

Class: MSc

Subject : Research Methodology

M.M.S./PGDM

Introduction to Statistics

Elements and Variables

- Entity under study represent a set, group or collection of Elements.
 - Elements are like items on which data are collected.
 - Measurable, quantifiable, countable or classifiable attribute which varies from one element to another element in a group is termed as variable.
 - Variable is attribute or characteristic of interest for the element.
-

Example of Data Set

- Entity: A group of students.

Elements	Variable or Attribute				
Student	Gender	Grade	Marks scored in Physics	Marks scored in Chemistry	Marks scored in Mathematics
Radha	F	AA	95	93	89
Rahul	M	BB	65	63	66
Raj	M	AB	91	87	88
Reena	F	AB	85	82	87
Reshma	F	BA	78	75	81
Rishit	M	BA	73	78	76
Rohit	M	AB	81	85	84

Observation

- Data is generated from actual measurement on each variable for every element of the entity under study.
 - The set of measurements for a particular element is called as observation.
-

Scales of Measurements

- Scale of measurement represents the nature of data obtained on a particular variable of an element.
 - It indicates the most appropriate data summarization and statistical analyses.
 - Four scales of measurement are
 - Nominal
 - Ordinal
 - Interval
 - Ratio
-

Scales of Measurements – Nominal Scale

Data for a variable consists of labels or names used to identify an attribute of the element.

Examples:

- Gender of an employee
- Designation of Faculty in a college

Generally nominal data is non-numeric, but some times numeric codes are used to label the attribute. In that case, the scale of measurement remains nominal.

Scales of Measurements – Ordinal Scale

Data for a variable exhibit all the properties of nominal data and in addition, have natural ordering or rank.

Examples:

- Ratings in survey or feedback
- Economic status: High, medium, low
- Education level: High school, collegiate

Ordinal data may be non-numeric or numeric.

Comparison – Nominal and Ordinal Data

- Arithmetic operations cannot be performed on nominal data.
Ordinal data provide sequence but arithmetic operations cannot be performed even if data appears numeric.
 - Nominal data cannot be used to compare the elements with one another but helps to compare elements with one another.
-

Scales of Measurements – Interval Scale

Data for a variable have the properties of ordinal data and the interval between values is expressed in terms of a fixed unit of measure. Interval data are always numeric in nature.

Examples:

- Scores of students in qualifying or entrance examination.

The students can be ranked on the basis of scores. The differences between the scores are meaningful.

Scales of Measurements – Ratio Scale

Data for a variable exhibit all the properties of interval data and the ratio of two values is meaningful. Ratio data are always numeric in nature.

Examples:

- Height, weight, cost and time.

Rohit's weight is double the weight of Rahul.

This scale requires zero value be included to indicate that nothing exist for the variable at zero point.

Types of Data

- Categorical or Qualitative data –Data that use labels or names to identify an attribute of each element.
- Quantitative data - Data that use numeric values to indicate how much or how many of something.

The statistical method appropriate for summarising data depends upon the type of the data.

Categorical or Qualitative Data

Data that can be grouped by specific category.

Categorical data uses either nominal or ordinal scale of measurement.

Variables taking categorical data are categorical variables.

Quantitative Data

Quantitative data uses either interval or ratio scale of measurement.

How much? Continuous Data - weight, length

How many? Discrete Data – units sold

Discrete data takes only certain values.

Variables taking quantitative data are quantitative variables.

Arithmetic operations are meaningful only if the data are quantitative.

Activity – Identify Type of Data

Data from production records –

Name of employee

- Categorical data, Nominal scale

Product number

- Categorical data, Nominal scale

Quantity produced

- Quantitative data, Ratio scale

Labour cost

- Quantitative data, Ratio scale
-

Cross-sectional Data

Cross sectional Data are data collected at the same or approximately same point in time.

Examples

- Gross annual income of employees during financial year of 2022-23
 - Sale of leather jackets at different Killer outlets in Mumbai during November and December 2022.
-

Time Series Data

Time series Data are collected over several time periods.

Examples

- Average annual rainfall for last 10 years
 - Three monthly Hba1c count for a diabetic patient for last 2 years
-

Sources of Data

Existing sources

- Internal Data - Databases available within organization like database of employees, customers, transactions.
- External Data – Database services provided by specialized companies.

Example – *brightdata.com*

Sources of Data

Statistical Studies

If data needed for particular application are not available through existing sources, those can be often obtained by conducting statistical study.

Statistical study may be

- Experimental
 - Observational
-

Experimental Study

It involves identification of the variables of interest – independent and dependent. The independent variable is then altered in controlled manner to measure the change in values of dependent variable.

Example – Impact of new drug developed by Pharmaceutical company on blood pressure. Blood pressure is dependent variable while dosage level of drug is independent variable.

Observational or Non-Experimental Study

In observational or non-experimental statistical study, no variables of interest are identified.

Surveys are the most common type of observational studies. They are conducted through personal interviews or by requesting respondents to answer the well designed questionnaire.

Economic Consideration

Time and cost are always involved in collecting data and carrying out statistical analysis.

Decision maker should always consider the fact that the cost of data acquisition and the subsequent statistical analysis should not exceed the savings generated by using the information to make a better decision.

Data Acquisition Errors

There is always a possibility of data errors in statistics.

Using erroneous data can be worse than not using any data at all.

Error in data acquisition occurs whenever data value obtained is not equal to actual value. Great care needs to be taken in collecting and recording the data.
