Effect of vitamin C on tooth cell growth

Timo Voipio

14 Aug 2016

Overview

This paper explores the `ToothGrowth` dataset from the R `datasets` package. Based on the data, we determine that larger dosages of vitamin C is associated with longer odontoblasts, and that for lower concentrations delivering the vitamin via orange juice seems to be associated with longer odontoblasts, compared to ascorbic acid.

Data description

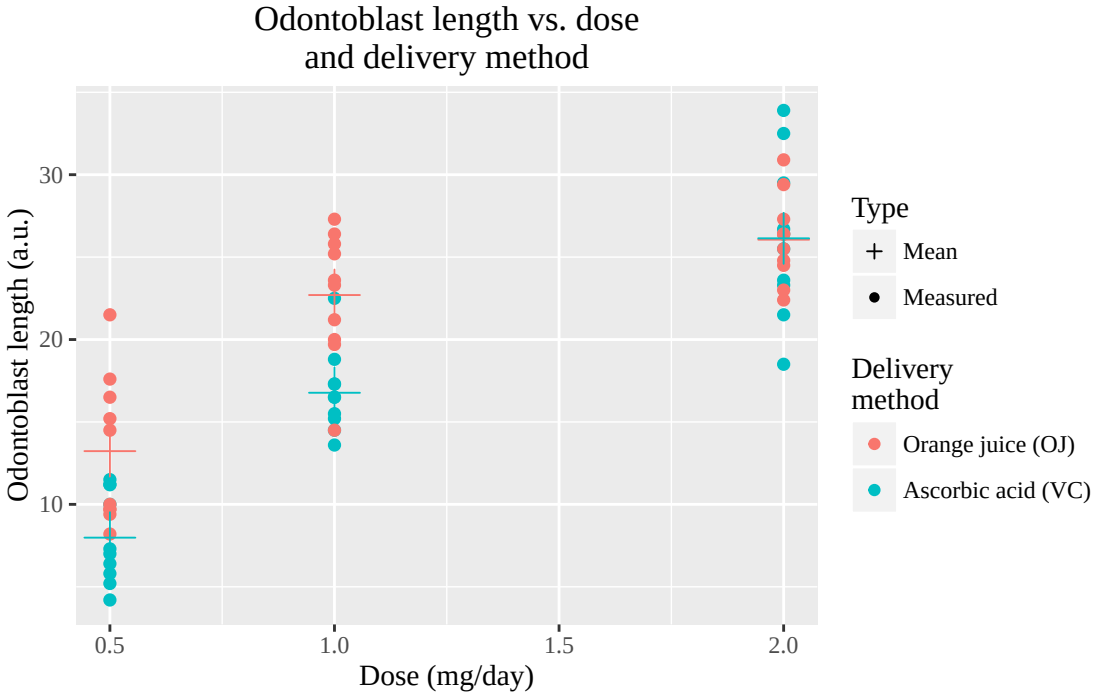
The dataset is described in its documentation as follows:

> The Effect of Vitamin C on Tooth Growth in Guinea Pigs. 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, 1.0, and 2.0 mg/day) by one of two delivery methods, (orange juice or ascorbic acid (a form of vitamin C and coded as VC).

Delivery via orange juice is coded with "OJ". There are ten measurements of the response (odontoblast length) for each of the six possible combinations of dose and delivery method.

Exploratory analysis

The measured odontoblast lengths are shown below as a function of the vitamin C dose. Delivery method is indicated by marker color. The mean of each of the six dose–method groups is indicated with a cross, colored according to the method.



We observe that the odontoblasts seem to be longer in those gerbils which had larger daily dose of vitamin C. For doses of 0.5 and 1.0 mg/day, also the delivery method seems to have an effect, with the orange juice being associated with longer odontoblasts.

Hypothesis testing of the mean odontoblast length

We investigate the differences in the mean lengths of different dose–method groups by using the two-sided t-test ($\alpha = 0.05$). In order to control false discovery rate, the p values are adjusted using the Benjamini–Hochberg method, R function `p.adjust(p, method = "BH")`. Only the dose–method pairs where either the dose or the method are the same are considered, i.e., if both the dose and the delivery method differ the difference of the means and the associated p value is not calculated.

The differences in the means are given in the table below. Gray cells correspond to pairs which are not considered. Cells with green background indicate that the difference in the means is statistically significant, while red background indicates the converse. The difference of the mean is given first, positive value indicates that the mean of the group named on the same row is larger than the mean of the group named at the top of the column. The number after the semicolon is the (unadjusted) p value.

