Orange juice or vitamin C - what's a key to sabre-toothed guinea pigs?

Ivana Lipnerová

December 25, 2015

Report

Overview

This study analyse the ToothGrowth data in the R datasets package (McNeil D. R., 1977)[1], aiming for comparing benefits of two diets on growth of teeth. The goals are to provide simple exploratory data analysis, compare two diets and draw a conclusion based on performed tests.

[1] McNeil, D. R. (1977) *Interactive Data Analysis*. New York: Wiley.

Loading the data

```
# Load libraries
library(dplyr); library(ggplot2); library(gridExtra); library(knitr)

# Load dataset
library(datasets)
data(ToothGrowth)
a <- (ToothGrowth) #give it easily typed name
```

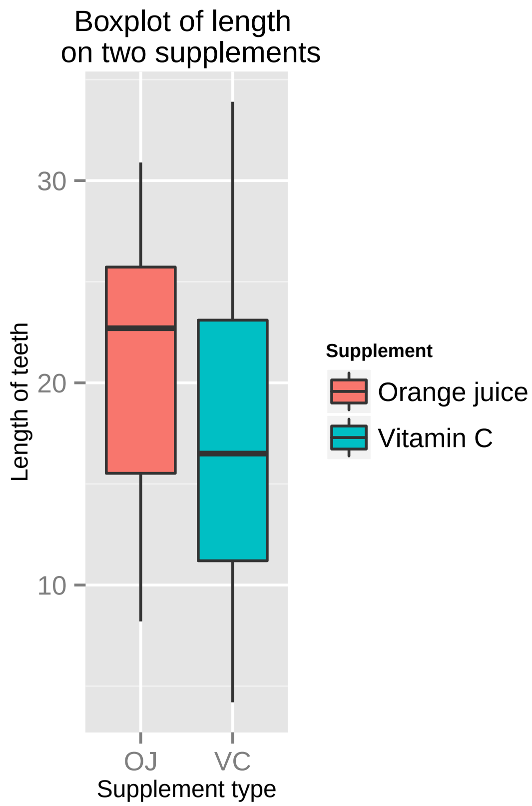
Basic exploratory data analyses

The data contains measurements of the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Animals were divided to 6 groups per 10 pigs, each group was fed with one of two supplements, orange juice (OJ) or ascorbic acid (vitamin C, VC), in one of three doses (0.5, 1, or 2 mg/day). The output data has format of 60 observations and 3 variables - **len** for length of odontoblasts, **sup** for supplement type, and **dose** for dose recieved.

```
str(a) #structure of data
```

```
## '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 ...
```

Boxplots of the data showing the differences and distributions of lengths of odontoblasts for each supplement on the left and for each supplement per dose on the right (see Appendix for code):



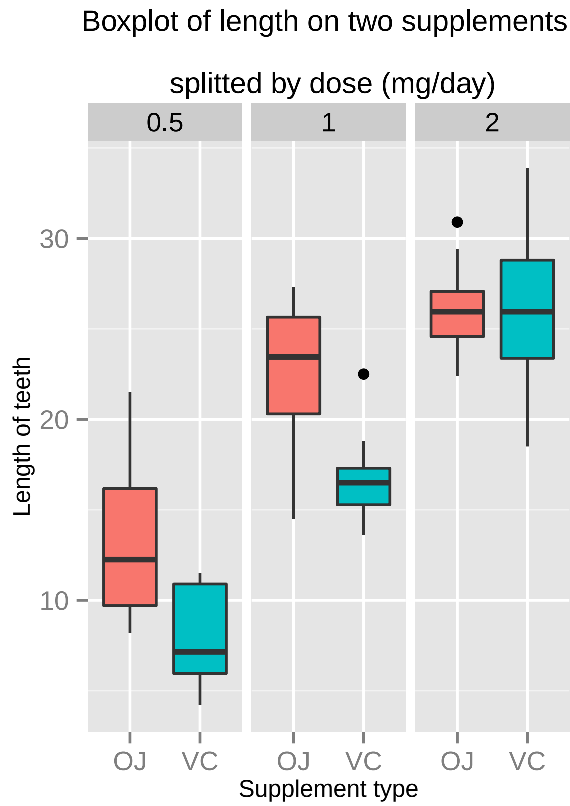


Table of means of lengths of odontoblasts for each dose of both supplements:

```
a_tab<-group_by(a, dose, supp) %>% summarise(mean(len))
kable(x=a_tab, digits=1, align="c")
```

dose	supp	mean(len)
0.5	OJ	13.2
0.5	VC	8.0
1.0	OJ	22.7
1.0	VC	16.8
2.0	OJ	26.1
2.0	VC	26.1

Testing

I decide to test the difference in teeth growth between supplementing orange juice (OJ) and vitamin C (VC) via two sample t-tests (R function *t-test*). Based on EDA I suppose that orange juice gives better results.

