# Statistical_Inference

Reza Mofidi

15/06/2020

## Peer Graded Assignment: Statistical Inference Course (Part-2)

```
library(dbplyr)
library(ggplot2)
if(!file.exists("~/data")){dir.create("~/data")}
```

## Basic Inferential Data Analysis

In the second portion of the project, we're going to analyze the ToothGrowth data in the R datasets package.This database is one of the standard datasets available on R. It is standard 60 observation dataset of 3 variables. The dataset examines the length of odontoblasts in relation to the dose of vitamin C administeredin 2 methods OJ and VC. the doses of vitamin C are o.5, 1 and 2 mgs.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:dbplyr':
##
##     ident, sql
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
data(ToothGrowth)
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
```

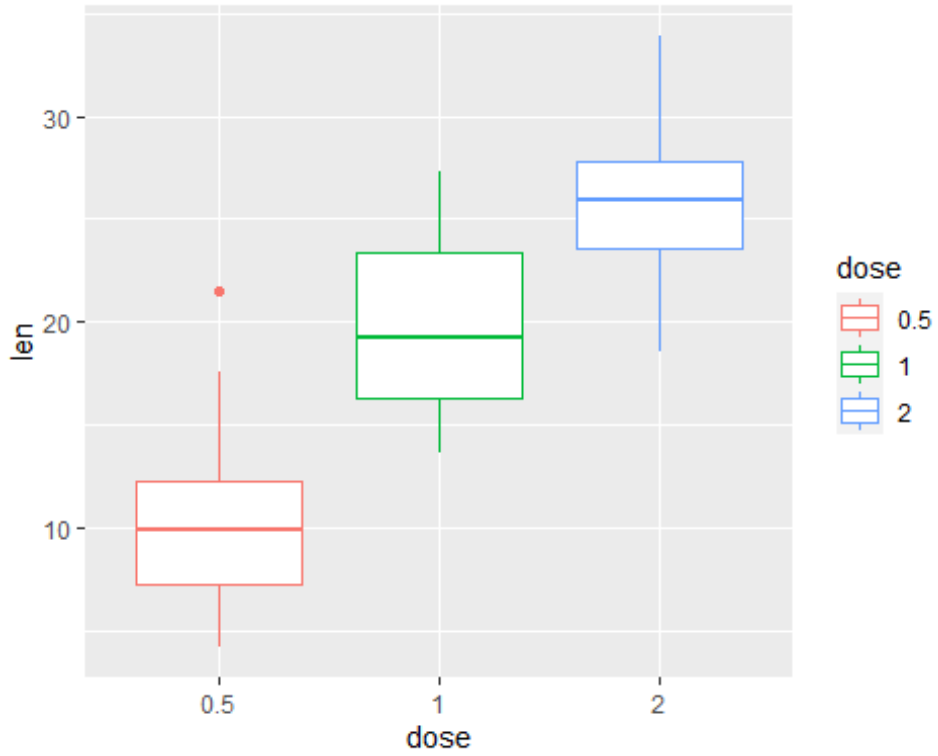
Next we summarize the dataset by dose and supplement. This returns the data classfied into 6 rows which are classified by the 2 metods of delivery of vitamin C (OJ and VC) and the 3 different doses (0.5, 1 and 2mgs) the mean Odontoblast length and standard deviation for each group is provided.

```
ToothGrowth %>%
     group_by(supp,dose) %>%
    summarize(lenmean=mean(len), lensd=sd(len), count= n())
```

```
## # A tibble: 6 x 5
## # Groups:   supp [2]
##   supp   dose lenmean lensd count
##   <fct> <dbl>   <dbl> <dbl> <int>
## 1 OJ      0.5   13.2   4.46    10
## 2 OJ      1     22.7   3.91    10
## 3 OJ      2     26.1   2.66    10
## 4 VC      0.5    7.98  2.75    10
## 5 VC      1     16.8   2.52    10
## 6 VC      2     26.1   4.80    10
```

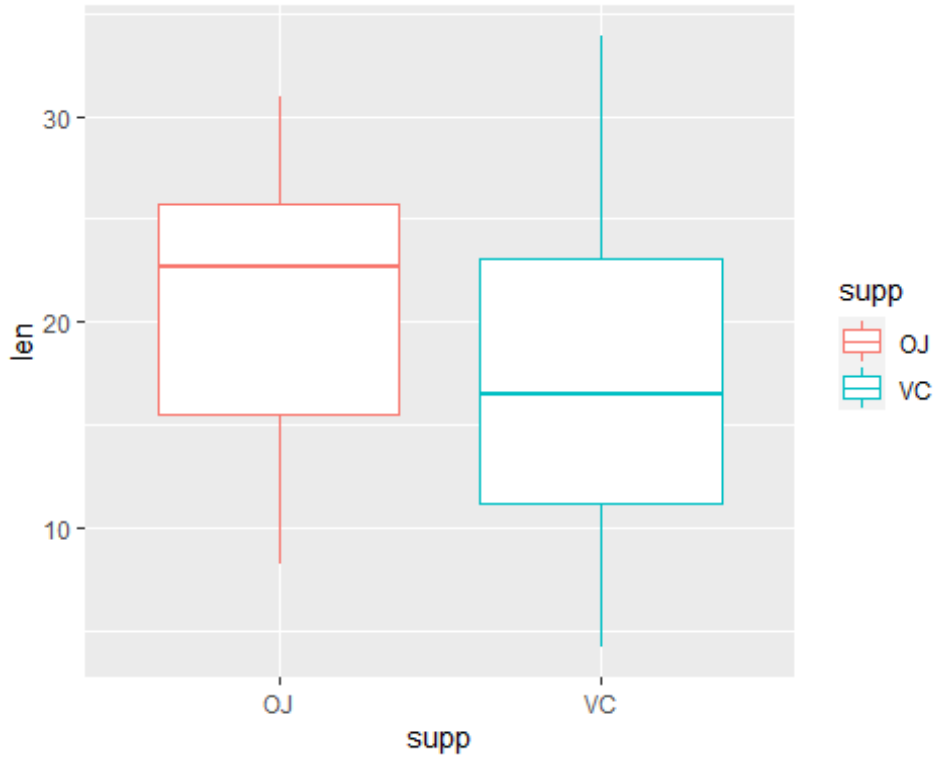
The following box plot illustrates the differences between these groups:

