Guinea Pig Tooth Growth Experiment

Juan Bosco Mendoza Vega

18 de septiembre de 2016

R Markdown

EXploratory statistics

```
dim(ToothGrowth)
```

```
## [1] 60  3
```

```
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
```

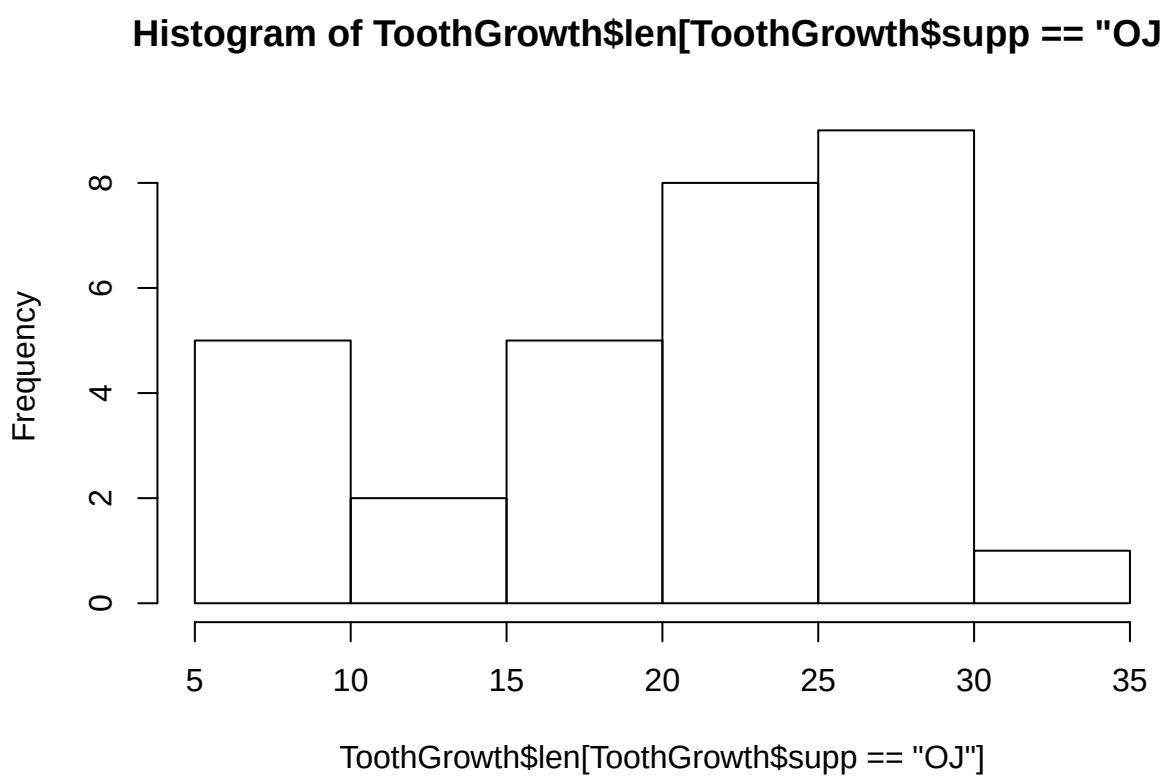
Let's now see if the *len* column has a normal distribution with the Shapiro test. The null hypothesis for this test is that the given data **is normally distributed**.

```
shapiro.test(ToothGrowth$len)
```

