Unit I

FY KKSU - Statistics

Definition

- **A branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data.**

What is Statistics

- Statistics is a set of decision-making techniques which helps businessmen in making suitable policies from the available data. In fact, every businessman needs a sound background of statistics as well as of mathematics.
- The purpose of statistics and mathematics is to manipulate, summarize and investigate data so that the useful decision-making results can be executed.

Uses of Statistics in Business

- With the help of statistical methods, quantitative information about production, sale, purchase, finance, etc. can be obtained. This type of information helps businessmen in formulating suitable policies.

- By using the techniques of time series analysis which are based on statistical methods, the businessman can predict the effect of a large number of variables with a fair degree of accuracy.

- In business decision theory, most of the statistics techniques are used in taking a business decision which helps us in doing the business without uncertainty.

- Nowadays, a large part of modern business is being organised around systems of statistical analysis and control.

Continue

- Statistics is the basis of economics. The consumer's maximum satisfaction can be determined on the basis of data pertaining to income and expenditure. The various laws of demand depend on the data concerning price and quantity. The price of a commodity is well determined on the basis of data relating to its buyers, sellers, etc.

Importance of Statistics

- The importance of statistics in the following major areas:

1. Importance of Statistics in Business and Industry
2. Importance in the Field of Science and Research
3. Importance in the Field of Banking
4. Importance to the State
5. Importance in planning

Limitations of Statistics

(1) Statistics laws are true on average. Statistics are aggregates of facts, so a single observation is not a statistic. Statistics deal with groups and aggregates only.

(2) Statistical methods are best applicable to quantitative data.

(3) Statistics cannot be applied to heterogeneous data.

(4) If sufficient care is not exercised in collecting, analyzing and interpreting the data, statistical results might be misleading.

(5) Only a person who has an expert knowledge of statistics can handle statistical data efficiently.

(6) Some errors are possible in statistical decisions. In particular, inferential statistics involves certain errors. We do not know whether an error has been committed or not.

Primary & Secondary Data Definitions

- **Primary Data:** Data that has been generated by the researcher himself/herself, surveys, interviews, experiments, specially designed for understanding and solving the research problem at hand.
- **Secondary Data:** Using existing data generated by large government Institutions, healthcare facilities etc. as part of organizational record keeping. The data is then extracted from more varied datafiles.

BASIS FOR COMPARISON	PRIMARY DATA	SECONDARY DATA
Meaning	Primary data refers to the first hand data gathered by the researcher himself.	Secondary data means data collected by someone else earlier.
Data	Real time data	Past data
Process	Very involved	Quick and easy
Source	Surveys, observations, experiments, questionnaire, personal interview, etc.	Government publications, websites, books, journal articles, internal records etc.
Cost effectiveness	Expensive	Economical
Collection time	Long	Short
Specific	Always specific to the researcher's needs.	May or may not be specific to the researcher's need.
Available in	Crude form	Refined form
Accuracy and Reliability	More	Relatively less

What are the Sources of Data?

- **Sources of Data**

The sources of data can be classified into two types: statistical and non-statistical. Statistical sources refer to data that is gathered for some official purposes, incorporate censuses, and officially administered surveys. Non-statistical sources refer to the collection of data for other administrative purposes or for the private sector.

What are the different sources of data?

- **Internal sources**

a) When data is collected from reports and records of the organisation itself, they are known as the internal sources.

b) For example, a company publishes its annual report' on profit and loss, total sales, loans, wages, etc.

- **External sources**

a) When data is collected from sources outside the organisation, they are known as the external sources. For example, if a tour and travel company obtains information on Karnataka tourism from Karnataka Transport Corporation, it would be known as an external source of data.

Types of Data

- **A) Primary data**
- Primary data means first-hand information collected by an investigator.
- It is collected for the first time.
- It is original and more reliable.
- For example, the population census conducted by the government of India after every ten years is primary data.

