HANDOUT 3

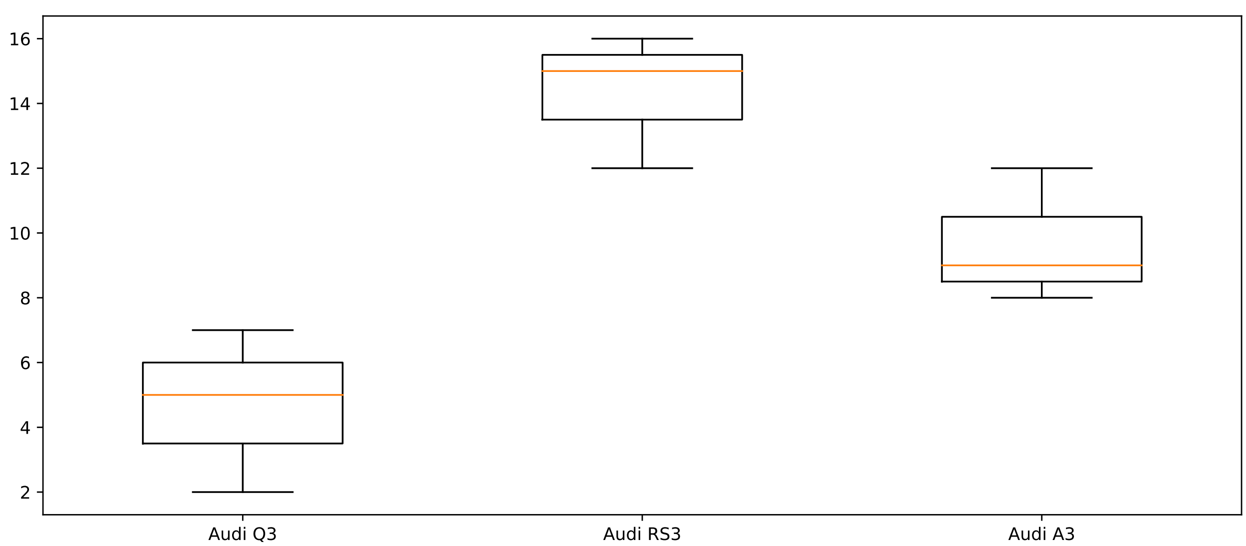
This handout will focus on the **Analysis of Variance** (ANOVA). This technique measures whether the overall variation among the groups/categories of the independent variable is plausible, or is likely to happen. We use ANOVA to identify the components of a total variance which includes the variance of group effects, the interaction effects, and the errors.

1 One-Way ANOVA

- Assumptions:
  - Independent: The samples are independent. Obtaining one sample does not affect obtaining another sample.
  - Normality: The populations where the samples were drawn must be normally distributed.
  - Homogeneous variance: The variances of the populations is equal.
- Roadmap:
  1. State the hypothesis.
  2. Calculate the means of each group.
  3. Construct the ANOVA table.
  4. Calculate the degrees of freedom (DF).
  5. Calculate the sum of squares (SS).
  6. Calculate mean squares (MS).
  7. Calculate the F statistic.
  8. Find the critical F.
  9. Compare F statistic and critical F.
- Data: 9 consumers rate three *Audi* models on a scale from 0 to 15.

Models	Q3	Q3	Q3	RS3	RS3	RS3	A3	A3	A3
Ratings	7	5	2	16	12	15	12	8	9

We can use the following **box plot**[1] to describe the data:



[1] This figure is visualized by `boxplot` in `pyplot`. It can also be plotted with `seaborn` in Python.

1.1 State the Hypothesis

In this question, the **independent** variable is the *Model* (Q3, RS3, A3). The **dependent** variable is the *ratings* given by the consumers.

H_0 : All population rating means are equal ($\mu_{\text{Q3}} = \mu_{\text{RS3}} = \mu_{\text{A3}}$.)
(There's no relationship between models and ratings)

H_1 : At least one population mean (rating average) is different (not all mean ratings are equal).
(There exists a relationship between models and ratings)

1.2 Calculate the Means of Each Group

$$\text{Mean rating for Q3 } (\overline{X}_{\text{Q3}}) = \frac{1}{N_{\text{Q3}}}\sum_{i=1}^{N_{\text{Q3}}} Q3_i = \frac{7+5+2}{3} = 4.6667$$

$$\text{Mean rating for RS3 } (\overline{X}_{\text{RS3}}) = \frac{1}{N_{\text{RS3}}}\sum_{j=1}^{N_{\text{RS3}}} RS3_i = \frac{16+12+15}{3} = 14.3333$$

$$\text{Mean rating for A3 } (\overline{X}_{\text{A3}}) = \frac{1}{N_{\text{A3}}}\sum_{k=1}^{N_{\text{A3}}} A3_i = \frac{12+8+9}{3} = 9.6667$$

