Basic Inferential Data Analysis of ToothGrowth data

Vasudha Singh

December 6, 2018

Overview

In this project , we are going to analyse the ToothGrowth dataset and gain some important information about data. Then we will do the comparison with the confidence intervals in order to make conclusion about Toothgrowth.

Load the ToothGrowth data and perform some basic exploratory data analyses

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

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

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

Basic summary of the data

ToothGrowth data set consists of 3 variables and 60 observations:
- 3 variables is “length”, “dosage”, “supp(OJ-orange juice, VC-vitamin c)”
- length and dosage are ‘numeric variables’
- supp are ‘factor variables’

Summary of dataset

```
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 ...
```

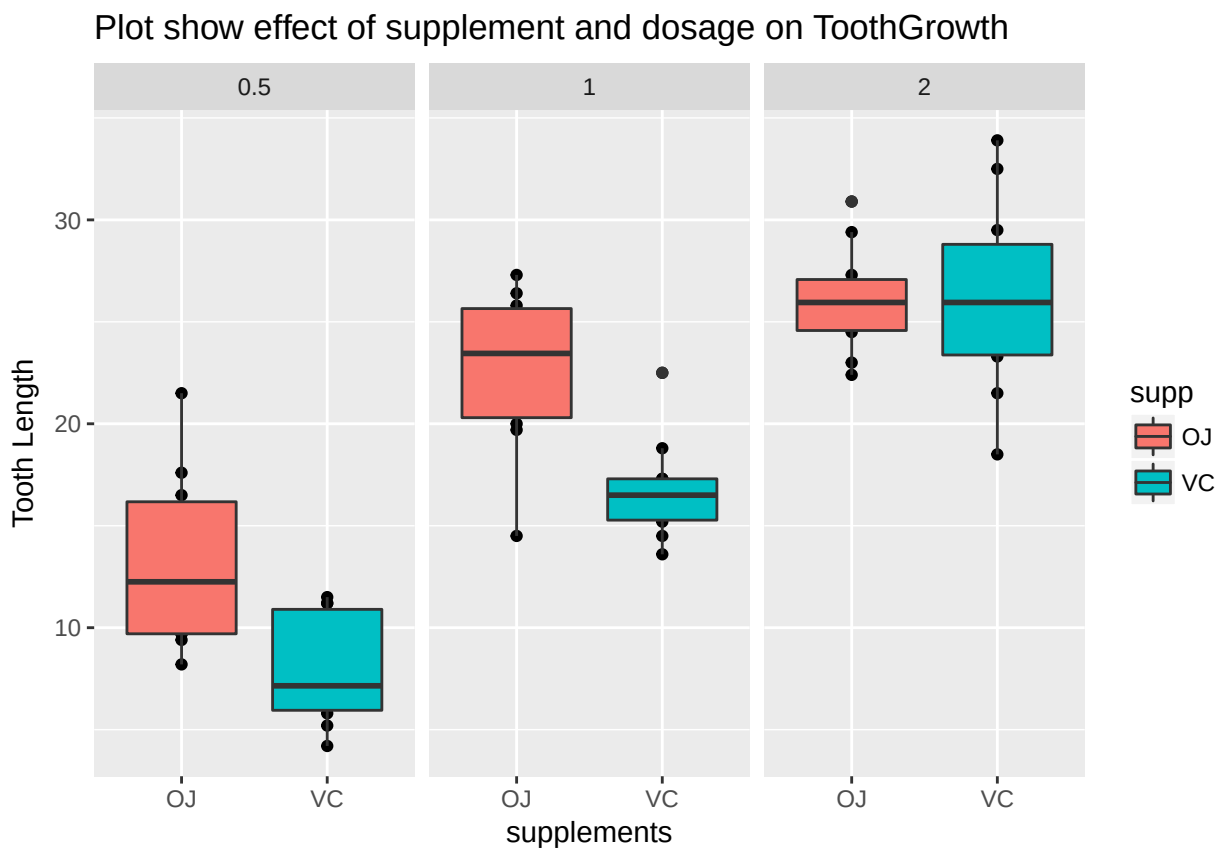
```r
head(ToothGrowth)
```

```
##    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
```

```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
```

```r
qplot(x=supp, y=len, data=ToothGrowth, facets= ~dose, xlab="supplements", ylab="Tooth Length", main="Pl
```



This boxplot show the effect of supplements and dosage on ToothGrowth length.while increasing the dose, the supplements of orange juice is more effective on toothgrowth than Vitamin C. But at 2.0 mg/day dose both the supplement are equally effective.

Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose

This is based on 2 sample T-testing and on the basis of that we assume the hypothesis.

Hypothesis

Null hypothesis: we test whether the supplement of OJ or VC deliver the same TOoth growth for different doses

For Dose=0.5

```
t_test_dose0.5<-t.test(len~supp, paired=FALSE, data=subset(ToothGrowth, dose == 0.5))
t_test_dose0.5
```

```
##
##  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
```

Dose of 0.5 mg/ml have a p-value lower than 0.05, which shows there is a difference in means. so the NUll Hypothesis can be rejected.

For Dose= 1

```
t_test_dose1<-t.test(len~supp, paired=FALSE, data=subset(ToothGrowth, dose == 1))
t_test_dose1
```

```
##
##  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
```

Dose of 1 mg/ml have a p-value lower than 0.05, which shows there is a difference in means. so the NUll Hypothesis can be rejected.

For Dose=2

```
t_test_dose2<-t.test(len~supp, paired=FALSE, data=subset(ToothGrowth, dose == 2))
t_test_dose2
```

```
##
##  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
```

Dose of 2 mg/ml have a p-value greater than 0.05, which shows there is a NO difference in means. so the NUll Hypothesis can NOT be rejected.

For dataset=ToothGrowth

```
t_test<-t.test(len~supp, paired=FALSE, data=ToothGrowth)
t_test
```

```
##
##  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
```

Dose of whole dataset have a p-value greater than 0.05, which shows there is a difference in means. so the NUll Hypothesis can be rejected.

State your conclusions and the assumptions needed for your conclusions

Assumption:

-The whole population is normally distributed

Conclusion:

The conclusion is when the dose is 0.5mg/ml or 1 mg/ml , there is a difference in means (i.e. there is a difference between the teethgrowth after taking supplement OJ and VC). while taking the dose 2.0 mg/ml, there is no difference in means.