Stat_inf_pr2

Harsha

Sunday, December 27, 2015

Project 2

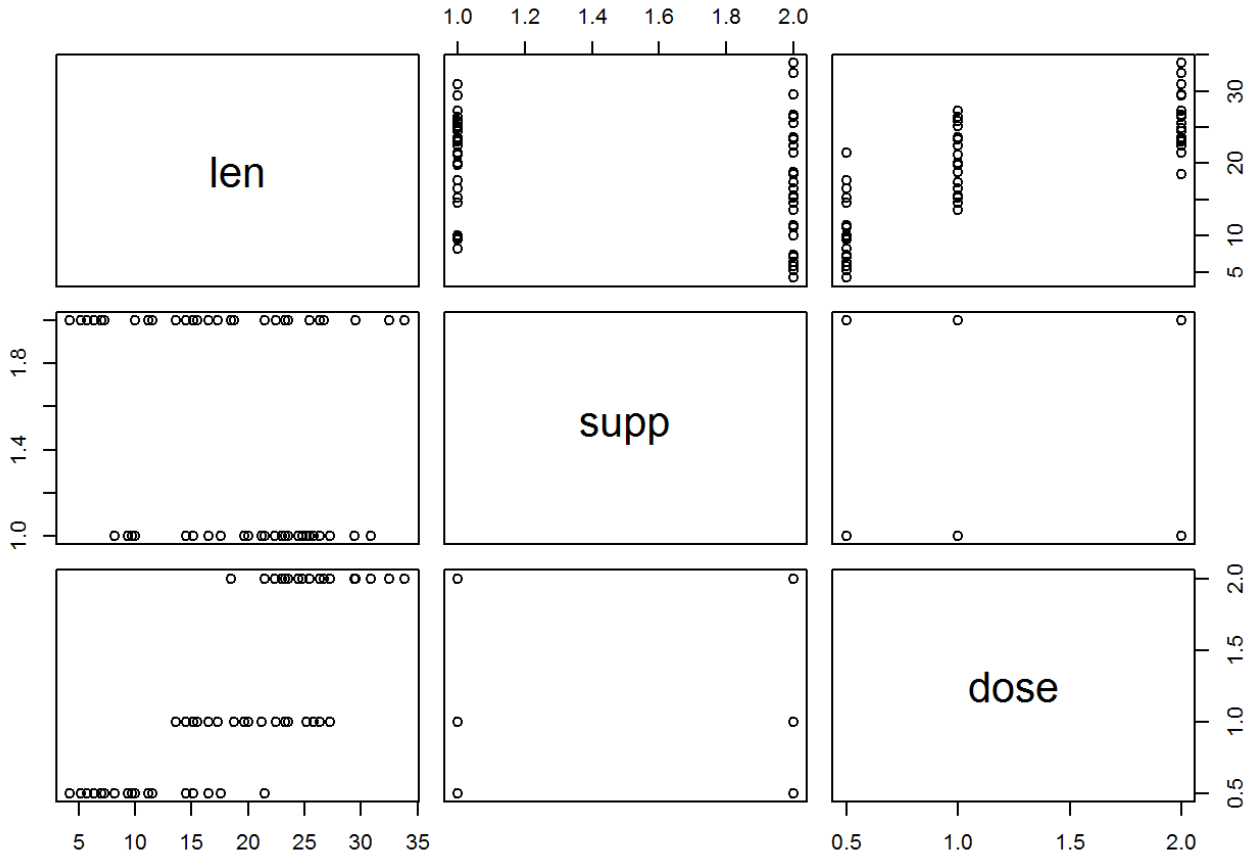
Problem statement

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. (Only use the techniques from class, even if there's other approaches worth considering) 4) State your conclusions and the assumptions needed for your conclusions.

Q1 and 2

```
library(lattice)

data(ToothGrowth)
plot(ToothGrowth)
```



From the plot it's evident that 1.5 dosage value does not exist. This can be removed from the data by converting as factors.

