Statistical Inference: Course Project - Part 2

Nikita Kirnosov

Contents

Overview

The response is the length of odontoblasts (teeth) in each of 10 guinea pigs at each of three dose levels of Vitamin C (0.5, 1, and 2 mg) with each of two delivery methods (orange juice or ascorbic acid) is under study in present work. The resulting data set has the following characteristics:

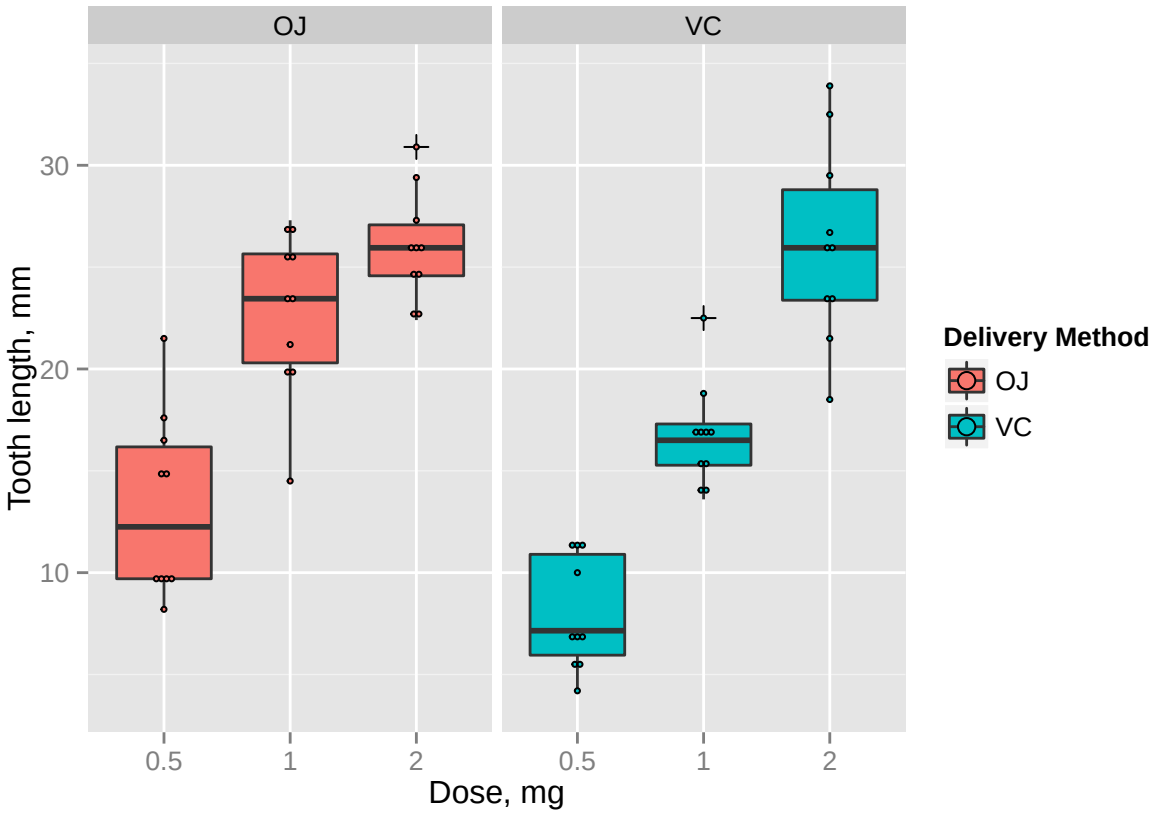
```
##       len     supp         dose
##  Min.   : 4   OJ:30   Min.   :0.50
##  1st Qu.:13   VC:30   1st Qu.:0.50
##  Median :19           Median :1.00
##  Mean   :19           Mean   :1.17
##  3rd Qu.:25           3rd Qu.:2.00
##  Max.   :34           Max.   :2.00
```

In current paper we aim to elucidate the dependences of the teeth length on

- vitamin C dose,
- delivery method.

Results

The graph below indicates for both methods of delivery the dose increase results in teeth length inctrease. It also appears that for doses of 0.5 and 1.0 mg orange juice reponse is stronger than for ascorbic acid. Let us test this observation.



To decide whether there is a real difference between the groups by dose level and delivery method, let us a series of two-sided unpaired t-tests to obtain the confidence intervals and p-values. Here we will assume unequal variance, which is suggested by the graph.

Our null hypothesis will be that there is no difference in the means between the two groups. Let us see whether this is true.

Table 1: *Comparison of delivery methods by dose*

Name	Dose_1	Dose_2	Dose_3
dose, mg	0.500	1.000	2.000
t	3.170	4.033	-0.046
p-value	0.006	0.001	0.964
95% confidence interval	[1.719 , 8.781]	[2.802 , 9.058]	[-3.798 , 3.638]

Table 1 presents t- and p-values for each vitamin C dose level as well as corresponding 95% confidence intervals. In order to be able to decide whether different dose levels yield the results which are statistically significantly different for different delivery methods, we should adjust the p-values. Here we chose to use Holm's method:

```
round(p.adjust(p_values, method = "holm"),3)
```

```
## [1] 0.012 0.003 0.964
```

Conclusion

- At the 0.5 and 1 mg dose levels, there is a statistically significant difference between the means of the two delivery options, since the adjusted p-values are significant at the $\alpha = 0.05$ level, and the 95% confidence intervals do not include zero.
- The 2 mg dose level does not indicate the statistically significant difference in delivery methods, but it has to be noted that could be a consequence of the extremely low test power due to a small number of data points.
- The possible conclusion can be made that orange juice is a more efficient source of vitamin C for guinea pigs teeth growth. The probable explanation is that it contains other nutrients used in this physiological process. It is also clear that the increase of the vitamin C dose from 0.5 mg to 1.0 mg results in positive response regardless of the supplementation method.

Appendix

R code used:

```
library(knitr)
library(pander)
library(ggplot2)
library(datasets)

opts_chunk$set(echo = TRUE, message = FALSE, warning = FALSE)
knit_hooks$set(inline = function(x) {
   if (is.numeric(x)) round(x, 2)})
options(digits=2)

summary (ToothGrowth)

ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) +
  geom_boxplot(outlier.colour = "black",outlier.size = 3,outlier.shape=3) +
  geom_dotplot(binaxis="y",binwidth = 1,stackdir="center",dotsize=0.25)+
  facet_grid(. ~ supp) +
  xlab("Dose, mg") +
  ylab("Tooth length, mm") +
  guides(fill=guide_legend(title="Delivery Method"))

t_test <- function(x) {
    dt <- t.test(len ~ supp, data=subset(ToothGrowth, dose==x),
                 paired=FALSE, var.equal=FALSE)
    return(c(sprintf("%.3f", round(x,1)),
             sprintf("%.3f", round(dt$stat,3)),
             sprintf("%.3f", round(dt$p.value,3)),
             paste("[",round(dt$conf.int[1],3),",",round(dt$conf.int[2],3),"]")))
    }

df <- data.frame (Name=c("dose, mg", "t", "p-value",
                         "95% confidence interval"),
                  Dose_1=t_test(0.5),Dose_2=t_test(1),Dose_3=t_test(2))
pander(df, split.tables=120,
       caption="*Comparison of delivery methods by dose*")
```

```r
x <- c(as.numeric(as.character(df[3,2])),
       as.numeric(as.character(df[3,3])),
       as.numeric(as.character(df[3,4])))

round(p.adjust(x, method = "bonferroni"),3)
```