

**THE USE OF MICROSOFT EXCEL FOR DATA ANALYSIS:  
DESCRIPTIVE STATISTICS**

**BY**

**ENGR. A. A. ABDULLAH, PhD  
DEPARTMENT OF MECHANICAL ENGINEERING  
FEDERAL UNIVERSITY OF TECHNOLOGY, MINNA**

**A TECHNICAL PAPER PRESENTED AT THE MONTHLY MEETING OF  
NIGERIA INSTITUTION OF MECHANICAL ENGINEERS (NIMECHE),  
MINNA CHAPTER, HELD ON SATURDAY 10<sup>TH</sup> AUGUST, 2019, AT THE  
NIGERIA SOCIETY OF ENGINEERS (NSE) SECRETARIAT, MINNA.**

**AUGUST, 2019**

# **THE USE OF MICROSOFT EXCEL FOR DATA ANALYSIS: DESCRIPTIVE STATISTICS**

Engr. A.A. Abdullah, PhD  
Department of Mechanical Engineering  
Federal University of Technology, Minna  
aliuaabdullah@futminna.edu.ng  
+2348034044953

## **Executive Summary**

This article explores the application of Microsoft Excel as a tool for performing descriptive statistical analysis. Descriptive statistics provide a foundation for understanding datasets through a number of measures which include: mean, median, and standard deviation. Microsoft Excel offers accessible and efficient tools for conducting various data analyses. Case studies in academic performance and manufacturing output demonstrate its practical relevance. Data analysis tools in Microsoft excel enables professionals to extract meaningful insights, identify trends, and support decision-making processes efficiently.

## **1. Introduction**

In today's data-driven world, the ability to analyse and interpret data efficiently is a fundamental skill across engineering, science, business, and social sciences (Provost and Fawcett, 2013). Data analysis enables professionals to extract meaningful insights, identify trends, and support decision-making processes. Among the various tools available for data analysis, Microsoft Excel remains one of the most widely used due to its accessibility, flexibility, and powerful built-in functions and various tool packs. Descriptive statistics focus on analysis of raw data into meaningful measures such as mean, median, standard deviation, and distribution patterns which cover presentation, visualisation and interpretation (Triola and Iossi).

This technical session is design to be hand-on exercise as participant are expected to perform statistical data analysis operations with Microsoft Excel without requiring advanced programming knowledge, making it especially useful for all member.

## 2. Aim and Objectives of the Presentation

To demonstrate the application of Microsoft Excel in performing descriptive statistical analysis for effective data interpretation. The objectives are to:

- (i.) explain the concept and importance of descriptive statistics
- (ii.) highlight key descriptive statistical measures
- (iii.) demonstrate how Microsoft Excel can be used for data analysis
- (iv.) apply Microsoft Excel tools to real-life case studies

## 3. Microsoft Excel and Descriptive Statistics

Microsoft Excel provides a user-friendly software developed by Microsoft for window operating system. To open/ lunch Microsoft Excel in windows go *Start – Programs – Microsoft Office – Excel*, as shown in Figure 1.