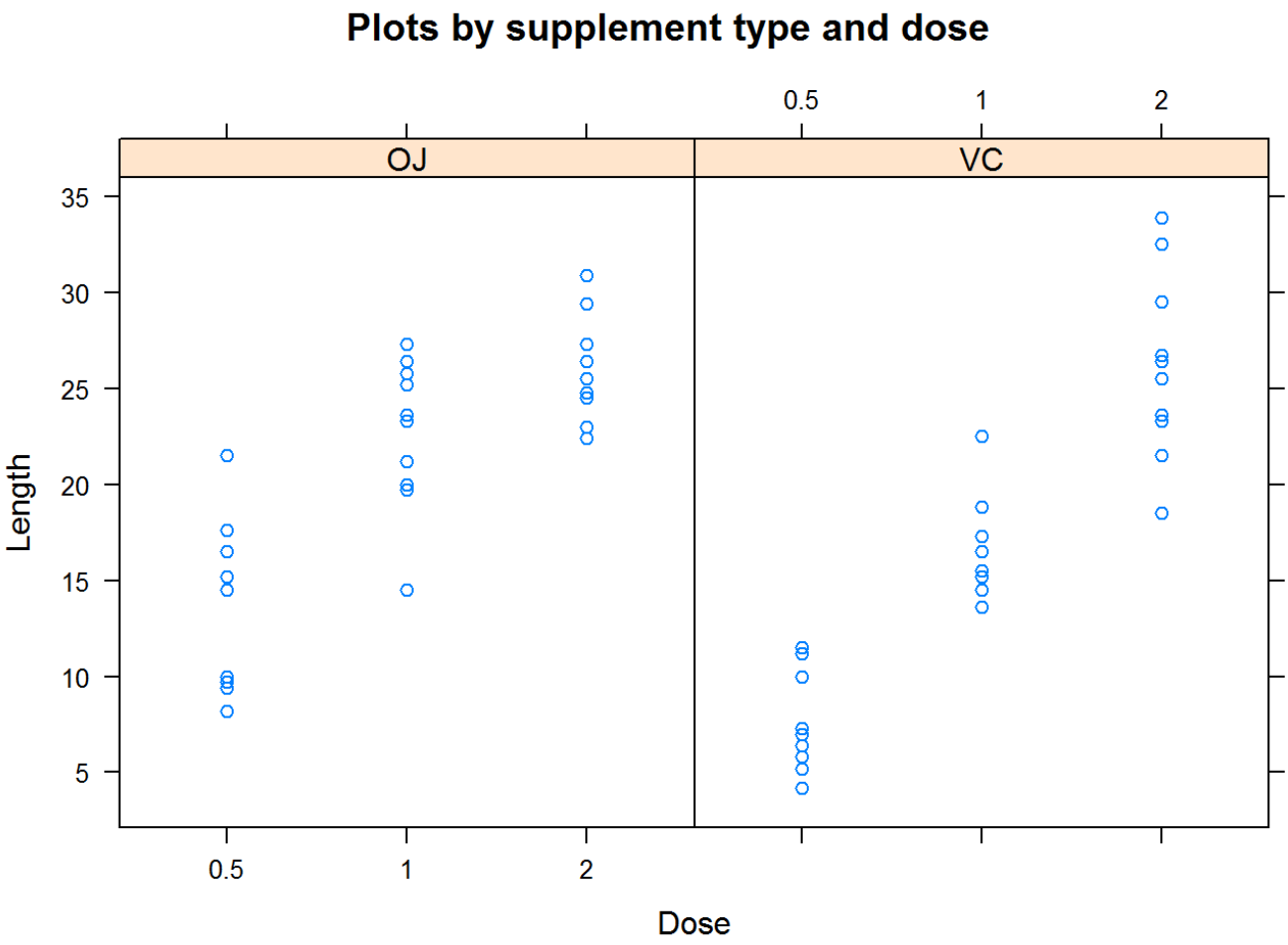
```
ToothGrowth$dose<-as.factor(ToothGrowth$dose)
summary(ToothGrowth$dose)
```

```
## 0.5   1   2
##  20  20  20
```

```
summary(ToothGrowth)
```

```
##       len        supp     dose
##  Min.   : 4.20   OJ:30   0.5:20
##  1st Qu.:13.07   VC:30   1  :20
##  Median :19.25           2  :20
##  Mean   :18.81
##  3rd Qu.:25.27
##  Max.   :33.90
```

```
xyplot(len~dose|supp, ToothGrowth,
       main="Plots by supplement type and dose",
       ylab="Length", xlab="Dose")
```



It would help if we get the mean based on the supplement

The box plot that shows the distribution of the quartiles would help in analysing the data

