University of Salford, MSc Data Science

Module: Applied Statistics and Data Visualisation

Date: Trimester 1, 2025-2026

Session: Workshop Week 7

Topic: One-Way and Two-Way ANOVA

Tools: RStudio

Instructors: Dr Kaveh Kiani, Nathan Topping

Objectives:

After completing this workshop, you will be able to:

- Perform One-Way Analysis of Variance Test (One-Way ANOVA)
- Perform Two-Way Analysis of Variance Test (Two-Way ANOVA)

Table of Contents

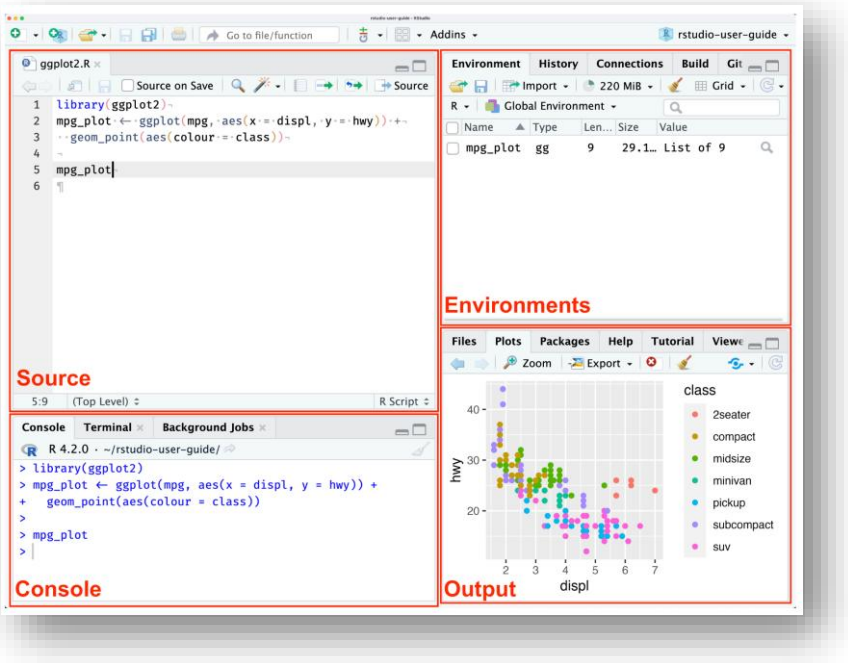
For your own benefit, we strongly encourage you to type out the R scripts yourself in RStudio, rather than copying and pasting them directly from the workshop notes. Typing the code helps reinforce key concepts, improves your understanding of the syntax, and strengthens your coding skills. While copying and pasting may seem faster, actively engaging with the code will lead to a deeper understanding and make you more proficient in R programming over time.

Take your time, experiment, and learn by doing!

Part 1: Working in RStudio

Take sometime later to read about the RStudio pane layout, but for now, just focus on knowing that there are **4 main panes**, **Source, Environments, Console and Output**

https://docs.posit.co/ide/user/ide/guide/ui/ui-panes.html

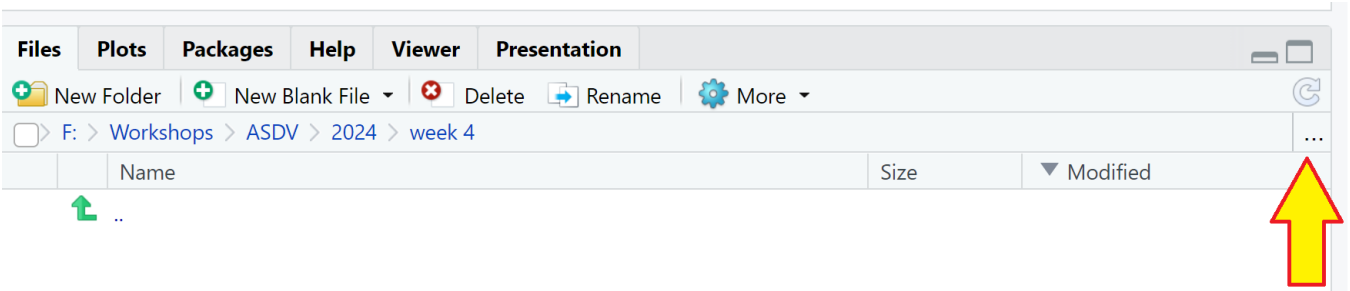


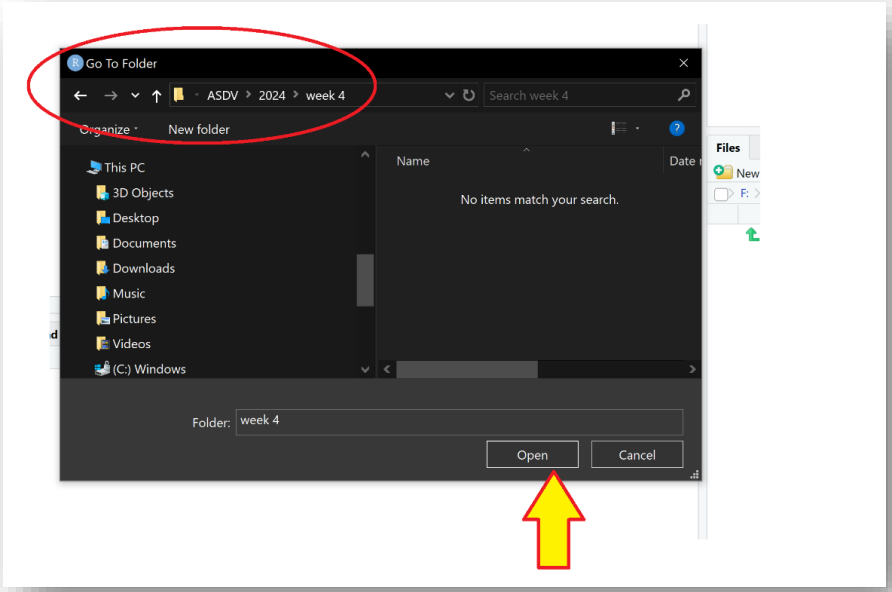
You may work through the workshops as follows:

