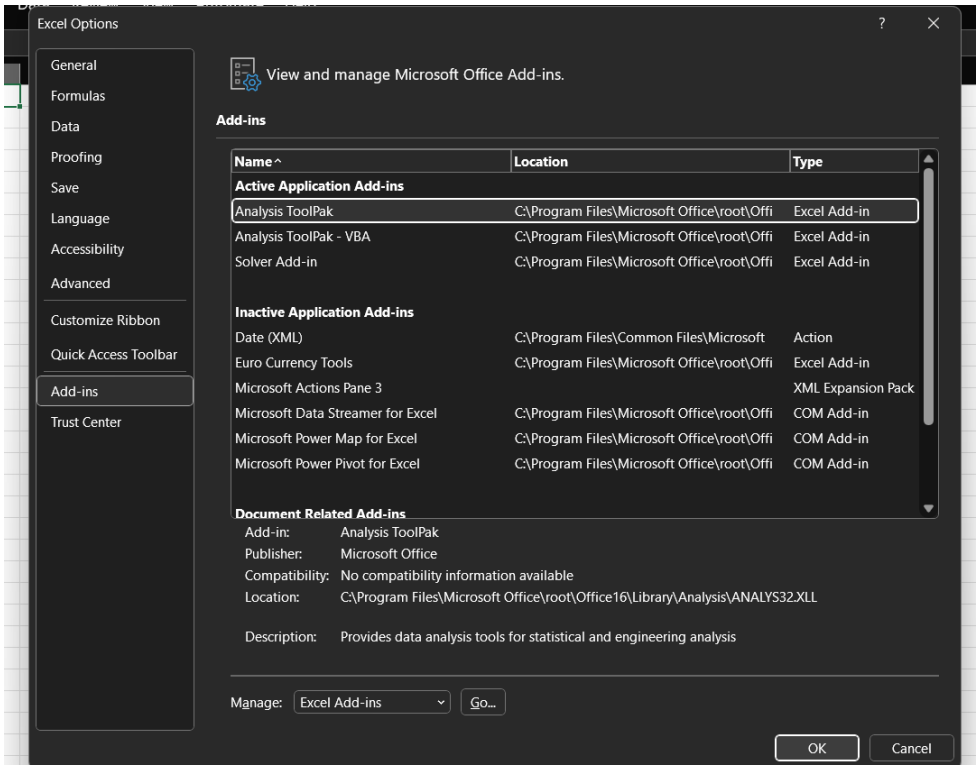
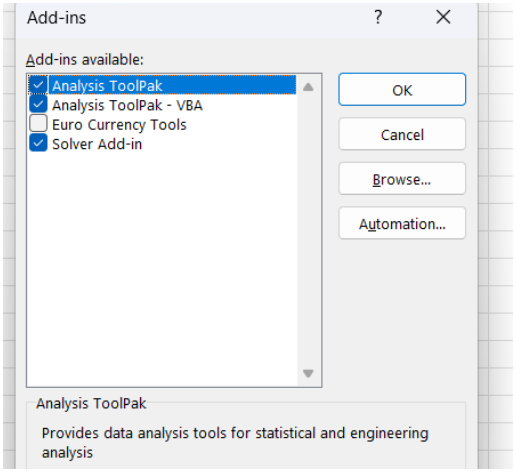
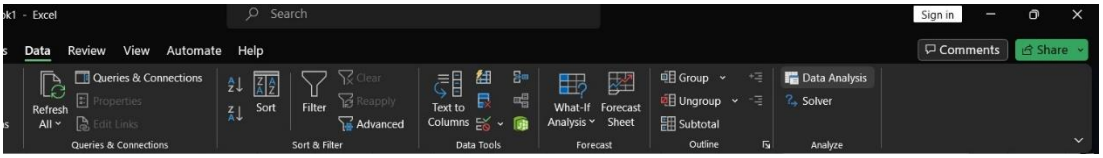
Practical: 01

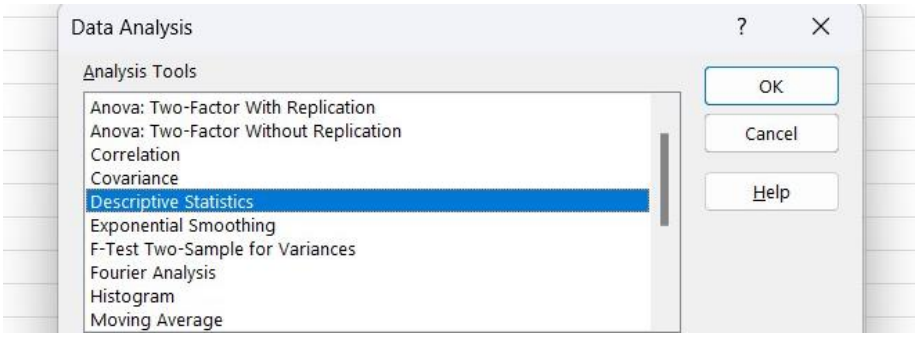
A. Write a program for obtaining descriptive statistics of data.

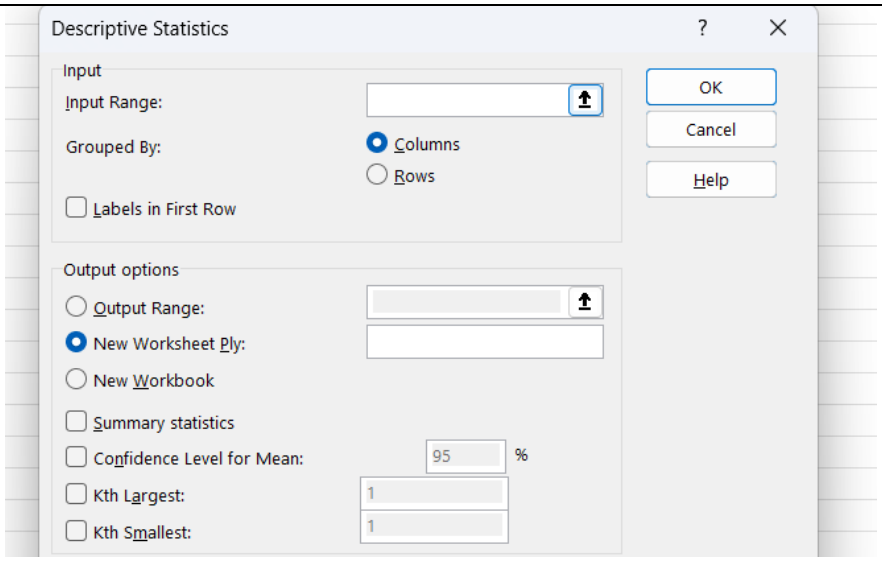
Go to File Menu > Options Add-Ins > Select Analysis ToolPak > press Ok.



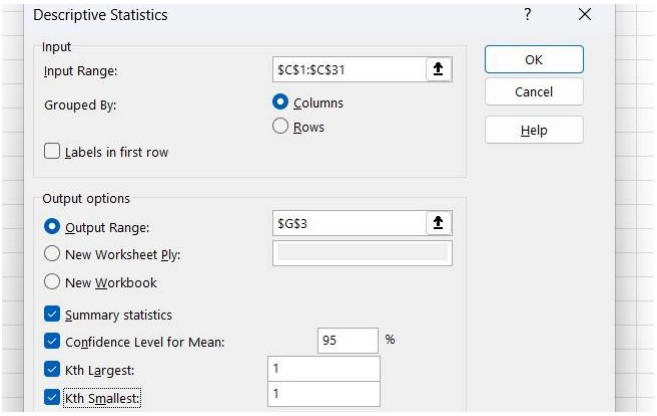




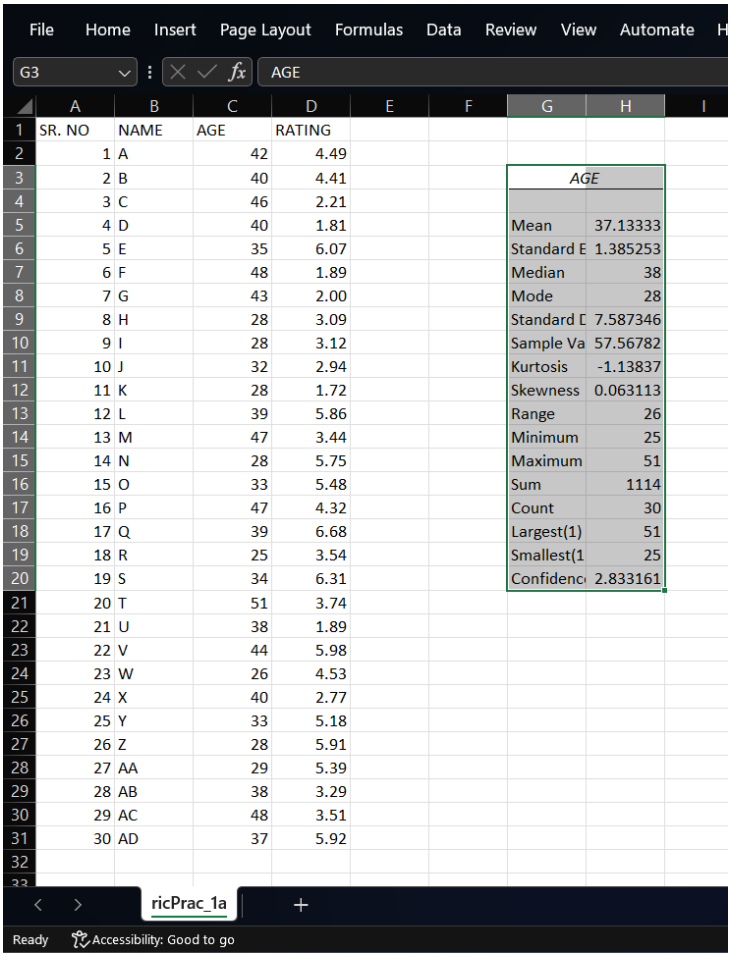




Select the data range from excel worksheet.



Output:



File Home Insert Page Layout Formulas Data Review View Automate

G3 | AGE

	A	B	C	D	E	F	G	H	I
1	SR. NO	NAME	AGE	RATING					
2	1	A	42	4.49					
3	2	B	40	4.41			AGE		
4	3	C	46	2.21					
5	4	D	40	1.81			Mean	37.13333	
6	5	E	35	6.07			Standard E	1.385253	
7	6	F	48	1.89			Median	38	
8	7	G	43	2.00			Mode	28	
9	8	H	28	3.09			Standard D	7.587346	
10	9	I	28	3.12			Sample Va	57.56782	
11	10	J	32	2.94			Kurtosis	-1.13837	
12	11	K	28	1.72			Skewness	0.063113	
13	12	L	39	5.86			Range	26	
14	13	M	47	3.44			Minimum	25	
15	14	N	28	5.75			Maximum	51	
16	15	O	33	5.48			Sum	1114	
17	16	P	47	4.32			Count	30	
18	17	Q	39	6.68			Largest(1)	51	
19	18	R	25	3.54			Smallest(1	25	
20	19	S	34	6.31			Confidenc	2.833161	
21	20	T	51	3.74					
22	21	U	38	1.89					
23	22	V	44	5.98					
24	23	W	26	4.53					
25	24	X	40	2.77					
26	25	Y	33	5.18					
27	26	Z	28	5.91					
28	27	AA	29	5.39					
29	28	AB	38	3.29					
30	29	AC	48	3.51					
31	30	AD	37	5.92					
32									

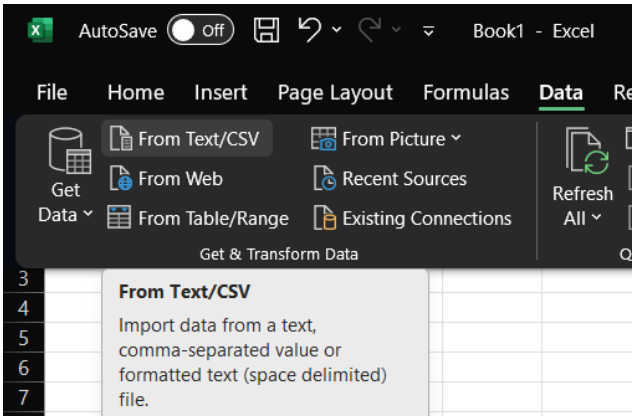
ricPrac_1a

Ready Accessibility: Good to go

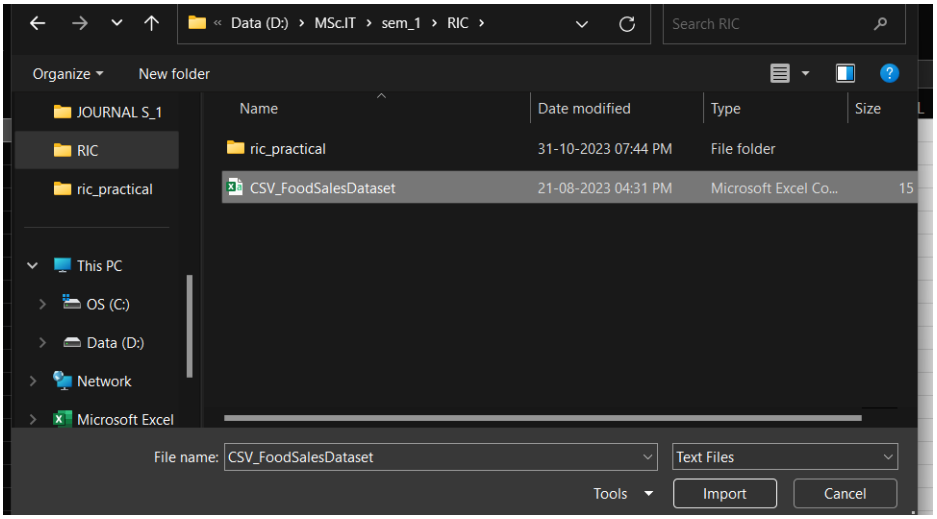
Practical: 02

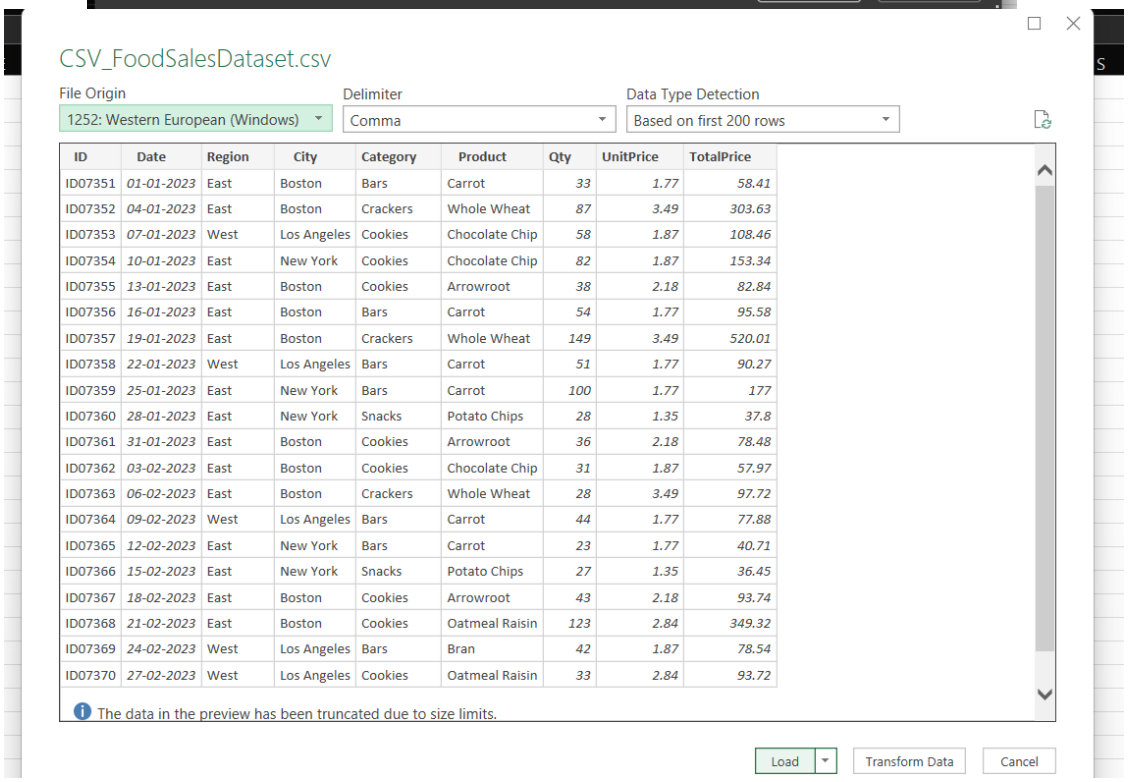
➢ Import data from different data sources.

