TWO WAY ANOVA (Factorial Analysis)

- A two-way ANOVA is used to estimate how the mean of a quantitative variable or numerical variable changes according to the levels of two categorical variables.
- Basically, it tells how 2 categorical features have impact on a numerical feature.

A two-way ANOVA with interaction tests three null hypotheses at the same time:

- There is no difference in group means at any level of the first independent variable.
- There is no difference in group means at any level of the second independent variable.
- The effect of one independent variable does not depend on the effect of the other independent variable (a.k.a. no interaction effect).

A two-way ANOVA without interaction (a.k.a. an additive two-way ANOVA) only tests the first two of these hypotheses.

Two-way ANOVA hypotheses

In our crop yield experiment, we can test three hypotheses using two-way ANOVA:

Null hypothesis (H_0)	Alternate hypothesis (H_a)
There is no difference in average yield for any fertilizer type.	There is a difference in average yield by fertilizer type.
There is no difference in average yield at either planting density.	There is a difference in average yield by planting density.
The effect of one independent variable on average yield does not depend on the effect of the other independent variable (a.k.a. no interaction effect).	There is an interaction effect between planting density and fertilizer type on average yield.

Your independent variables should not be dependent on one another (i.e. one should not cause the other). This is impossible to test with categorical variables – it can only be ensured by good experimental design.

In addition, your dependent variable should represent unique observations – that is, your observations should not be grouped within locations or individuals.

NUMERICAL EXAMPLE:

Gender	Score	Age Group
Boys	4	10 Year Olds
Boys	6	10 Year Olds
Boys	8	10 Year Olds
Girls	4	10 Year Olds
Girls	8	10 Year Olds
Girls	9	10 Year Olds
Boys	6	11 Year Olds
Boys	6	11 Year Olds
Boys	9	11 Year Olds
Girls	7	11 Year Olds
Girls	10	11 Year Olds
Girls	13	11 Year Olds
Boys	8	12 Year Olds
Boys	9	12 Year Olds
Boys	13	12 Year Olds
Girls	12	12 Year Olds
Girls	14	12 Year Olds
Girls	16	12 Year Olds

- So the above dataset is used to demonstrate two way ANOVA test.
- In the above dataset we want to test whether Gender has impact on score or age group has impact on score or both has the impact on score.
- Here gender and age group are independent variables or factors and score is dependent variable.

H_0: Gender will have no significant effect on students score.

H_0: Age will have no significant effect on students score.

H_0: Gender and Age interaction will have
no significant effect on students score.

- These are the 3 different null hypothesis for the test as already mentioned.

1.Divide the dataset with respect to gender feature.

Gender	Score	Age Group
Boys	4	10 Year Olds
Boys	6	10 Year Olds
Boys	8	10 Year Olds
Boys	6	11 Year Olds
Boys	6	11 Year Olds
Boys	9	11 Year Olds
Boys	8	12 Year Olds
Boys	9	12 Year Olds
Boys	13	12 Year Olds

Gender	Score	Age Group
Girls	4	10 Year Olds
Girls	8	10 Year Olds
Girls	9	10 Year Olds
Girls	7	11 Year Olds
Girls	10	11 Year Olds
Girls	13	11 Year Olds
Girls	12	12 Year Olds
Girls	14	12 Year Olds
Girls	16	12 Year Olds

2.Just write the above same table in classified form as below and calculate mean for each category.

Boys

	10 Year Olds	11 Year Olds	12 Year Olds
	4	6	8
	6	6	9
	8	9	13
means	6	7	10

Girls

10 Year Olds	11 Year Olds	12 Year Olds
4	7	12
8	10	14
9	13	16
7	10	14

3.Now calculate the whole mean table (take the mean values from above 2 tables and form whole mean table)

	Mean Table			Marginal Mean
	10 Year Olds	11 Year Olds	12 Year Olds	Average
Boys	6	7	10	7.7
Girls	7	10	14	10.3
Average	6.5	8.5	12	9

4.These are the factors which we need to determine to perform test

Sum of Squares 1st Factor (Gender)

+

Sum of Squares 2nd Factor (Age)

+

Sum of Squares Within (Error)

+

Sum of Squares Both Factors

Sum of Squares Total

Sum of squares 1st factor (Gender):

Sum of Squares 1st Factor (Gender)

