Statistical Inference Assignment_Part2

Edie Hawthorne

7/21/2018

Part 2: Basic Inferential Data Analysis

This part is the analysis of the ToothGrowth Data in the R datasets package.

Overview

This analysis is to use the confidence intervals to compare tooth growth by supp (supplement type) and dose (Dose in milligrams/day) in the ToothGrowth dataset.

Basic exploratory data analysis

After leading the ToothGrowth dataset, we found out there are 60 observations and three variables in this data frame. The variables are the Tooth Length (len), the supplement type VC or Oj (supp) which is a factor, and the dose in milligrams/day (dose).

```
library(datasets)
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 ...
```

Summary

From the summary of the dataset, we can see the 5 statistical numbers of each variable.

```
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
```

Subset

Subset the ToothGrowth data by supp (Supplement type) of VC and OJ as well as the dose (Dose in milligrams/day) of 0.5, 1 and 2.

```
group_vc_0.5 <- ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == "0.5","len
"]
group_vc_1 <- ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == "1","len"]
group_vc_2 <- ToothGrowth[ToothGrowth$supp == "VC" & ToothGrowth$dose == "2","len"]
group_oj_0.5 <- ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == "0.5","len
"]
group_oj_1 <- ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == "1","len"]
group_oj_2 <- ToothGrowth[ToothGrowth$supp == "OJ" & ToothGrowth$dose == "2","len"]
```

Confidence Intervals

Take the confidence intervals of each subset groups by the supplement type factor and dose. From the confidence interval matrix comparason we can see, we will reject the group_0.5 abd group_1 which are the supplement and does with 0.5 milligrams/day. We fail to reject group_2 which is the supplement and does with 2 milligrams/day because it's confidence interval contains mean 0.

```
library(dplyr)
confs <- as.data.frame(
    matrix(
        c(t.test(group_oj_0.5, group_vc_0.5, paired = FALSE, var.equal = FALSE)$conf,
          t.test(group_oj_1, group_vc_1, paired = FALSE, var.equal = FALSE)$conf,
          t.test(group_oj_2, group_vc_2, paired = FALSE, var.equal = FALSE)$conf),
        ncol=2,
        byrow=TRUE
    )
)
confs <- rename(confs, lower = V1, upper = V2)
confs$group <- c("group_0.5", "group_1", "group_2")

confs <- confs %>% select(group, lower, upper)
```

Comparison of the confidence Intervals steps:

group	lower	upper
group_0.5	1.719057	8.780943
group_1	2.802148	9.057852
group_2	-3.798071	3.638070