Tooth growth analysis

Jan Herman

October 25, 2015

Overview

This short analysis concerns the dataset `ToothGrowth` – one of the standard datasets shipped with R. The dataset contains information about tooth growth described by length of odontoblasts (cells responsible for tooth growth; measured in μm) 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).

We will perform basic exploratory analysis of the data and then tests hypotheses about an impact of supplement types and different doses on odontoblasts lenghts.

Exploratory analysis

Throughout the analyses, we will need three R packages – `dplyr` for data manipulation, `ggplot2` for figure plotting and `knitr` for making pdf output (and displaying summary tables by the `kable()` function). We will load them now.

```r
library(dplyr, warn.conflicts = FALSE)
library(ggplot2)
library(knitr)
```

Next we load the dataset, rename the supplement factor levels and do basic summary.

```r
data(ToothGrowth)
ToothGrowth <- ToothGrowth %>%
    mutate(supp = factor(supp, labels = c("Orange juice", "Ascorbic acid")))
summary(ToothGrowth)
```

```
##       len                   supp         dose      
##  Min.   : 4.20   Orange juice :30   Min.   :0.500  
##  1st Qu.:13.07   Ascorbic acid: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  
```

Then we plot the first (and only) exploratory graph. Since throughout the analyses we don't have any ambitions to uncover a quantitative relationship between dose and odontoblasts length, we treat dose as an (ordered) factor, rather than as a numeric value.

```r
ggplot(data = ToothGrowth, aes(x = factor(dose), y = len)) +
    geom_point(size = 4, shape = 21) +
    facet_grid(. ~ supp) +
    ggtitle("Odontoblasts length by dose and supplement") +
```

```
    labs(x = "Dose [mg]",
         y = expression(paste("Odontoblasts length [", mu, "m]"))) +
    theme_bw()
```

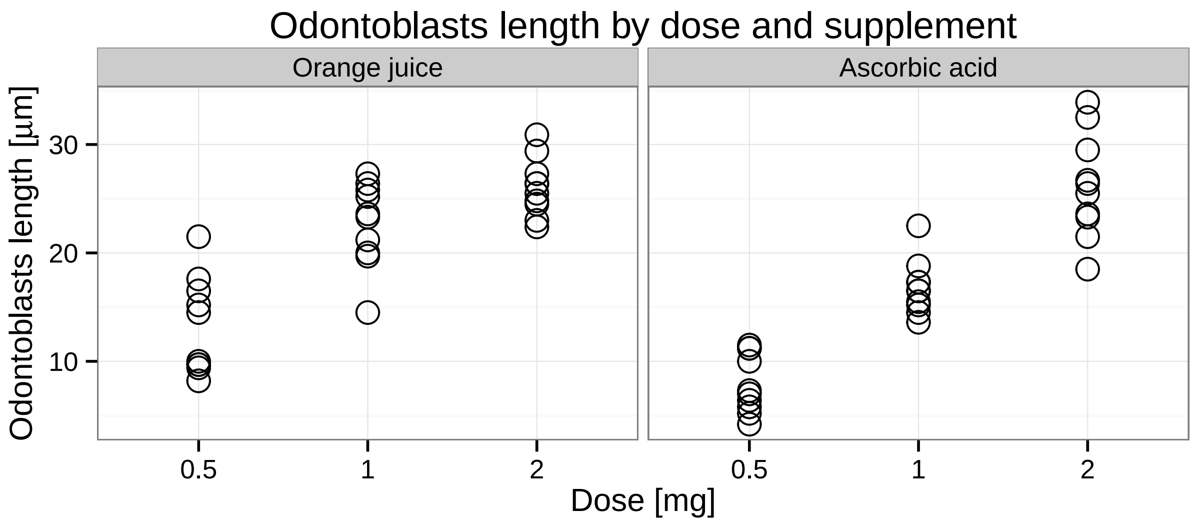

Figure 1: Exploratory graph – odontoblasts length by dose and supplement type

Summaries

In this section, we present short summaries of tooth growth dataset, possibly grouped by dose and/or supplement type. In each table we report median, sample mean and sample standard deviation.

```
summary_dataset <- tbl_df(ToothGrowth) %>%
    summarize(median = median(len), mean = mean(len), sd = sd(len))
kable(summary_dataset, col.names = c("Median", "Mean", "Standard deviation"),
      caption = "Overall summary")
```

Table 1: Overall summary

Median	Mean	Standard deviation
19.25	18.81333	7.649315

```
summary_by_dose <- tbl_df(ToothGrowth) %>%
    group_by(dose) %>%
    summarize(median = median(len), mean = mean(len), sd = sd(len))
kable(summary_by_dose, col.names = c("Dose", "Median", "Mean", "Standard deviation"),
      caption = "Summary by dose")
```

Table 2: Summary by dose

Dose	Median	Mean	Standard deviation
0.5	9.85	10.605	4.499763
1.0	19.25	19.735	4.415436
2.0	25.95	26.100	3.774150

```
summary_by_supp <- tbl_df(ToothGrowth) %>%
    group_by(supp) %>%
    summarize(median = median(len), mean = mean(len), sd = sd(len))
kable(summary_by_supp,
      col.names = c("Supplement type", "Median", "Mean", "Standard deviation"),
      caption = "Summary by supplement type")
```

Table 3: Summary by supplement type

Supplement type	Median	Mean	Standard deviation
Orange juice	22.7	20.66333	6.605561
Ascorbic acid	16.5	16.96333	8.266029

```
summary_by_supp_and_dose <- tbl_df(ToothGrowth) %>%
    group_by(supp, dose) %>%
    summarize(median = median(len), mean = mean(len), sd = sd(len))
kable(summary_by_supp_and_dose,
      col.names = c("Supplement type", "Dose", "Median", "Mean", "Standard deviation"),
      caption = "Summary by supplement type and dose")
```

Table 4: Summary by supplement type and dose

Supplement type	Dose	Median	Mean	Standard deviation
Orange juice	0.5	12.25	13.23	4.459708
Orange juice	1.0	23.45	22.70	3.910953
Orange juice	2.0	25.95	26.06	2.655058
Ascorbic acid	0.5	7.15	7.98	2.746634
Ascorbic acid	1.0	16.50	16.77	2.515309
Ascorbic acid	2.0	25.95	26.14	4.797731

Hypotheses testing

In all hypotheses testing, we will use the Student's t-tests, since we have only a small number of samples and cannot deduce the distribution they are drawn from.

