Report - Analysis of the ToothGrowth dataset

Hauer-Glocke

12 Juli 2017

Overview of report

I conducted the following steps, which I will elaborate in the following:

1. Loading of the ToothGrowth data and performing of basic exploratory data analyses
2. Provide a basic summary of the data.
3. Use confidence intervals and/or hypothesis tests to comparetooth growth by supp and dose.
4. State your conclusions and the assumptions needed for your conclusions.

For the analysis and functions it is necessary to load the following packages:

library(ggplot2)

library(dplyr)

Exploratory Data Analysis and Data Summary

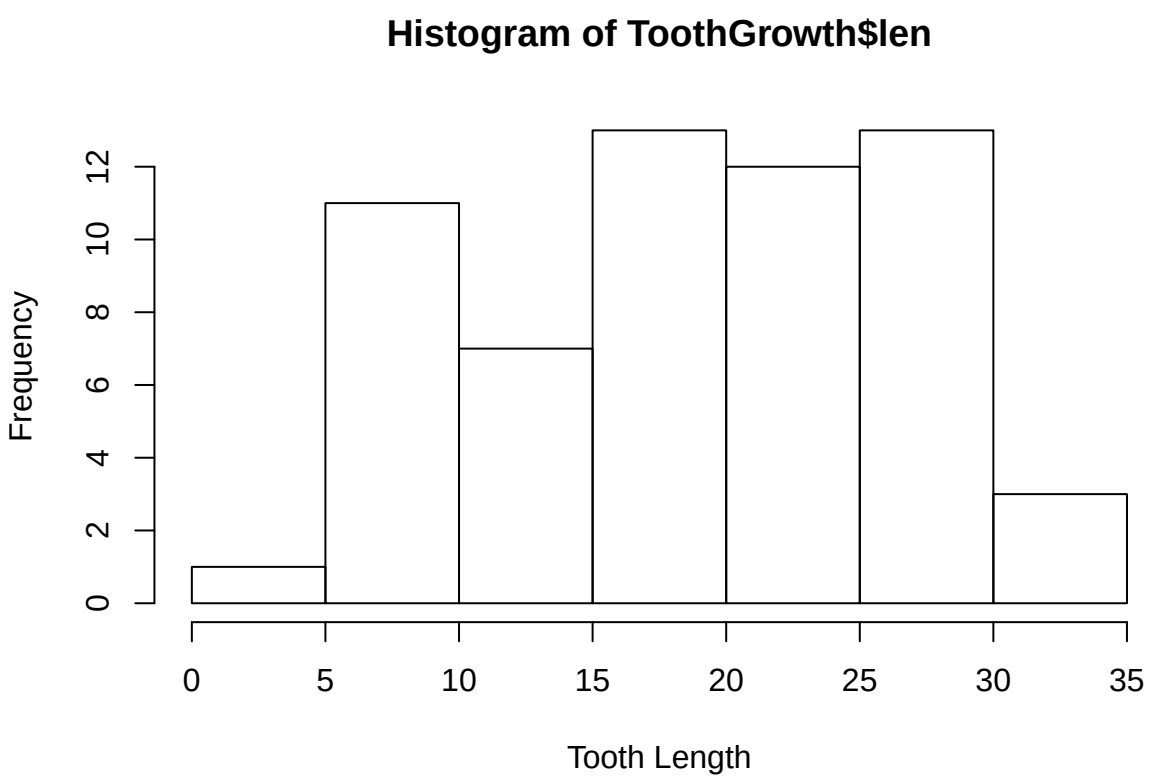
```
#Loading and reading data. The dosage has 3 levels. The data was read as numeric, so I changed it to fa
data("ToothGrowth")
ToothGrowth$dose <- as.factor(ToothGrowth$dose)

#See the different classes and variables sizes
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: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 1 ...
```