Score	Boys Mean		Grand Mean		Score	Girls Mean		Grand Mean	
4	7.7	-	9	$= (-1.3)^2 = 1.8$	4	10.3	-	9	$= (1.3)^2 = 1.8$
6	7.7	-	9	$= (-1.3)^2 = 1.8$	8	10.3	-	9	$= (1.3)^2 = 1.8$
8	7.7	-	9	$= (-1.3)^2 = 1.8$	9	10.3	-	9	$= (1.3)^2 = 1.8$
6	7.7	-	9	$= (-1.3)^2 = 1.8$	7	10.3	-	9	$= (1.3)^2 = 1.8$
6	7.7	-	9	$= (-1.3)^2 = 1.8$	10	10.3	-	9	$= (1.3)^2 = 1.8$
9	7.7	-	9	$= (-1.3)^2 = 1.8$	13	10.3	-	9	$= (1.3)^2 = 1.8$
8	7.7	-	9	$= (-1.3)^2 = 1.8$	12	10.3	-	9	$= (1.3)^2 = 1.8$
9	7.7	-	9	$= (-1.3)^2 = 1.8$	14	10.3	-	9	$= (1.3)^2 = 1.8$
13	7.7	-	9	$= (-1.3)^2 = 1.8$	16	10.3	-	9	$= (1.3)^2 = 1.8$
			sum of squares	= 16				sum of squares	= 16

sum of squares for 1st Factor Gender = 16 + 16 = 32

The boys mean (7.7) is the marginal mean of boys and grand mean (9) is the average of marginal mean of boys and girls.

The girls mean (10.3) is the marginal mean of girls and grand mean (9) is the average of marginal mean of boys and girls.

So now add both sum of squares in order to get sum of squares of 1st factor(gender)

Sum of squares 2nd factor (Age):

Sum of Squares 2nd Factor (Age)

Boys

4	6.5 - 9 = $(-2.5)^2$ = 6.3
6	6.5 - 9 = $(-2.5)^2$ = 6.3
8	6.5 - 9 = $(-2.5)^2$ = 6.3
6	8.5 - 9 = $(-.5)^2$ = .25
6	8.5 - 9 = $(-.5)^2$ = .25
9	8.5 - 9 = $(-.5)^2$ = .25
8	12 - 9 = $(3)^2$ = 9.0
9	12 - 9 = $(3)^2$ = 9.0
13	12 - 9 = $(3)^2$ = 9.0

sum of squares = 46.5

Girls

4	6.5 - 9 = $(-2.5)^2$ = 6.3
8	6.5 - 9 = $(-2.5)^2$ = 6.3
9	6.5 - 9 = $(-2.5)^2$ = 6.3
7	8.5 - 9 = $(-.5)^2$ = .25
10	8.5 - 9 = $(-.5)^2$ = .25
13	8.5 - 9 = $(-.5)^2$ = .25
12	12 - 9 = $(3)^2$ = 9.0
14	12 - 9 = $(3)^2$ = 9.0
16	12 - 9 = $(3)^2$ = 9.0

sum of squares = 46.5

sum of squares for 2nd Factor = 93.0
Age

6.5 is avg of 10-year-old boys and girls

8.5 is avg of 11-year-old boys and girls

12 is avg of 12-year-old boys and girls

9 is the avg of mean of 10,11,12-year-old boys and girls.

Finally add both sum of squares in order to get sum of squares for 2nd factor.

Sum of squares within (error):

Sum of Squares Within (Error)

Boys

4	- 6 = $(-2.0)^2$ = 4.0
6	- 6 = $(0)^2$ = 0.0
8	- 6 = $(2.0)^2$ = 4.0
6	- 7 = $(-2.0)^2$ = 4.0
6	- 7 = $(-1.0)^2$ = 1.0
9	- 7 = $(2.0)^2$ = 4.0
8	- 10 = $(-2.0)^2$ = 4.0
9	- 10 = $(-1.0)^2$ = 1.0
13	- 10 = $(3.0)^2$ = 9.0

sum of squares = 28.0

Girls

4	- 7 = $(-3.0)^2$ = 9.0
8	- 7 = $(1.0)^2$ = 1.0
9	- 7 = $(2.0)^2$ = 4.0
7	- 10 = $(-3.0)^2$ = 9.0
10	- 10 = $(0)^2$ = 0.0
13	- 10 = $(3.0)^2$ = 9.0
12	- 14 = $(-2.0)^2$ = 4.0
14	- 14 = $(0)^2$ = 0.0
16	- 14 = $(2.0)^2$ = 4.0

sum of squares = 40.0

total sum of squares within = 68

In boys table 6,7,10 are means of 10,11,12-year-old boys respectively

In girls table 7,10,14 are means of 10,11,12-year-old girls respectively

Finally add both the sum of squares in order to get the sum of squares within.

