Statistical Inference: Peer Assessment Part 2

Part 2: Basic Inferential Analysis

Overview

In the second part of the project, the ToothGrowth data in the R datasets package will be analysed. The toothgrowth dataset shows the effect of Vitamin C on Tooth Growth in Guinea Pigs.

1. Load the ToothGrowth data and perform some basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)
4. State your conclusions and the assumptions needed for your conclusions.

Load the ToothGrowth data and perform basic exploratory data analysis

load libraries

```
library(knitr)
library(ggplot2)
library(datasets)
```

basic exploratory data analysis

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

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

```
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
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

```r
head(ToothGrowth, 10)
```

```
##     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
## 7  11.2   VC  0.5
## 8  11.2   VC  0.5
## 9   5.2   VC  0.5
## 10  7.0   VC  0.5
```

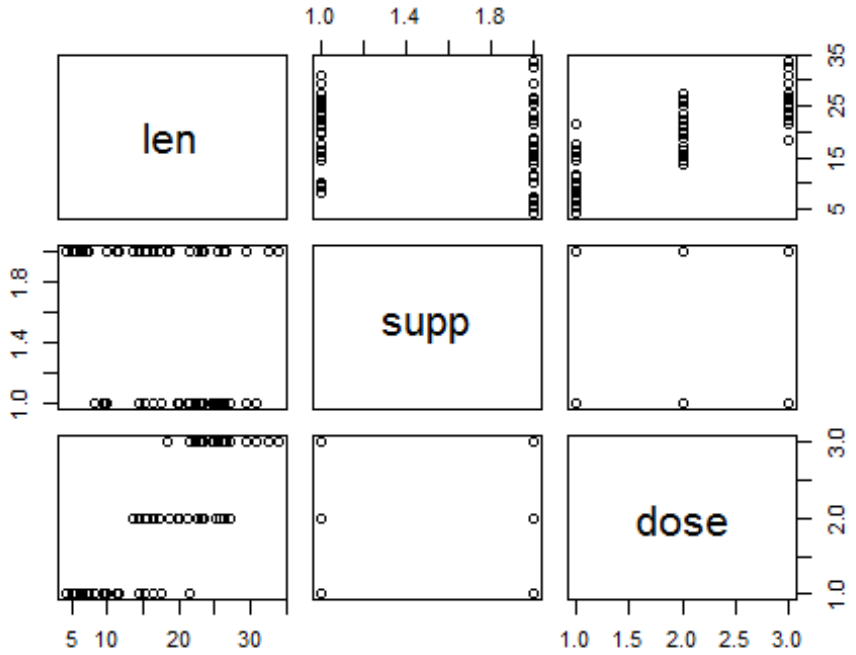
```r
tail(ToothGrowth, 10)
```

```
##     len supp dose
## 51 25.5   OJ    2
## 52 26.4   OJ    2
## 53 22.4   OJ    2
## 54 24.5   OJ    2
## 55 24.8   OJ    2
## 56 30.9   OJ    2
## 57 26.4   OJ    2
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

```r
table(ToothGrowth$supp, ToothGrowth$dose)
```

```
##
##      0.5  1  2
##   OJ  10 10 10
##   VC  10 10 10
```

```
pairs(ToothGrowth)
```