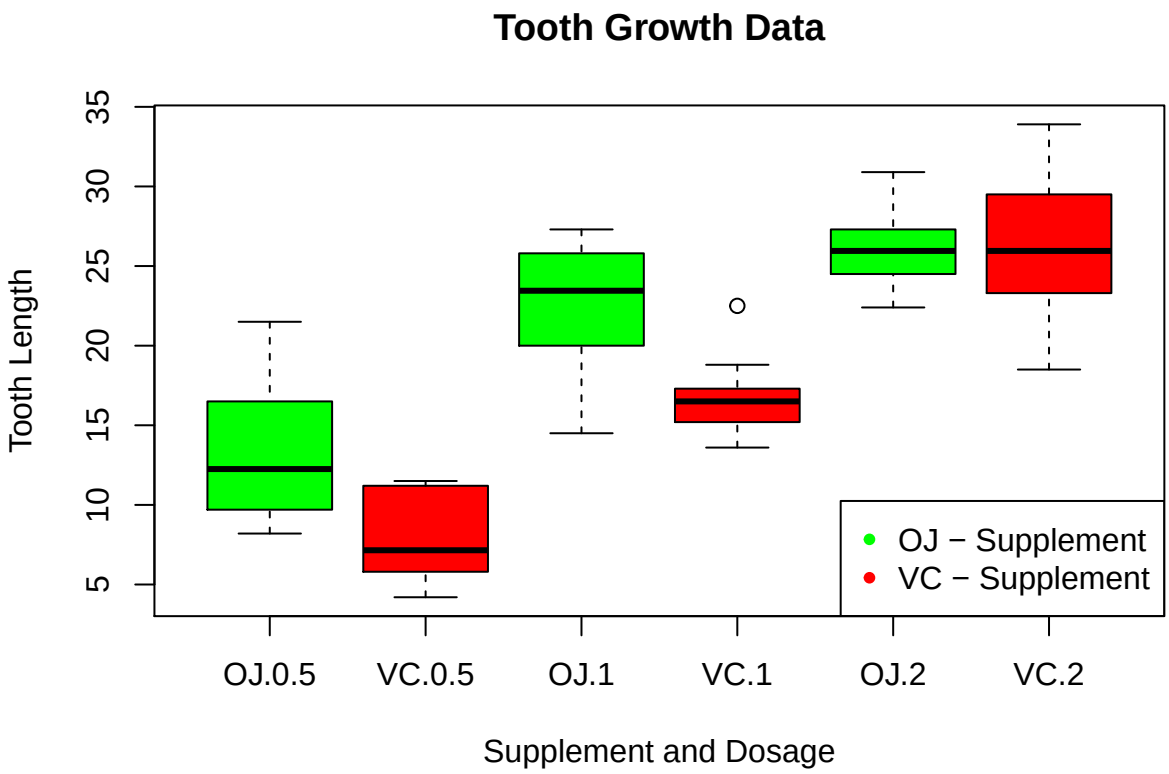
```
#Overview of the different values of the three variables.
summary(ToothGrowth)
```

```
##       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
```

```
#Shows the distribution of our outcome variable "Tooth Length"
hist(ToothGrowth$len, xlab = "Tooth Length")
```




```
#This boxplot shows the interactions between the variables and its subsequent values.
boxplot(len ~ supp*dose,
        data=ToothGrowth,
        main="Tooth Growth Data",
        xlab="Supplement and Dosage",
        ylab="Tooth Length",
        col=c("green","red"))
legend("bottomright", pch=20, legend=c("OJ - Supplement","VC - Supplement"),col=c("green","red"))
```



The tooth length can have value of 4.2 to 33.9. The boxplot suggests that there is an effect of the dosage of vitamin C. A difference in the supplements via vitamin C (VC) treatment or orange juice (OJ) is not visible from the analysis.

Hypothteses Tests and Confidence Intervals

Subsequently, I will discuss two hypotheses and show the relevant confidence intervals for my tests:

1. Is there a difference between the supplement Orange Juice (OJ) and the Vitamin C treatment via ACA (VC).
2. The Dosage of Vitamin C (dose) influences the Growth of the teeth length.

Hypothesis 1 - Difference in Supplements

The means in Tooth Length of the two supplements (OJ, VC) are 20.66 and 16.96 respectively, suggusting higher teeth growth under the OJ supplement.

Therefore, I test the Null Hypothesis of equal means with the alternative that the OJ supplement leads to higher teeth lengths.