Sum of squares total:

Score	Grand Mean	(Score - Grand Mean)^2
4	- 9	$=(-5)^2 = 25.0$
6	- 9	$=(-3)^2 = 9.0$
8	- 9	$=(-1)^2 = 1.0$
6	- 9	$=(-3)^2 = 9.0$
6	- 9	$=(-3)^2 = 9.0$
9	- 9	$=(0)^2 = 0.0$
8	- 9	$=(-1)^2 = 1.0$
9	- 9	$=(0)^2 = 0.0$
13	- 9	$=(4)^2 = 16.0$
4	- 9	$=(-5)^2 = 25.0$
8	- 9	$=(1)^2 = 1.0$
9	- 9	$=(0)^2 = 0.0$
7	- 9	$=(-2)^2 = 4.0$
10	- 9	$=(1)^2 = 1.0$
13	- 9	$=(4)^2 = 16.0$
12	- 9	$=(-3)^2 = 9.0$
14	- 9	$=(5)^2 = 25.0$
16	- 9	$=(7)^2 = 49.0$
		200

9 is the grand mean of boys and girls.

Still, we didn't find the sum of squares of both the factors so to find that

Sum of Squares Total - Sum of Squares 1st Factor - Sum of Squares 2nd Factor - Sum of Squares Within

200 - 32 - 93 - 68 = Sum of Square Both Factors

7

So sum of squares of both factors is 7.

Final view of calculations done so far:

Sum of Squares 1st Factor (Gender)	32
+	
Sum of Squares 2nd Factor (Age)	93
+	
Sum of Squares Within (Error)	68
+	
Sum of Squares Both Factors	7
Sum of Squares Total	200

Interpretation of results:

	Sum of Squares	Degrees of Freedom	Mean Square	F ratio
Sum of Squares 1st Factor (Gender)	32	1	32	5.64
Sum of Squares 2nd Factor (Age)	93	2	46.50	8.20
Sum of Square Both Factors	7	2	3.5	.62
Sum of Squares Within (Error)	68	12	5.67	
Sum of Squares Total	200	17		

Degree of freedom for sum of squares gender

- In gender there are 2 categories that is boys and girls. So the df will be n-1 that is 2-1=1
- So the degree of freedom for sum of squares gender is 1

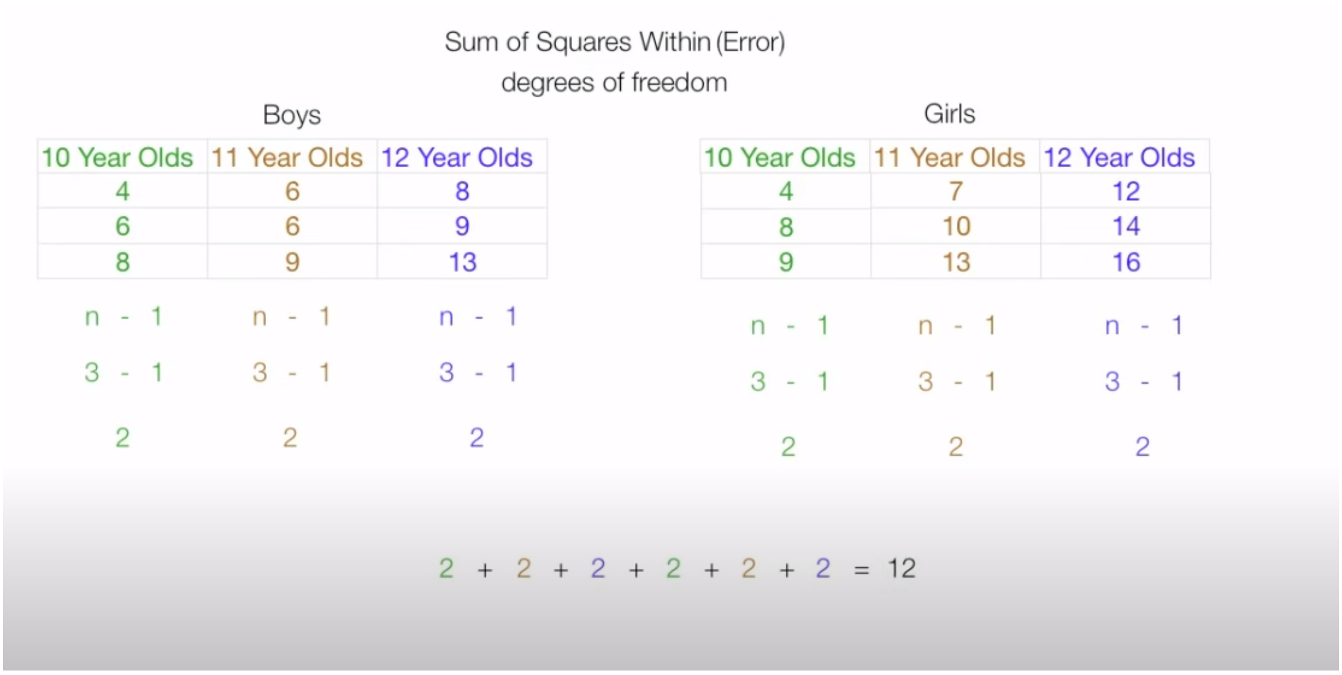
Degree of freedom for sum of squares Age