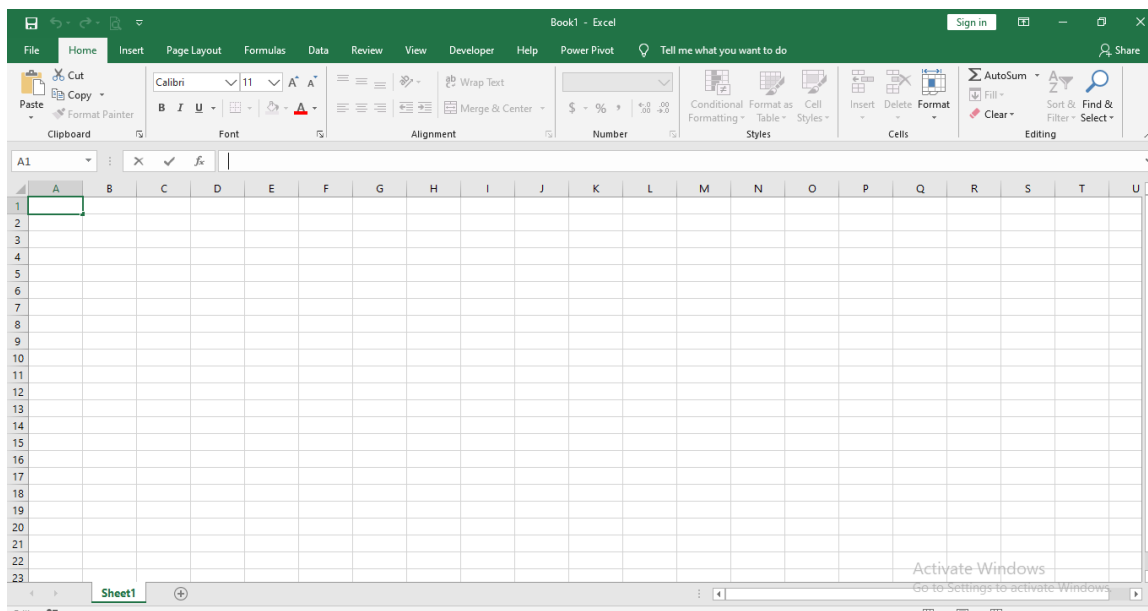


Figure 1: Microsoft Excel Interface of Version 2016

When it opens you will see a blank worksheet (sheet1), which consists of alphabetically titled columns and numbered rows. Every sheet of Microsoft Excel has 256 columns and 65,536 rows and each cell is referenced by its coordinates of columns and rows. For instance, **A1** is the cell located in column A and row 1; **B7** is the cell in column B and row 7. You can reference a range

of cells, for example **C1:C5** are cells in columns C and rows 1 to 5. You can also reference a matrix, **A10:C15**, are cells in columns A, B and C and rows 10 to 15.

Excel stores your work in a workbook, each workbook has one or more worksheets (and/or charts) which you can view by clicking on the sheet tab (lower left corner of the active (current) sheet).

You can type anything on a cell, in general you can enter text (or labels), numbers, formulas (starting with the “=” sign), and logical values (as in ‘true’ or ‘false’).

Click on a cell and start typing, once you finish typing press ‘enter’ (to move to the next cell below) or “tab” (to move to the next cell to the right). Alternatively, data can be imported from various source and format. Commonly used file formats into Excel are .csv, .txt, or database files like Access or SQL. It is important to understanding how to properly bring external data into Microsoft Excel is crucial for ensuring the integrity and accuracy of the data analysis. Likewise, navigating through a excel worksheet the following shortcuts (key) can be utilised using keyboard:

- (i.) **“Home”**: Moves to the first column in the current row
- (ii.) **“End + Right Arrow”**: Moves to the last filled cell in the current row
- (iii.) **“End + Down Arrow”**: Moves to the last filled cell in the current column
- (iv.) **“Ctrl + Home”**: Moves to cell A1
- (v.) **“Ctrl + End”**: Moves to the last cell in your document (not the last cell of the current sheet)
- (vi.) **“Ctrl + Shift + End”**: Selects everything between the active cell to the last cell in the document

Typically, the first things to do with new data is to get to know it by asking some general questions like but not limited to the following:

- What variables are included? What information are we getting?
- What is the format of the variables: string, numeric, etc.?
- What type of variables: categorical, continuous, and discrete?
- Is this sample or population data?

After looking at the data one may want to know some further details or can start answering some of these questions by looking directly at the table, for some other questions you may have to do

some calculations by obtaining a set of descriptive statistics. These statistics are a collection of measurements of basically two things: location and variability. Location tells you the central value (the mean is the most common measure, other are mode, median) of your variables. Variability refers to the spread of the data from the centre value (i.e. variance, standard deviation and range).

Microsoft Excel has built-in formulas, functions, commands and toolpak that help with organising and manipulating data directly in the cells: these tools play a crucial role in data handling within worksheet, making it easier to organise, analyse, and manipulate datasets according to specific needs. Tables 1 to 3 present some commonly used Text functions, formulas and commands. Other are Lookup and Reference functions, Logical function, Aggregation functions, and others.

