Tooth Growth Data Analysis

yangru1q1

2019-06-27

```r
library(tidyverse)
```

Synopsis

In this project, we are going to explore the ToothGrowth dataset in the R datasets package. The dataset contains 60 observations regarding to the length of **odontoblasts** (cell responsible for tooch growth) for guinea pigs. Each animal recieved 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). So, there are total **15** treatment combinations, by different delivery method with different amout. The main goal for this project is to understand if the growth of tooth length for pigs differ by each treatment combinations. From this project, we can conclude that the tooth length does effected by different treatment combinations; Furthermore, mean tooth length for animals who recieve more than 1 mg orange juice (1 or 2 mg) per day seems have no difference with the mean tooth length for those animals who recieve 2 mg vitamin C per day.

Data Exploration

```r
data("ToothGrowth")
```

Notice that "dose" is numeric in the dataset, we need to transform "dose" column into factor.

```r
str(ToothGrowth)
```

```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

Summary across groups of treament

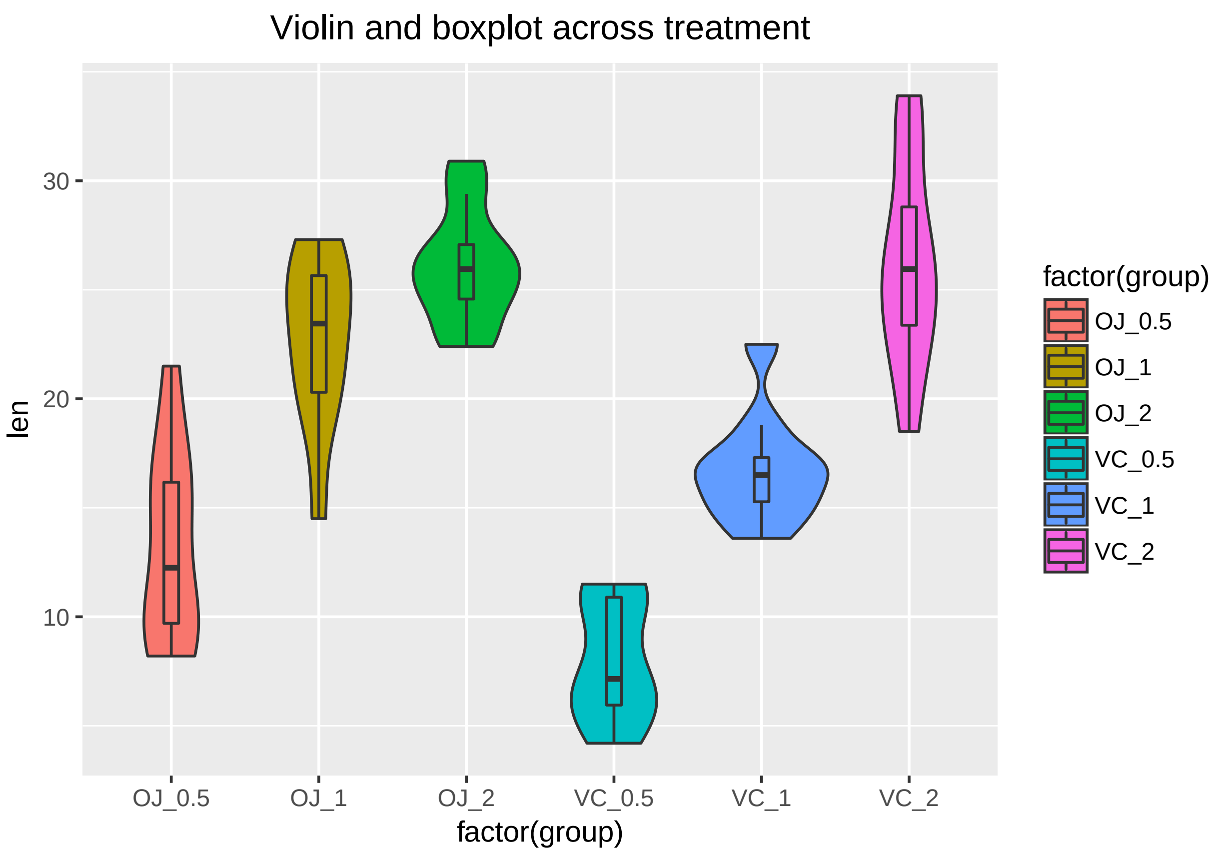
```r
kable(ToothGrowth %>%
                group_by(supp, dose) %>%
                summarize(min_len = min(len),
                          median_len = median(len),
                          max_len = max(len),
                          mean_len = mean(len),
                          sd_len = sd(len))
)
```

supp	dose	min_len	median_len	max_len	mean_len	sd_len
OJ	0.5	8.2	12.25	21.5	13.23	4.459708
OJ	1.0	14.5	23.45	27.3	22.70	3.910953
OJ	2.0	22.4	25.95	30.9	26.06	2.655058
VC	0.5	4.2	7.15	11.5	7.98	2.746634
VC	1.0	13.6	16.50	22.5	16.77	2.515309

supp	dose	min_len	median_len	max_len	mean_len	sd_len
VC	2.0	18.5	25.95	33.9	26.14	4.797731