H0 <- mean(oj$len) = mean(vc$len)

H1 <- mean(oj$len) > mean(vc$len)

```
oj <- ToothGrowth %>% filter(supp=="OJ")
vc <- ToothGrowth %>% filter(supp=="VC")
```

```r
t.test(oj$len, vc$len, alternative = "greater", paired = FALSE, var.equal = FALSE, conf.level = 0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  oj$len and vc$len
## t = 1.9153, df = 55.309, p-value = 0.03032
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  0.4682687       Inf
## sample estimates:
## mean of x mean of y
##  20.66333  16.96333
```

The P-Value of 0.03032 suggest a rejecting of the Null Hypothesis on a 5%-level. Concluding, that the supplement OJ leads to a higher teeth length than the supplement VC.

The 95%-confidence interval to support the alternative hypthesis is suggusted to be 0.4682687 to infinity in the difference between supplement OJ and supplement VC.

```r
oj_2 <- ToothGrowth %>% filter(supp=="OJ") %>% filter(dose==2)
vc_2 <- ToothGrowth %>% filter(supp=="VC") %>% filter(dose==2)

t.test(oj_2$len, vc_2$len, alternative = "two.sided", paired = FALSE, var.equal = FALSE, conf.level = 0
```

```
##
##  Welch Two Sample t-test
##
## data:  oj_2$len and vc_2$len
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean of x mean of y
##     26.06     26.14
```

Nonetheless, the preceding test result in a P-Value of 0.9639, which suggusts a failed rejection of the Null Hypothesis, that there are differences in the supplements, when the Dosage level is 2.

The 95%-confidence interval for the null hypthesis is subsequently -3.79807 to 3.63807 in the hypothesis test.

Hypothesis 2 - Difference in Dosage

The means in Tooth Length for the three dosage levels (0.5, 1, 2) are 10.61, 19.74 and 26.1 respectively, suggusting higher teeth growth with increasing level of dosage.

Therefore, I test the Null Hypothesis of equal means with the alternative that the higher dosage leads to higher teeth lengths.

First: > H0 <- mean(dose_10$len) = mean(dose_05$len)

H1 <- mean(dose_10$len) > mean(dose_10$len)

Second: > H0 <- mean(dose_20$len) = mean(dose_10$len)

H1 <- mean(dose_20$len) > mean(dose_10$len)

```r
dose_05 <- ToothGrowth %>% filter(dose==0.5)
dose_10 <- ToothGrowth %>% filter(dose==1)
dose_20 <- ToothGrowth %>% filter(dose==2)

#The first hypothesis test:
t.test(dose_10$len, dose_05$len, alternative = "greater", paired = FALSE, var.equal = FALSE, conf.level
```

```
##
##  Welch Two Sample t-test
##
## data:  dose_10$len and dose_05$len
## t = 6.4766, df = 37.986, p-value = 6.342e-08
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  6.753323      Inf
## sample estimates:
## mean of x mean of y
##    19.735    10.605
```

```r
#The second hypthesis test:
t.test(dose_20$len, dose_10$len, alternative = "greater", paired = FALSE, var.equal = FALSE, conf.level
```

```
##
##  Welch Two Sample t-test
##
## data:  dose_20$len and dose_10$len
## t = 4.9005, df = 37.101, p-value = 9.532e-06
## alternative hypothesis: true difference in means is greater than 0
## 95 percent confidence interval:
##  4.17387     Inf
## sample estimates:
## mean of x mean of y
##    26.100    19.735
```

The hypothesis tests show a P-Value of 6.342e-08 and 9.532e-06 respectively, suggusting a rejection of the null hypotheses. Therefore, I conclude that an increase in the dosage level leads to an increase in teeth length.

The tests also suggust confidence intervals for the alternative hypotheses (H1) on a 0.95 level to be true: 6.753323 to infinity and 4.17387 to infinity respectively.

Assumptions and Conclusions

The following conclusion are drawn in consideration of these assumptions:

1. Tooth Growth follows a normal distribution.
2. All observations are independent and identically distributed (i.i.d.)

In summary, my analysis support the following conclusions:

- The supplement OJ has got a higher influence on teeth growth then the supplement VC for dosage 0.5mg and 1mg. The data does not support this finding for any increment in dosage.
- An increment in dosage of vitamin (regardless from 0.5mg to 1mg or 1mg to 2mg) leads to an increase in teeth length.