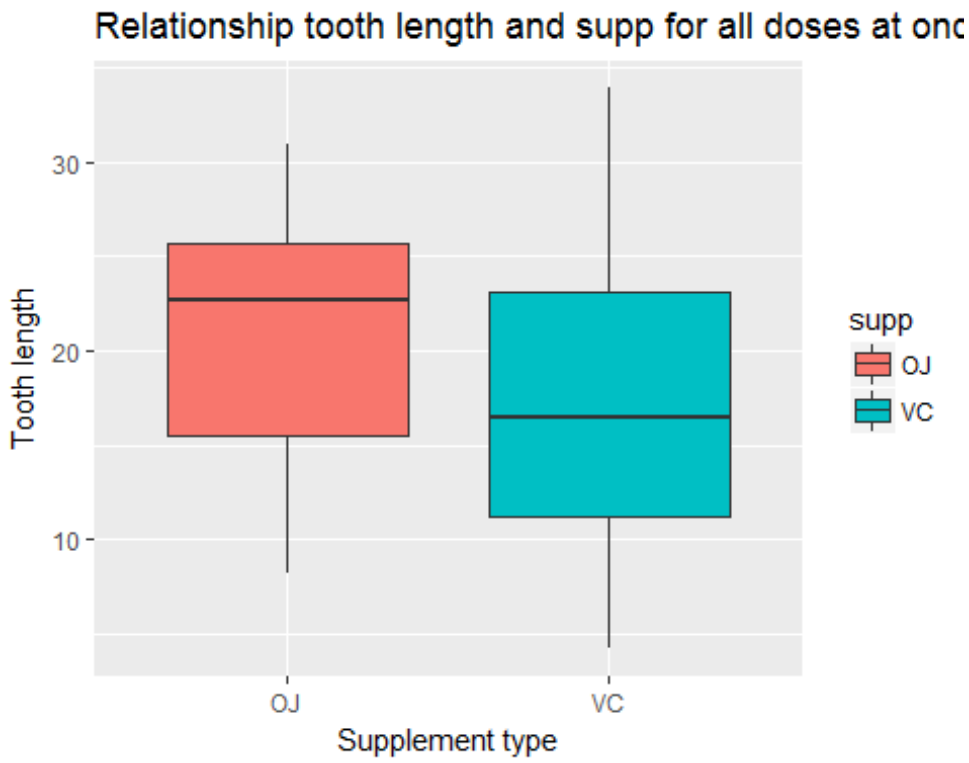


From the matrix of scatterplots made with the function "pairs()" it seems to be interesting to make some boxplots with ggplot2 in order to see the relationship between Tooth length (len), Dose levels of vitamin C (dose in mg/day) and supplement type (supp): VC = ascorbic acid and OJ = orange juice.

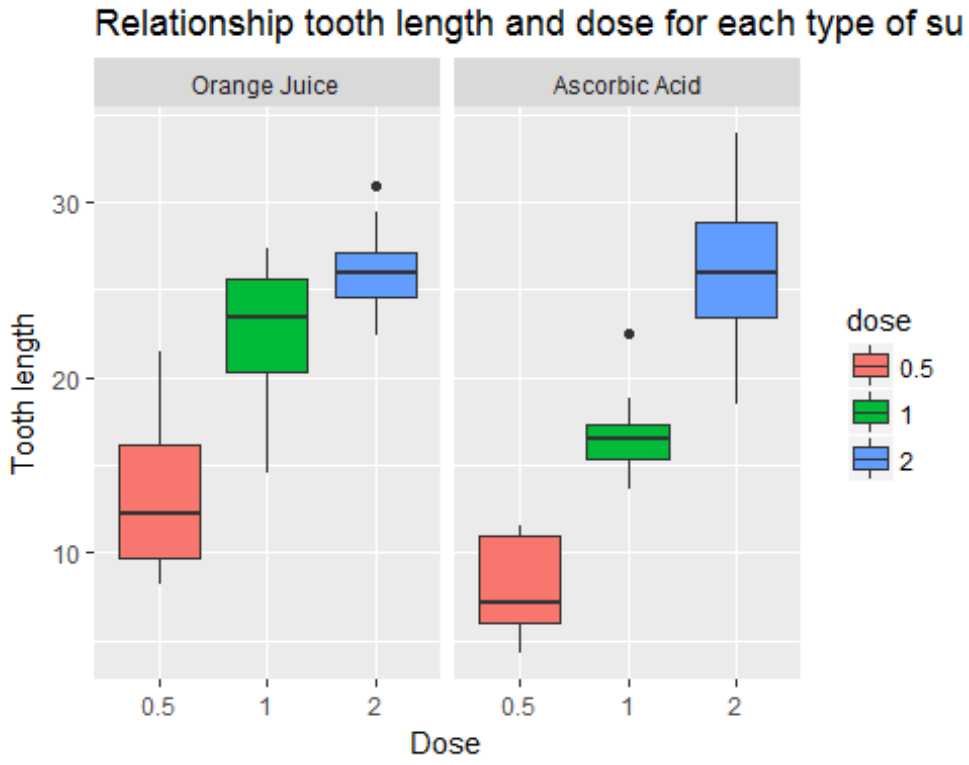
```
box_plot <- ggplot(ToothGrowth, aes(x = supp, y = len, fill = supp)) +
  geom_boxplot() +
  labs(title = "Relationship tooth length and supp for all doses at once",
       x = "Supplement type", y = "Tooth length")

levels(ToothGrowth$supp) <- c("Orange Juice", "Ascorbic Acid")
box_plot_panels <- ggplot(ToothGrowth, aes(x = dose, y = len, fill = dose)) +
  geom_boxplot() +
  facet_grid(. ~ supp) +
  labs(title = "Relationship tooth length and dose for each type of supplemen
t",
       x = "Dose", y = "Tooth length")

box_plot
```




```
box_plot_panels
```



