Statistics Inference Project Part 2

Rajaram

02/08/2020

Basic Inferential Data Analysis

In this part of project, we analyze ToothGrowth data from datasets package in R. We are going to check if there is any effect of tooth growth based on dosage of Vitamin C or supplement type of the dose.

Load Data

We are going to load the data from datasets package and check the structure of data.

```
library(datasets)
toothgrowth <- ToothGrowth
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 ...
```

Data Check and Transformation

We are going to check the data for completness and perform transformation before EDA.

```
sum(!complete.cases(toothgrowth))
```

```
## [1] 0
```

There is no row with NA. Lets have a look at the data summary.

```
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
```

Data Summary looks ok. We will modify the dose to factor as we have only 3 values.

```
unique(toothgrowth$dose)
```

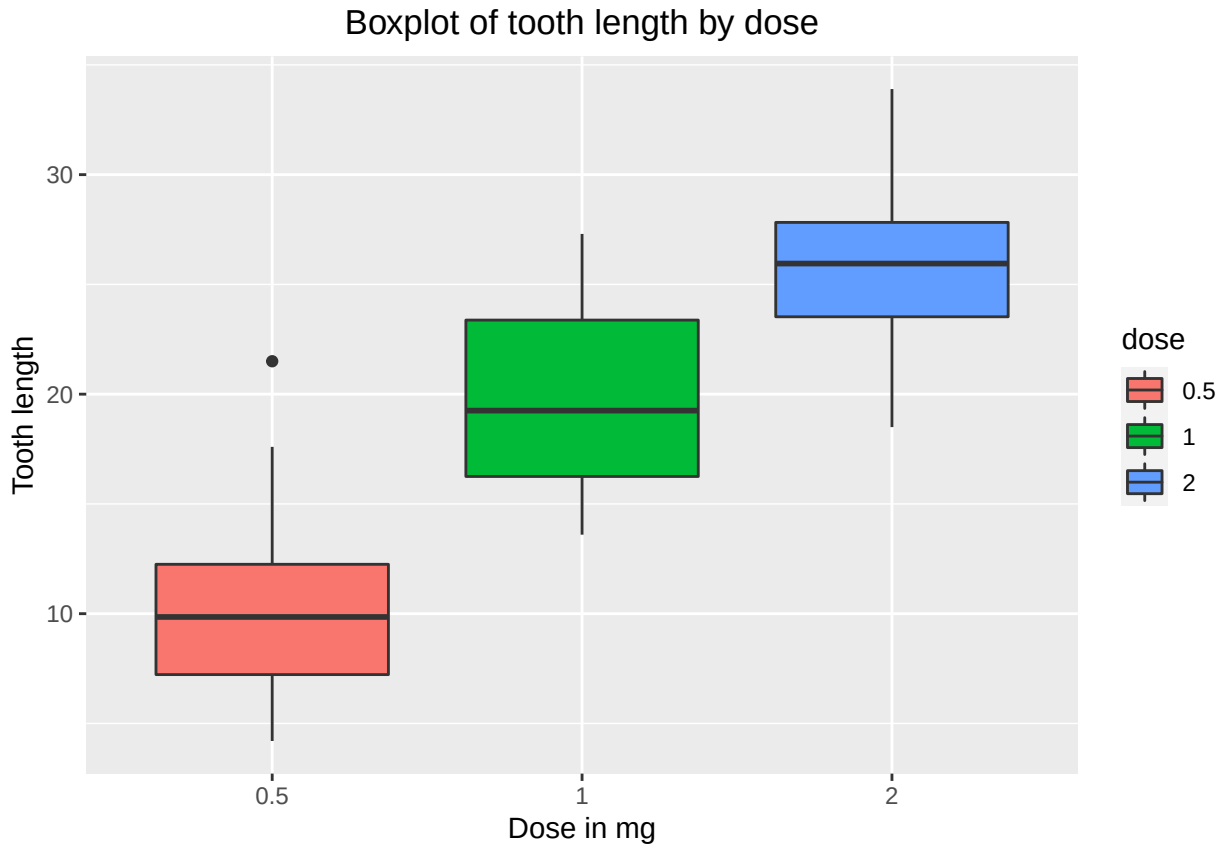
```
## [1] 0.5 1.0 2.0
```

```
toothgrowth$dose <- as.factor(toothgrowth$dose)
```

Exploratory data analysis

We will have a look at the relationship between length vs dose/supplement type in boxplot.

