

Stat. Inf. Part 2: ToothGrowth

Jean Marie Cimula

5 March 2016

Overview

Now in the second portion of the class, 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.

Load the ToothGrowth data and perform exploratory data analyses

Below type, nature and summary of data in our dataset.

Basic summary of the data

```
library(datasets)
data(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 ...
```

```
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(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
```

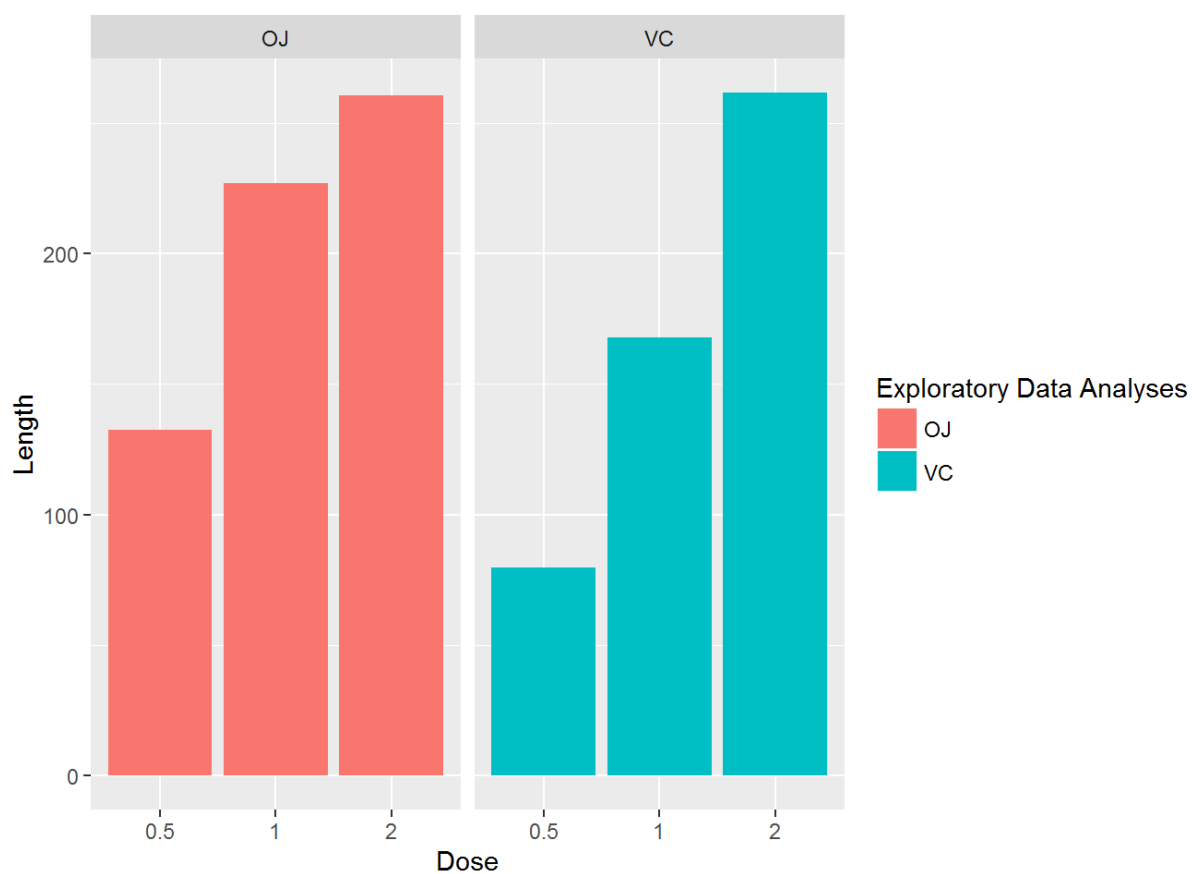
Numbers of rows : 60 and columns : 3

Loading some libraries

```
library(ggplot2)
library(GGally)
```

Below we perform exploratory data analyses

```
library(ggplot2)
ggplot(data=ToothGrowth, aes(x=as.factor(dose), y=len, fill=supp)) + geom_bar(
  stat="identity",) +
  facet_grid(. ~ supp) + xlab("Dose") + ylab("Length") +
  guides(fill=guide_legend(title="Exploratory Data Analyses"))
```



Provide a basic summary of the data

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

```
table(ToothGrowth$supp, ToothGrowth$dose)
```

```
##
##           0.5  1  2
##  OJ    10 10 10
##  VC    10 10 10
```

Use confidence intervals and/or hypothesis tests to compare tooth growth by supp and dose

Using the package GGally to summary all hypothesis tests and confidence intervals for the variables supp and dose in a single view

```
an <- aov(len ~ supp * dose, data=ToothGrowth)
summary(an)
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## supp       1  205.4    205.4   15.572 0.000231 ***
## dose       2 2426.4   1213.2   92.000 < 2e-16 ***
## supp:dose   2  108.3     54.2    4.107 0.021860 *
## Residuals  54  712.1     13.2
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
confint(an)
```

```
##           2.5 %    97.5 %
## (Intercept) 10.9276907 15.532309
## suppVC      -8.5059571 -1.994043
## dose1        6.2140429 12.725957
## dose2        9.5740429 16.085957
## suppVC:dose1 -5.2846186  3.924619
## suppVC:dose2  0.7253814  9.934619
```

```
print(model.tables(an,"means"),digits=3)
```

```
## Tables of means
## Grand mean
##
## 18.81333
##
##  supp
## supp
##      OJ      VC
## 20.66 16.96
##
##  dose
## dose
##   0.5      1      2
## 10.60 19.73 26.10
##
##  supp:dose
##      dose
## supp 0.5      1      2
##   OJ 13.23 22.70 26.06
##   VC  7.98 16.77 26.14
```

State your conclusions and the assumptions needed for your conclusions

Based on the result below our assumptions are that the distribution of the mean is normal. There is no doubt that OC and VJ have obvious different impact on ToothGrowth.