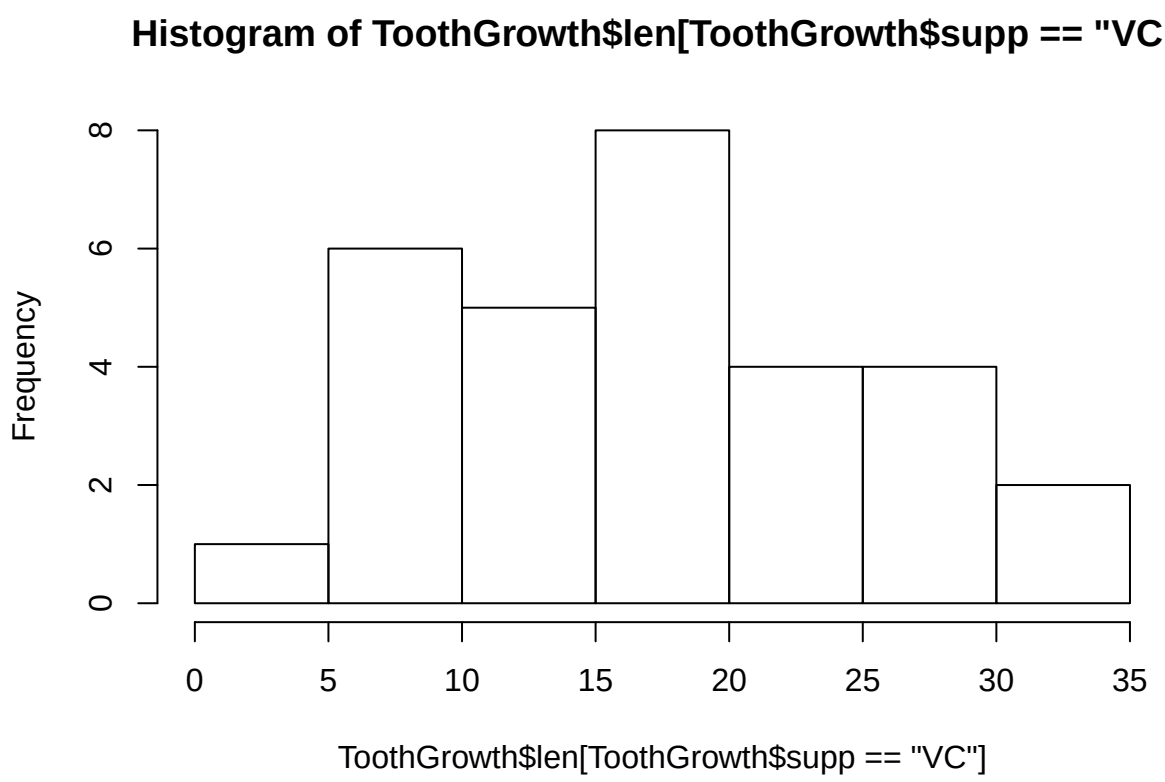
```
##
##  Shapiro-Wilk normality test
##
## data:  ToothGrowth$len
## W = 0.96743, p-value = 0.1091
```

Since we didn't get p-value less or equal to 0.05 that allows is to reject the null hypothesis, we can assume *len* is normally distributed.