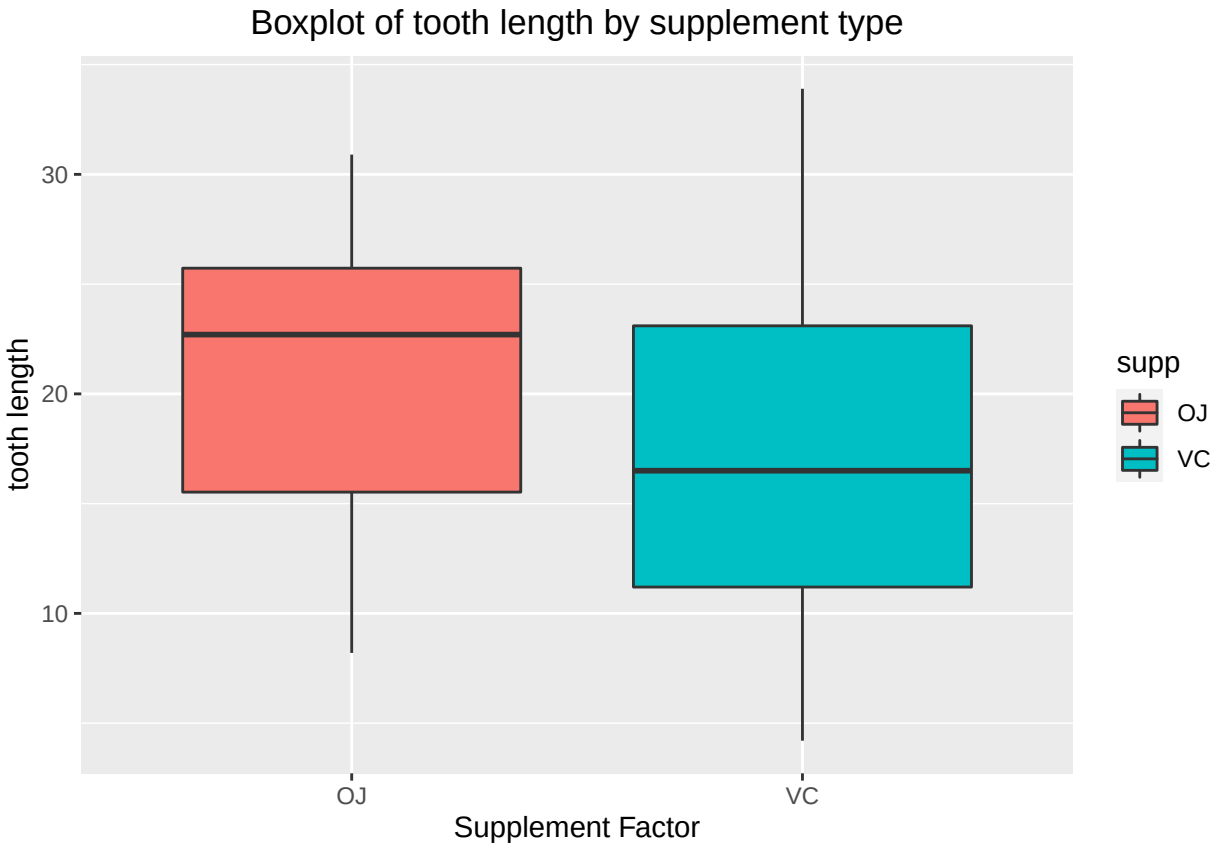
```
library(ggplot2)


ggplot(toothgrowth, aes(x=dose, y=len, fill=dose))+
  geom_boxplot()+
  labs(title ="Boxplot of tooth length by dose", x="Dose in mg",y="Tooth length")+
  theme(plot.title = element_text(hjust=0.5))
```



From the above plot, we can infer higher the dose, longer the teeth are.Also, we can notice tooth length of 1 mg dose is twice as much as of tooth length of 0.5 mg. But the same is not the case between 2 mg and 1 mg.The box position is diffrent between the doses.

Lets now have a look at supplement type.

```r
ggplot(toothgrowth, aes(x=supp,y=len,fill=supp))+
  geom_boxplot()+
  labs(title ="Boxplot of tooth length by supplement type", x="Supplement Factor",y="tooth length")+
  theme(plot.title = element_text(hjust=0.5))
```



The shape of the box look similar, however there is a difference in median line. Median for Oranje Juicie is higher, which means high values are more than low values.

Hypothesis Tests

Now, we will test if supplement type has any effect on the tooth growth.

H0: The supplement type doesnt have any effect on tooth growth

Ha: The supplement type has effect on tooth growth

```r
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
```

The above tests shows the folllowing :

1. P-value is more than significance level of 0.05.
2. Confidence Interval contains 0, which shows test is not really significant.

There are not enough evidences to reject the null hypothesis. We fail to reject the null hypothesis that the diffrent supplement type does not have any effect on tooth growth.

Lets test if dose has any effect on the tooth growth.

H0: The dosage doesnt have any effect on tooth growth

Ha: The dosage has effect on tooth growth

```
doses_0.5_1.0 <- toothgrowth[toothgrowth$dose %in% c(0.5,1.0),]
doses_1.0_2.0 <- toothgrowth[toothgrowth$dose %in% c(1.0,2.0),]
doses_0.5_2.0 <- toothgrowth[toothgrowth$dose %in% c(0.5,2.0),]
t.test(len ~ dose, data = doses_0.5_1.0)
```

```
##
##  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 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
```

```
t.test(len ~ dose, data = doses_1.0_2.0)
```

```
##
##  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 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
```

```
t.test(len ~ dose, data = doses_0.5_2.0)
```

```
##
##  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 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
```

For all the dose level pairs, p-value is less than significance level of 0.05 and the confidence interval doesn't contain zero. Hence, we can reject the null hypothesis.

Conclusions

1. Supplement type has no effect on tooth length
2. Dose levels has an effect on tooth length

Assumptions

1. The expriment was done with random assignment of guinea pigs with one of the dose levels and by one of the supplement types.
2. 60 guinea pigs are reprsentative of the population