1. Create a folder in the university PCs F drive or your personal OneDrive or your laptop to save datasets and R scripts. A good practice is to create a folder named **ASDV** (or similar) for all the R workshops that you will have throughout the trimester and then create a subfolder for each week. So, you can have **ASDV\Week7** for this week. Each week, you should **download the data for the week's workshop from Blackboard** and save it in this folder, as well as saving your scripts and the workshop materials in this folder too.

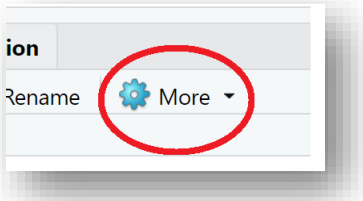
2. Start the RStudio

3. In the **Output Pane** (on the bottom right side of the RStudio window) click on the ellipse sign (…) to navigate to and select the folder where you saved your data file (ASDV\week7) and open the folder.

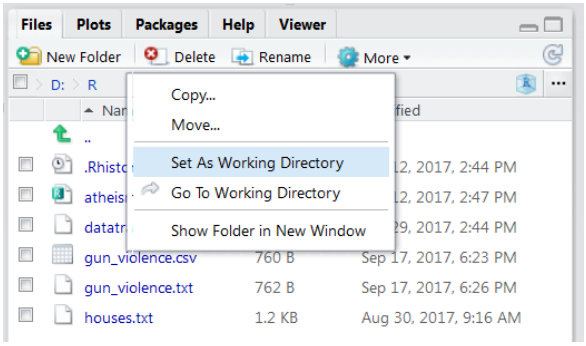




4. Then select the **More** tab



And then **Set As Working Directory**



5. From the **Source Pane** open a new R script window:

File → New File → R script

Part 2: Tests, at a Glance and Packages

Test	Predictor variable, response variable and research question(s)	Assumptions
One-Way ANOVA	Dependent variable, numerical variable Independent variable, one categorical variable with more than two levels	6 assumptions
Two-Way ANOVA (balanced data)	Dependent variable, numerical variable Independent variable, more than one balanced categorical variable with more than two levels	6 assumptions
Two-Way ANOVA (unbalanced data)	Dependent variable, numerical variable Independent variable, more than one unbalanced categorical variable with more than two levels	6 assumptions
Kruskal-Wallis One-Way ANOVA	Dependent variable, numerical variable Independent variable, one categorical variable with more than two levels	No assumptions

Packages

Packages are collections of R functions, data, and compiled code in a well-defined format. A package is either a **Base Package** or a **User Contributed Package**.

There are a set of **Base (or Standard) Packages** which are considered part of the R source code and automatically available as part of the R installation and we can directly use functions from the standard base packages.

However, the more we work with R, we will come to realize that there are many users contributed packages that have been created to add specific functionality. There are 10,000+ **User Contributed Packages** and growing. To use **User Contributed Packages** will require installation (Many packages can be installed from the **CRAN** repositories).

We will be using following packages throughout the workshop today.

1. datarium:

Is a data bank for statistical analysis and visualization.

2. RVAideMemoire:

Contains diverse more or less complicated functions, written to simplify user's life: simplifications of existing functions, basic but not implemented tests, easy-to-use tools, bridges between functions of different packages.

3. Car:

In R, the car package (Companion to Applied Regression) provides functions for regression analysis and diagnostics, particularly useful for linear and generalized linear models. The package, developed to accompany the textbook Companion to Applied Regression by John Fox and Sanford Weisberg.

4. carData:

The carData package in R provides a collection of datasets often used in applied regression and analysis of variance, making it valuable for teaching and learning statistical methods. It includes data from real-world studies across diverse fields, allowing users to explore meaningful examples. The package is frequently paired with the car package for enhanced analysis, visualization, and model diagnostics

5. stats:

This package contains functions for statistical calculations and random number generation. This package has been written by R Core Team and contributors worldwide.

6. graphics:

This package contains functions for 'base' graphics. Base graphics are traditional S-like graphics, as opposed to the more recent grid graphics.

Packages 1 to 4 are **User Contributed Packages**, and we must install and call them before using. Package 5 and 6, are a **Base Package**, and they are ready to use.

First, please install and load all packages at once (it is possible your machine has some of the packages now and it will ask for re-installation and you can refuse it).

```
install.packages("datarium")
install.packages("RVAideMemoire")
install.packages("car")
install.packages("carData")



library(datarium)
library(RVAideMemoire)
library(car)
library(carData)
```

Part 3: Analysis of Variance (ANOVA)

1. What is ANOVA

ANOVA is used when we want to experiment effect of two or more levels of **one or two (or even three) independent variables** on **a dependent variable**.