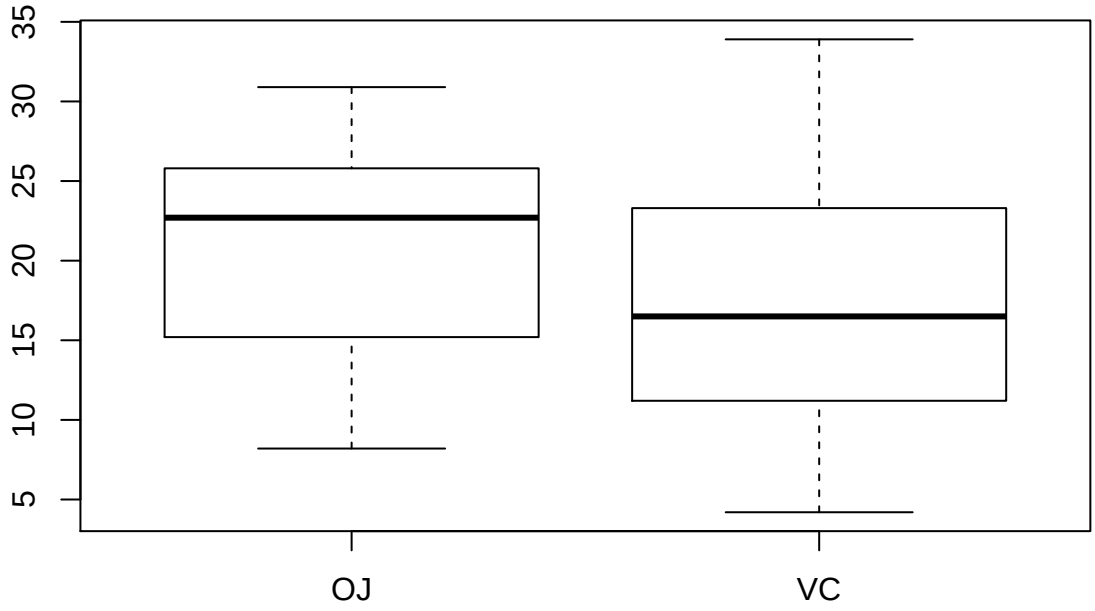
```
hist(ToothGrowth$len[ToothGrowth$supp == "OJ"])
```




```
hist(ToothGrowth$len[ToothGrowth$supp == "VC"])
```



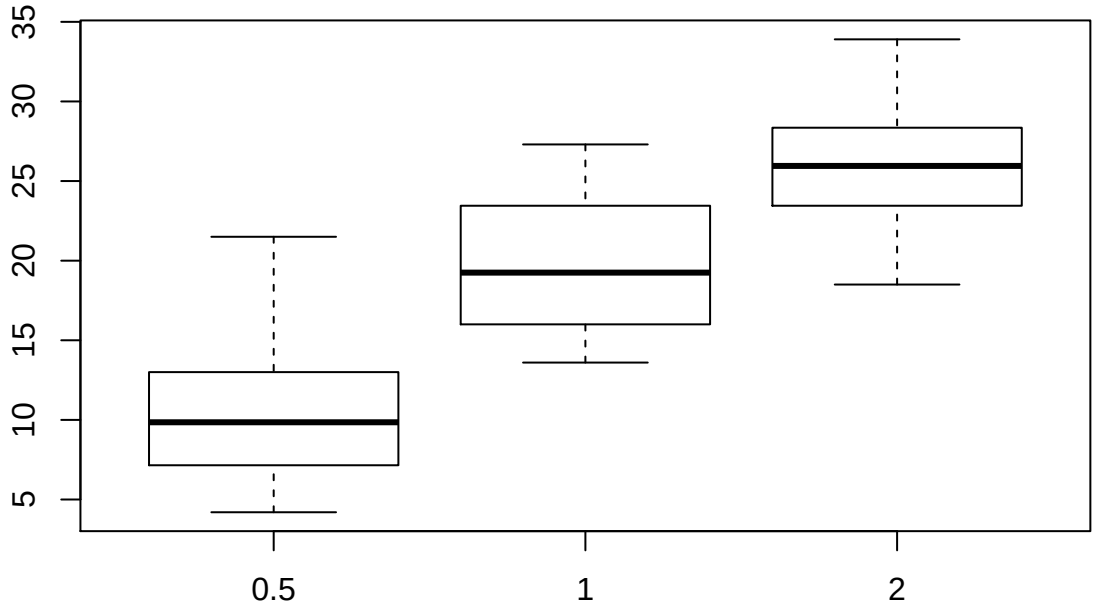

```
tooth_split <- split(ToothGrowth, ToothGrowth$supp)

boxplot(len ~ supp, ToothGrowth)
```



