

### UNIT 3 : MEASURE OF LOCATION AND DISPERSION

#### 1. What is Central Tendency?

The central tendency measure is defined as **the number used to represent the center or middle of a set of data values.**

#### 2. Mention three main measures of Central tendencies.

The 3 most common measures of central tendency are **the mode, median, and mean**

#### 3. Define Mean. Write it's Excel formula.

Mean is the **average value** of the given numbers and is calculated by dividing the sum of given numbers by the total number of numbers.

To find the mean in Excel, we use the formula “=AVERAGE(array of numbers)”

#### 4. Define Median. Write it's Excel formula.

Median can be defined as the **middle value** of a group of numbers.

To find median in Excel, we use the formula “=MEDIAN(number1, [number2], ...)”

#### 5. Define Mode. Write it's Excel formula.

A mode is defined as the value that has a **higher frequency** in a given set of values. It is the value that **appears the most number of times.**

To find mode in Excel , we use the formula “=MODE(number1, [number2], ...)”

#### 6. Define statistical dispersion.

In statistics, the extent to which the numerical data are distributed or squeezed about an average value is called statistical dispersion. In short, it is the distribution of data.

#### 7. Name the different types of measures of Statistical Dispersion.

There are two types of measures of statistical dispersion.

1. Absolute measures of dispersion.
2. Relative measures of dispersion.

#### 8. Mention the different Absolute measures of Dispersion for the given data.

The different types of Absolute measures of dispersion are

- i) Range
- ii) Variance
- iii) Quartile deviation
- iv) Mean Deviation
- v) Standard Deviation

#### 9. Mention the different Relative measures of Dispersion for the given data

The different types of Relative measures of dispersion are

- i) Coefficient of Range
- ii) Coefficient of variation
- iii) Coefficient of Mean deviation
- iv) Coefficient of Quartile deviation

#### 10. Define the following with Excel formula.

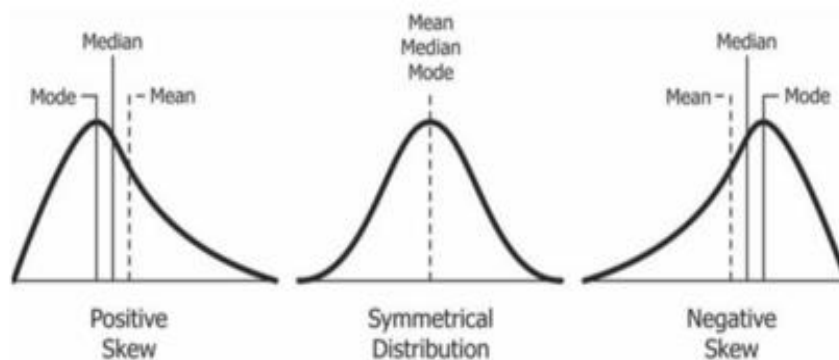
- a) Range    b) Quartile    c) Quartile deviation.    d) Variance    e) Standard Deviation.

- a) Range : It is the difference between the maximum value and the minimum value given in a data set.  
Range = Maximum value – minimum value.
- b) Quartile : Quartile refers to the values which divide the set into quarters.
- c) Quartile deviation : It is half of the difference between First and Third quartile  $((Q_3 - Q_1)/2)$ .
- d) Variance: It gives us the info about how far the data is spread from its mean.
- e) Standard deviation : In statistics, the standard deviation is a measure of the amount of variation or dispersion of a set of values.

A low standard deviation indicates that the values tend to be close to the mean (also called the expected value) of the set, while a high standard deviation indicates that the values are spread out over a wider range.

**11. Explain skewness and kurtosis graphs in MS Excel.**

skewness is a measure of symmetry. If the skewness  $S$  is zero then the distribution represented by  $S$  is perfectly symmetric. If the skewness is negative, then the distribution is skewed to the left, while if the skew is positive then the distribution is skewed to the right



Kurtosis in statistics describes the distribution of the data set. It shows to what extent the data set points of a particular distribution differ from the data of a normal distribution. It is also used to determine whether a distribution contains extreme values.