Continue

- **B) Secondary data**
- Secondary data refers to second-hand information.
- It is not originally collected and rather obtained from already published or unpublished sources.
- For example, the address of a person taken from the telephone directory or the phone number of a company taken from Just Dial are secondary data.

Methods of Collecting Primary Data

- Direct personal investigation
- Indirect oral investigation
- Information through correspondents
- Telephonic interview
- Mailed questionnaire
- The questionnaire filled by enumerators

Sampling Method

- A sampling method is a process for choosing sample members from a population. Three (3) common sampling methods are:
- Simple random sampling
- Stratified sampling
- Cluster sampling

Census Method

- A census method is that process of the statistical list where all members of a population are analysed. The population relates to the set of all observations under concern. For instance, if you want to carry out a study to find out student's feedback about the amenities of your school, then all the students of your school would form a component of the 'population' for your study

Explain the census method of collection of data. Give its advantages and disadvantages

- A statistical investigation in which the data are collected for each and every element/unit of the population is termed as census method.
- It is also known as ‘complete enumeration’ or ‘100% enumeration’ or ‘complete survey’.
- It is useful when case intensive study is required or the area is limited.
- Examples:

1. Demographic data on birth and death rates, literacy, workforce, life expectancy, size and composition of a population
2. The census of India is conducted every 10 years

(B) The following are the advantages of the census method.	
(1) Intensive study	• It provides intensive and in-depth information covering many facets of the problems. • Example: In a population census, not only the number of persons is counted, but the information is also collected on various other parameters like the number of males and females, age, education, marital status, occupational level, income health conditions, etc.
(2) Results are more accurate and reliable	• Since, in this type of investigation, every item of the universe is taken into account, the conclusions are more accurate and reliable.
(B) The following are the disadvantages of the census method.	
(1) Costly method	• Since the data are obtained for or from each and every unit of the population, it is a very expensive method of investigation, especially in case of a large population size.
(2) Needs more time and manpower	• Since a large volume of data is to be collected, more time and manpower is required for its collection, analysis, and interpretation.
(3) Not suitable for the large population	• This method is meaningless in the case of an infinite universe where the number of items is unlimited.

Explain the sampling method of collection of data. Briefly discuss its advantages and disadvantages.

- The sampling method is the one in which only some of the representative items of the population are selected and the data are collected from these.
- Instead of collecting information for and from all the units of population, we select a sample, i.e., only a few items of the population.
- Conclusions derived from the small sample are generalised for the whole population.

Advantages

- It is more economical than the census method, as the task of collection and analysis of data is limited only to a small sample, i.e., a few units of the population.

- In this method, only a few units are analysed, hence we get quicker results.

- Due to the small size of the sample, it is possible to cross-check the data to test the reliability.

- Due to scientific nature, the sampling method can be used to cross-check the results of the census method.

Disadvantages

- The conclusions of the sampling method are based only on the results of a few items taken from the whole population. Thus, these may not be cent percent correct.

- It is a special technique and beyond the capacity of every person.
- Its use requires specialised knowledge and training.

Census Method	Sampling Method
The extensive enquiry is conducted at each and every unit of the population.	The limited enquiry is conducted as only a few units of the population are studied.
More Time, Money, and Labour It requires a large amount of money, time, and labour.	Less Time, Money, and Labour Relatively less money, time, and labour are required.
It is more suitable if the population is heterogeneous in nature.	It is more suitable if the population is homogeneous in nature.
The results are quite reliable and accurate under the census method.	The results of the sampling method are less reliable because a high degree of accuracy is not achieved.
It is very difficult to organise and supervise the census method.	The sampling method is comparatively easy to organise and supervise.
Under this method, the results of the investigation cannot be verified.	Under this method, the results can be tested by taking out another small sample.
The census method is an old method of investigation and not a very scientific method.	The sampling method is a new and practicable method. It is a scientific method.