Odontoblasts length by supplement type

At first we will concern on comparing odontoblasts length on both types of supplement. Denote μ_{OJ} the mean of odontoblasts length of pigs with orange juice supplement, μ_{VC} the mean of odontoblasts length of pigs with ascorbic acid supplement. Recall from previous section, that $\bar{x}_{\mathrm{OJ}} = 20.663$ and $\bar{x}_{\mathrm{VC}} = 16.963$ are the corresponding sample means. The null hypothesis H_0 states that $\mu_{\mathrm{OJ}} = \mu_{\mathrm{VC}}$, the alternative hypothesis H_1 states that the (population) means are unequal.

We will use both sided student's t-test (as we don't know which supplement should have larger impact on odontoblasts length), with two groups (the pigs in experiment are all different, hence impairable), with possible unequal variances of groups and the (default) 0.95 confidence level:

```
dose_hyp <- t.test(len ~ supp, data = ToothGrowth, var.equal = FALSE, paired = FALSE)
dose_hyp
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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 in group Orange juice mean in group Ascorbic acid
##                    20.66333                    16.96333
```

From the verbose output we see that we cannot reject H_0. The confidence interval `dose_hyp$conf.int`, which is $(-0.171, 7.571)$, does contain zero. Also note that the p-value is 0.0606345, i.e. greater than $0.05 = 1 - 0.95$, which is the bound for the 0.95 confidence level.

Odontoblasts length by dose

Let us denote $\mu_{d=0.5}$, $\mu_{d=1}$ and $\mu_{d=2}$ the population means of pigs by the specified dose (0.5 mg, 1 mg and 2 mg, respectively). Recall that the corresponding sample means are $\bar{x}_{d=0.5} = 10.605$, $\bar{x}_{d=1} = 19.735$ and $\bar{x}_{d=2} = 26.1$, respectively.

We feel that there is a strong positive correlation between dose and tooth growth, let us make it clear using one-sided student's t-tests. We will make three of them, one for each pair of doses.

Let H_0^1 be the null hypothesis stating $\mu_{d=0.5} = \mu_{d=1}$ and the alternative H_1^1 will be of the form $\mu_{d=0.5} < \mu_{d=1}$. As before, consider the two groups unpaired and their variances unequal. Set the confidence level to 0.99.

```
supp_hyp_1 <- t.test(len ~ dose, data = filter(ToothGrowth, dose %in% c(0.5, 1)),
                     var.equal = FALSE, paired = FALSE,
                     alternative = "less", conf.level = 0.99)
```

The confidence interval (`supp_hyp_1$conf.int`) is $(-\infty, -5.706)$ does not contain 0, so we reject H_0^1 in favor of H_1^1. For curiosity, the p-value corresponding for the given test is 6.3415036×10^{-8} hence we should reject the H_0^1 even if we set the confidence level to 0.9999999.

Without much more effort we can state the hypotheses and test them for the remaining pairs of doses. The meta-settings (two group t-test, one-sided, unequal variances, unpaired groups, confidence level 0.99) remain the same.

Let H_0^2 be the null hypothesis stating $\mu_{d=1} = \mu_{d=2}$ and the alternative H_1^2 be of the form $\mu_{d=1} < \mu_{d=2}$. Similarly denote by H_0^3 the null hypothesis $\mu_{d=0.5} = \mu_{d=2}$ and the alternative H_1^3 states $\mu_{d=0.5} < \mu_{d=2}$.

```r
supp_hyp_2 <- t.test(len ~ dose, data = filter(ToothGrowth, dose %in% c(1, 2)),
                     var.equal = FALSE, paired = FALSE,
                     alternative = "less", conf.level = 0.99)
supp_hyp_3 <- t.test(len ~ dose, data = filter(ToothGrowth, dose %in% c(0.5, 2)),
                     var.equal = FALSE, paired = FALSE,
                     alternative = "less", conf.level = 0.99)
```

We will reject both the null hypotheses in favor of the alternatives. The evidence are the confidence intervals $(-\infty, -3.207)$ for the second test and $(-\infty, -12.301)$ for the last one. Corresponding the p-values are 9.5321476×10^{-6}, respectively $2.1987625 \times 10^{-14}$ [sic].

Conclusions

Concerning an impact of supplement type on the odontoblasts length we cannot reject the hypothesis that the population means of both group are equal (at the 0.95 confidence level). In another words at this confidence level we can't tell if the supplement type of vitamin C has an impact on tooth growth of guinea pigs.

On the other hand, we have shown that the odontoblasts length **does** depend on the dose of vitamin C recieved by guinea pigs (at the 0.99 confidence level and possibly a lot higher, as is suggested by reported p-values). All of three tests addressing this dependency showed that larger dose implies longer odontoblasts.

Notice

The author apologizes for exceeding the maximum page size of the report. But in his opinion splitting the report into text and supplementary part each of 3 pages would decrease the readibility of the analysis.

Another excuse for the bad English.

The knitr source code of this analysis is available at github.