Example 1) Suppose that it is your objective to compare the fuel consumption recorded for three different makes of cars, A-cars, B-cars, and C-cars.

Dependent variable: Fuel consumption per mile

One Independent variable: Types of cars (3 levels/categories)

Example 2) Suppose that it is your objective to compare the fuel consumption recorded for three different makes of cars, A-cars, B-cars, and C-cars and 2 fuel types, Petrol, and Diesel.

Dependent variable: Fuel consumption per mile

1st Independent variable: Types of cars (3 levels/categories)

2nd Independent variable: Types of fuel (2 levels/categories)

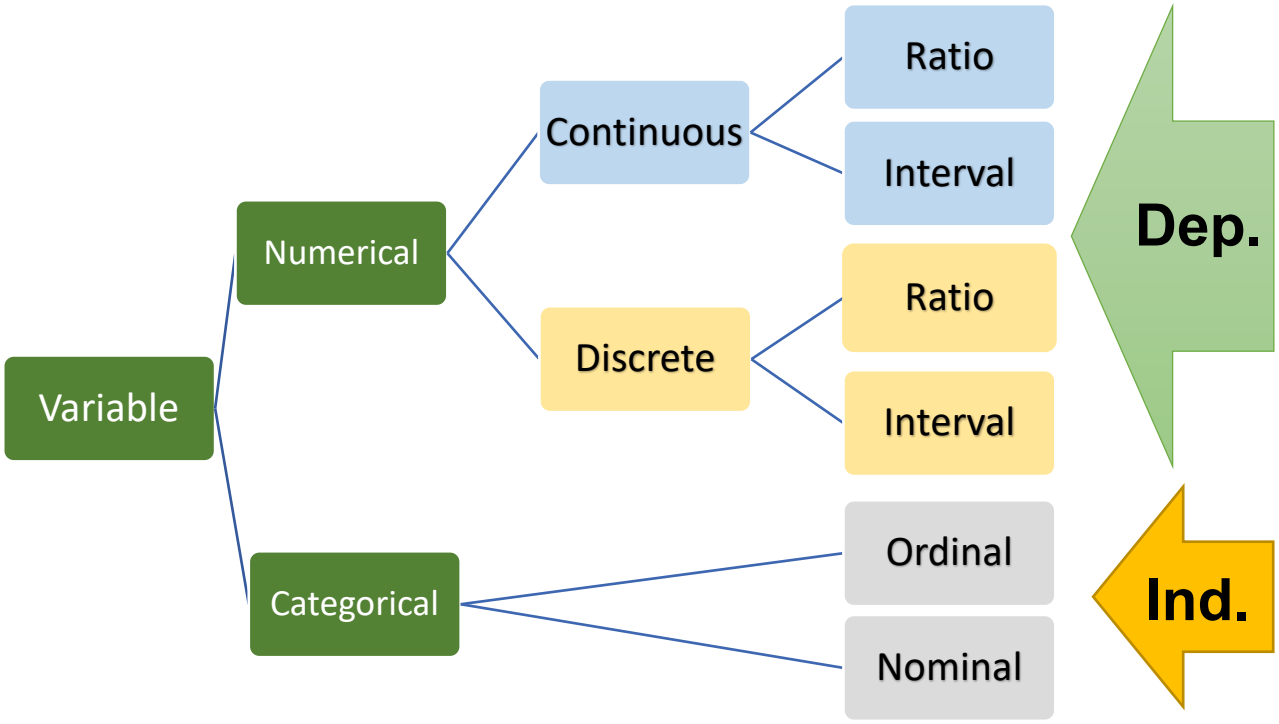
2. Data Type for ANOVA

In ANOVA, **the dependent variable** must be a **Numerical** (ratio or interval)**. The independent variable(s)** must be **categorical** (nominal or ordinal). Let's review the example 1 again.

Example 1) Suppose that it is your objective to compare the fuel consumption recorded for three different makes of cars, A-cars, B-cars, and C-cars.

Dependent variable: Fuel consumption per mile (numerical variable)

Independent variable: Types of cars (categorical variable)



3. ANOVA Assumptions

Like the t-test, ANOVA is also a parametric test and has some assumptions. To perform ANOVA, the data needs to "pass" six assumptions to give us a valid result. However, even when our data fails certain assumptions, there is often a solution to overcome this failure.

Assumption 1: Dependent variable should be continuous.

Assumption 2: Independent variables should be categorical with two or more categories.

Assumption 3: Observations should be independent, which means that there is no relationship between the observations in each group or between the groups themselves.

Assumption 4: There should be no significant outliers.

Assumption 5: Dependent variable should be approximately normally distributed for each category of the independent variable.

Assumption 6: Variances of the dependent variable within each category should be homogeneous.

If your data fails to the 6th assumption, you should run a **Welch’s ANOVA** instead of the ordinary ANOVA. Welch's ANOVA can be used for both one-way and two-way ANOVA. It is a robust version of ANOVA that is less sensitive to violations of the assumption of equal variances.