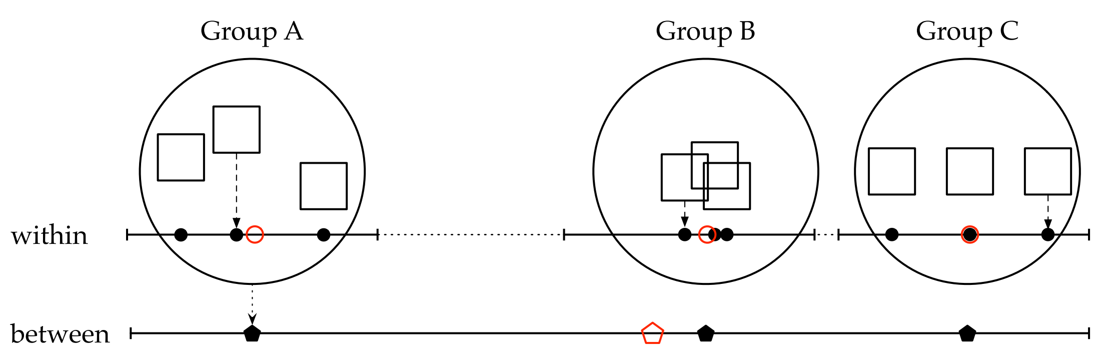
$$\text{Total mean rating } (\overline{X}) = \frac{1}{N}\sum_{m=1}^{N} \text{model}_m = \frac{7+5+\cdots+16+12+\cdots+8+9}{9} = 9.5556$$

1.3 Construct the ANOVA Table

- The **variance "between" groups** measures the **main effect**.
- The **variance "within" groups** measures the **errors**.

	Degree of freedom(DF)	Sum of Squares(SS)	Mean Squares(MS)	F statistic(F)	Critical F
Between					
Within					
Total					

We can visualize the concept of *within* and *between* with the **vertical projections** in the following graph:



1.4 Calculate the Degrees of Freedom

DF_{total} = **Total** degree of freedom = (total numbers of observations - 1) = 9 - 1 = 8.

$\text{DF}_{\text{between}}$ = Degree of freedom for **between** = (numbers of groups/category - 1) = 3 - 1 = 2.

$\text{DF}_{\text{within}}$ = Degree of freedom for **within** = (total degree of freedom - Degree of freedom for between) = 8 - 2 = 6.

	Degree of freedom(DF)	Sum of Squares(SS)	Mean Squares(MS)	F statistic(F)	Critical F
Between	**2**				
Within	**6**				
Total	**8**				

1.5 Calculate the Sum of Squares (squared deviation)

Models	Q3	Q3	Q3	RS3	RS3	RS3	A3	A3	A3
Ratings	7	5	2	16	12	15	12	8	9

★ $\overline{X}_{\text{Q3}} = 4.6667$, $\overline{X}_{\text{RS3}} = 14.3333$, $\overline{X}_{\text{A3}} = 9.6667$, $\overline{X} = 9.5556$

- **Total** sum of squares:

$$\begin{aligned} \text{SS}_{\text{total}} &= \sum_{p=1}^{N} \left(X_p - \overline{X}\right)^2 \\ &= (7-9.5556)^2 + (5-9.5556)^2 + (2-9.5556)^2 + \cdots + (8-9.5556)^2 + (9-9.5556)^2 \\ &= 170.2222 \end{aligned}$$

- **Between** groups ($q =$ Q3, RS3, A3) sum of squares:

$$\begin{aligned} \text{SS}_{\text{between}} &= \sum_{q \in \{\text{Q3, RS3, A3}\}} \left(\overline{X}_q - \overline{X}\right)^2 \cdot f_q \text{ (number of oberservtaions in that group/category)} \\ &= (4.6667-9.5556)^2 \cdot 3 + (14.3333-9.5556)^2 \cdot 3 + (9.6667-9.5556)^2 \cdot 3 \\ &= 140.2222 \end{aligned}$$