**Table 1: Text Functions for Data Handling & Pre-Processing**

S/N	Function	Description	Example
1	LEFT	Extracts a specified number of characters from the beginning of a text string.	=LEFT(A1, 5)
2	RIGHT	Extracts characters from the end of a text string.	=RIGHT(A1, 3)
3	MID	Extracts a substring from the middle of a text string.	=MID(A1, 3, 4)
4	TRIM	Removes extra spaces from a text string.	=TRIM(A1)
5	UPPER	Converts text to uppercase.	=UPPER(A1)
6	LOWER	Converts text to lowercase.	=LOWER(A1)
7	PROPER	Capitalizes the first letter of each word.	=PROPER(A1)
8	TEXTJOIN	Joins text from multiple cells using a specified delimiter.	=TEXTJOIN(";", TRUE, A1:A5)

Some formulate for data handling and pre-processing in Microsoft Excel are presented in Table 2.

**Table 2: Formulas for Data Handling & Pre-Processing**

S/N	Formulas	Description	Example
1	CONCATENATE	Combines the contents of two or more cells into one	=CONCATENATE(A1, " ", B1)
2	IFERROR	Returns a specified value if a formula results in an error.	=IFERROR(VLOOKUP(A2, B2:B10, 1, FALSE), "Not Found")
3	INDEX and MATCH	An alternative to VLOOKUP for more flexible data referencing.	=INDEX(B2:B10, MATCH(A2, A1:A10, 0))
4	SUMPRODUCT	Multiplies corresponding values in ranges or arrays and returns the sum of the products.	=SUMPRODUCT(A1:A10, B1:B10)
5	ARRAY	Allows for complex calculations on multiple values at once.	{=SUM(A1:A10 * B1:B10)} (entered using Ctrl + Shift + Enter)

In addition to functions and formulas, Excel has command features for various tasks. These are features you can access via Excel's ribbon, context menus, or keyboard shortcuts for organising and manipulating data, some of these commands are presented in Table3.

**Table 3: Commands for Data Handling & Pre-Processing**

S/N	Command	Description	How to Use
1	SORT	Organizes data in ascending or descending order based on one or more columns	Go to the "Data" tab and select "Sort."
2	FILTER	Displays only the rows that meet specific criteria.	Go to the "Data" tab and click "Filter."
3	REMOVE DUPLICATES	Eliminates duplicate entries in a dataset, ensuring each record is unique.	Go to the "Data" tab and click "Remove Duplicates."
4	TEXT TO COLUMNS	Splits data from one column into multiple columns based on delimiters (such as commas or spaces).	Go to the "Data" tab and select "Text to Columns."

S/N	Command	Description	How to Use
5	FLASH FILL	Automatically fills in data based on patterns detected from your initial entries.	Start typing a pattern and press Ctrl + E to trigger Flash Fill.
6	GROUP AND OUTLINE	Organizes data into groups that can be expanded or collapsed.	Select rows/columns, right-click, and choose "Group."
7	REMOVE BLANK CELLS	Select and remove blank cells in a range.	Use the "Go To Special" command (F5 or Ctrl + G), select "Blanks," and delete or replace them.
8	CONSOLIDATE	Combines data from multiple ranges into one summary report.	Go to the "Data" tab and select "Consolidate."
9	DATA VALIDATION	Restricts the type of data that can be entered into a cell.	Go to the "Data" tab and click on "Data Validation."
10	PIVOTTABLES	Summarizes and analyses large datasets by dynamically grouping and aggregating data.	Go to the "Insert" tab and click on "PivotTable."

Furthermore, Microsoft excel has a special tool for data analysis called “Analysis ToolPak” (Carr, 2008). This can be accessed under data menu. Meanwhile, if you do not see “data analysis” option you need to install it, go to Tools – Add-Ins, a window will pop-up and check the “Analysis ToolPak” option, then press OK. However, a simplify steps to Activate Data Analysis Toolpak in Microsoft Excel 2016 is as follows:

- (i.) Open Excel.
- (ii.) Go to File > Options.
- (iii.) In the Excel Options dialog box, select Add-ins.
- (iv.) At the bottom, next to Manage, select Excel Add-ins from the dropdown, then click Go.
- (v.) In the Add-Ins box, check Analysis ToolPak.
- (vi.) Click OK.