Part 4: One-Way ANOVA

The Framework for One-Way Analysis of Variance

Suppose that we have independent random samples of $n_1, n_2, \ldots, n_K$ observations from K populations. If the population means are denoted $\mu_1, \mu_2, \ldots, \mu_K$, the one-way analysis of variance framework is designed to test the null hypothesis:

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_K$$

$$H_1: \mu_i \neq \mu_j \text{ For at least one pair } \mu_i, \mu_j$$

To keep the one-way ANOVA definition simple and free of extra technical details, you can memorise example 1 as a typical case of a One-Way ANOVA.

Recall example 1**,** Suppose that it is your objective to compare the fuel consumption recorded for three different makes of cars, A-cars, B-cars, and C-cars.

Dependent variable: Fuel consumption per mile

One Independent variable: Types of cars (3 categories)

One-Way ANOVA for Heart Attack Data:

To demonstrate the Analytical steps of ANOVA we will use ***Heart Attack*** dataset from the **datarium package**.

Data consists of measures of cholesterol concentration in 72 participants treated with three different drugs. The aim is to examine the potential of new class of drugs in lowering the cholesterol concentration and consequently reducing heart attack. The participants include 36 males and 36 females. Males and females were further (equally) subdivided into whether they were at low or high risk of heart attack. This data set is suitable for one way and two ANOVA test.

It contains the following independent (categorical) variables:

- **Gender**, which has two categories: "male" and "female",
- **Risk** which has two levels: "low" and "high",
- **Drug**, which has three categories: "A", "B" and "C".

Dependent Variable (continuous) is the **cholesterol level** of the patients.

To run a One-way ANOVA lets choose **one** independent variable (ideally to have more than 2 levels).

Objective: To examine the potential of new class of drugs in lowering the cholesterol concentration and consequently reducing heart attack.

Independent Variable: **Drug**, which has three categories: "A", "B" and "C".

Dependent Variable: **Cholesterol level** of the patients.

In this case, the Null Hypothesis is that the different drugs have similar effects and none of them has superiority in lowering the cholesterol level.

$$H_0: \mu_A = \mu_B = \mu_C$$

But the pharmaceutical company claims its new drug (drug C) is better than the other two drugs on the market (drugs A and B). So, first, we want to see if any difference between drugs is there or not. And if we see any significant differences then we can dig deeper to find the differences. Therefore, the Alternative Hypothesis can be formulated as:

$$H_1: \mu_A \neq \mu_B \neq \mu_C$$

First let's see few rows of the dataset (from the datarium package):

```
head(as.data.frame(heartattack))
```

```
##    gender risk drug cholesterol id
## 1   male  low    A    5.238730  1
## 2   male  low    A    5.075693  2
## 3   male  low    A    4.678653  3
## 4   male  low    A    5.361076  4
## 5   male  low    A    4.957381  5
## 6   male  low    A    4.828897  6
```

The next step is checking the assumptions:

Assumptions are:

A1: Dependent variable should be continuous.

A2: Independent variables should be categorical with two or more categories.

A3: Observations should be independent, which means that there is no relationship between the observations in each group or between the groups themselves.

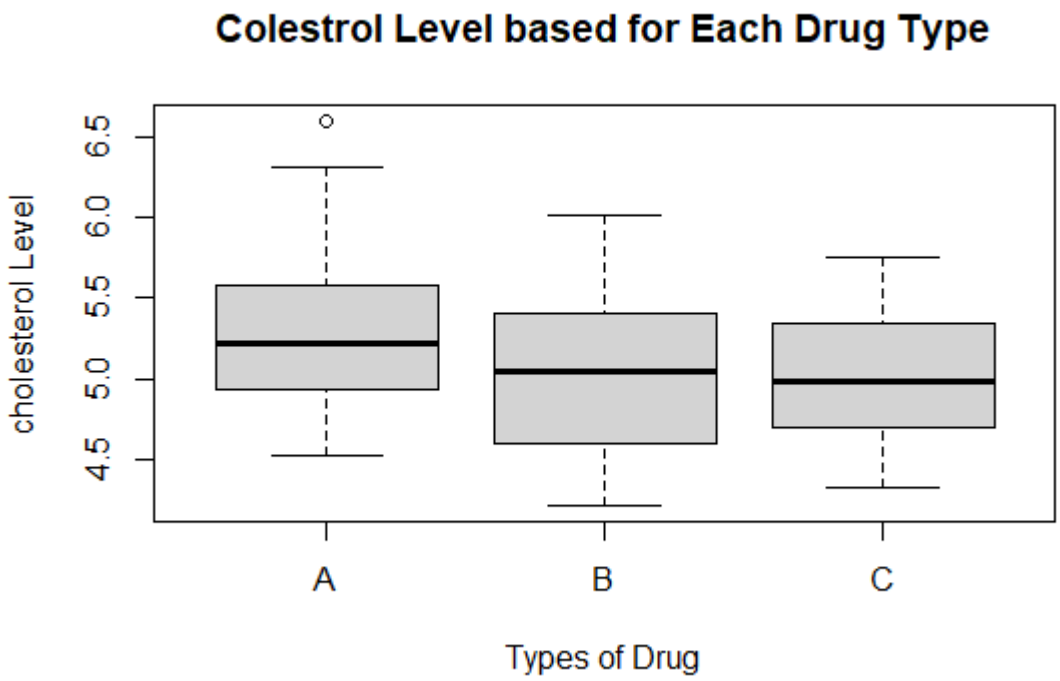
A4: There should be no significant outliers.

A5: Dependent variable should be approximately normally distributed for each category of the independent variable.

A6: Variances of the dependent variable within each category should be homogeneous.

