Effect of Vitamin C Supplement on Guinea Pig Tooth Growth

Heli Pykälä

2016-05-25

Overview

In this report we research the effect of delivery method and dosage of vitamin C on guinea pig tooth growth. We compare the data samples using permutation tests. Based on the data we found that increasing the dosage from 0.5 mg/day up to 2.0 mg/day seem to have positive effect on the tooth growth.

Data

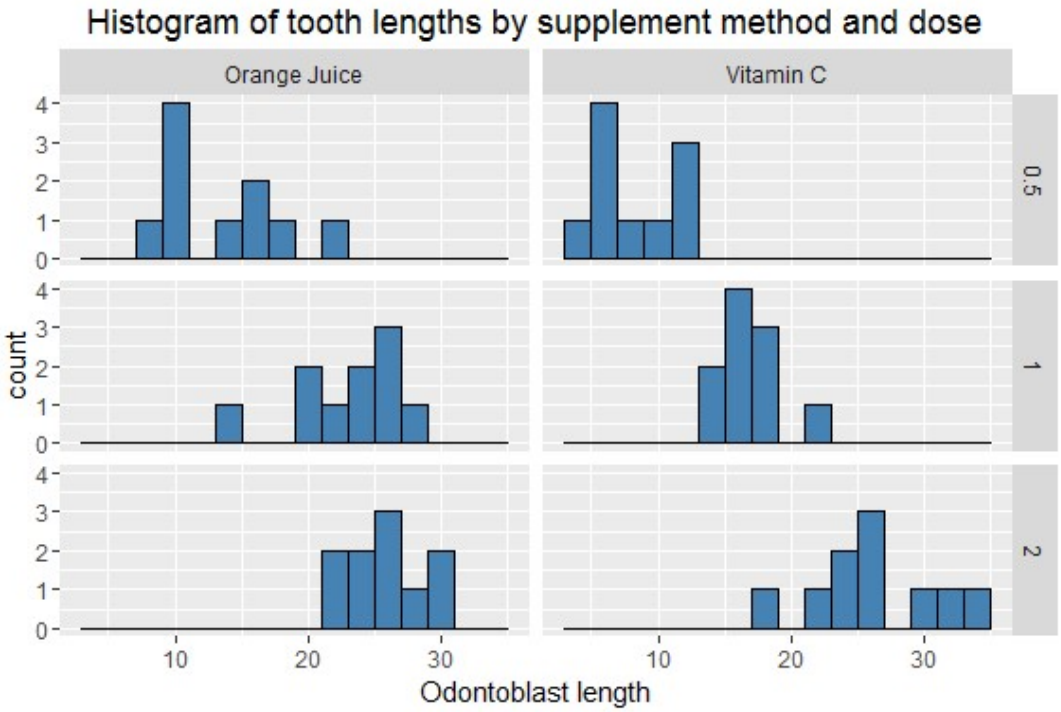
We use ToothGrowth data set from the R library data sets. Description of the data: The response is the length of odontoblasts (cells responsible for tooth growth) in 60 guinea pigs. Each animal received one of three dose levels of vitamin C (0.5, 1, and 2 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC).

Gather the data and factorize the dosage since we will compare groups instead of considering the dose as a quantitative variable. Next explore some characteristics of the data set.

```
library(datasets)
library(ggplot2)
tooth <- ToothGrowth
tooth$dose <- factor(tooth$dose)
levels(tooth$supp) <- c("Orange Juice", "Vitamin C")

summary(tooth)
```

```
##       len                  supp     dose
##  Min.   : 4.20   Orange Juice:30   0.5:20
##  1st Qu.:13.07   Vitamin C   :30   1  :20
##  Median :19.25                     2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

```
ggplot(data = tooth, aes(x = len)) +
    geom_histogram(binwidth = 2, colour = "black", fill = "steelblue") +
    facet_grid(facets = dose ~ supp) +
    ggtitle("Histogram of tooth lengths by supplement method and dose") +
    xlab("Odontoblast length")
```



From histograms it seems clear that the dosage of the vitamin C has correlation with the tooth growth. The supplement type doesn't have such a clear connection. We also see that the distributions don't seem to be very normal, but as the sample size is only 10 that is difficult to say for sure. In addition, if we combine the data by margins so that we can look at dosage or delivery method independently the distribution is pretty far from normal as it won't be even unimodal.

Hypothesis testing

In the data exploration we concluded that assuming normally distributed samples doesn't seem to be valid. Student's t-test is quite robust even when the data aren't normally distributed, but to be on the safe side we use a non parametric permutation test to avoid having to make assumptions about the underlying distribution. It suffices that the observations are independent which is the case according to the data set description. In the permutation test our null hypothesis is that the different samples have same average. If the null hypothesis is valid the group label is irrelevant and we can permutate our samples. We use function permTest to calculate two sided p values for the tests. For details of the function, please refer to the appendix. Source the R function and set seed for reproducibility.

