Final assignment of the Statistical Inference course. Part 2

Ivan Tiunov

In this part of the final assignment we are going to analyze the ToothGrowth data and compare tooth growth by supplement and dose using t-tests.

Required packages:

```
library(ggplot2)
library(dplyr, warn.conflicts = FALSE)
```

1. Exploratory analyses of the data

```
data("ToothGrowth")
dim(ToothGrowth)
```

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

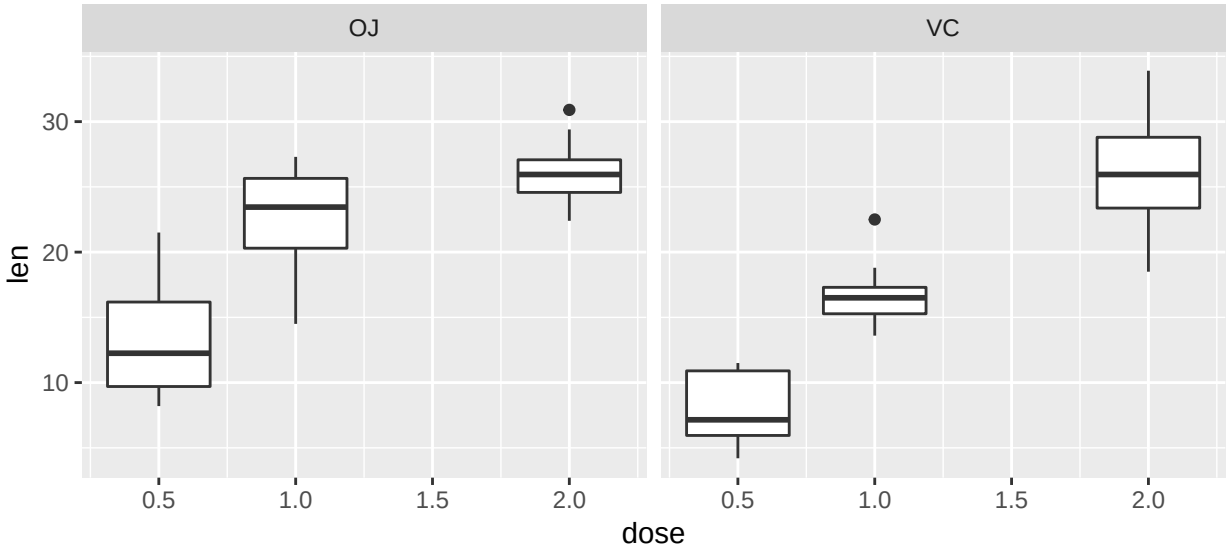
```
str(ToothGrowth)
```

```
## '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 ...
```

```
table(ToothGrowth$dose)
```

```
##
## 0.5   1   2
##  20  20  20
```

```
ggplot(data = ToothGrowth, aes(x = dose, y = len)) +
    geom_boxplot(aes(group = dose)) + facet_grid(.~supp)
```



2. Summary of the data

The data describe 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 (`OJ`) or ascorbic acid (`VC`). Each dose-supplement combination consists of ten data points. We can see from the plot that tooth length seems to increase with higher dose of both supplements. We cas assume that the `OJ` has higher effect on tooth growth for doses 0.5 and 1 mg/day.

3. Comparison of tooth length by supplement and dose

We conduct a number of t-test to see which supplement has larger effect on tooth growth and if there is a statistical difference in tooth growth for different doses within each supplement. We assume independence and different variance between each of six groups in the study.

```
knitr::kable(
    group_by(ToothGrowth, dose, supp) %>%
        summarise(Variance = round(var(len), 3)))
```

dose	supp	Variance
0.5	OJ	19.889
0.5	VC	7.544
1.0	OJ	15.296
1.0	VC	6.327
2.0	OJ	7.049
2.0	VC	23.018

First, let's compare the effect of each supplement at the same dose. Our null hypothesis is that there is no difference between supplements. Alternative hypothesis is that supplement `OJ` has greater effect on tooth growth.

```
d <- ToothGrowth  # to make name shorter
supOJ <- d$supp == "OJ"
supVC <- d$supp == "VC"

sup.05 <- t.test(d$len[supOJ & d$dose == 0.5],
                 d$len[supVC & d$dose == 0.5], alternative = "g")
sup.10 <- t.test(d$len[supOJ & d$dose == 1],
                 d$len[supVC & d$dose == 1], alternative = "g")
sup.20 <- t.test(d$len[supOJ & d$dose == 2],
                 d$len[supVC & d$dose == 2], alternative = "g")
knitr::kable(
    data.frame(Dose = c(0.5, 1.0, 2.0),
               p.value = round(c(sup.05$p.value, sup.10$p.value, sup.20$p.value),3),
               CI_low = round(c(sup.05$conf.int[1], sup.10$conf.int[1], sup.20$conf.int[1]), 3)),
    col.names = c("Dose", "p-value", "$CI_{\\text{low}}$")
)
```

Dose	p-value	CI_{low}
0.5	0.003	2.346
1.0	0.001	3.356
2.0	0.518	-3.133

The table shows p-palues and lower bounds of condidence interval for out t-tests. As can be seen, p-values for doses of 0.5 and 1 mg/day are much lower than alpha values for out confidence level (95 %), and confidence interval do not include zero. So, we can reject our null hypothesis and conclude that orange juce has larger effect on tooth growth that ascorbic acid at doses of 0.5 and 1 mg/day. However, there is no significant difference between oranje juice and ascorbic acid at dose of 2 mg/day due to the high p-value, that is much larger than out alpha.

Now, let's run t-tests to see if `OJ` or `VC` at doses of 2 mg/day have greater effect on tooth growth than `OJ` at concentration 1 mg/day. First test: $H_0 : \mu_{OJ,2} = \mu_{OJ,1}, H_a : \mu_{OJ,2} > \mu_{OJ,1}$:

```
oj_vs_oj <- t.test(d$len[supOJ & d$dose == 2],
                   d$len[supOJ & d$dose == 1], alternative = "g")
```

Second test: $H_0 : \mu_{VC,2} = \mu_{OJ,1}, H_a : \mu_{VC,2} > \mu_{OJ,1}$:

```
vc_vs_oj <- t.test(d$len[supVC & d$dose == 2],
                   d$len[supOJ & d$dose == 1], alternative = "g")
knitr::kable(
    data.frame(H0 = c("$\\mu_{OJ, 2} = \\mu_{OJ,1}$",
                      c("$\\mu_{VC, 2} = \\mu_{OJ,1}$")),
               p.value = round(c(oj_vs_oj$p.value, vc_vs_oj$p.value), 3),
               CI_low = round(c(oj_vs_oj$conf.int[1], vc_vs_oj$conf.int[1]),3)),
    col.names = c("$H_0$", "p-value", "$CI_{\\text{low}}$")
)
```

H_0	p-value	CI_{low}
$\mu_{OJ,2} = \mu_{OJ,1}$	0.020	0.749
$\mu_{VC,2} = \mu_{OJ,1}$	0.048	0.038

The table shows p-values and lower bounds for both hypothesis. In this case we can sagely reject the null hypothesis in favor of the alternative due to the low p-values, and conclude that orange juice and ascorbic acid at dose of 2 mg/day have greater effect on tooth growth that oranje juice at dose of 1 mg/day.