A1 to A3 are correct by design. To check A4 you should draw the box plot to see the potential outliers.

```
boxplot(cholesterol ~ drug , data=heartattack, names=c("A", "B", "C"),
        xlab="Types of Drug", ylab="cholesterol Level",
        main="Colestrol Level for Each Drug Type")
```



There is one outlier in the group A, but it seems not an influential outlier and we can confirm the A5 assumption.

Remark1: If you see an influential point and you suspect it is an outlier, don't simply remove it. First try to explore whether the value is logical and acceptable and if you can't confirm the value of the suspected data point, only then remove it.

The A5 assumption is to check the normality. We want to implement **Shapiro-Wilk** test that you have learnt before. The only difference is we must check the cholesterol variable normality in all 3 levels of the drug.

Recall the Shapiro-Wilk test hypotheses:

H_0: the variable follows a normal distribution

H_1: the variable does NOT follow a normal distribution

We need to use a new package called **`RVAideMemoire`**

```
byf.shapiro(cholesterol ~ drug, data=heartattack)


## Shapiro-Wilk normality tests
## data:  cholesterol by drug
##       W   p-value
```

```
## A 0.9388  0.1538
## B 0.9741  0.7675
## C 0.9653  0.5537
```

All p-values are greater than 0.05 and we shouldn't reject the null hypothesis. We can confirm that cholesterol level has normally distributed through each level of the Drug.

The last assumption, A6, is to check homogeneity of variances in three levels of the Drug variable. We use Bartlett test from the **default package, stats**.

```
bartlett.test(cholesterol ~ drug, data=heartattack)


## Bartlett test of homogeneity of variances
## data:  cholesterol by drug
## Bartlett's K-squared = 3.1208, df = 2, p-value = 0.21
```

From the output, the p-value is 0.21 and it is not less than the significance level of 0.05. This means that there is no evidence to suggest that the cholesterol level variances are significantly different for the three drug levels.

Remark 2: After (or before) running a test if you want to see which package has been linked to the test then run the following command.

```
find("bartlett.test")
[1] "package:stats"
```

Now we have checked all 6 assumptions and the data has passed this stage. The final step is to conduct the ANOVA test to see is there any difference between the 3 drugs or not. We will use the default package **stats**.

```
oneway.test(cholesterol ~ drug,data=heartattack, var.equal = TRUE)
## One-way analysis of means
## data:  cholesterol and drug
## F = 2.6305, num df = 2, denom df = 69, p-value = 0.07926
```

The p-value is greater than 0.05 and it indicates that we cannot reject the null hypothesis H_0: $\mu_A = \mu_B = \mu_C$, in other words, we couldn't confirm that there is a significant difference between the three drugs.

Remark 3: But what if null hypothesis was rejected? Then there would be two possible

decisions.

- To stop the test and conclude that 3 drugs are not similar.
- To run **post-hoc** tests to find out which drug(s) is(are) different from the others.

We recommend searching and reading about the post-hoc test for ANOVA analysis in R. A good online source to start is:

https://statsandr.com/blog/anova-in-r/#anova-in-r

It is important to keep in mind that doing a post-hoc test could be a possible action that you have to take in your final assignment.

Remark 4: At the following command that you ran:

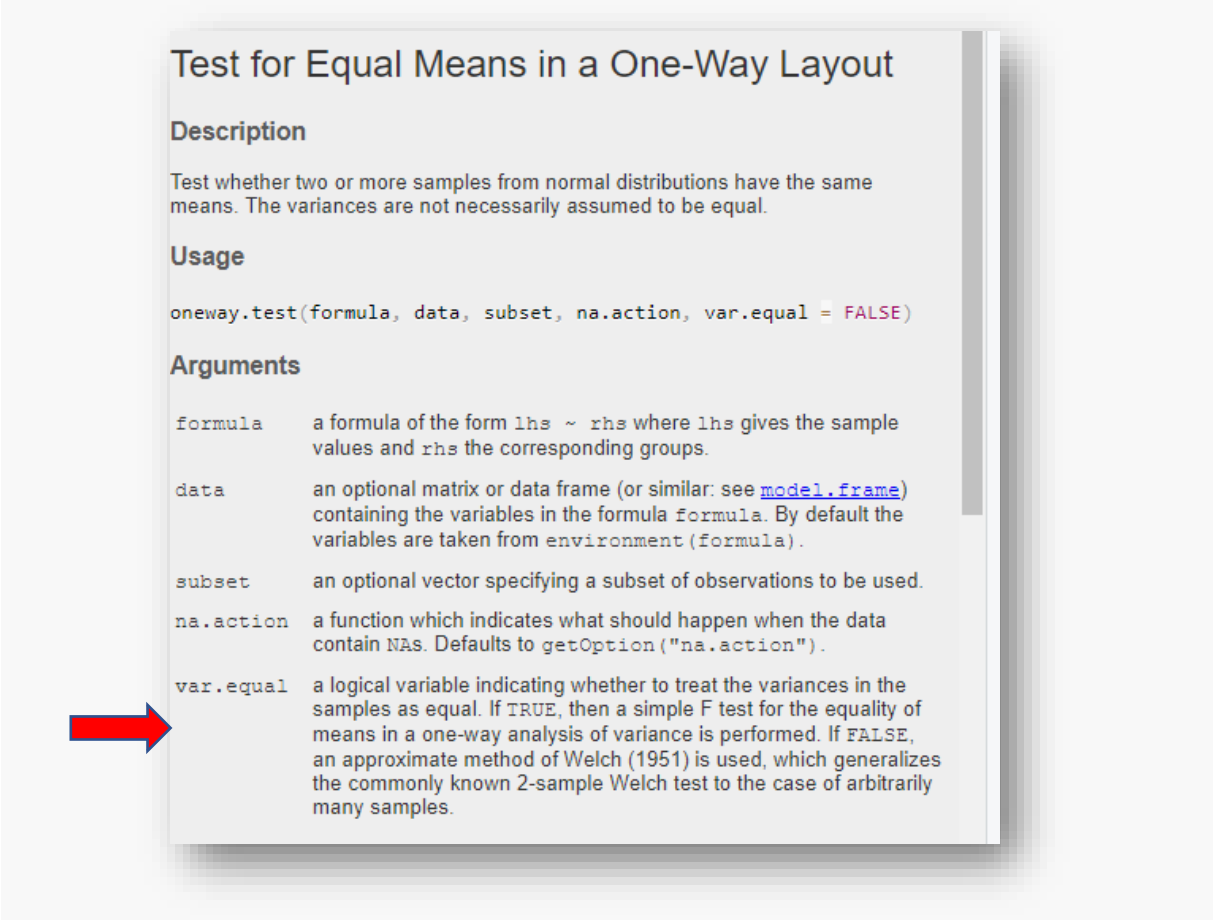
```
oneway.test(cholesterol ~ drug,data=heartattack, var.equal = TRUE)
```

