Tooth Growth Exploration

James Whedbee

September 25, 2015

The purpose of this report is to investigate the Tooth Growth data set in the R datasets package.

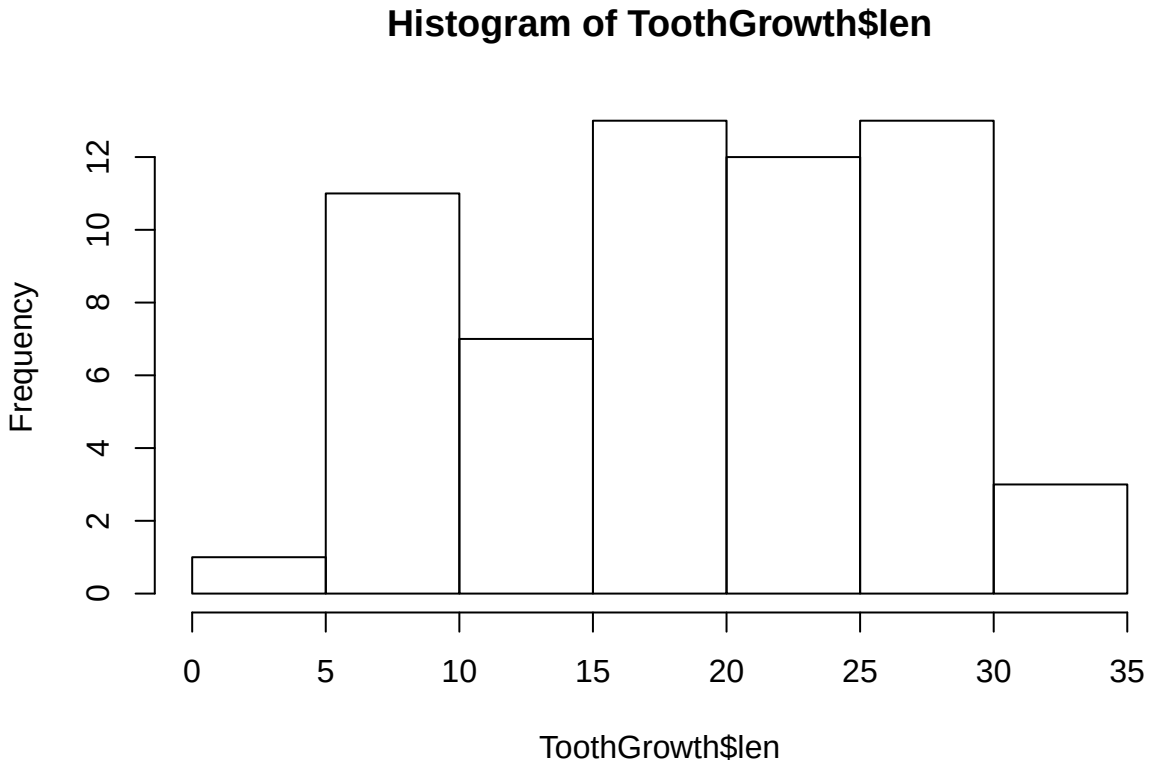
Exploratory Analysis

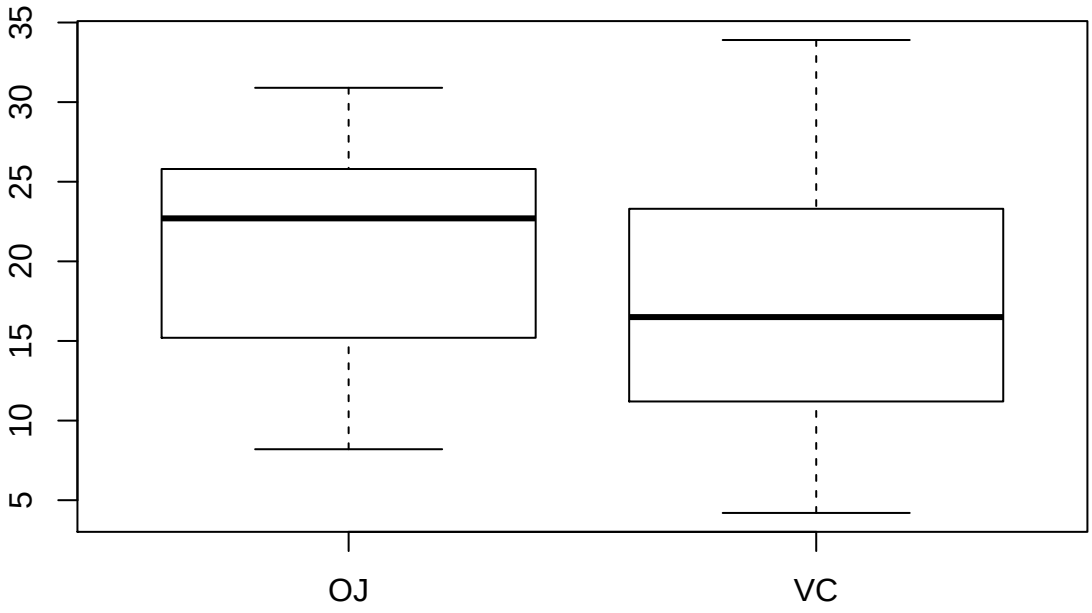
First, here is a brief summary of the Tooth Growth data:

