Topic – ANOVA (One way & two way)

Guided By- Sanket B
Github ID : bsanketm

ANOVA

- A statistical method in which the variation in a set of observations is divided into distinct components.

- i.e. – explained variation and unexplained variation (due to random/chance factor).

- The explained variation have a statistical influence on the given data set, while the random factors do not.

ASSUMPTIONS :

These following are the primary assumptions in ANOVA

1. The responses for each level have a normal population distribution.
2. All population's have same variance.
3. All samples are drawn independently for each other.
4. Within each sample, the observations are sampled randomly and independently of each other.
5. Factor effects are additive.

WHY ANOVA ?

- Analysts use the ANOVA test to determine the influence that independent variables have on the dependent variable in a regression study.

- To determine if there is a statistically significant difference between two or more categorical groups by testing for differences of means using variance.

One way ANOVA

▶ The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups with one dependent variable. Two population testing in statistics can be performed using an ANOVA test.

▶ ex-Testing the relationship between shoe brand (Nike, Adidas, Bata, puma) and race finish times by a person in a marathon.

▶ Here variable is shoe brand.

SCENARIO I

- A logical reasoning test was conducted on students of a college from different fields of studies (FOS).(Arts , Commerce , Science)

- The following are the resultant marks obtained by the students in a the test.

Raw data	
ARTS	
	4
	6
	8
	4
	8
	9
COMMERCE	
	6
	6
	9
	7
	10
	13
SCIENCE	
	8
	9
	13
	12
	14
	16

•Here is Akash, a statistician, who wants to find out whether the field of study (FOS) had any significance on marks of the students in this test.
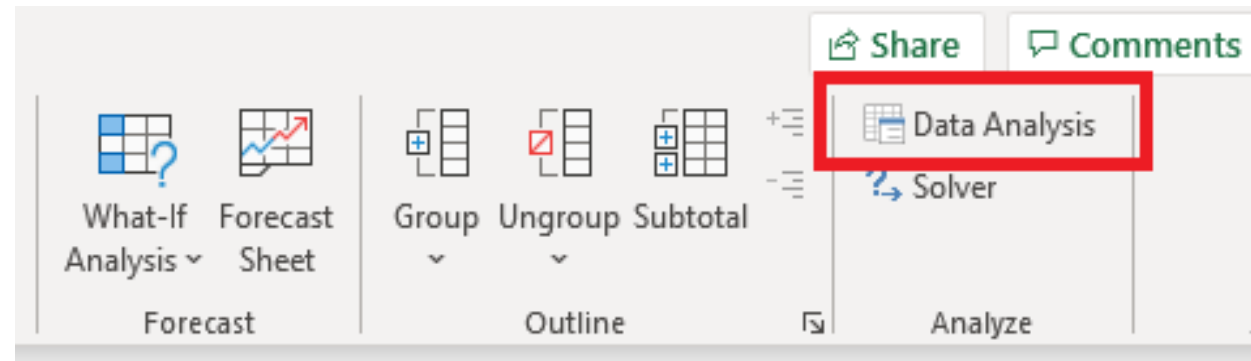
Hi I am Akash

STEPS :

1.First, we need to arrange our data appropriately in order to do one-way analysis of variance in Excel.

