UNIT 19 ADVANCED OPTIONS IN SPREADSHEETS

Structure

19.0 OBJECTIVES

After completing this unit, you should be able to:

- understand the process to sort the data in Excel;
- understand the basic features of Excel such as sorting of data in worksheets and computation of data with formulas;
- understand the filtering of data as per the parameters provided;
- search the data using lookup tables;
- analyze the data using descriptive statistics;
- use correlation and regression for decision making for the given dataset; and
- solve the problems using hypothesis testing.

19.1 INTRODUCTION

Spreadsheets are grid-based files designed to organize information and perform calculations with scalable entries. People all around the world use spreadsheets to create tables for any personal or business need. However, spreadsheets have grown from simple grids to powerful tools, functioning like databases or apps that perform numerous calculations on a single sheet. In businesses spreadsheets can be used to determine mortgage payments over time or to help in calculating the depreciation of assets and to see how it will affect business taxes. The data between several sheets can be combined to visualize it in color coded tables for an at-a-glance understanding.

With over 400 functions, MS Excel is more or less the most comprehensive spreadsheet option when it comes to pure calculations. It also has strong visualization abilities, including conditional formatting, Pivot Tables, SmartArt, graphs, and charts. Home and business users alike can create powerful spreadsheets and reports to track data and inform their decisions. Advanced options in spreadsheets such as data searching, filtering, sorting, frequency distribution, descriptive statistics, referencing etc. as explained in detail in the further sections of this unit helps businesses greatly in speedy calculations.

Various advanced, spreadsheet features are required for those aspirants who would like to go for office automation. Organizations need most of the work is to be done without any human errors, thus task automation is the only way out. Thus, understanding and implementation of processes and formulas in this chapter shall help the users.

19.2 SORTING DATA

Sorting is a basic but, an important feature in MS Excel. The raw data is always required to be sorted, before it can be analysed and interpreted further. To apply sorting on data, there are numerous ways available in MS Excel. Sorting can be done, either on the whole sheet or on some specific cell range. Different kinds of Sorting like Alphabetical, Chronological or by Color are available in MS Excel.

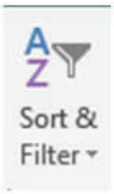
To sort a particular set of data, follow the below steps:

1) Select the range of data to be sorted.

	A	B	C	D	E	F
1	Product	Quantity	Price	Amount		
2	Laptop	8	35000	280000		
3	Desktop	5	30000	150000		
4	Smart Phones	10	25000	250000		
5	Ear Phones	20	600	12000		
6	Head Phones	10	1000	10000		
7	Covers & Cases	20	300	6000		
8	Chargers & Data Cables	20	200	4000		
9						
10						
11						

Fig. 19.1: Selecting the Data

2) Click on the Sort & Filter Button under the Menu Ribbon.



3) Once clicked, a drop-down list would appear, stating the order of sorting to be applied on the data, whether "A to Z", "Z to A" or a Custom List. Select the desired type. Let's choose "A to Z".

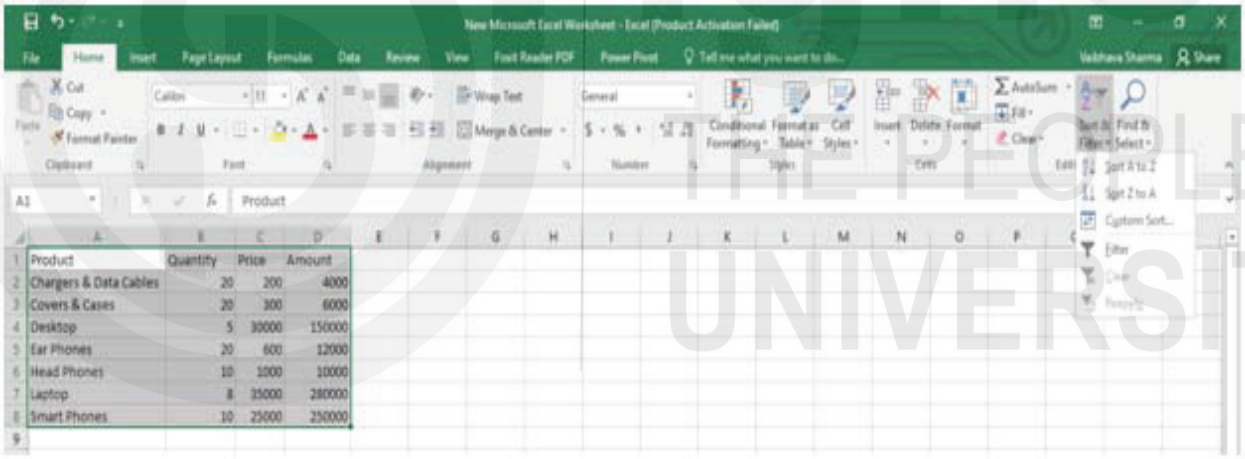


Fig. 19.2: Drop-Down List of Sort & Filter

4) The data would then appear like in Figure 19.3.

	A	B	C	D	E
1	Product	Quantity	Price	Amount	
2	Chargers & Data Cables	20	200	4000	
3	Covers & Cases	20	300	6000	
4	Desktop	5	30000	150000	
5	Ear Phones	20	600	12000	
6	Head Phones	10	1000	10000	
7	Laptop	8	35000	280000	
8	Smart Phones	10	25000	250000	
9					
10					
11					

Fig. 19.3: Result of Sort A to Z

5) Alternatively, a keyboard shortcut of "Ctrl+Shift+L" can also be used after Step 1.

6) Then each column of the header row would get separate drop-down list. Then using that list, the data can be sorted according to any column head, whether Smallest to Largest or Largest to Smallest.

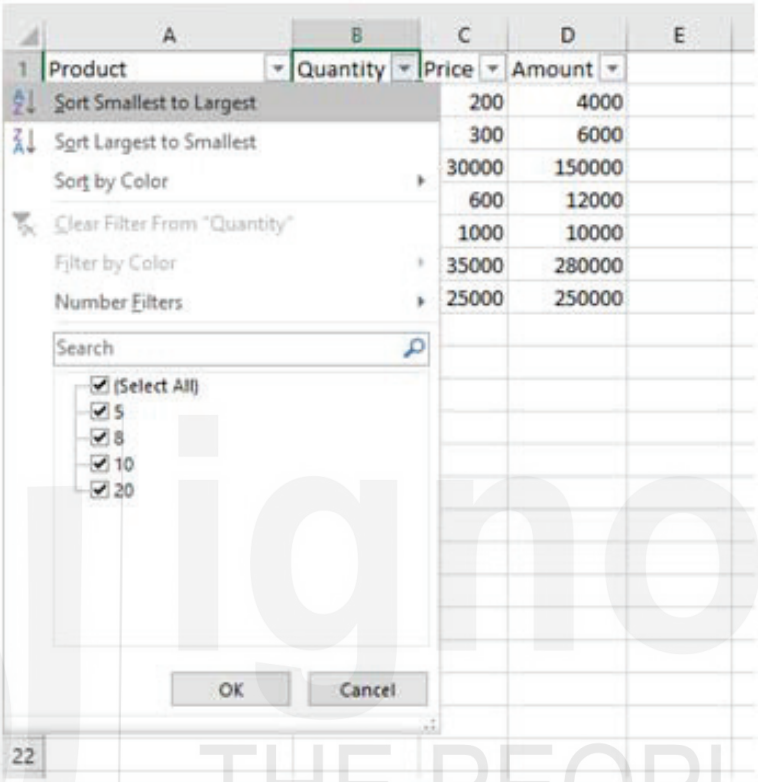


Fig. 19.4: Sort using "Ctrl+Shift+L"

19.3 FILTERING DATA

Like sorting, filtering the data is equally important for analyzing it effectively. Sort & Filter go hand in hand. They are usually used together. The ways to apply a filter are same as applying a sort on the data. Similar to sorting, there are some other filter options also available, like Number Filters and Text Color Filters and Cell Color Filters.

