Course Project Part 2 - Analyze the ToothGrowth data

Tatiana Lezhneva

Overveiw

There is a dataset **ToothGrows** which shows the effect of vitamin C on tooth growth in guinea pigs. The response 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 (OJ) or ascorbic acid (a form of vitamin C and coded as VC). We are going to do some exploratory data analyses and then compare tooth growth by dilivery method and dose.

1. Exploratory data analyses

```
rawdata <- ToothGrowth
str(rawdata)
```

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

So we have 60 observations on 3 variables: Tooth length, Supplement type and Dose in milligramm/day. Transformormation of variable Dose to factor:

```
rawdata$dose <- as.factor(rawdata$dose)
str(rawdata)
```

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

Summary statictics:

```
summary(rawdata)
```

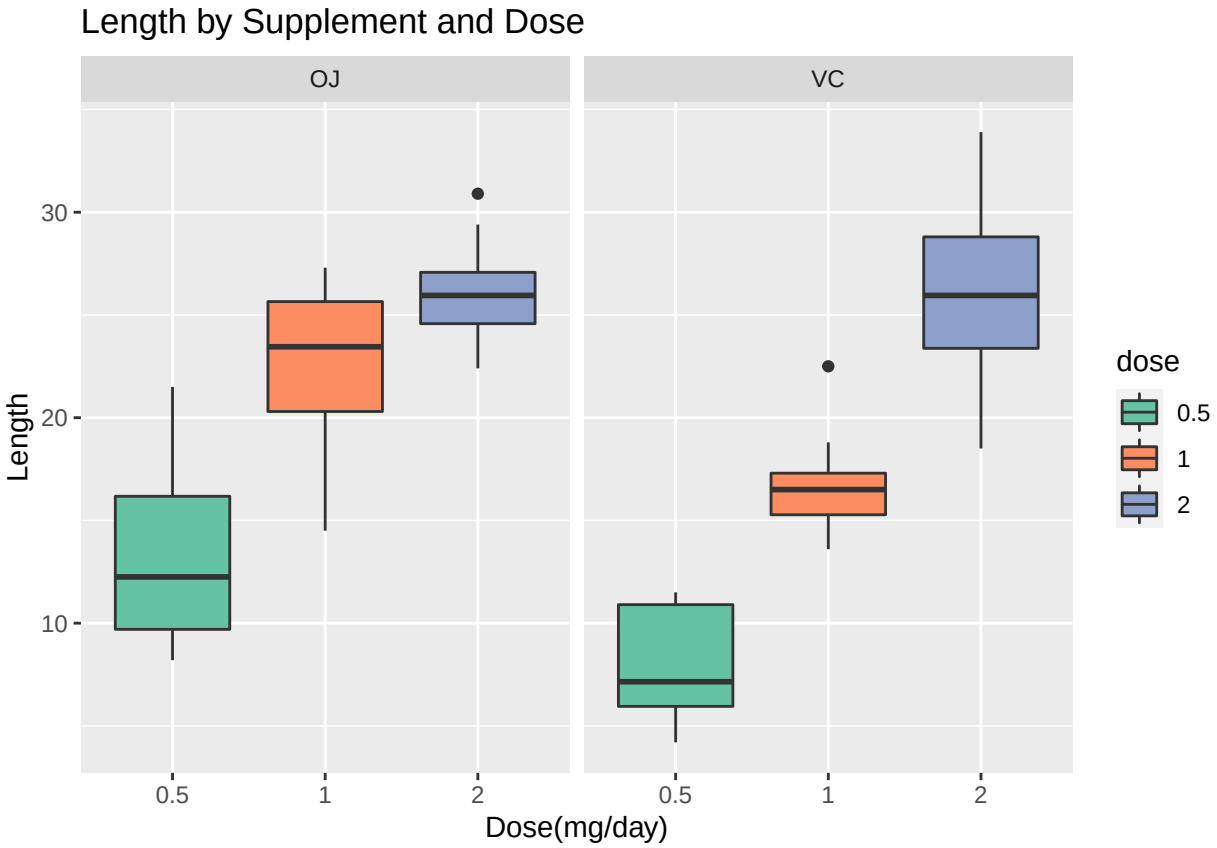
```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

Lets plot boxplots length by dose for every delivery method

```
library(ggplot2)

ggplot(rawdata, aes(y = len, x = dose)) +
    geom_boxplot(aes(fill = dose)) +
    facet_wrap(~ supp ) +
```

```
    scale_fill_brewer(palette = "Set2") +
    labs(x = "Dose(mg/day)", y = "Length") +
    ggtitle("Length by Supplement and Dose ")
```



Looks like the amount of dose has an impact on length unlike delivery method but lets test both hypotheses.

2. Hythoteses Testing

As we are going to use T-test, lets assume:

- all observations are independant
- pigs were chosen randomly.

2.1. Impact of Delivery Method of Vitamin C

H_0: The means of different delivery method groups are the same.

H_a: The means of different delivery method groups are different.

```
t.test(len ~ supp, data = rawdata)
```

```
##
##  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
```

2.2. Impact of Vitamin C Dose

H_0: The means of groups (dose = 0.5 and 2) are the same.

H_a: The means of groups (dose = 0.5 and 2) are different.

```
t.test(rawdata$len[rawdata$supp == "OJ"], rawdata$len[rawdata$supp == "VC"])
```

```
##
##  Welch Two Sample t-test
##
## data:  rawdata$len[rawdata$supp == "OJ"] and rawdata$len[rawdata$supp == "VC"]
## 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 of x mean of y
##  20.66333  16.96333
```

```
t.test(len ~ dose, data = rawdata[rawdata$dose %in% c(0.5, 2),])
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -11.799, df = 36.883, p-value = 4.398e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -18.15617 -12.83383
## sample estimates:
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

Conclusions

In the first testing of delivery methods, we've got the p-value 0.06. At first glance the p-value is greater than 5% significant level and says that we can't reject H_0. But on the other hand p-value is so close to the significant level that I would prefer to say that more observations are needed. In the second testing about vitamin C dose we've got p-value close to 0 that says we have reasons to reject H_0 and the amount of vitamin C dose does have an impact on tooth growth.