	VC 0.5	OJ 0.5	VC 1.0	OJ 1.0	VC 2.0
OJ 0.5	$5.25; 6.36 \times 10^{-3}$				
VC 1.0	$8.79; 6.81 \times 10^{-7}$				
OJ 1.0		$9.47; 8.78 \times 10^{-5}$	$5.93; 1.04 \times 10^{-3}$		
VC 2.0	$18.16; 4.68 \times 10^{-8}$		$9.37; 9.16 \times 10^{-5}$		
OJ 2.0		$12.83; 1.32 \times 10^{-6}$		$3.36; 3.92 \times 10^{-2}$	$-0.08; 9.64 \times 10^{-1}$

We observe that for both delivery methods and all combinations of dose, the difference in the means is statistically significant. Additionally, for daily doses of 0.5 and 1.0 mg, the gerbils to whom the vitamin was delivered via orange juice had, on average, longer odontoblast cells than those who were given ascorbic acid. For the daily dose of 2.0 mg, the difference between the delivery methods was not statistically significant.

Permutation testing

Finally, we evaluate the different delivery methods via permutation testing at each dose level. For each 3 dose levels, the delivery method labels are reassigned at random 10000 times, then the difference between the means of the reassigned groups is calculated and compared to the difference in the original data. The fraction of the permutations where the difference of means is greater than the actual sample mean serves as an estimate of the significance of the measured difference.

The following table shows the result of the permutation testing. From the previous section our hypothesis is that orange juice (OJ) is associated with higher mean than ascorbic acid (VC), so for each permutation we have subtracted the mean of the data labeled `VC` from the mean of the data assigned the label `OJ`. The observed difference for each dose is given in the column $\hat{\Delta}$. The mean and standard deviation of the differences obtained by permuting the labels are given in the columns $\bar{\Delta}$ and σ, respectively. The rightmost column $\hat{\Delta} > \bar{\Delta}$ shows the fraction of the permutations which resulted in larger difference of means than present in the original data.

Dose	$\hat{\Delta}$	$\bar{\Delta}$	σ	$\bar{\Delta} > \hat{\Delta}$
0.5	5.25	0.01	2.01	0.0031
1.0	5.93	0.01	1.97	0.0010
2.0	-0.08	0.03	1.70	0.5247

Small values of $\hat{\Delta} > \bar{\Delta}$ indicate that the difference in the means between the delivery methods, as in the original data, is statistically significant. We see that for the two smaller doses the difference between the delivery methods, is significant, but the largest dose approximately half of the label permutations result in a difference of means larger than in the original data.

Conclusions

We conclude that, according to our analysis, larger doses of vitamin C are associated with longer odontoblasts. For daily doses of 0.5 mg and 1.0 mg, delivering the vitamin via orange juice instead of as ascorbic acid is associated higher mean odontoblast length. These results are statistically significant ($\alpha < 0.05$); the Benjamini–Hochberg correction was used to control the false discovery rate. For the largest reported dose, 2.0 mg/day, the difference between the delivery methods is not statistically significant. Our conclusions are based on the assumption that the odontoblast length is normally distributed and thus the *t*-test may be used.

Appendix

This appendix includes the R code used to conduct the simulation experiment and to format the results. The simulation was performed using R version 3.2.3 (2015-12-10) on OS X 10.10.5 (Yosemite) [x86_64-apple-darwin13.4.0 (64-bit)].

Source code

The following shows the source code used in the analysis. Due to length constraints, some code sections related strictly to presenting the results have been omitted. Complete R Markdown sourcecode is available in GitHub, https://github.com/tvoipio/JHU_SIProject

```
library(ggplot2)
library(datasets)
library(xtable)

# Temporarily disable warnings while dplyr is loaded
oldw <- getOption("warn")
options(warn = -1)
library(dplyr)
options(warn = oldw)

data("ToothGrowth")
#tg <- ToothGrowth[, c("len", "supp", "dose")]

# Set random seed for repeatability
set.seed(160808)
# Determine the mean odontoblast length for each combination of dose
# and delivery method
tgmeans <- group_by(ToothGrowth, supp, dose) %>%
    summarize(len = mean(len)) %>%
    ungroup()
# Ensure that the columns of tgmeans are in the same order as in tg
tgmeans <- tgmeans[, names(ToothGrowth)]

# Combine the measurement results and the computed means for plotting
tgc <- rbind(ToothGrowth, tgmeans)

# Add a factor variable to indicate whether the row is measured data
# or a mean
tgc <- cbind(tgc,
             type = factor(c(rep("Measured", times = nrow(ToothGrowth)),
                             rep("Mean", times = nrow(tgmeans)))))

# Prepare a plot of length vs dose using ggplot2
gt <- ggplot(tgc, aes(x = dose, y = len))
```