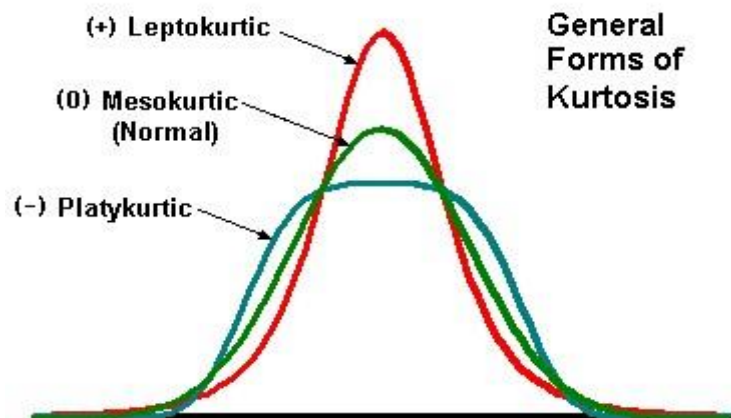
There are 3 types of Kurtosis.

1. Mesokurtic
2. Leptokurtic
3. Platykurtic

If the Kurtosis of data falls close to zero or equals zero, it is referred to as Mesokurtic. It means that the data set follows a normal distribution

If the Kurtosis of data falls more than zero (kurtosis + ve), it is referred to as leptokurtic. Leptokurtic has steep curves on both sides, indicating the large population of outliers in the data set.

If the Kurtosis of data falls less than zero (kurtosis - ve), it is referred to as Platykurtic. Platykurtic has pale or flat curve, and that curve indicates the small number of outliers in a distribution.



**UNIT 4 : INTRODUCTION TO PYTHON PROGRAMMING****1. What is Programming ?**

Programming refers to a technological process for telling a computer which tasks to perform in order to solve problems.

**2. What is Python ?**

Python is a high-level, general-purpose programming language.

**3. Who invented Python Programme?**

Python was created by Guido van Rossum, and first released on February 20, 1991.

**4. Mention the advantages of learning Python.**

Advantages of Python are

- 1. Easy to Read, Learn and Write:** Python is a **high-level programming language** that has English-like syntax. This makes it easier to read and understand the code.
- 2. Improved Productivity :** Python is a very **productive language**. Due to the simplicity of Python, developers can focus on solving the problem. They don't need to spend too much time in understanding the **syntax** or **behavior** of the programming language. You write less code and get more things done.
- 3. Interpreted Language :** Python is an interpreted language which means that Python directly **executes the code** line by line. In case of any error, it stops further execution and reports back the error which has occurred. Python shows only one error even if the program has multiple errors. This makes **debugging** easier.
- 4. Dynamically Typed :** Python doesn't know the type of variable until we run the code. It automatically assigns the data type during **execution**. The programmer doesn't need to worry about declaring variables and their data types.
- 5. Free and Open-Source :** Python comes under the **OSI approved** open-source license. This makes it **free to use** and **distribute**. You can download the source code, modify it and even distribute your version of Python. This is useful for organizations that want to modify some specific behavior and use their version for development.
- 6. Vast Libraries Support :** The standard library of Python is huge, you can find almost all the functions needed for your task. So, you don't have to depend on external libraries. But even if you do, a **Python package manager (pip)** makes things easier to import other great packages from the **Python package index (PyPi)**. It consists of over 200,000 packages.
- 7. Portability :** In many languages like C/C++, you need to change your **code** to run the program on different platforms. That is not the same with Python. You only write once and run it anywhere.

**5. Mention the steps of Python installation.**

Step 1 : Open a web browser and go to <https://www.python.org/downloads/>.

Step 2 : Download the latest version of python for windows 10 (Ex :3.10.5(64 bit))

Step 3 : Run the .exe file that you just downloaded.

