Statistical Inference Course Project

Veasna Kheng

23/07/2020

Part 2: Basic Inferential Data Analysis

Overview

This second part of the project reports the basic inferential analysis of the ToothGrowth in the R datasets package.

1. Load the ToothGrowth Data and perform some basic exploratory analysis

```r
data("ToothGrowth")
head(ToothGrowth, 5)
```

```
##    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
```

```r
str(ToothGrowth)
```

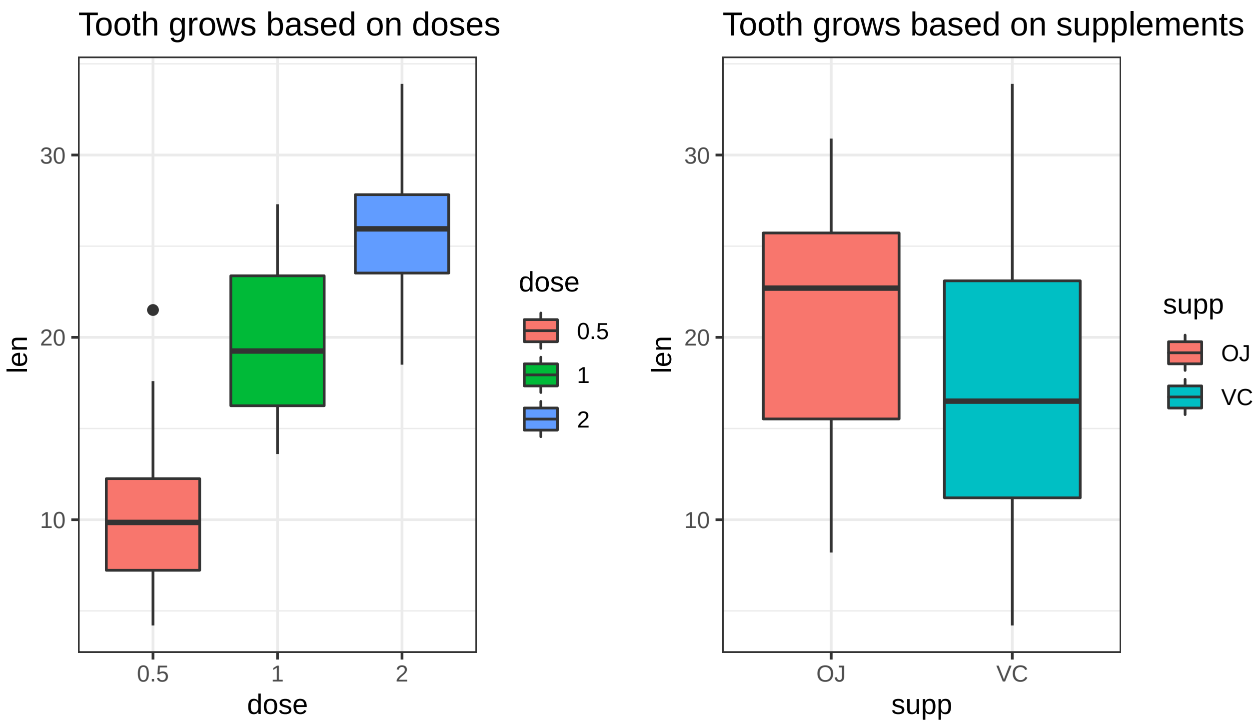
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```r
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  
```

```r
library(ggplot2)
library(cowplot)
f1 <- ggplot(ToothGrowth, aes(supp, len, fill = supp)) +
  geom_boxplot(show.legend = T)+
  labs(title = "Tooth grows based on supplements") +
  theme_bw()
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
f2 <- ggplot(ToothGrowth, aes(dose, len, fill = dose)) +
```

```r
  geom_boxplot() +
  labs(title = "Tooth grows based on doses") +
  theme_bw()
plot_grid(f2,f1)
```



The left panel of the boxplot above shows that dose of vitamins has positive effects on tooth growth: The higher dose, the faster tooth grows. The right panel indicates that OJ's average effect appear to be higher than that of VC. These observations lead the following hypothesis tests.

2. Hypothesis Tests

The tests are based on the following assumptions:

- All subjects are i.i.d variables
- Variances of tooth growth (len variable) differ according to a type of vitamin.
- The distribution of sample means are approximately to normal according to the CLT.

2.1 The test is based on a type of supplement

- H_o: average effect of OJ is not different from that of VC
- H_a: average effect of OJ is greater than that of VC

To proceed this one sided-test, we first need to calculate means and variances of these two independent samples: one is based on OJ and the other based on VC.

```r
# separate data
S1 <- ToothGrowth[ToothGrowth$supp == "OJ",]
S2 <- ToothGrowth[ToothGrowth$supp == "VC",]
# mean and variance of sample 1 based on OJ
m1 <- mean(S1$len)
v1 <- var(S1$len)
# mean and variance of sample 1 based on VC
```

```
m2 <- mean(S2$len)
v2 <- var(S2$len)
```

Next, we can estimate a test statistic by

$$t = \frac{m1 - m2}{\sqrt{\frac{v1}{n1} + \frac{v2}{n2}}},$$

where $n1$ and $n2$ are sizes of sample 1 and 2, respectively; and t has distribution $T(df)$ where degree of freedom

$$df = \frac{\left(\frac{v1}{n1} + \frac{v2}{n2}\right)^2}{\frac{(v1/n1)^2}{n1-1} + \frac{(v2/n2)^2}{n2-1}}$$

```
# calculate a test statistic and degree of freedom
n1 <- length(S1$len)
n2 <- length(S2$len)
t <- (m1 - m2) / sqrt(v1/n1 + v2/n2)
df <- (v1/n1 + v2/n2)^2 / ( (v1/n1)^2 / (n1-1) + (v2/n2)^2 / (n2-1) )
# confident interval with alpha = 0.05
m1-m2 + c(-1,1) * qt(0.95, df) * sqrt(v1/n1 + v2/n2)
```

```
## [1] 0.4682687 6.9317313
```

Since this interval does not contain zero, we have enough evidence to reject the null and conclude that the average effect of OJ on tooth growth is greater than that of VC.

2.2 The test is based on doses

a. **Dose 0.5 Versus Dose 1**

- H_o: effect of dose 0.5 is not different from that of dose 1
- H_a: effect of dose 0.5 is smaller than that of dose 1

```
d0 <- ToothGrowth$len[ToothGrowth$dose == 0.5]
d1 <- ToothGrowth$len[ToothGrowth$dose == 1]
# p-value is less than 5 percent level?
t.test(d1,d0, alternative = "greater", paired = FALSE, var.equal = FALSE,
       conf.level = 0.95)$p.value < 0.05
```

```
## [1] TRUE
```

Because p-value is less than 5% level, we have enough evidence to reject the null and conclude that the effect of dose 1 on toot growth is greater than that of dose 0.5.

b. **Dose 1 Versus Dose 2**

- H_o: effect of dose 1 is not different from that of dose 1
- H_a: effect of dose 1 is smaller than that of dose 2

```
d2 <- ToothGrowth$len[ToothGrowth$dose == 2]
# p-value is less than 5 percent level?
t.test(d2,d1, alternative = "greater", paired = FALSE, var.equal = FALSE,
       conf.level = 0.95)$p.value < 0.05
```

```
## [1] TRUE
```

Because p-value is less than 5% level, we have enough evidence to reject the null and conclude that the effect of dose 2 on toot growth is greater than that of dose 1.