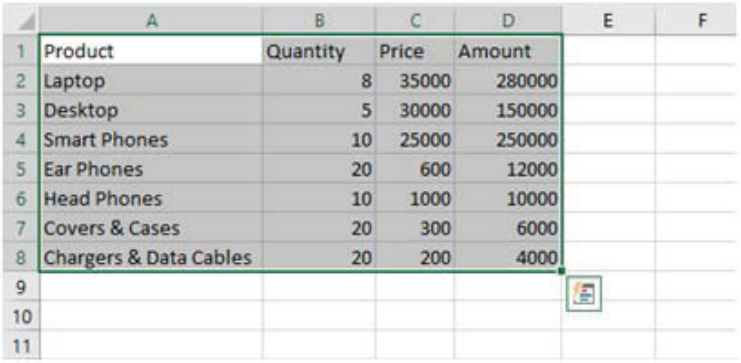
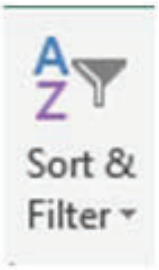


Fig. 19.5: Selecting the data

To filter a particular set of data, follow the below steps:

1) Select the range of data to be sorted.

2) Click on the Sort & Filter Button under the Menu Ribbon.



3) Once clicked, a drop-down list would appear, an option for Filter would appear, click that option for applying filter on your dataset.

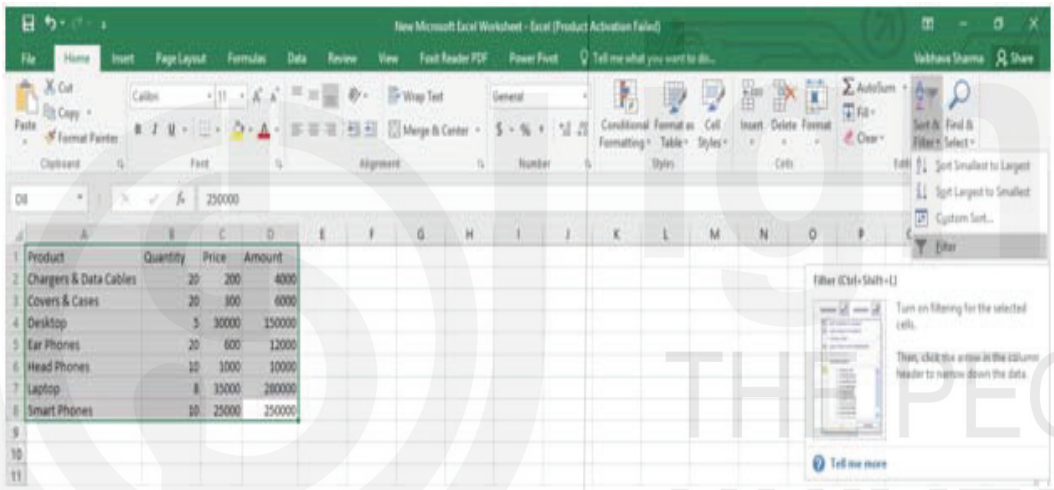


Fig. 19.6: Drop-Down list of Sort & Filter

4) Once clicked, the dataset would appear like in figure19.7 below:

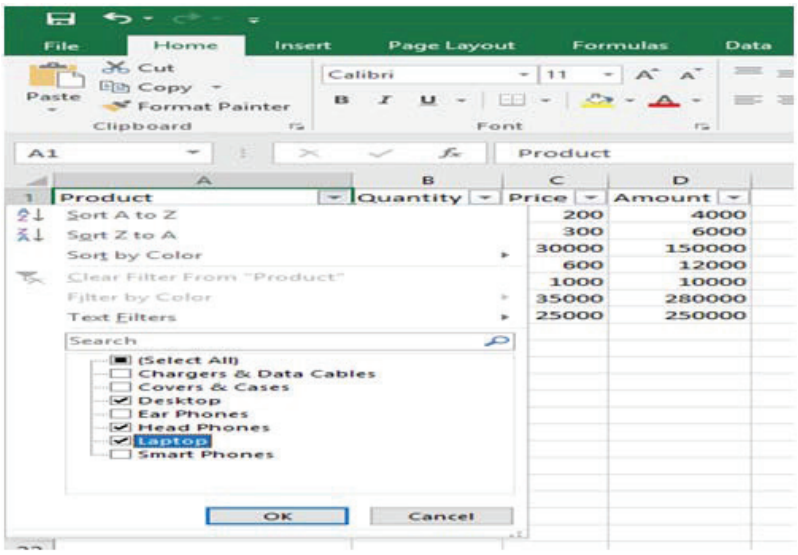


Fig. 19.7: Result of Sort &Filter Option

5) Then we can use the drop-down list on the header row of the data set to filter the data accordingly. For example, as shown in figure 19.8 below:

	A	B	C	D
1	Product	Quantity	Price	Amount
2	Chargers & Data Cables	20	200	4000
3	Covers & Cases	20	300	6000
4	Desktop	5	30000	150000
5	Ear Phones	20	600	12000
6	Head Phones	10	1000	10000
7	Laptop	8	35000	280000
8	Smart Phones	10	25000	250000
9				
10				
11				
12				

Fig. 19.8: Selected Items to filter

6) As the items are selected in figure 19.8, this will result as in figure 19.9 given below:

	A	B	C	D
1	Product	Quantity	Price	Amount
4	Desktop	5	30000	150000
6	Head Phones	10	1000	10000
7	Laptop	8	35000	280000
9				
10				

Fig. 19.9: Result of Filter

7) Alternatively, a keyboard shortcut of "Ctrl+Shift+L" can also be used to apply filters to a dataset.

19.4 SEARCHING DATA

The dataset in MS Excel are usually large and the analysis becomes cumbersome as the dataset increases. If we need to search for some particular item in the whole dataset, MS Excel includes an option to find.

Use the following steps to find an item in the dataset:

1) The dataset consists of the "ID"of students and their respective marks in each of the three subjects.

2) Now, for example if we need to find the marks for a particular student, we can use Find option in MS Excel for that.

3) Use the Find & Select option provided in the Menu Ribbon under the Home Tab.

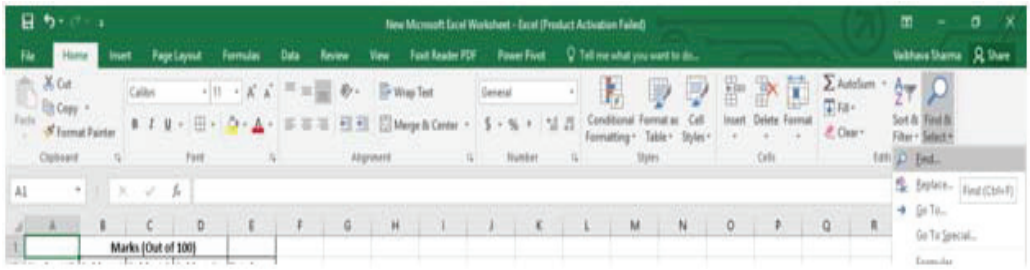


Fig. 19.10: Find & Select Option

4) Click on the Find option. A dialog box would appear as shown in figure 19.11 below:

