Tooth Growth Analysis

Mariusz Musiał

22 listopada 2015

Overview

In this paper we report the results of the analysis of growth of odontoblasts in the groups of 10 guinea pigs, each supplied with three different doses of Vitamic C and two different delivery methods. We will explore the data availiable in ToothGrowth data set, as well as formulate, test and briefly discuss a few hypoteses based on this exploration.

Summary of the dataset

The data we are going to analyse is available in ToothGrowth data frame. After loading relevant libraries and the data frame itselt, we'll check the structure of the data (see code chunk 'str' in Appendix):

```
## dose :  num [1:60] 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
## len :  num [1:60] 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 ...
```

From the output of the `ls.str` function we can see, that this data frame consists of 60 observations of 3 variables:

- `len` - numerical variable representing the length of odontoblasts, in μm,
- `supp` - factor variable with two levels, representing the delivery method,
- `dose` - numerical variable representing vitamin C dose level in mg.

Delivery method is coded as "OJ", meaning Orange Juice and "VC", meaning ascrobic acid.

The `dose` variable needs some further invetigation, namely what levels are present in the data frame (see code chunk 'unique' in Appendix).

```
## [1] 0.5 1.0 2.0
```

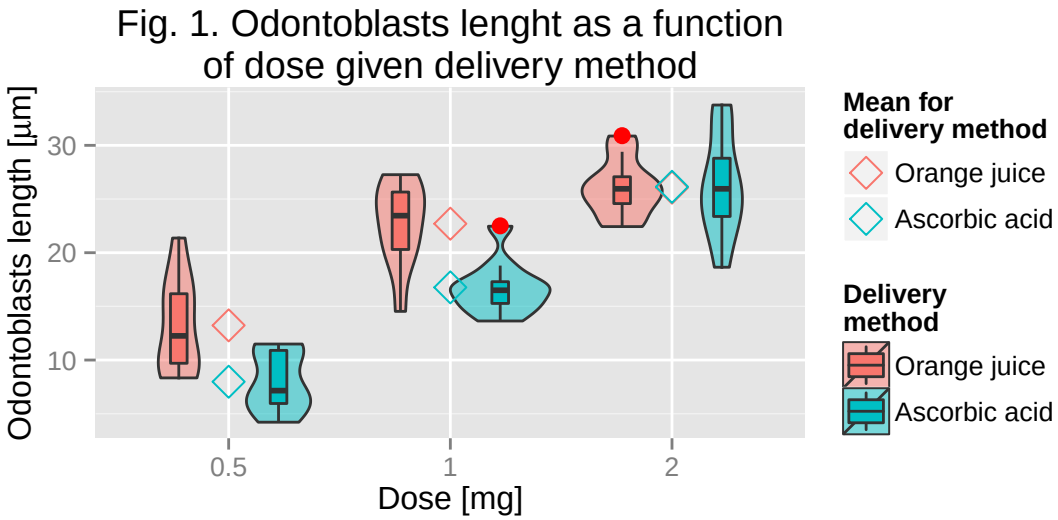
We can see, that there are three different dose levels in the data: 0.5 mg, 1 mg and 2 mg.

It is worth to check how many observations we have for each delivery method and dose level (see code chunk 'table' in Appendix):

```
##     dose
## supp 0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

The above table indicates that we have 10 observations for each combination of delivery method and dose level, summing up to a total of 60 observations.

Having verified the structure of the data we proceed to visualise it in the form of combined violin and box plot, what will allow for better understanding of the influence of dose levels and delivery methods on odontoblasts length. Code for generating this plot is provided in the Appendix as code chunk 'boxviolin'.



Figire 1 reveals possible direct proportional relationship between dose level and odontoblasts length for both delivery methods with clearer advantage of OJ for dose levels of 0.5 mg and 1 mg and less clear for 2 mg dose. Furthermore, it seems that there could be no benefit of increasing the dose beyond 1 mg for OJ delivery method.

Both medians and means appear to follow the same relationship to dose and method. From that figure we can also appreciate the approximate distribution of odontoblasts lenght data for each case and the fact of having outliers for two combinations of dose/delivery, marked with red dots.

Analysis of odontoblasts lenght in function of Vitamim C dose and delivery method

The analysis of data will consist of two groups. First we test the claim that there is advantage of delivery of vitamin C by means of Orange Juice over ascrobic acid for all the doses.

In the second part we test the claim that there is no statistically significant difference in influence on mean odontoblast lenght for 2 mg dose of orange juice over 1 mg dose.

For the purpose of the analysis and posterior discussion we assume the following:

- There is approximately normal distributions of data in each group
- The variances are unequal in each tested group (possibly due to unequal supply of vitamin C)
- Other diet-related variables are kept constant in all groups
- Uncontrolled variables are randomized across the groups

The following table summarizes the tests.

Table 1: Test group 1.

Dose	Analysis	Null hypothesis (regarding odontoblast length)	Results
0.5 mg	lenght ~ delivery method	The true difference of means between OJ and VC is 0	Table 3
1 mg	lenght ~ delivery method	The true difference of means between OJ and VC is 0	Table 4
2 mg	lenght ~ delivery method	The true difference of means between OJ and VC is 0	Table 5

Table 2: Test group 2.

Delivery	Analysis	Null hypothesis (regarding odontoblast length)	Results
OJ	lenght ~ dose (1 - 2 mg)	The true difference of means between 1 and 2 mg is 0	Table 6

Results of tests (see code chunk 'ttest' in Appendix):

Table 3: Welch Two Sample t-test: `len` by `supp`

Test statistic	df	P value	Alternative hypothesis
3.17	14.97	*0.006359* * *	two.sided

Table 4: Welch Two Sample t-test: `len` by `supp`

Test statistic	df	P value	Alternative hypothesis
4.033	15.36	*0.001038* * *	two.sided

Table 5: Welch Two Sample t-test: `len` by `supp`

Test statistic	df	P value	Alternative hypothesis
-0.04614	14.04	*0.9639*	two.sided

Table 6: Welch Two Sample t-test: `len` by `dose`

Test statistic	df	P value	Alternative hypothesis
-2.248	15.84	*0.0392* *	two.sided

Results and conclusions

The p-values of the t tests presented in Tables 3 and 4 are smaller than $\alpha = 0.05$ so we reject the null hypotheses. Therefore we conclude that for doses of 0.5 and 1 mg there is advantage of delivering vitamin C by means of orange juice for the growth of odontoblasts. There is however no proof in the data that this delivery method for dose of 2 mg has advantages over delivering ascrobic acid, as seen in table 5 (where the p-value is significantly grater than 0.05, thus we failed to reject the null hypothesis).

The result presented in Table 6 shows, that we reject the null hypothesis stating that there is no statistically significant difference in means of odontoblast lengths between 1 mg and 2 mg doses of orange juice (p-value less than 0.05). If this has practical importance remains to be seen. Please refer to Appendix for CI review for this case.

Appendix

Orange Juice dose 1 mg vs 2 mg - CI review

Based on the results presented in Table 6 we rejected the null hypothesis stating that there is no statistically significant difference in means of odontoblast lengths between 1 mg and 2 mg doses of orange juice, due to its p-value being less than 0.05. That would mean that supplying doses greater than 1 mg of vitamin C by means of orange juice might have practical impacts as to the benefits to the health of the subject.

Let's investigate the t test result for this case in more detail.