Go to Data > click on “From Text/CSV”.



Select file you require > click on import





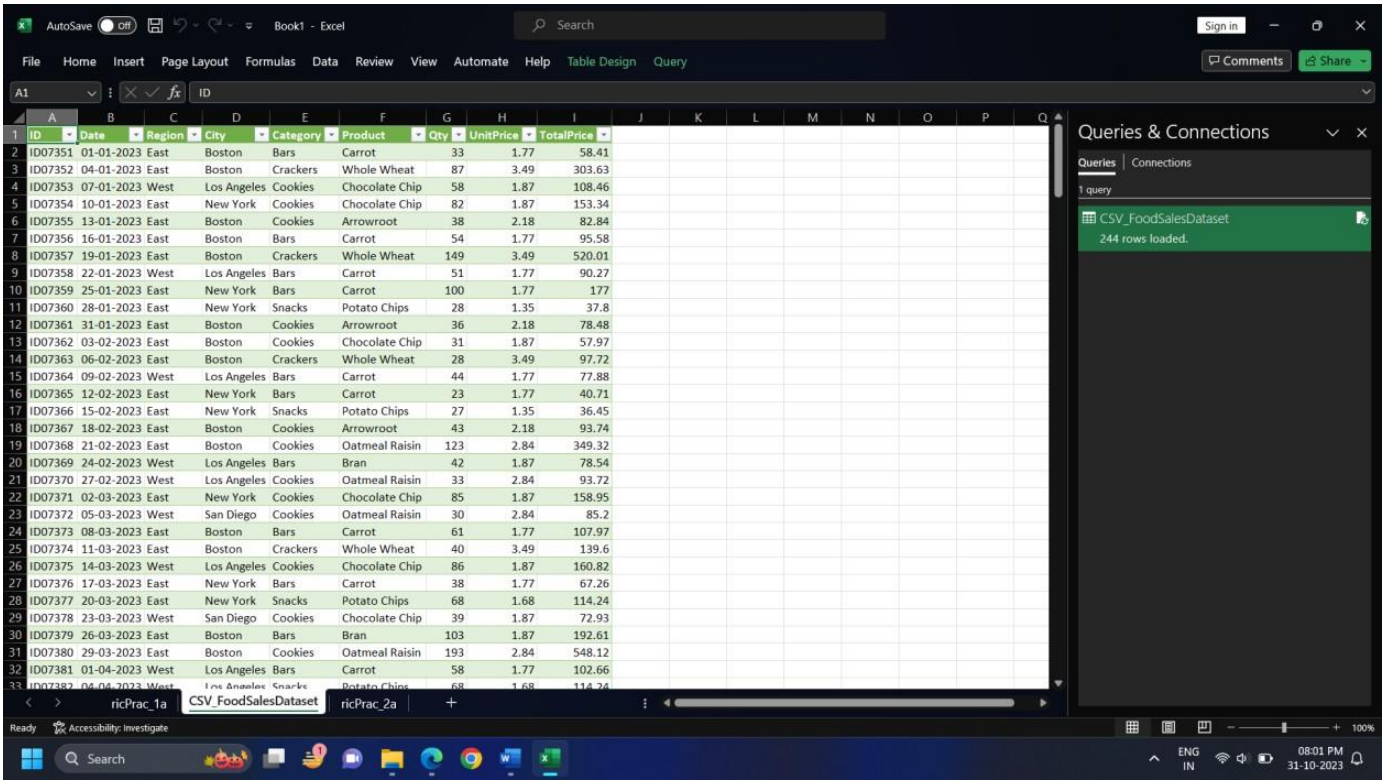
CSV_FoodSalesDataset.csv

File Origin: 1252: Western European (Windows) | Delimiter: Comma | Data Type Detection: Based on first 200 rows

ID	Date	Region	City	Category	Product	Qty	UnitPrice	TotalPrice
ID07351	01-01-2023	East	Boston	Bars	Carrot	33	1.77	58.41
ID07352	04-01-2023	East	Boston	Crackers	Whole Wheat	87	3.49	303.63
ID07353	07-01-2023	West	Los Angeles	Cookies	Chocolate Chip	58	1.87	108.46
ID07354	10-01-2023	East	New York	Cookies	Chocolate Chip	82	1.87	153.34
ID07355	13-01-2023	East	Boston	Cookies	Arrowroot	38	2.18	82.84
ID07356	16-01-2023	East	Boston	Bars	Carrot	54	1.77	95.58
ID07357	19-01-2023	East	Boston	Crackers	Whole Wheat	149	3.49	520.01
ID07358	22-01-2023	West	Los Angeles	Bars	Carrot	51	1.77	90.27
ID07359	25-01-2023	East	New York	Bars	Carrot	100	1.77	177
ID07360	28-01-2023	East	New York	Snacks	Potato Chips	28	1.35	37.8
ID07361	31-01-2023	East	Boston	Cookies	Arrowroot	36	2.18	78.48
ID07362	03-02-2023	East	Boston	Cookies	Chocolate Chip	31	1.87	57.97
ID07363	06-02-2023	East	Boston	Crackers	Whole Wheat	28	3.49	97.72
ID07364	09-02-2023	West	Los Angeles	Bars	Carrot	44	1.77	77.88
ID07365	12-02-2023	East	New York	Bars	Carrot	23	1.77	40.71
ID07366	15-02-2023	East	New York	Snacks	Potato Chips	27	1.35	36.45
ID07367	18-02-2023	East	Boston	Cookies	Arrowroot	43	2.18	93.74
ID07368	21-02-2023	East	Boston	Cookies	Oatmeal Raisin	123	2.84	349.32
ID07369	24-02-2023	West	Los Angeles	Bars	Bran	42	1.87	78.54
ID07370	27-02-2023	West	Los Angeles	Cookies	Oatmeal Raisin	33	2.84	93.72

The data in the preview has been truncated due to size limits.

Load | Transform Data | Cancel

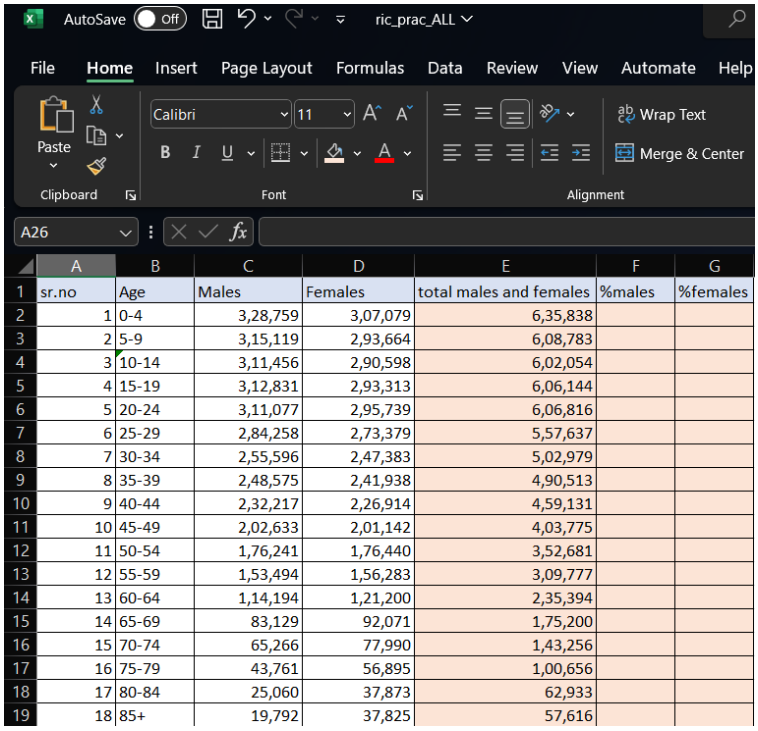


ID	Date	Region	City	Category	Product	Qty	UnitPrice	TotalPrice
ID07351	01-01-2023	East	Boston	Bars	Carrot	33	1.77	58.41
ID07352	04-01-2023	East	Boston	Crackers	Whole Wheat	87	3.49	303.63
ID07353	07-01-2023	West	Los Angeles	Cookies	Chocolate Chip	58	1.87	108.46
ID07354	10-01-2023	East	New York	Cookies	Chocolate Chip	82	1.87	153.34
ID07355	13-01-2023	East	Boston	Cookies	Arrowroot	38	2.18	82.84
ID07356	16-01-2023	East	Boston	Bars	Carrot	54	1.77	95.58
ID07357	19-01-2023	East	Boston	Crackers	Whole Wheat	149	3.49	520.01
ID07358	22-01-2023	West	Los Angeles	Bars	Carrot	51	1.77	90.27
ID07359	25-01-2023	East	New York	Bars	Carrot	100	1.77	177
ID07360	28-01-2023	East	New York	Snacks	Potato Chips	28	1.35	37.8
ID07361	31-01-2023	East	Boston	Cookies	Arrowroot	36	2.18	78.48
ID07362	03-02-2023	East	Boston	Cookies	Chocolate Chip	31	1.87	57.97
ID07363	06-02-2023	East	Boston	Crackers	Whole Wheat	28	3.49	97.72
ID07364	09-02-2023	West	Los Angeles	Bars	Carrot	44	1.77	77.88
ID07365	12-02-2023	East	New York	Bars	Carrot	23	1.77	40.71
ID07366	15-02-2023	East	New York	Snacks	Potato Chips	27	1.35	36.45
ID07367	18-02-2023	East	Boston	Cookies	Arrowroot	43	2.18	93.74
ID07368	21-02-2023	East	Boston	Cookies	Oatmeal Raisin	123	2.84	349.32
ID07369	24-02-2023	West	Los Angeles	Bars	Bran	42	1.87	78.54
ID07370	27-02-2023	West	Los Angeles	Cookies	Oatmeal Raisin	33	2.84	93.72
ID07371	02-03-2023	East	New York	Cookies	Chocolate Chip	85	1.87	158.95
ID07372	05-03-2023	West	San Diego	Cookies	Oatmeal Raisin	30	2.84	85.2
ID07373	08-03-2023	East	Boston	Bars	Carrot	61	1.77	107.97
ID07374	11-03-2023	East	Boston	Crackers	Whole Wheat	40	3.49	139.6
ID07375	14-03-2023	West	Los Angeles	Cookies	Chocolate Chip	86	1.87	160.82
ID07376	17-03-2023	East	New York	Bars	Carrot	38	1.77	67.26
ID07377	20-03-2023	East	New York	Snacks	Potato Chips	68	1.68	114.24
ID07378	23-03-2023	West	San Diego	Cookies	Chocolate Chip	39	1.87	72.93
ID07379	26-03-2023	East	Boston	Bars	Bran	103	1.87	192.61
ID07380	29-03-2023	East	Boston	Cookies	Oatmeal Raisin	193	2.84	548.12
ID07381	01-04-2023	West	Los Angeles	Bars	Carrot	58	1.77	102.66
ID07382	04-04-2023	West	Los Angeles	Snacks	Potato Chips	68	1.68	114.24

Practical: 03

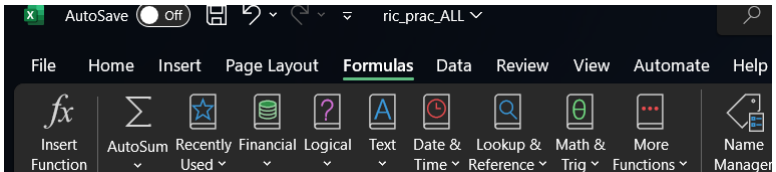
➢ Perform analysis of given secondary data.

Determine your research question - Knowing exactly what you are looking for.
Locating data - Knowing what is out there and whether you can gain access to it. Aquick Internet search, possibly with the help of a librarian, will reveal a wealth of options.
Evaluating relevance of the data - Considering things like the data's original purpose, when it was collected, population, sampling strategy/sample, data collection protocols, operationalization of concepts, questions asked, and form/shape of the data.
Assessing credibility of the data - Establishing the credentials of the original researchers, searching for full explication of methods including any problems encountered, determining how consistent the data is with data from other sources, anddiscovering whether the data has been used in any credible published research.
Analysis - This will generally involve a range of statistical processes.
Example: Analyze the given Population Census Data for Planning and Decision Making byusing the size and composition of populations.



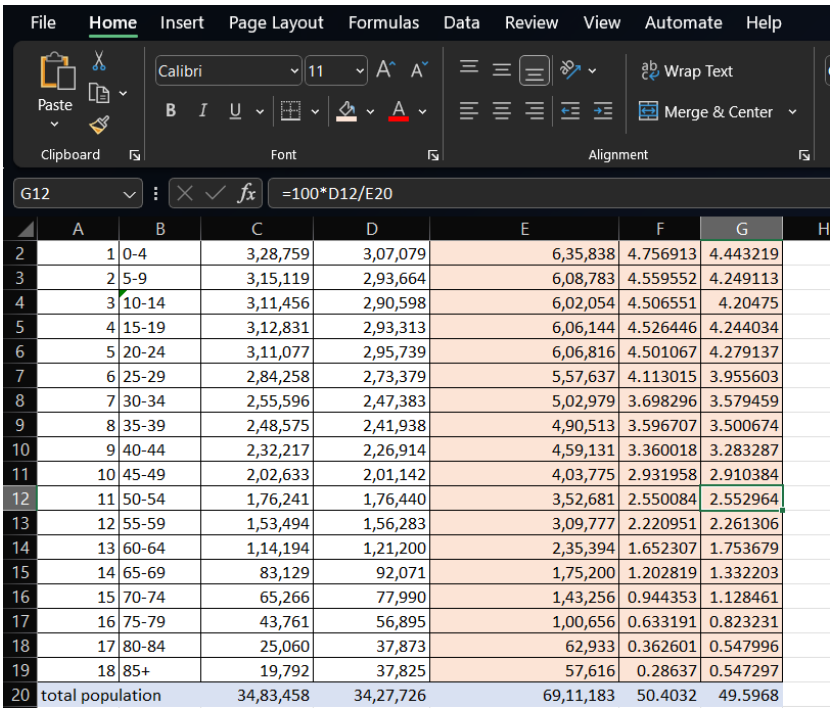
	A	B	C	D	E	F	G
1	sr.no	Age	Males	Females	total males and females	%males	%females
2	1	0-4	3,28,759	3,07,079	6,35,838		
3	2	5-9	3,15,119	2,93,664	6,08,783		
4	3	10-14	3,11,456	2,90,598	6,02,054		
5	4	15-19	3,12,831	2,93,313	6,06,144		
6	5	20-24	3,11,077	2,95,739	6,06,816		
7	6	25-29	2,84,258	2,73,379	5,57,637		
8	7	30-34	2,55,596	2,47,383	5,02,979		
9	8	35-39	2,48,575	2,41,938	4,90,513		
10	9	40-44	2,32,217	2,26,914	4,59,131		
11	10	45-49	2,02,633	2,01,142	4,03,775		
12	11	50-54	1,76,241	1,76,440	3,52,681		
13	12	55-59	1,53,494	1,56,283	3,09,777		
14	13	60-64	1,14,194	1,21,200	2,35,394		
15	14	65-69	83,129	92,071	1,75,200		
16	15	70-74	65,266	77,990	1,43,256		
17	16	75-79	43,761	56,895	1,00,656		
18	17	80-84	25,060	37,873	62,933		
19	18	85+	19,792	37,825	57,616		

Click on cell **A20** and label it as "total population".
Click on **C20** and click on *AutoSum* > a range will be selected *automatically* > clickenter. Similarly follow this for cell **D20** and **E20**.



