UNIT 19 ADVANCED OPTIONS IN SPREADSHEETS

Structure

19.0 Objectives
19.1 Introduction
19.2 Sorting Data
19.3 Filtering Data
19.4 Searching Data
  19.4.1 Lookup
  19.4.2 Referencing
19.5 Frequency Distribution Using Array Formulas
19.6 Loading Data analysis ToolPak
19.7 Descriptive Statistics
19.8 Correlation & Regression
19.9 Hypothesis Testing
19.10 Let Us Sum Up
19.11 Key Words
19.12 Terminal Questions

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.
Fig. 19.1: Selecting the Data

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

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

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”.

4) The data would then appear like in Figure 19.3.
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"](image)

### 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](image)
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.

4) Once clicked, the dataset would appear like in figure 19.7 below:
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:

![Fig. 19.8: Selected Items to filter](image)

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

![Fig. 19.9: Result of Filter](image)

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

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.
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:

![Fig. 19.13: Search results](image)

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](image)

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:

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:

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**

<table>
<thead>
<tr>
<th>Formula</th>
<th>Refers to</th>
<th>returns</th>
</tr>
</thead>
<tbody>
<tr>
<td>=b4</td>
<td>Cell B4</td>
<td>The value in the Cell B4.</td>
</tr>
<tr>
<td>=a4:c8</td>
<td>Cells from A4 through C8</td>
<td>Values in all the cells. To use this referencing, we need to press “Ctrl+Shift+Enter” when the formula is entered.</td>
</tr>
<tr>
<td>=Sheet2!B1</td>
<td>Cell B1 on Sheet2</td>
<td>The value in B1 on Sheet2.</td>
</tr>
</tbody>
</table>

![Image of Excel spreadsheet showing formulas and cell references](image.png)

**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
Advanced Options in Spreadsheets

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

![Fig. 19.18: "=A4:C8"](image)

3) “=Sheet2!B1” :-

![Fig. 19.19: "=Sheet2!B1"](image)

### 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.

![Fig. 19.20: Sample dataset](image)
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)

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.

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

Fig.19.22: Data Analysis ToolPak
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 (Figure 19.24):

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.

![Descriptive Statistics Dialog box]

**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.

![Descriptive Statistics Options]

**Fig. 19.26: Descriptive Statistics Options**

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

![Example for Correlation](image)
2) Now, similar to Descriptive Statistics we will use the Data Analysis ToolPak.

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

![Correlation Dialog Box](image)

**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)

![Correlation Result](image)

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

   Fig. 19.33: Summary Output

   b. Residual Output

   ![Residual Output Table]

   Fig. 19.34: Residual Output
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.

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

<table>
<thead>
<tr>
<th></th>
<th>Player 1</th>
<th>Player 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean</td>
<td>56.700</td>
<td>52.150</td>
</tr>
<tr>
<td>Variance</td>
<td>1044.432</td>
<td>1084.766</td>
</tr>
<tr>
<td>Observations</td>
<td>20.000</td>
<td>20.000</td>
</tr>
<tr>
<td>Pearson Correlation</td>
<td>-0.121</td>
<td></td>
</tr>
<tr>
<td>Hypothesized Mean Difference</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>df</td>
<td>19.000</td>
<td></td>
</tr>
<tr>
<td>t Stat</td>
<td>0.416</td>
<td></td>
</tr>
<tr>
<td>P(T&lt;=t) one-tail</td>
<td>0.341</td>
<td></td>
</tr>
<tr>
<td>t Critical one-tail</td>
<td>1.729</td>
<td></td>
</tr>
<tr>
<td>P(T&lt;=t) two-tail</td>
<td>0.682</td>
<td></td>
</tr>
<tr>
<td>t Critical two-tail</td>
<td>2.093</td>
<td></td>
</tr>
</tbody>
</table>

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-
183
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.

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Year</td>
<td>Output</td>
<td>Estimated Output</td>
<td>Slope</td>
<td>?</td>
</tr>
<tr>
<td>2</td>
<td>2001</td>
<td>100</td>
<td>?</td>
<td>Intercept</td>
<td>?</td>
</tr>
<tr>
<td>3</td>
<td>2002</td>
<td>125</td>
<td>?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2003</td>
<td>190</td>
<td>?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>2004</td>
<td>210</td>
<td>?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

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.