Basic summary of the data

The basic exploratory data analysis above shows that the ToothGrowth data has 60 observations and 3 variables: len, supp and dose. Len is numeric, Supp is a factor with 2 levels (OJ and VC) and Dose is numeric with three levels of vitamin C (0,5; 1 and 2 mg/day).

The first boxplot seems to show that tooth growth increases more with orange juice than with ascorbic acid for all dose levels at once.

Furthermore, it seems that the mean toothlength is higher with orange juice than with ascorbic acid for the dose levels 0,5 and 1 mg/day. For dose level 2 mg/day the mean tooth lengths seem to be close to each other for orange juice and ascorbic acid.

Let's test this in the paragraph below.

Confidence intervals and hypothesis tests

Hypothesis 1

H0: the supplement type has no impact on the tooth growth for all doses at once. Ha: the supplement orange juice has more impact on the tooth growth for all doses at once.

```
OrJ <- subset(ToothGrowth, supp == "Orange Juice")
AsA <- subset(ToothGrowth, supp == "Ascorbic Acid")
t.test(OrJ$len, AsA$len, paired = FALSE, alternative = "greater")
```

```
##
##  Welch Two Sample t-test
##
## data:  OrJ$len and AsA$len
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687       Inf
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

The null hypothesis, H0, can be rejected on basis of the t-test results. The hypothesis Ha, the alternative, can be accepted.

Hypothesis 2 to 4

For the dose levels 0.5, 1 and 2 mg/day, the next 3 tests will be done. H0: the supplement type has no impact on the tooth growth for the tested dose level. Ha: the supplement orange juice has more impact on the tooth growth for the tested dose level.

Because there is multiple-testing a BH (Benjamin & Hochberg) correction will be used as adjustment method.

```
Hyp <- c()
Hyp[1] <- t.test(OrJ[OrJ$dose == 0.5,]$len, AsA[AsA$dose == 0.5,]$len, paired
= FALSE, alternative = "greater")$p.value
Hyp[2] <- t.test(OrJ[OrJ$dose == 1,]$len, AsA[AsA$dose == 1,]$len, paired = F
ALSE, alternative = "greater")$p.value
Hyp[3] <- t.test(OrJ[OrJ$dose == 2,]$len, AsA[AsA$dose == 2,]$len, paired = F
ALSE, alternative = "greater")$p.value
p.adjust(Hyp, method = "BH")
```

```
## [1] 0.004768955 0.001557564 0.518074206
```

See the next paragraph for the conclusions.

Conclusions and assumptions

Assumptions

- The subjects (guinea pigs) that got the different 3 dose levels in combination with the different 2 supplement types are supposed to be different. So the 6 groups are independent and not paired (paired = FALSE)
- The mean tooth length is supposed to be distributed normal
- The guinea pigs are representative for the total population of guinea pigs

Conclusions

With the t-tests is shown that:

- the supplement orange juice has more impact on the tooth growth for all doses at once
- the supplement orange juice has more impact on the tooth growth for dose level 0.5 mg/day, because the null hypothesis is rejected by the p-value of 0.004768955
- the supplement orange juice has more impact on the tooth growth for dose level 1 mg/day, because the null hypothesis is rejected by the p-value of 0.001557564
- the supplement type has no impact on the tooth growth for dose level 2 mg/day, because the null hypothesis is accepted by the p-value of 0.518074206.