- In Age there are 3 categories that is 10,11,12-year-old. So, the df will be n-1 that is 3-1=2
- So the degree of freedom for sum of squares gender is 2

Degree of freedom for sum of squares both

- It can be calculated by multiplying df of both the features that 2*1=2
- So the degree of freedom for sum of squares both is 2

Degree of freedom for sum of squares both



- So the degree of freedom for sum of squares within is 12.

Mean square value can be calculated by

Sum of square / Degree of freedom

F-Ratio

F Ratio for sum of squares gender

$$\text{F test} = \frac{(\text{sum of square gender / degree of freedom gender})}{(\text{Sum of square within / degree of freedom within})}$$

F test= 5.64

According to F table the F value pr critical value for df(1,12) is 4.75

5.64 > 4.75

Therefore 5.64 falls on rejection region

So reject the null hypothesis.

F Ratio for sum of squares Age

$$F \text{ test} = \frac{(\text{sum of square age / degree of freedom age})}{(\text{Sum of square within / degree of freedom within})}$$

F test= 8.20

According to F table the F value pr critical value for df (2,12) is 3.89

8.20> 3.89

Therefore 8.20 falls on rejection region

So reject the null hypothesis.

F Ratio for sum of squares Both

$$F \text{ test} = \frac{(\text{sum of square both / degree of freedom both})}{(\text{Sum of square within / degree of freedom within})}$$

F test= 0.62

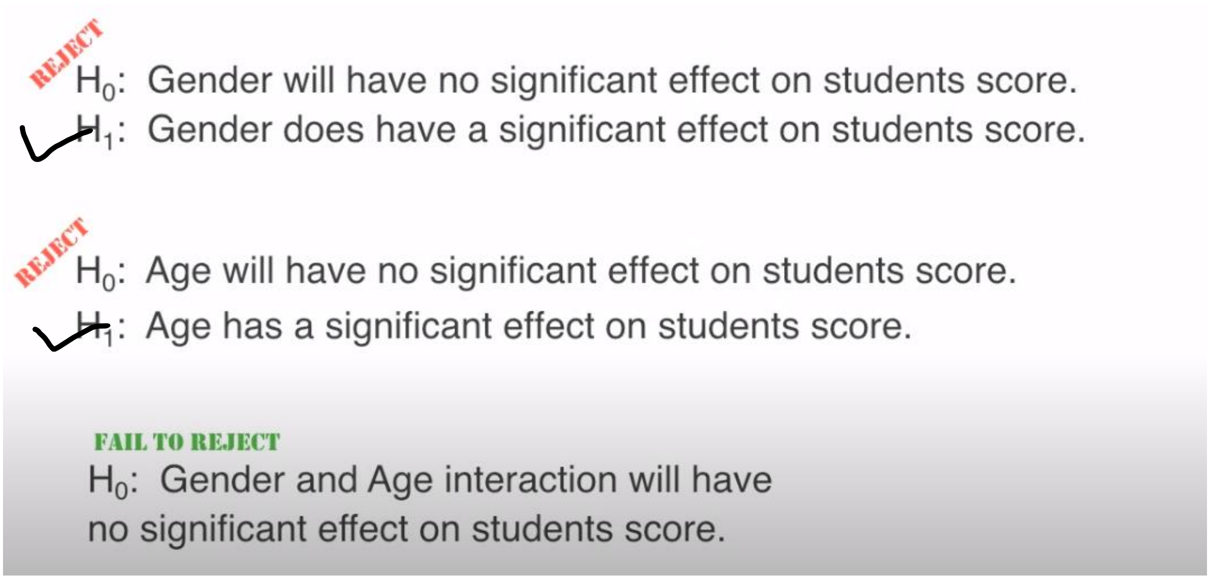
According to F table the F value pr critical value for df (2,12) is 3.89

0.62 < 3.89

Therefore 0.62 not in rejection region

So accept the null hypothesis.

FINAL RESULT



So gender have effect on score , age have effect on score but not together (means both gender and age effect the score) because there is no enough evidence to reject the third null hypothesis.

Links:

[Introduction to Two Way ANOVA (Factorial Analysis) - YouTube]

[How to Calculate a Two Way ANOVA (factorial analysis) - YouTube]

[How to Interpret the Results of A Two Way ANOVA (Factorial) - YouTube]