UNIT – 2

1. **Accepts of method validation:**

   Method validation ensures that the research method or technique used to collect or analyze data is **scientifically sound and reliable**.

   Accepts / Criteria:

   - **Accuracy**: The method must measure what it is intended to measure.
   - **Reliability**: Results should be consistent over time and across different conditions.
   - **Validity**: Both internal and external validity should be ensured.
     - *Internal validity*: The results are due to the variables tested.
     - *External validity*: The findings can be generalized to other settings.
   - **Reproducibility**: Other researchers should be able to replicate the method and obtain similar results.
   - **Sensitivity and Specificity**: The method should detect small differences and differentiate between conditions accurately.
   - **Robustness**: The method should work under a variety of conditions.

2. **Observation**

   Observation is a fundamental technique used to gather primary data, especially in qualitative research.

   Accepts / Criteria:

   a. **Objectivity**: The observer must remain unbiased and avoid inserting personal feelings.
   b. **Systematic Approach**: Observations should follow a structured and pre-defined protocol.
   c. **Controlled Conditions (in structured observation):** The environment should be controlled to avoid outside influences.
   d. **Ethical Conduct:** Ensure informed consent, privacy, and non-intrusive behavior.
   e. Recording Accuracy: Observations should be documented promptly and precisely.
   f. **Reliability and Validity:** The observations should be repeatable and relevant to the research objectives.

3. **Data Collection**

   Data collection is the process of gathering information to address research questions or hypotheses.

   Accepts / Criteria:

a. **Relevance**: Data collected must be directly related to the research objectives.
b. **Completeness**: Ensure all necessary data is collected for comprehensive analysis.
c. **Accuracy**: Avoid errors in measurement, recording, or reporting.
d. **Timeliness**: Data should be collected within the appropriate time frame.
e. **Reliability**: Data must be consistent when collected under similar conditions.
f. **Ethical Standards**: Consent, confidentiality, and proper use of data should be maintained.
g. **Appropriate Tools and Techniques**: Use the right tools (e.g., questionnaires, interviews, sensors) suitable for the research type.

4. Methods of Data Collection:

a. Observations

i. **Definition**: Observation involves systematically watching and recording behavior, events, or conditions as they occur in natural or controlled settings.
ii. **Types**: Participant observation (researcher actively engages) and non-participant observation (researcher does not interact).
iii. **Advantages**: Provides contextually rich data, ideal for studying behaviors in real-time.
iv. **Example**: Observing classroom dynamics to assess teacher-student interaction.

b. Interviews

i. **Definition**: Interviews are a qualitative data collection method involving direct interaction between the interviewer and respondent. They can be structured, semi-structured, or unstructured, depending on the study.
ii. **Types**: Structured (fixed questions), semi-structured (some flexibility), unstructured (free-flowing conversation).
iii. **Advantages**: Provides in-depth insights, allows for probing questions.
iv. **Example**: An interview with a healthcare provider to understand patient challenges and care practices.

c. Surveys and Questionnaires

i. **Definition**: Surveys and questionnaires involve asking a set of questions to a target audience, either in person, by mail, online, or over the phone.

ii. **Types**: Surveys can be structured (fixed response options) or unstructured (open-ended responses).
iii. **Advantages**: Cost-effective, efficient for collecting large amounts of data.
iv. **Example**: A customer satisfaction survey to assess user experience with a product.

d. <u>Focus Groups</u>

i. **Definition**: Focus groups are a qualitative data collection method involving a small group discussion guided by a moderator to gather participants' views on a specific topic.
ii. **Advantages**: Offers diverse perspectives, helps identify trends or common themes.
iii. **Example**: Conducting a focus group with parents to discuss concerns about remote learning.

e. <u>Experiments</u>

i. **Definition**: Experiments are controlled studies where researchers manipulate one or more variables to observe their effect on other variables. This method is common in scientific and psychological research.
ii. **Advantages**: Provides causal relationships, highly controlled and replicable.
iii. **Example**: Testing the effects of a new drug on patient recovery rates in a clinical trial.

5. <u>Sampling Methods:</u>

Sampling is the process of selecting a subset of individuals or items from a larger population to make inferences about that population. Researchers use sampling to collect data more efficiently and to generalize findings to the entire group without surveying everyone.

a. **<u>Key Objectives of Sampling</u>**:
   1. **Reduce Costs and Time**: Sampling allows for efficient data collection by focusing on a representative subset.
   2. **Improve Accuracy**: Smaller, well-designed samples can lead to more accurate, focused data collection.

3. **Ensure Representativeness**: By carefully selecting a sample, researchers can ensure that the findings are relevant to the larger population.

b. <u>Types of Sampling Methods</u>

Sampling methods can be broadly classified into two categories: **probability sampling** and **non-probability sampling**.

i. <u>Probability Sampling:</u>

In probability sampling, every individual or item in the population has a known, non-zero chance of being selected. This type of sampling is often used when researchers aim for unbiased, generalizable results.

