BASIC ANOVA TEST

***Question***

Does the type of treatment significantly affect the weight of plants?

***Hypothesis***

Null Hypothesis : The type of treatment does not have a significant effect on the weight of plants.  
Alternative Hypothesis : The type of treatment has a significant effect on the weight of plants.

# Step 1: Load the Necessary Libraries

# Data Wrangling libraries   
  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.1 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.4 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(dplyr)

# Step 2: Load and Explore the Datasete

# Load the dataset  
PlantGrowth

## weight group  
## 1 4.17 ctrl  
## 2 5.58 ctrl  
## 3 5.18 ctrl  
## 4 6.11 ctrl  
## 5 4.50 ctrl  
## 6 4.61 ctrl  
## 7 5.17 ctrl  
## 8 4.53 ctrl  
## 9 5.33 ctrl  
## 10 5.14 ctrl  
## 11 4.81 trt1  
## 12 4.17 trt1  
## 13 4.41 trt1  
## 14 3.59 trt1  
## 15 5.87 trt1  
## 16 3.83 trt1  
## 17 6.03 trt1  
## 18 4.89 trt1  
## 19 4.32 trt1  
## 20 4.69 trt1  
## 21 6.31 trt2  
## 22 5.12 trt2  
## 23 5.54 trt2  
## 24 5.50 trt2  
## 25 5.37 trt2  
## 26 5.29 trt2  
## 27 4.92 trt2  
## 28 6.15 trt2  
## 29 5.80 trt2  
## 30 5.26 trt2

str(PlantGrowth) #This is an R inbuilt dataset

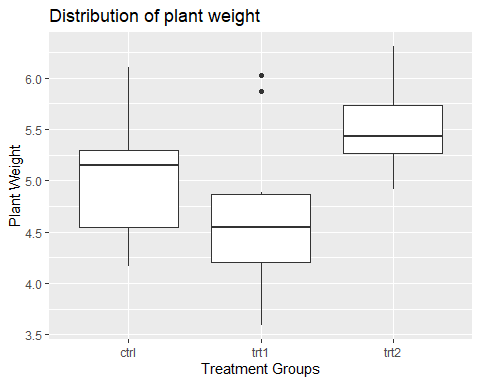
## 'data.frame': 30 obs. of 2 variables:  
## $ weight: num 4.17 5.58 5.18 6.11 4.5 4.61 5.17 4.53 5.33 5.14 ...  
## $ group : Factor w/ 3 levels "ctrl","trt1",..: 1 1 1 1 1 1 1 1 1 1 ...

The PlantGrowth dataset has 2 variables (weight and group). weight is a numerical variable and group is a factor variable with 3 levels/categories.  
There are 30 observations in this dataset.

# Step 3: Visualize the Distribution

The aim is to test if the average weights of plants in the 3 treatment groups are equal or not.  
So, the x-variable is ‘group’, the categorical variable, and the y-variable is ‘weight’, the numerical variable.  
Since there is a categorical variable and a numerical variable, visualize the distribution of these variables using a boxplot.

ggplot(PlantGrowth, aes(x = group, y = weight))+  
 geom\_boxplot()+  
 labs(title = "Distribution of plant weight ",  
 x = "Treatment Groups",  
 y = "Plant Weight")



Plants treated with Treatment2 has a higher mean weight compared to plants treated with Treatment1 and the Control group.

# Step 4: Run the ANOVA Test

summary(aov(PlantGrowth$weight~PlantGrowth$group))

## Df Sum Sq Mean Sq F value Pr(>F)   
## PlantGrowth$group 2 3.766 1.8832 4.846 0.0159 \*  
## Residuals 27 10.492 0.3886   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

qf(0.95, df1 = 2, df2 = 27) # 0.95 = 1 - alpha

## [1] 3.354131

qf(0.99, df1 = 2, df2 = 27) # 0.95 = 1 - alpha

## [1] 5.488118

# Step 5: ANOVA Test; Interpretation

Suppose , fail to reject the null hypothesis as the p-value from the ANOVA test is 0.0159 The F (2,27) statistic, 4.846 < 5.488118. There isn’t sufficient evidence to conclude that the mean plant weights from the 3 treatment groups are different at 1% significance level.

Suppose then reject the null hypothesis as the p-value from the ANOVA test is 0.0159 is < 0.05 The F (2,27) statistic, 4.846 > 3.354131 There is evidence to conclude that the mean plant weights from the 3 treatment groups are different at 5% significance level.