2.We can run the ANOVA model once we have organized our data in this way , Open Data Analysis under the Data tab.
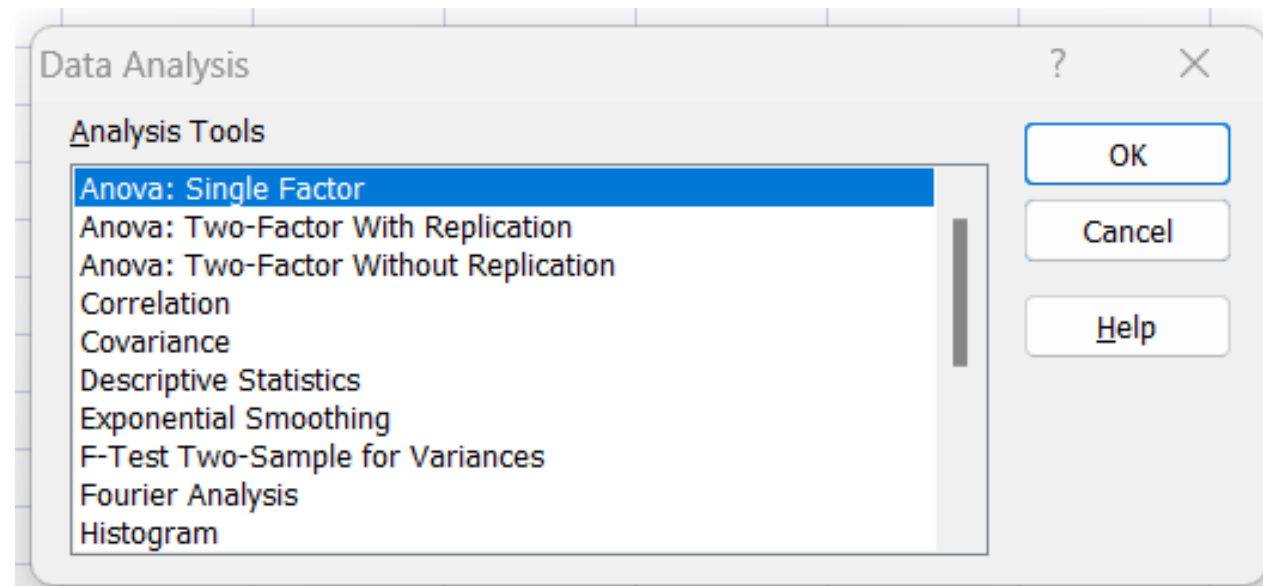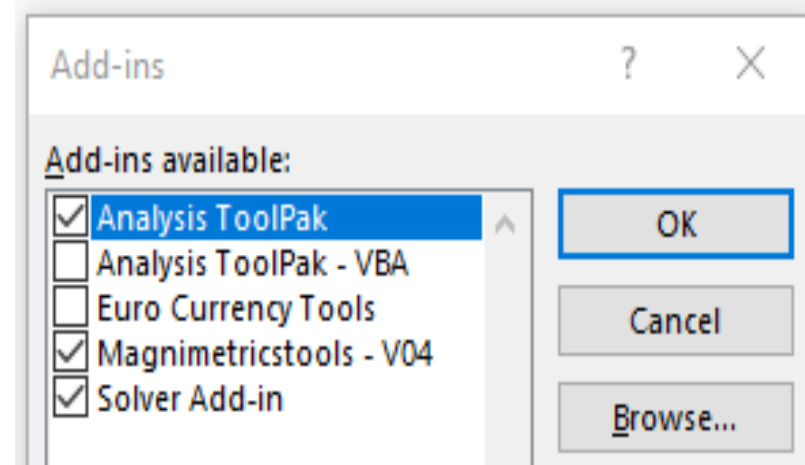
	A	B	C
1	ARTS	COMMERCE	SCIENCE
2	4	6	8
3	6	6	9
4	8	9	13
5	4	7	12
6	8	10	14
7	9	13	16

3.Remember to enable Analysis Tool -Pack in the Add-Ins preferences if you don't see it there.

4.We're going to choose ANOVA: Single Factor (One-Way) from the Data Analysis menu.

7.P-value equals 0.009176. Given that this is less than the 0.05 Alpha threshold, the outcome is statistically significant.

Conclusion – We conclude that field of studies was significant factor to score good marks by the students in logical reasoning test.

Anova: Single Factor

SUMMARY

Groups	Count	Sum	Average	Variance
ARTS	6	39	6.5	4.7
COMMERCE	6	51	8.5	7.5
SCIENCE	6	72	12	9.2

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	93	2	46.5	6.518692	0.009176	3.68232
Within Groups	107	15	7.133333			
Total	200	17				

NOTE - For further comparing the different FOS and to see if they differ significantly with each other by finding critical difference .

Uses of one way ANOVA

- IN RESEARCH , EX - THE RESEARCHERS CAN TAKE NOTE OF THE SUGAR LEVELS BEFORE AND AFTER MEDICATION AND AFTER MEDICATION FOR EACH MEDICINE AND THEN TO UNDERSTAND WHETHER THERE IS A STATISTICALLY SIGNIFICANT DIFFERENCE IN THE MEAN RESULTS FROM THE MEDICATIONS.

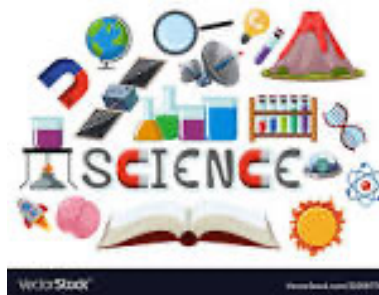- IN AGRICULTURE , EX - FARMER CAN FIND OUT WHICH OF THE DIFFERENT FERTILIZERS LEADS TO THE HIGHEST CROP YIELD.

