title: " Basic Inferential Data Analysis Instructions"

author: "dinh quy kien"

date: "May 09, 2023"

output:

word_document: default

pdf_document: default

html_document: default

Overview The project aim is to analyze the ToothGrowth data in the R datasets package.

Load the necessary packages

```
library(ggplot2)
library(tinytex)
library(datasets)
```

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

```
data(ToothGrowth)
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: num  0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 0.5 ...
```

```
head(ToothGrowth, 4)
```

```
##    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
```

```
tail(ToothGrowth, 4)
```

```
##     len supp dose
## 57 26.4   OJ    2
## 58 27.3   OJ    2
## 59 29.4   OJ    2
## 60 23.0   OJ    2
```

Summary of the data

```
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
```

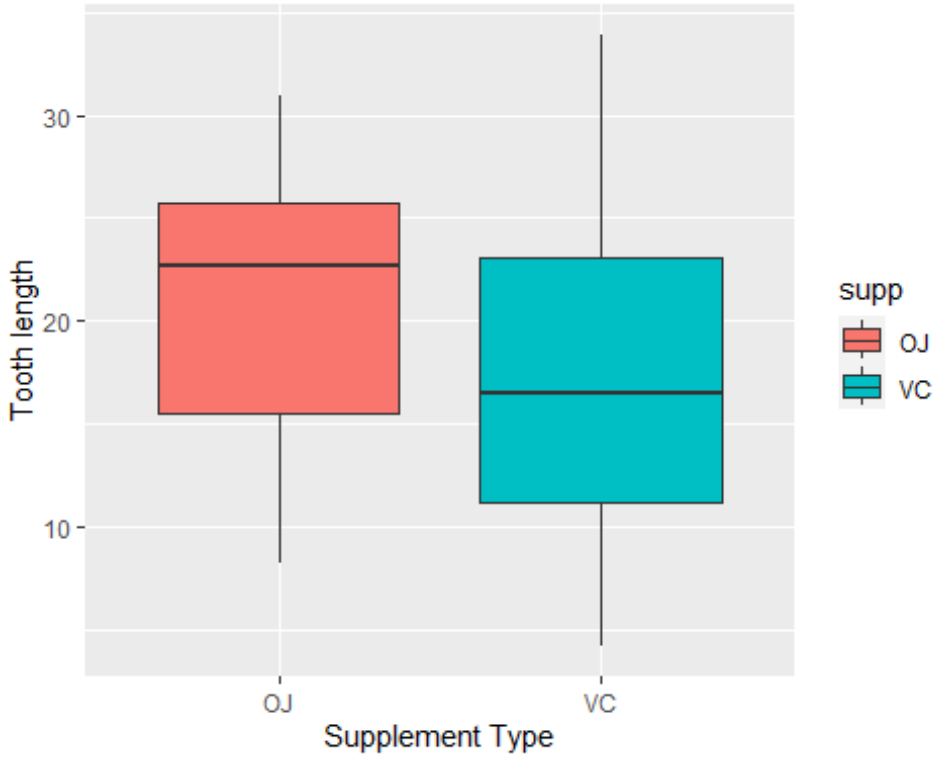
2.Basic summary of the data

```
# Calculatiing the mean of len based on the supplement methods
Supplement_mean = split(ToothGrowth$len, ToothGrowth$supp)
sapply(Supplement_mean, mean)
```

```
##       OJ       VC
## 20.66333 16.96333
```

Graph

```
ggplot(aes(x=supp, y=len), data=ToothGrowth) + geom_boxplot(aes(fill=supp))+
        xlab("Supplement Type") +ylab("Tooth length")
```



