# Analyzing Tooth Growth

*Ronald Hensbergen*

*29 november 2016*

## Overview

The source of this data set is origined from a study of the length of odontoblasts, which can be directly related to tooth growth in guinea pigs. Details of the study can be found at the R help file for this data set.

## Exploratory Analysis

First I declared all different parts of the data as different variables.

```
data("ToothGrowth")
ToothGrowth$dose <- as.factor(ToothGrowth$dose)

TG.0.5 <- ToothGrowth[ToothGrowth$dose  == 0.5, ]
TG.1 <- ToothGrowth[ToothGrowth$dose  == 1, ]
TG.2 <- ToothGrowth[ToothGrowth$dose  == 2, ]
TG.OJ <- ToothGrowth[ToothGrowth$supp == "OJ", ]
TG.VC <- ToothGrowth[ToothGrowth$supp == "VC", ]

TG.OJ.0.5 <- ToothGrowth[ToothGrowth$dose  == 0.5 & ToothGrowth$supp == "OJ", ]
TG.OJ.1 <- ToothGrowth[ToothGrowth$dose  == 1 & ToothGrowth$supp == "OJ", ]
TG.OJ.2 <- ToothGrowth[ToothGrowth$dose  == 2 & ToothGrowth$supp == "OJ", ]
TG.OJ.N2 <- ToothGrowth[ToothGrowth$dose != 2 & ToothGrowth$supp == "OJ", ]
TG.VC.0.5 <- ToothGrowth[ToothGrowth$dose  == 0.5 & ToothGrowth$supp == "VC", ]
TG.VC.1 <- ToothGrowth[ToothGrowth$dose  == 1 & ToothGrowth$supp == "VC", ]
TG.VC.2 <- ToothGrowth[ToothGrowth$dose  == 2 & ToothGrowth$supp == "VC", ]
TG.VC.N2 <- ToothGrowth[ToothGrowth$dose != 2 & ToothGrowth$supp == "VC", ]
```

I started off with some basic analysis, like the column definitions and the summary of the data.

```
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 ...
```

```
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
```

Also I did some Student tests (T-Tests)

```
t.test(len ~ supp, data = ToothGrowth)
```

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## 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 in group OJ mean in group VC
##         20.66333         16.96333
```

```
t.test(TG.1$len,TG.2$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  TG.1$len and TG.2$len
## 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(TG.0.5$len,TG.1$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  TG.0.5$len and TG.1$len
## 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:
##  -11.983781  -6.276219
## sample estimates:
## mean of x mean of y
##    10.605    19.735
```

```
t.test(TG.OJ.2$len,TG.VC.2$len)
```

```
##
##  Welch Two Sample t-test
##
## data:  TG.OJ.2$len and TG.VC.2$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
```

```
t.test(TG.OJ.N2$len,TG.VC.N2$len)
```

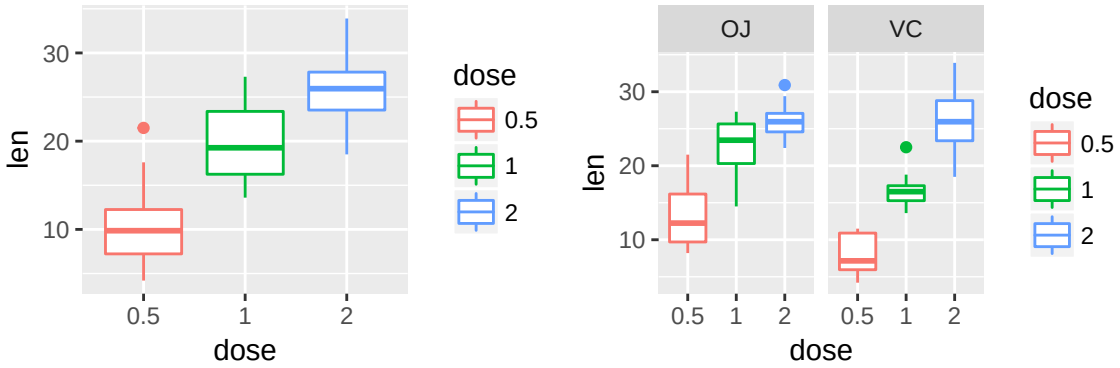
```
##
##  Welch Two Sample t-test
##
## data:  TG.OJ.N2$len and TG.VC.N2$len
## t = 3.0503, df = 36.553, p-value = 0.004239
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  1.875234 9.304766
## sample estimates:
## mean of x mean of y
##    17.965    12.375
```

Above shows that there is no statistical evidence that the means of the Orange Juice and Vitamine C are different (first T-Test). This is mostly due to the OJ and VC samples of a dose of 2 (4th T-Test). The P-Value here is 0.96, where the means are almost equal (26.06 and 26.14). If that dose is excluded (last T-test), then the P-value is relatively low again (0.004) and there is statistical evidence.

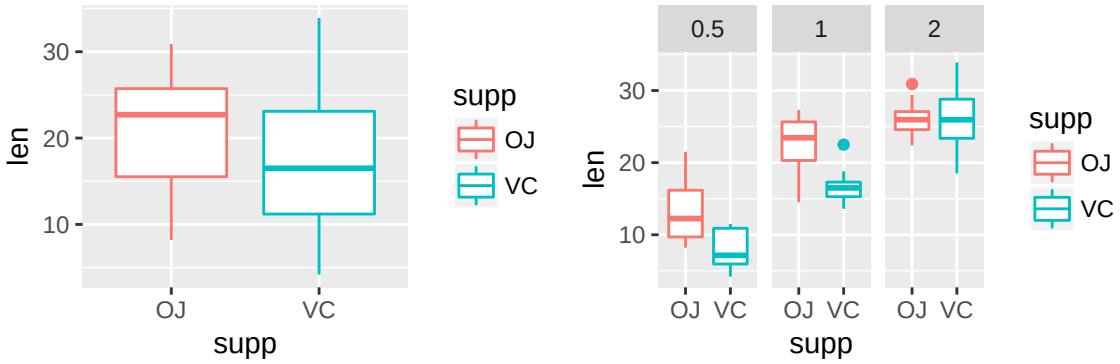
## Visualisations

Often visualisations explain better what's going on than just some formulas and text...

