Inferential Exploratory Data Analysis

Stephen Sani

April 4, 2016

1. Loading the data

The ToothGrowt data was loaded

```
data("ToothGrowth")
```

2. Summarizing and analysing the data

The first six rows of the data was viewed

```
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
```

The summary of the ToothGrowth data was obtained

```
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
```

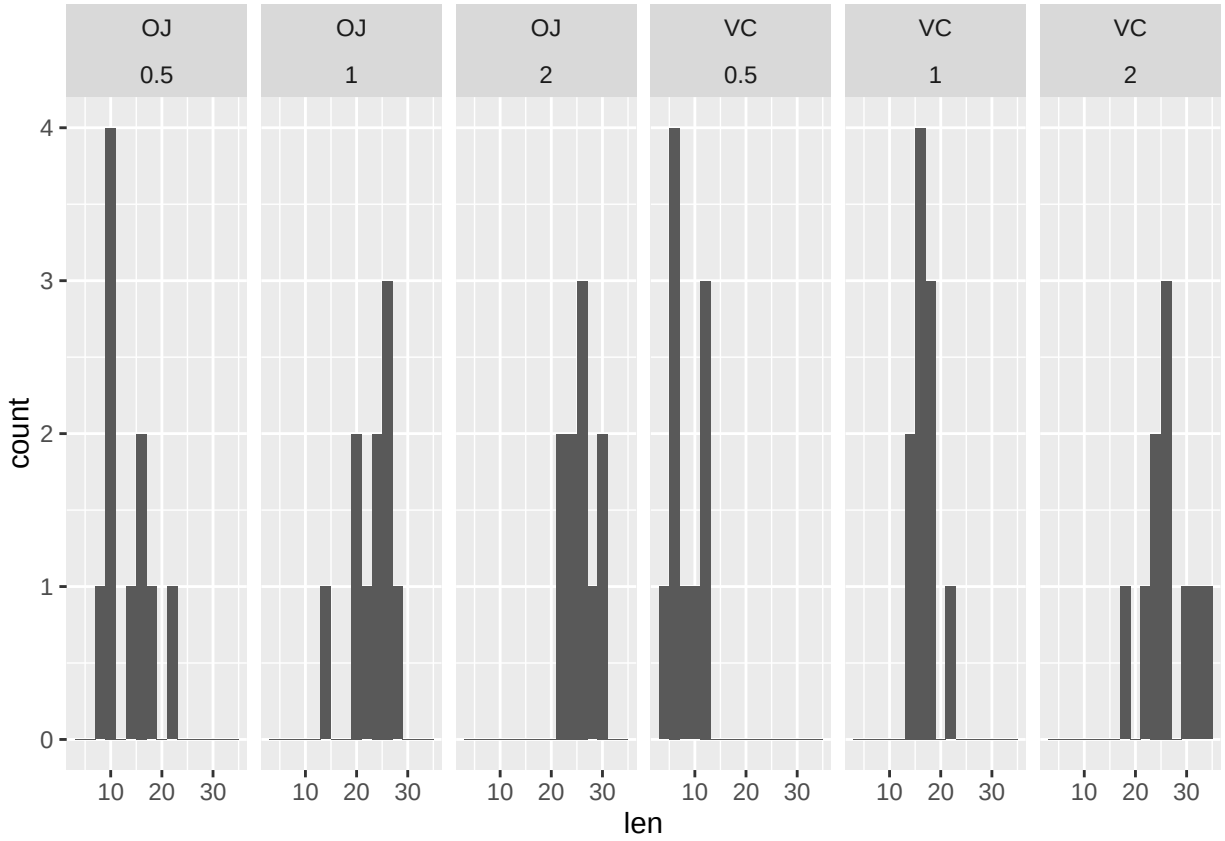
The structure of the data was obtained

```
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 ...
```

The distribution of the data was ploted by supp and dose variables

```
library(ggplot2)
qplot(len, data=ToothGrowth, facets = .~supp+dose, binwidth=2)
```



3. Confidence interval

(a.) Confidence interval for the two "supp" levels

The dataframe was subset by the two "supp" levels "OJ" and "VC"

```
supp1<-ToothGrowth[ToothGrowth$supp=="OJ", "len"]
supp2<-ToothGrowth[ToothGrowth$supp=="VC", "len"]
```

Confidence interval for the test statistic was obtained for H(null)=0

```
t.test(supp1, supp2, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] -0.1710156  7.5710156
## attr(,"conf.level")
## [1] 0.95
```

Since the interval includes 0, we fail to reject the Null hypothesis

(b.) Confidence interval for combinations of dose levels

The data frame was subset by the dose levels

```
x1<-ToothGrowth[ToothGrowth$dose=="0.5", "len"]
x2<-ToothGrowth[ToothGrowth$dose=="1", "len"]
x3<-ToothGrowth[ToothGrowth$dose=="2", "len"]
```

Confidence interval for the test statistic between "0.5" and "1" was obtained for H(null)=0

```
t.test(x1, x2, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] -11.983781  -6.276219
## attr(,"conf.level")
## [1] 0.95
```

Since the interval does not include 0, We reject the null hypothesis

Confidence interval for the test statistic between "0.5" and "2" was obtained for H(null)=0

```
t.test(x1, x3, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] -18.15617 -12.83383
## attr(,"conf.level")
## [1] 0.95
```

Since the interval does not include 0, We reject the null hypothesis

Confidence interval for the test statistic between "1" and "2" was obtained for H(null)=0

```
t.test(x2, x3, paired = FALSE, var.equal = FALSE)$conf
```

```
## [1] -8.996481 -3.733519
## attr(,"conf.level")
## [1] 0.95
```

Since the interval does not include 0, We reject the null hypothesis