ToothGrowth Data Analisys

Oleg Tsarev

OVERVIEW

In this report we will investigate the effect of vitamin C on tooth growth in Guinea Pigs. 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) - in our dataset column "dose" with each of two delivery methods - orange juice or ascorbic acid - column "supp". https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/ToothGrowth.html.

LOADING THE TEETH GROWTH DATA AND SHORT SUMMARY

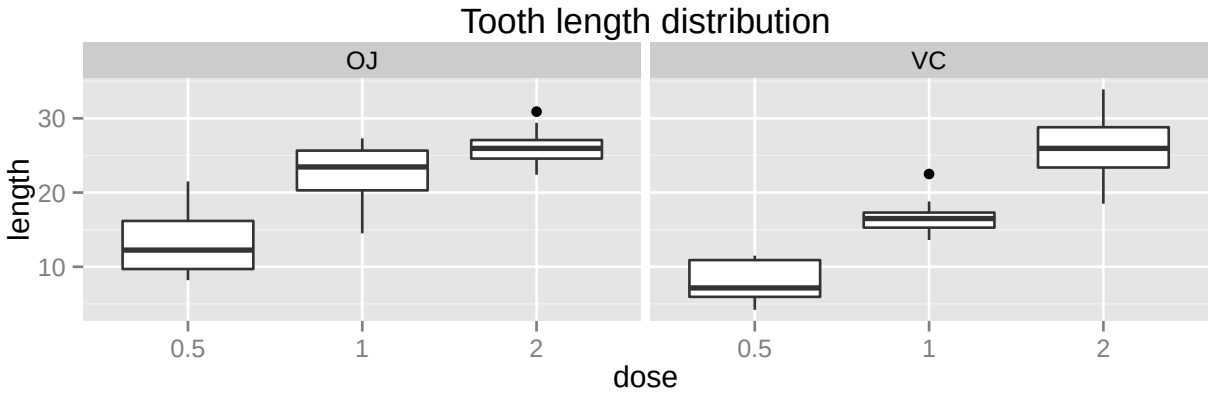
```
data(ToothGrowth)
summary(ToothGrowth)
```

```
##       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
```

```
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 ...
```

```
library(ggplot2)
q <- ggplot(data = ToothGrowth, aes(x = factor(dose), y = len))
q <- q + geom_boxplot() + facet_grid(. ~ supp)
q + ggtitle('Tooth length distribution') + labs(x = "dose", y = "length")
```



So, as you can see above, seems that with dose 0.5 and 1 mg orange juice average teeth length is higher than ascorbid acid (same doses), but with dose 2 mg average teeth lengths are about the same. Let's investigate it.

CONFIDENCE INTERVALS AND HYPOTHESIS TESTS

Let's use confidence intervals and hypothesis tests to compare tooth growth by supp and dose. To do so we create matrix of Two Sample t-tests (code below in the Appendix). We assume that the observations are statistically independent, so we use an independent group Student's T interval (not a paired one) and use 5% cut-off mark for rejection of the null hypothesis.

Tooth growth comparison by supp

```
library(knitr)
opts_chunk$set(size = 'large')
mtrx_test["OJ  all","VC  all"][[1]] # t-test (not a paired) with all doses of supp
```

```
##
##  Welch Two Sample t-test
##
## data:  q1$len and q2$len
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

The null hypothesis here is that the difference between the mean of tooth length with orange juice (all doses) and the mean of tooth length with ascorbic acid (all doses) is zero. As you can see above p-value $> 5\%$, therefore we fail to reject the null hypothesis. We can say that with 95% probability we sure that the difference between mentioned means (20.66 and 16.96) is between -0.17 and 7.57. There are a negative values, so we can not say that orange juice **always** is more effective in terms of tooth growth than ascorbic acid.

Tooth growth comparison by supp and with dose 0.5 mg

```
mtrx_test["OJ  0.5","VC  0.5"][[1]] # t-test (not a paired) with dose 0.5 mg
```

```
##
##  Welch Two Sample t-test
##
## data:  q1$len and q2$len
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean of x mean of y
##     13.23      7.98
```

The null hypothesis here is that the difference between the mean of tooth length with orange juice (0.5 mg) and the mean of tooth length with ascorbic acid (0.5 mg) is zero. As you can see above p-value $< 5\%$, therefore we reject the null hypothesis. We can say that with 95% probability we sure that the difference between mentioned means (13.23 and 7.98) is between 1.72 and 8.78. Therefore we can say that with dose 0.5 mg orange juice is more effective in terms of tooth growth than ascorbic acid.

Tooth growth comparison by supp and with dose 1 mg

