Statistical Inference Project - Part 2

Prashant Chand

2 Jan 2020

Part 2: Basic Inferential Data Analysis

Synopsis:

In this second part of the project we're going to analyze the ToothGrowth dataset in the R datasets package.This data is about the growth in length of teeths of Guinea pigs when they are provided with the certain amount of doses of vitamin C. Vitamin C is provided in 2 differrent forms i.e OJ(Orange Juice) and VC (Ascorbi Acid, a form of Vitamin C).The no of doses provided to the pigs are 0.5,1 and 2.

Load the ToothGrowth data and required packages

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

First few rows of the data to get an insight on how the data looks like.

```
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
```

Summary of the data

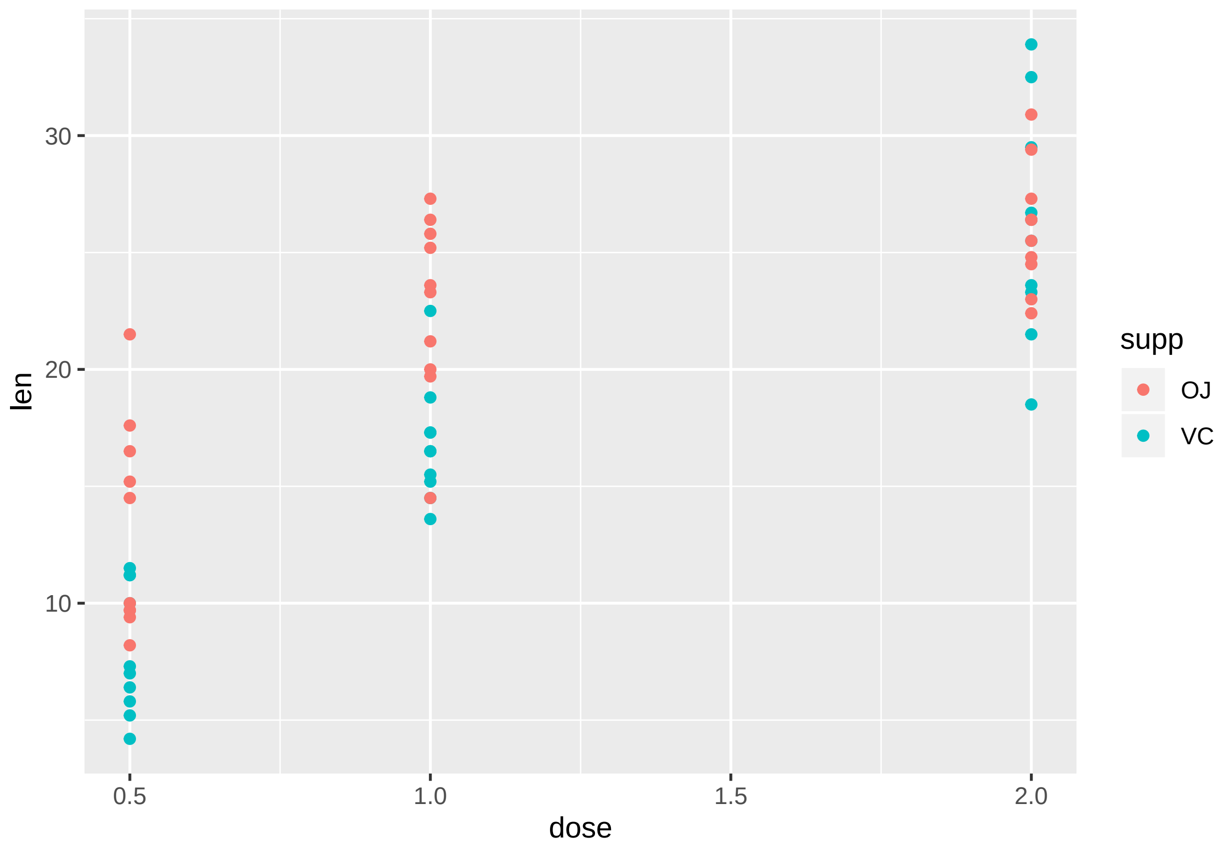
```
str(ToothGrowth)
```

```
## '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 ...
```

As we can see from the summary of the data there are 60 obs. of 3 variables i.e len,supp and dose.The len variable contains the length of the tooths of the pigs.The supp variable is a factor variable which has 2 levels i.e OJ and VC.The dose variable has 3 values i.e 0.5,1 and 2.

