Part 2: Basic Inferential Data Analysis

Olha Klishchuk

2023-10-12

Overview

```
library(datasets)
suppressPackageStartupMessages(library(dplyr))
```

Data
A data frame with 60 observations on 3 variables.

[,1] len numeric Tooth length.
[,2] supp factor Supplement type (vitamin C or other J).
[,3] dose numeric Dose in milligrams/day.

```
data(ToothGrowth)
head(ToothGrowth)
```

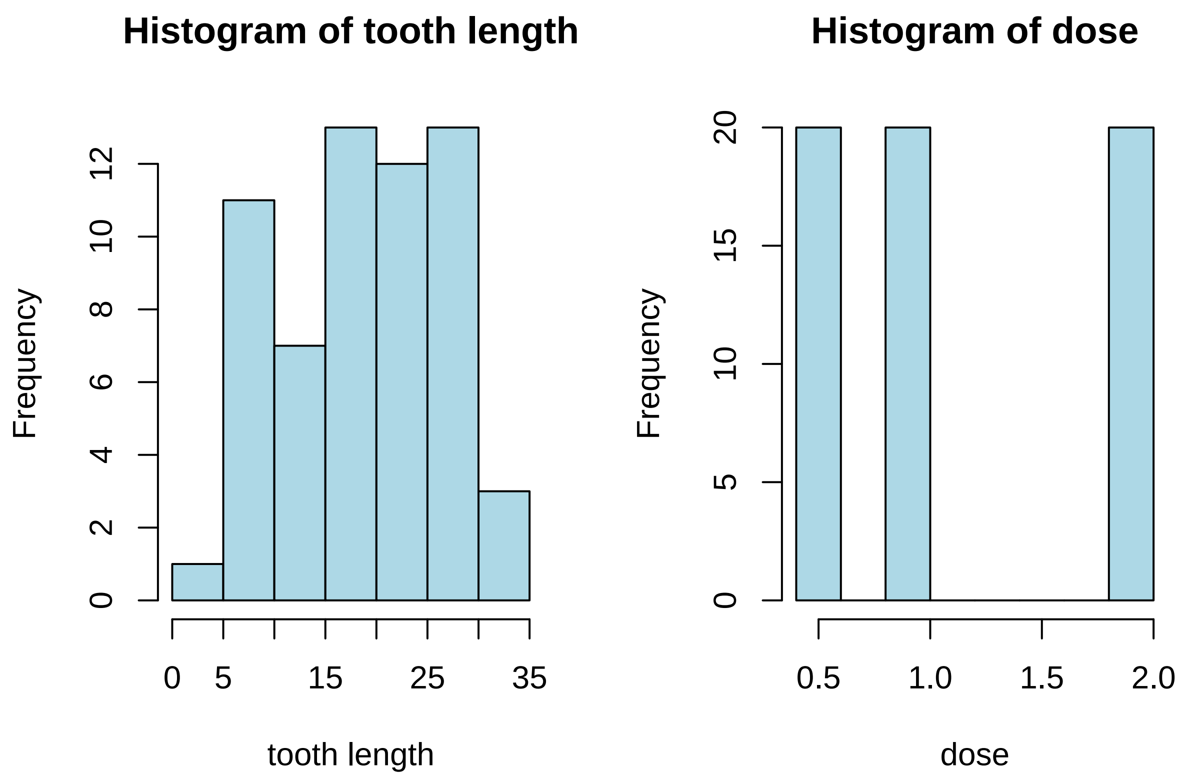
```
##    len supp dose
## 1  4.2   VC  0.5
## 2 11.5   VC  0.5
## 3  7.3   VC  0.5
## 4  5.8   VC  0.5
## 5  6.4   VC  0.5
## 6 10.0   VC  0.5
```

Description

```
# Descriptive Statistics and EDA
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
df<-ToothGrowth
options(repr.plot.height = 6)
par(mfcol = c(1,2))
with(df, hist(len, breaks = 10, main = "Histogram of tooth length",
              xlab = "tooth length", ylab = "Frequency", col = "lightblue"))
with(df, hist(dose, xlab = "dose", ylab = "Frequency", col = "lightblue"))
```



Plot 2. Distribution of data by key experimental variables

So from descriptive analysis, which you can find in appendix 2.1, we find out that mean effect from vitamin C injection was 16.96 cm of teeth growth, while in other treatment we obtained - 20.66 cm. Lets test hypothesis that differences between these two treatments are significant.