Graphical illustrations of the steps are shown in Figures 2 to 5.

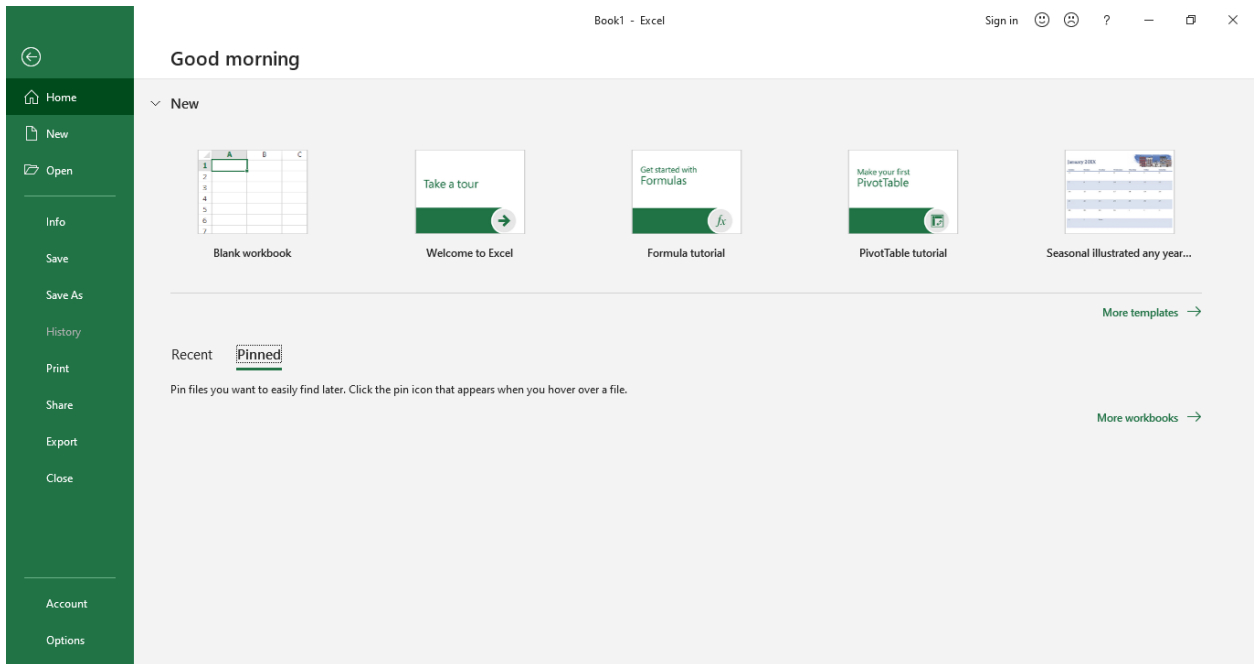


Figure 3: Options from file menu

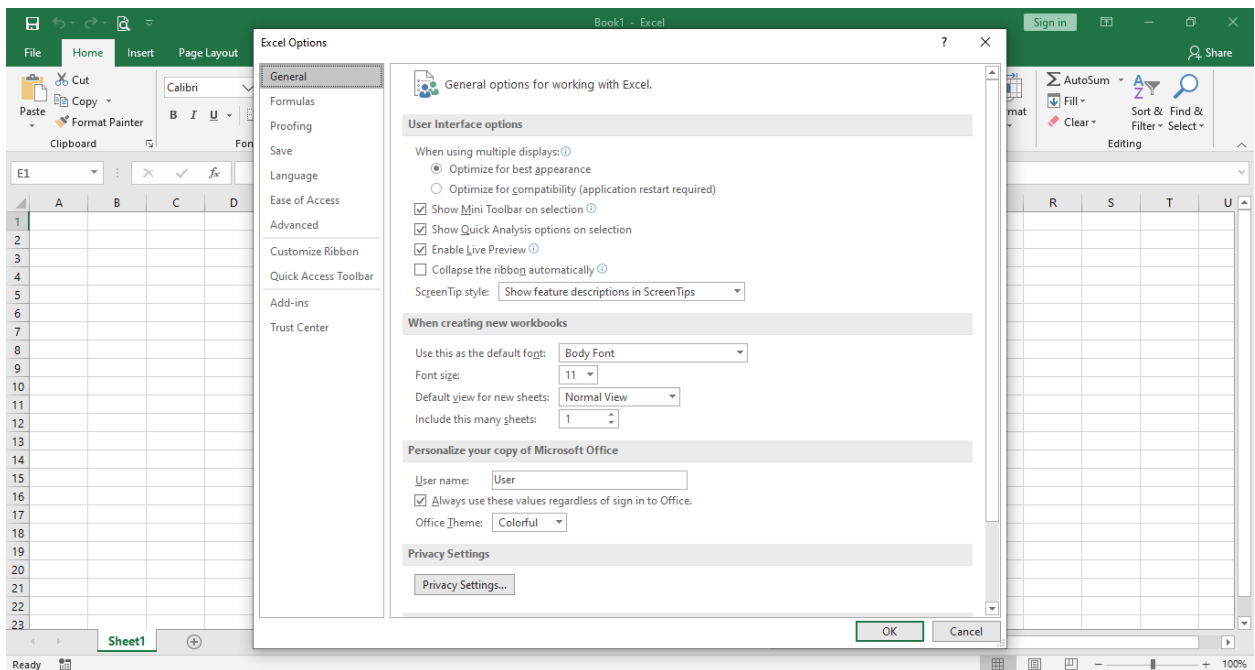


Figure 3: Excel options interface to access Add-ins

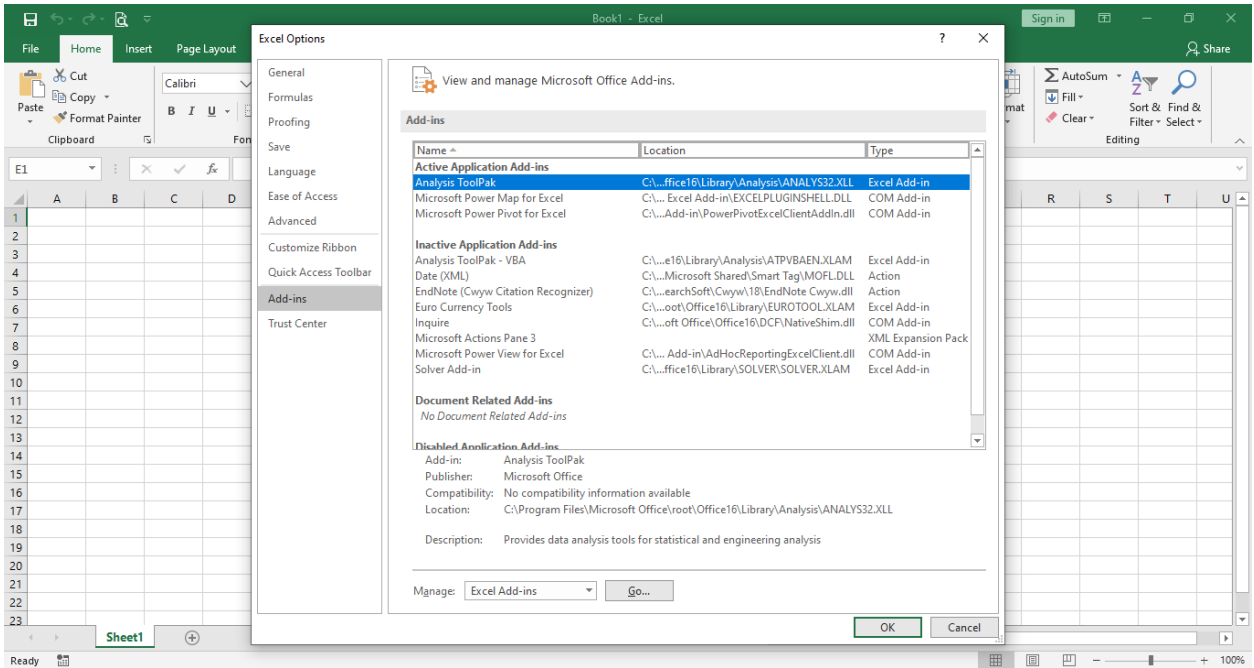


Figure 4: Add-ins interface to access Analysis Toolpak

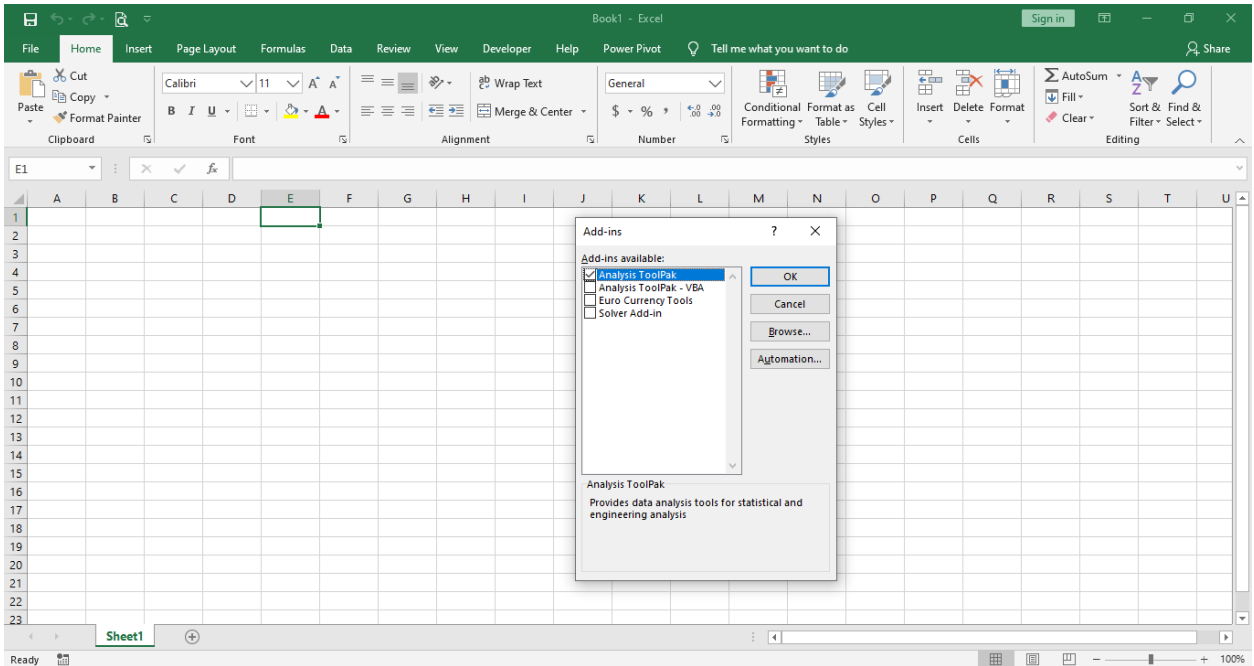


Figure 5: Add-Ins dialog box to select Analysis ToolPak

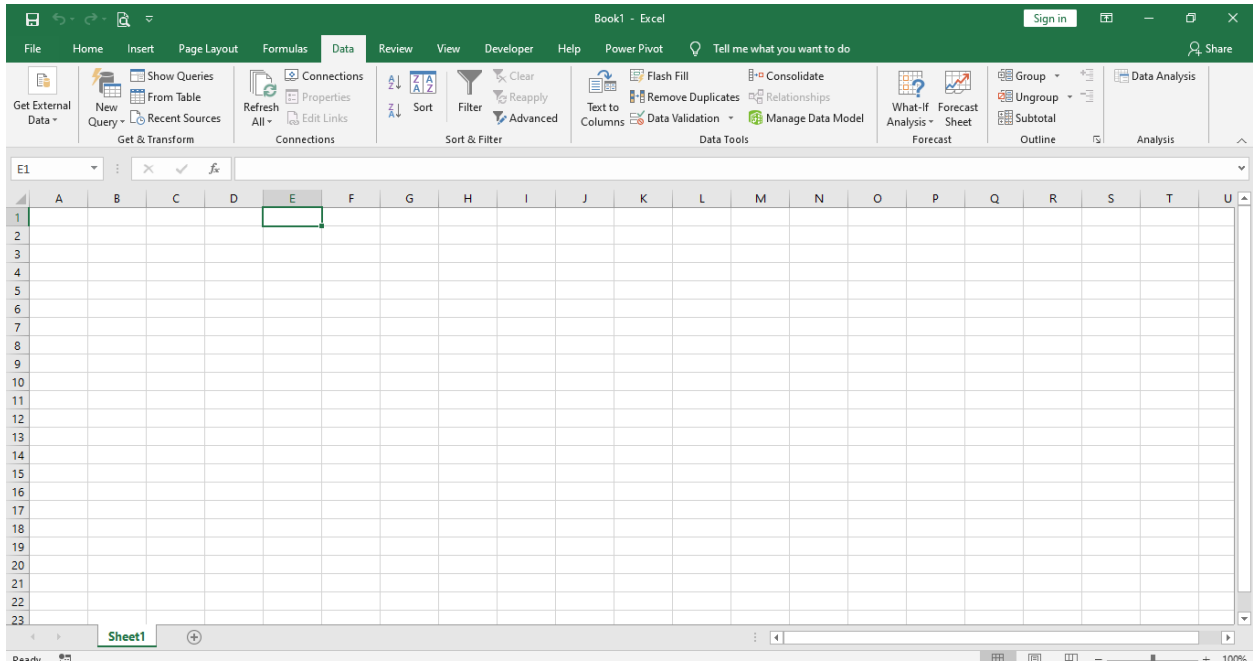


Figure 5: Data Analysis under Analysis

With above setup Microsoft Excel is ready for statistical analysis, it is basically the study of what causes such variability. It covers the following:

- (i.) Measures of Central Tendency: Mean, mode and median of a given dataset.
