Statistical Inference Course Project - Part II - Basic Inferential Data Analysis

Huseyin Emre Balli

December 21, 2018

Overview

Now in the second portion of the project, 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/or hypothesis tests to compare tooth growth by supp and dose.
4. Conclusions

Load the ToothGrowth data and perform some basic exploratory data analyses & Provide a basic summary of the data.

Load ToothGrowth data

```
data("ToothGrowth")
```

Summary of data

```
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
```

First few rows

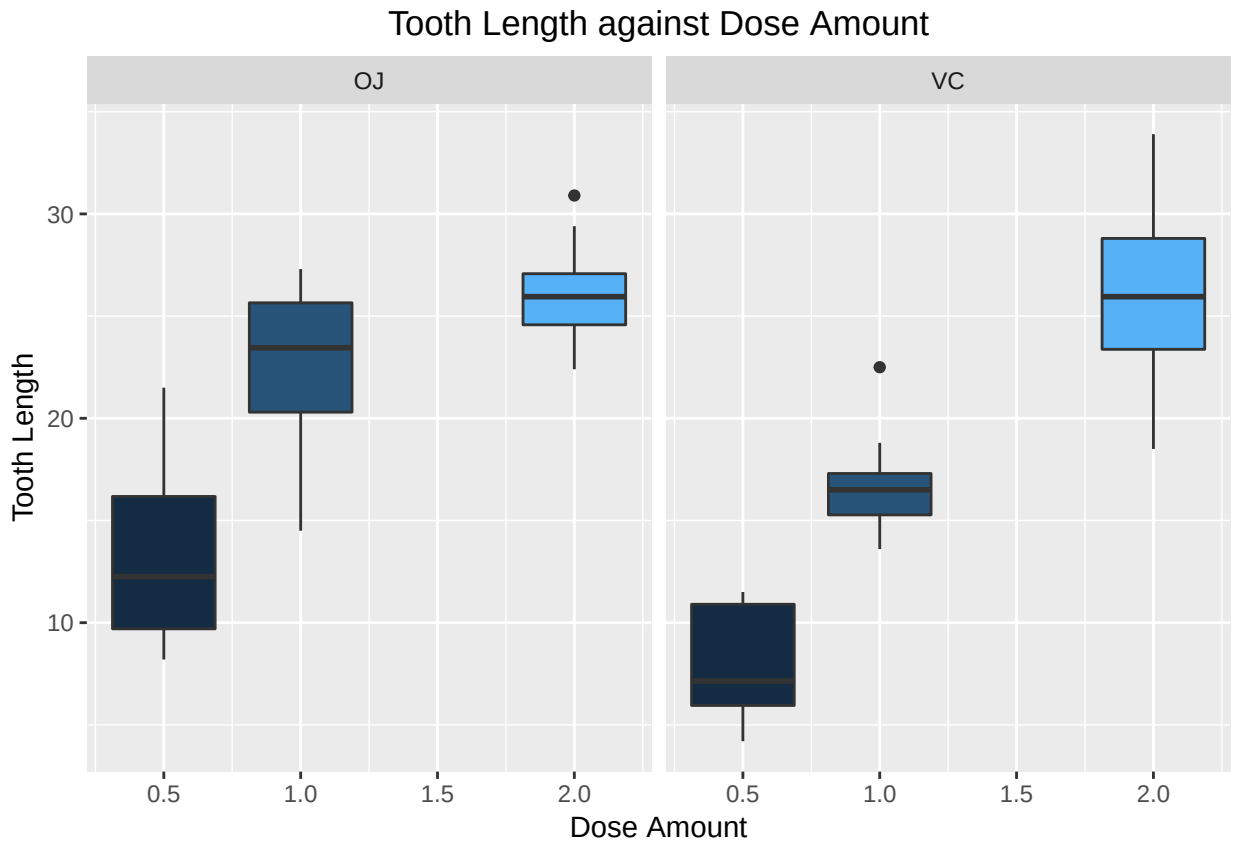
```
head(ToothGrowth)
```

```
##    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
```

Exploratory Plot

```
library(ggplot2)
ggplot(data = ToothGrowth, aes(x = dose, y = len)) +
               geom_boxplot(aes(fill = dose, group = dose)) +
```

```
                facet_grid( ~ supp) +
                        labs(x = "Dose Amount", y = "Tooth Length",
                             title = "Tooth Length against Dose Amount") +
                                theme(plot.title = element_text(hjust = 0.5),
                                      legend.position = "none")
```



Run t-tests for different variables

```
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
t1 <- t.test(data = ToothGrowth, len ~ supp)
ToothGrowth_0.5_1.0 <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5, 1.0))
t2 <- t.test(len ~ dose,data = ToothGrowth_0.5_1.0)
ToothGrowth_0.5_2.0 <- subset(ToothGrowth, ToothGrowth$dose %in% c(0.5, 2.0))
t3 <- t.test(len ~ dose,data = ToothGrowth_0.5_2.0)
ToothGrowth_1.0_2.0 <- subset(ToothGrowth, ToothGrowth$dose %in% c(1.0, 2.0))
t4 <- t.test(len ~ dose,data = ToothGrowth_1.0_2.0)
```

Conclusions

Look at the results

```
t1$estimate
```

```
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

```
t1$p.value
```

```
## [1] 0.06063451
```

```
t1$conf.int
```

```
## [1] -0.1710156  7.5710156
## attr(,"conf.level")
## [1] 0.95
```

The p-value of this test is **0.06**. Since the p-value is greater than 0.05 and the confidence interval of the test contains zero we can say that supplement types seems to have no impact on Tooth growth based on this test.

```
t2$estimate
```

```
## mean in group 0.5   mean in group 1
##            10.605            19.735
```

```
t2$p.value
```

```
## [1] 1.268301e-07
```

```
t2$conf.int
```

```
## [1] -11.983781  -6.276219
## attr(,"conf.level")
## [1] 0.95
```

```
t3$estimate
```

```
## mean in group 0.5   mean in group 2
##            10.605            26.100
```

```
t3$p.value
```

```
## [1] 4.397525e-14
```

```
t3$conf.int
```

```
## [1] -18.15617 -12.83383
## attr(,"conf.level")
## [1] 0.95
```

```
t4$estimate
```

```
## mean in group 1 mean in group 2
##          19.735          26.100
```

```
t4$p.value
```

```
## [1] 1.90643e-05
```

```
t4$conf.int
```

```
## [1] -8.996481 -3.733519
## attr(,"conf.level")
## [1] 0.95
```

As can be seen, the p-value of each test was essentially zero and the confidence interval of each test does not cross over zero (0).

Based on this result we can assume that the average tooth length increases with an inceasing dose, and therefore the null hypothesis can be rejected.

1. The sample is representative of the population

2. The distribution of the sample means follows the Central Limit Theorem

3. The 95% confidence interval is adapted as the standard

With the results of t-tests above, we can conclude that supplement delivery method has no effect on tooth length with 95% confidence interval. However increased dosages result in increased tooth length.