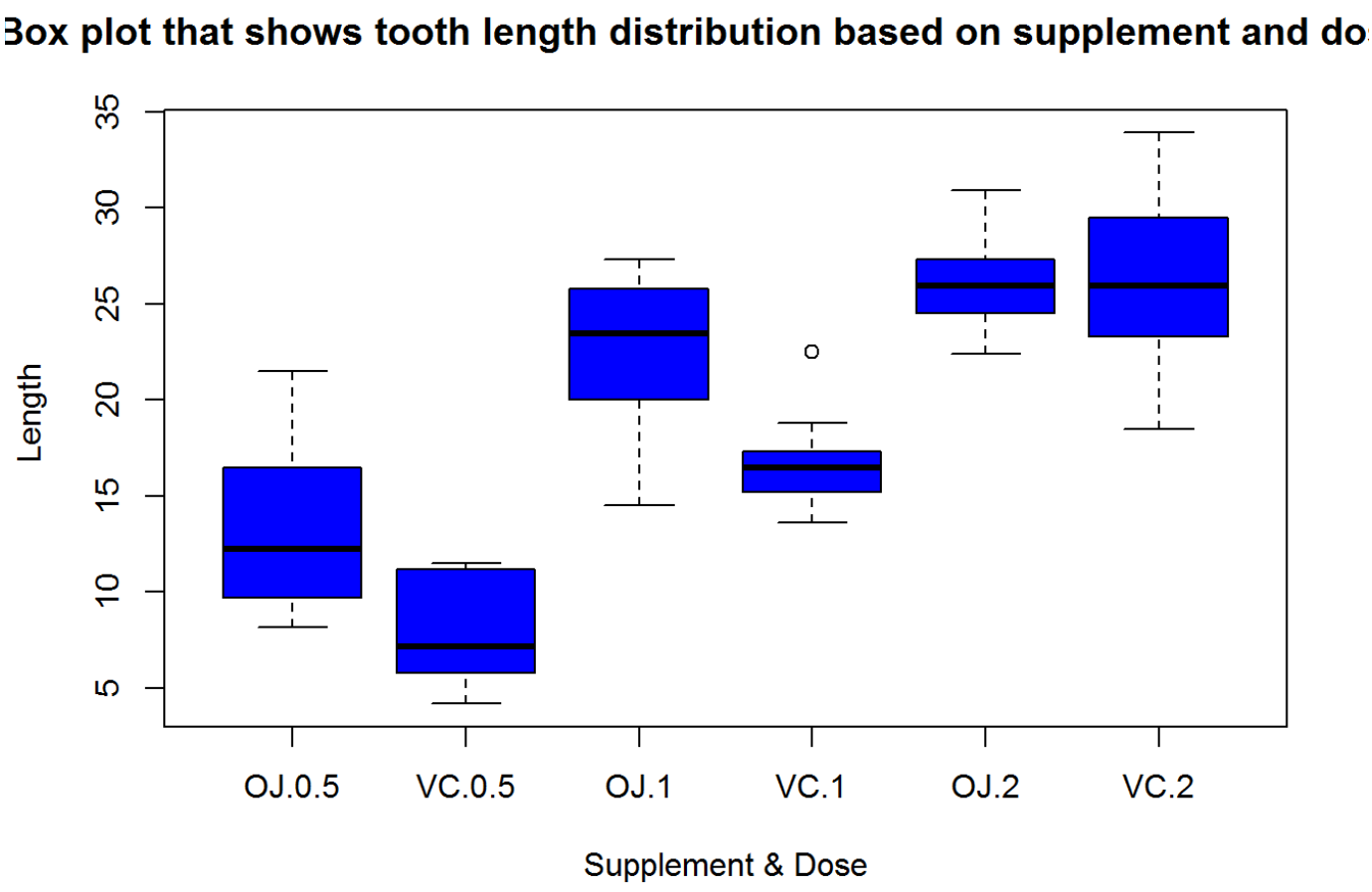
```
split_data_supp = split(ToothGrowth$len, ToothGrowth$supp)
sapply(split_data_supp, mean)
```

```
##       OJ       VC
## 20.66333 16.96333
```

```
sapply(split_data_supp, var)
```

```
##       OJ       VC
## 43.63344 68.32723
```

```
boxplot(len ~ supp * dose, ToothGrowth, col="blue", ylab=" Length", xlab="Supplement &
Dose",main="Box plot that shows tooth length distribution based on supplement and dosa
ge")
```



Q3

The distribution shows the effect of dosage but the same cannot be said about the supplement. However t.test is to be used to confirm. The null hypothesis is that suplement or dosage doesn't make a difference

First let's see whether supplement has made any difference. Since variances are far, var.equal = FALSE

```
t.test(ToothGrowth$len[ToothGrowth$supp=="OJ"], ToothGrowth$len[ToothGrowth$supp=="V
C"], paired = FALSE, var.equal = FALSE)
```

```
##
##  Welch Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$supp == "OJ"] and ToothGrowth$len[ToothGrowth$su
pp == "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
```

P value is 0.06. So we cannot conclude about the supplement.

Now let's check the dosage quantity

```
t.test(ToothGrowth$len[ToothGrowth$dose==2], ToothGrowth$len[ToothGrowth$dose==1], pai
red = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 2] and ToothGrowth$len[ToothGrowth$dose
== 1]
## t = 4.9005, df = 38, p-value = 1.811e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.735613 8.994387
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

```
t.test(ToothGrowth$len[ToothGrowth$dose==2], ToothGrowth$len[ToothGrowth$dose==0.5], p
aired = FALSE, var.equal = TRUE)
```

```
##
##  Two Sample t-test
##
## data:  ToothGrowth$len[ToothGrowth$dose == 2] and ToothGrowth$len[ToothGrowth$dose
== 0.5]
## t = 11.799, df = 38, p-value = 2.838e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  12.83648 18.15352
## sample estimates:
## mean of x mean of y
##    26.100    10.605
```

The results show that there's an effect of the dosage quantity on tooth length.

Conclusion

The dosage quanity has an effect on the tooth length (increases with increase) but the same cannot be said about the supplement.

Q4

Assumptions

Experiment is assumed to be randomly assigning the dosage and supplement to the subject.