```
tt <- t.test(len ~ dose, data = ToothGrowth,
                    subset = (ToothGrowth$dose != .5 & ToothGrowth$supp == 'OJ'))
tt
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -2.2478, df = 15.842, p-value = 0.0392
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -6.5314425 -0.1885575
## sample estimates:
## mean in group 1 mean in group 2
##           22.70           26.06
```

It can be seen that the p-value is only marginal smaller than the limit of 0.05. We should therefore also analyse the Confidence Interval to assess whether there are any implications of alternative hypothesis from practical importance point of view.

The 95% CI for the difference in the means of odontoblast length, when supplied with the doses of 1 and 2 mg of orange juice ranges from **-6.53** to **-0.19**. As the lower difference limit in means (absolute value) is close to 0, we cannot state, based on the data, that the difference in doses has, or has not, any practical importance.

For more in-depth explanation of statistical significance versus practical importance see Jerry Dallal's The Little Handbook of Statistical Practice.

Code chunk for printing the structre of ToothGrowth data frame

```
# Code chunk 'str'
library(datasets)
library(ggplot2)
library(pander)
data("ToothGrowth")

stru <- ls.str(ToothGrowth)
```

Code chunk for determining number of doses in the data frame

```
# Code chunk 'unique'
uni <- unique(ToothGrowth$dose)
```

Code chunk for verification of number of observations per combination of dose and delivery method

```
# Code chunk 'table'
tab <- table(ToothGrowth[-1])
```

Code chunk for plotting combined box and violin plot

```
# Code chunk 'boxviolin'
ge <- ggplot(ToothGrowth, aes(factor(dose), len))
ge <- ge + geom_violin(aes(fill = factor(supp)), alpha =.5)
ge <- ge + geom_boxplot(aes(fill = factor(supp)), width = .2,
                        position = position_dodge(.9), outlier.colour = "red",
                        outlier.size = 3)
ge <- ge + xlab("Dose [mg]")
ge <- ge + ylab(expression(paste("Odontoblasts length [", mu,"m]")))
ge <- ge +
    ggtitle("Fig. 1. Odontoblasts lenght as a function\nof dose given delivery method")
ge <- ge + stat_summary(aes(color = supp), fun.y=mean, geom="point", shape=5, size=4)
ge <- ge + scale_color_discrete(name = "Mean for\ndelivery method",
                                breaks=c("OJ", "VC"),
                                labels=c("Orange juice", "Ascorbic acid"))
ge <- ge + scale_fill_discrete(name = "Delivery\nmethod", breaks=c("OJ", "VC"),
                               labels=c("Orange juice", "Ascorbic acid"))
```

Code chunk for generating tables with the results of t tests

```
# Code chunk 'ttest'
p1 <- pander(t.test(len ~ supp, data = ToothGrowth,
                    subset = (ToothGrowth$dose == .5)))
p2 <- pander(t.test(len ~ supp, data = ToothGrowth,
                    subset = (ToothGrowth$dose == 1)))
p3 <- pander(t.test(len ~ supp, data = ToothGrowth,
                    subset = (ToothGrowth$dose == 2)))

p4 <- pander(t.test(len ~ dose, data = ToothGrowth,
                    subset = (ToothGrowth$dose != .5 & ToothGrowth$supp == 'OJ')))
```