Classification of Data and Tabular Presentation

- Classification of data is also used in tabular presentation and is of four types; viz., Geographical or Spatial Classification, Chronological or Temporal Classification, Qualitative Classification, and Quantitative Classification.
- **1. Spatial Classification of Data and Tabular Presentation**
- Spatial Classification of data means to classify data based on the geographical location, place, or region such as state, district, town, city, country, etc.

For example

- a number of students from different states at Delhi University. The Tabular presentation of the same can be shown as follows:

State	Number of Students (in '000)
Haryana	2,456
Maharashtra	1,808
Bihar	336
Assam	1,921
Sikkim	5,287

2. Temporal Classification of Data and Tabular Presentation

- Temporal Classification of data means to classify data based on the time period. It means that time becomes the classifying variable in the case of temporal classification

For example

- the sale of Laptops by a manufacturer in different years. The tabular presentation of the same can be shown as follows:

Year	Sale (Units)
2015	25,000
2016	46,000
2017	70,000
2018	90,000
2019	1,00,000

3. Qualitative Classification of Data and Tabular Presentation

- Qualitative Classification of data means to classify data based on qualitative characteristics or attributes

For example

- data of the students of Class XI can be classified on qualitative attributes like male or female, and Commerce or Science. The tabular presentation of the same can be shown as follows:
-

Sex	Number of Students	
	Commerce	Science
Female	32	26
Male	15	30
Total	**47**	**56**

4. Quantitative Classification of Data and Tabular Presentation

- Quantitative Classification of data means to classify data based on the quantitative characteristics.

For example

- data on the number of players playing different sports in a school. The tabular presentation of the same can be shown as follows:

Sports	Number of Players
Cricket	25
Football	36
Table Tennis	13
Basketball	27

Presentation of Statistical Data

- Data presentation is defined as the process of using various graphical formats to visually represent the relationship between two or more data sets so that an informed decision can be made based on them.
- Statistics is all about data. Presenting data effectively and efficiently is an art. You may have uncovered many truths that are complex and need long explanations while writing. This is where the importance of the presentation of data comes in. You have to present your findings in such a way that the readers can go through them quickly and understand each and every point that you wanted to showcase. As time progressed and new and complex research started happening, people realized the importance of the presentation of data to make sense of the findings.

Types of Data Presentation

- Broadly speaking, there are three methods of data presentation:
  - Textual
  - Tabular
  - Diagrammatic

Textual Ways of Presenting Data

- Out of the different methods of data presentation, this is the simplest one. You just write your findings in a coherent manner and your job is done. The demerit of this method is that one has to read the whole text to get a clear picture. Yes, the introduction, summary, and conclusion can help condense the information.

Tabular Ways of Data Presentation and Analysis

- To avoid the complexities involved in the textual way of data presentation, people use tables and charts to present data. In this method, data is presented in rows and columns - just like you see in a cricket match showing who made how many runs. Each row and column have an attribute (name, year, sex, age, and other things like these). It is against these attributes that data is written within a cell.

Diagrammatic Presentation: Graphical Presentation of Data in Statistics

- **Bar Diagram**
- **Pie Chart**
- **Histogram**
- **Frequency Polygon**
- **Ogive**

What are the 4 types of Tabular Presentation?

The tabular presentation method can be further divided into 4 categories:

- Qualitative
- Quantitative
- Temporal
- Spatial

Qualitative classification is done when the attributes in the table are some kind of ‘quality’ or feature. Suppose you want to make a table where you would show how many batsmen made half-centuries and how many batsmen made centuries in IPL 2020. Notice that the data would have only numbers - no age, sex, height is needed. This type of tabulation is called quantitative tabulation.

If you want to make a table that would inform which year’s world cup, which team won. The classifying variable, here, is year or time. This kind of classification is called Temporal classification.

If you want to list the top 5 coldest places in the world. The classifying variable here would be a place in each case. This kind of classification is called Spatial Classification.

Frequency Distribution

- Frequency distributions are **portrayed as frequency tables or charts**. Frequency distributions can show either the actual number of observations falling in each range or the percentage of observations.

Thank You