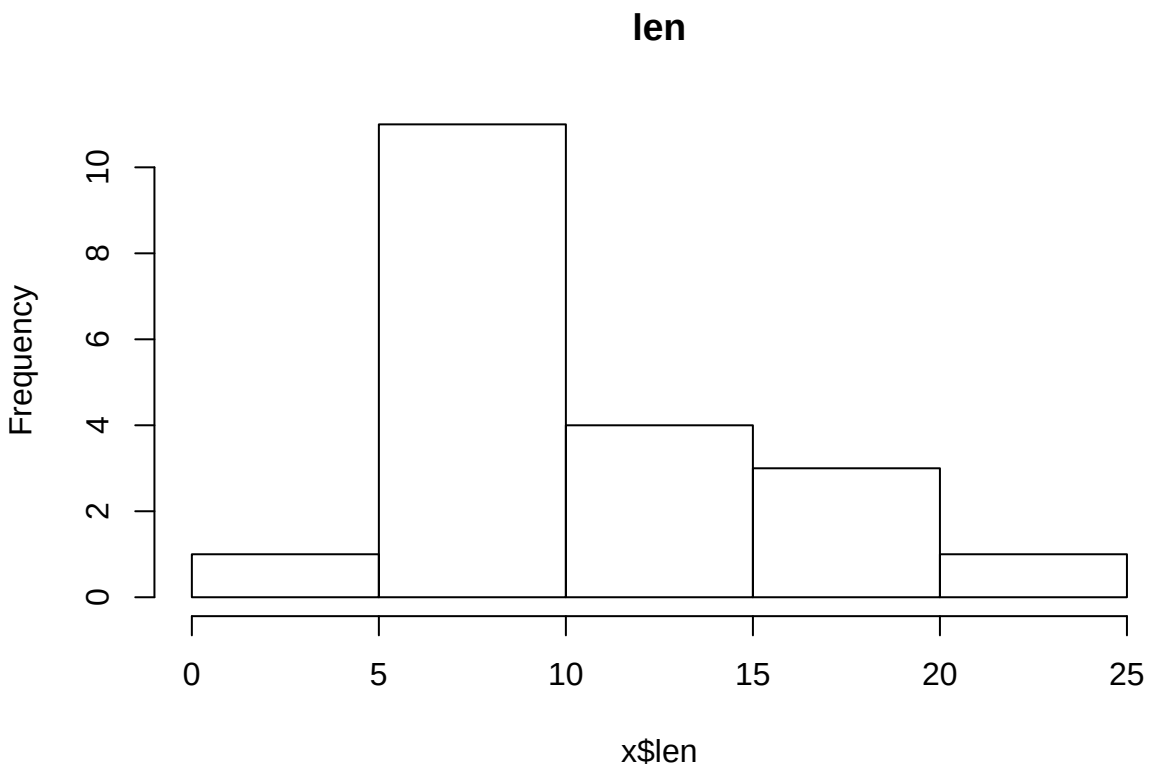

```
t.test(
    len ~ supp,
    data = ToothGrowth,
    var.equal = T, paired = F, alternative = "two.sided"
       )
```