Two way ANOVA (Without Replication)

▶ A statistical test to assess the difference between the means of more than two groups is the two-way analysis of variance (ANOVA). It is also referred to as a two-factor Factorial ANOVA.

▶ ex- Testing the relationship between shoe brand (Nike, Adidas, Puma, Bata), runner age group (junior, senior, master's), and race finishing times in a marathon.

▶ Here there are two variables (factors) shoe brand and runner age group

SCENARIO II

• Akash further wanted to know whether the "college/school" a new factor and "FOS" had any significance in scoring good marks. So he took the marks of a student from each school and each FOS.
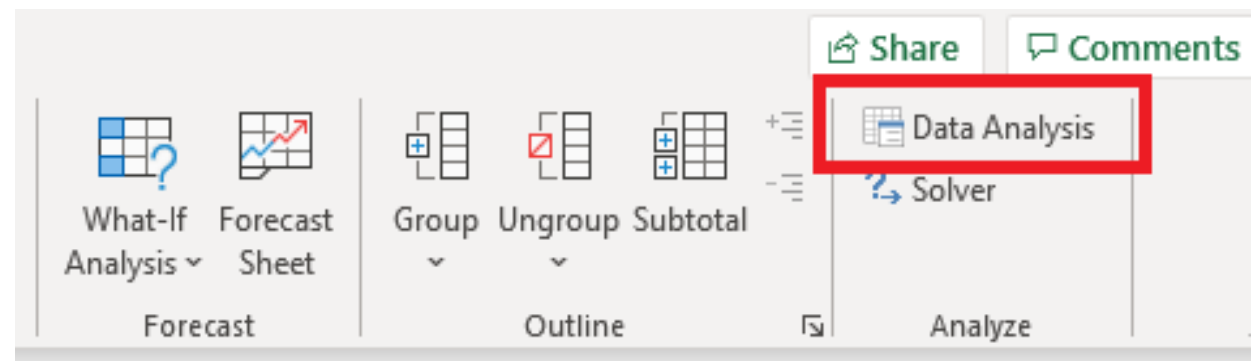
STEPS:

1.First, we need to arrange our data appropriately in order to do a two-factor analysis of variance in Excel.

2.We can run the ANOVA model once we have organized our data in this way and made sure that there are an equal number of observations for each combination of the two categories. Open Data Analysis under the Data tab.

	A	B	C	D
1	School	ARTS	COMMERCE	SCIENCE
2	A	4	6	8
3	B	6	6	9
4	C	8	9	13
5	D	4	7	12
6	E	8	10	14
7	F	9	13	16
8				

Share   Comments

What-If Analysis   Forecast Sheet   Group   Ungroup   Subtotal   Data Analysis   Solver
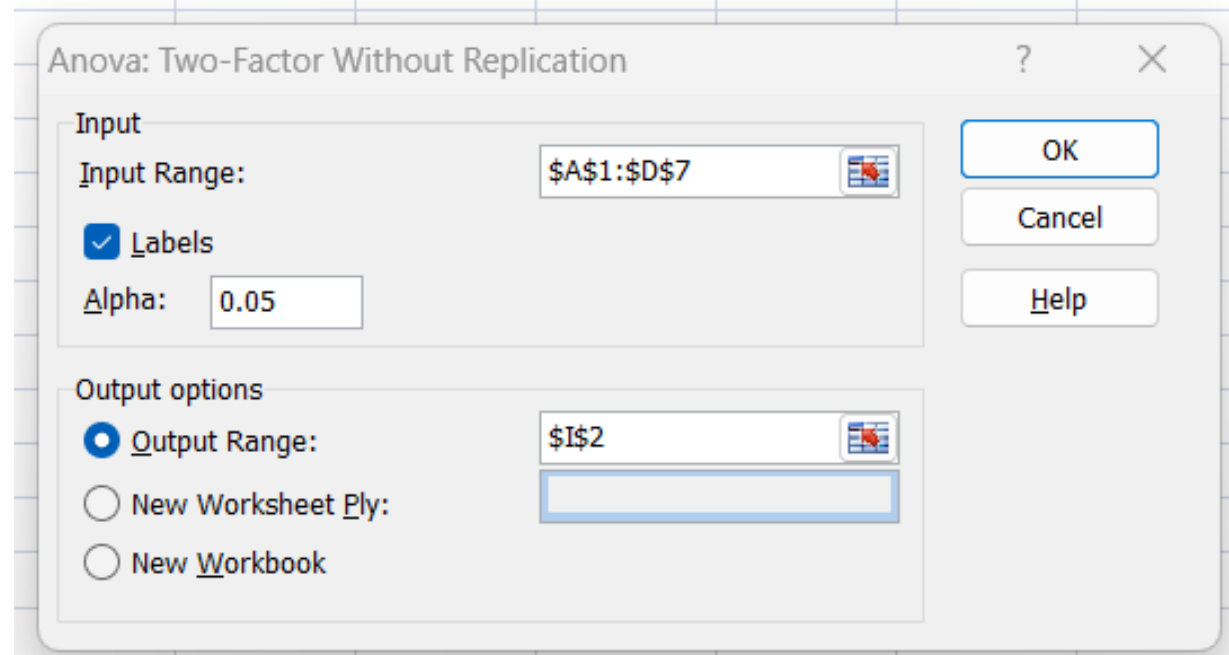
Forecast   Outline   Analyze

