Hypothersis test on ToothGrowth data

Michael Lagunov

8/15/2020

Synopsis

In this part an exploratory and hypothesis test analysis conducted on the Tooth Growth dataset.

The dataset shows 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) (From the help page)

Data and libraries setting

```
knitr::opts_chunk$set(echo = TRUE, dev = "cairo_pdf")
library(ggplot2) #beautiful plot
library(kableExtra) #beautiful table
library(extrafont) #Times New Roman font for the plot
```

```
## Registering fonts with R
```

```
loadfonts(device = "win", quiet = TRUE) #for extrafont

data(ToothGrowth)
```

Exploring Analysis

First, see the dimensions and structure using dim and stf functions respectively

```
dim(ToothGrowth)
```

```
## [1] 60  3
```

```
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 ...
```

60 observations and 3 variables.

From the help page we see, that len represent Tooth length, supp - Supplement type (Orange juice - OJ, or ascorbic acid - VC) and dose - Dose in milligrams/day

Let's see some piece of data

```
head(ToothGrowth)
```

```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

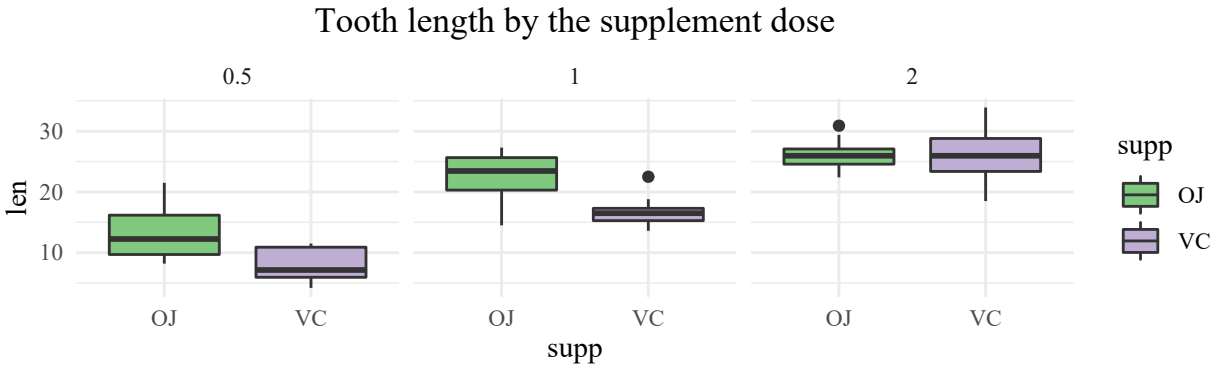
```
summary(ToothGrowth)
```

```
##       len        supp         dose
##  Min.   : 4.20   OJ:30   Min.   :0.500
##  1st Qu.:13.07   VC:30   1st Qu.:0.500
##  Median :19.25           Median :1.000
##  Mean   :18.81           Mean   :1.167
##  3rd Qu.:25.27           3rd Qu.:2.000
##  Max.   :33.90           Max.   :2.000
```

There are two equal groups of pigs, separated by the supplement, and the dose ranged from .5 to 2 with the median of 1 milligrams/day

Finally, make a plot of the tooth length by dose

```
ggplot(aes(x = supp, y = len), data = ToothGrowth) +
    geom_boxplot(aes(fill = supp)) + facet_wrap(~ dose) +
  ggtitle("Tooth length by the supplement dose") +
  theme_minimal() +
  theme(text=element_text(family="Times New Roman"),
        plot.title = element_text(hjust = 0.5)) +
  scale_fill_brewer(palette="Accent")
```



From the plot we see positive correlation between the amount of dose and the tooth growth. Also, orange juice seems to provide with better effect except the dose of 2, although not obvious whether the effect significant. For that, make a hypothesis test

Hypothesis test

H0 - there is no difference in the mean H1 - the mean is different alpha = 5%

```
t_full <- t.test(len ~ supp, data = ToothGrowth)
t_full
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 55.309, p-value = 0.06063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1710156  7.5710156
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

p-value is 0.0606345 which is larger than 0.05, so the NULL hypothesis cannot be rejected. Also, confidence interval is from -0.1710156 to 7.5710156, zero is falling into it.

For the sake of interest, make a hypothesis tests based on the specific supplement dose amount

```
dose_05 <- subset(ToothGrowth, dose == 0.5)
dose_1 <- subset(ToothGrowth, dose == 1)
dose_2 <- subset(ToothGrowth, dose == 2)

t_05 <- t.test(len ~ supp, data = dose_05)
t_1 <- t.test(len ~ supp, data = dose_1)
t_2 <- t.test(len ~ supp, data = dose_2)

kable(data.frame("Dose" = c(0.5, 1, 2),
           "P-value" = c(t_05$p.value, t_1$p.value, t_2$p.value),
           "Conf.Interval.Low" = c(t_05$conf.int[1], t_1$conf.int[1], t_2$conf.int[1]),
           "Conf.Interval.High" = c(t_05$conf.int[2], t_1$conf.int[2], t_2$conf.int[2])),
      "latex", booktabs = T) %>%
  kable_styling(position = "center")
```

Dose	P.value	Conf.Interval.Low	Conf.Interval.High
0.5	0.0063586	1.719057	8.780943
1.0	0.0010384	2.802148	9.057852
2.0	0.9638516	-3.798071	3.638070

From the splitted test we see, that for the dose of 0.5 and 1 the p-value is very low, but for the last dose of 2 p-value is extremely high. This is quite surprising, ascorbic acid showed very good results on the high dose. I would recommend to conduct the experiment with more participants

Conclusion

Results showed that dosage affect the tooth length, although supplement difference doesn't provide enough significance level with p-value of 0.06. On the high dosage of different supplement there is no difference, whereas on the lower dosage the significant level is very high