Var.equal = TRUE is for the assumption that you checked, and you saw that the variances are equal.

But if A6 assumption fails then you can put `var.equal = FALSE` and Welch's ANOVA will be run.

To know more about oneway.test arguments including var.equal run help:

```
?oneway.test
```



Test for Equal Means in a One-Way Layout

Description

Test whether two or more samples from normal distributions have the same means. The variances are not necessarily assumed to be equal.

Usage

```
oneway.test(formula, data, subset, na.action, var.equal = FALSE)
```

Arguments

formula	a formula of the form `lhs ~ rhs` where `lhs` gives the sample values and `rhs` the corresponding groups.
data	an optional matrix or data frame (or similar: see `model.frame`) containing the variables in the formula `formula`. By default the variables are taken from `environment(formula)`.
subset	an optional vector specifying a subset of observations to be used.
na.action	a function which indicates what should happen when the data contain `NA`s. Defaults to `getOption("na.action")`.
var.equal	a logical variable indicating whether to treat the variances in the samples as equal. If `TRUE`, then a simple F test for the equality of means in a one-way analysis of variance is performed. If `FALSE`, an approximate method of Welch (1951) is used, which generalizes the commonly known 2-sample Welch test to the case of arbitrarily many samples.

Remark 5: If you remember the **F distribution** from the previous workshops, we have mentioned that F is a building block distribution for some of the hypothesis testing algorithms. Here you found out that the F statistics have been used for ANOVA analysis.

```
## F = 2.6305, num df = 2, denom df = 69, p-value = 0.07926
```

Part 5: Two-Way ANOVA (for balanced data)

Like our approach in one-way ANOVA, example 2 is a distinct example of Two-Way ANOVA

Recall example 2, suppose that it is your objective to compare the fuel consumption recorded for three different makes of cars, A-cars, B-cars, and C-cars and 2 fuel types, Petrol, and Diesel.

Dependent variable: Fuel consumption per mile

1st Independent variable: Types of cars (3 categories)

2nd Independent variable: Types of fuel (2 categories)

Sometimes second independent variable is called **Block Variable** because it divides the data (including the first variable) into blocks for example:

1st Block is:	**2nd Block is:**
Fuel Type 1	**Fuel Type 2**
Car A	Car A
Car B	Car B
Car C	Car C

Two-Way ANOVA for Heart Attack Data:

Here again we like to examine the ***Heart Attack*** data but will add a second factor, **Gender**.

The objective of the two-way ANOVA test is to see how levels of Drugs and levels of Gender together can make any significant effect.

From One-way ANOVA you find out that there is no significant difference between the 3 drugs to reduce the cholesterol level but what if drugs have a significant difference for females or males?

We can define all possible effects on the cholesterol level variation as:

$$Variation\ in\ \boldsymbol{Cholesterol} = \beta_1 \boldsymbol{Drug} + \beta_2 \boldsymbol{Gender} + \beta_3(\boldsymbol{Drug} \times \boldsymbol{Gender}) + Error$$

There are 3 effects $\begin{cases} Drug\ effect \\ Gender\ effect \\ \textbf{\textit{Interaction}}\ between\ drug\ and\ gender \end{cases}$

3 null hypotheses based on three possible effects can be formulated as:

1. There is no difference in the means of drug levels (drugs' means are equal)
2. There is no difference in the means of gender levels (genders' means are equal)
3. There is no interaction between drug and gender

The alternative hypothesis for cases 1 and 2 is: the means are not equal.

The alternative hypothesis for case 3 is: there is an interaction between drug and gender.

But what is the meaning of **Balanced Data**?