	A	B	C	D	E	F	G
1	sr.no	Age	Males	Females	total males and females	%males	%females
2	1	0-4	3,28,759	3,07,079	6,35,838		
3	2	5-9	3,15,119	2,93,664	6,08,783		
4	3	10-14	3,11,456	2,90,598	6,02,054		
5	4	15-19	3,12,831	2,93,313	6,06,144		
6	5	20-24	3,11,077	2,95,739	6,06,816		
7	6	25-29	2,84,258	2,73,379	5,57,637		
8	7	30-34	2,55,596	2,47,383	5,02,979		
9	8	35-39	2,48,575	2,41,938	4,90,513		
10	9	40-44	2,32,217	2,26,914	4,59,131		
11	10	45-49	2,02,633	2,01,142	4,03,775		
12	11	50-54	1,76,241	1,76,440	3,52,681		
13	12	55-59	1,53,494	1,56,283	3,09,777		
14	13	60-64	1,14,194	1,21,200	2,35,394		
15	14	65-69	83,129	92,071	1,75,200		
16	15	70-74	65,266	77,990	1,43,256		
17	16	75-79	43,761	56,895	1,00,656		
18	17	80-84	25,060	37,873	62,933		
19	18	85+	19,792	37,825	57,616		
20	total population		34,83,458	34,27,726	69,11,183		

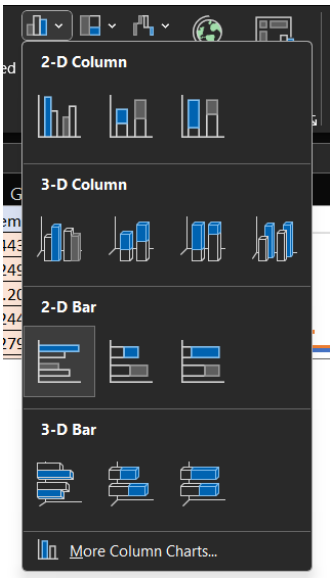
To calculate percent of males in cell F2, enter formula=100*C2/E20. And copy theformula in cell F2 down to cell F19.
Similarly, to calculate the percent of females in cell G2, enter the formula =100*D2/E20.Copy the formula in cell G2 down to cell G19.

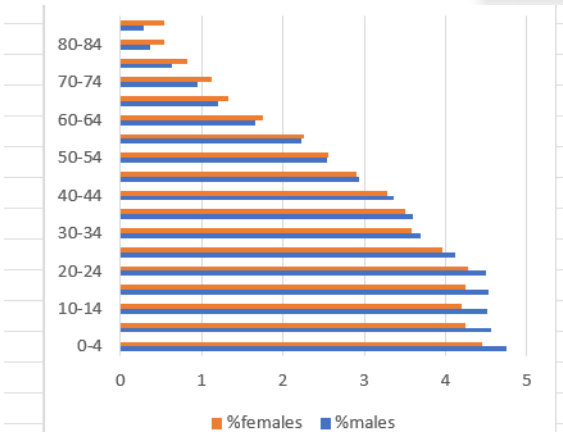


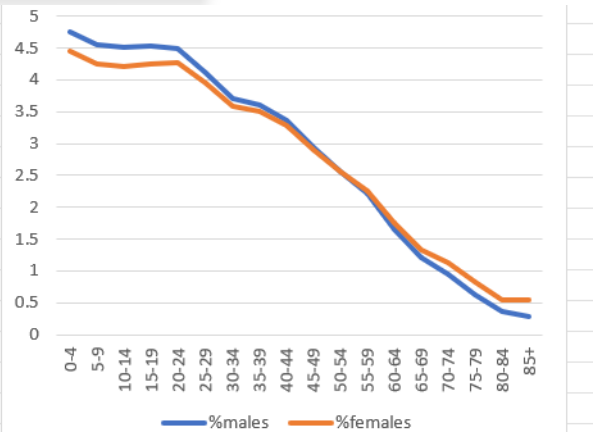
	A	B	C	D	E	F	G
2	1	0-4	3,28,759	3,07,079	6,35,838	4.756913	4.443219
3	2	5-9	3,15,119	2,93,664	6,08,783	4.559552	4.249113
4	3	10-14	3,11,456	2,90,598	6,02,054	4.506551	4.20475
5	4	15-19	3,12,831	2,93,313	6,06,144	4.526446	4.244034
6	5	20-24	3,11,077	2,95,739	6,06,816	4.501067	4.279137
7	6	25-29	2,84,258	2,73,379	5,57,637	4.113015	3.955603
8	7	30-34	2,55,596	2,47,383	5,02,979	3.698296	3.579459
9	8	35-39	2,48,575	2,41,938	4,90,513	3.596707	3.500674
10	9	40-44	2,32,217	2,26,914	4,59,131	3.360018	3.283287
11	10	45-49	2,02,633	2,01,142	4,03,775	2.931958	2.910384
12	11	50-54	1,76,241	1,76,440	3,52,681	2.550084	2.552964
13	12	55-59	1,53,494	1,56,283	3,09,777	2.220951	2.261306
14	13	60-64	1,14,194	1,21,200	2,35,394	1.652307	1.753679
15	14	65-69	83,129	92,071	1,75,200	1.202819	1.332203
16	15	70-74	65,266	77,990	1,43,256	0.944353	1.128461
17	16	75-79	43,761	56,895	1,00,656	0.633191	0.823231
18	17	80-84	25,060	37,873	62,933	0.362601	0.547996
19	18	85+	19,792	37,825	57,616	0.28637	0.547297
20	total population		34,83,458	34,27,726	69,11,183	50.4032	49.5968

We need to show a chart with two series of data (% male and % female) and the age labels incolumn A as the Category X-axis labels. Highlight the range A2:A19, hold down the CTRL key and highlight the range E2:F19 Under inset tab, under horizontal bar charts select clustered bar chart and under line or area chart select line chart.

Output:







Practical: 04

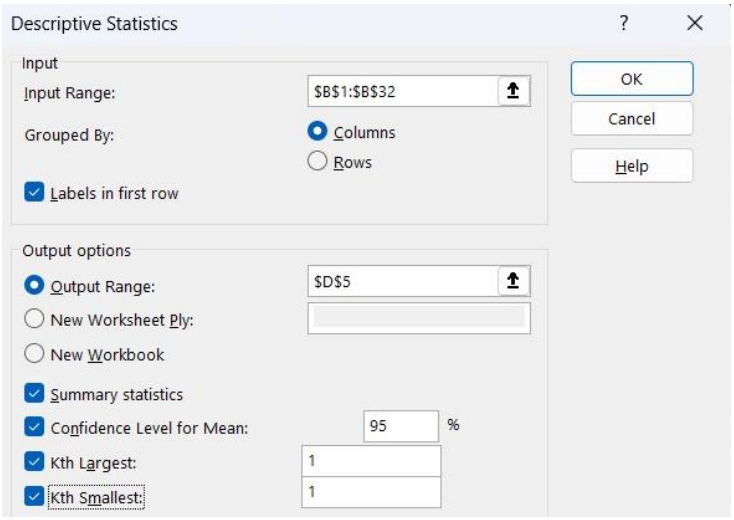
A. Perform testing of hypothesis using one sample t-test.

One sample t-test: The One Sample T Test determines whether the sample mean is statistically different from a known or hypothesized population mean. The One Sample T Testis a parametric test.

No. Of Bars	Protein Values
1	20.7
2	20.75
3	22.14
4	22.12
5	27.46
6	22.91
7	19.56
8	22.15
9	25.34
10	21.1
11	19.85
12	20.33
13	18.04
14	21.29
15	21.54
16	24.12
17	24.75
18	21.08
19	19.95
20	25.06
21	21.39
22	19.72
23	22.44
24	22.33
25	18.28
26	19.08
27	25.79
28	16.26
29	19.88
30	20.53
31	17.46

Here,
H0 is population mean is 20.
H1 is population mean is not 20.

First, we require descriptive statistics for above data.



Protein Values	
Mean	21.4
Standard Error	0.4564972
Median	21.1
Mode	#N/A
Standard Deviation	2.5416687
Sample Variance	6.46008
Kurtosis	0.1453746
Skewness	0.3907659
Range	11.2
Minimum	16.26
Maximum	27.46
Sum	663.4
Count	31
Largest(1)	27.46
Smallest(1)	16.26
Confidence Level(95.0%)	0.9322916

We require Mean, Standard Error, Standard Deviation and Count.
We require to degrees of freedom, hypothesized mean, alpha value, t-statistic, p-value.
We calculate degrees of freedom as function= E19 -1.
Hypothesized mean is 20 and Alpha value is taken as 0.05.
t-statistic is calculated as function = (E7-E25)/E8.
P value is calculatedas function = TDIST (E27, E24, 2)

	D	E
6		
7	Mean	21.4
8	Standard Error	0.4564972
9	Median	21.1
10	Mode	#N/A
11	Standard Deviation	2.5416687
12	Sample Variance	6.46008
13	Kurtosis	0.1453746
14	Skewness	0.3907659
15	Range	11.2
16	Minimum	16.26
17	Maximum	27.46
18	Sum	663.4
19	Count	31
20	Largest(1)	27.46
21	Smallest(1)	16.26
22	Confidence Level(95.0%)	0.9322916
23		
24	degrees of freedom	30
25	hypothesized mean	20
26	alpha value	0.05
27	t-statistic	3.0668316
28	p-value	0.0045526

Then for final step of T-test, we use function =IF(E28>E26,"H0 accepted, H1 rejected","H1accepted")

Output:

	D	E	F
23			
24	degrees of freedom	30	
25	hypothesized mean	20	
26	alpha value	0.05	
27	t-statistic	3.0668316	
28	p-value	0.0045526	
29			
30	H1 accepted		
31			
32			

B. Perform testing of hypothesis using two sample t-test.

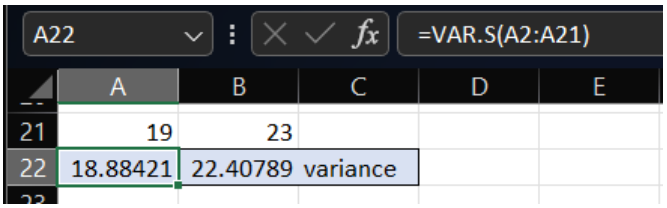
Suppose researchers want to know whether or not two different species of plants in a particular country have the same mean height. Because it would take too long to go around and measure every single plant, they decide to collect a sample of 20 plants from each species.

The following data shows the height (in inches) for each plant in each sample:

species 1	species 2
12	10
12	24
20	17
12	23
24	19
25	16
10	10
15	14
16	19
21	14
21	12
11	21
12	12
16	12
15	11
17	21
17	18
14	23
19	16
19	23

When we conduct a two-sample t-test, we must first decide if we will assume that the two populations have equal or unequal variances. As a rule of thumb, we can assume the populations have equal variances if the ratio of the larger sample variance to the smaller sample variance is less than 4:1.

To calculate the variance of two samples Click on cell A22 and type = VAR.S(A2:A21) Click on cell B22 and type = VAR.S(B2:B21)

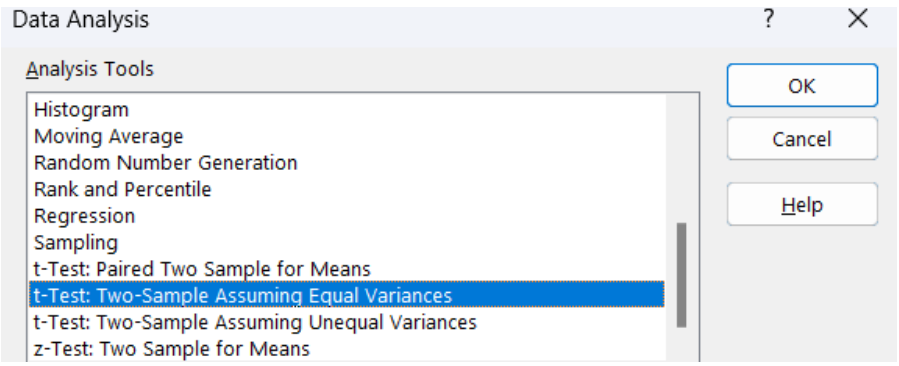


The ratio of the larger sample variance to the smaller sample variance is 1.1865, which is less than 4. This means we can assume that the population variances are equal.

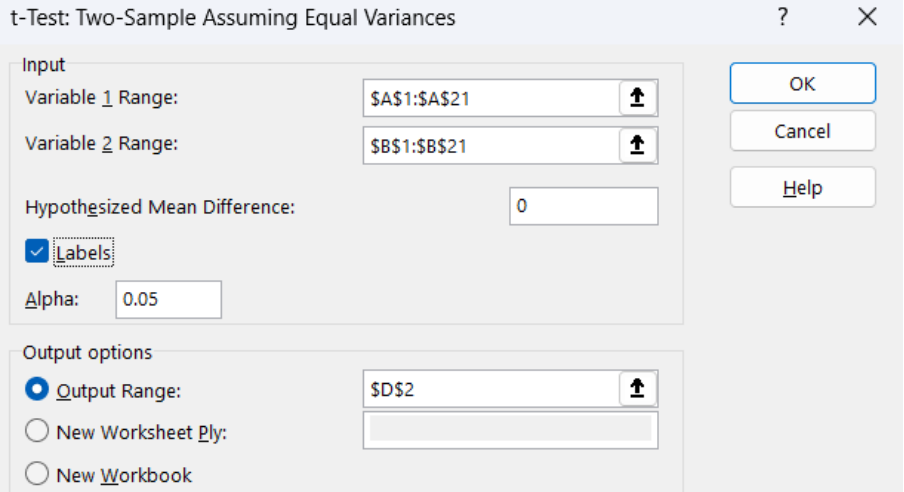
H0= Two different species of plants in a particular country have the same mean height.

H1= Two different species of plants in a particular country do not have the same mean height.

Go to Data tab > Data Analysis



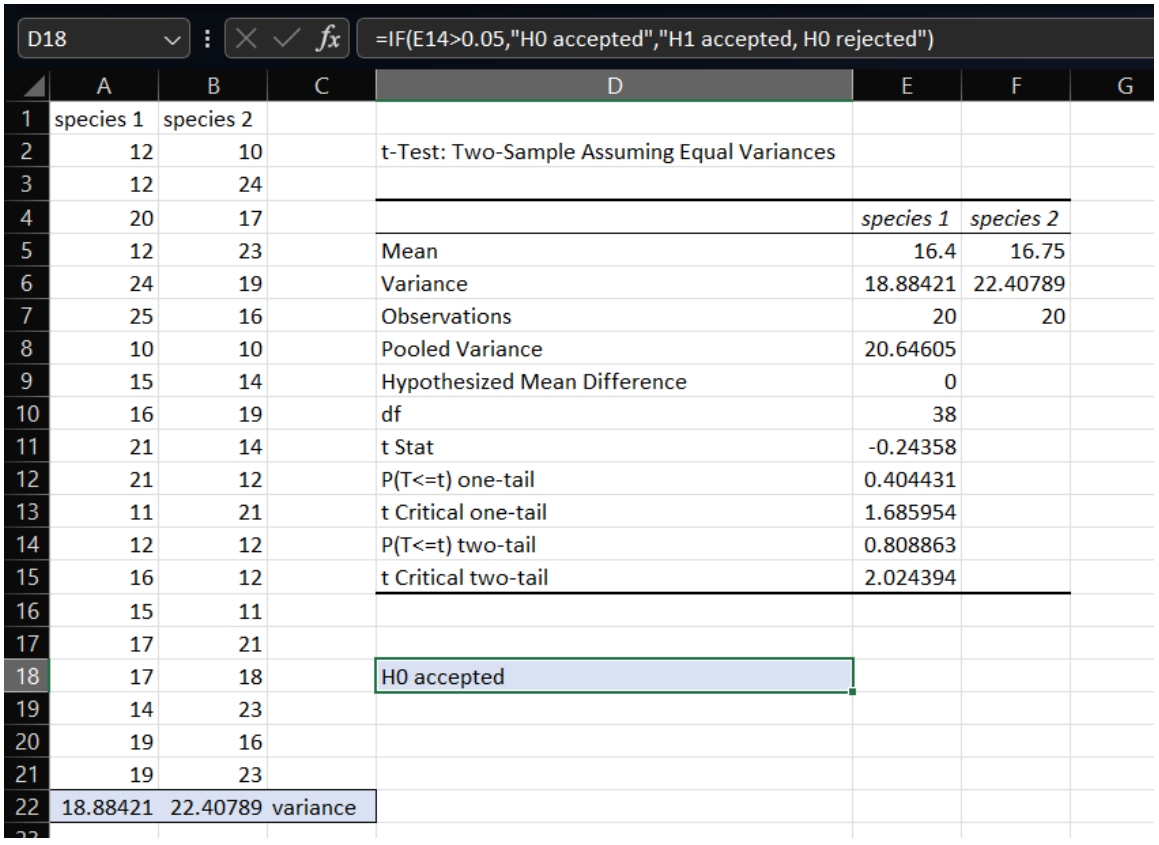
Click on OK.



