Statistical Inference -Part-2

Satyajit Das

Monday,June 22, 2020

This is the second part of Statistical Inference Project.

Requirements Now in the second portion of the class, we're going to analyze the ToothGrowth data in the R datasets package.

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 hypothesis tests to compare tooth growth by supp and dose. (Use the techniques from class even if there's other approaches worth considering)

4.State your conclusions and the assumptions needed for your conclusions.

Data Analysis

Load the data

```
library(ggplot2)
data(ToothGrowth)
```

1. Summary of the data

Take a look at first few rows

```
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
```

Type of Supplements and Dosages

```
table(ToothGrowth$supp) # Supplement
```

```
##
## OJ VC
## 30 30
```

```r
table(ToothGrowth$dose) # Dosage
```

```
##
## 0.5   1   2
##  20  20  20
```

Summary of ToothGrowth

```r
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
```

Correlation in ToothGrowth data

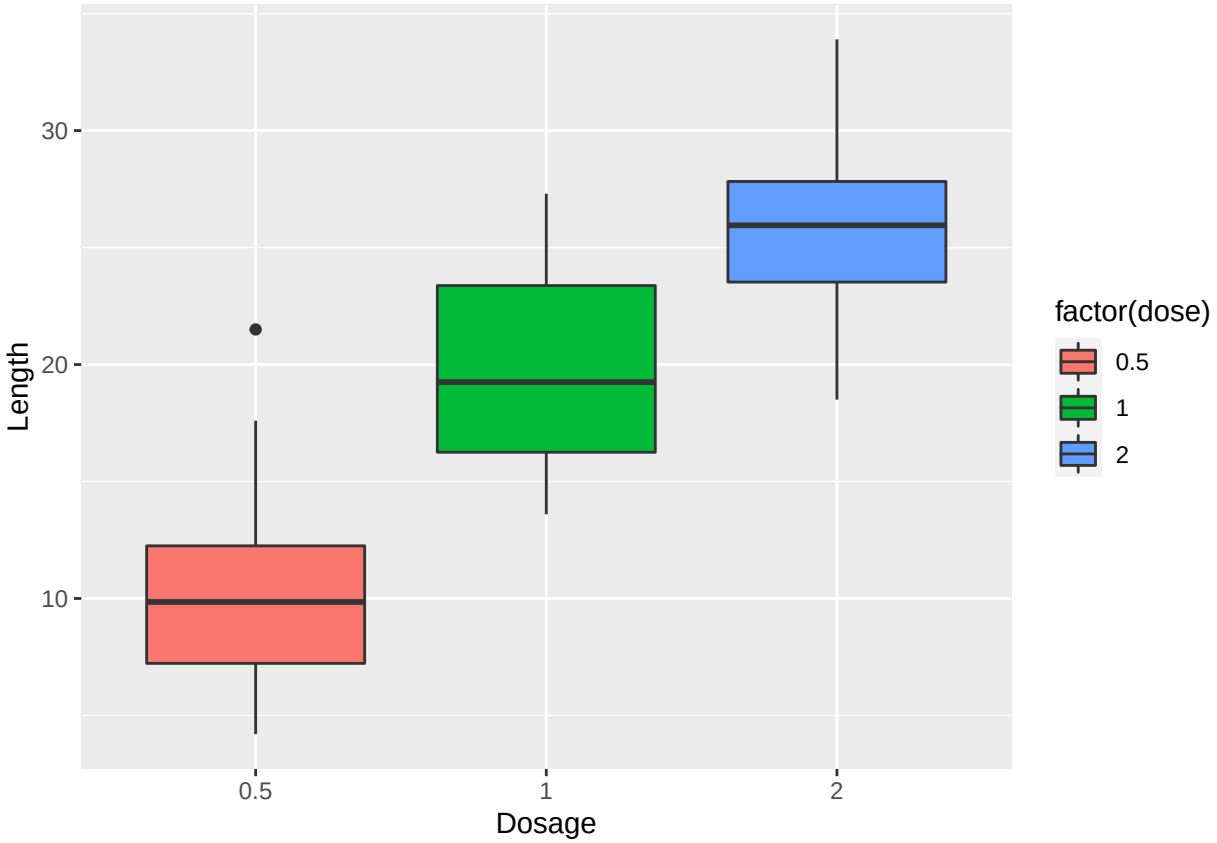
```r
cor(ToothGrowth[sapply(ToothGrowth, is.numeric)])
```

```
##            len      dose
## len  1.0000000 0.8026913
## dose 0.8026913 1.0000000
```

We can see that the length of the tooth and the dosage have a high correlation

Plotting

```r
p1 <- ggplot(ToothGrowth, aes(x =factor(dose), y = len, fill = factor(dose)))
p1 + geom_boxplot() + xlab("Dosage") + ylab("Length")
```



Again, the boxplot shows more clearly that as the dosage increases, the length of the tooth increases.

2. Use confidence intervals and hypothesis tests to compare tooth growth by supp and dose

The 95% Confidence Intervals of Tooth Grow

```
Interval <- (mean(ToothGrowth$len) + c(-1, 1) * qnorm(0.975) * sd(ToothGrowth$len)/sqrt(length(ToothGrow
Interval
```

```
## [1] 16.87783 20.74884
```

T-Test by length and Supplements

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

```
##
##  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
```

This T-Test shows that we cannot reject the Null hypothesis.

T-Test of Dosage Level

Subset Data

```
dos1 <- subset(ToothGrowth, dose == 0.5)
dos2 <- subset(ToothGrowth, dose == 1)
dos3 <- subset(ToothGrowth, dose == 2)
```

T-Test

```
t.test(len~supp, data=dos1) # Small Dosage = 0.5
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

```
t.test(len~supp, data=dos2) # Medium Dosage = 1
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

```
t.test(len~supp, data=dos3) # Big Dosage = 2
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

- For Small and Medium dosage, their p-values are small, so we can reject the null hypothesis
- For Big Dosage, we cannot reject the null hypothesis

3. State your conclusions and the assumptions needed for your conclusions

- The Tooth Length Growth is not controlled by Supplements.
- For dosage level 0.5 and 1, the Orange Juice has a higher effect on the length of tooth of Guinea Pigs than the Vitamin C.
- For dosage level 2, there are so such difference between Orange Juice and Vitamin C