3.Remember to enable Analysis Tool-Pack in the Add-Ins preferences if you don't see it there.

4.We're going to choose ANOVA: Two-Factor Without Replication from the Data Analysis menu.
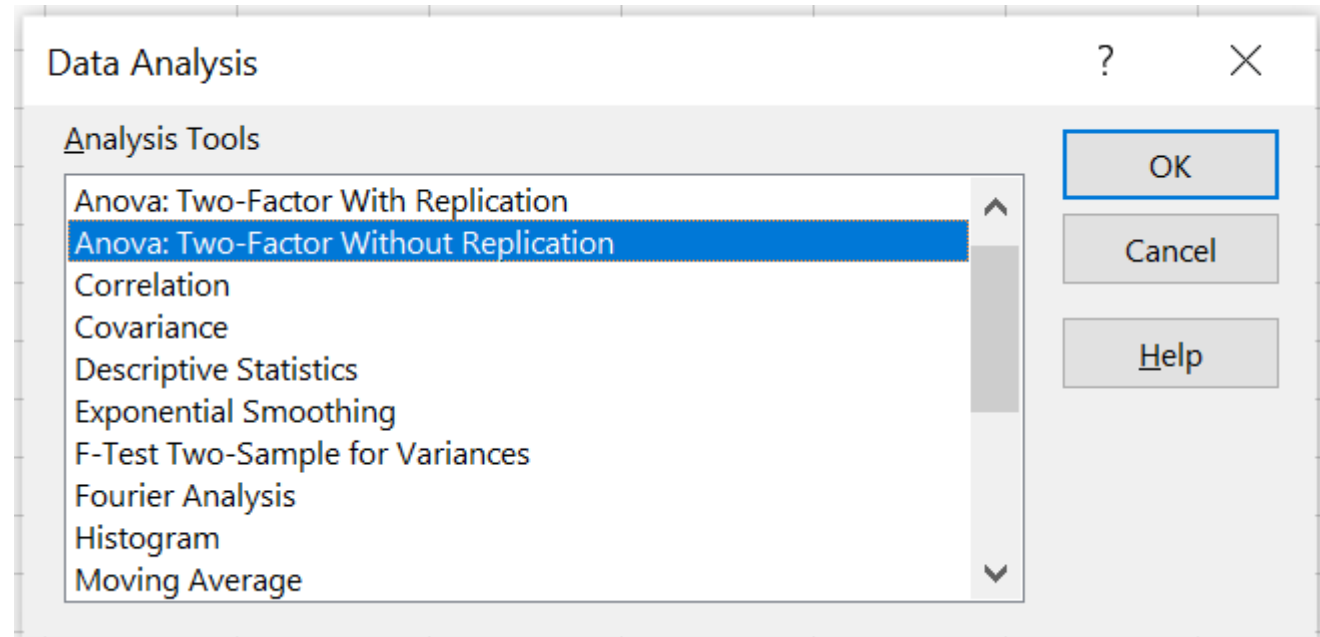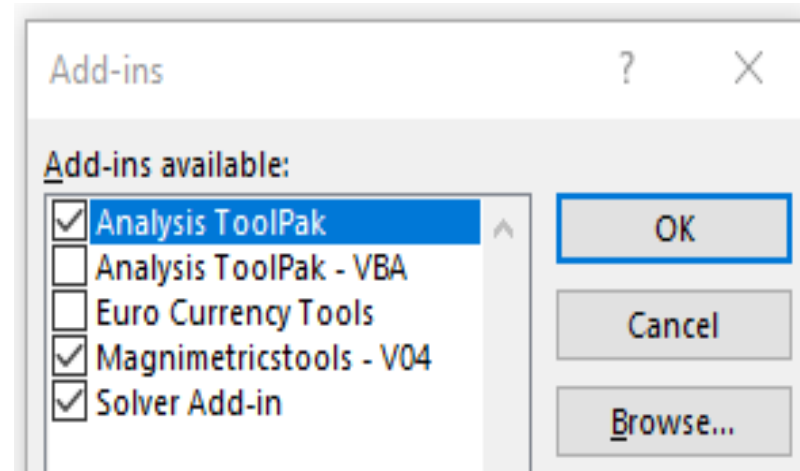
5.The screen after that requests that we enter certain information. Choose the input range first . You can leave the Alpha value at 0.05, which is often what we do in financial modelling and analysis.