- **Within** groups ($q =$ Q3, RS3, A3) sum of squares ($\text{SS}_{\text{within}} = \text{SS}_{\text{error}}$):

$$\begin{aligned} \text{SS}_{\text{within}} &= \sum_{q \in \{\text{Q3, RS3, A3}\}} \sum_{i=1}^{N_q} \left(X_{q,i} - \overline{X}_q\right)^2 \\ &= \left(X_{\text{Q3},1} - \overline{X}_{\text{Q3}}\right)^2 + \left(X_{\text{Q3},2} - \overline{X}_{\text{Q3}}\right)^2 + \left(X_{\text{Q3},3} - \overline{X}_{\text{Q3}}\right)^2 + \\ &\quad \left(X_{\text{RS3},1} - \overline{X}_{\text{RS3}}\right)^2 + \left(X_{\text{RS3},2} - \overline{X}_{\text{RS3}}\right)^2 + \left(X_{\text{RS3},3} - \overline{X}_{\text{RS3}}\right)^2 + \\ &\quad \left(X_{\text{A3},1} - \overline{X}_{\text{A3}}\right)^2 + \left(X_{\text{A3},2} - \overline{X}_{\text{A3}}\right)^2 + \left(X_{\text{A3},3} - \overline{X}_{\text{A3}}\right)^2 \\ &= (7-4.6667)^2 + (5-4.6667)^2 + (2-4.6667)^2 + \\ &\quad (16-14.3333)^2 + (12-14.3333)^2 + (15-14.3333)^2 + \\ &\quad (12-9.6667)^2 + (8-9.6667)^2 + (9-9.6667)^2 \\ &= 30 \end{aligned}$$

	Degree of freedom(DF)	Sum of Squares(SS)	Mean Squares(MS)	F statistic(F)	Critical F
Between	2	**140.2222**			
Within	6	**30**			
Total	8	**170.2222**			

1.6 Calculate the Mean Squares

- Mean square: the average squared deviation per degree of freedom.

$$MS_{\text{between}} = \frac{SS_{\text{between}}}{DF_{\text{between}}} = \frac{140.2222}{2} = 70.1111$$
$$MS_{\text{within}} = \frac{SS_{\text{within}}}{DF_{\text{within}}} = \frac{30}{6} = 5$$

	Degree of freedom(DF)	Sum of Squares(SS)	Mean Squares(MS)	F statistic(F)	Critical F
Between	2	140.2222	**70.1111**		
Within	6	30	**5**		
Total	8	170.2222			

1.7 Calculate the F Statistic

$$\text{F Statistic} = \frac{MS_{\text{between}}}{MS_{\text{within}}} = \frac{70.1111}{5} = 14.0222$$

	Degree of freedom(DF)	Sum of Squares(SS)	Mean Squares(MS)	F statistic(F)	Critical F
Between	2	140.2222	70.1111	**14.0222**	
Within	6	30	5		
Total	8	170.2222			

1.8 Find the Critical F

- We will conduct a **F test** at **1% significance level (α)**.

$$\text{Critical } F_{\mathbf{0.01},\ DF_{\textbf{between}}(\text{F table } \textbf{column}),\ DF_{\textbf{within}}(\text{F table } \textbf{row})} = F_{0.01,2,6} = 10.92$$

	Degree of freedom(DF)	Sum of Squares(SS)	Mean Squares(MS)	F statistic(F)	Critical F
Between	2	140.2222	70.1111	14.0222	**10.92**
Within	6	30	5		
Total	8	170.2222			

1.9 Compare F Statistic and Critical F

The F test concludes the following:

F statistic = 14.0222 $>$ 10.92 = **critical F**

This means that the **P-value** for this F test is **less** than the **significance level** (α). Hence, we reject the null hypothesis (H_0) and conclude that the population's mean ratings for these three groups/models (Q3, RS3, A3) are not all the same.

2 Relation Strength Measurements

- **Omega squared** (ω^2): measures the relationship strength between the independent variable (group/car models) and the dependent variables (ratings).

$$\omega^2 = \frac{SS_{\text{between}} - (DF_{\text{between}} \cdot MS_{\text{within}})}{SS_{\text{total}} + MS_{\text{within}}} = \frac{140.2222 - (2 \cdot 5)}{170.2222 + 5} = \frac{130.2222}{175.2222} = 0.7432$$

- **Eta squared** (η^2): measures the relation strength in a sample.

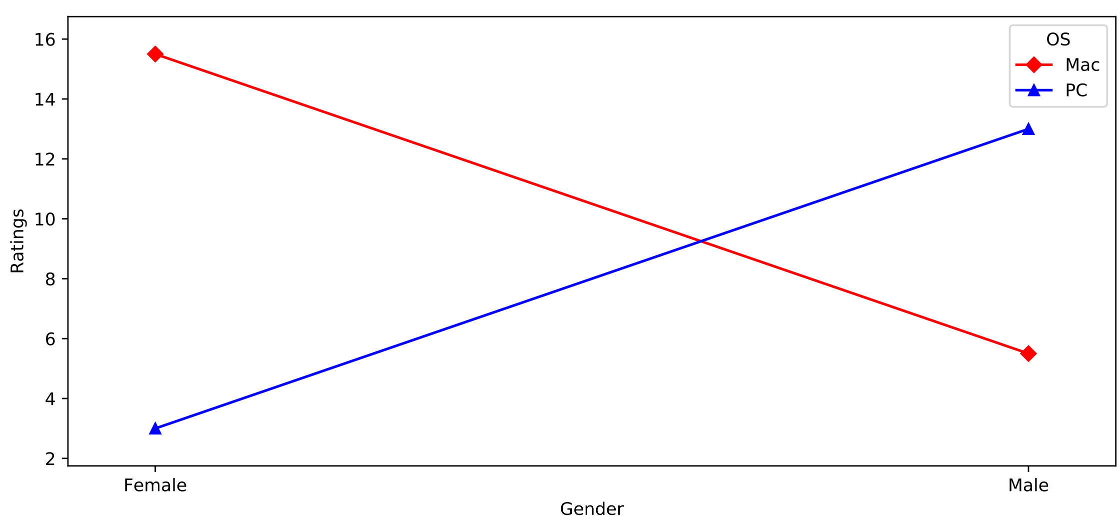
$$\eta^2 = \frac{SS_{\text{between}}}{SS_{\text{total}}} = \frac{140.2222}{170.2222} = 0.8238$$

