The Vitamin C intake of the Guinea Pig: data analysis

Ilya Maltsev

March 13, 2017

Overview

The Vitamin C intake of the Guinea Pig can be observed by the growth of the odontoblasts of the incisor tooth (original paper). The observed dataset contains measured length of odontoblasts for one of three dose levels by one of two delivery methods. The project focuses on comparing length versus dose and supplement.

It is proved that the bigger is dose the larger is length for any supplement. 'Orange juice' supplement results in larger length for the doses 0.5 and 1, no significant difference is found for the dose 2.

Data processing

Requred libraries:

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

Load the data:

```
data("ToothGrowth")
ToothGrowth <- tbl_df(ToothGrowth)
```

Basic summary

```
table(is.na.data.frame(ToothGrowth))
```

```
##
## FALSE
##   180
```

There are no missed values in the data.

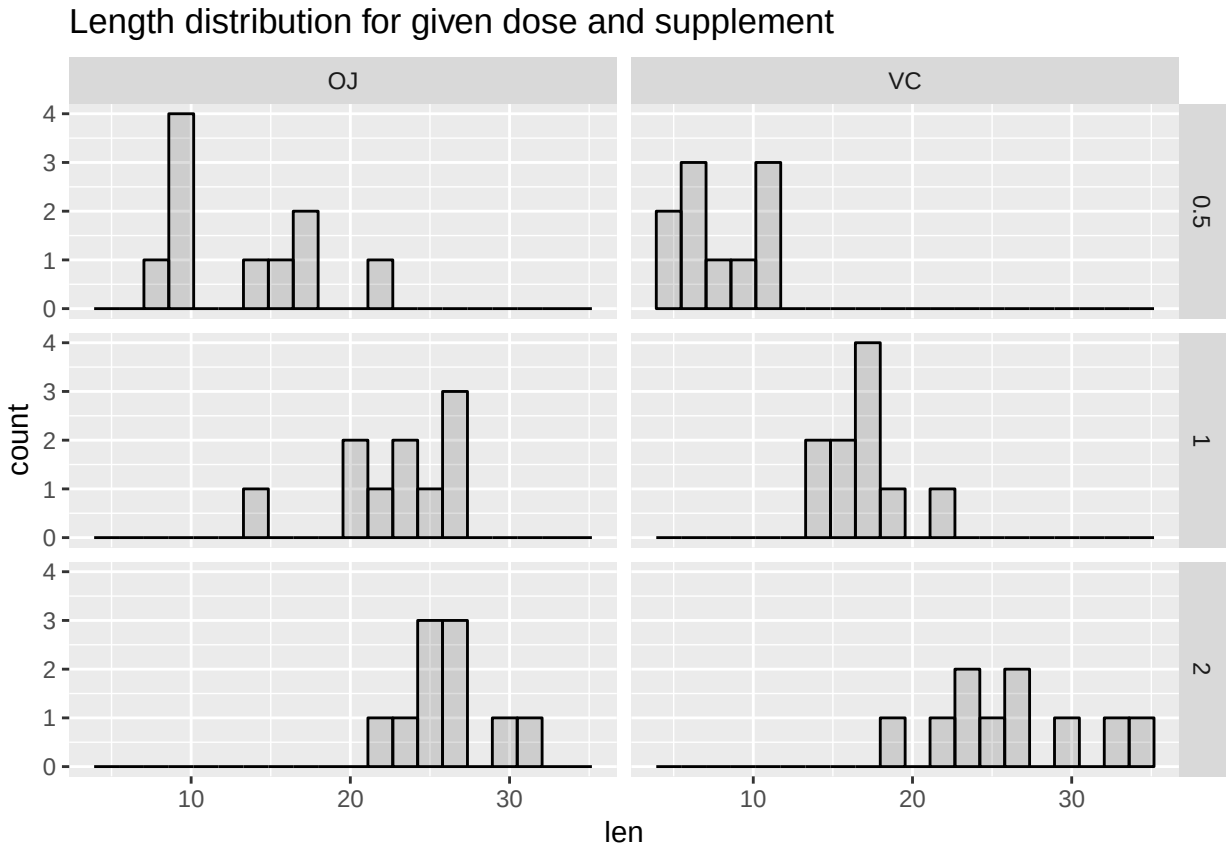
Basic properties of the data:

```
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
```

Number of available observations of 'orange juice' delivery method is 30 and 30 for 'vitamin c' delivery method

Exploratory analysis



It is obvious from the plot that the dose affects the length, the bigger is dose the larger is length. The distribution of lengths given dose and supplement does not seem to match any known distribution.