6.Once you click OK, Excel will create a output with all the necessary statistical data for the Two-Way ANOVA (without replication) model.

7.P-value equals 0.000118. Given that this is less than the 0.05 Alpha threshold, and $F(17.45455) > Fcrit(3.325835)$. The outcome is statistically significant .

8. P-value equals 0.0000132. Given that this is less than the 0.05 Alpha threshold and $F(42.27273) > Fcrit(4.102821)$ the outcome is statistically significant.

Conclusion – We concluded that field of studies and which College the student went both were significant factors in determining the marks of the students in the logical reasoning test.

Anova: Two-Factor Without Replication

SUMMARY	Count	Sum	Average	Variance
A	3	18	6	4
B	3	21	7	3
C	3	30	10	7
D	3	23	7.666667	16.33333
E	3	32	10.66667	9.333333
F	3	38	12.66667	12.33333
ARTS	6	39	6.5	4.7
COMMERCE	6	51	8.5	7.5
SCIENCE	6	72	12	9.2

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Rows	96	5	19.2	17.45455	0.000118	3.325835
Columns	93	2	46.5	42.27273	1.32E-05	4.102821
Error	11	10	1.1			
Total	200	17				

NOTE – Further we can check whether individual FOS and individual School differ significantly by finding critical difference.
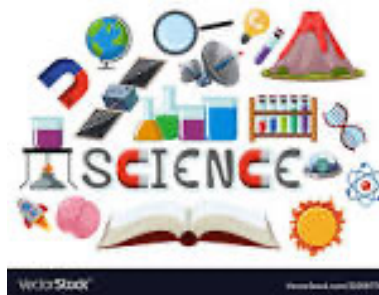
Two way ANOVA (With Replication)

▶ It is also a statistical test to assess the difference between the means of more than two groups is the two-way analysis of variance (ANOVA). Unlike ANOVA(without replication)where the experiment is replicated only once, here the experiment is replicated multiple times.

▶ ex- Testing the relationship between shoe brand (Nike, Adidas, Puma, Bata), 3 people of each runner age group (junior, senior, master's), and race finishing times in a marathon.

▶ Here there are two variables (factors) shoe brand and runner age group , replication is 3 times (3 runners of each age group).

SCENARIO III

- Akash wasn't satisfied with the result, he wanted more accurate design so he took 3 students from each school and each FOS (replicated the treatment three times).
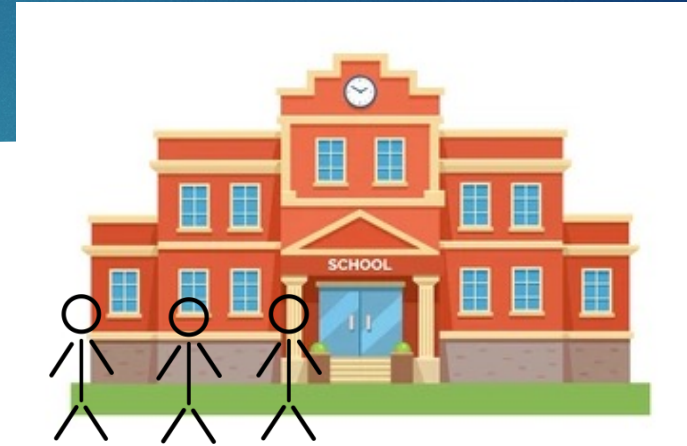
STEPS:

1.First, we need to arrange our data appropriately in order to do a two-factor analysis of variance in Excel.

