Basic Inferential Data Analysis

Volha Leusha

In this paper analysis of the ToothGrowth dataset from Rdataset package is presented and basic summary of data is provided. In second part confidence intervals and hypothesis tests are used to compare tooth growth by supp and dose, with subsequent conclusions on results.

Dataset Base Analysis

1. Load dataset from datasets library:

```
#Load the ToothGrowth data:
library(datasets)
data("ToothGrowth")
```

2. Basic summary of data exploratory data analysis:

```
##    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
```

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

```
##       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
```

The dataset 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).

Dataset has 60 observarions of 3 variables:

- supp: factor with only two possible supplement types: OJ, VC with equal proportion
- dose: numerical categorical variable that can take only 3 unique values (milligrams/day)

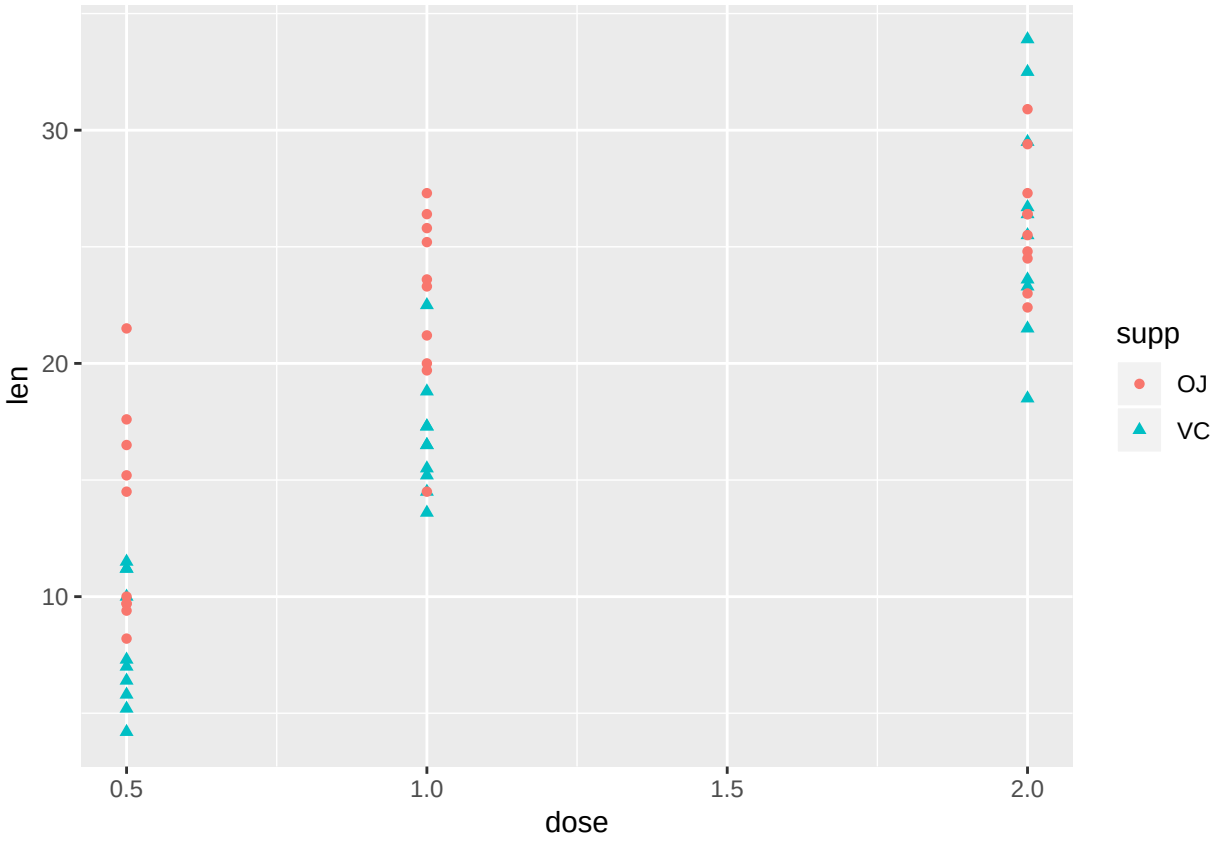
```
## [1] 0.5 1.0 2.0
```

- len: numerical tooth length in range from 4.2 to 33.90 with mean 18.81 and standard deviation

```
## [1] 7.649315
```