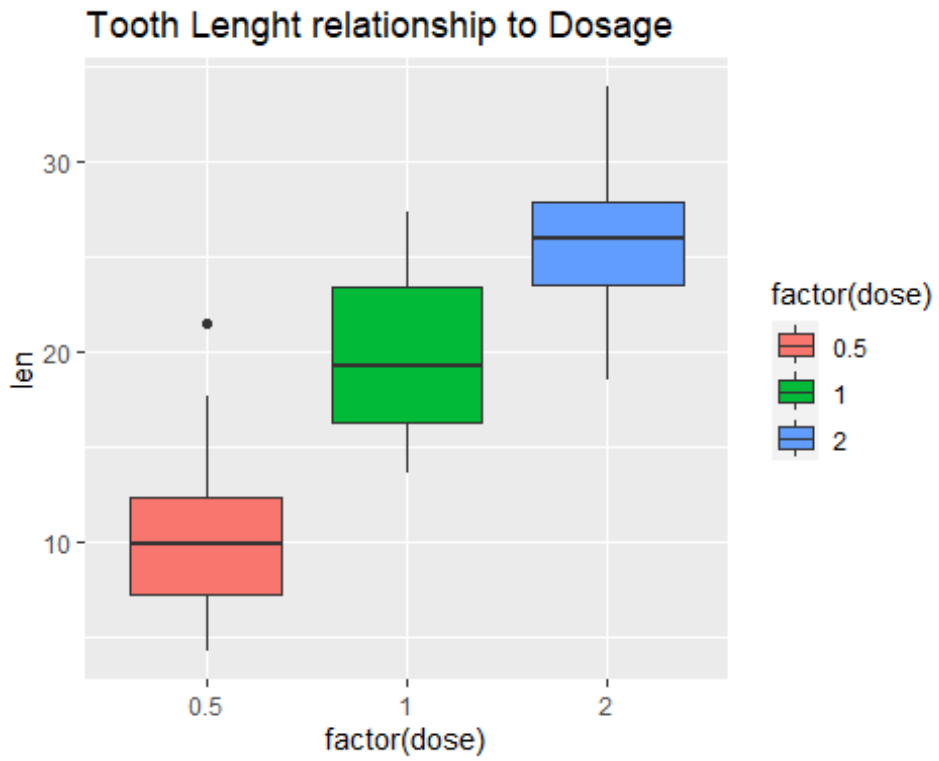
3. Using confidence intervals to compare growth of tooth by supplement dose

```
unique(ToothGrowth$dose)
```

```
## [1] 0.5 1.0 2.0
```

There are 3 dose groups: 0.5, 1, and 2 Graph shows relationship between Tooth length to Dose

```
g <- ggplot(aes(x = factor(dose), y = len), data = ToothGrowth) +
    geom_boxplot(aes(fill = factor(dose)))
g <- g + labs(title="Tooth Lenght relationship to Dosage")
print(g)
```



T-test for dose 0.5 mg:

```
t.test(len ~ supp, ToothGrowth[ToothGrowth$dose == .5, ])

##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 3.1697, df = 14.969, p-value = 0.006359
## alternative hypothesis: true difference in means between group OJ and grou
p VC is not equal to 0
## 95 percent confidence interval:
##  1.719057 8.780943
## sample estimates:
## mean in group OJ mean in group VC
##            13.23             7.98
```

T-test for dose 1 mg:

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

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = 4.0328, df = 15.358, p-value = 0.001038
## alternative hypothesis: true difference in means between group OJ and grou
p VC is not equal to 0
## 95 percent confidence interval:
##  2.802148 9.057852
## sample estimates:
## mean in group OJ mean in group VC
##            22.70            16.77
```

T-test for dose 2 mg:

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

```
##
##  Welch Two Sample t-test
##
## data:  len by supp
## t = -0.046136, df = 14.04, p-value = 0.9639
## alternative hypothesis: true difference in means between group OJ and grou
p VC is not equal to 0
## 95 percent confidence interval:
##  -3.79807  3.63807
## sample estimates:
## mean in group OJ mean in group VC
##            26.06            26.14
```

4.Conclusion:

For all three dosages, the p-value of this test is is less than 0.5, an evidence that we can reject the null hypothesis. We can infer that supplement type has no effect on tooth growth, and increasing the dose level leads to increased tooth growth.