```
source("permTest.R")
set.seed(123)
```

In all tests we use two sided tests so the hypotheses will be:

$H_0: \mu_1 = \mu_2$

$H_a: \mu_1 \neq \mu_2$

Delivery method

First we study the difference between delivery methods. The null hypothesis is that both delivery methods have the same average and the alternative hypothesis is that they are different.

```
pt1 <- permTest(tooth$len,tooth$supp, "Vitamin C", "Orange Juice")
data.frame("Difference of means" =  pt1$origStat, "p-value" = pt1$p)

##   Difference.of.means p.value
## 1                 3.7  0.0603
```

the p-value of the test is 0.06 We can't reject the null hypothesis on 0.05 significance level.

Dosage

Next we perform two more permutation tests to see if there's statistically significant difference in tooth growth with different dosages of the vitamin C. First we test if there is difference between groups that received the dose 0.5 or 1.0 and then between groups 1.0 and 2.0.

```
pt2 <- permTest(tooth$len,tooth$dose, 0.5, 1.0)
data.frame("Difference of means" =  pt2$origStat, "p-value" = pt2$p)

##   Difference.of.means p.value
## 1                9.13       0
```

The p-value is close to zero so we can reject the null hypothesis on 0.05 significance level or any other reasonable critical value. The difference between groups is 9.13 so we can conclude that increasing the dosage from 0.5 to 1.0 increases the tooth growth.

```
pt3 <- permTest(tooth$len,tooth$dose, 1.0, 2.0)
data.frame("Difference of means" =  pt3$origStat, "p-value" = pt3$p)

##   Difference.of.means p.value
## 1               6.365       0
```

The p-value is close to zero so we can reject the null hypothesis on 0.05 significance level. The difference between groups is 6.37 so we can conclude that increasing the dosage from 1.0 to 2.0 increases the tooth growth.

Multiple testing correction for controlling family wise error rate is not necessary in this case. Of the three tests one was not significant anyway and other two had so small p-values that they would survive even the strictest control procedures.

Conclusion

Based on the tooth growth data we found that the dosage of vitamin C correlates with the length of odontoplasts of guinea pigs. We obtained statistically significant result that guinea pigs getting 1.0 mg of vitamin C per day had longer odontoplasts than the group getting 0.5 mg/day. The group getting 2.0 mg per day had even better tooth growth. However based on this data we found no statistically significant evidence that the delivery method has effect on the tooth growth.

Appendix

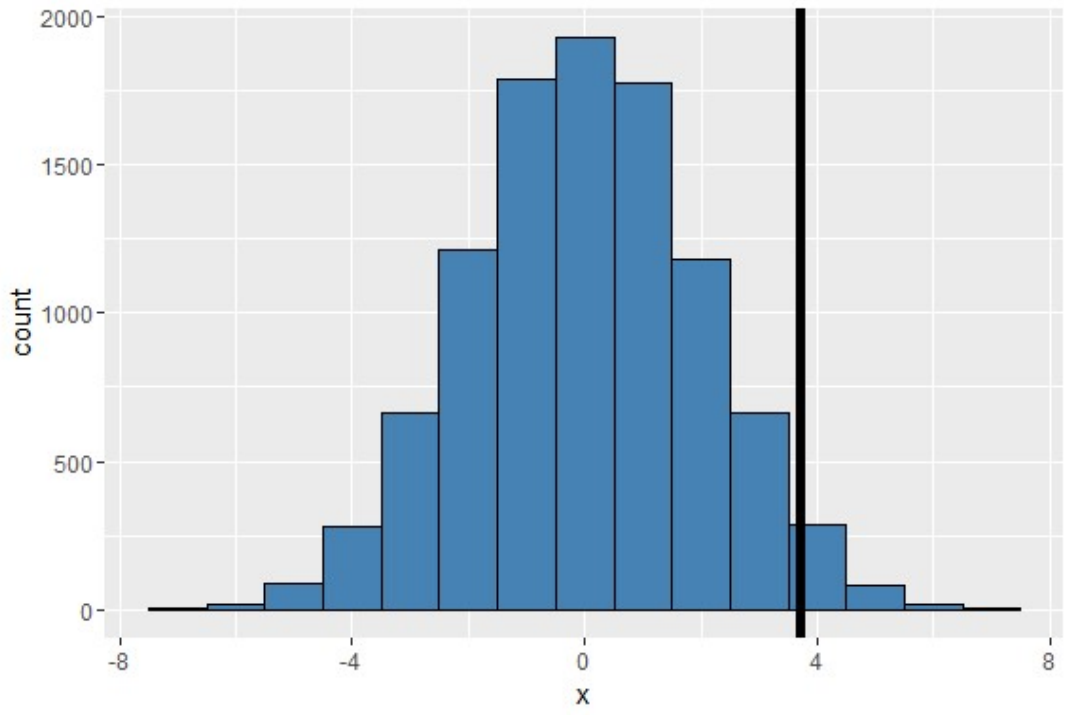
Source code for function permTest

