Notebook for Statistical Inference Course Project 02

Samuel B Baco

```
library(tidyverse)
```

PART 02: BASIC INFERENTIAL DATA ANALYSIS

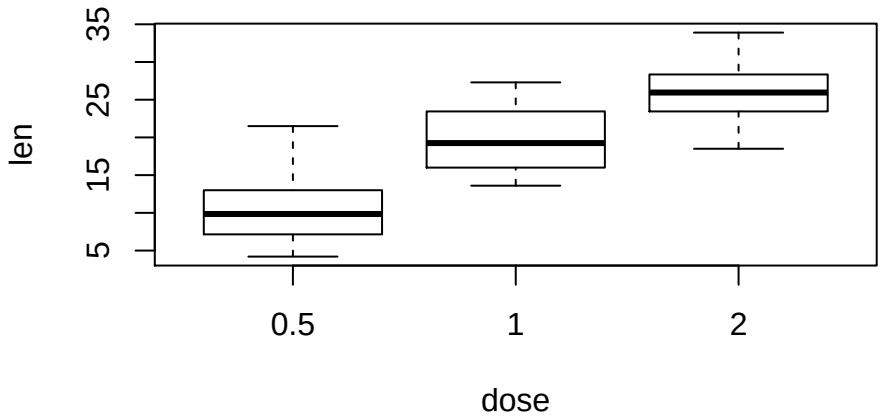
This part consist on analysing the ToothGrowth dataset from *datasets* package. Below there is a summary of the dataset.

```
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
```

It is possible to see that that are 3 possible values from **dose** variable: 0.5, 1, 2.

```
boxplot(len ~ as.factor(dose), data = ToothGrowth, xlab = "dose")
```



There seems to be a considerable difference on variable **len** as **dose** increases. To make sure, it is important to construct the confidence intervals (with 95%) and make sure the extremes does not superimpose themselves.

```
ToothGrowth %>%
        group_by(dose) %>%
        summarise(LCIL = mean(len) - qt(0.975, df = (length(len)-1) * sd(len)/sqrt((length(len)))),
                  UCIL = mean(len) + qt(0.975, df = (length(len)-1)) * sd(len)/sqrt((length(len))))
```

```
## # A tibble: 3 x 3
##    dose  LCIL  UCIL
##   <dbl> <dbl> <dbl>
## 1   0.5  8.51  12.7
```

```
## 2   1   17.6   21.8
## 3   2   24.0   27.9
```

As no UCIL (Upper confidence interval limit) is bigger than LCIL (Lower confidence interval limit), it is possible to say that all diferente values of **dose** procuce statiscal different means for **len**, considering a alpha of 5%.