ToothGrow Analysis

Enelen Brinshaw

Contents

Overview

This report analyses the effect of different doses and methods of delivery of vitamin C in guinea pigs' odontoblasts (cells responsible for tooth growth). There are three variables in our dataset, `len`, the Tooth length, `supp`, the supplement type, which can be either orange juice (OJ) or ascorbidc acid (VC), and `dose`, the dose in milligrams/day. There are 60 observations, ten for each combination of `supp` and `dose`.

Exploratory Analysis

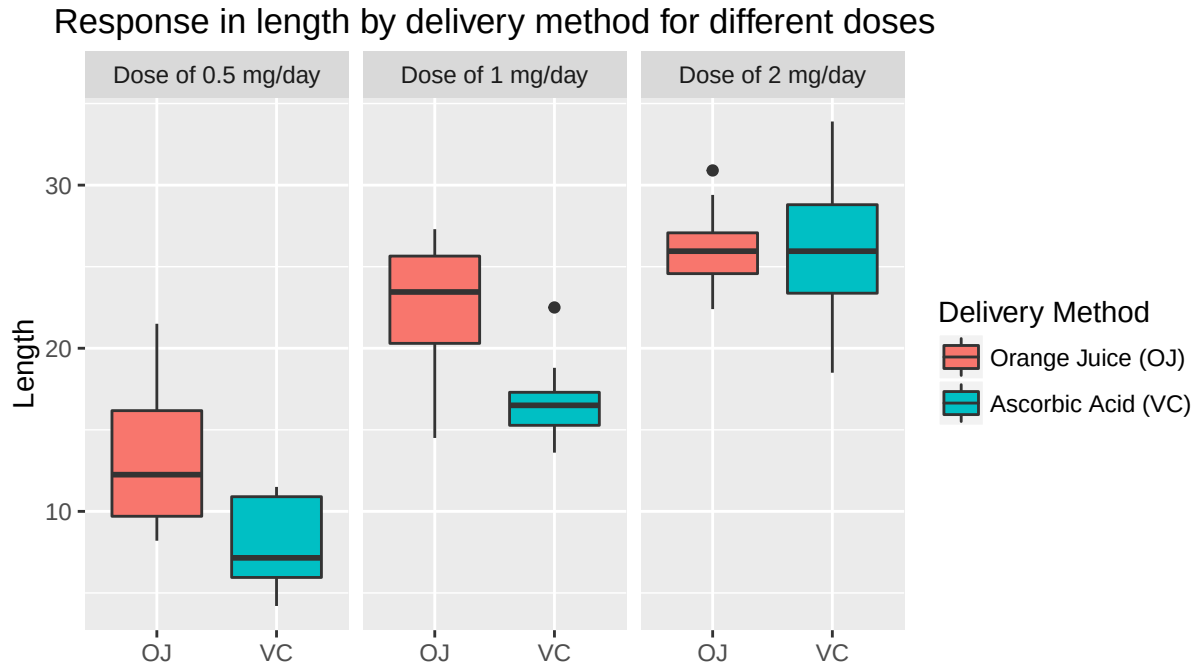
Let us first explore the dataset to see the general trends.

```
data("ToothGrowth")
df <- ToothGrowth
df$dose <- as.factor(df$dose)
require(ggplot2)
ggplot(ToothGrowth,aes(dose,len,colour = supp)) +
     geom_point() + geom_smooth() +
     scale_colour_discrete(name = "Delivery Method",
                           labels = c("Orange Juice (OJ)", "Ascorbic Acid (VC)")) +
     labs(x = "Dose (mg/day)", y = "Length",
          title = "Response in Length of guinea pigs to Vitamin C")
```

There is a clear advantage of giving the dose via orange juice rather than ascorbic acid,

although the advantage is not present for the highest dose of 2mg/day. Let us look at a few more graphs to gain more understanding.

```
ggplot(df, aes(supp, len, group = supp, fill = supp)) +
     geom_boxplot() +
     facet_wrap(facets = ~dose, labeller = doses) + # user-defined function
     scale_fill_discrete(name = "Delivery Method",
                         labels = c("Orange Juice (OJ)", "Ascorbic Acid (VC)")) +
     labs(y = "Length", x = "",
          title = "Response in length by delivery method for different doses")
```



Again we can see Orange juice seems to be better than Ascorbic acid except at the highest dose.

Statistical tests of significance

Let us now confirm whether these differences are significant or due to luck by conducting t-tests, and developing confidence interval for the differences.

```
res <- t.test(df$len[df$supp == "OJ"], df$len[df$supp == "VC"])
require(pander)
htest(res, "")  # another user defined function see appendix
```

Table 1: Welch Two Sample t-test: (OJ−VC)

df	Test Statistic	p-Value	Confidence Interval	
55.31	1.915	0.06063	-0.171	7.571

The test fails at 95%, but considering there was no difference at the highest dose, this is well within the realms of possibility. Let us redo the tests for each dose.

```
k <- split(df,f=df$dose)
tests <- lapply(k, function(x) t.test(x$len[x$supp == "OJ"],
                                      x$len[x$supp=="VC"]))
htest(tests[[1]], "Dose = 0.5")
htest(tests[[2]], "Dose = 1.0")
htest(tests[[3]], "Dose = 2.0")
```

Table 2: Welch Two Sample t-test: (OJ−VC) Dose = 0.5

df	Test Statistic	p-Value	Confidence Interval	
14.97	3.17	0.006359	1.719	8.781

Table 3: Welch Two Sample t-test: (OJ−VC) Dose = 1.0

df	Test Statistic	p-Value	Confidence Interval	
15.36	4.033	0.001038	2.802	9.058

Table 4: Welch Two Sample t-test: (OJ−VC) Dose = 2.0

df	Test Statistic	p-Value	Confidence Interval	
14.04	-0.04614	0.9639	-3.798	3.638

The t-tests confirm our initial exploration. There is a significant difference between tooth length when dose is supplied by Orange juice versus ascorbic acid, but not at the dose level of 2 mg/day (p-value = 0.96 ; interval contains 0). At the level of 0.5mg/day, we are 95% sure, that the orange juice on average produces a greater response in length by an amount between 1.72 to 8.78 when compared to doses given by ascorbic acid. The confidence interval for difference in length for dose of 1 mg/day ranges from 2.8 to 9.1.

Of course, these are all under the assumptions that the data follows a t-distribution, the samples are not paired, and the variance of the two groups in not equal.

Appendix

Code for user-defined functions used in the assignment.

```
# labeller function to label facets
doses <- function(variable, value){
     paste("Dose of", value, "mg/day")
}

# function to print t.test in pander with confidence interval
htest <- function(res, cap = NULL){
     conf <- res$conf.int # extract confidence interval
     attributes(conf) <- NULL # remove its attributes
     conf <- as.character(round(conf, 3)) # and convert to string
     # create a data frame for pander
     temp <- data.frame(df = res$parameter,
                        `Test Statistic` = res$statistic,
                        `p-Value` = res$p.value,
                        `Confidence Interval` = paste(conf[1],
                                                      "$\\quad$",
                                                      conf[2]),
                        check.names = FALSE,
                        row.names = NULL)
     if(is.null(cap)) # if no caption is given use command given
          cap <- paste0("`", res$data.name, "`")
     else
          cap <- paste("(OJ$-$VC)", cap)
     pander(temp, caption = paste0(res$method, ": ", cap))
}
```