

New
syllabus
2021-22



Chapter 10
**Data
Handling**

Informatics Practices
Class XI (As per CBSE Board)

Visit : python.mykvs.in for regular updates

Data Handling



Data is a collection of facts, such as numbers, words, measurements, observations or just descriptions of things.

Data vs Information

Data are raw numbers or other findings which, by themselves, are of limited value.

Information is data that has been converted into a meaningful and useful context.

Computers are being used extensively nowadays in everyday life/every field In the form of laptop, desktop, smartphone, gadgets etc to handle such data and convert these in information.

E.g. -1. A day's temperature, humidity, wind and speed of recorded are Data While percentage of weather as cold or warm is an Information. **2.** Students' names in a class are Data While names of students in alphabetic order are Information.



Purpose/importance of Data

- To make informed decision
- To find solution to the problem
- To stop the guessing game
- To get the result we want
- To improve people's lives
- To be strategic in our approaches
- To know what we are doing



Types of data

Structured data is the data which conforms to a well define structure or a data model and can be easily accessed and used by a person or a computer program.

unstructured data usually refers to information that doesn't reside in a traditional row-column database. Unstructured data files often include text and multimedia content.



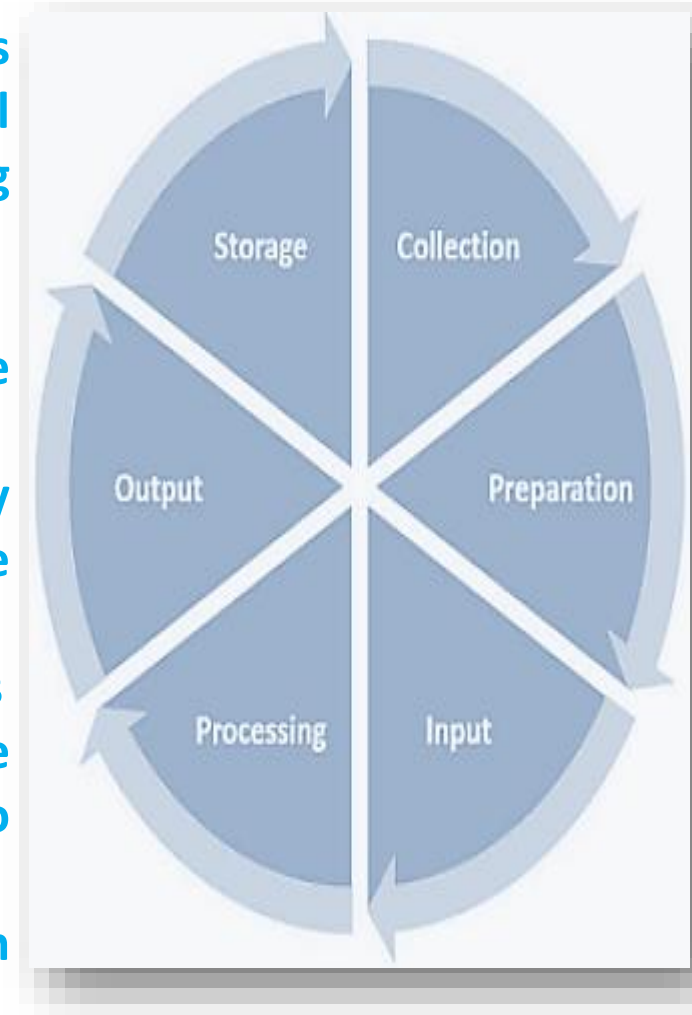
Difference between structured & unstructured data

PROPERTIES	STRUCTURED DATA	UNSTRUCTURED DATA
Technology	Based on Relational database table	Based on character and binary data
Transaction management	Matured transaction	No transaction management
Version management	Versioning over tuples,row,tables	Versioned as a whole
Flexibility	It is sehema(table structure) dependent and less flexible	it very flexible and there is absence of schema(table structure)
Scalability	It is very difficult to scale DB schema	It is very scalable
Robustness	Very robust	—
Query performance	Structured query allow complex joining	Only textual query are possible
Example	Excel files or SQL databases.	emails, videos, audio files, web pages, and social media messages

Data Processing Cycle

The data processing cycle is the set of operations used to transform data into useful information. This cycle involves the following steps:

- Collection of data
- Preparation of the data into a format suitable for data entry, as well as error checking
- Entry of the data into the system, which may involve manual data entry, scanning, machine encoding, and so forth
- Processing of the data with computer programs
- Transmitting the resulting information to the user, typically via screen or printed report, so that it can be acted upon
- Storing the input data and output information for future use





Statistical method for data handling

Mean- Mean is an average of all the numbers. The steps required to calculate a mean are:

- sum up all the values of a target variable in the dataset
- divide the sum by the number of values

For **example**, take this list of numbers: 10, 10, 20, 40, 70. The mean (average) is found by adding all of the numbers together and dividing by the number of items in the set: $(10 + 10 + 20 + 40 + 70) / 5 = 30$.



Statistical method for data handling

Median - Median is the middle value of a sorted list of numbers.

The steps required to get a median from a list of numbers are:

- sort the numbers from smallest to highest
- if the list has an odd number of values, the value in the middle position is the median
- if the list has an even number of values, the average of the two values in the middle will be the median

For **example**, take this list of numbers: 10, 10, 20, 40, 70.

The median is found by ordering the set from lowest to highest and finding the exact middle. The median is just the middle number: 20.



Statistical method for data handling

Mode - To find the mode, or modal value, it is best to put the numbers in order. Then count how many of each number. A number that appears most often is the mode. e.g. {19, 8, 29, 35, 19, 28, 15}. Arrange them in order: {8, 15, 19, 19, 28, 29, 35}. 19 appears twice, all the rest appear only once, so 19 is the mode.

Having two modes is called "bimodal". Having more than two modes is called "multimodal".



Standard deviation means measure the amount of variation / dispersion of a set of values. A low standard deviation means the values tend to be close to the mean in a set and a high standard deviation means the values are spread out over a wider range. Standard deviation is the most important concepts as far as finance is concerned. Finance and banking is all about measuring the risk and standard deviation measures risk. Standard deviation is used by all portfolio managers to measure and track risk.

Steps to calculate the standard deviation:

1. Work out the Mean (the simple average of the numbers)
2. Then for each number: subtract the Mean and square the result
3. Then work out the mean of those squared differences.
4. Take the square root of that and we are done!

Data Handling



E.g. Std deviation for (9, 2, 12, 4, 5, 7)

Step 1. Work out the mean $-(9+2+12+4+5+7) / 6 = 39/6 = 6.5$

Step 2. Then for each number: subtract the Mean and square the result $-(9 - 6.5)^2 = (2.5)^2 = 6.25$, $(2 - 6.5)^2 = (-4.5)^2 = 20.25$

Perform same operation for all remaining numbers.

Step 3. Then work out the mean of those squared differences.

Sum = $6.25 + 20.25 + 2.25 + 6.25 + 30.25 + 0.25 = 65.5$

Divide by N-1: $(1/5) \times 65.5 = 13.1$ (This value is "**Sample Variance**")

Step 4. Take the square root of that: $s = \sqrt{13.1} = 3.619...$ (stddev)

formula for Standard Deviation

Above e.g. is for practice

purpose otherwise stddev is performed for large amount of data

$$s = \sqrt{\frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2}$$