Click on OK.

t-Test: Two-Sample Assuming Equal Variances		
	species 1	*species 2*
Mean	16.4	16.75
Variance	18.88421	22.40789
Observations	20	20
Pooled Variance	20.64605	
Hypothesized Mean Difference	0	
df	38	
t Stat	-0.24358	
P(T<=t) one-tail	0.404431	
t Critical one-tail	1.685954	
P(T<=t) two-tail	0.808863	
t Critical two-tail	2.024394	

Click on D18 type =IF(E14>0.05,"H0 accepted","H1 accepted, H0 rejected")



	A	B	C	D	E	F	G
1	species 1	species 2					
2	12	10		t-Test: Two-Sample Assuming Equal Variances			
3	12	24					
4	20	17			*species 1*	*species 2*	
5	12	23		Mean	16.4	16.75	
6	24	19		Variance	18.88421	22.40789	
7	25	16		Observations	20	20	
8	10	10		Pooled Variance	20.64605		
9	15	14		Hypothesized Mean Difference	0		
10	16	19		df	38		
11	21	14		t Stat	-0.24358		
12	21	12		P(T<=t) one-tail	0.404431		
13	11	21		t Critical one-tail	1.685954		
14	12	12		P(T<=t) two-tail	0.808863		
15	16	12		t Critical two-tail	2.024394		
16	15	11					
17	17	21					
18	17	18		H0 accepted			
19	14	23					
20	19	16					
21	19	23					
22	18.88421	22.40789	variance				

C. Perform testing of hypothesis using paired t-test.

The T distribution provides a good way to perform one sample tests on the mean when the population variance is not known provided the population is normal or the sample is sufficiently large so that the Central Limit Theorem applies.

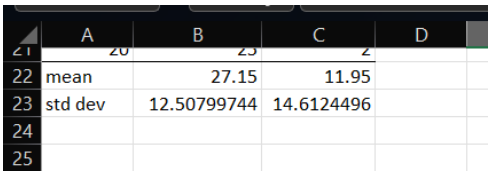
Paired Sample t Test

Example: A college Principal informed classroom teachers that some of their students showed unusual potential for intellectual gains. One months later the students identified to teachers as having potential for unusual intellectual gains showed significantly greater gains performance on a test said to measure IQ than did students who were not so identified. Below are the data for the students:

	experimental	comparison
1	35	2
2	40	27
3	12	38
4	15	31
5	21	1
6	14	19
7	46	1
8	10	34
9	28	3
10	48	1
11	16	2
12	30	3
13	32	2
14	48	1
15	31	2
16	22	1
17	12	3
18	39	29
19	19	37
20	25	2

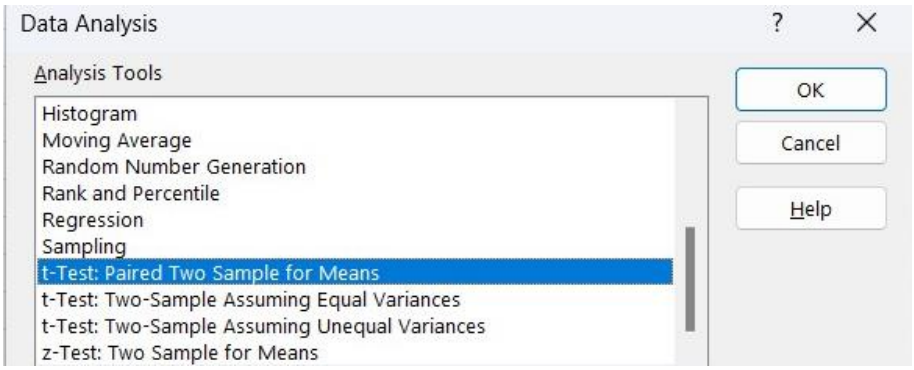
H0 represents that the difference in gain scores is not likely the result of the experimental treatment.
H1 represents that the difference in gain scores is likely the result of the experimental treatment and not the result of chance variation.

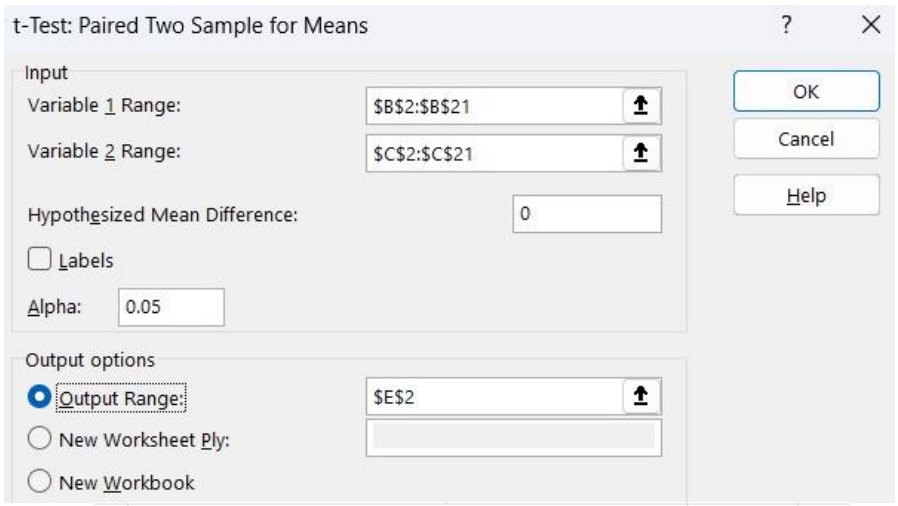
Experimental Data
To calculate Standard Mean, go to cell B22 and type =AVERAGE (B2:B21) To calculate Standard Deviation, go to cell B23 and type =STDEV (B2:B21)
Comparison Data
To calculate Standard Mean, go to cell C22 and type =AVERAGE (C2:C21) To calculate Standard Deviation, go to cell C23 and type =STDEV (C2:C21)



	A	B	C	D
22	mean	27.15	11.95	
23	std dev	12.50799744	14.6124496	
24				
25				

Find T-test Statistics, go to Data > Data Analysis> Click on Ok.

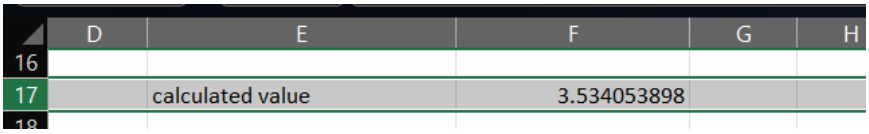




t-Test: Paired Two Sample for Means

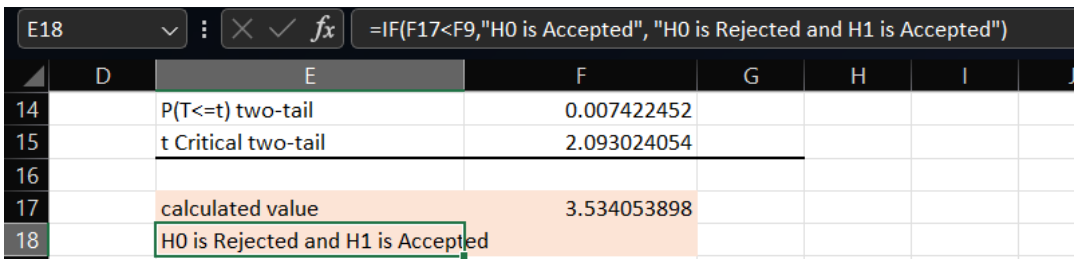
	Variable 1	*Variable 2*
Mean	27.15	11.95
Variance	156.45	213.5237
Observations	20	20
Pearson Correlation	-0.395904927	
Hypothesized Mean Difference	0	
df	19	
t Stat	2.996289153	
P(T<=t) one-tail	0.003711226	
t Critical one-tail	1.729132812	
P(T<=t) two-tail	0.007422452	
t Critical two-tail	2.093024054	

To calculate the T-Test square value go to cell F17 and type =(B22- C22)/SQRT(((B23*B23)/COUNT (B2:B21)) +((C23*C23)/ COUNT (C2:C21)))



	D	E	F	G	H
16					
17		calculated value	3.534053898		

Now go to cell F18 and type =IF (F17<F9,"H0 is Accepted", "H0 is Rejected and H1 is Accepted")



E18 =IF(F17<F9,"H0 is Accepted", "H0 is Rejected and H1 is Accepted")

	D	E	F	G	H	I
14		P(T<=t) two-tail	0.007422452			
15		t Critical two-tail	2.093024054			
16						
17		calculated value	3.534053898			
18		H0 is Rejected and H1 is Accepted				

Our calculated value is larger than the tabled value at alpha = 0.05, so we reject the null hypothesis and accept the alternative hypothesis, namely, that the difference in gain scores is likely the result of the experimental treatment and not the result of chance variation.

Practical: 05

A. Perform testing of hypothesis using chi-squared goodness-of-fit test.

Problem: A system administrator needs to upgrade the computers for his division. He wantsto know what sort of computer system his workers prefer. He gives three choices: Windows,Mac, or Linux. Test the hypothesis or theory that an equal percentage of the population prefers each type of computer.

system	Oi	Ei
windows	20	33.33333
mac	60	33.33333
linux	20	33.33333

H0: The population distribution of the variable is the same as the proposed distribution.
H1: The distributions are different.
To calculate the Chi –Squared value for Windows go to cell D2 and type =((B2-C2)^2)/C2
To calculate the Chi –Squared value for mac go to cell D3 and type =((B3-C3)^2)/C3
To calculate the Chi –Squared value for linux go to cell D4 and type =((B4-C4)^2)/C4
Go to Cell D5 for "sum{[(Oi-Ei)^2]/Ei}" and type=SUM(D2:D4)
To get the table value for Chi-Square for α = 0.05 and dof = 2, go to cell D7 and type = CHIINV(0.05,2)
At cell D8 type =IF(D5>D7, "H0 Accepted","H0 Rejected")

Output:

	A	B	C	D
1	system	Oi	Ei	sum{[(Oi-Ei)^2]/Ei}
2	windows	20	33.333333	5.333333333
3	mac	60	33.333333	21.33333333
4	linux	20	33.333333	5.333333333
5	total	100	100	32
6				
7			table value	5.991464547
8				H0 Accepted
9				

B. Perform testing of hypothesis using chi-squared test of independence.

In a study to understand the performance of M. Sc. IT Part -1 class, a college selects a random sample of 100 students. Each student was asked his grade obtained in B. Sc. IT. Thesample is as given below.

Sr. No	Roll No	Name	Gen	Grd
1	1	Gaborone	m	O
2	2	Francesco	m	O
3	5	Niamey	m	O
4	13	Maxixe	m	O
5	16	Tema	m	O
6	17	Kumasi	m	O
7	34	Blida	m	O
8	35	Oran	m	O
9	38	Saefda	m	O
10	42	sonam	m	O
11	43	Annaba	m	O
12	45	Bejaefa	m	O
13	48	Medea	m	O
14	49	Djelfa	m	O
15	50	Tipaza	m	O
16	51	Bechar	m	O
17	54	Mostag	m	O
18	55	Tiaret	m	O
19	56	Bouira	m	O
20	59	Tebessa	m	O
21	61	Harrach	m	O
22	62	Mila	m	O
23	65	Fouka	m	O
24	66	El Eulma	m	O
25	68	Abbes	m	O
26	69	Jijel	m	O
27	70	Guelma	m	O
28	85	Khechna	m	O
29	87	Kiffan	m	O
30	88	Lakhdaria	m	O
31	6	Maputo	m	D
32	12	Lichinga	m	D
33	15	Garcia	m	D
34	19	Accra	m	D
35	27	Wa	m	D
36	28	Navrongo	m	D
37	37	Mascara	m	D
38	44	Batna	m	D
39	57	El Biar	m	D
40	60	Boufarik	m	D
41	63	Oued	m	D
42	64	Ahras	m	D
43	71	Befda	m	D
44	86	Birtouta	m	D
45	18	Takoradi	m	C
46	22	Cape Coast	m	C
47	29	Kwabeng	m	C
48	30	Algiers	m	C
49	31	Laghouat	m	C
50	39	Relizane	m	C
51	52	Setif	m	C
52	53	Biskra	m	C
53	67	Kolea	m	C
54	100	AefnFakroun	m	C
55	26	Nima	m	B
56	32	TiziOuzou	m	B
57	33	Chlef	m	B
58	89	M'sila	m	A
59	96	Heliopolis	m	A
60	97	Berrouaghia	m	A
61	98	Sougueur	m	A
62	3	Maun	f	O
63	7	Tete	f	O
64	9	Chimoio	f	O
65	11	Pemba	f	O
66	14	Chibuto	f	O
67	25	Mampong	f	O
68	36	Tlemcen	f	O
69	40	Adrar	f	O
70	41	Tindouf	f	O
71	46	Skikda	f	O
72	47	Ouargla	f	O
73	10	Matola	f	D
74	20	Legon	f	D
75	21	Sunyani	f	D
76	72	Teenas	f	D
77	73	Kouba	f	D
78	75	HussenDey	f	D
79	77	Khenchela	f	D
80	82	HassiBahbah	f	D
81	84	Baraki	f	D
82	91	Boudouaou	f	D
83	95	Tadjenanet	f	D
84	4	Molepolole	f	C
85	8	Quelimane	f	C
86	23	Bolgatanga	f	C
87	58	Mohammadia	f	C
88	83	Merouana	f	C
89	24	Ashaiman	f	B
90	76	N'gaous	f	B
91	90	Oued	f	B
92	92	BordjMenael	f	B
93	93	Boukhari	f	B
94	74	Reghaa	f	A
95	78	Cheria	f	A
96	79	Mouzaa	f	A
97	80	Meskiana	f	A
98	81	Miliana	f	A
99	94	Sig	f	A
100	99	Kadiria	f	A

H0: The performance of girl students is same as boy students.
H1: The performance of boy and girl students are different.

Open Excel Workbook

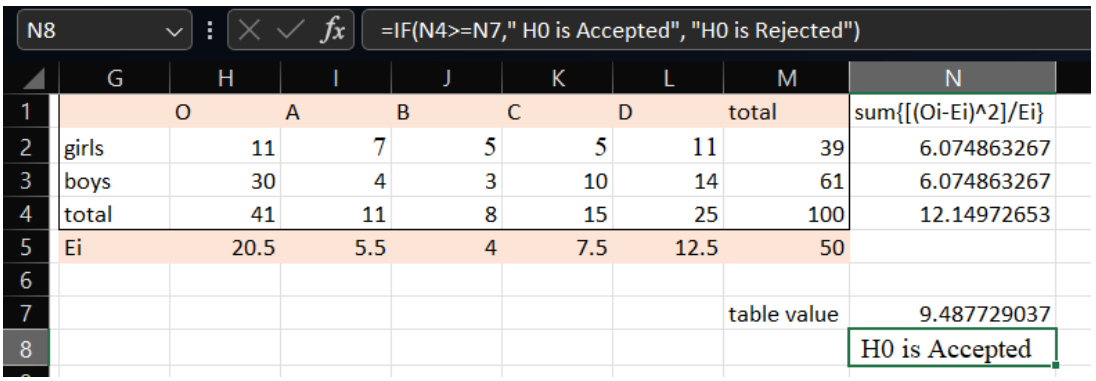
	O	A	B	C	D	total	sum{[(Oi-Ei)^2]/Ei}
girls	11	7	5	5	11	39	6.074863267
boys	30	4	3	10	14	61	6.074863267
total	41	11	8	15	25	100	12.14972653
Ei	20.5	5.5	4	7.5	12.5	50	

To prepare a contingency table as shown above.To calculate Girls Std with “O” Grade
Go to Cell H2 and type =COUNTIFS(D2:D101,"f",E2:E101,"O")
To calculate Girls Students with “A” Grade
Go to Cell I2 and type =COUNTIFS(D2:D101,"f",E2:E101,"A")
To calculate Girls Students with “B” Grade
Go to Cell J2 and type =COUNTIFS(D2:D101,"f",E2:E101,"B")
To calculate Girls Students with “C” Grade
Go to Cell K2 and type =COUNTIFS(D2:D101,"f",E2:E101,"C")
To calculate Girls Students with “D” Grade
Go to Cell L2 and type =COUNTIFS(D2:D101,"m",E2:E101,"D")

To calculate Boys Students with “O” Grade
Go to Cell H3 and type =COUNTIFS(D2:D101,"m",E2:E101,"O")
To calculate Boys Students with “A” Grade
Go to Cell I3 and type =COUNTIFS(D2:D101,"m",E2:E101,"A")
To calculate Boys Students with “B” Grade
Go to Cell J3 and type =COUNTIFS(D2:D101,"m",E2:E101,"B")
To calculate Boys Students with “C” Grade
Go to Cell K3 and type =COUNTIFS(D2:D101,"m",E2:E101,"C")
To calculate Boys Students with “D” Grade
Go to Cell L3 and type =COUNTIFS(D2:D101,"m",E2:E101,"D")
Use AutoSum to get total values.