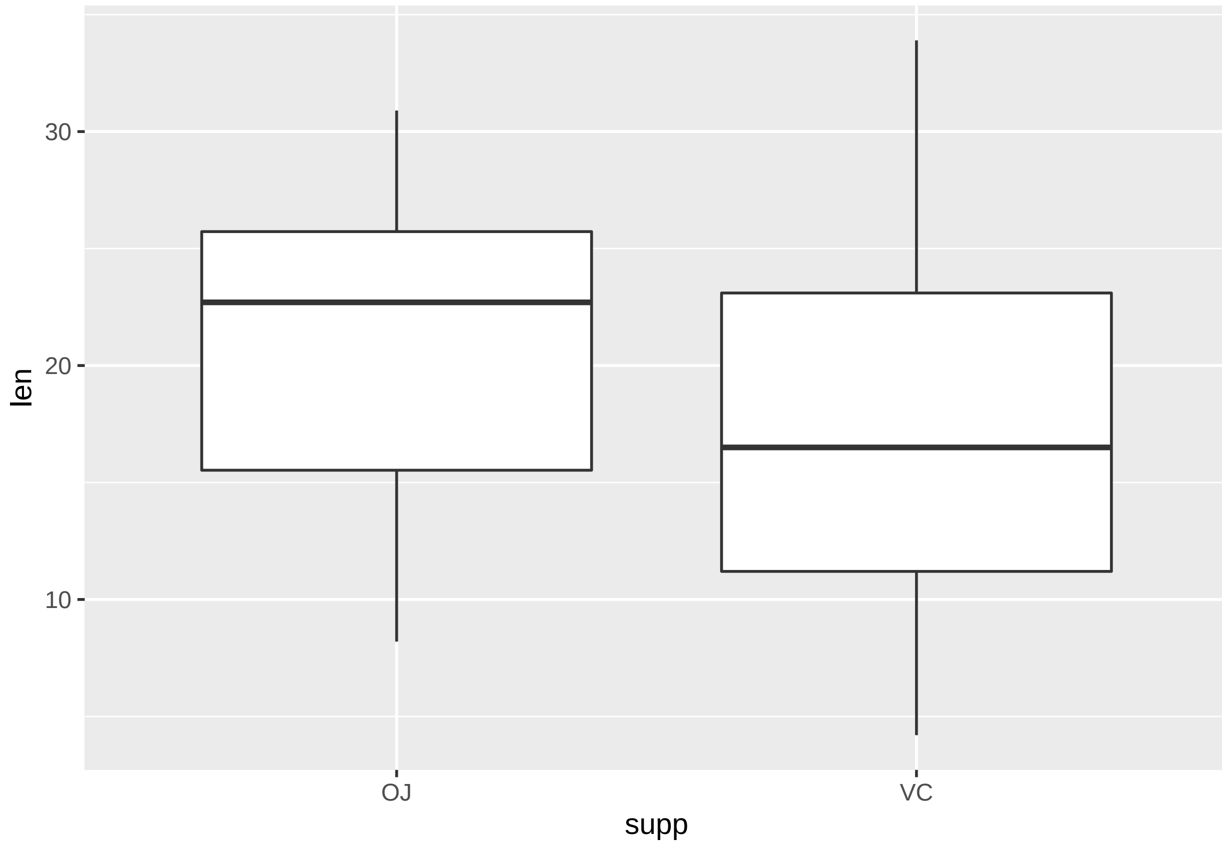
Basic exploratory data analyses

```r
g<-ggplot(data=ToothGrowth,aes(x=dose,y=len,col=supp))
g+geom_point()
```



It can be seen clearly that the growth in the length of tooths of the pigs who are provided with 2 doses of vitamin C has significantly more growth than the ones who are provide with less dose of Vitamin C.It can also be observed that the ones who are given OJ has more growth than those who are given VC.

```r
g<-ggplot(data = ToothGrowth,aes(x=supp,y=len))
g+geom_boxplot()
```



It can be seen that the median of the length of the pigs who are given OJ(orange Juice) is around 23 and the ones who are given VC the median of these pigs is around 16.So we can conclude here that the pigs who are given the dose s of OJ the growth in their length is significantly larger which we will further prove in our hypothesis testing.

Using hypothesis testing to compare tooth growth by supp and dose.

I)First we will be doing the hypothesis testing of the difference in means of the length of the pigs with respect to the supplement types given to them.
Ho:True Diffrence in means is 0
Ha:True Differnce in means is not equal to 0

```
OJ<-ToothGrowth[ToothGrowth$supp=="OJ",]$len
VC<-ToothGrowth[ToothGrowth$supp=="VC",]$len
t.test(OJ,VC,paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  OJ and VC
## 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 of x mean of y
##  20.66333  16.96333
```

From the t-test results it can be observed that the p value is 0.06063 which is slightly more than the significance level of 0.05.Hence we cannot reject the null hypothesis.So we can say that the supplement types has no impact on the tooth growth based on this test. We can also observe from the result that the mean of OJ is more than the mean of VC.But we are not rejecting the null hypothesis here so we cannot confirm that the growth in the tooth of pigs who are given OJ is more than the ones who are given VC.

II)Second hypothesis testing will be based on the no of doses given to the pigs and see the impact of the doses on them.
a)The hypothesis testing between data of the pigs who are given 0.5 and 1 doses.

```
point5_dose<-ToothGrowth[ToothGrowth$dose==0.5,]$len
one_dose<-ToothGrowth[ToothGrowth$dose==1,]$len
t.test(point5_dose,one_dose,paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  point5_dose and one_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 of x mean of y
##    10.605    19.735
```

The p-value is very small and the confidence interval does not contain 0,so we can reject the null hypothesis and can say that as the no of dose increase the length of the tooths also increase.

b)The hypothesis testing between data of the pigs who are given 1 and 2 doses.

```
two_dose<-ToothGrowth[ToothGrowth$dose==2,]$len
t.test(one_dose,two_dose,paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  one_dose and two_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 of x mean of y
##    19.735    26.100
```

Again the p-value is very small and the confidence interval does not contain 0,so we can reject the null hypothesis.Hence we can say that the increase in no of doses have significant impcat on the growth of the tooth.

c)The hypothesis testing between data of the pigs who are given 0.5 and 2 doses.

```r
t.test(point5_dose,two_dose,paired = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  point5_dose and two_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 of x mean of y
##    10.605    26.100
```

The p-value is very small and we can reject the null hypothesis.

Conclusion

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.

Given the following assumptions: 1)The sample is representative of the population 2)The distribution of the sample means follows the Central Limit Theorem