Part 2:Basic Inferential Data Analysis Instructions

Nour Qweder

8/13/2020

Now in the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package. ## 1.1.) Data analysis Load the ToothGrowth data and perform some basic exploratory data analyses Provide a basic summary of the data.

```
set.seed(12345)
data("ToothGrowth")
attach(ToothGrowth)
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
```

```
sapply(ToothGrowth, class)
```

```
##       len      supp      dose
## "numeric"  "factor" "numeric"
```

```
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
```

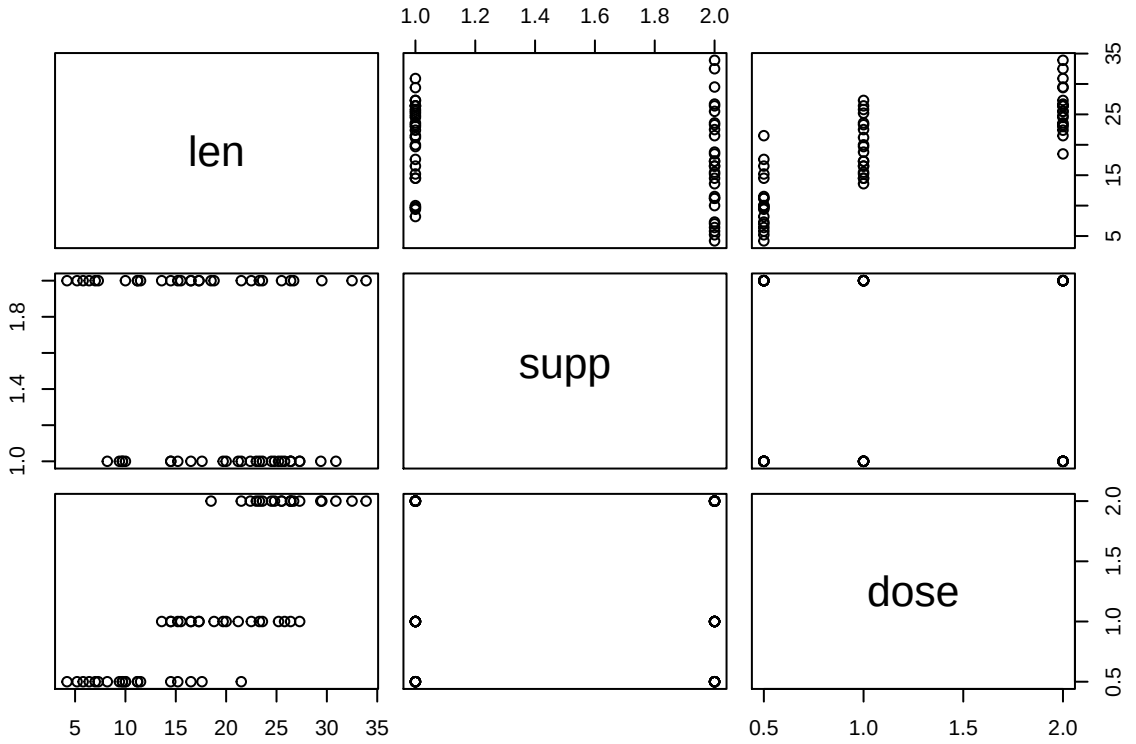
```
n = dim(ToothGrowth)# 60 3
sapply(ToothGrowth,function(x) sum(is.na(x))) #see if there are NAN values per column
```

```
##  len supp dose
##    0    0    0
```

```
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 ...
```

```
plot(ToothGrowth)
```