To calculate Ei
On H5 type =H4/2 On I5 type =I4/2 On J5 type =J4/2 On K5 type =K4/2 On L5 type =L4/2 On M5 type =M4/2
Now calculate “sum{[(Oi-Ei)^2]/Ei}”
Go to cell N2 and type
=SUM((H2-H5)^2/H5,(I2-I5)^2/I5,(J2-J5)^2/J5,(K2-K5)^2/K5,(L2-L5)^2/L5)
Go to cell N3 and type
=SUM((H3-H5)^2/H5,(I3-I5)^2/I5,(J3-J5)^2/J5,(K3-K5)^2/K5,(L3-L5)^2/L5)
To get the table value go to cell N7 and type =CHIINV(0.05,4)
Go to cell N8 and type =IF(N4>=N7," H0 is Accepted", "H0 is Rejected")

Output:



Practical: 06

➢ **Perform testing of hypothesis using Z-test.**

One sample:

IQ
82
82
85
87
87
88
92
92
94
94
95
96
97
97
97
99
99
101
101
103
103
105
107
109
109
109
110
112
112
113
114
115

Given:

population mean = 100,
population std dev = 15.

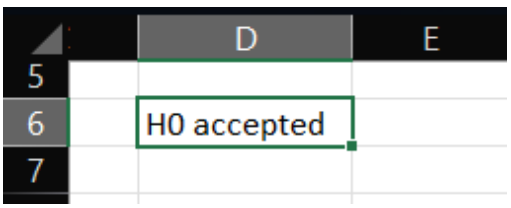
H0: population mean is equal to 100.
H1: population mean is not equal to 100.

Click on D2 and type =Z.TEST(A2:A33,100,15)

*(Please note that this gives us value for a one-tailed test. To obtain P-value for two-tailed test click on D4 and type =D2*2. This will give us an approximate P-value for two-tailed Z-test.)*

	0.565525	one tail
P-value	1.131049	two tail

Go to D6 and type =IF(D4>0.05,"H0 accepted","H1 accepted,H0 rejected")



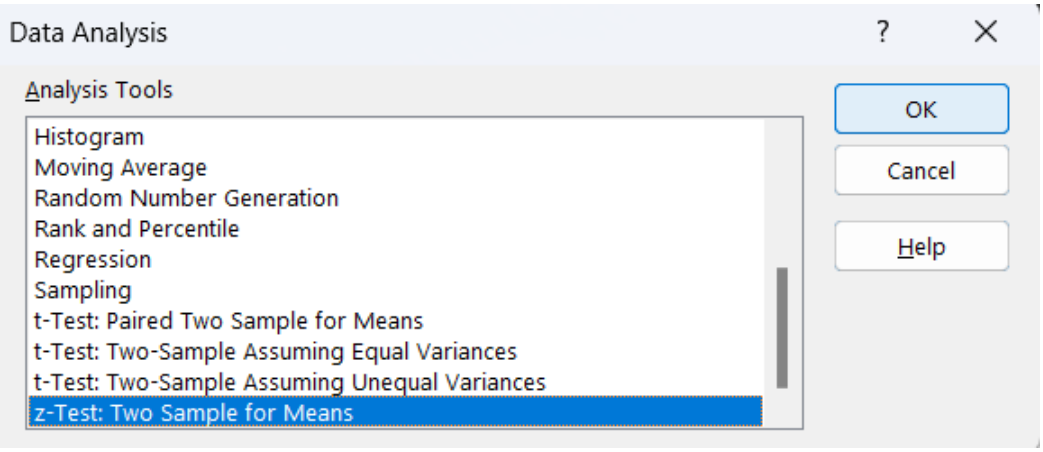
Two sample:

CITY A	CITY B
82	88
84	88
85	90
89	91
89	91
90	91
90	93
90	93
91	95
91	95
92	99
94	99
94	102
94	102
98	105
98	107
99	108
99	109
105	109
106	114
106	115
109	116
109	117
109	117
110	119
112	123
112	128
113	129
114	130
114	133

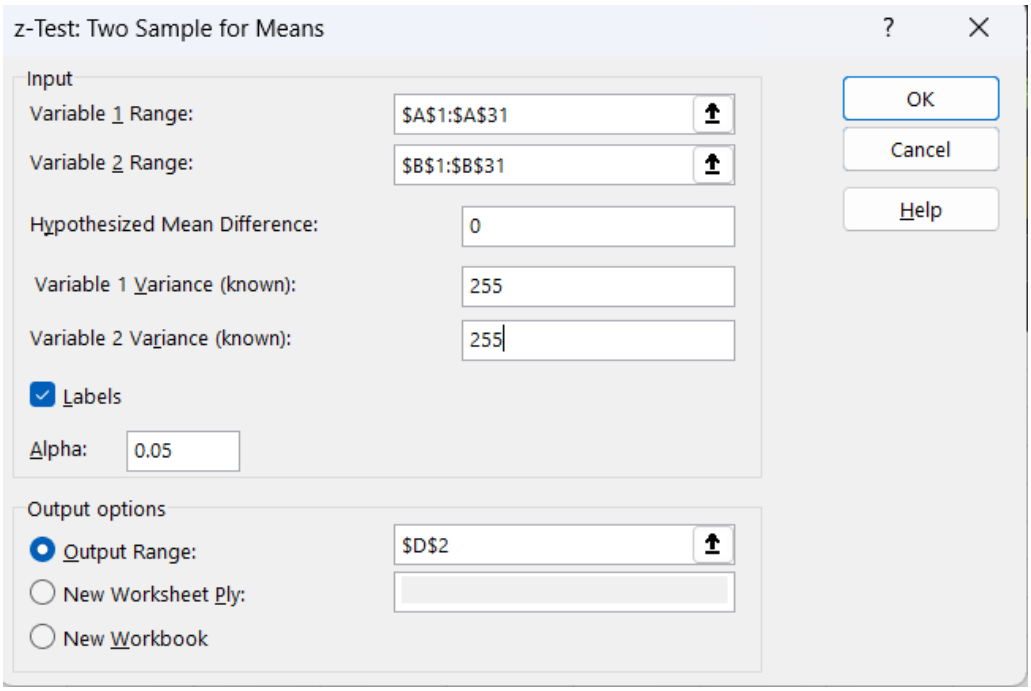
Given: variance for city
a: 225 Variance for city b: 225

H0: population mean for city a = population mean for city b
H1: population mean for city a is not equal to population mean for city b.

Go to Data tab > Data Analysis



Click on OK.



Click on OK.

D2 | z-Test: Two Sample for Means

	A	B	C	D	E	F
1	CITY A	CITY B				
2	82	88		z-Test: Two Sample for Means		
3	84	88				
4	85	90			*CITY A*	*CITY B*
5	89	91		Mean	98.93333	106.5333
6	89	91		Known Variance	255	255
7	90	91		Observations	30	30
8	90	93		Hypothesized Mean Difference	0	
9	90	93		z	-1.84327	
10	91	95		P(Z<=z) one-tail	0.032645	
11	91	95		z Critical one-tail	1.644854	
12	92	99		P(Z<=z) two-tail	0.065289	
13	94	99		z Critical two-tail	1.959964	
14	94	102				

Click on D15 and type =IF(E12>0.05,"H0 accepted","H1 accepted, H0 rejected")

Output:

	A	B	C	D	E	F
1	CITY A	CITY B				
2	82	88		z-Test: Two Sample for Means		
3	84	88				
4	85	90			*CITY A*	*CITY B*
5	89	91		Mean	98.93333	106.5333
6	89	91		Known Variance	255	255
7	90	91		Observations	30	30
8	90	93		Hypothesized Mean Difference	0	
9	90	93		z	-1.84327	
10	91	95		P(Z<=z) one-tail	0.032645	
11	91	95		z Critical one-tail	1.644854	
12	92	99		P(Z<=z) two-tail	0.065289	
13	94	99		z Critical two-tail	1.959964	
14	94	102				
15	94	102		H0 accepted		
16	98	105				

Practical: 07

A. Perform testing of hypothesis using One-way ANOVA.

ANOVA assumptions:

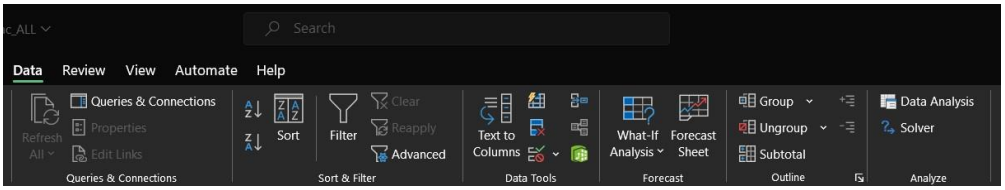
- The dependent variable (none, low medium and high in our example) should becontinuous.
- The independent variables (daily and weekly in our example) should be two or morecategorical groups.
- There must be different participants in each group with no participant being in morethan one group.
- The dependent variable should be approximately normally distributed for eachcategory.
- Variances of each group are approximately equal.

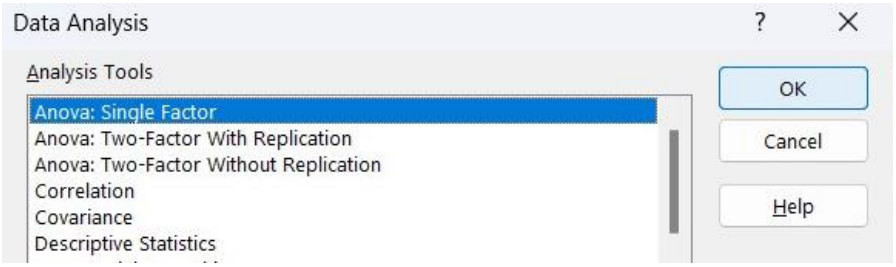
Below you can find the salaries of people who have a degree in economics, medicine orhistory.

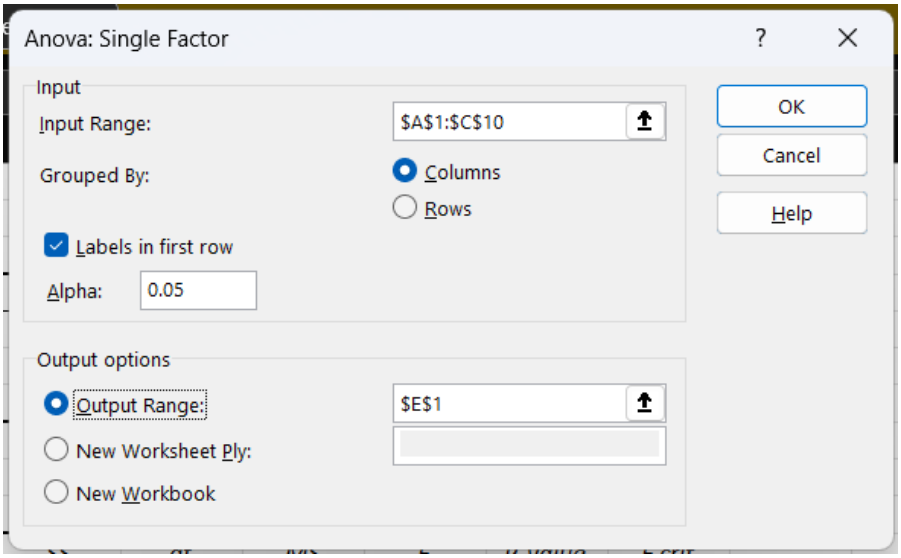
economics	medicine	history
42	69	35
53	54	40
49	58	53
53	64	42
43	64	50
44	55	39
45	56	55
52		39
54		40

H0 – all means are same (μ1 = μ2 = μ3).
H1 – at least one mean is different.

To perform ANOVA, go to Data > Data Analysis







Input Range: A1:C10
Output Range: E1

	D	E	F	G	H	I	J	K
1		Anova: Single Factor						
2								
3		SUMMARY						
4		*Groups*	*Count*	*Sum*	*Average*	*Variance*		
5		economics	9	435	48.33333	23.5		
6		medicine	7	420	60	32.33333		
7		history	9	393	43.66667	50.5		
8								
9								
10		ANOVA						
11		*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
12		Between Groups	1085.84	2	542.92	15.19623	7.16E-05	3.443357
13		Within Groups	786	22	35.72727			
14								
15		Total	1871.84	24				

To take a decision, in cell B14 type =IF(J11<C13,"H0 accepted","H1 accepted, H0 rejected")
Since the resulting p value is less than 0.05. The null hypothesis (H0) is rejected andconcluded that at least one mean is different.

Output:

	A	B	C	D
1	economics	medicine	history	
2	42	69	35	
3	53	54	40	
4	49	58	53	
5	53	64	42	
6	43	64	50	
7	44	55	39	
8	45	56	55	
9	52		39	
10	54		40	
11				
12				
13	significance level		0.05	
14		H1 accepted, H0 rejected		

B. Perform testing of hypothesis using Two-way ANOVA.

A two-way ANOVA ("analysis of variance") is used to determine whether or not there is astatistically significant difference between the means of three or more independent groupsthat have been split on two variables (sometimes called "factors").

The results of a two-way ANOVA to be valid, the following assumptions should be met:

- **Normality** – The response variable is approximately normally distributed for eachgroup.
- **Equal Variances** – The variances for each group should be roughly equal.
- **Independence** – The observations in each group are independent of each other andthe observations within groups were obtained by a random sample.

Example:

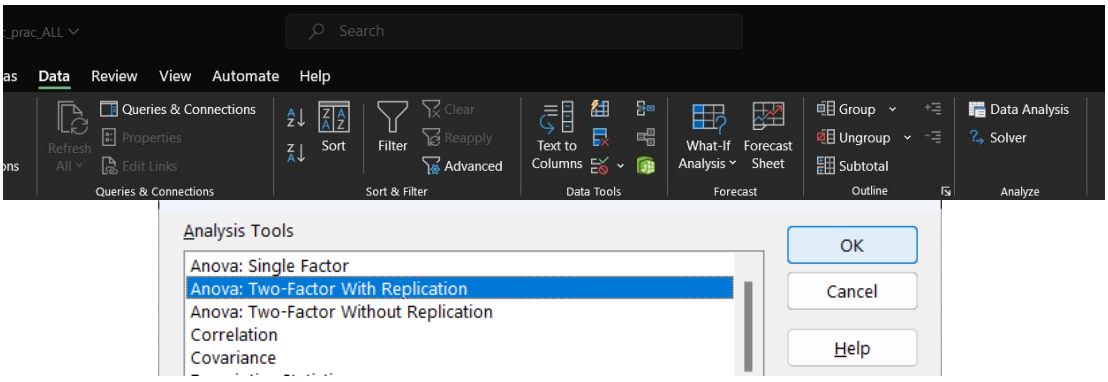
A botanist wants to know whether plant growth is influenced by sunlight exposure and watering frequency. She plants 40 seeds and lets grow for two months under differentconditions for sunlight
exposure and watering frequency. After two months, she records theheight of each plant. The results are shown below:

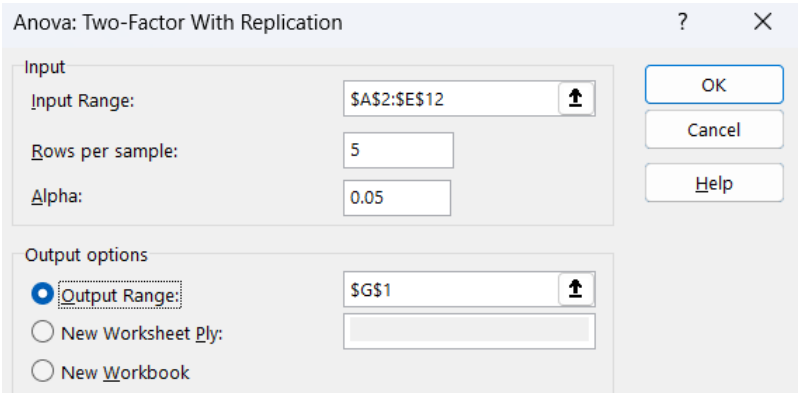
Water Frequency	Sunlight Exposure			
	None	Low	Medium	High
Daily	4.8	5	6.4	6.3
	4.4	5.2	6.2	6.4
	3.2	5.6	4.7	5.6
	3.9	4.3	5.5	4.8
	4.4	4.8	5.8	5.8
Weekly	4.4	4.9	5.8	6
	4.2	5.3	6.2	4.9
	3.8	5.7	6.3	4.6
	3.7	5.4	6.5	5.6
	3.9	4.8	5.5	5.5

H0 represents whether a plant watered daily or weekly has impact on how sunlightexposure affects a plant.