```r
# Function to perform two sided permutation test to test if
# the mean of variable y is different between groups g1 and g2
# y: vector containing samples for the variable
# group: vector containing group values for y
# g1: value of the first group
# g2: value of the second group
permTest <- function(y, group, g1, g2, n = 10000) {

    # test validity of parameters
    if(length(y) != length(group)) {
        message("ERROR: y and group have different lengths")
        return()
    }

    if(!any(group == g1)) {
        message("no samples in g1")
        return()
    }

    if(!any(group == g2)) {
        message("no samples in g2")
        return()
    }

    #subset only the groups being tested
    gsub <- group[group %in% c(g1,g2)]
    ysub <- y[group %in% c(g1,g2)]

    # test statistic is the difference of the mean of y in groups g1 and g2
    testStat <- function(var,gr) {mean(var[gr==g2]) - mean(var[gr == g1])}

    #original permutation
    origStat <- testStat(ysub,gsub)

    #test statistic for n permutations
    permStats <- sapply(1:n, function(i) testStat(ysub, sample(gsub)))

    #p value for two sided test is calculated as the percentage of permutations
    #having more extreme value than the original regardless of the sign
    p <- mean(abs(permStats) > abs(origStat))

    # return p-value, test statistics of permutations,
    # and the original test statistic
    list(p = p, permStats = permStats, origStat = origStat)
}
```

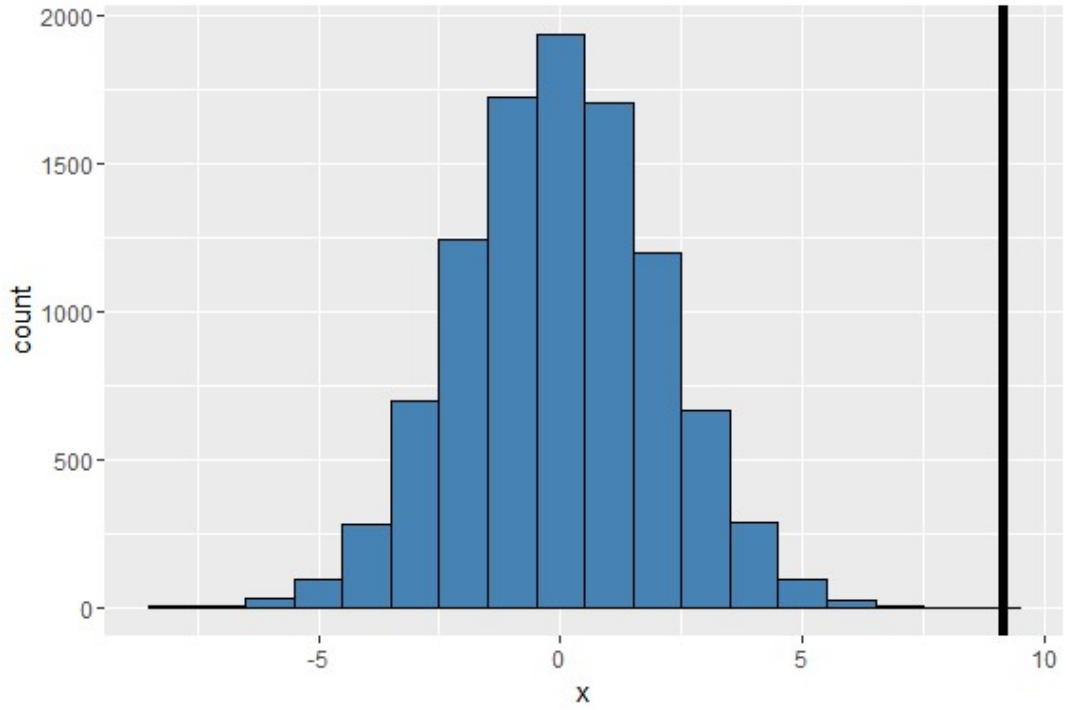
Histograms for permutation test results

Here we display histograms for the permutation test results. The histogram shows the distribution of the test statistic and the vertical line the test statistic for the original distribution. The histograms are consistent with the calculated p-values.

```
dt1 <- data.frame(x=pt1$permStats)
ggplot(dt1,aes(x=x)) +
    geom_histogram(colour = "black", fill = "steelblue", binwidth = 1) +
    geom_vline(xintercept = pt1$origStat, size = 2, color = "black")
```




```
dt2 <- data.frame(x=pt2$permStats)
ggplot(dt2,aes(x=x)) +
    geom_histogram(colour = "black", fill = "steelblue", binwidth = 1) +
    geom_vline(xintercept = pt2$origStat, size = 2, color = "black")
```




```
dt3 <- data.frame(x=pt3$permStats)
ggplot(dt3,aes(x=x)) +
    geom_histogram(colour = "black", fill = "steelblue", binwidth = 1) +
    geom_vline(xintercept = pt3$origStat, size = 2, color = "black")
```