hence one may reject the null hypothesis that there is no difference between the supplements. Additionally, given that the interval is positive, one may conclude that at low doses, orange juice is more effective in promoting tooth growth.

null	lower_interval	upper_interval
OJ 0.5 = VC 0.5	1.72	8.78
OJ 1 = VC 1	2.80	9.06
OJ 2 = VC 2	-3.80	3.64

However, when the dose is 2 mg, the interval contains 0, vis-a-vis, one fails to reject the null hypothesis of insignificance difference at the 95% confidence level. Therefore, we may conclude that there is insufficient evidence to determine which of the two supplements is more effective than the other in promoting tooth growth.

Conclusion and Assumptions

Therefore, based on the aforementioned results, higher doses of vitamin C promotes tooth growth in guinea pigs, with orange juice being better as a supplement than ascorbic acid at lower doses.

The conclusions above however, are reached under the assumption that variances are not constant across test groups (by supplement, as seen by the spread of the data in the exploratory table), and that test subjects are independent and not paired (the data did not contain subject identifiers for pairing).