```
TGgroup <- ToothGrowth %>%
                mutate(group = paste(supp, dose, sep = "_")) %>%
                select(len, group)

dodge <- position_dodge(width = 0.4)
TGgroup %>%
        ggplot(aes(x = factor(group), y = len, fill = factor(group))) +
        geom_violin(position = dodge) +
        geom_boxplot(width=.1, outlier.colour=NA, position = dodge) +
        labs(title = "Violin and boxplot across treatment") +
        theme(plot.title = element_text(hjust = 0.5))
```



We can visualize from above plot, the costant variance across each treatments have been violated. We should consider using t-test with unequal variance.

Data Analysis

Let's build a new dataset contain all p-values for unequal variance two sample t-test and visualize it. For two different treatments group of observation, g_1 and g_2, the t-test is basically testing two groups have equal means.

$$H_0 : \mu_{g_1} = \mu_{g_2} \quad v.s. \quad H_a : \mu_{g_1} \neq \mu_{g_2}$$

Since multiple comparisons have performed, let's consider to use multiple standards whether we should reject the alternative hypothesis, namely "No Correction", "Bonferroni" and "BH".

```
groups <- unique(TGgroup$group)
pVals <- data.frame()
n <- 1
for(i in 1:length(groups)){
        for(j in seq(i+1, 6)){
                pVals[n,1] = groups[i]; pVals[n,2] = groups[j]
                subdata = subset(TGgroup, group %in% c(groups[i], groups[j]))
                pVals[n,3] = t.test(len~group, paired = FALSE, var.equal = FALSE,
                               data = subdata)$p.value[1]
                n = n + 1
        }
        if(n==16) break
}
colnames(pVals) <- c("group1", "group2", "pvalue")
pVals <- pVals %>% arrange(pvalue) %>%
        mutate(p.Sig = ifelse(pvalue<=0.05, "Reject", "Not reject"),
               Bonferronni = ifelse(pvalue <= 0.05 / 15, "Reject", "Not reject"),
               BH = ifelse(pvalue <= 1:15 / 15 * 0.05, "Reject", "Not Reject"))
kable(pVals, digits = 12)
```

group1	group2	pvalue	p.Sig	Bonferronni	BH
VC_0.5	OJ_2	1.400000e-11	Reject	Reject	Reject
VC_0.5	OJ_1	3.655200e-08	Reject	Reject	Reject
VC_0.5	VC_2	4.681600e-08	Reject	Reject	Reject
VC_1	OJ_2	2.361070e-07	Reject	Reject	Reject
VC_0.5	VC_1	6.811020e-07	Reject	Reject	Reject
OJ_0.5	OJ_2	1.323784e-06	Reject	Reject	Reject
VC_2	OJ_0.5	7.196254e-06	Reject	Reject	Reject
OJ_0.5	OJ_1	8.784919e-05	Reject	Reject	Reject
VC_1	VC_2	9.155603e-05	Reject	Reject	Reject
VC_1	OJ_1	1.038376e-03	Reject	Reject	Reject
VC_0.5	OJ_0.5	6.358607e-03	Reject	Not reject	Reject
OJ_1	OJ_2	3.919514e-02	Reject	Not reject	Reject
VC_1	OJ_0.5	4.601033e-02	Reject	Not reject	Not Reject
VC_2	OJ_1	9.652612e-02	Not reject	Not reject	Not Reject
VC_2	OJ_2	9.638516e-01	Not reject	Not reject	Not Reject

Observed that if we use the p-value for two sample t-test, 13 out of 15 treatment groups will have different tooth length means at 95 percent confidence level; If we control the family-wise error rate, 10 out of 15 treatment groups will have different tooth length means at 95 percent confidence level; if we control the false-discovery rate, 13 out of 15 treatment groups will have different tooth length means at 95 percent confidence level.

Conclusion

1. Mean tooth length for animals who recieve more than 1 mg orange juice (1 or 2 mg) per day seems have no difference with the mean tooth length for those animals who recieve 2 mg vitamin C per day.
2. Amount of VC or orange juice recieved by animal daliy does influence the tootch length.