```
library(ggplot2)
library(gridExtra)
g1 <- ggplot(aes(y=len,x=dose, col=dose), data=ToothGrowth)
g1 <- g1 + geom_boxplot()
g1a<- g1 + facet_grid(. ~ supp)
grid.arrange(g1,g1a,ncol=2)
```



```
g2 <- ggplot(aes(y=len,x=supp,col=supp), data=ToothGrowth)
g2 <- g2 + geom_boxplot()
g2a <- g2 + facet_grid(. ~ dose)
grid.arrange(g2,g2a,ncol=2)
```




## Conclusions

This is only a very brief analysis and by no means comprehensive. This already can lead to some first conclusions: - Higher doses of either Vitamine C or Orange Juice lead to faster growth of odontoblasts at Guinea Pigs. - There isn't a statistical evidence that the means of Vitamine C and Orange Juice are different. - Previous point is mostly due to the equal means of Vitamine C and Orange Juise with a dose of 2. With that part of the data excluded there is statistical evidence again for a difference in means.

### Session information

Last section of this paper is the Session Info where all the necessary information is stored for reproducibility.

```
library(devtools)
devtools::session_info()
```

```
## Session info -------------------------------------------------------------

##  setting  value
##  version  R version 3.3.2 (2016-10-31)
##  system   x86_64, mingw32
##  ui       RTerm
##  language (EN)
##  collate  Dutch_Netherlands.1252
##  tz       Europe/Berlin
##  date     2016-12-11

## Packages -----------------------------------------------------------------

##  package    * version date       source
##  assertthat   0.1     2013-12-06 CRAN (R 3.3.1)
##  backports    1.0.4   2016-10-24 CRAN (R 3.3.2)
##  colorspace   1.3-1   2016-11-18 CRAN (R 3.3.2)
##  devtools   * 1.12.0  2016-06-24 CRAN (R 3.3.2)
```

```
##  digest       0.6.10  2016-08-02 CRAN (R 3.3.1)
##  evaluate     0.10    2016-10-11 CRAN (R 3.3.1)
##  ggplot2    * 2.2.0   2016-11-11 CRAN (R 3.3.1)
##  gridExtra  * 2.2.1   2016-02-29 CRAN (R 3.3.1)
##  gtable       0.2.0   2016-02-26 CRAN (R 3.3.1)
##  htmltools    0.3.5   2016-03-21 CRAN (R 3.3.1)
##  knitr        1.15.1  2016-11-22 CRAN (R 3.3.2)
##  labeling     0.3     2014-08-23 CRAN (R 3.3.1)
##  lazyeval     0.2.0   2016-06-12 CRAN (R 3.3.1)
##  magrittr     1.5     2014-11-22 CRAN (R 3.3.1)
##  memoise      1.0.0   2016-01-29 CRAN (R 3.3.1)
##  munsell      0.4.3   2016-02-13 CRAN (R 3.3.1)
##  plyr         1.8.4   2016-06-08 CRAN (R 3.3.1)
##  Rcpp         0.12.8  2016-11-17 CRAN (R 3.3.2)
##  reshape2     1.4.2   2016-10-22 CRAN (R 3.3.1)
##  rmarkdown    1.2     2016-11-21 CRAN (R 3.3.2)
##  rprojroot    1.1     2016-10-29 CRAN (R 3.3.2)
##  scales       0.4.1   2016-11-09 CRAN (R 3.3.2)
##  stringi      1.1.2   2016-10-01 CRAN (R 3.3.1)
##  stringr      1.1.0   2016-08-19 CRAN (R 3.3.1)
##  tibble       1.2     2016-08-26 CRAN (R 3.3.1)
##  withr        1.0.2   2016-06-20 CRAN (R 3.3.1)
##  yaml         2.1.14  2016-11-12 CRAN (R 3.3.2)
```