Hypothesis Testing

```r
# Hypothesis test Ha: m2-m1>0
# Paired test by supplement used
t.test(df[df[['supp']]=='OJ','len'], df[df[['supp']]=='VC','len'], paired = T)$ a
```

```
## [1] "two.sided"
```

```r
#Grouping data by dosage
dose_0.5v1<-
```

```r
df%>%
 group_by(dose)%>%
 mutate(dose=dose, supp = supp, len = len)%>%
 filter(dose %in% c(.5,1))

dose_0.5v2<-
df%>%
 group_by(dose)%>%
 mutate(dose=dose, supp = supp, len = len)%>%
 filter(dose != 1)

dose_1v2<-
df%>%
 group_by(dose)%>%
 mutate(dose=dose, supp = supp, len = len)%>%
 filter(dose > .5)
```

```r
# Hypothesis test Ha: B!=0
# Paired test by dosage
data_test<-
data.frame(statistics=c(t.test(len~dose, dose_0.5v1, paired = F)$statistic,
           t.test(len~dose, dose_0.5v2, paired = F)$statistic,
           t.test(len~dose, dose_1v2, paired = F)$statistic),
p.value = c(t.test(len~dose, dose_0.5v1, paired = F)$p.value,
           t.test(len~dose, dose_0.5v2, paired = F)$p.value,
           t.test(len~dose, dose_1v2, paired = F)$p.value),
           row.names = c('0.5v1', '0.5v2','1v2'))
data_test
```

```
##       statistics      p.value
## 0.5v1  -6.476648 1.268301e-07
## 0.5v2 -11.799046 4.397525e-14
## 1v2    -4.900484 1.906430e-05
```

Therefore, the type of supplement is used in treatment appeared to be significant and level of efficiency is tooth growth on 3.7 cm more then on vitamin C using.

```r
#paired T-confidence intervals
round(t.test(df[df[['supp']]=='OJ','len'], df[df[['supp']]=='VC','len'], paired
             = FALSE, conf.level = .95, var.equal = TRUE)$conf, 5)
```

```
## [1] -0.16701  7.56701
## attr(,"conf.level")
## [1] 0.95
```

Summary

- using other viitamins (A, D, E, alpha tocopherol) the tooth growth in guinea pigs were more significant then under solely supplementing vitamin C.
- The mean difference in about 3.7 cm.
- the standard deviation in observation is lower under more diet on -1.6605

References

- Crampton, E. W. (1947). The growth of the odontoblast of the incisor teeth as a criterion of vitamin C intake of the guinea pig. The Journal of Nutrition, 33(5), 491–504. doi:10.1093/jn/33.5.491.

Appendix

B

B.1 Descriptives

```
df<-ToothGrowth
mean_vc<-
    df%>%
    group_by(supp)%>%
    filter(supp == 'VC')%>%
    mutate(mean_len = mean(len), mean_dose = mean(dose))%>%
    select(-c(len, dose))

sd_vc<-
    df%>%
    group_by(supp)%>%
    filter(supp == 'VC')%>%
    mutate(sd_len = sd(len), sd_dose = sd(dose))%>%
    select(-c(len,dose))
cat('mean: ',as.numeric(mean_vc[1,]), 'sd:', as.numeric(sd_vc[1,]))
```

```
## mean:  2 16.96333 1.166667 sd: 2 8.266029 0.6342703
```

```
mean_oj<-
    df%>%
    group_by(supp)%>%
    filter(supp != 'VC')%>%
    mutate(mean_len = mean(len), mean_dose = mean(dose))%>%
    select(-c(len,dose))

sd_oj<-
    df%>%
    group_by(supp)%>%
    filter(supp != 'VC')%>%
    mutate(sd_len = sd(len), sd_dose = sd(dose))%>%
    select(-c(len,dose))
cat('mean:', as.numeric(mean_oj[1,]), 'sd: ',  as.numeric(sd_oj[1,]))
```

```
## mean: 1 20.66333 1.166667 sd:  1 6.605561 0.6342703
```

B.2 Hypothesis test

$H_a : B! = 0$ ##### Paired test by dosage

```r
t.test(len~dose, dose_0.5v1, paired = F)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means between group 0.5 and group 1 is not equal to 0
## 95 percent confidence interval:
##  -11.983781  -6.276219
## sample estimates:
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

```r
t.test(len~dose, dose_0.5v2, paired = F)
```

```
##
##  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 between group 0.5 and group 2 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
```

```r
t.test(len~dose, dose_1v2, paired = F)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by dose
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means between group 1 and group 2 is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean in group 1 mean in group 2
##          19.735          26.100
```