H1 represents whether a plant watered daily or weekly has no impact, how sunlightexposure affects a plant.

Go to Data tab > Data Analysis





Input Range - A2:E12
Rows Per Sample – 5
Alpha – 0.05
Output Range - G1

Anova: Two-Factor With Replication

SUMMARY	none	low	medium	high	Total
	daily				
Count	5	5	5	5	20
Sum	20.7	24.9	28.6	28.9	103.1
Average	4.14	4.98	5.72	5.78	5.155
Variance	0.378	0.232	0.447	0.412	0.775237
weekly					
Count	5	5	5	5	20
Sum	20	26.1	30.3	26.6	103
Average	4	5.22	6.06	5.32	5.15
Variance	0.085	0.137	0.163	0.317	0.722632
Total					
Count	10	10	10	10	
Sum	40.7	51	58.9	55.5	
Average	4.07	5.1	5.89	5.55	
Variance	0.211222	0.18	0.303222	0.382778	

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Sample	0.00025	1	0.00025	0.000921	0.975975	4.149097
Columns	18.76475	3	6.254917	23.04898	3.9E-08	2.90112
Interaction	1.01075	3	0.336917	1.241517	0.310898	2.90112
Within	8.684	32	0.271375			
Total	28.45975	39				

To take a decision, click on cell B15 and type==IF(L24<B14,"H0 accepted","H1 accepted,H0 rejected")

Output:

	A	B	C	D	E
1		sunlight exposure			
2	water frequency	none	low	medium	high
3	daily	4.8	5	6.4	6.3
4		4.4	5.2	6.2	6.4
5		3.2	5.6	4.7	5.6
6		3.9	4.3	5.5	4.8
7		4.4	4.8	5.8	5.8
8	weekly	4.4	4.9	5.8	6
9		4.2	5.3	6.2	4.9
10		3.8	5.7	6.3	4.6
11		3.7	5.4	6.5	5.6
12		3.9	4.8	5.5	5.5
13					
14	significance level	0.05			
15	H1 accepted, H0 rejected				

Practical: 08

A. Perform the Random sampling for the given data and analyze it.

Example: A test conducted of 40 marks for a class of 100. We want a sample that representsthe class. Data for the same is given below.

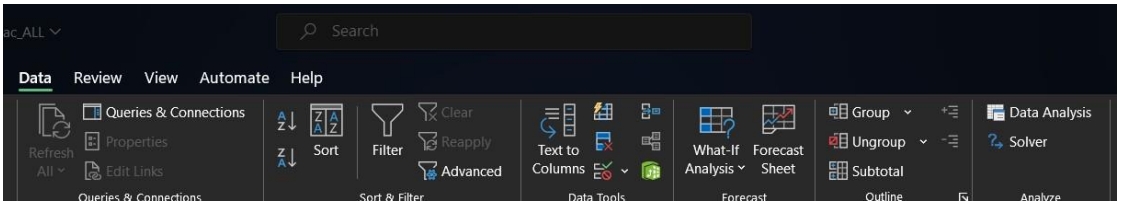
rollno	marks
1	19
2	29
3	8
4	27
5	38
6	5
7	36
8	24
9	23
10	12
11	33
12	30
13	27
14	13
15	22
16	10
17	36
18	17
19	26
20	10
21	17
22	12
23	4
24	22
25	23

26	1
27	9
28	12
29	3
30	8
31	18
32	5
33	32
34	15
35	26
36	5
37	24
38	2
39	29
40	4
41	31
42	31
43	13
44	7
45	31
46	31
47	8
48	13
49	20
50	13

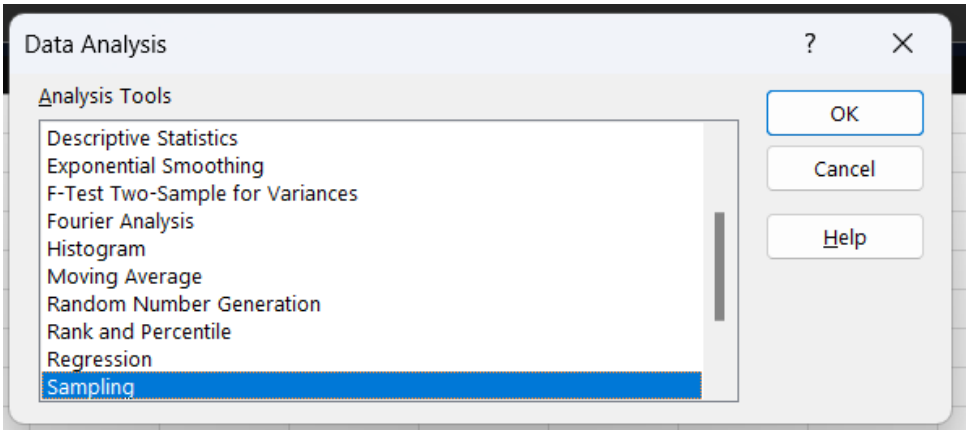
51	36
52	6
53	26
54	0
55	4
56	4
57	19
58	38
59	40
60	13
61	25
62	30
63	21
64	5
65	21
66	33
67	29
68	1
69	7
70	19
71	24
72	33
73	21
74	7
75	34

76	25
77	33
78	10
79	26
80	17
81	34
82	18
83	19
84	22
85	3
86	31
87	4
88	31
89	25
90	25
91	28
92	8
93	13
94	9
95	1
96	25
97	39
98	2
99	33
100	38

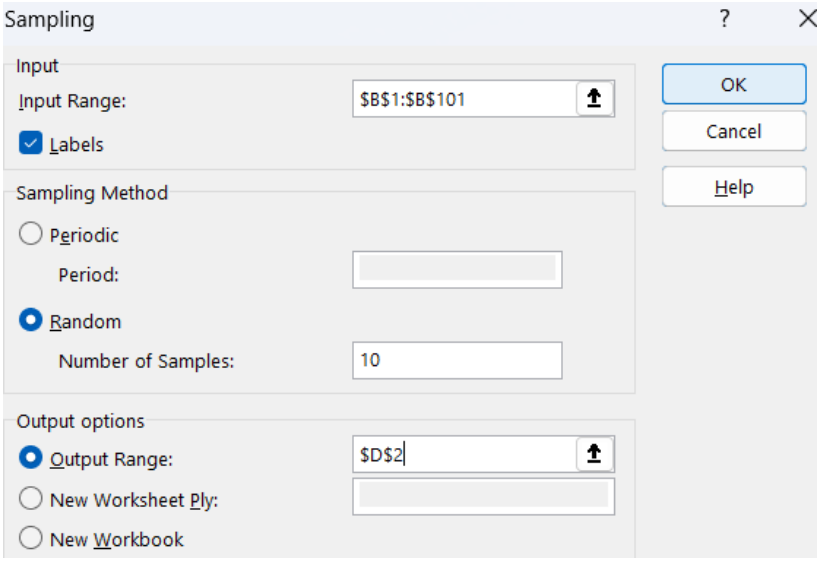
Go to Data tab > Data Analysis



Select sampling > ok

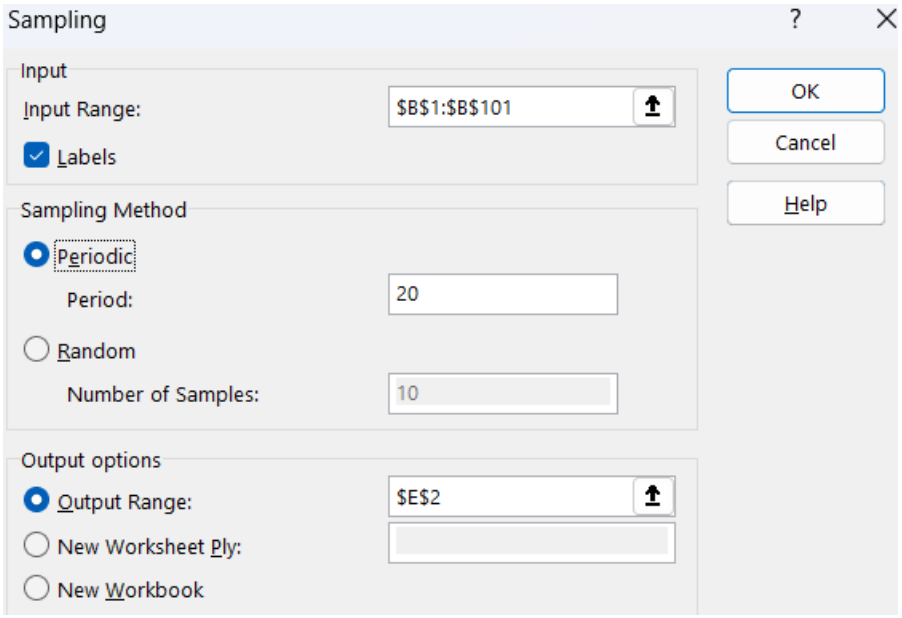


- **For random sampling**



Select **input range** - B1:B101
Make sure you have checked checkbox for labels.
Select Sampling method as Random.
Number of samples – 10
Select **output range** – D2.

- **For periodic sampling**



Select **input range** - B1:B101
Make sure you have checked checkbox for labels.
Select Sampling method as Periodic.
period – 20

if you want a specific number of samples and you want to know what to type in this field then use the formula: total number of samples/samples required.

Example: Select output range – E2

Output:

	random sample	periodic sample
2	27	10
3	33	4
4	18	13
5	31	17
6	18	38
7	13	
8	31	
9	38	
10	25	
11	27	

B. Perform the Periodic sampling for the given data and analyze it.

Example: A test conducted of 40 marks for a class of 100. We want a sample that represent intervals such as below 10, between 11 to 20, between 21 to 30, greater than 30, etc. Data forthe same is given below.

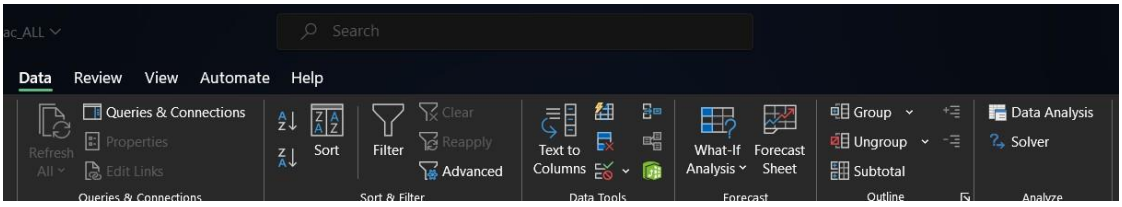
rollno	marks
1	19
2	29
3	8
4	27
5	38
6	5
7	36
8	24
9	23
10	12
11	33
12	30
13	27
14	13
15	22
16	10
17	36
18	17
19	26
20	10
21	17
22	12
23	4
24	22
25	23

26	1
27	9
28	12
29	3
30	8
31	18
32	5
33	32
34	15
35	26
36	5
37	24
38	2
39	29
40	4
41	31
42	31
43	13
44	7
45	31
46	31
47	8
48	13
49	20
50	13

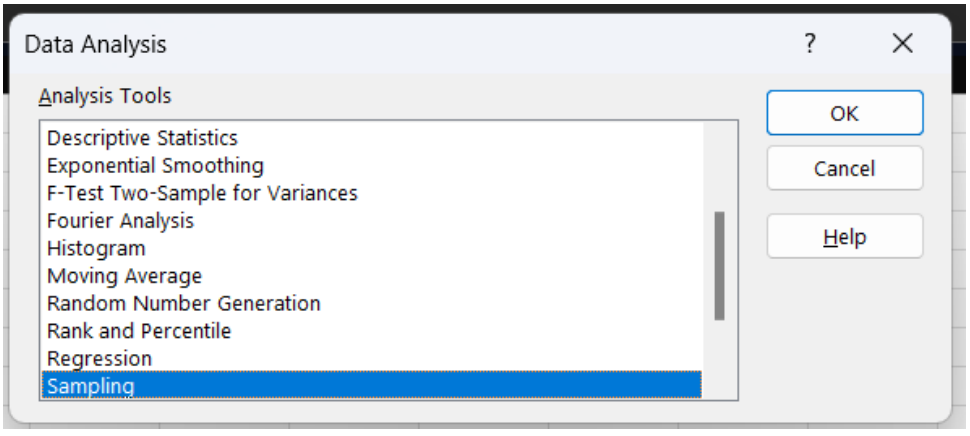
51	36
52	6
53	26
54	0
55	4
56	4
57	19
58	38
59	40
60	13
61	25
62	30
63	21
64	5
65	21
66	33
67	29
68	1
69	7
70	19
71	24
72	33
73	21
74	7
75	34

76	25
77	33
78	10
79	26
80	17
81	34
82	18
83	19
84	22
85	3
86	31
87	4
88	31
89	25
90	25
91	28
92	8
93	13
94	9
95	1
96	25
97	39
98	2
99	33
100	38

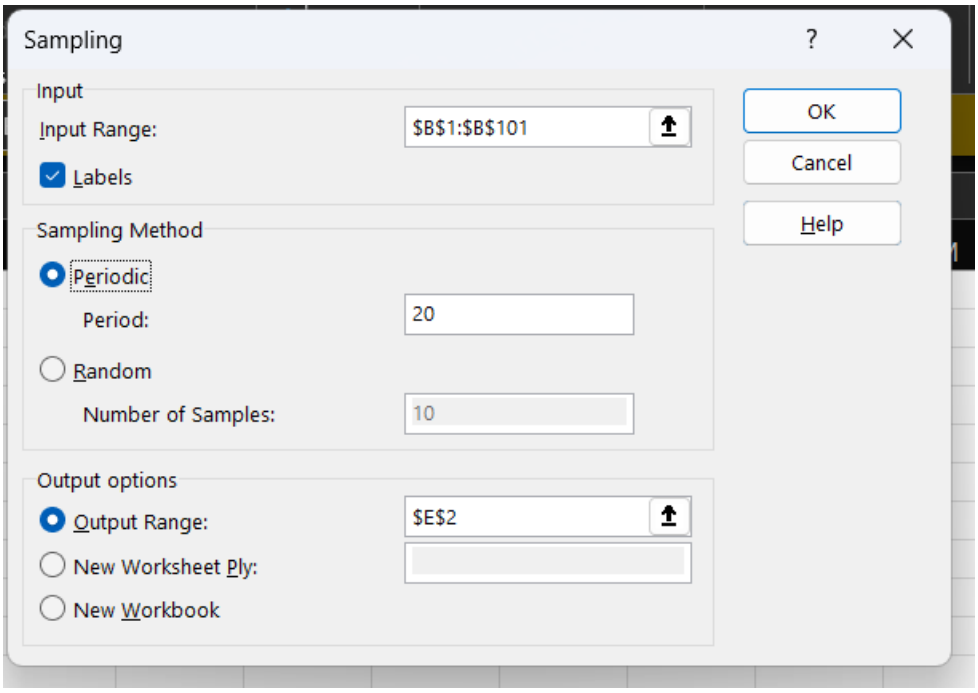
Go to Data tab > Data Analysis



Select sampling > ok



For periodic sampling



Select input range - B1:B101

Make sure you have checked checkbox for labels. Select

Sampling method as Periodic.

period – 20

if you want a specific number of samples and you want to know what to type in this field then use the formula: total number of samples/samples required.

Example: Select **output range –** E2

Output:

E
periodic sample
10
4
13
17
38

Practical: 09

- **Compute different types of correlation.**

Correlation is a statistical term describing the degree to which two variables move in coordination with one another. If the two variables move in the same direction, then those variables are said to have a positive correlation. If they move in opposite directions, then they have a negative correlation.