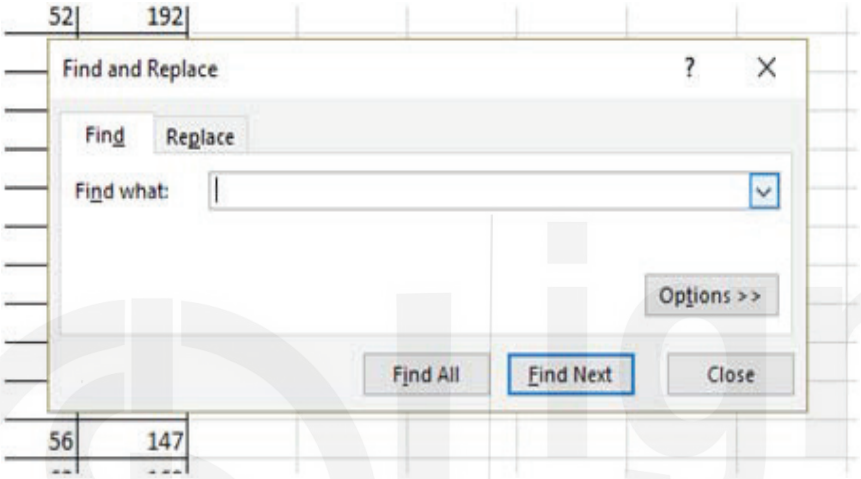


Fig. 19.11: Find & Replace Dialog Box

5) Now we can search for whatever we want to find. Say, we need to know the marks of the student with ID – 4776. We will type in "4776" in the search bar of the dialog box.

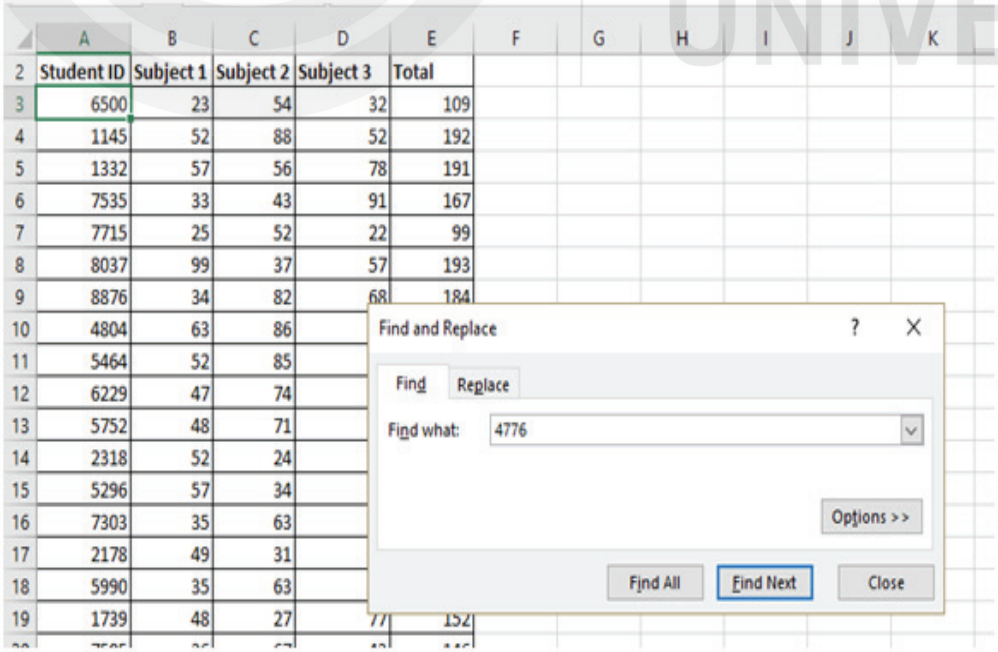


Fig. 19.12: Searching in Find & Replace

6) Once we press "Find Next", it will search for the item in the dataset. If the item exists in the dataset, it will show that item, as in figure 19.13 below:

82	2329	94	79	95	268
83	3585	96	86	90	272
84	4054	100	42	28	170
85	6352	76	21	71	168
86	2462	66	86	47	199
87	2618	38	81	55	174
88	4776	55	74	27	156
89	1625	81	68	83	232
90	7309	22	86	94	202
91	5637	37	91	25	153
92	2827	64	53	31	148
93	6819	83	32	79	194
94	4174	75	58	73	206

Fig. 19.13: Search results

7) If the item does not exist in the dataset, it will return with an error prompt as in figure 19.14 below:

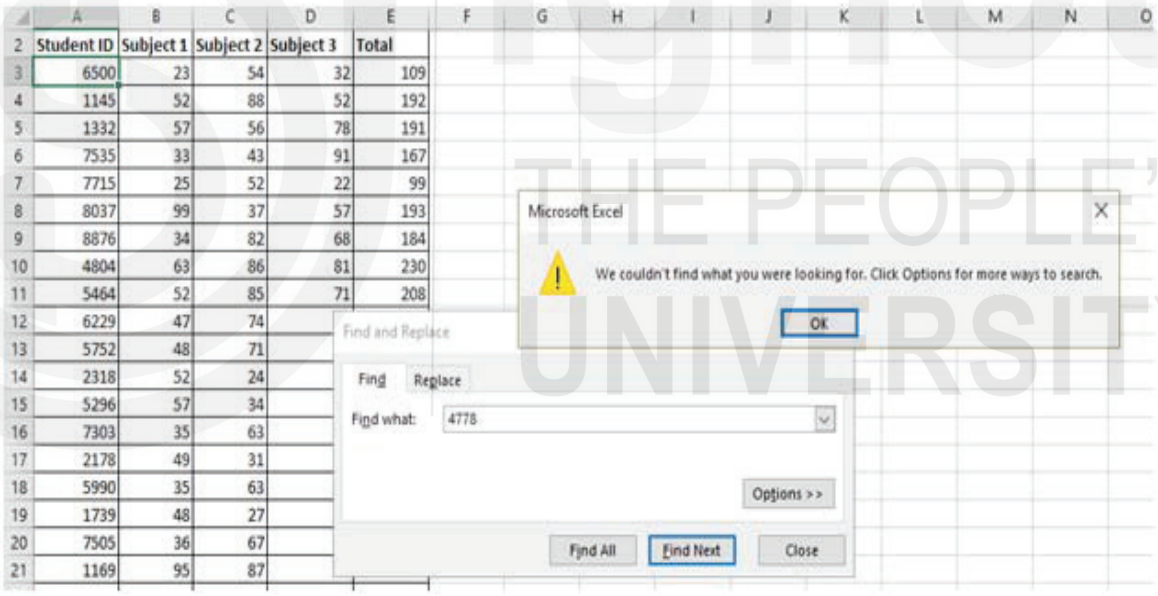


Fig. 19.14: Search Error Prompt

8) Alternatively, we can use a keyboard shortcut "Ctrl+F", to trigger the Find & Replace dialog box and find an item from the dataset.

19.4.1 Lookup

LOOKUP function in MS Excel is used to look in a single row or a column and to find a value from the same position in another row or a column. Its functionality is limited, because it is not entirely eligible to handle situations where an array is involved. Therefore, we use the other two powerful LOOKUP functions. There are mainly two types of Lookup functions in MS

Excel, namely "HLOOKUP" and "VLOOKUP". HLOOKUP is the Horizontal Lookup and VLOOKUP is the Vertical Lookup.