2.We can run the ANOVA model once we have organized our data in this way and made sure that there are an equal number of observations for each combination of the two categories. Open Data Analysis under the Data tab.
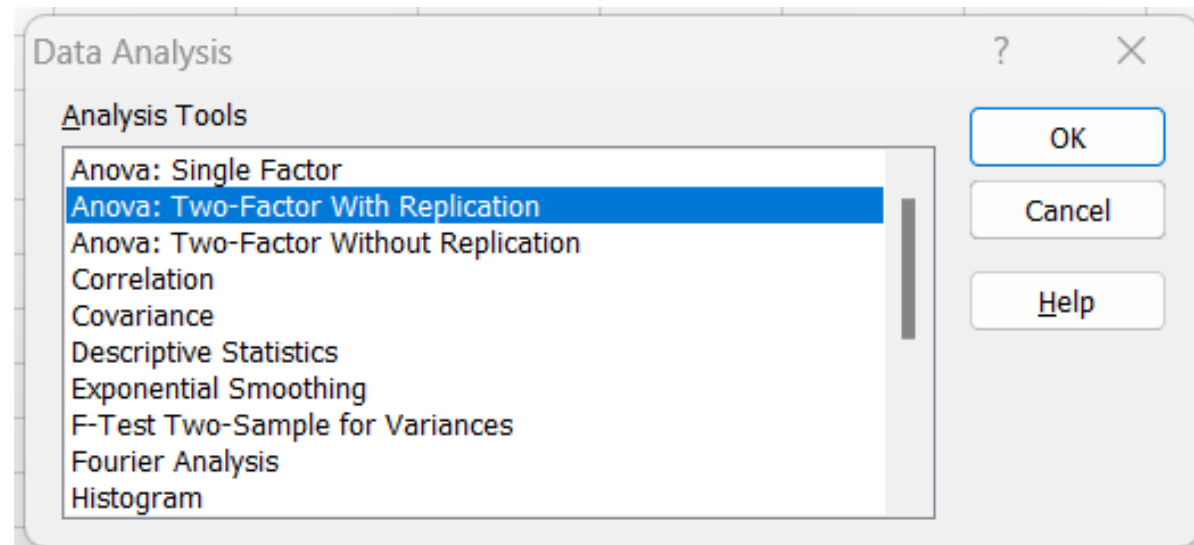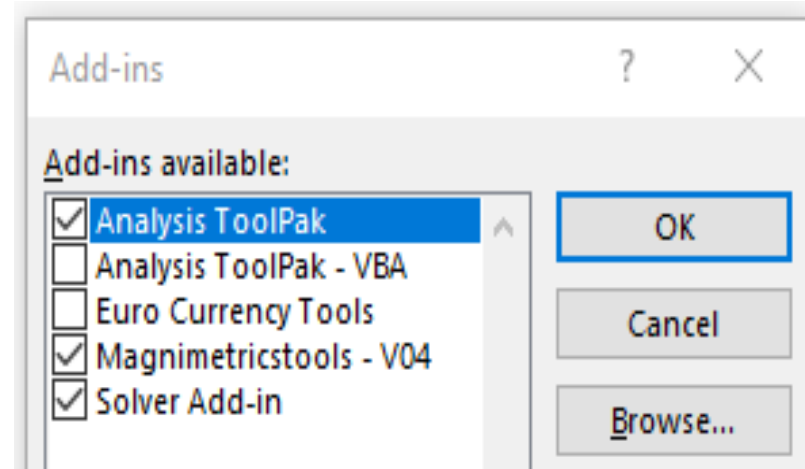
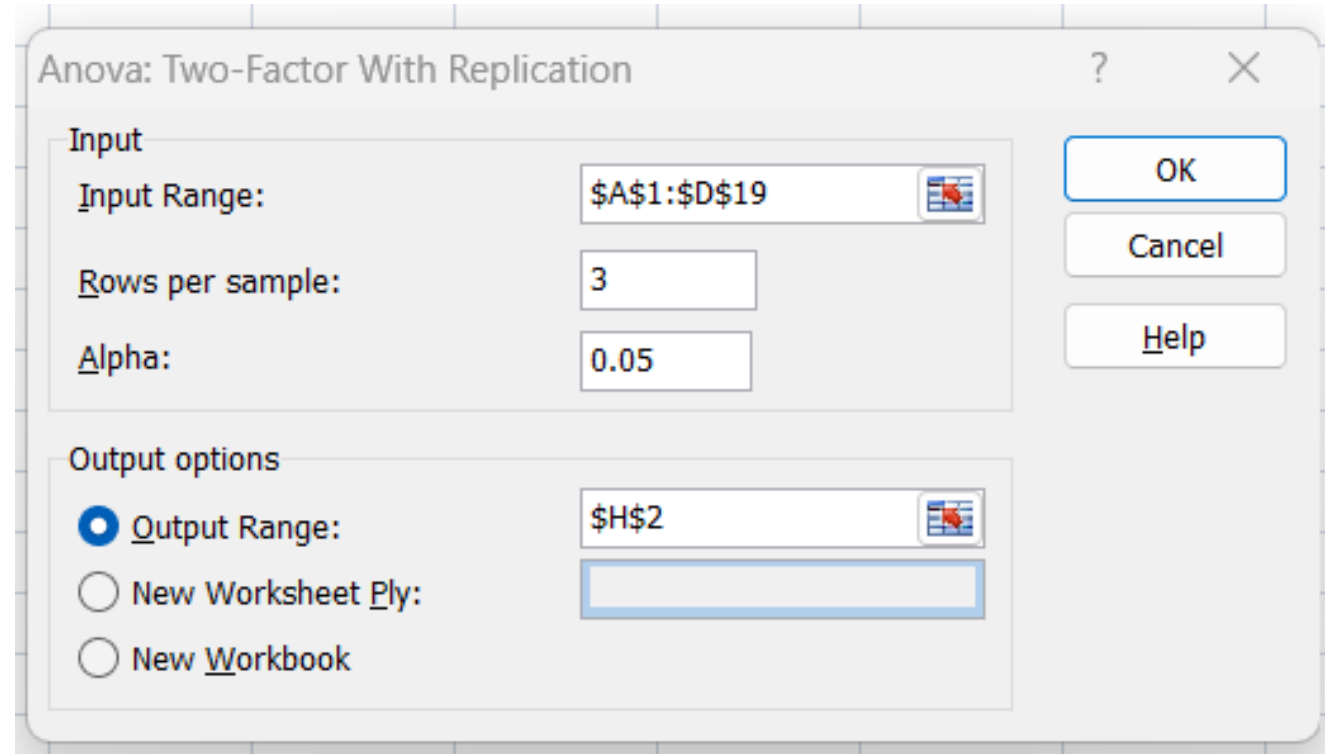|	A	B	C	D