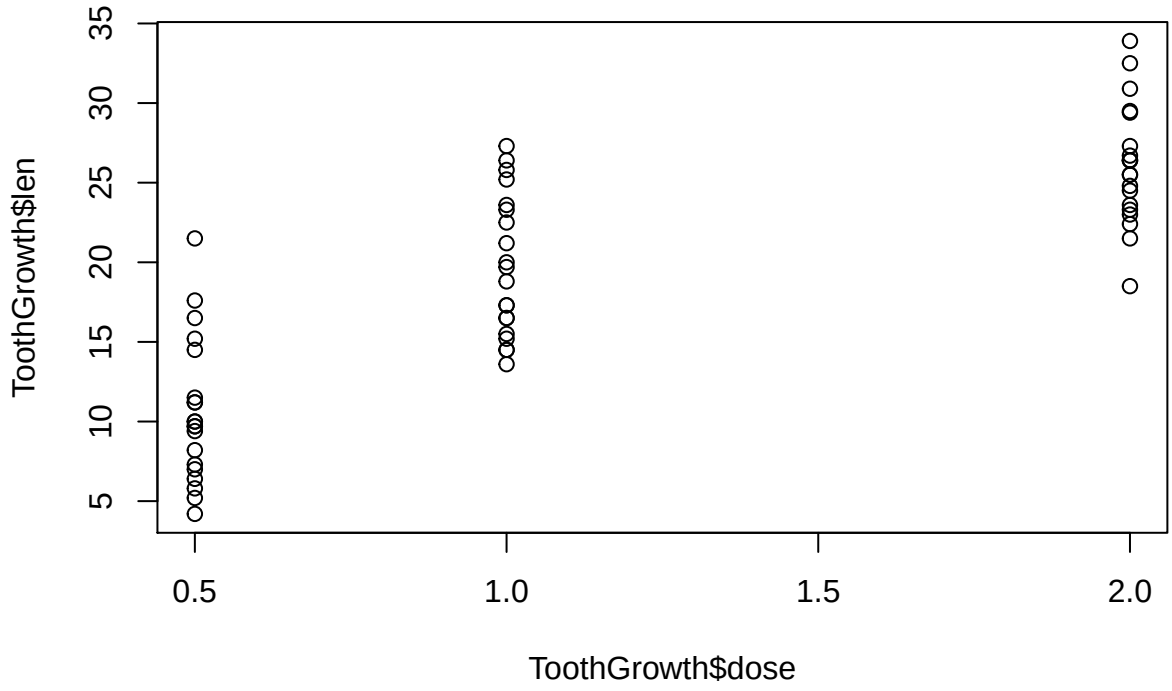
```
## 'data.frame':    60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 2 ...
##  $ dose: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
##       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
```

Below, you can see a plot of the distribution of the len variable, a plot of len split by supp, and a plot of len split by dose, respectively:







Comparing len by supp and dose

By supp

Based on our exploratory analysis, it seems possible that length is greater when supp is "OJ" than when it is "VJ". To test this, a t-test was performed to compare the difference in means between the supps, with the null hypothesis that there was no difference in means.

```
lenOfOJ <- ToothGrowth$len[ToothGrowth$supp=="OJ"]
lenOfVC <- ToothGrowth$len[ToothGrowth$supp=="VC"]
t.test(lenOfOJ,lenOfVC)
```

```
##
##  Welch Two Sample t-test
##
## data:  lenOfOJ and lenOfVC
## 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
```

The 95% confidence interval does contain 0, so for now we accept the null hypothesis. The p-value was very close to .05, so it seems worthwhile to gather more data, which could help stengthen the evidence to reject the null hypothesis.

By dose

Based on our exploratory analysis, it seems possible that as length increases as dose increases. To test this, a t-test was performed to compare the difference in means between the doses, with the null hypothesis that there was no difference in means.

```
lenOfDoseHalf <- ToothGrowth$len[ToothGrowth$dose==.5]
lenOfDose1 <- ToothGrowth$len[ToothGrowth$dose==1]
lenOfDose2 <- ToothGrowth$len[ToothGrowth$dose==2]
t.test(lenOfDose1,lenOfDose2)
```

```
##
##  Welch Two Sample t-test
##
## data:  lenOfDose1 and lenOfDose2
## t = -4.9005, df = 37.101, p-value = 1.906e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -8.996481 -3.733519
## sample estimates:
## mean of x mean of y
##    19.735    26.100
```

```
t.test(lenOfDose1,lenOfDoseHalf)
```

```
##
##  Welch Two Sample t-test
##
## data:  lenOfDose1 and lenOfDoseHalf
## t = 6.4766, df = 37.986, p-value = 1.268e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   6.276219 11.983781
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

Neither 95% confidence interval contains 0, so it is likely that there is a difference in len when split by dose. This conclusion is based on the assumption that the variance of the groups in question is the same. It was also assumed that there are no unobserved covariates which are contaminating the results.