Below is an example of "VLOOKUP". Let us say we want to look for the total marks for a particular Student ID. We will use the formula as shown in figure19.15 below:

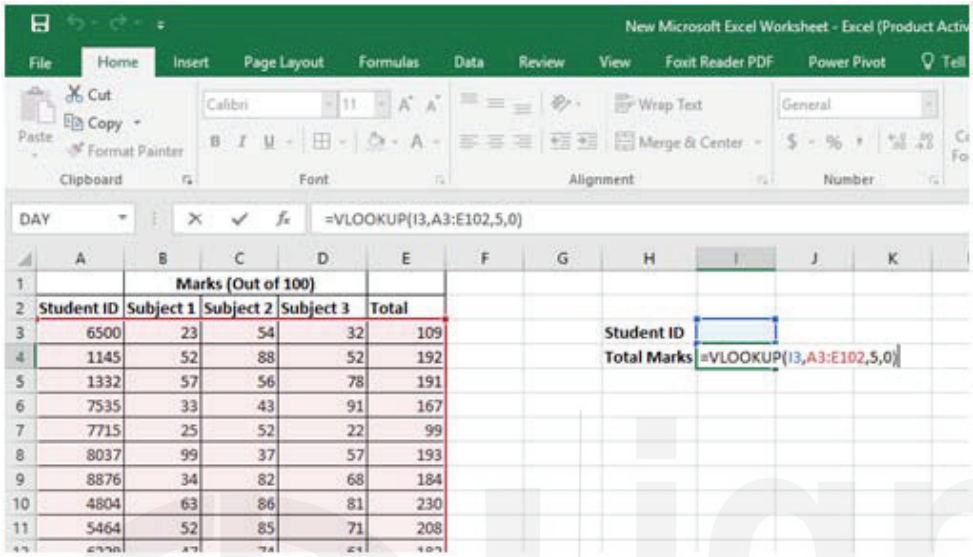


Fig. 19.15: VLOOKUP Formula

The syntax for VLOOKUP is as follows:

=VLOOKUP(lookup_value, table_array, col_index_num, [range_lookup])

The formula will return the total marks in cell "I4", referring to the Student ID entered in Cell "I3". For example, we put in "1073" as student ID. It returned with the total marks as it can be seen in the figure 19.16 below:

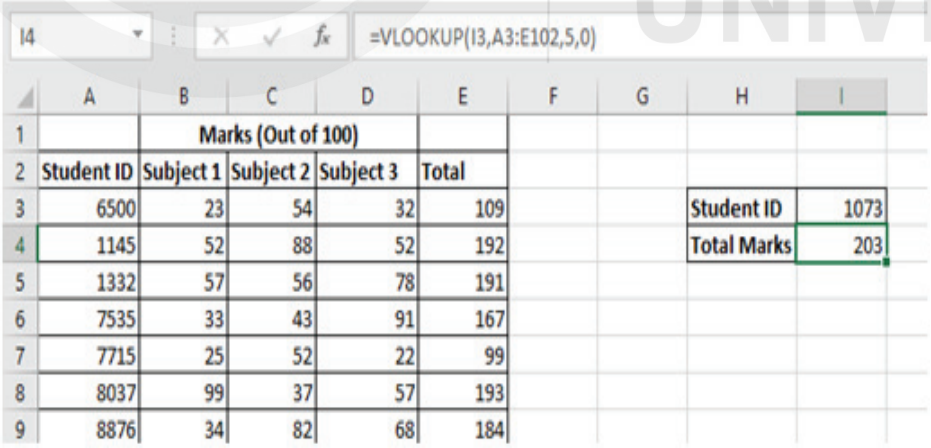


Fig. 19.16: Result of VLOOKUP

19.4.2 Referencing

Cell referencing is an important feature of MS Excel. It refers to a cell or a range of cells on a sheet, which is to be used in a formula so that MS Excel can find those values, that is required by the formula to be used.

A cell reference can be used to refer to:

1) Data from one or more contiguous cells on the worksheet

2) Data contained in different areas of a worksheet.

3) Data on the other worksheets in the same workbook.

Below are some examples of the cell referencing formulas that are used in MS Excel:

Table 19.1: Cell Referencing Formulas

Formula	Refers to	returns
=b4	Cell B4	The value in the Cell B4.
=a4:c8	Cells from A4 through C8	Values in all the cells. To use this referencing, we need to press "Ctrl+Shift+Enter" when the formula is entered.
=Sheet2!B1	Cell B1 on Sheet2	The value in B1 on Sheet2.

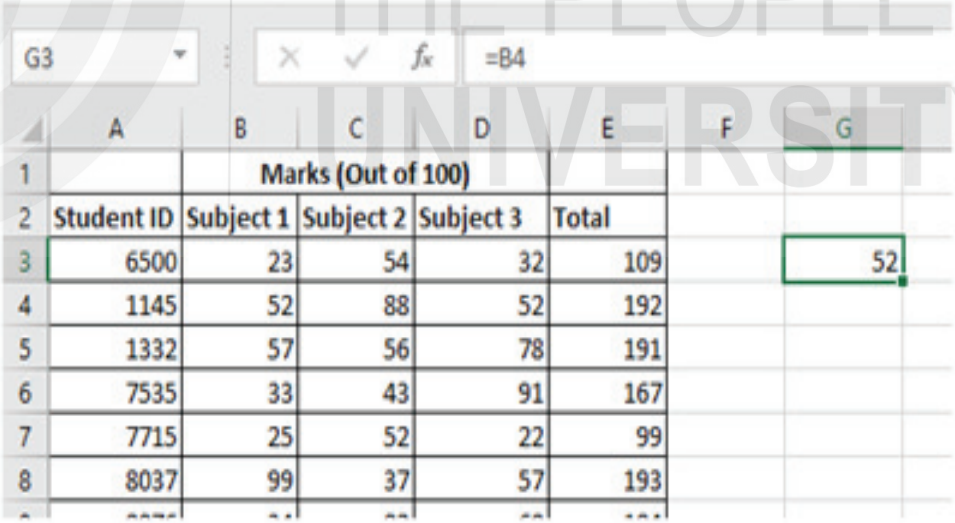


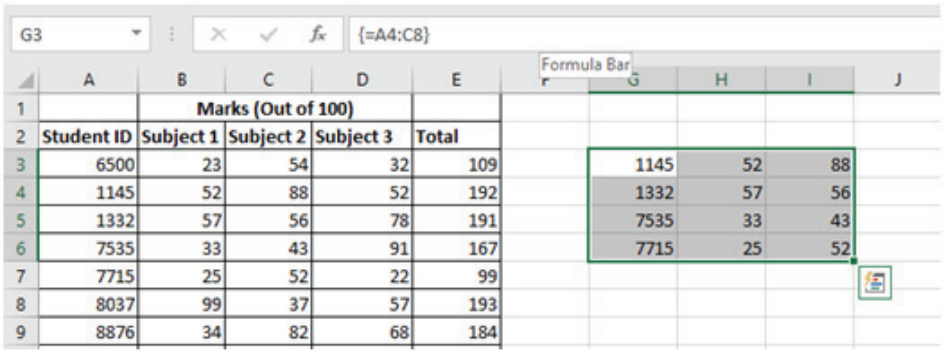
Fig. 19.17: "=B4"

1) "=B4" :-

2) "=A4:C8" :-

For using this cell reference, we first need to select the area in which we need the array to fill the values. Since it is A4:C8, it means three columns and four

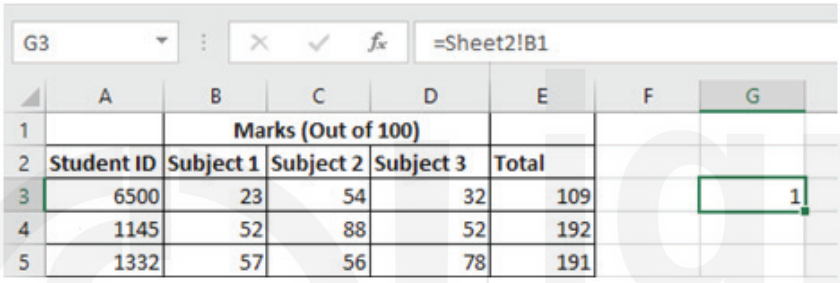
rows. Therefore, we will select such an area with three columns and four rows.