Tooth growth comparisson by supp and dose

Plot dataset grouped by supplement type and dosage:



From the graph and table it can be seen that there is potential difference in tooth length between supplements, suggesting that pigs on supplement OJ have longer teeth than on VC. Also, the bigger the dossage, the longer are teeth. To prove above assumptions, three unpaired t-tests are performed, taking into account following assumptions:

- The sample is representative of the population
- Data follows bell-shaped distribution curve
- Standard deviations of samples are approximately equal

1. First test:
   Ho: there is no difference between tooth length with different supplements
   Ha: there is difference between tooth length with different supplements
   alpha =.05

```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1670064  7.5670064
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

Confidence interval includes zero (and p-value is greater than alpha) consequently at 95% confidence level, there is no significant difference of the two means.

2. Second test:
   Ho: there is no difference between dose0.5 and 1.0 for tooth length
   Ha: tooth length for dose 1.0 is bigger than for dose 0.5
   alpha= .05

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 1] and ToothGrowth$len[ToothGrowth$dose == 0.5]
## t = 6.4766, df = 38, p-value = 6.331e-08
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  6.753344      Inf
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

Confidence interval is fully above zero (and p-value is less than alpha) consequently at 95% confidence level null hypothesis can be rejected and it can be assumed that Ha is true.

3. Third test:
   Ho: there is no difference between dose 1.0 and 2.0 for tooth length
   Ha: tooth length for dose 2.0 is higher than for dose 1.0
   alpha= .05

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 2] and ToothGrowth$len[ToothGrowth$dose == 1]
## t = 4.9005, df = 38, p-value = 9.054e-06
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  4.175196      Inf
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

Confidence interval is fully above zero (and p-value is less than alpha) consequently at 95% confidence level null hypothesis can be rejected and it can be assumed that Ha is true.

Conclusions:

1. There is no enough evidence to say that supplement type effects tooth length

2. Dosage increase results in tooth length increase

Appendix

Following R code is used to generate results in the paper:

```
#Load the ToothGrowth data:

library(datasets)
```

```r
data("ToothGrowth")

#first glance on the data:
head(ToothGrowth)

#look into data description:
str(ToothGrowth)
summary(ToothGrowth)

#check for unique dose values
unique(ToothGrowth$dose)

#find standard deviation for length
sd(ToothGrowth$len)

#open ggplot2 library:
library(ggplot2)

#Plot dataset grouped by supplement type and dosage
ggplot(ToothGrowth, aes(x=dose, y=len, shape=supp, color=supp)) +geom_point()+
    theme(plot.margin = margin(3,3,3,3, "cm"))
```

```r
#1. First test:
#Ho: there is no difference between tooth length with different supplements
#Ha: tooth length for supplement OJ in higher
##------------------------------------------------------------------------------
# perform t-test for length vs supplement type:
t.test(len~supp, data= ToothGrowth, var.equal = TRUE, paired= FALSE)

#Second test:
#Ho: there is no difference between dose0.5 and 1.0 for tooth length
#Ha: tooth length for dose 1.0 is higher than for dose 0.5
#------------------------------------------------------------------------------
## perform t-test for length vs dosage 1.0 or 2.0:
t.test(ToothGrowth$len[ToothGrowth$dose==1],ToothGrowth$len[ToothGrowth$dose==.5],
       var.equal =TRUE, alternative = 'greater')

# Third test:
#Ho: there is no difference between dose 1.0 and 2.0 for tooth length
#Ha: tooth length for dose 2.0 is higher than for dose 1.0
#------------------------------------------------------------------------------
## perform t-test for length vs dosage 1.0 or 2.0:
t.test(ToothGrowth$len[ToothGrowth$dose==2],ToothGrowth$len[ToothGrowth$dose==1],
       var.equal =TRUE, alternative = 'greater')
```