(code block omitted)

```r
# Create a data frame of the possible combinations of delivery method
# and dose, one combination per row
# The leftmost variable variest fastest, the rightmost slowest
metxdose <- expand.grid(supp = unique(tgc$supp), dose = unique(tgc$dose))

# Set the significance level
signif.level <- 0.05

# Choose the p value adjustment method to control the false discovery
# rate (using p.adjust())
#adj.method <- "bonferroni"
adj.method <- "BH"

# Indices of the elements of a lower triangular matrix (main diagonal omitted)
# Thanks to: http://stackoverflow.com/a/20898910
# We want to consider each dose--method pair only once
rows <- nrow(metxdose)
rowinds <- rev(abs(sequence(seq.int(rows - 1)) - rows) + 1)
colinds <- rep.int(seq.int(rows - 1), rev(seq.int(rows - 1)))
idx <- cbind(rowinds, colinds)

# Form a list of the index pairs; each element of the list identifies
# a dose-method combination, i.e., a row in metxdose
idxl <- lapply(seq_len(nrow(idx)), function(i) as.vector(idx[i,]))

# Remove from consideration the indices which consider to pairs
# where both the dose and the delivery method differ
onediffers <- sapply(idxl, function(x) any(metxdose[x[1], ]
                                           == metxdose[x[2], ]))
idxl <- idxl[onediffers]

# Function to calculate the two-sided t-test between the identified
# pair of dose-method combinations. The calculated difference is
# data.r-data.c, where data.c is the combination corresponding to the
# index on the column in the metxdose matrix and data.r corresponds to
# the row.
calculate_stat <- function(l)
{
    data.r <- ToothGrowth[ToothGrowth$supp == metxdose[l[1], "supp"] &
                  ToothGrowth$dose == metxdose[l[1], "dose"], "len"]
    data.c <- ToothGrowth[ToothGrowth$supp == metxdose[l[2], "supp"] &
                  ToothGrowth$dose == metxdose[l[2], "dose"], "len"]
    t.test(data.r, data.c, paired = FALSE, var.equal = FALSE)
}

# Perform the t tests for each element in idxl, and for convenience
# separate the p values of the calculated differences of means
# and the differences themselves into separate vectors
t.test.results <- lapply(idxl, calculate_stat)
p.values <- sapply(t.test.results, function(res) res$p.value)
meandeltas <- sapply(t.test.results, function(res) -diff(res$estimate))
names(meandeltas) <- NULL

# Calculate the adjusted p values
p.adj.values <- p.adjust(p.values, method = adj.method)

# Create descriptive yet concise row and column names for use in the
# result matrix
```

```
rowcolnames <- apply(metxdose, 1, paste, collapse = " ")

# Construct empty matrices into which the calculated p values, adjusted
# p values, mean deltas and whether the result is significant or not are placed
# Since the matrix data is given as a 0-length vector, they are filled
# with NAs.
p.values.mat <- matrix(data = numeric(), nrow = rows, ncol = rows,
                       dimnames = list(rowcolnames, rowcolnames))
p.adj.values.mat <- p.values.mat
meandeltas.mat <- p.values.mat
p.signif.mat <- matrix(data = logical(), nrow = rows, ncol = rows,
                       dimnames = list(rowcolnames, rowcolnames))

# Populate the matrices created above
for (i in seq_along(idxl))
{
    row <- idxl[[i]][1]
    col <- idxl[[i]][2]
    p.values.mat[row, col] <- p.values[i]
    p.adj.values.mat[row, col] <- p.adj.values[i]
    meandeltas.mat[row, col] <- meandeltas[i]
    p.signif.mat[row, col] <- p.adj.values[i] < signif.level
}
rm(row, col)
```

(code block omitted)

```
Nperm <- 10000
doses <- unique(ToothGrowth$dose)
# Create a data frame for storing the results of the permutation tests
perms <- data.frame(matrix(numeric(Nperm*length(doses)*2), ncol = 2))
names(perms) <- c("deltamean", "dose")
perms[, "greater"] <- logical()

testStat <- function(x, suppgrp)
{
    mean(x[suppgrp == "OJ", "len"]) - mean(x[suppgrp == "VC", "len"])
}

obsdeltas <- numeric(3)

for (i in seq_along(doses))
{
    dosei <- doses[i]
    tgdose <- subset(ToothGrowth, dose == dosei)

    obsdeltas[i] <- testStat(tgdose, tgdose$supp)

    offset <- (i-1)*Nperm
    perms[(offset + 1:Nperm), "deltamean"] <-
        sapply(seq(Nperm), function(x) testStat(tgdose, sample(tgdose$supp)))
    perms[(offset + 1:Nperm), "dose"] <- dosei
    perms[(offset + 1:Nperm), "greater"] <-
        perms[(offset + 1:Nperm), "deltamean"] > obsdeltas[i]
}

dosetab <- perms %>%
    group_by(dose) %>%
```

```
    summarize(meandelta = mean(deltamean), sddelta = sd(deltamean),
              greater = mean(greater))
dosetab <- cbind(dosetab, obsdelta = obsdeltas)
dosetab <- select(dosetab, dose, obsdelta, everything())
names(dosetab) <- c("Dose", "$\\hat{\\Delta}$", "$\\bar{\\Delta}$",
                    "$\\sigma$", "$\\bar{\\Delta} > \\hat{\\Delta}$")
print(xtable(dosetab, digits = c(1, 1, 2, 2, 2, 4)),
      include.rownames = FALSE, sanitize.colnames.function = identity)
```