```
library(ggplot2)
ToothGrowth$dose <- as.factor(ToothGrowth$dose)
ggplot(ToothGrowth, aes(x=dose, y=len, color=dose)) +
  geom_boxplot()
```



There appears to be a clear dose dependent relationship between the length of odontoblasts and the dose of vitamin C administered (Above). There does not appear to be a similar relationship between the supp vatiable and the length of the odontoblasts (len). Below:

```
ToothGrowth$supp <- as.factor(ToothGrowth$supp)
ggplot(ToothGrowth, aes(x=supp, y=len, color=supp)) +
  geom_boxplot()
```




Lets examine the above hypotheses:

### Tooth growth by method of delivery (supp)

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

## [1] -0.1710156  7.5710156
## attr(,"conf.level")
## [1] 0.95

t.test(len ~ supp, data = ToothGrowth)$p.value

## [1] 0.06063451
```

The above data suggests that the differences tooth growth by (supp) or method of vit C delivery does not appear to be statistically significant as the confidence intervals cross 0 and the p value is higher than 0.5.

### Tooth growth by Vitamin C dose

In order to be able to make this comparison with T test we repeat the comparison between the methods of delivery at the 3 dosage levels:

*Dosage level 2 mg:*

```
t.test(len ~ supp, data =ToothGrowth[ToothGrowth$dose == 2,])$conf.int
```

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

```
t.test(len ~ supp, data =ToothGrowth[ToothGrowth$dose == 2,])$p.value
```

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

*Dosage level 1 mg:*

```
t.test(len ~ supp, data =ToothGrowth[ToothGrowth$dose == 1,])$conf.int
```

```
## [1] 2.802148 9.057852
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(len ~ supp, data =ToothGrowth[ToothGrowth$dose == 1,])$p.value
```

```
## [1] 0.001038376
```

*Dosage level 0.5 mg:*

```
t.test(len ~ supp, data =ToothGrowth[ToothGrowth$dose == 0.5,])$conf.int
```

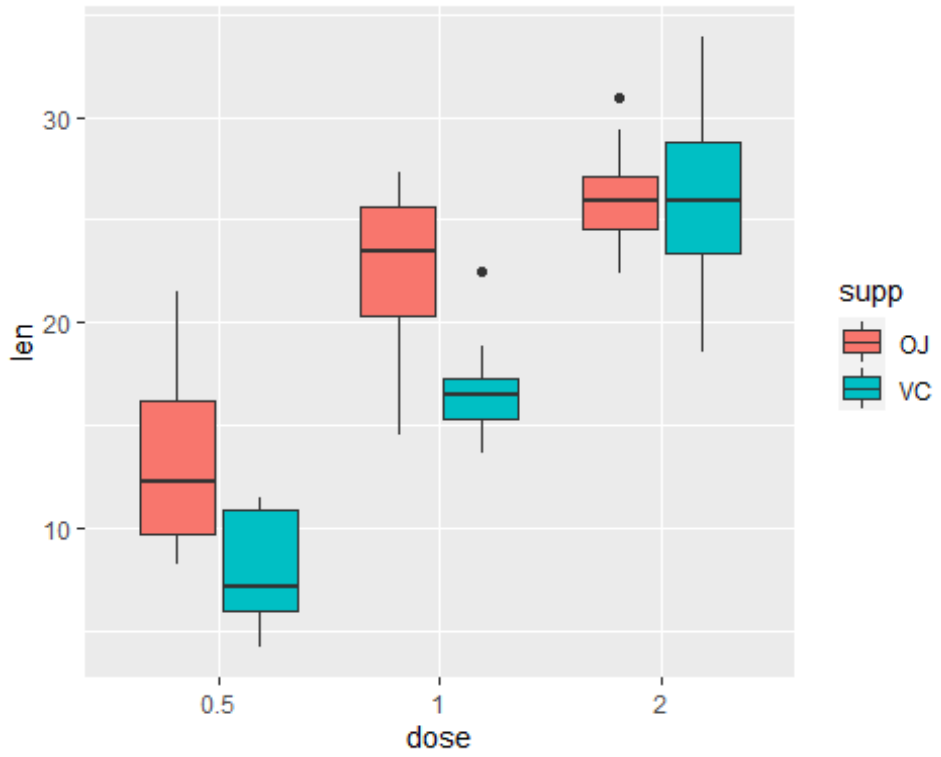
```
## [1] 1.719057 8.780943
## attr(,"conf.level")
## [1] 0.95
```

```
t.test(len ~ supp, data =ToothGrowth[ToothGrowth$dose == 0.5,])$p.value
```

```
## [1] 0.006358607
```