```
mtrx_test["OJ  1  ","VC  1  "][[1]] # t-test (not a paired) with dose 1 mg
```

```
##
##  Welch Two Sample t-test
##
## data:  q1$len and q2$len
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean of x mean of y
##     22.70     16.77
```

The null hypothesis here is that the difference between the mean of tooth length with orange juice (1 mg) and the mean of tooth length with ascorbic acid (1 mg) is zero. As you can see above p-value $< 5\%$, therefore we reject the null hypothesis. We can say that with 95% probability we sure that the difference between mentioned means (22.70 and 16.77) is between 2.80 and 9.06. Therefore we can say that with dose 1 mg orange juice is more effective in terms of tooth growth than ascorbic acid.

Tooth growth comparison by supp and with dose 2 mg

```
mtrx_test["OJ  2  ","VC  2  "][[1]] # t-test (not a paired) with dose 2 mg
```

```
##
##  Welch Two Sample t-test
##
## data:  q1$len and q2$len
## 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 null hypothesis here is that the difference between the mean of tooth length with orange juice (2 mg) and the mean of tooth length with ascorbic acid (2 mg) is zero. As you can see above p-value $>$ 5%, therefore we fail to reject the null hypothesis. We can say that with 95% probability we sure that the difference between mentioned means (26.06 and 26.14) is between -3.80 and 3.64, so the interval is almost symetric around zero and we can say that orange juice is about the same effective in terms of tooth growth as ascorbic acid.

CONCLUSIONS AND NEEDED ASSUMPTIONS

Conclusions:
With doses 1 mg and 0.5 mg orange juice is more effective in terms of tooth growth than ascorbic acid.
With dose 2 mg orange juice is about the same effective in terms of tooth growth as ascorbic acid.
Assumptions:
a) Observations of Tooth Growth are statistically independent;
b) These observations are unpaired.
c) Large doses are more effective in terms of tooth growth.

APPENDIX

```
# create matrix as a 12x12 array of lists - to store t-test results
mtrx_test <- array(list(),c(12,12),dimnames = list(1:12,1:12))
# create special matrix as a 12x12 array of lists - to store p-values
p.v <- array(numeric(),c(12,12),dimnames = list(1:12,1:12))
# create columns and rows of mentioned matrixes
k <- 1
for (i in c("VC ","OJ ","all")){
    for (j in c("0.5","1  ","2  ","all")){
        colnames(mtrx_test)[k] <- paste(i,j)
        rownames(mtrx_test)[k] <- paste(i,j)
        colnames(p.v)[k] <- paste(i,j)
        rownames(p.v)[k] <- paste(i,j)
        k <- k +1
    }
}
# fill matrixes with t-tests results
for (i in 1:nrow(mtrx_test)){
    for (j in 1:ncol(mtrx_test)){
        q1_s <- gsub(" ","",substr(rownames(mtrx_test)[i],1,3))
        q1_d <- gsub(" ","",substr(rownames(mtrx_test)[i],5,nchar(rownames(mtrx_test)[i])))
        if(q1_s == "all") q1_s <- c("VC","OJ")
        if(q1_d == "all") q1_d <- c("0.5","1","2")
        q2_s <- gsub(" ","",substr(colnames(mtrx_test)[j],1,3))
        q2_d <- gsub(" ","",substr(colnames(mtrx_test)[j],5,nchar(colnames(mtrx_test)[j])))
        if(q2_s == "all") q2_s <- c("VC","OJ")
        if(q2_d == "all") q2_d <- c("0.5","1","2")
        q1 <- ToothGrowth[ToothGrowth$supp %in% q1_s & ToothGrowth$dose %in% q1_d,]
        q2 <- ToothGrowth[ToothGrowth$supp %in% q2_s & ToothGrowth$dose %in% q2_d,]
        # t.test results are stored in matrixes
        mtrx_test[i,j][[1]] <- t.test(q1$len,q2$len,paired = F)
        p.v[i,j] <- sprintf("%3.2e",mtrx_test[i,j][[1]]$p.value)
    }
}
# delete redundant variables
rm(list=ls()[! ls() %in% c("mtrx_test","ToothGrowth","p.v")])
```