Step 4 : Install Python 3.10.5(64 bit) by clicking on Install now

**6. What are Python Libraries?**

Normally, a library is a collection of books or is a room or place where many books are stored to be used later. Similarly a Python library is simply a collection of codes or modules of codes that we can use in a program for specific operations. We use libraries so that we don't need to write the code again in our program that is already available.

### 7. Mention the packages/Libraries for Python.

The commonly used Python libraries are

1. **TensorFlow:** This library was developed by Google in collaboration with the Brain Team. It is an open-source library used for high-level computations. It is also used in machine learning and deep learning algorithms. It contains a large number of tensor operations. Researchers also use this Python library to solve complex computations in Mathematics and Physics.
2. **Matplotlib:** This library is responsible for plotting numerical data. And that's why it is used in data analysis. It is also an open-source library and plots high-defined figures like pie charts, histograms, scatterplots, graphs, etc.
3. **Pandas:** Pandas are an important library for data scientists. It is an open-source machine learning library that provides flexible high-level data structures and a variety of analysis tools. It eases data analysis, data manipulation, and cleaning of data. Pandas support operations like Sorting, Re-indexing, Iteration, Concatenation, Conversion of data, Visualizations, Aggregations, etc.
4. **NumPy:** The name "Numpy" stands for "Numerical Python". It is the commonly used library. It is a popular machine learning library that supports large matrices and multi-dimensional data. It consists of in-built mathematical functions for easy computations. Even libraries like TensorFlow use Numpy internally to perform several operations on tensors. Array Interface is one of the key features of this library.
5. **SciPy:** The name "SciPy" stands for "Scientific Python". It is an open-source library used for high-level scientific computations. This library is built over an extension of Numpy. It works with Numpy to handle complex computations. While Numpy allows sorting and indexing of array data, the numerical data code is stored in SciPy. It is also widely used by application developers and engineers.
6. **Scrapy:** It is an open-source library that is used for extracting data from websites. It provides very fast web crawling and high-level screen scraping. It can also be used for data mining and automated testing of data.
7. **Scikit-learn:** It is a famous Python library to work with complex data. Scikit-learn is an open-source library that supports machine learning. It supports various supervised and unsupervised algorithms like linear regression, classification, clustering, etc. This library works in association with Numpy and SciPy.
8. **PyGame:** This library provides an easy interface to the Standard Directmedia Library (SDL) platform-independent graphics, audio, and input libraries. It is used for developing video games using computer graphics and audio libraries along with Python programming language.
9. **PyTorch:** PyTorch is the largest machine learning library that optimizes tensor computations. It has rich APIs to perform tensor computations with strong GPU acceleration. It also helps to solve application issues related to neural networks.
10. **PyBrain:** The name "PyBrain" stands for Python Based Reinforcement Learning, Artificial Intelligence, and Neural Networks library. It is an open-source library built for beginners in the field of Machine Learning. It provides fast and easy-to-use algorithms for machine learning tasks. It is so flexible and easily understandable and that's why is really helpful for developers that are new in research fields.

### 8. What is Anaconda in python ?

Anaconda Python is a free, open-source platform that helps to write and execute code in the programming language Python. It is popular for data analysis and scientific computing. It comes with many preinstalled libraries which include NumPy, SciPy, Matplotlib, Pandas, Cython, Spyder and Jupyter.

### 9. How do you execute Python Syntax?

Python syntax can be executed by writing directly in the Command Line:

```
>>> print("Hello, World!")  
Hello, World!
```

**10. How do you start Python in Windows?**

- Step 1 :- Start -> "All Programs"-> Python3.10-> IDLE Python (3.10 64 bit)  
 Step 2:- The IDLE shell 3.10.5 window will open  
 Step 3 :- Click "File"-> New File-> A new untitled window will open.  
 Step 4 :- Type the program code in editor window  
 Step 5 :- Click "File"-> save -> give the File name -> save. Then Click "Run"-> "Run module".  
 Step 6 :- The result get displayed in IDLE shell 3.10.5 window .