Assumptions for testing:

- Guinea pigs were selected randomly and each combination of dose and supplement fed to different group of pigs, thus they form independent groups (t.tests default *paired=FALSE* is used)
- Lengths of odontoblasts have normal distribution for tested sets of data (see Appendix for normality tests)

- Different variances for each dataset are assumed, as viewed as less harmful then doing otherwise (t.tests default *var.equal=FALSE* is used)

```
# Prepare table of results:
test_res <- NULL

# Test OJ vs VC overall:
test_res<-cbind(t.test(len ~ factor(supp), data=a, paired=F)$p.value,
    t.test(len ~ factor(supp), data=a, paired=F)$conf.int[1],
    t.test(len ~ factor(supp), data=a, paired=F)$conf.int[2])

# Test OJ vs VC for each dose:
for (i in c(0.5, 1, 2)) { #loop for obtaining just wanted numbers
  test_res <- rbind(test_res,
              cbind(t.test(len ~ factor(supp), data=a[a$dose== i, ], paired=F)$p.value,
                    t.test(len ~ factor(supp), data=a[a$dose== i, ], paired=F)$conf.int[1],
                    t.test(len ~ factor(supp), data=a[a$dose== i, ], paired=F)$conf.int[2]
                    ))
  }

# Format table of results:
test_res<-as.data.frame(test_res)
row.names(test_res)<-c("OJ vs VC overall", "OJ vs VC - dose 0.5",
              "OJ vs VC - dose 1.0", "OJ vs VC - dose 2.0")
colnames(test_res)<-c("p-value", "95% CI (-)", "95% CI (+)")

# Print table of results:
kable(test_res, digits=3, row.names=T, align="c",
      caption="Summary table of test results")
```

Table 2: Summary table of test results

	p-value	95% CI (-)	95% CI (+)
OJ vs VC overall	0.061	-0.171	7.571
OJ vs VC - dose 0.5	0.006	1.719	8.781
OJ vs VC - dose 1.0	0.001	2.802	9.058
OJ vs VC - dose 2.0	0.964	-3.798	3.638

Conclusion

There is no statistically significant overall difference between orange juice and vitamin C, however for lower doses (0.5 mg/day and 1 mg/day) the orange juice has more prominent effect on growth of guinea pigs odontoblasts. High dose of 2 mg/day is sufficient to achieve maximum odontoblasts growth no matter the supplement recieved. It also suggests that such a dose of vitamin C is high enough to compensate for whichever advantage the orange juice has at lower dose levels.

So, to increase chance of creating a breed of sabre-toothed guinea pigs, one should feed them up with either orange juice or ascorbic acid.

Appendix

Boxplots code

Code for creating boxplots on page 2:

```
# Overall boxplot:
p1 <- ggplot(a, aes(x=supp, y=len)) + geom_boxplot(aes(fill=supp)) +
  labs(title="Boxplot of length \n on two supplements",
       x="Supplement type",
       y="Length of teeth") +
  scale_fill_discrete(name="Supplement",
                      labels=c("Orange juice", "Vitamin C")) +
  theme(title = element_text(size = rel(0.75), hjust = 0.5))

# Dose boxplot:
p2 <- ggplot(a, aes(y=len, x=as.factor(supp))) +
  geom_boxplot(aes(fill=supp)) +
  guides(fill=FALSE) +
  facet_grid(.~dose) +
  labs(title="Boxplot of length on two supplements
       \n splitted by dose (mg/day)",
       x="Supplement type",
       y="Length of teeth") +
  theme(title = element_text(size = rel(0.75), hjust = 0.5))

# Arrange them:
grid.arrange(p1, p2, ncol=2, nrow=1)
```

Normality testing

I utilize the Shapiro-Wilk test of normality to test whether the data follow normal distribution (or at least are not so far from normality). Note that for Shapiro-Wilk test of normality the p-value lower than 0.01 indicates non-normal distribution of tested data.

```
# Crude code for generating table:
shap_tests<-as.data.frame(shapiro.test(a$len[a$supp=="OJ" & a$dose == 0.5])[c(4, 2)])
shap_tests<-rbind(shap_tests, as.data.frame(shapiro.test(a$len[a$supp=="OJ" &
                                                                  a$dose == 1])[c(4, 2)]))
shap_tests<-rbind(shap_tests, as.data.frame(shapiro.test(a$len[a$supp=="OJ" &
                                                                  a$dose == 2])[c(4, 2)]))
shap_tests<-rbind(shap_tests, as.data.frame(shapiro.test(a$len[a$supp=="VC" &
                                                                  a$dose == 0.5])[c(4, 2)]))
shap_tests<-rbind(shap_tests, as.data.frame(shapiro.test(a$len[a$supp=="VC" &
                                                                  a$dose == 1])[c(4, 2)]))
shap_tests<-rbind(shap_tests, as.data.frame(shapiro.test(a$len[a$supp=="VC" &
                                                                  a$dose == 2])[c(4, 2)]))
shap_tests<-rbind(shap_tests, as.data.frame(shapiro.test(a$len[a$supp=="OJ"])[c(4, 2)]))
shap_tests<-rbind(shap_tests, as.data.frame(shapiro.test(a$len[a$supp=="VC"])[c(4, 2)] ))

kable(shap_tests, digits = 3, align="l", col.names = c("Tested dataset", "p-value"))
```

Tested dataset	p-value
a*len*[*a*supp == “OJ” & a$dose == 0.5]	0.182
a*len*[*a*supp == “OJ” & a$dose == 1]	0.415
a*len*[*a*supp == “OJ” & a$dose == 2]	0.815
a*len*[*a*supp == “VC” & a$dose == 0.5]	0.170
a*len*[*a*supp == “VC” & a$dose == 1]	0.270
a*len*[*a*supp == “VC” & a$dose == 2]	0.919
a*len*[*a*supp == “OJ”]	0.024
a*len*[*a*supp == “VC”]	0.428

All datasets, except dataset containing whole OJ related data, are insignificant, thus normally distributed. The exception is only weakly significant and therefore taken as approximately normal for purpose of t-test.