Examples of Probability Sampling:

a. Simple random sampling:

i. Technique: Each individual in the population has an equal chance of being selected. Researchers use random number generators or random selection tools to choose participants.

ii. Example: A school administrator randomly selects 50 students from a list of all students to survey about cafeteria satisfaction.

b. **Stratified Sampling**

i. **Technique**: The population is divided into subgroups (strata) based on a characteristic (e.g., age, gender), and random samples are taken from each subgroup.

ii. **Example**: In a study on employee satisfaction, researchers divide employees into departments (e.g., sales, HR, finance) and randomly select employees from each department.

c. **Systematic Sampling**

i. **Technique**: A starting point is randomly selected, and then every kth individual is chosen from a list. This method is often used when there's a fixed pattern or order in the population list.

ii. **Example**: A researcher wants to survey a population of 1,000 people and decides to select every 10th person on a sorted list after a random start.

d. **Cluster Sampling**

i. **Technique**: The population is divided into clusters (groups) that are randomly selected. All individuals within selected clusters are then included in the sample.
ii. **Example**: In a national health study, a researcher randomly selects specific cities (clusters) and surveys all residents within those cities.

ii. <u>Non-Probability Sampling</u>

In non-probability sampling, individuals are selected based on specific characteristics or convenience rather than random selection. This method is suitable for exploratory research where generalizability is less critical.

Examples of Non-Probability Sampling:

a. **Convenience Sampling**
   i. **Technique**: Participants are selected based on availability or ease of access, making it a fast and easy sampling method.
   ii. **Example**: A psychology student surveys classmates because they are easily accessible and available for quick data collection.

b. **Quota Sampling**
   i. **Technique**: The population is divided into categories (e.g., age, gender), and a specified number of participants from each category is chosen non-randomly.
   ii. **Example**: A researcher studying consumer preferences might set a quota to survey 50 men and 50 women in a shopping mall.

c. **Snowball Sampling**
   i. **Technique**: Participants recruit other participants, making it useful for studying hard-to-reach populations.
   ii. **Example**: In a study on experiences of ex-convicts, initial participants refer other ex-convicts they know, expanding the sample.

d. **Purposive Sampling**

i. **Technique**: Participants are selected based on specific criteria or characteristics relevant to the study's purpose.
ii. **Example**: In a study on the effects of leadership training, a researcher selects participants who hold managerial positions to gain insights specific to leaders.

6. **Methods of Data Processing**

Data processing is concerned with editing, coding, classifying, tabulating and charting and diagramming research data. The essence of **data processing in research** is data reduction. Data reduction involves winnowing out the irrelevant from the relevant data and establishing order from chaos and giving shape to a mass of data. **Data processing in research** consists of five important steps. They are:

a. Editing of Data

Editing is the first step in **data processing**. Editing is the process of examining the data collected in questionnaires/schedules to detect errors and omissions and to see that they are corrected and the schedules are ready for tabulation. When the whole data collection is over a final and a thorough check up is made. Mildred B. Parten in his book points out that the editor is responsible for seeing that the data are;

1. Accurate as possible,
2. Consistent with other facts secured,
3. Uniformly entered,
4. As complete as possible,
5. Acceptable for tabulation and arranged to facilitate coding tabulation.

b. Coding of Data

Coding is necessary for efficient analysis and through it the several replies may be reduced to a small number of classes which contain the critical information required for analysis. Coding decisions should usually be taken at the designing stage of the questionnaire. This makes it possible to pre-code the questionnaire choices and which in turn is helpful for computer tabulation as one can straight forward key punch from the original questionnaires. But in case of hand coding some standard method may be used. One such standard method is to code in the margin with a colored pencil. The other method can be to transcribe the data from the questionnaire to a coding sheet. Whatever method is adopted, one should

see that coding errors are altogether eliminated or reduced to the minimum level.

c. Classification of Data

Classification or categorization is the process of grouping the statistical data under various understandable homogeneous groups for the purpose of convenient interpretation. A uniformity of attributes is the basic criterion for classification; and the grouping of data is made according to similarity. Classification becomes necessary when there is a diversity in the data collected for meaningless for meaningful presentation and analysis. However, it is meaningless in respect of homogeneous data. A good classification should have the characteristics of clarity, homogeneity, equality of scale, purposefulness and accuracy.

Objectives of Classification are below:

i. The complex scattered and haphazard data is organized into concise, logical and intelligible form.
ii. It is possible to make the characteristics of similarities and dis – similarities clear.
iii. Comparative studies is possible.
iv. Understanding of the significance is made easier and thereby good deal of human energy is saved.
v. Underlying unity amongst different items is made clear and expressed.
vi. Data is so arranged that analysis and generalization becomes possible.

d. Data Diagrams

Diagrams are charts and graphs used to present data. These facilitate getting the attention of the reader more. These help presenting data more effectively. Creative presentation of data is possible. The data diagrams classified into:

i. **Charts:** A chart is a diagrammatic form of data presentation. Bar charts, rectangles, squares and circles can be used to present data. Bar charts are uni-dimensional, while rectangular, squares and circles are two-dimensional.
ii. **Graphs:** The method of presenting numerical data in visual form is called graph, A graph gives relationship between two variables by means of either a curve or a straight line. Graphs may be divided into two categories. (1) Graphs of Time Series and (2) Graphs of Frequency Distribution. In graphs of time series one of the factors is time and

other or others is / are the study factors. Graphs on frequency show the distribution of by income, age, etc. of executives and so on.

