Statistical Inference Course Project, Part 2

Josias Alvarenga

April 9, 2017

Overview

Part 2: Basic Inferential Data Analysis Instructions

Now in the second portion of the project, we're going to 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.

Question 1

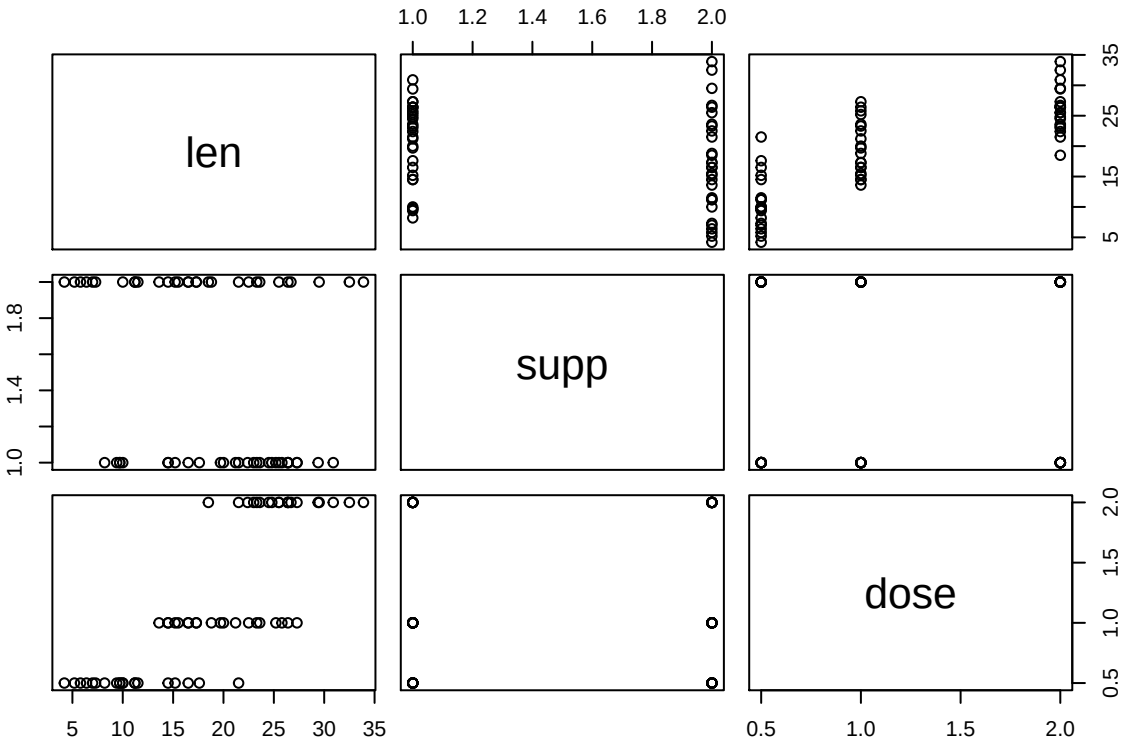
```
data(ToothGrowth)
head(ToothGrowth)
```

```
##    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
```

```
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)
```



Based on the data set analysis we can verify that the variable dose is numeric and assumes the values 0.5, 1.0 and 2.0 mg. So, I will transform it to a factor variable.

