

“Statistical Inference - ToothGrowth Data”

David Contreras Carretero

29/10/2020

Part 2: Basic Inferential Data Analysis Instructions

Now in the second portion of the project, 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.
4. State your conclusions and the assumptions needed for your conclusions.

1. Load the ToothGrowth data and perform some basic exploratory data analyses

```
#load datasets & packages
library(datasets)
library(ggplot2)
library(dplyr)
```

```
#load data
data(ToothGrowth)
```

```
TG <- ToothGrowth
```

```
dim(TG)
```

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

```
head(TG)
```

```
##      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
```

```
TG$dose <-factor(TG$dose)
str(TG)
```

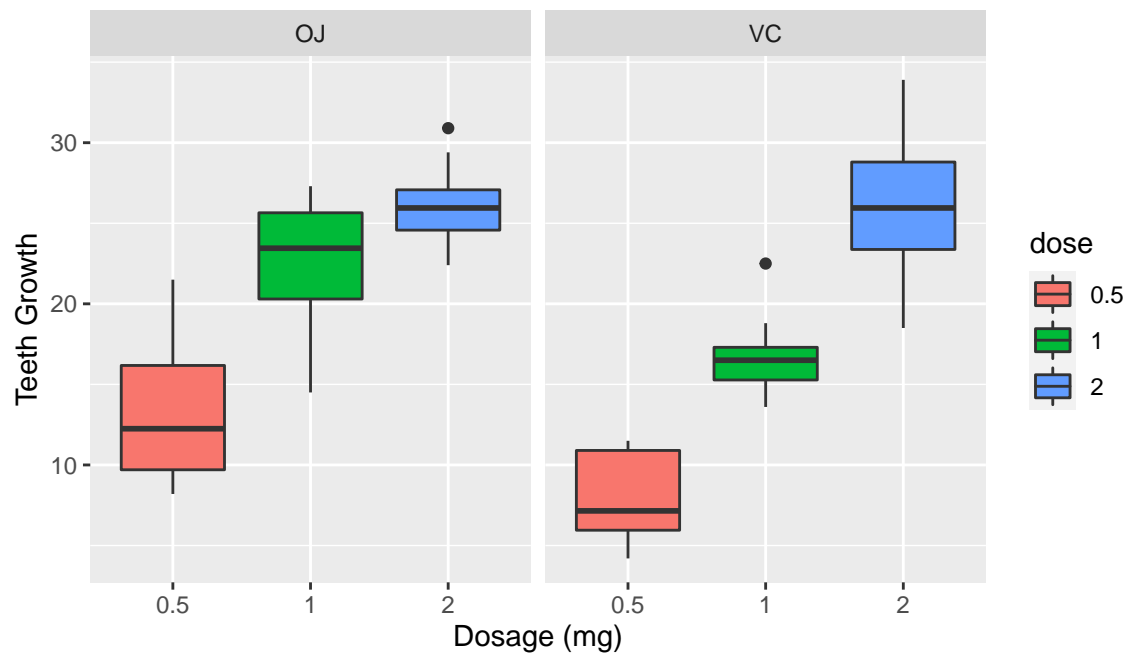
```
## 'data.frame':   60 obs. of  3 variables:
##  $ len : num  4.2 11.5 7.3 5.8 6.4 10 11.2 11.2 5.2 7 ...
##  $ supp: Factor w/ 2 levels "OJ","VC": 2 2 2 2 2 2 2 2 2 ...
##  $ dose: Factor w/ 3 levels "0.5","1","2": 1 1 1 1 1 1 1 1 1 ...
```

2. Provide a basic summary of the data.

```
## # A tibble: 3 x 3
##   dose MeanLen SDLen
##   <fct>   <dbl> <dbl>
## 1 0.5     10.6  4.50
## 2 1      19.7  4.42
## 3 2      26.1  3.77
```

```
## # A tibble: 6 x 4
## # Groups:   supp [2]
##   supp dose MeanLen SDLen
##   <fct> <fct>   <dbl> <dbl>
## 1 OJ   0.5     13.2  4.46
## 2 OJ   1      22.7  3.91
## 3 OJ   2      26.1  2.66
## 4 VC   0.5     7.98  2.75
## 5 VC   1      16.8  2.52
## 6 VC   2      26.1  4.80
```

```
plot <- ggplot(TG,
               aes(x=dose,y=len,fill=dose))
plot + geom_boxplot(notch=F) + facet_grid(~supp) +
  scale_x_discrete("Dosage (mg)") +
  scale_y_continuous("Teeth Growth")
```



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

```
TGLow <-rbind(TG[(TG$dose == 0.5) & (TG$supp == "OJ"),],
             TG[(TG$dose == 1) & (TG$supp == "OJ"),],
             TG[(TG$dose == 0.5) & (TG$supp == "VC"),],
             TG[(TG$dose == 1) & (TG$supp == "VC"),])
testlow <- t.test(len ~ supp, data =TGLow,var.equal = FALSE)
testlow$p.value;testlow$conf.int
```

```
## [1] 0.00423861
```

```
## [1] 1.875234 9.304766
## attr(,"conf.level")
## [1] 0.95
```

```
TGHigh <-rbind(TG[(TG$dose == 2) & (TG$supp == "OJ"),],
              TG[(TG$dose == 2) & (TG$supp == "VC"),])
testHigh <- t.test(len ~ supp, data =TGHigh,var.equal = FALSE)
testHigh$p.value;testHigh$conf.int
```

```
## [1] 0.9638516
```

```
## [1] -3.79807 3.63807
## attr(,"conf.level")
## [1] 0.95
```

4. State your conclusions and the assumptions needed for your conclusions.

Basic exploratory data analysis show that we have 60 observations of three variables (len, supp, dose), that seems to record tooth growth (len) of subjects using two different supplements (supp), "OJ" and "VC", in three different dosifications 0.5, 1 and 2 (assuming miligrams). We assume that the test has been performed in random unrelated subjects in each different group of supplement and dosage.

Summarizing data we observe that the subjects using supplement "OJ" had best results than users of "VC". With both supplements the growth increases as it does the dosage. At dosage of 2mg toothgrowth seems to be similar with both supplements although we observe more variability in "VC" data.

So based on these summaries our null hypothesis would be that at lower dosage(0.5,1 mg) OJ works better than "VC", but at a dosage of 2mg they are equally valid.

At a confidence level of confidence interval 1.8752343, 9.3047657 doesn't contain 0, so mean growth is constantly higher using "OJ" supplement. P-value 0.0042386 is below 0.5 threshold. That confirms hypothesis of "OJ" working better than "VC" at lower dosage.

At a confidence level of confidence interval -3.7980705, 3.6380705 contains 0, meaning that average tooth growth is similar using "OJ" or "VC" supplement. P-value 0.9638516 is close to 1. Confirming the hypothesis of "OJ" and "VC" having similar results at a dosage of 2mg.