3 Two-Way ANOVA

- Assumptions:
  - Independent
  - Normality
  - Homogeneous variance
  - Balance data: the groups/categories must have the same sample size.
- Roadmap:
  1. State the hypothesis.
  2. Calculate the means.
  3. Construct the ANOVA table.
  4. Calculate the degrees of freedom (DF).
  5. Calculate the sum of squares (SS).
  6. Calculate mean squares (MS).
  7. Calculate the F statistics.
  8. Find the critical F.
  9. Compare F statistic and critical F for each effect.
- Data: 2 operating systems, PC (Windows) and Mac, are rated by 2 genders - male and female consumers. This dataset can be shown explicitly:

OS	Gender	Ratings
PC	Male	12
PC	Male	14
PC	Female	0
PC	Female	6
Mac	Male	8
Mac	Male	3
Mac	Female	13
Mac	Female	18

The two **independent** variables are "OS" and "Gender." The **dependent** variable is "Ratings."



3.1 State the Hypothesis

There are **3 sets** of hypothesis:

► **Set 1:**

$$\text{H}_0 : \text{The population means of } \textbf{OS} \text{ ratings are equal } (\mu_{\text{PC}} = \mu_{\text{Mac}}).$$
$$\text{H}_1 : \text{At least one mean ratings is different } (\mu_{\text{PC}} \neq \mu_{\text{Mac}}).$$

► **Set 2:**

$$\text{H}_0 : \text{The population means of } \textbf{Gender} \text{ ratings are equal } (\mu_{\text{male}} = \mu_{\text{female}}).$$
$$\text{H}_1 : \text{At least one mean ratings is different } (\mu_{\text{male}} \neq \mu_{\text{female}}).$$

► **Set 3:**

$$\text{H}_0 : \text{There is } \textbf{no} \text{ replationship between OS and gender.}$$
$$\text{H}_1 : \text{There } \textbf{exists} \text{ replationship between OS and gender.}$$

3.2 Calculate the Means

OS	Gender	Ratings
PC	Male	12
PC	Male	14
PC	Female	0
PC	Female	6
Mac	Male	8
Mac	Male	3
Mac	Female	13
Mac	Female	18

$$\text{PC mean rating from male } \left(\overline{X}^{\text{PC}}_{\text{male}}\right) = \frac{1}{N^{\text{PC}}_{\text{male}}} \sum_{i \in \text{male}}^{N^{\text{PC}}_{\text{male}}} X^{\text{PC}}_i = \frac{12+14}{2} = 13$$

$$\text{PC mean rating from female } \left(\overline{X}^{\text{PC}}_{\text{female}}\right) = \frac{1}{N^{\text{PC}}_{\text{female}}} \sum_{j \in \text{female}}^{N^{\text{PC}}_{\text{female}}} X^{\text{PC}}_j = \frac{0+6}{2} = 3$$

$$\text{Mac mean rating from male } \left(\overline{X}^{\text{Mac}}_{\text{male}}\right) = \frac{1}{N^{\text{Mac}}_{\text{male}}} \sum_{m \in \text{male}}^{N^{\text{Mac}}_{\text{male}}} X^{\text{Mac}}_m = \frac{8+3}{2} = 5.5$$

$$\text{Mac mean rating from female } \left(\overline{X}^{\text{Mac}}_{\text{female}}\right) = \frac{1}{N^{\text{Mac}}_{\text{female}}} \sum_{n \in \text{female}}^{N^{\text{Mac}}_{\text{female}}} X^{\text{Mac}}_n = \frac{13+18}{2} = 15.5$$

$$\text{Total mean rating } \left(\overline{X}\right) = \frac{1}{N} \sum_{s=1}^{N} X_s = \frac{12+14+\cdots+13+18}{8} = \frac{74}{8} = 9.25$$