7. **Data Analysis**

Data analysis in research is the systematic use of statistical and analytical tools to describe, summarize, and draw conclusions from datasets. This process involves organizing, analyzing, modeling, and transforming data to identify trends, establish connections, and inform decision-making.

- **Main Goals**: Describe data, make inferences, predict future events, and provide data-driven recommendations.
- **Stages of Data Analysis**: Data collection, preprocessing, exploratory data analysis, model building and testing, interpretation, and reporting.

Types of Data Analysis

a. **Descriptive Analysis**

Descriptive analysis focuses on summarizing and describing the features of a dataset. It provides a snapshot of the data, highlighting central tendencies, dispersion, and overall patterns.

- **Central Tendency Measures**: Mean, median, and mode are used to identify the central point of the dataset.
- **Dispersion Measures**: Range, variance, and standard deviation help in understanding the spread of the data.
- **Frequency Distribution**: This shows how often each value in a dataset occurs.

b. **Inferential Analysis**

Inferential analysis allows researchers to make predictions or inferences about a population based on a sample of data. It is used to test hypotheses and determine the relationships between variables.

- **Hypothesis Testing**: Techniques like t-tests, chi-square tests, and ANOVA are used to test assumptions about a population.
- **Regression Analysis**: This method examines the relationship between dependent and independent variables.
- **Confidence Intervals**: These provide a range of values within which the true population parameter is expected to lie.

c. **Exploratory Data Analysis (EDA)**

EDA is an approach to analyzing data sets to summarize their main characteristics, often with visual methods. It helps in discovering patterns, spotting anomalies, and checking assumptions with the help of graphical representations.

- **Visual Techniques**: Histograms, box plots, scatter plots, and bar charts are commonly used in EDA.

- **Summary Statistics**: Basic statistical measures are used to describe the dataset.

d. Predictive Analysis

Predictive analysis uses statistical techniques and machine learning algorithms to predict future outcomes based on historical data.

- **Machine Learning Models**: Algorithms like linear regression, decision trees, and neural networks are employed to make predictions.
- **Time Series Analysis**: This method analyzes data points collected or recorded at specific time intervals to forecast future trends.

e. Causal Analysis

Causal analysis aims to identify cause-and-effect relationships between variables. It helps in understanding the impact of one variable on another.

- **Experiments**: Controlled experiments are designed to test the causality.
- **Quasi-Experimental Designs**: These are used when controlled experiments are not feasible.

8. Hypothesis Testing

Hypothesis method ***compares two opposite statements about a population and uses sample data to decide which one is more likely to be correct***.To test this assumption we first take a sample from the population and analyze it and use the results of the analysis to decide if the claim is valid or not.

a. Defining Hypotheses

i. **Null hypothesis (H0):** The null hypothesis is the starting assumption in statistics. It says there is no relationship between groups. For Example A company claims its average production is 50 units per day then here:
Null Hypothesis: H_o: The mean number of daily visits (μ) = 50.

ii. **Alternative hypothesis (H1):** The alternative hypothesis is the opposite of the null hypothesis it suggests there is a difference between groups. **like** The company's production is not equal to 50 units per day then the alternative hypothesis would be:
H_1: The mean number of daily visits (μ) ≠ 50.

b. Types of Hypothesis Testing

It involves basically two types of testing:

i. One-Tailed Test

1. A one-tailed test is used when **we expect a change in only one direction—either an increase or a decrease but not both**. Let's say if we're analyzing data to see if a new algorithm improves accuracy we would only focus on whether the accuracy goes up not down.
2. The test looks at just one side of the data to decide if the result is enough to reject the null hypothesis. If the data falls in the critical region on that side then we reject the null hypothesis.
3. There are two types of one-tailed test:
    a. **Left-Tailed (Left-Sided) Test:** If the alternative hypothesis say that the true parameter value is less than the null hypothesis. then it is a Left tailed test. Example: H0:$\mu \geq 50$ and H1: $\mu < 50$
    b. **Right-Tailed (Right-Sided) Test**: when the alternative hypothesis say that the true parameter value is greater than the null hypothesis then it is called Right Tailed test. Example: H0 : $\mu \leq 50$ and H1:$\mu > 50$

ii. Two-Tailed Test

1. A two-tailed test is used when **we want to check for a significant difference in both directions—whether the result is greater than or less than a specific value**. We use this test when we don't have a specific expectation about the direction of change.
2. If we are testing whether a new marketing strategy affects sales we want to know if sales increase or decrease so we look at both possibilities.
3. Example: H0: $\mu = \mu = 50$ and H1: $\mu \neq 50$