> **Data is balanced when number of subjects are equal inside each group.**

Let's examine the Heart attack data.

```
summary(heartattack)
##    gender      risk    drug    cholesterol          id
## male  :36   high:36   A:24   Min.   :4.210   Min.   : 1.00
## female:36   low :36   B:24   1st Qu.:4.812   1st Qu.:18.75
##                       C:24   Median :5.146   Median :36.50
##                              Mean   :5.129   Mean   :36.50
##                              3rd Qu.:5.382   3rd Qu.:54.25
##                              Max.   :6.605   Max.   :72.00
```

We can confirm that the data is balanced (36 for each gender, 36 for each risk, 24 for each drug).

Before running the test, we should check the 6 assumptions.

> **Now it is your turn to test the assumptions. Return to the previous sections and review the assumptions and procedures. You must check the assumptions of Normality and Homogeneity of Variances (Assumptions A5 and A6) for each cell in the two-way table—that is, for all combinations of the two factor levels. In the heart attack data, with three levels of drug and two levels of gender, we have 3 * 2 = 6**

combinations for the drug and gender factors.

```
byf.shapiro(cholesterol ~ drug*gender, data=heartattack)

bartlett.test(cholesterol ~ interaction(drug,gender), data=heartattack)
```

If the data passed the assumptions' checking stage, then you should use **stats package** to get the job done. Use the following code to test both drug and gender and their interaction effects (pay attention to the * sing in the code).

```
result_aov1 <- aov(cholesterol ~ drug * gender, data=heartattack)

summary(result_aov1)


##             Df Sum Sq Mean Sq  F value Pr(>F)
## drug         2  1.235  0.6177   2.856 0.0646 .
## gender       1  1.367  1.3672   6.322 0.0144 *
## drug:gender  2  0.564  0.2818   1.303 0.2786
## Residuals   66 14.273  0.2163
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- We don't reject the null hypothesis for drug at 0.05 significance level, but we can reject it at 0.1 significance level (look at the black dot beside the 0.0646 and then look at the underneath of the result '.' 0.1)
- We reject the null hypothesis for gender at 0.05 level.
- We don't reject the null hypothesis for the interaction effect at 0.05 level.

Therefore, we can say after bringing the gender into the test we can see there is a difference between genders but there is no drug effect (all 3 drugs are still the same) and there is not any interaction between drug and gender. Possibly we can conclude that gender's effect on the cholesterol level is only because of the physiological effects, not drugs.

Here we can run a post-hoc test to see which levels of the gender have made these differences (however in this dataset the gender variable has only two levels and which level has made this difference is meaningless because both can make a difference in opposite directions. But still, we can explore which gender has a lower/upper cholesterol level):

Use ***TukyHSD test*** for the post-hoc analysis purpose:

```
TukeyHSD(result_aov1, which = "gender")


#  Tukey multiple comparisons of means
#   95% family-wise confidence level
#Fit: aov(formula = cholesterol ~ drug * gender, data = heartattack)


#$gender
#                   diff        lwr         upr     p adj
# female-male -0.2755964 -0.4944374 -0.05675534 0.0143703
```

95% confidence interval for the (Female mean) – (male mean) is

`(-0.4944374 -0.05675534)` which is always negative. This means cholesterol level of females are lower than males.

Remark 6: if you don't want to test the interaction effect, use the following code:

```
result_aov2 <- aov(cholesterol ~ drug + gender ,data=heartattack)
summary (result_aov2)
```

You must run this test and explore the conclusions that can be drawn from the outcomes.

Part 6: Two-Way ANOVA (for unbalanced data)

Data is **unbalanced** when number of subjects are **unequal** in each group. Let's explore the job satisfaction data (from datarium package). It contains the job satisfaction score organized by gender and education level.

```
summary(jobsatisfaction)
##       id         gender     education_level     score
## 1       : 1   male  :28   school    :19     Min.   : 4.780
## 2       : 1   female:30   college   :19     1st Qu.: 5.800
## 3       : 1               university:20     Median : 6.380
## 4       : 1                                 Mean   : 6.963
## 5       : 1                                 3rd Qu.: 8.515
## 6       : 1                                 Max.   :10.000
## (Other):52
```

The data is unbalanced, see the gender and educational level. Install ***car package*** to run a two-way ANOVA for unbalanced data. There are three types of analysis (Type I, Type II, and Type III) but we recommend using Type III.

```
unb_anova <- aov(score ~ education_level * gender, data = jobsatisfaction)
Anova(unb_anova, type = "III")

## Anova Table (Type III tests)
## Response: score
##                         Sum Sq Df  F value    Pr(>F)
## (Intercept)             265.038  1 876.0876 < 2.2e-16 ***
## education_level          80.128  2 132.4321 < 2.2e-16 ***
## gender                    0.468  1   1.5471  0.219147
## education_level:gender    4.440  2   7.3379  0.001559 **
## Residuals                15.731 52
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Try to interpret the results.

- Is the education_level effect significant?
- Is the gender effect significant?
- Is the interaction effect significant?
- Is there any possible next step?

Remark 7: Hopefully you learnt One-Way and Two-Way ANOVA. You should know that the higher ways are possible, and you can conduct Three-Way ANOVA and above but these tests are rarely used because the results of the tests are complex and difficult to interpret.

Part 7: Non-Parametric Version of the One-Way ANOVA Test

The nonparametric version of the ANOVA is the Kruskal-Wallis test. It is used to compare the medians of three or more independent groups. The Kruskal-Wallis test is a rank-based test, which means that it does not make any assumptions about the distribution of the data.