Assume the lengths follow to normal distribution (null hypothesis). Number of observations is small so the hypothesis can be verified by Shapiro–Wilk test. Try to reject the null hypothesis given an α level of 0.05.

```
alpha <- 0.05
# if reject = FALSE then we failed to reject
ToothGrowth %>% group_by(supp, dose) %>%
    summarise(pvalue = shapiro.test(len)$p.value, reject = pvalue <= alpha)
```

```
## Source: local data frame [6 x 4]
## Groups: supp [?]
##
##     supp  dose    pvalue reject
##   <fctr> <dbl>     <dbl>  <lgl>
## 1     OJ   0.5 0.1820408  FALSE
## 2     OJ   1.0 0.4152983  FALSE
## 3     OJ   2.0 0.8147908  FALSE
## 4     VC   0.5 0.1695627  FALSE
## 5     VC   1.0 0.2697855  FALSE
## 6     VC   2.0 0.9194497  FALSE
```

We failed to reject the hypothesis for all combinations of supplement and dose. It means we can assume that the lengths distribution is normal.

Dosage and supplement effect

Let the null hypothesis is some (supp2, dose2) combination has greater effect than some other combination (supp1, dose1). Let the probability of rejection (α, Type 1 error rate) is 0.05. We would reject the null hypothesis if right bound of confidence interval is greater than zero or if p-value is greater than α.

```
alpha <- 0.05
data_frame(dose1 = c(0.5,  1,    0.5,  1,    0.5,  1,    2),
           dose2 = c(1,    2,    1,    2,    0.5,  1,    2),
           supp1 = c("OJ", "OJ", "VC", "VC", "VC", "VC", "VC"),
           supp2 = c("OJ", "OJ", "VC", "VC", "OJ", "OJ", "OJ")) %>%
    rowwise() %>% # evaluate by row
    mutate(htest = list(t.test(
        y = (ToothGrowth %>% filter(supp == supp2) %>% filter(dose == dose2))$len,
        x = (ToothGrowth %>% filter(supp == supp1) %>% filter(dose == dose1))$len,
        alternative = "less", paired = FALSE,
        var.equal = FALSE, conf.level = 1 - alpha))) %>% # run test
    mutate(pvalue = htest$p.value) %>% # extract p-value
    mutate(left = htest[[4]][1]) %>% # extract left bound of conf.int
    mutate(right = htest[[4]][2]) %>% # extract right bound of conf.int
    mutate(reject = pvalue > alpha | right > 0) %>% # make a decision
    select(-htest)
```

```
## Source: local data frame [7 x 8]
## Groups: <by row>
##
## # A tibble: 7 <U+00D7> 8
##   dose1 dose2 supp1 supp2       pvalue  left      right reject
##   <dbl> <dbl> <chr> <chr>        <dbl> <dbl>      <dbl>  <lgl>
## 1   0.5   1.0    OJ    OJ 4.392460e-05  -Inf -6.2143160  FALSE
## 2   1.0   2.0    OJ    OJ 1.959757e-02  -Inf -0.7486236  FALSE
## 3   0.5   1.0    VC    VC 3.405509e-07  -Inf -6.7468669  FALSE
## 4   1.0   2.0    VC    VC 4.577802e-05  -Inf -6.3465248  FALSE
## 5   0.5   0.5    VC    OJ 3.179303e-03  -Inf -2.3460403  FALSE
## 6   1.0   1.0    VC    OJ 5.191879e-04  -Inf -3.3561576  FALSE
## 7   2.0   2.0    VC    OJ 5.180742e-01  -Inf  3.1334996   TRUE
```

The test shows that increasing of the dose given a supplement results in increasing of the length in all cases. Length effect of 'orange juice' supplement is bigger than 'vitamin c' one for the doses 0.5 and 1. The test cannot distinguish length effect of the supplements for dose equal to 2 for given level of significance.

Conclusion

The distribution of length given supplement and dose can be assumed normal with the significance level $\alpha = 0.5$. This is proved by Shapiro-Wilk test.

T confidence interval test shows the bigger is dose the larger is length and 'orange juice' supplement results in larger length for the doses 0.5 and 1. Significance level is $\alpha = 0.5$. No significant difference is found for the dose 2.