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

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

Question 2

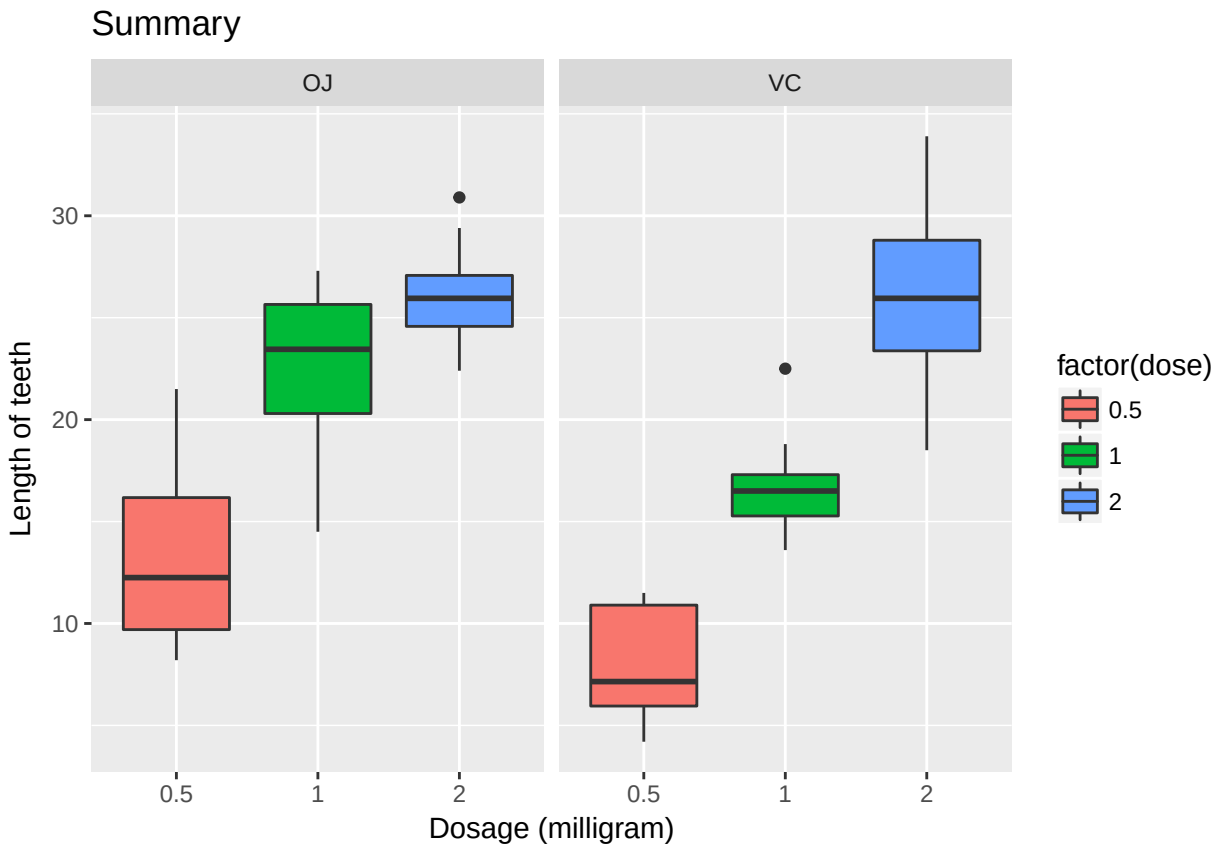
```
summary(ToothGrowth)
```

```
##       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
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
plot <- ggplot(ToothGrowth, aes(x=factor(dose), y=len, fill=factor(dose)))
plot + geom_boxplot(notch=FALSE) + facet_grid(.~supp) +
        scale_x_discrete("Dosage (milligram)") +
        scale_y_continuous("Length of teeth") +
        ggtitle("Summary")
```



Question 3

```
supp.equal <- t.test(len~supp, paired=F, var.equal=T, data=ToothGrowth)
supp.unequal <- t.test(len~supp, paired=F, var.equal=F, data=ToothGrowth)
supp.result <- data.frame("P-Value"=c(supp.equal$p.value, supp.unequal$p.value),
        "Conf Low"=c(supp.equal$conf[1],supp.unequal$conf[1]),
        "Conf High"=c(supp.equal$conf[2],supp.unequal$conf[2]),
        row.names=c("Equal Var","Unequal Var"))
supp.result
```

```
##                P.Value   Conf.Low Conf.High
## Equal Var   0.06039337 -0.1670064  7.567006
## Unequal Var 0.06063451 -0.1710156  7.571016
```

Question 4

As doses are larger, the impact on tooth growth are also larger.