1	School	ARTS	COMMERCE	SCIENCE
2	A	4	6	8
3	A	8	9	6
4	A	7	4	5
5	B	6	6	9
6	B	3	5	7
7	B	3	7	7
8	C	8	9	13
9	C	6	14	15
10	C	6	11	11
11	D	4	7	12
12	D	18	7	8
13	D	12	5	13
14	E	8	10	14
15	E	6	5	6
16	E	4	5	7
17	F	9	13	16
18	F	4	9	8
19	F	10	6	6
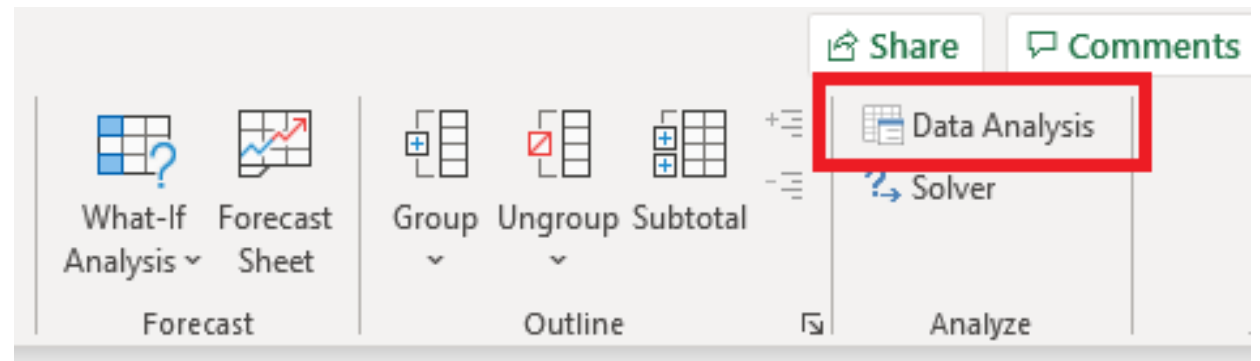
3.Remember to enable Analysis Tool-Pack in the Add-Ins preferences if you don't see it there.