	A	B	C	D	E	G	H	I
1		Marks (Out of 100)						
2	Student ID	Subject 1	Subject 2	Subject 3	Total			
3	6500	23	54	32	109	1145	52	88
4	1145	52	88	52	192	1332	57	56
5	1332	57	56	78	191	7535	33	43
6	7535	33	43	91	167	7715	25	52
7	7715	25	52	22	99			
8	8037	99	37	57	193			
9	8876	34	82	68	184			

Fig. 19.18: "=A4:C8"

3) "=Sheet2!B1" :-



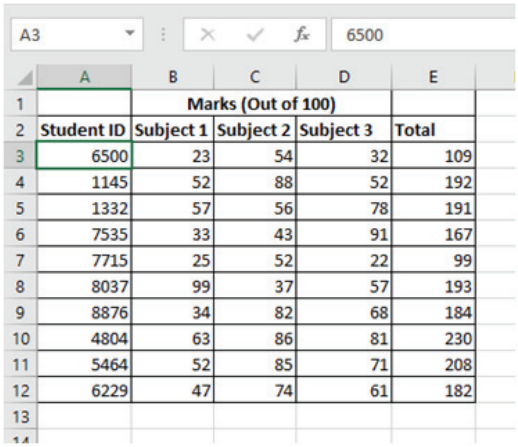
	A	B	C	D	E	F	G
1		Marks (Out of 100)					
2	Student ID	Subject 1	Subject 2	Subject 3	Total		
3	6500	23	54	32	109		1
4	1145	52	88	52	192		
5	1332	57	56	78	191		

Fig. 19.19: "=Sheet2!B1"

19.5 FREQUENCY DISTRIBUTION USING ARRAY FORMULAS

Frequency distribution table is a useful statistic that shows the separate values for various outcomes in a sample data set. The values are the number of times that particular outcome has occurred in the sample data set. Using MS Excel's formula "FREQUENCY", we can create a frequency distribution table for a particular data set.

For example, let us take the below data set in figure 19.20 as a sample.



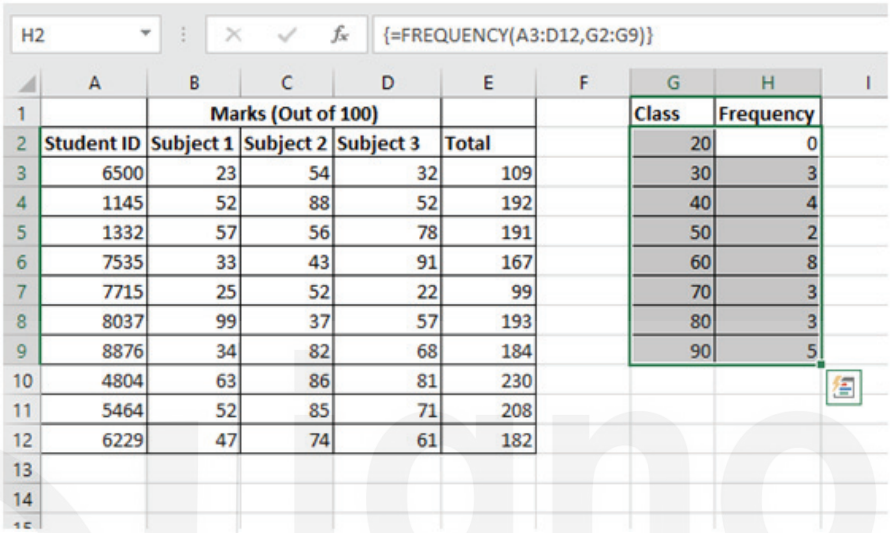
	A	B	C	D	E
1		Marks (Out of 100)			
2	Student ID	Subject 1	Subject 2	Subject 3	Total
3	6500	23	54	32	109
4	1145	52	88	52	192
5	1332	57	56	78	191
6	7535	33	43	91	167
7	7715	25	52	22	99
8	8037	99	37	57	193
9	8876	34	82	68	184
10	4804	63	86	81	230
11	5464	52	85	71	208
12	6229	47	74	61	182
13					

Fig. 19.20: Sample dataset

Now we will use the "FREQUENCY" Formula on this data set. Remember we need to use the array formula here, that is, we need to select the area of cells for the frequency table to set in and use "Ctrl+Shift+Enter" to enter the formula.

Syntax for FREQUENCY is as follows:

=FREQUENCY(data_array, bins_array)



	A	B	C	D	E	F	G	H	I
1		Marks (Out of 100)					Class	Frequency	
2	Student ID	Subject 1	Subject 2	Subject 3	Total		20	0	
3	6500	23	54	32	109		30	3	
4	1145	52	88	52	192		40	4	
5	1332	57	56	78	191		50	2	
6	7535	33	43	91	167		60	8	
7	7715	25	52	22	99		70	3	
8	8037	99	37	57	193		80	3	
9	8876	34	82	68	184		90	5	
10	4804	63	86	81	230				
11	5464	52	85	71	208				
12	6229	47	74	61	182				
13									
14									

Fig. 19.21: Frequency Table

As we can see in the figure 19.21, a frequency table is created with class in the first column and frequency of that class in the second column.

Check Your Progress A

1) What are the various data filter options available in MS Excel?

……

……

……

2) What is the syntax for VLOOKUP?

……

……

……

……

……

3) Give some examples of cell referencing.

……

……

……

4) Write the syntax for FREQUENCY.

……

……

……

……

……

19.6 LOADING DATA ANALYSIS TOOLPAK

Data Analysis ToolPak is available in MS Excel to develop and solve complex and more detailed statistical problems. We provide the data set and parameters to the system and it uses the appropriate statistical or engineering tool to solve the problem.

Some tools also make the use of charts for better analysis and the best possible solution for that particular problem. Data analysis can only be used on one worksheet at a time.

See the figure 19.22 for where to find the Data Analysis ToolPak in MS Excel.

1) Go To Data tab from the Menu Ribbon
2) Click on the "Data Analysis" Option.

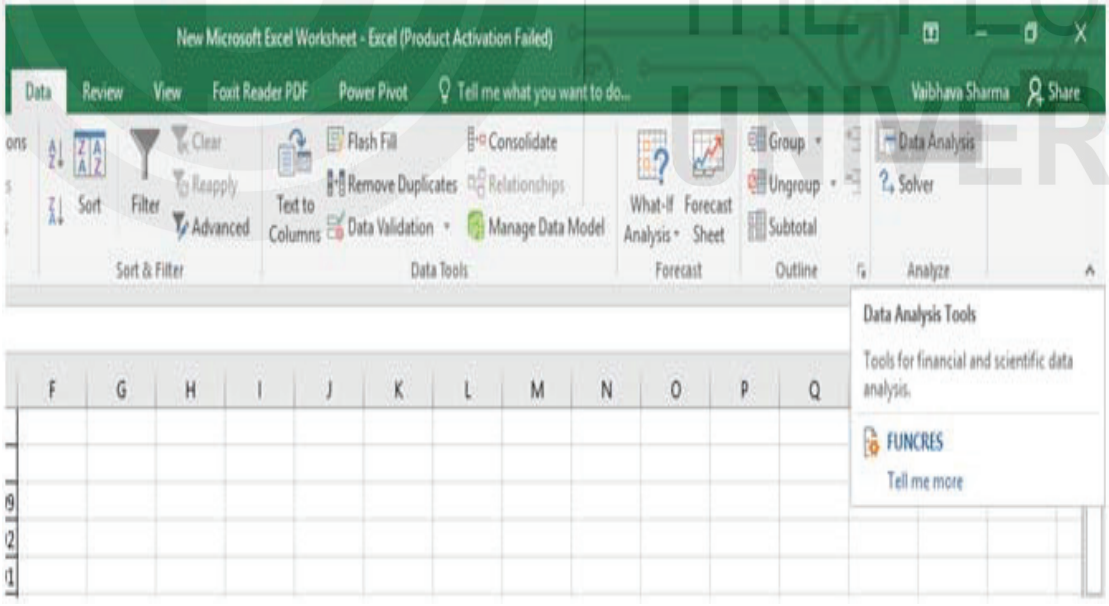


Fig.19.22: Data Analysis ToolPak

3) After clicking on that option, a dialog box appears, as shown in the below figure 19.23.