- (ii.) Measures of Dispersion: Range, variance, and standard deviation.
- (iii.) Measures of Distribution Shape: Skewness, and kurtosis.

Descriptive data analysis is the foundational step in understanding the characteristics and structure of a dataset. It involves summarising and organising data using statistical measures. In addition, tables, charts, and graphical representations and visualisation, this analysis provides an overview of the data's central tendencies and dispersion as well as patterns. The interpretation of these descriptive statistics is crucial, as it offers insights into trends, anomalies, and relationships within the data, guiding further analysis and decision-making processes.

#### 4. Case Study

As a pilot project, a private company involved in hardware and maintenance related activity planned to launch its own brand of assembled computers. The company identified a small colony

having 12 families in all. Each of the family is having members of different age group. For the sake of their brand performance related study, they decided to issue one computer system to each family as a single unit. They got the system installed with several applications suitable for each age group. The application installed are: Computer Games, Movies, Application Software, E-Books, Internet, and Security related software.

The company wants to study the *application usage pattern and the satisfaction level of the families, so that they can customize their product accordingly*. In order to monitor the system usage time, they installed an automatic clock with the system, which records the usage time. This will help them to analyse the questionnaire feedback, which they wish to get filled, at the end of their study. Raw data collected is shown in Table A1(see Appendix). Response of the questionnaire is coded and shown in Table 4

**Table 4: Data coded for analysis**

Family	System Utilization Time[hours]	System Utilization Activity						Product Satisfaction	Product Purchase
		1	2	3	4	5	6		
1	30	1	1	1	0	0	0	0	0
2	30	0	0	0	1	0	1	1	1
3	60	1	1	0	0	0	0	2	1
4	45	0	0	0	0	1	0	-1	0
5	30	0	0	0	0	0	1	1	0
6	60	0	1	0	0	0	0	2	1
7	30	0	0	0	1	0	0	1	0
8	45	0	0	1	1	0	0	-1	0
9	15	0	0	0	0	0	1	1	1
10	60	0	1	0	0	0	0	2	1
11	180	1	1	1	1	0	0	2	1
12	120	1	1	0	1	0	0	2	1

The output from Microsoft Excel of analysing dataset in Table 4 is shown in Table 5.