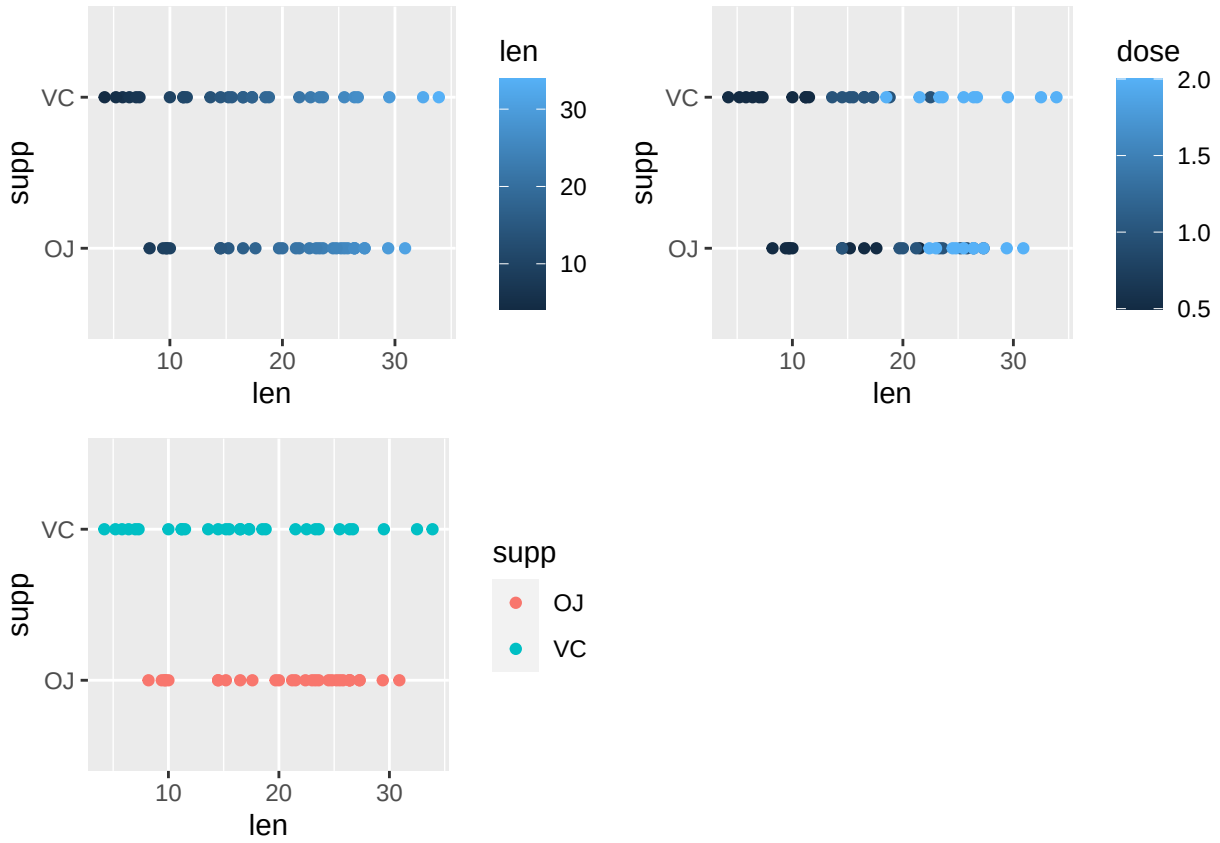


From previous analysis, it can be noticed that *ToothGrowth* dataset with 60 observations and 3 features *len as a number value, supp as a factor and dose as a number value.*

1.2.) confidence intervals

Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose. (Only use the techniques from class, even if there's other approaches worth considering)

From the previous test, it can be noticed that $p\_value = 0.6 > 0.5$; Based on t-test, it can be obtained that supplement seems to have no impact on the growth of tooth



Hypothesis

Length can be affected by dose

Application methods have no impact on tooth growth. H_0 : Both group have the same mean. H_A : Means are different.

T-test

```
ToothGrowth_sub <- subset(ToothGrowth, ToothGrowth$dose %in% c(1.0,0.5))
t.test(len~dose,data=ToothGrowth_sub)
```

```
##
##  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
```

1.3.) conclusions

From the previous test, it can be noticed that $p\_value = 1.268e\text{-}07 < 0.05$; Based on t-test, it can be obtained that does seems to have an impact on the growth of tooth, so H_0 can not be rejected.