4.We're going to choose ANOVA: Two-Factor With Replication from the Data Analysis menu.

5.The screen after that requests that we enter certain information. Choose the input range first. Then, keep in mind to correctly include the Rows per sample number, which represents the quantity of data within each factor combination You can leave the Alpha value at 0.05, which is often what we do in financial modelling and analysis.

6.Once you click OK, Excel will create a output with all the necessary statistical data for the Two-Way ANOVA : with replication model.

Anova: Two-Factor With Replication   ?   ✕

Input

Input Range: $A$1:$D$19

Rows per sample: 3

Alpha: 0.05

OK

Cancel

Help

Output options

● Output Range: $H$2

○ New Worksheet Ply:

○ New Workbook

7. In sample we see that P-value equals 0.01753 which is less than the 0.05 Alpha threshold and F(3.18625)>Fcrit(2.47717),I.e. the outcome is statistically significant, this means there is significant effect of just college as a own variable A,B,C,D,E (schools differ significantly) .

8. In columns we see that P-value equals 0.05257 which is greater than the 0.05 Alpha threshold and F(3.20039)<Fcrit(3.25945) the outcome is statistically significant, this means there is significant effect of FOS as a own variable (FOS do not differ significantly) .

9. In sample we see that P-value equals 0.4557 which is less than 0.05 Alpha threshold and F(1.00786)<Fcrit(2.10605) , i.e. the outcome is statistically significant , this means there is significant effect of the scores of the students from different college with respect to the FOS (interaction effect of the those variable is significant).

Conclusion – scores depend on which college you go to and also at which college you chose to do your FOS.

Anova: Two-Factor With Replication

SUMMARY		ARTS	COMMER	SCIENCE	Total
A					
Count		3	3	3	9
Sum		19	19	19	57
Average		6.3333	6.3333	6.3333	6.3333
Variance		4.3333	6.3333	2.3333	3.25
B					
Count		3	3	3	9
Sum		12	18	23	53
Average		4	6	7.6667	5.8889
Variance		3	1	1.3333	3.8611
C					
Count		3	3	3	9
Sum		20	34	39	93
Average		6.6667	11.333	13	10.333
Variance		1.3333	6.3333	4	11
D					
Count		3	3	3	9
Sum		34	19	33	86
Average		11.333	6.3333	11	9.5556
Variance		49.333	1.3333	7	20.278
E					
Count		3	3	3	9
Sum		18	20	27	65
Average		6	6.6667	9	7.2222
Variance		4	8.3333	19	9.6944
F					
Count		3	3	3	9
Sum		23	28	30	81
Average		7.6667	9.3333	10	9
Variance		10.333	12.333	28	13.75
Total					
Count		18	18	18	
Sum		126	138	171	
Average		7	7.6667	9.5	
Variance		13.765	8.3529	12.265	

ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Sample	150.167	5	30.0333	3.18625	0.01753	2.47717
Columns	60.3333	2	30.1667	3.20039	0.05257	3.25945
Interaction	95	10	9.5	1.00786	0.4557	2.10605
Within	339.333	36	9.42593			
Total	644.833	53				

Uses of two way ANOVA

• IN AGRICULTURAL FIELD , EX - FARMER CAN TEST THE EFFECT OF FERTILIZER AND WATERING FREQUENCY ON THE CROP YIELD.

• IN BUSINESS , EX - ALLOWS COMPANY TO COMPARE WORKER PRODUCTIVITY BASED ON TWO INDEPENDENT VARIABLES SUCH AS DEPARTMENT AND GENDER.

• IN RESEARCH , EX - TESTING OF A MEDICINAL DRUG AND ITS EFFECT BASED ON GENDER AND AGE.

Over all conclusion

•THE MORE FACTORS WE TAKE INTO ACCOUNT MORE OUR MODEL BECOMES ACCURATE TWO WAY(WR)> TWO WAY(WOR)>ONE WAY.

•THE MORE UNITS WE PERFORM THE EXPERIMENT ON THE MORE OUR MODEL BECOMES ACCURATE TWO WAY(WR)> TWO WAY(WOR).

Conclusion – We concluded that field of studies and which College the student went both were significant factors in determining the marks of the students in the logical reasoning test.

One way

Conclusion – We concluded that field of studies and which College the student went both were significant factors in determining the marks of the students in the logical reasoning test.

Two way without replication

Conclusion – scores depend on which college you go to and also at which college you chose to do your FOS.

Two way with replication

Thank you

SPECIAL THANKS TO "AKASH"

Darshan Pagar(TY1G2233703)

Disha Shinde(TY1G2233691)

Rama Mahanavar(TY1G2233668)

Muskan Mansuri(TY1G2233670)

Anurag Surve(TY1G2233695)