The differences in tooth growth (len values) between OJ and VC delivery methods are statistically significant for the dosage subgroups 0.5 mgs and 1 mgs but not at 2 mg dosage values. This is illustrated in the following box plot:

```
ggplot(ToothGrowth, aes(x=dose, y=len, fill=supp)) +
  geom_boxplot()
```

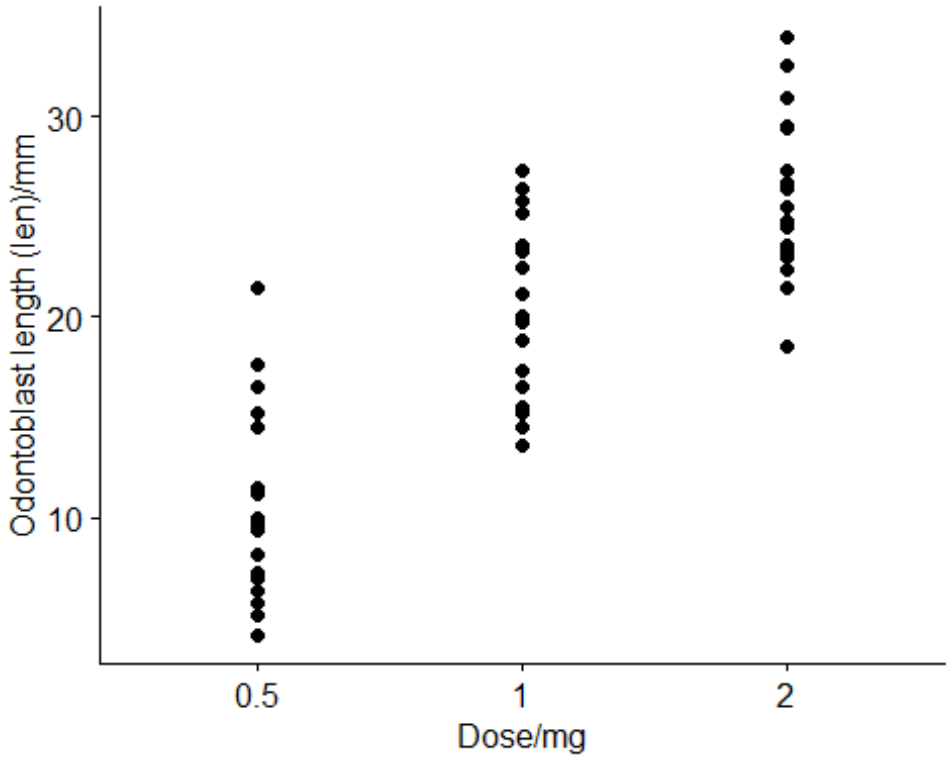



### Examining the correlation between VitC dosage and tooth growth.

A close correlation is observed between len (tooth growth) and dosage of vitamin C (dose).This relationship is illustrated by the following plot as well as Pearson correlation coefficient of 0.83, 95 percent confidence intervals: (0.74-0.90)

```
library(ggpubr)
ggscatter(ToothGrowth, x = "dose", y = "len",
          add = "reg.line", conf.int = TRUE,
          cor.coef = TRUE, cor.method = "pearson",
          xlab = "Dose/mg", ylab = "Odontoblast length (len)/mm")
```

```
## `geom_smooth()` using formula 'y ~ x'
```




```
ToothGrowth$dose1<- as.numeric(ToothGrowth$dose)
correl <- cor.test(ToothGrowth$dose1, ToothGrowth$len,
                   method = "pearson")
correl

##
##  Pearson's product-moment correlation
##
## data:  ToothGrowth$dose1 and ToothGrowth$len
## t = 11.509, df = 58, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7358578 0.8977675
## sample estimates:
##       cor
## 0.8339558
```

## Conclusions

On the face of it there is not an statistically significant difference in tooth growth based on the method of delivery of Vitamin C (supp). This may be a type II error due to small sample size.

It appears that by combining the supp and dose as classifiers it is possible to see that at lower doses of vitamin C (0.5 mg and 1 mg) the method of delivery may be an important variable in tooth growth or len value, but not at the highest dose of 2mgs.

A close correlation is observed between the dosage of vitamin C (as a continuous variable) and tooth growth as examined by odontoblast length (len).

This study is limited by the small sample size and the limited amount of data available. There is also no explanation of how data was collected, so the results must be interpreted with caution.