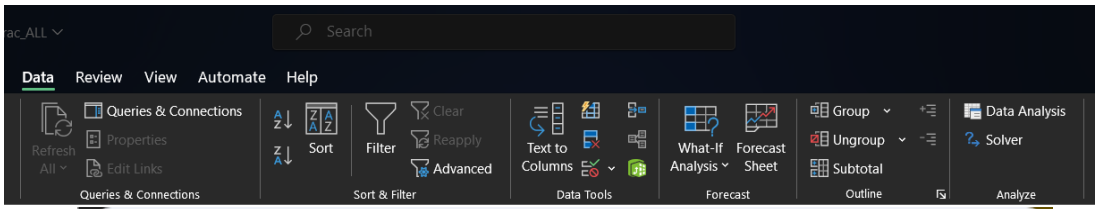
In Excel,

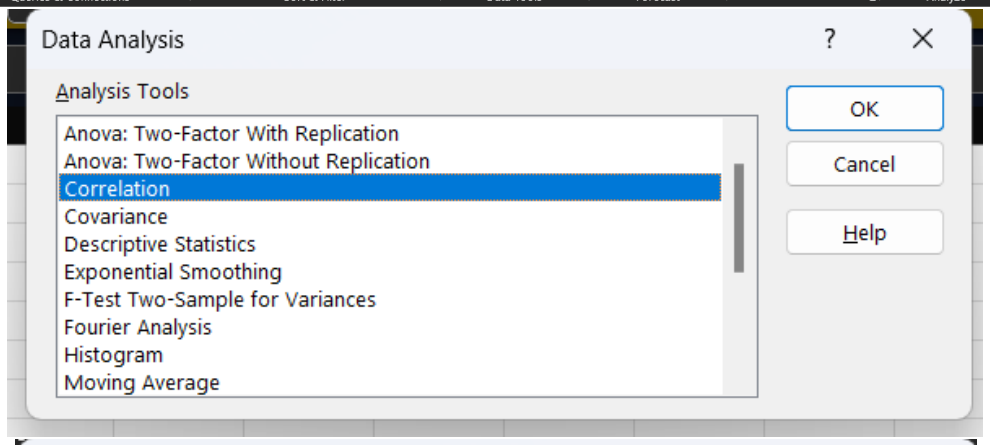
"1" represents positive correlation.
"0" represents no correlation.
"-1" represents negative correlation.

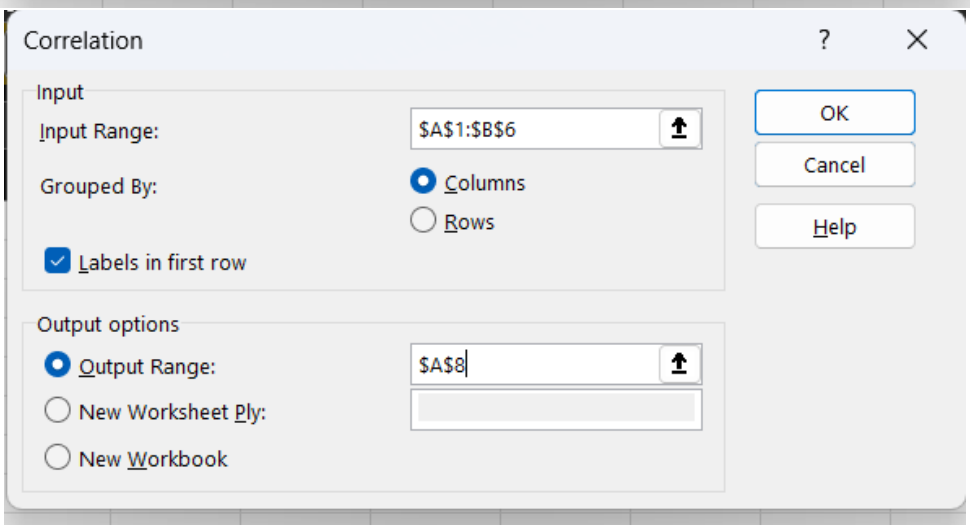
- **Positive correlation**

X	Y
0	2
10	12
2	4
12	14
6	8

Go to Data tab > Data Analysis > Correlation > Click Ok.

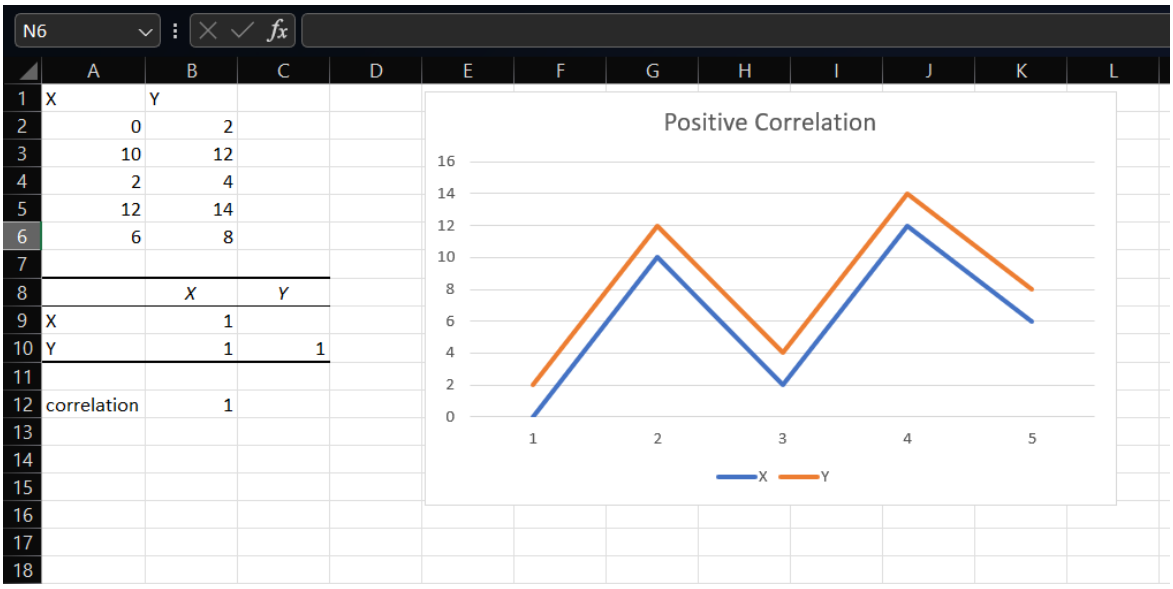






Select **input range** – A1:B6.

Select radio button Columns.
Check checkbox labels in first row.
Select **output range** – A8.
Click on Ok.
Another way to calculate correlation is to type =CORREL(A2:A6,B2:B6) in cell B12.
To plot a graph, go to Insert tab > Charts > Line Chart > Line.

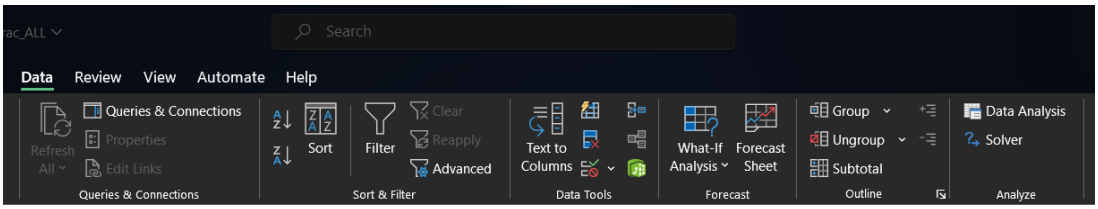


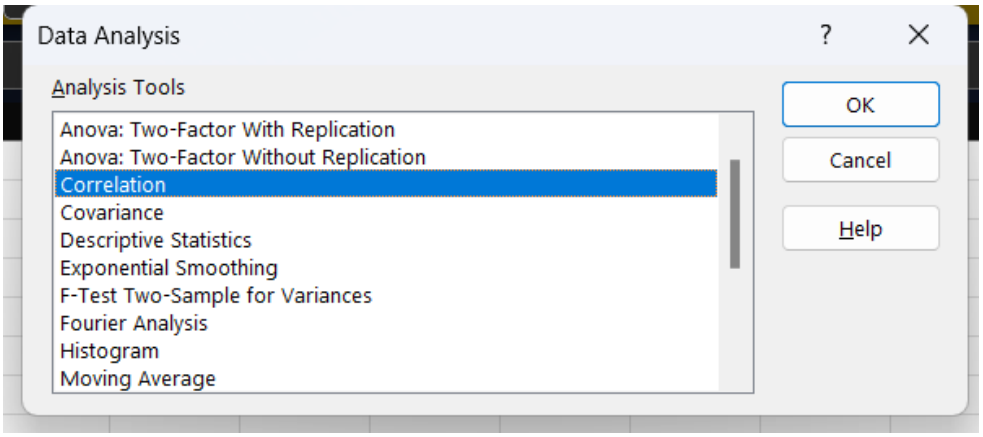
	A	B	C
1	X	Y	
2	0	2	
3	10	12	
4	2	4	
5	12	14	
6	6	8	
7			
8		*X*	*Y*
9	X	1	
10	Y	1	1
11			
12	correlation	1	

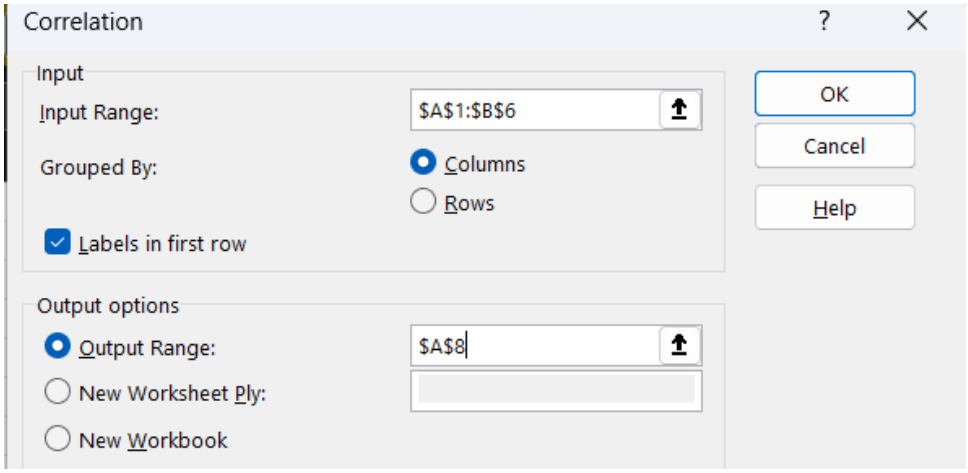
- **Negative correlation**

X	Z
0	2
10	-8
2	0
12	-10
6	-4

Go to Data tab > Data Analysis > Correlation > Click Ok.







Select input range – A1:B6.

Select radio button Columns.

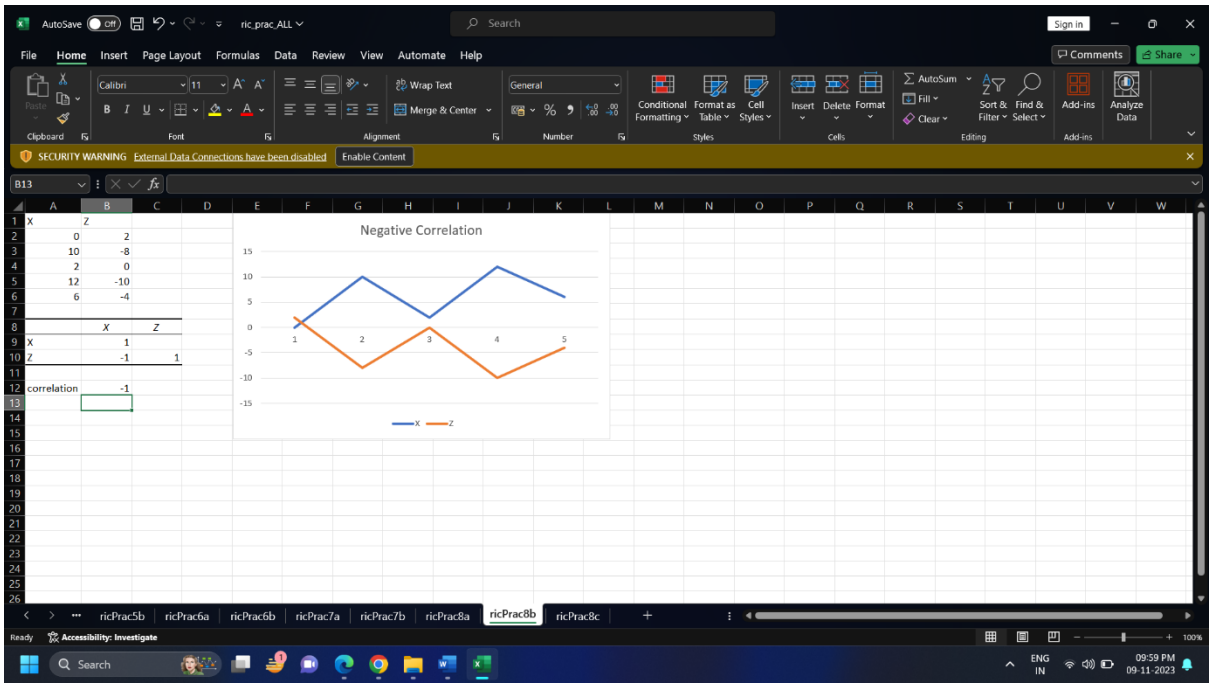
Check checkbox labels in first row.

Select output range – A8.

Click on Ok.

Another way to calculate correlation is to type =CORREL(A2:A6,B2:B6) in cell B12.

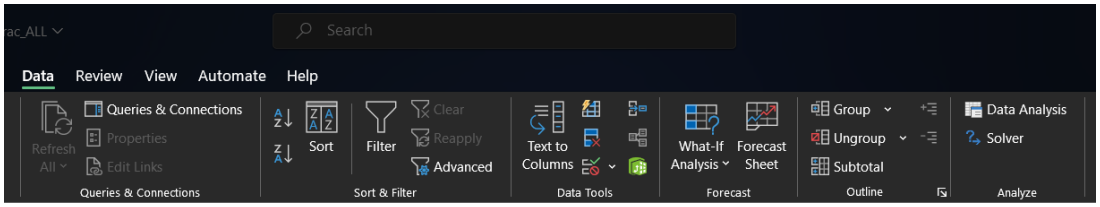
To plot a graph, go to Insert tab > Charts > Line Chart > Line.