$$\text{Mean rating for PC } \left(\overline{X}_{\text{PC}}\right) = \frac{1}{N_{\text{PC}}} \sum_{m=1}^{N_{\text{PC}}} X^{\text{PC}}_m = \frac{12+14+0+6}{4} = 8$$

$$\text{Mean rating for Mac } \left(\overline{X}_{\text{Mac}}\right) = \frac{1}{N_{\text{Mac}}} \sum_{n=1}^{N_{\text{Mac}}} X^{\text{Mac}}_n = \frac{8+3+13+18}{4} = 10.5$$

$$\text{Mean rating from male } \left(\overline{X}_{\text{male}}\right) = \frac{1}{N_{\text{male}}} \sum_{i=1}^{N_{\text{male}}} X^{\text{male}}_i = \frac{12+14+8+3}{4} = 9.25$$

$$\text{Mean rating from female } \left(\overline{X}_{\text{female}}\right) = \frac{1}{N_{\text{female}}} \sum_{j=1}^{N_{\text{female}}} X^{\text{female}}_j = \frac{0+6+13+18}{4} = 9.25$$

3.3 Construct the ANOVA Table

	Degree of freedom(DF)	Sum of Squares(SS)	Mean Squares(MS)	F Statistic(F)	Critical F
OS					
Gender					
Interaction btw OS and Gender					
error					
Total					

3.4 Calculate Degrees of Freedom

$$
\begin{aligned}
\text{DF}_{\text{total}} &= \text{(numbers of observations - 1) = 8 - 1 = 7.} \\
\text{DF}_{\text{OS}} &= \text{(numbers of OS categories - 1) = 2 - 1 = 1.} \\
\text{DF}_{\text{Gender}} &= \text{(numbers of Gender categories - 1) = 2 - 1 = 1.} \\
\text{DF}_{\text{interaction}} &= (\text{DF}_{\text{OS}} \cdot \text{DF}_{\text{Gender}}) = 1 \cdot 1 = 1. \\
\text{DF}_{\text{error}} &= \text{DF}_{\text{within}} = (\text{DF}_{\text{total}} - \text{DF}_{\text{OS}} - \text{DF}_{\text{Gender}} - \text{DF}_{\text{interaction}}) = 7 - 1 - 1 - 1 = 4.
\end{aligned}
$$

	Degree of freedom(DF)	Sum of Squares(SS)	Mean Squares(MS)	F Statistic(F)	Critical F
OS	**1**				
Gender	**1**				
Interaction btw OS and Gender	**1**				
error	**4**				
Total	**7**				

3.5 Calculate the Sum of Squares

OS	Gender	Ratings
PC	Male	12
PC	Male	14
PC	Female	0
PC	Female	6
Mac	Male	8
Mac	Male	3
Mac	Female	13
Mac	Female	18

★ $\overline{X}^{\text{PC}}_{\text{male}} = 13,\quad \overline{X}^{\text{PC}}_{\text{female}} = 3,\quad \overline{X}^{\text{Mac}}_{\text{male}} = 5.5,\quad \overline{X}^{\text{Mac}}_{\text{female}} = 15.5,\quad \overline{X} = 9.25$

★ $\overline{X}_{\text{PC}} = 8,\quad \overline{X}_{\text{Mac}} = 10.5,\quad ,\overline{X}_{\text{male}} = 9.25,\quad \overline{X}_{\text{female}} = 9.25$

- **Total** sum of squares:

$$
\begin{aligned}
\text{SS}_{\text{total}} &= \sum_{p=1}^{N} (X_p - \overline{X})^2 \\
&= (12-9.25)^2 + (14-9.25)^2 + (0-9.25)^2 + \cdots + (13-9.25)^2 + (18-9.25)^2 \\
&= 257.5
\end{aligned}
$$

- **Error** sum of squares:

$$\begin{aligned}
SS_{\text{error}} = SS_{\text{within}} &= \sum_{p\in\{\text{OS}\}} \sum_{q\in\{\text{Gender}\}} \sum_{i=1}^{N_q^p} \left(X_{q,i}^p - \overline{X}_q^p\right)^2 \\
&= \left(X_{\text{male},1}^{\text{PC}} - \overline{X}_{\text{male}}^{\text{PC}}\right)^2 + \left(X_{\text{male},2}^{\text{PC}} - \overline{X}_{\text{male}}^{\text{PC}}\right)^2 + \left(X_{\text{female},1}^{\text{PC}} - \overline{X}_{\text{female}}^{\text{PC}}\right)^2 + \left(X_{\text{female},2}^{\text{PC}} - \overline{X}_{\text{female}}^{\text{PC}}\right)^2 + \\
&\quad \left(X_{\text{male},1}^{\text{Mac}} - \overline{X}_{\text{male}}^{\text{Mac}}\right)^2 + \left(X_{\text{male},2}^{\text{Mac}} - \overline{X}_{\text{male}}^{\text{Mac}}\right)^2 + \left(X_{\text{female},1}^{\text{Mac}} - \overline{X}_{\text{female}}^{\text{Mac}}\right)^2 + \left(X_{\text{female},2}^{\text{Mac}} - \overline{X}_{\text{female}}^{\text{Mac}}\right)^2 \\
&= (12-13)^2 + (14-13)^2 + (0-3)^2 + (6-3)^2 + \\
&\quad (8-5.5)^2 + (3-5.5)^2 + (13-15.5)^2 + (18-15.5)^2 \\
&= 45
\end{aligned}$$