**11. What is Indentation in Python ?**

Leading white space (spaces and tabs) at the beginning of a statement is called Indentation. Python uses Indentation to indicate a block of code.

Ex:

```
if a==1:( BLOCK 1)
    print(a)( BLOCK 2)
        if b==2:( BLOCK 3)
            print(b)( BLOCK 2 (continuation))
print("end")( BLOCK 1 (continuation))
```

**12. Mention the different quotation used in Python? Give Examples.**

Python accepts single quotes( ' ), double quotes( " ) and triple ( ''' or """) to represent a string in Python.

- Single quotes are used for single word  
**Ex : GPT= 'Gauribidanur '**
- Double quotes are used for sentences  
**Ex : GPT = " Government Polytechnic Gauribidanur"**
- Triple quotes are used to represent a multi-line string.  
**Ex : GPT = """Government Polytechnic Gauribidanur is a technical institute under collegiate and technical education """**

**13. How do the comments in Python starts?**

Comments in Python begin with a hash mark (#) and whitespace character and continue to the end of the line. Python ignores everything after the hash mark and up to the end of the line. It can be inserted anywhere in your code, even in line with other code.

```
Ex : 1. # This is a comment
     2. print("This will run.") # This won't run
```

**14. What is a variable in Python?**

A Python variable is a reserved memory location to store values. In other words, a variable in a python program gives data to the computer for processing.

**15. Mention the rules for Python variables.**

1. A Variable name should start with letter (a-z , A-Z) or underscore (\_).  
EX: age , \_age , Age
2. In variable name, no special characters allowed other than underscore (\_).
3. Variables are case sensitive. (age, Age, AGE are three different variables.)
4. A Variable name can not start with a number.
5. Variable name should not be a Python keyword.  
Ex : pass, break, continue.. etc are reserved for special meaning in Python. So, we should not declare keyword as a variable name.

### 16. What are output variables in Python?

The Python print statement is often used to output variables. To combine both text and a variable, Python uses the + character.

```
Ex : 1) x = "awesome"  
      print("Python is", x)
```

*output : Python is awesome*

### 17. Mention the different data types of Python.

Different data types in Python are Numbers, List, Tuple, Strings, Dictionary, etc.

### 18. Explain the following

- a) Numbers in Python    b) String in Python    c) Python Lists    d) Python Dictionary

#### a) **Numbers in Python:**

There are three numerical types in Python.

- i) **Int** :- Int, or integer, is a whole number, positive or negative, without decimals, of unlimited length.  
Ex: x = 35656222554887711
- ii) **Float** :- Float, or "floating point number" is a number, positive or negative, containing one or more decimals. Ex : x = 1.10
- iii) **Complex** :- Complex numbers are written with a+ib form where a is real part and b is imaginary part .

**b) String in Python :** String is a collection of alphabets, words or other characters. It is one of the primitive data structures and are the building blocks for data manipulation. Python has a built-in string class named **str** . Strings in python are surrounded by either single quotation marks, or double quotation marks.

**c) Python Lists :** A list in Python is used to store the sequence of various types of data. A list can be defined as a collection of values or items of different types. The items in the list are separated with the comma (,) and enclosed with the square brackets [].

**d) Python Dictionary:** A dictionary is a kind of data structure that stores items in key-value pairs. A key is a unique identifier for an item, and a value is the data associated with that key. Dictionaries often store information such as words and definitions, but they can be used for much more. Dictionaries are mutable in Python, which means they can be changed after they are created. They are also unordered, indicating the items in a dictionary are not stored in any particular order.

### 19. Mention the different types of operators in Python.

The different types of operators in python are

1. Arithmetic operators
2. Comparison operators
3. Logical ( or Relational) operators
4. Assignment operators
5. Conditional ( or ternary ) operators