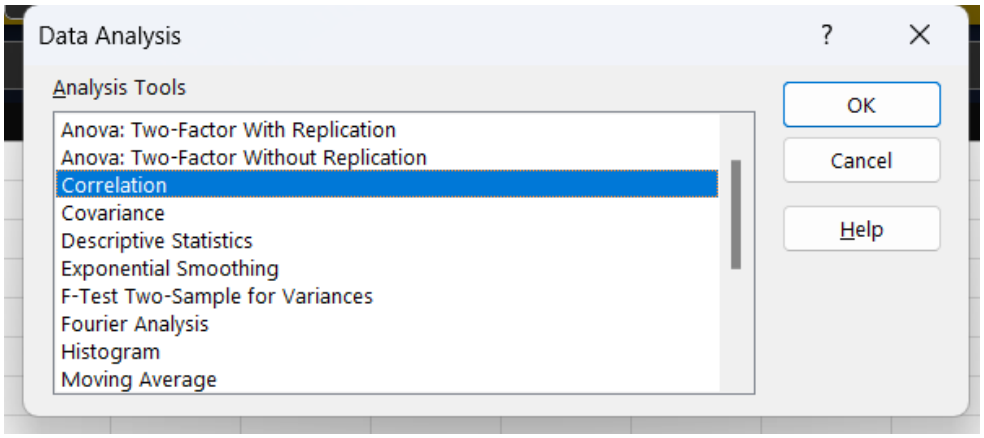


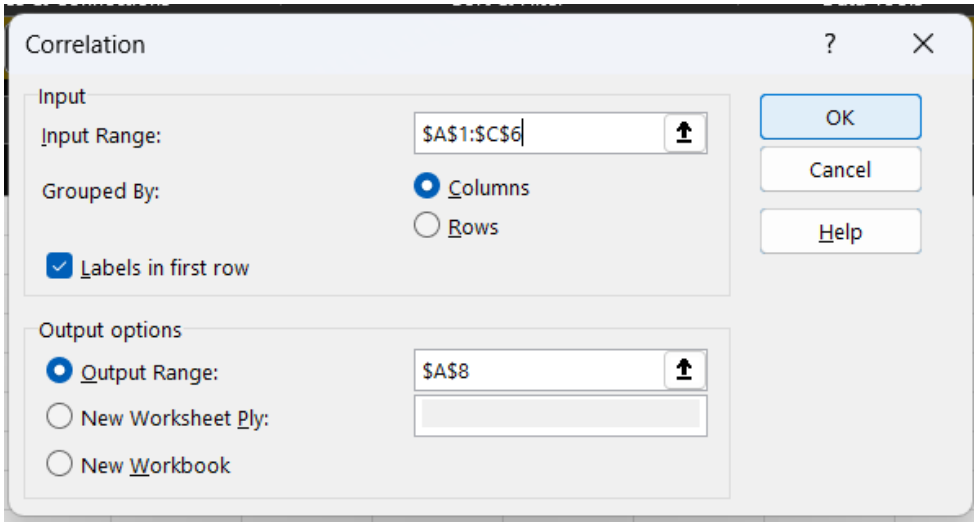
▪ Correlation for three or more variables

A	B	C
0	2	2
14	6	11
1	8	3
10	5	13
5	6	4

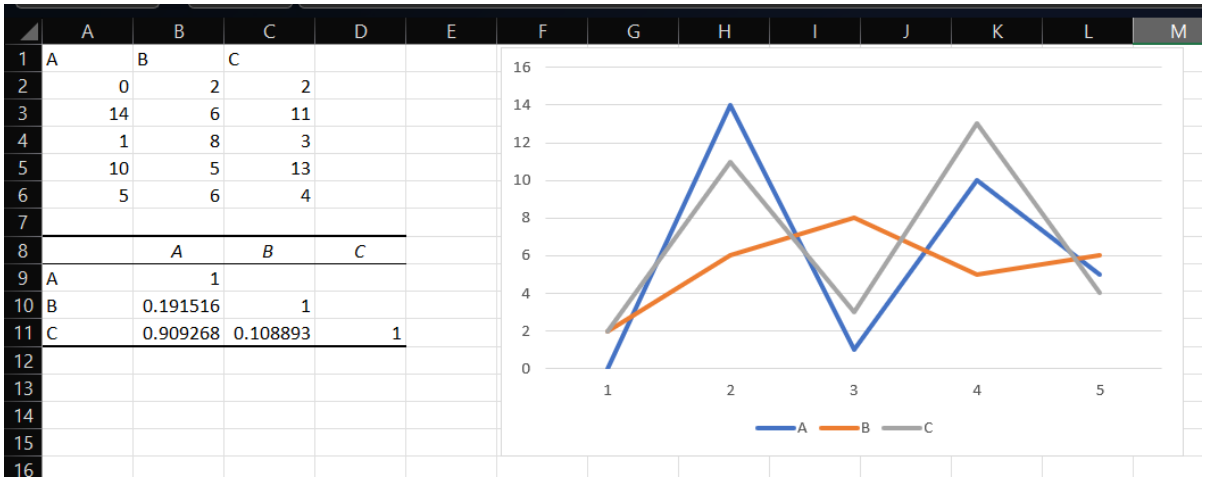
Go to Data tab > Data Analysis > Correlation > Click Ok.







Select input range – A1:C6.
Select radio button Columns.
Check checkbox labels in first row.
Select output range – A8.
Click on Ok.
To plot a graph, go to Insert tab > Charts > Line Chart > Line.



	A	B	C
1	A	B	C
2	0	2	2
3	14	6	11
4	1	8	3
5	10	5	13
6	5	6	4

		A	B	C
8				
9	A	1		
10	B	0.191516	1	
11	C	0.909268	0.108893	1

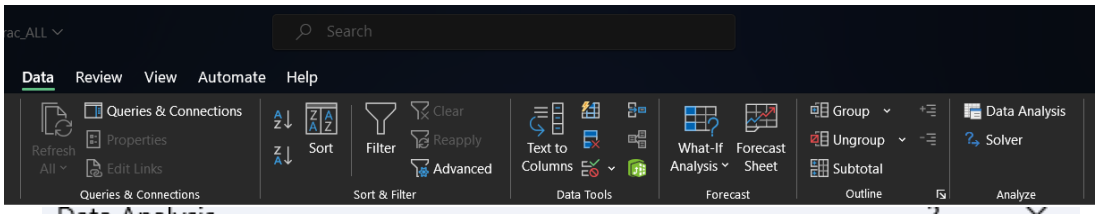
This example represents more real life situations.
Correlation between A and B is near “0”.
Correlation between A and C is near “1”.
Correlation between B and C is near “0”.

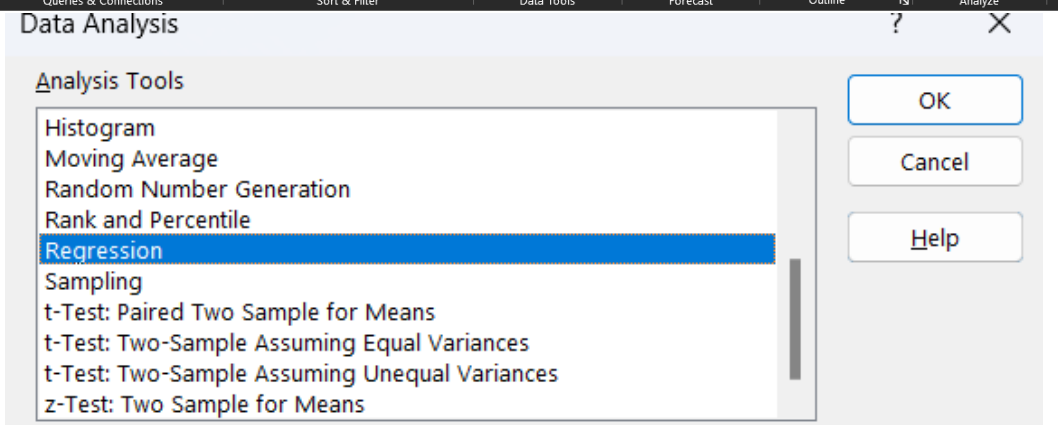
Practical: 10

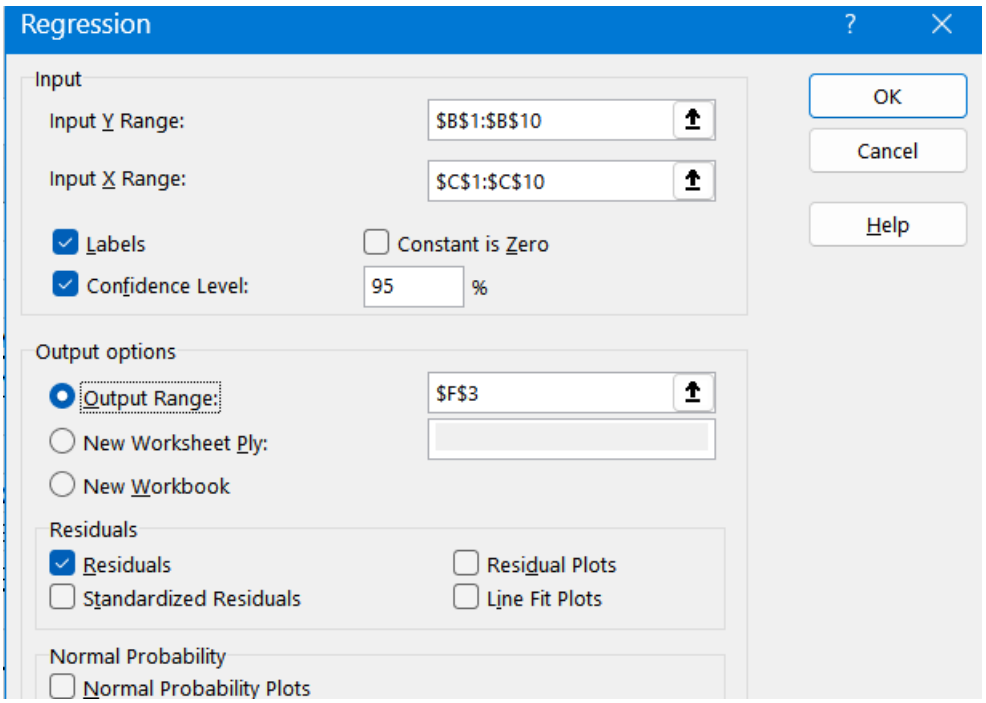
A. Perform linear regression for prediction.

Car_Name	Selling_Price	Present_Price
ritz	3.35	5.59
sx4	4.75	9.54
ciaz	7.25	9.85
wagon r	2.85	4.15
swift	4.6	6.87
vitara brezza	9.25	9.83
ciaz	6.75	8.12
s cross	6.5	8.61
ciaz	8.75	8.89
ciaz	7.45	8.92

Go to Data tab > Data Analysis > Regression > Click Ok.







Select **input Y range –** B1:B10.
Select **input X range** – C1:C10.
Check checkboxes for Labels and Confidence Level.
Select **output range** – F3.
Check checkbox Residuals.
Click Ok.

Output:

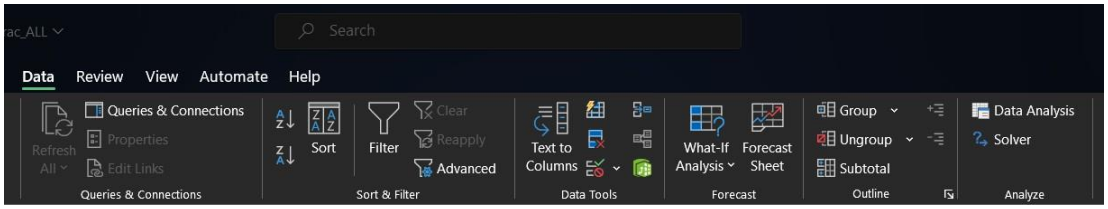
SUMMARY OUTPUT								
Regression Statistics								
Multiple R	0.806128							
R Square	0.649843							
Adjusted R	0.599821							
Standard E	1.431531							
Observatic	9							
ANOVA								
	df	*SS*	*MS*	*F*	*ignificance F*			
Regressior	1	26.62226	26.62226	12.99103	0.00869			
Residual	7	14.34496	2.04928					
Total	8	40.96722						
	Coefficients	*andard Erı*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*ower 95.0%*	*pper 95.0%*
Intercept	-1.21357	2.058974	-0.58941	0.574111	-6.08227	3.655128	-6.08227	3.655128
Present_Pı	0.909337	0.252292	3.604307	0.00869	0.312762	1.505913	0.312762	1.505913

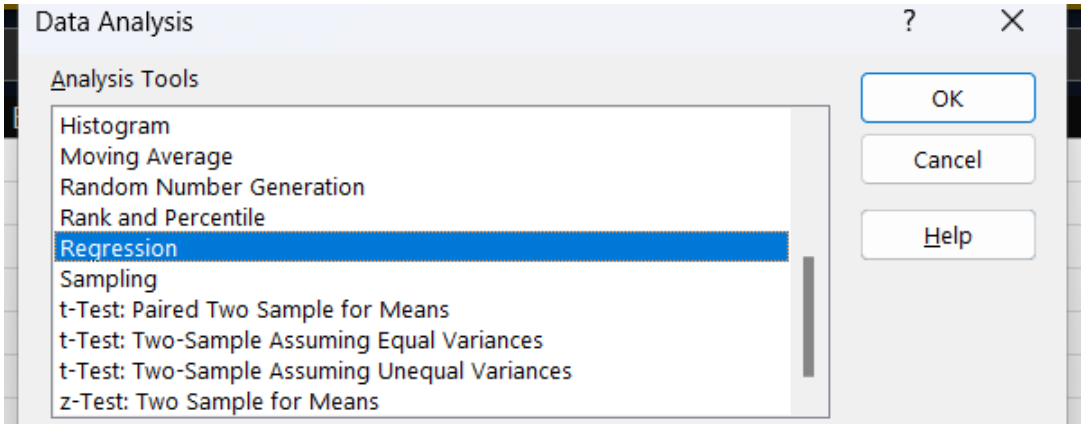
RESIDUAL OUTPUT		
Observatior	*ted Selling_*	*Residuals*
1	3.869624	-0.51962
2	7.461505	-2.71151
3	7.7434	-0.4934
4	2.560178	0.289822
5	5.033575	-0.43358
6	7.725213	1.524787
7	6.170247	0.579753
8	6.615822	-0.11582
9	6.870436	1.879564

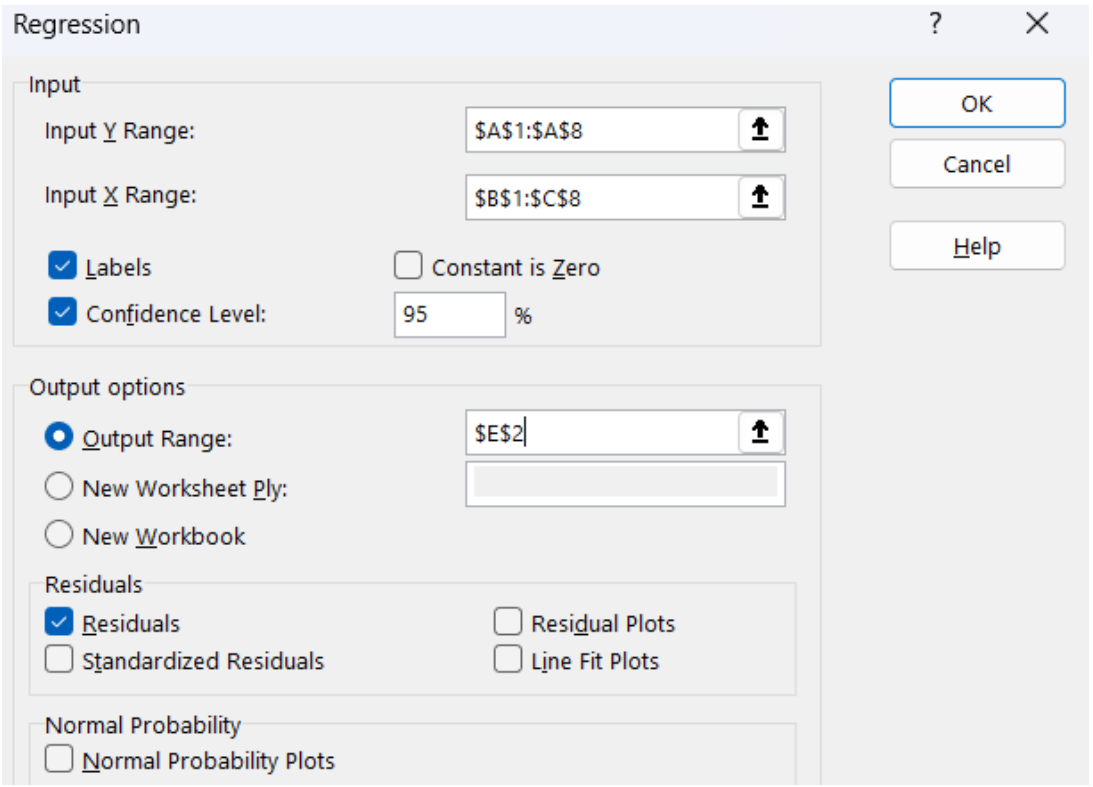
B. Perform multiple regression for prediction.

Quantity Sold	price in dollars	Advertising in dollars
8500	2	2800
4700	5	200
5800	3	400
7400	2	500
6200	5	3200
7300	3	1800
5600	4	900

Go to Data tab > Data Analysis > Regression > Click Ok.







Select **input Y range** – A1:A8.
Select **input X range** – B1:C8.
Check checkboxes for Labels and Confidence Level.
Select **output range –** E2. Check checkbox Residuals.
Click Ok.

Output:

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.980681431
R Square	0.961736068
Adjusted R Square	0.942604102
Standard Error	310.5239249
Observations	7

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	9694299.568	4847149.784	50.26854403	0.001464128
Residual	4	385700.4318	96425.10794		
Total	6	10080000			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	8536.213882	386.9117478	22.06243137	2.49812E-05	7461.974654	9610.453111	7461.974654	9610.453111
price in dollars	-835.7223514	99.65304469	-8.386320297	0.001106064	-1112.40356	-559.0411432	-1112.40356	-559.0411432
Advertising in dollars	0.592228496	0.104346803	5.675578729	0.004755309	0.302515325	0.881941666	0.302515325	0.881941666

RESIDUAL OUTPUT

Observation	*Predicted Quantity Sold*	*Residuals*
1	8523.008967	-23.00896712
2	4476.047825	223.9521754
3	6265.938227	-465.9382265
4	7160.883427	239.1165726
5	6252.733311	-52.73331119
6	7095.05812	204.9418798
7	5726.330123	-126.3301229