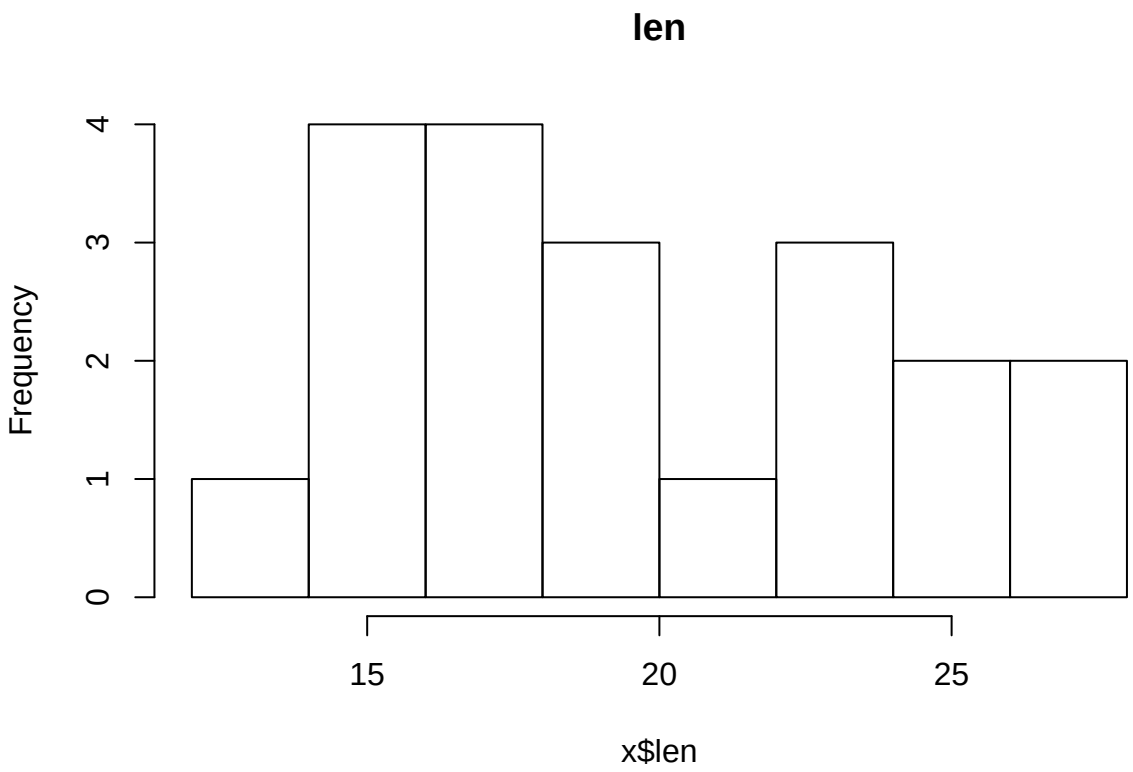
```
##
##  Two Sample t-test
##
## data:  len by supp
## t = 1.9153, df = 58, p-value = 0.06039
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.1670064  7.5670064
## sample estimates:
## mean in group OJ mean in group VC
##         20.66333         16.96333
```

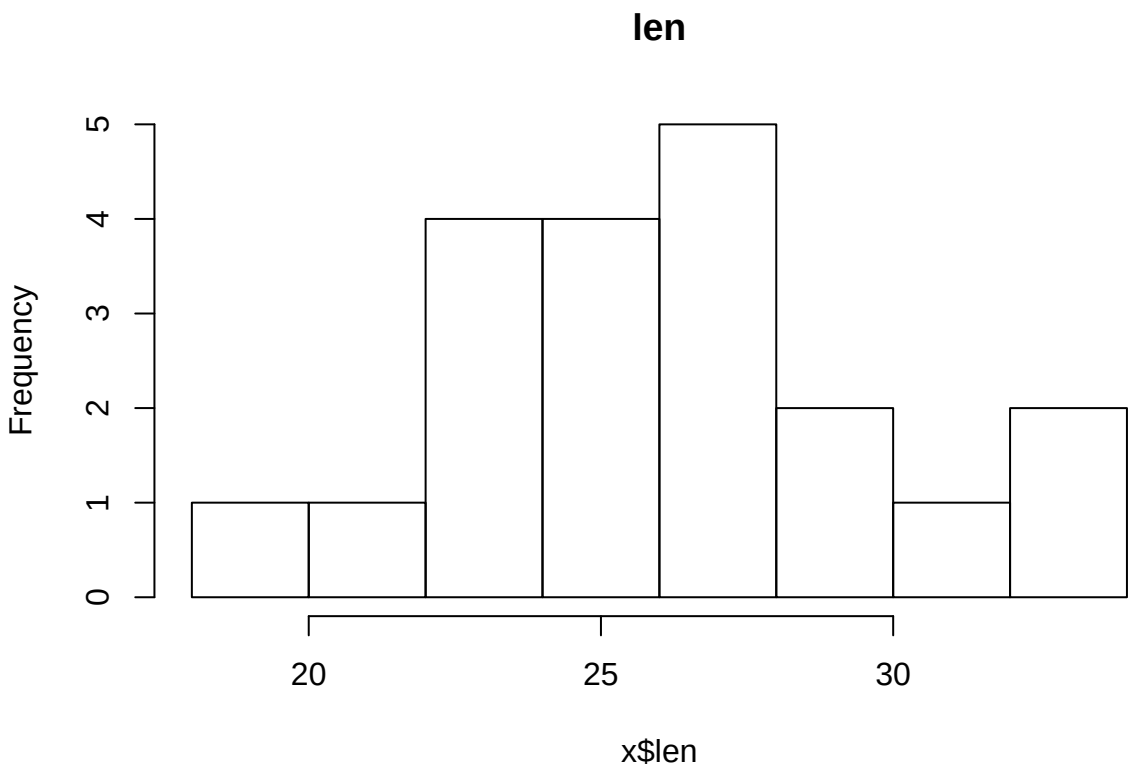
```
boxplot(len ~ dose, ToothGrowth)
```




```
tooth_new <- ToothGrowth

tooth_new$dose <- as.factor(tooth_new$dose)

tooth_split <-  split(tooth_new, tooth_new$dose)

lapply(tooth_split, function(x) hist(x$len, main = names(x[1])))
```