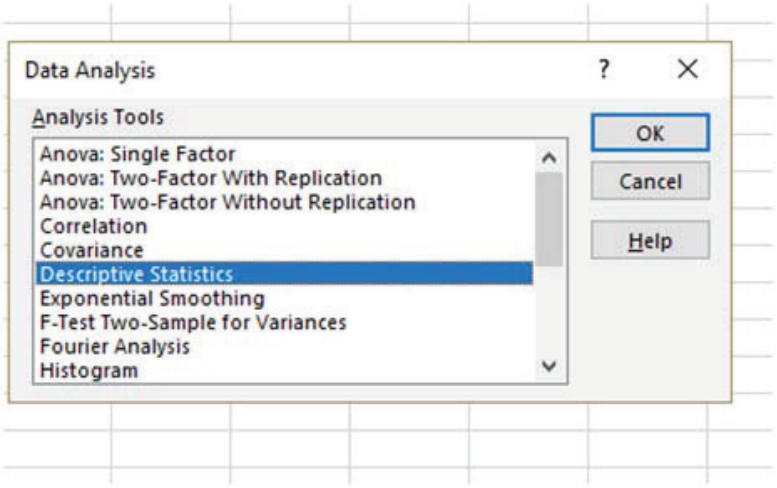


Fig. 19.23: Data Analysis Toolpak Dialog Box

4) From here, we can select the statistical tool that we want to apply on our data set.

19.7 DESCRIPTIVE STATISTICS

Descriptive statistics is a summary of statistical methods, which quantitatively describes or summarizes a sample data set. In MS Excel, it includes methods such as Mean, Standard Deviation, Median, Mode, Range, and Sample Variance and so on.

For example, we want to apply descriptive statistics on the below data set (Figurer. 19.24):-

	A
1	Player Runs in last 20 matches
2	154
3	180
4	118
5	182
6	135
7	70
8	126
9	21
10	158
11	24
12	86
13	49
14	158
15	123
16	21
17	26
18	90
19	5
20	137
21	109
22	

Fig. 19.24: Sample data set for Descriptive statistics

1) Click on Data Analysis tab from Data Tab in Menu Ribbon.

2) Then select "Descriptive Statistics" from the dialog box that appears on screen.

3) After that, following dialog box would appear.

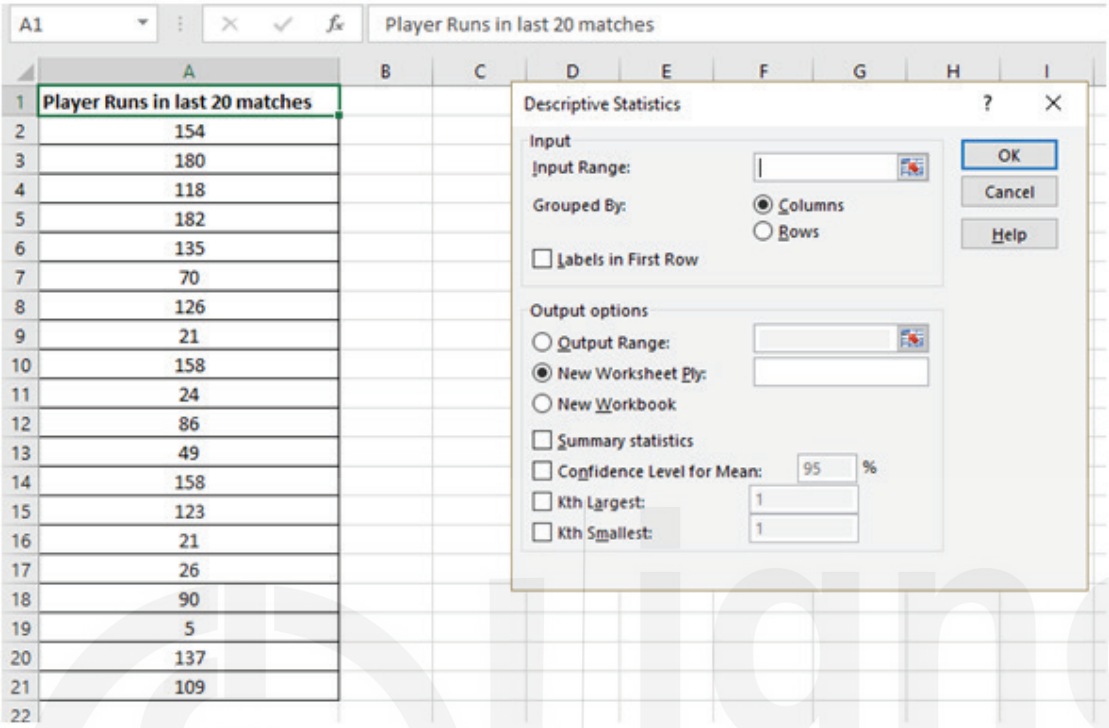


Fig. 19.25: Descriptive Statistics Dialog box

4) Then put in the set of options or as required for the analysis. For example, we have applied the below stated options for this analysis example.

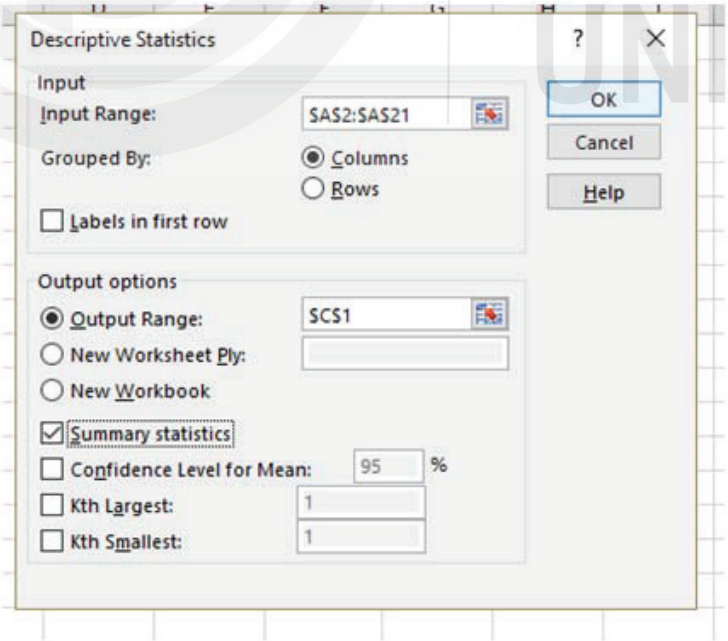


Fig. 19.26: Descriptive Statistics Options

5) After pressing OK, following table will be created.