To perform the Kruskal-Wallis test, you first need to rank all of the data values, regardless of group membership. Then, you calculate a test statistic that is based on the average ranks of the groups. The test statistic is compared to a critical value from a chi-squared distribution to determine if the differences between the groups are statistically significant.

The Kruskal-Wallis test is a powerful tool for comparing the medians of three or more groups, even when the data is not normally distributed. However, it is important to note that the Kruskal-Wallis test is less powerful than the ANOVA test when the data is normally distributed.

Here are some examples of when you might use the Kruskal-Wallis test:

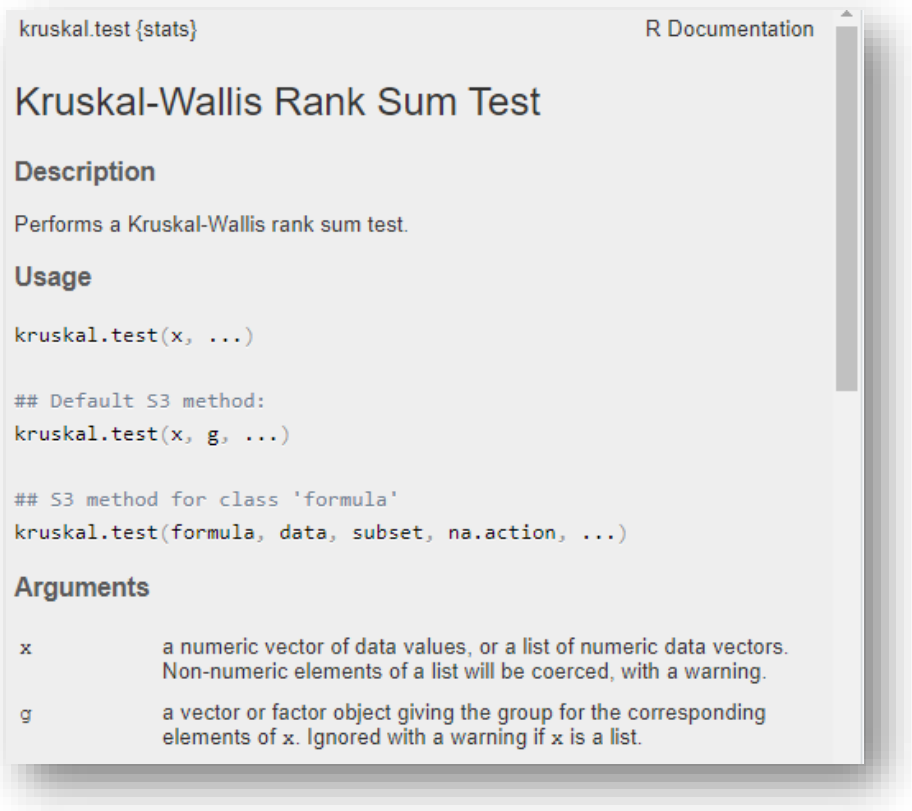
You are comparing the medians of three or more groups of students on a test score, but the test scores are not normally distributed.

You are comparing the medians of three or more groups of patients on a pain scale, but the pain scores are not normally distributed.

You are comparing the medians of three or more groups of animals on a weight measure, but the weight measures are not normally distributed.

We can use kruskal.test() function from the default R Package stats. Read about it:

```
?kruskal.test
```



kruskal.test {stats} R Documentation

Kruskal-Wallis Rank Sum Test

Description

Performs a Kruskal-Wallis rank sum test.

Usage

```
kruskal.test(x, ...)

## Default S3 method:
kruskal.test(x, g, ...)

## S3 method for class 'formula'
kruskal.test(formula, data, subset, na.action, ...)
```

Arguments

x	a numeric vector of data values, or a list of numeric data vectors. Non-numeric elements of a list will be coerced, with a warning.
g	a vector or factor object giving the group for the corresponding elements of x. Ignored with a warning if x is a list.

Let's run the test for heartattack data

```
kruskal.test(cholesterol ~ drug, data=heartattack)

#     Kruskal-Wallis rank sum test

# data:  cholesterol by drug

# Kruskal-Wallis chi-squared = 2.7976, df = 2, p-value = 0.2469
```

You can compare the result with the parametric test that you ran before:

```
oneway.test(cholesterol ~ drug,data=heartattack, var.equal = TRUE)

## One-way analysis of means

## data:  cholesterol and drug

## F = 2.6305, num df = 2, denom df = 69, p-value = 0.07926
```

In both tests we don't reject the null hypothesis.

The Null Hypothesis is that the different drugs have similar effects and none of them has superiority in lowering the cholesterol level.

$$H_0: \mu_A = \mu_B = \mu_C$$

Challenge: Utilise the built-in R data set called PlantGrowth, which includes plant weights gathered under a control and two distinct treatment conditions. Execute the Kruskal-Wallis test to compare the weights of plants across various groups. What conclusions can be drawn from the outcomes?

Here are few initial steps that you can start with:

```
plant <- PlantGrowth
head(plant,30)
summary(plant)
```

References:

1. https://cran.r-project.org/web/packages/datarium/datarium.pdf
2. https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html
3. https://www.statology.org/test-for-normality-in-r/
4. https://cran.microsoft.com/snapshot/2017-081/web/packages/RVAideMemoire/RVAideMemoire.pdf
5. https://www.tidyverse.org/packages/