len
Frequency
0
2
4
6
8
10
0
5
10
15
20
25
x$len


len
Frequency
0
1
2
3
4
15
20
25
x$len




```
## $`0.5`
## $breaks
## [1]  0  5 10 15 20 25
##
## $counts
## [1]  1 11  4  3  1
##
## $density
## [1] 0.01 0.11 0.04 0.03 0.01
##
## $mids
## [1]  2.5  7.5 12.5 17.5 22.5
##
## $xname
## [1] "x$len"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
##
## $`1`
## $breaks
## [1] 12 14 16 18 20 22 24 26 28
##
```

```
## $counts
## [1] 1 4 4 3 1 3 2 2
##
## $density
## [1] 0.025 0.100 0.100 0.075 0.025 0.075 0.050 0.050
##
## $mids
## [1] 13 15 17 19 21 23 25 27
##
## $xname
## [1] "x$len"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
##
## $`2`
## $breaks
## [1] 18 20 22 24 26 28 30 32 34
##
## $counts
## [1] 1 1 4 4 5 2 1 2
##
## $density
## [1] 0.025 0.025 0.100 0.100 0.125 0.050 0.025 0.050
##
## $mids
## [1] 19 21 23 25 27 29 31 33
##
## $xname
## [1] "x$len"
##
## $equidist
## [1] TRUE
##
## attr(,"class")
## [1] "histogram"
```

```
t.test(
    tooth_split$`0.5`$len,
    tooth_split$`1`$len,
    var.equal = T, paired = F, alternative = "two.sided"
)
```

```
##
##  Two Sample t-test
##
## data:  tooth_split$`0.5`$len and tooth_split$`1`$len
## t = -6.4766, df = 38, p-value = 1.266e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -11.983748  -6.276252
```

```
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

```
t.test(
    tooth_split$`0.5`$len,
    tooth_split$`2`$len,
    var.equal = T, paired = F, alternative = "two.sided"
)
```

```
##
##  Two Sample t-test
##
## data:  tooth_split$`0.5`$len and tooth_split$`2`$len
## 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:
##  -18.15352 -12.83648
## sample estimates:
## mean of x mean of y
##    10.605    26.100
```

```
t.test(
    tooth_split$`1`$len,
    tooth_split$`2`$len,
    var.equal = T, paired = F, alternative = "two.sided"
)
```

```
##
##  Two Sample t-test
##
## data:  tooth_split$`1`$len and tooth_split$`2`$len
## 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:
##  -8.994387 -3.735613
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```