- **OS** sum of squares:

$$\begin{aligned}
SS_{\text{OS}} &= \sum_{u\in\{\text{PC, Mac}\}} \left(\overline{X}_u - \overline{X}\right)^2 \cdot f_u \text{ (number of oberservtaions in that group/category)} \\
&= (8-9.25)^2 \cdot 4 + (10.5-9.25)^2 \cdot 4 \\
&= 12.5
\end{aligned}$$

- **Gender** sum of squares:

$$\begin{aligned}
SS_{\text{Gender}} &= \sum_{v\in\{\text{male, female}\}} \left(\overline{X}_v - \overline{X}\right)^2 \cdot f_v \text{ (number of oberservtaions in that group/category)} \\
&= (9.25-9.25)^2 \cdot 4 + (9.25-9.25)^2 \cdot 4 \\
&= 0
\end{aligned}$$

- **Interaction** sum of squares:

$$\begin{aligned}
SS_{\text{interaction}} &= SS_{\text{total}} - SS_{\text{OS}} - SS_{\text{Gender}} - SS_{\text{error}} \\
&= 257.5 - 12.5 - 0 - 45 \\
&= 200
\end{aligned}$$

	Degree of freedom(DF)	Sum of Squares(SS)	Mean Squares(MS)	F Statistic(F)	Critical F
OS	1	**12.5**			
Gender	1	**0**			
Interaction btw OS and Gender	1	**200**			
error	4	**45**			
Total	7	**257.5**			

3.6 Calculate Mean Squares

$$\begin{aligned}
MS_{\text{OS}} &= \frac{SS_{\text{OS}}}{DF_{\text{OS}}} = \frac{12.5}{1} = 12.5 \\
MS_{\text{Gender}} &= \frac{SS_{\text{Gender}}}{DF_{\text{Gender}}} = \frac{0}{1} = 0 \\
MS_{\text{interaction}} &= \frac{SS_{\text{interaction}}}{DF_{\text{interaction}}} = \frac{200}{1} = 200 \\
MS_{\text{error}} &= \frac{SS_{\text{error}}}{DF_{\text{error}}} = \frac{45}{4} = 11.25
\end{aligned}$$

	Degree of freedom(DF)	Sum of Squares(SS)	Mean Squares(MS)	F Statistic(F)	Critical F
OS	1	12.5	**12.5**		
Gender	1	0	**0**		
Interaction btw OS and Gender	1	200	**200**		
error	4	45	**11.25**		
Total	7	257.5			

3.7 Calculate the F Statistics

- Notice that since we have 3 sets of hypothesis, we will obtain 3 F statistics as well.

$$\text{F Statistic}_{\text{OS}} = \frac{\text{MS}_{\text{OS}}}{\text{MS}_{\text{error}}} = \frac{12.5}{11.25} = 1.1111$$
$$\text{F Statistic}_{\text{Gender}} = \frac{\text{MS}_{\text{Gender}}}{\text{MS}_{\text{error}}} = \frac{0}{11.25} = 0$$
$$\text{F Statistic}_{\text{interaction}} = \frac{\text{MS}_{\text{interaction}}}{\text{MS}_{\text{error}}} = \frac{200}{11.25} = 17.7778$$

3.8 Find the Critical F (from F table)

- We will conduct a **F test** at **1% significance level**:

$$\text{Critical F} = \text{F}_{\mathbf{0.01},\ \text{DF}_{\textbf{effect}}(\text{F table }\textbf{column}),\ \text{DF}_{\textbf{error}}(\text{F table }\textbf{row})}$$

- OS critical F: $F_{0.01,1,4} = 21.2 > 1.1111 = \text{ F statistic}_{\text{OS}}$
- Gender critical F: $F_{0.01,1,4} = 21.2 > 0 = \text{ F statistic}_{\text{Gender}}$
- Interaction critical F: $F_{0.01,1,4} = 21.2 > 17.7778 = \text{ F statistic}_{\text{interaction}}$

	Degree of freedom(DF)	Sum of Squares(SS)	Mean Squares(MS)	F Statistic(F)	Critical F
OS	1	12.5	12.5	**1.1111**	**21.2**
Gender	1	0	0	**0**	**21.2**
Interaction btw OS and Gender	1	200	200	**17.7778**	**21.2**
error	4	45	11.25		
Total	7	257.5			

Hence, we fail to reject the null hypothesis (H_0) of every set and conclude that the population means of OS ratings and Gender ratings are equal. Furthermore, there are no relationship between OS and Gender.