	A	B	C	D
1	Player Runs in last 20 matches		Column1	
2	154			
3	180		Mean	98.6
4	118		Standard Error	12.92578003
5	182		Median	113.5
6	135		Mode	21
7	70		Standard Deviation	57.80584563
8	126		Sample Variance	3341.515789
9	21		Kurtosis	-1.29819608
10	158		Skewness	-0.255357765
11	24		Range	177
12	86		Minimum	5
13	49		Maximum	182
14	158		Sum	1972
15	123		Count	20
16	21			
17	26			
18	90			
19	5			
20	137			
21	109			

Fig. 19.27: Descriptive Statistics Result

19.8 CORRELATION & REGRESSION

Correlation & Regression are also applied through the Data Analysis ToolPak from MS Excel.

To assess the strength of the linear relationship between a pair of variables, correlation coefficients are put to use. Similarly, to assess that how many of the independent variables are related to the dependent variable, and the strength of their bond, Regression Analysis is put to use.

Correlation:

1) Let us take the below stated example.

	A	B
1	Player Runs in last 20 matches	Team Total Runs
2	154	394
3	180	389
4	118	256
5	182	354
6	135	198
7	70	300
8	126	259
9	21	364
10	158	301
11	24	145
12	86	246
13	49	288
14	158	267
15	123	308
16	21	360
17	26	234
18	90	239
19	5	201
20	137	350
21	109	295
22		

Fig. 19.28: Example for Correlation

2) Now, similar to Descriptive Statistics we will use the Data Analysis ToolPak.

3) This time the following dialog box would show up.

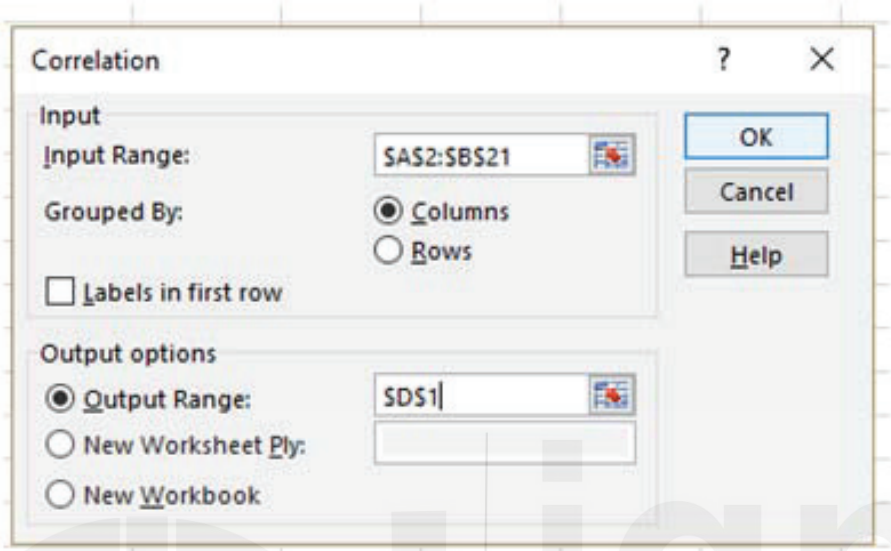


Fig. 19.29: Correlation Dialog Box

4) You can choose to keep the same settings as in Fig.19.29, or choose your own settings according to the requirements.

5) This would result in the following result (Fig.19.30)

	Column 1	*Column 2*
Column 1	1	
Column 2	0.394	1

Fig. 19.30: Correlation Result

6) The result shows a positive correlation of 0.394 in the Runs scored by the Player in a match and the Total runs of the team in that match.

Regression:

1) Let us take the below stated example. We want to predict the Team score if we know the runs scored by Player 1 and Player 2.

	A	B	C
1	Player 1	Player 2	Team Total Runs
2	41	42	394
3	45	30	389
4	45	76	256
5	16	30	354
6	61	93	198
7	108	10	300
8	17	28	259
9	70	64	364
10	113	2	301
11	87	30	145
12	62	2	246
13	60	64	288
14	62	89	267
15	116	45	308
16	49	76	360
17	20	84	234
18	83	114	239
19	48	92	201
20	19	46	350
21	12	26	295
22			

Fig. 19.31: Regression Example

2) Similar to Correlation, we will follow the same process to reach the below given dialog box. (Fig.19.32)

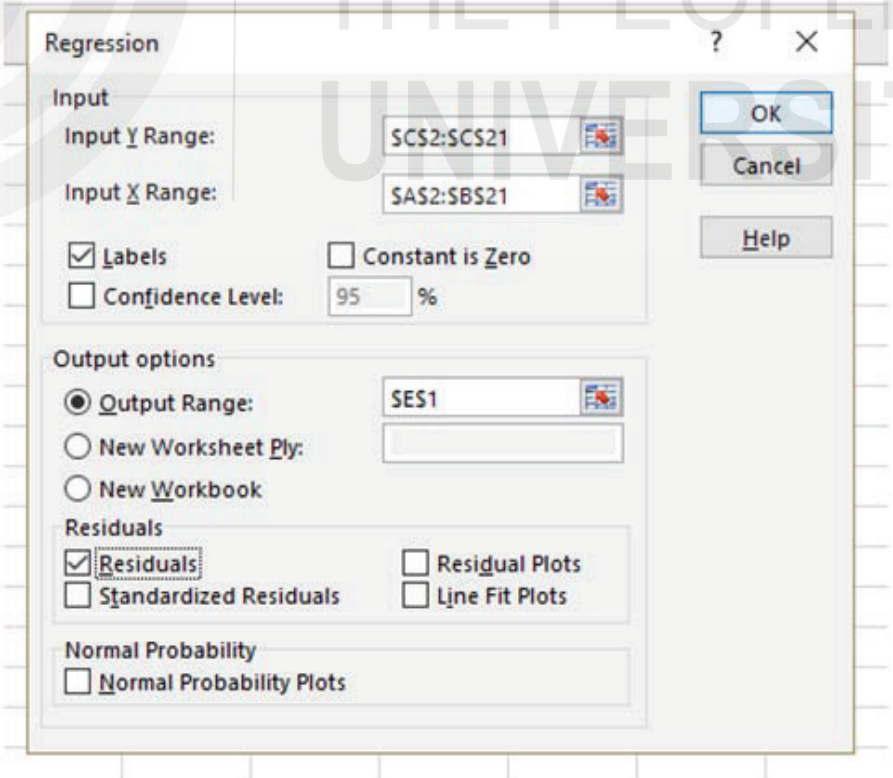


Fig. 19.32: Regression Dialog Box

3) You can choose to keep the same settings as in Fig.19.32, or choose your own settings according to the requirements.

4) After clicking OK, it will result in the following figure (Fig.19.33 & 19.34).

5) There will be two kinds of output.

   a. Summary Output

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.319
R Square	**0.102**
Adjusted R Square	-0.010
Standard Error	64.835
Observations	19

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	7638.212	3819.106	0.909	0.423
Residual	16	67256.946	4203.559		
Total	18	74895.158			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	331.158	41.133	8.051	0.000	243.960	418.356	243.960	418.356
41	-0.355	0.467	-0.759	0.459	-1.345	0.636	-1.345	0.636
42	-0.550	0.457	-1.204	0.246	-1.518	0.419	-1.518	0.419

Fig. 19.33: Summary Output

   b. Residual Output

RESIDUAL OUTPUT

Observation	*Predicted 394*	*Residuals*
1	299	90
2	273	-17
3	309	45
4	258	-60
5	287	13
6	310	-51
7	271	93
8	290	11
9	284	-139
10	308	-62
11	275	13
12	260	7
13	265	43
14	272	88
15	278	-44
16	239	0
17	264	-63
18	299	51
19	313	-18

Fig. 19.34: Residual Output

19.9 HYPOTHESIS TESTING

Hypothesis testing in MS Excel are available in various types (figure 19.35):

- t-Test: Paired two sample for means
- t-Test: Two-Sample Assuming Equal Variances
- t-Test: Two-Sample Assuming Unequal Variances
- z-Test: Two-Sample for Means

Hypothesis Testing is used when we need to check whether the solution is moving in the correct direction according to the problem, or not. It is a statistical method used to make decisions by using sample data, out of the population.