**Table 5: Results of the descriptive data analysis**

Statistics	System Utilization Time[hours]	System Utilization Activity						Product Satisfaction	Product Purchase
		1	2	3	4	5	6		
Mean	58.75	0.33	0.50	0.25	0.42	0.08	0.25	1.00	0.58
Standard Error	13.50	0.14	0.15	0.13	0.15	0.08	0.13	0.33	0.15
Median	45.00	0.00	0.50	0.00	0.00	0.00	0.00	1.00	1.00
Mode	30.00	0.00	1.00	0.00	0.00	0.00	0.00	2.00	1.00
Standard Deviation	46.76	0.49	0.52	0.45	0.51	0.29	0.45	1.13	0.51
Sample Variance	2186.93	0.24	0.27	0.20	0.27	0.08	0.20	1.27	0.27
Kurtosis	3.76	-1.65	-2.44	-0.33	-2.26	12.00	-0.33	-0.34	-2.26
Skewness	1.95	0.81	0.00	1.33	0.39	3.46	1.33	-0.91	-0.39
Range	165.00	1.00	1.00	1.00	1.00	1.00	1.00	3.00	1.00
Minimum	15.00	0.00	0.00	0.00	0.00	0.00	0.00	-1.00	0.00
Maximum	180.00	1.00	1.00	1.00	1.00	1.00	1.00	2.00	1.00
Sum	705.00	4.00	6.00	3.00	5.00	1.00	3.00	12.00	7.00
Count	12	12	12	12	12	12	12	12	12

## 5. Summary and Conclusion

A practical case study is provided to demonstrate descriptive data analysis using Microsoft Excel 2016 software. The study utilized Microsoft Excel's built-in functions and the Data Analysis Toolpak to evaluate descriptive data analysis of a dataset. The following conclusions are drawn:

- (i.) the concept and importance of descriptive statistics were explored.
- (ii.) the use of Microsoft Excel for data analysis has been simplified to participants.
- (iii.) Microsoft Excel Toolpak is utilised for analysis of case studies.

## References

- Carr, N. T. (2008). Using Microsoft Excel® to calculate descriptive statistics and create graphs. *Language Assessment Quarterly*, 5(1), 43-62.
- Provost, F., & Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. " O'Reilly Media, Inc."
- Triola, M. F., & Iossi, L. (2018). *Elementary statistics*. New York: Pearson.

## APPENDIX A QUESTIONNAIRE

1. Family No.: .....
2. What is the Average System Utilization time(minutes)?  
**15 30 45 60 75 90 105 120 135 150 165 180**
3. What are the Activities Performed by your family while utilizing the system?  
a) Computer Games, b) Internet, c) E-Books, d) Application Software's, e) Movies f)  
Security related software
4. What is the level of satisfaction, achieved by your family?  
-2 = Very dissatisfied, 1 = dissatisfied, 0 = indifference, 1 = satisfied, 2 = Very satisfied.
5. Would your family like to purchase the product?  
a) Yes b) No

Responses to the questions framed in the questionnaire, collected data is presented in Table A1:

Table A1: Field Response

<b>Famil y</b>	<b>System Utilization Time[hours]</b>	<b>System Utilization Activity</b>	<b>Product Satisfaction</b>	<b>Product Purchase</b>
1	30	1,2,3	0	N
2	30	4,6	1	Y
3	60	1,2	2	Y
4	45	5	-1	N
5	30	6	1	N
6	60	2	2	Y
7	30	4	1	N
8	45	3,4	-1	N
9	15	6	1	Y
10	60	2	2	Y
11	180	1,2,3,4	2	Y
12	120	1,2,4	2	Y