The rejection result obtained from the approach of comparing the obtained statistics to the critical F also implies that the same result should be obtained if we take the P-value approach. We can see that the P-values (obtained by statistical software):

$$\text{P-value}_{\text{OS}} = 0.3513$$
$$\text{P-value}_{\text{Gender}} = 1.0000$$
$$\text{P-value}_{\text{interaction}} = 0.0135$$

these three P-values are all **greater** than our test's significance level (α) which is 0.01. Hence, the approach of **comparing the P-value with α** also gives us the same "fail to reject the null" result.

4 The χ^2 and F Distributions

The χ^2 (chi-square) and the F distributions (probability measures) are important functions that help us to conduct important tests.

When we want to test whether two random variables are independent, we will use the χ^2 test. When we want to know whether or not the variances of the two population are equal and concerning whether or not three or more population means are equal, we will use the F test. Indeed, we can obtain these two distributions directly from a standard normal distribution.

Suppose, we have random variables that follows a standard normal distribution:

$$N_1, ..., N_m \overset{\text{i.i.d}}{\sim} N(0, 1).$$

Then a **χ^2 random variable with m degrees of freedom** can be obtained from the **sums** of m squared independent standard normal random variables:

$$\left(N_1^2 + N_2^2 + \cdots + N_m^2\right) \sim \chi_m^2$$

A **F random variable with degrees of freedom r and s** is the **ratio** of two independent χ^2 random variables divided by their degrees of freedom r and s:

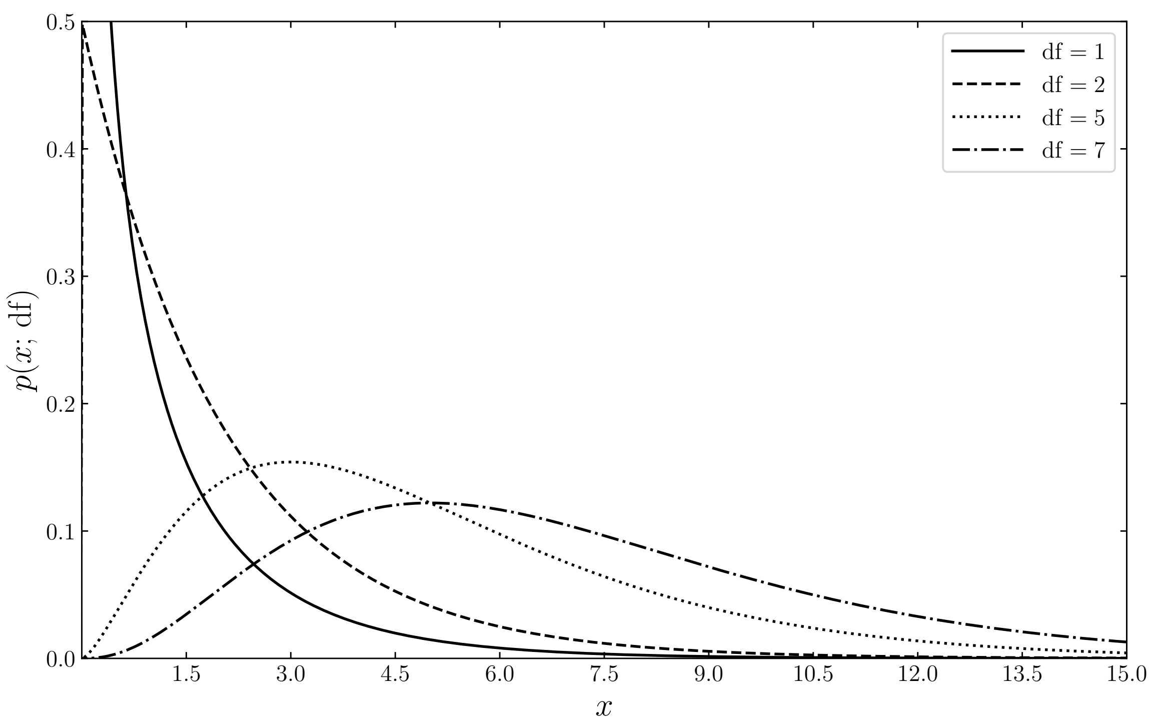
$$\text{Let } Y \sim \chi_r^2 \text{ and } Z \sim \chi_s^2 \text{ (indepedent), then } \frac{\frac{1}{r}Y}{\frac{1}{s}Z} \sim F_{r,s}.$$