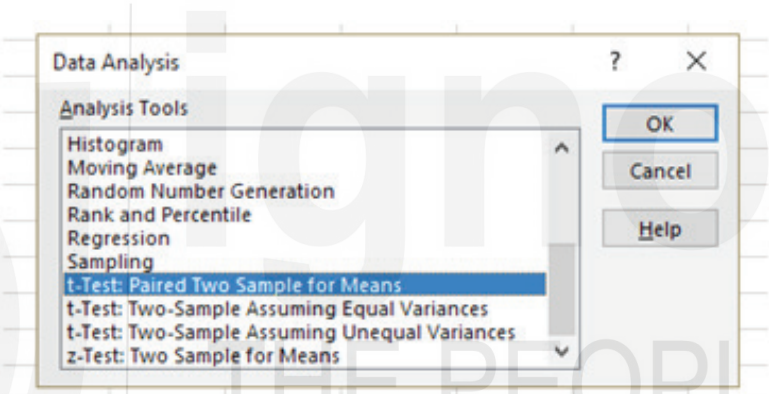


Fig. 19.35: Types of Hypothesis Testing

We will take the same example as taken for Regression in figure 19.31. We will use the t-Test: Paired Two Sample for Means for this example. After we select this option from the Data Analysis ToolPak, following dialog box would appear as shown in figure 19.36.

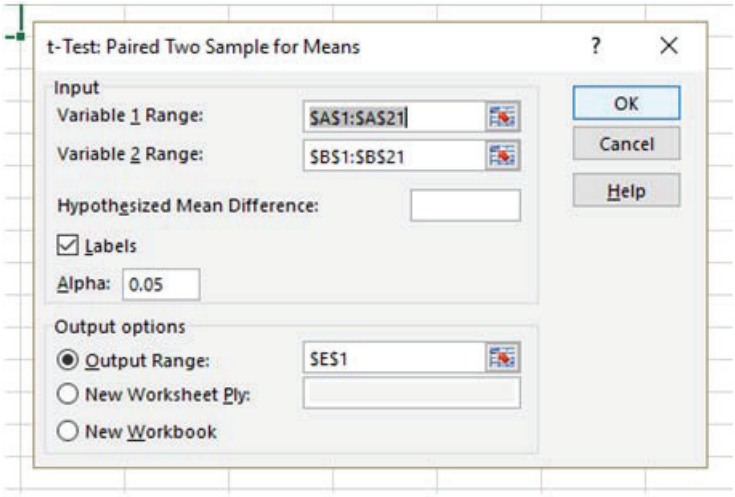


Fig. 19.36: Hypothesis Testing Dialog Box

The result of this selection would appear as follows: (Fig 19.37)

t-Test: Paired Two Sample for Means		
	Player 1	*Player 2*
Mean	56.700	52.150
Variance	1044.432	1084.766
Observations	20.000	20.000
Pearson Correlation	-0.121	
Hypothesized Mean Difference	0.000	
df	19.000	
t Stat	0.416	
P(T<=t) one-tail	0.341	
t Critical one-tail	1.729	
P(T<=t) two-tail	0.682	
t Critical two-tail	2.093	

Fig. 19.37: Result of Hypothesis Testing

It results that the Mean Runs scored by Player 1 is 56.7 and Player 2 has 52.15. The Pearson correlation is -0.121, which depicts that they have a weak inverse relation with each other.

Check Your Progress B

1) What is Data Analysis ToolPak?

……

……

……

……

……

……

2) What do you understand by descriptive statistics?

……

……

……

……

……

3) Distinguish between correlation & regression?

…….

…….

…….

…….

4) What do you understand by hypothesis testing?

…….

…….

…….

…….

…….

19.10 LET US SUM UP

Sorting is a basic but, an important feature in MS Excel. The raw data is always required to be sorted, before it can be analysed and interpreted further. To apply sorting on data, there are numerous ways available in MS Excel. Like sorting, filtering the data is equally important for analyzing it effectively. Sort & Filter go hand in hand. They are usually used together.

LOOKUP function in MS Excel is used to look in a single row or a column and to find a value from the same position in another row or a column. There are mainly two types of Lookup functions in MS Excel, namely "HLOOKUP" and "VLOOKUP". HLOOKUP is the Horizontal Lookup and VLOOKUP is the Vertical Lookup.

Data Analysis ToolPak is available in MS Excel to develop and solve complex and more detailed statistical problems. Descriptive statistics is a summary of statistical methods, which quantitatively describes or summarizes a sample data set. In MS Excel, it includes methods such as Mean, Standard Deviation, Median, Mode, Range, and Sample Variance and so on.

Correlation & Regression are applied through the Data Analysis ToolPak from MS Excel. To assess the strength of the linear relationship between a pair of variables, correlation coefficients are put to use. Similarly, to assess that how many of the independent variables are related to the dependent variable, and the strength of their bond, Regression Analysis is put to use.

Hypothesis Testing is used when we need to check whether the solution is moving in the correct direction according to the problem, or not. In MS Excel various types of Hypothesis testing are available; such as t-Test for Paired two sample for means, t-Test for Two-Sample Assuming Equal Variances, t-

Test for Two-Sample Assuming Unequal Variances, z-Test for Two-Sample for Means.

19.11 KEY WORDS

Lookup: LOOKUP function in MS Excel is used to look in a single row or a column and to find a value from the same position in another row or a column. HLOOKUP is the Horizontal Lookup and VLOOKUP is the Vertical Lookup.

Cell Referencing: Cell referencing is an important feature of MS Excel. It refers to a cell or a range of cells on a sheet, which is to be used in a formula so that MS Excel can find those values, that is required by the formula to be used.

Descriptive Statistics: Descriptive statistics is a summary of statistical methods, which quantitatively describes or summarizes a sample data set. They can be broken down into measures of central tendency and measures of variability (spread).

Frequency Distribution: Frequency distribution table is a useful statistic that shows the separate values for various outcomes in a sample data set. The values are the number of times that particular outcome has occurred in the sample data set.

Correlation: Correlation analysis is a method of statistical evaluation used to study the strength of a relationship between two, numerically measured, continuous variables. This particular type of analysis is useful when a researcher wants to establish if there are possible connections between variables.

Regression: Regression analysis is a set of statistical methods used for the estimation of relationships between a dependent variable and one or more independent variables. It can be utilized to assess the strength of the relationship between variables and for modeling the future relationship between them.

Hypothesis: Hypothesis Testing is used when we need to check whether the solution is moving in the correct direction according to the problem, or not. It is a statistical method used to make decisions by using sample data, out of the population.

19.12 TERMINAL QUESTIONS

1) Explain LOOKUP, VLOOKUP and HLOOKUP with examples.
2) What are various steps involved in data sorting?
3) Explain frequency distribution using array formulas.

4) Complete the following table for regression analysis. Write down the appropriate formula in the relevant cells.

	A	B	C	D	E
1	Year	Output	Estimated Output	Slope	?
2	2001	100	?	Intercept	?
3	2002	125	?		
4	2003	190	?		
5	2004	210	?		

5) What are the various types available in Excel for hypothesis testing?

6) What is the significance of Data Analysis ToolPak in MS Excel?

Note: These questions are helpful to understand this unit. Do efforts for writing the answer of these questions but do not send your answer to university. It is only for yours practice.