4.1 The χ^2 Distribution

Suppose random variable Y follows a χ_ν^2 distribution : $Y \sim \chi_\nu^2$. This distribution can be characterized by the following probability density function:

$$f(y) = \frac{y^{\frac{\nu}{2}} \cdot e^{-\frac{y}{2}}}{2^{\frac{\nu}{2}} \cdot \Gamma\left(\frac{\nu}{2}\right)}$$

where $\Gamma\left(\frac{\nu}{2}\right)$ is the gamma function: $\Gamma(\alpha) = (\alpha - 1)!$. Hence, we can see that the χ^2 distribution is determined by the degree of freedom ν. When we vary the ν, we can obtain different shapes of the density function which is illustrated as below:

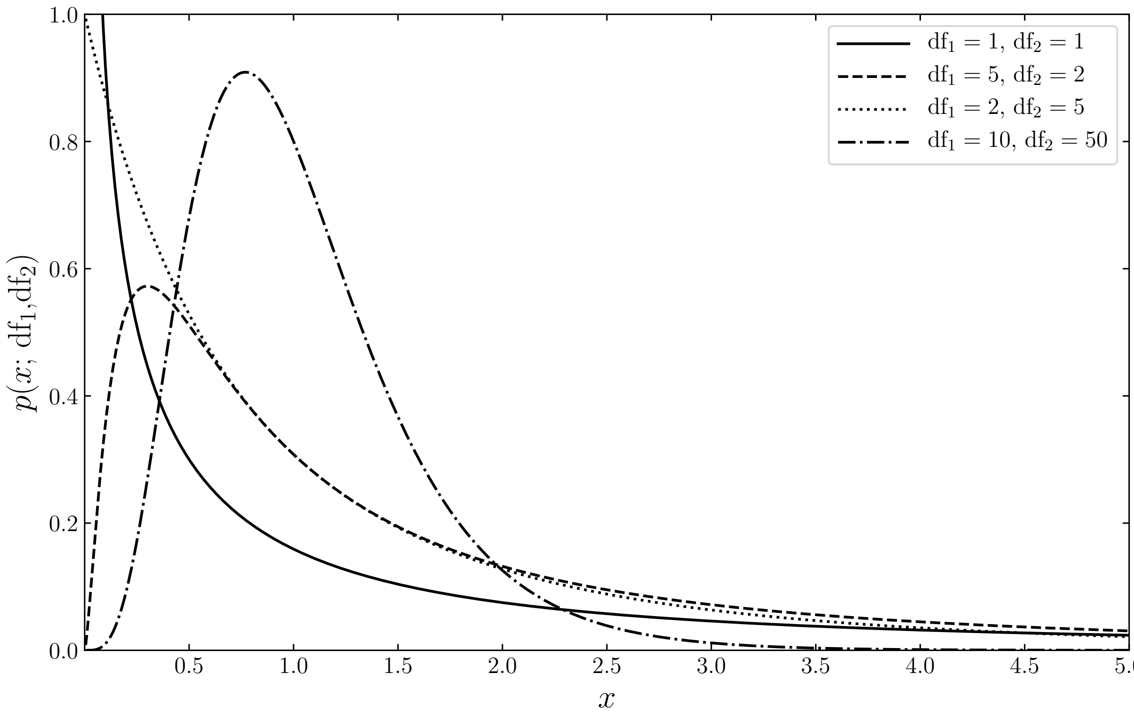


4.2 The F Distribution

The F distribution is determined by two degrees of freedom: F_{ν_1,ν_2}. Suppose random variable Y follows a F distribution: $Y \sim F_{\nu_1,\nu_2}$, then the density function of Y is the following:

$$f(y) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\cdot\Gamma\left(\frac{\nu_2}{2}\right)}\left(\frac{\nu_1}{\nu_2}\right)^{\frac{\nu_1}{2}}\cdot\left(1+\frac{\nu_1}{\nu_2}y\right)^{-\frac{\nu_1+\nu_2}{2}}$$

which shows that the shape of F density function is determined by two degrees of freedom. We can illustrate its density function with respect to different pairs of degrees of freedom as follows:



5 Comparison between one-way and two-way ANOVA

The main characteristics of variance analysis for these two different approaches can be summarized in the following table:

	One-Way	Two-Way
The Test	compare between the means of 3 or more groups/categories of data	compare between the means of 3 or more groups/categories of data where involves two independent variables
Number of independent variables	1	2
Compared object	means of 3 or more groups of an independent variable on a dependent variable	the effect of multiple groups of two independent variables on a dependent variable and on each other