Unit V: Measurement, Scaling and Sampling

Data Measurement Scale

Measurement may be defined as the assignment of numbers to objects or events according to certain rules. There are generally four types of measurement scales, which are as follows.

a) Nominal scale

Nominal scale is used for measuring variables which are qualitative in nature. It is the first level of measurement where labels are assigned to the attributes of the variables in the form of number. Numbers are used as mere identifiers and do not hold any numerical value & no arithmetic operations can be drawn upon them. It only satisfies the 'Identity' property of scale of measurement. Nominal scale is the simplest scale & is also called as the 'Categorical scale' because it represents only the names or categories. It is also called as least powerful level of measurement. The only statistical analysis that can be performed on a nominal scale is frequency count. Mode is used as a measure of central tendency.

For example: -

- jersey number of players in cricket team, types of hair color, PAN number, Telephone number etc.
- Another example, what is your gender? – Male (1) or Female (2)

b) Ordinal scale

Ordinal scale is used for measuring variable which are qualitative in nature. It is the second level of measurement where labels are assigned to the variable in the form of numbers & they are arranged in a proper order. Not only the numbers but also the order of the variables is important. That's why it is called as ordinal scale. It satisfies the 'Identity' & 'Magnitude/Order' property.

*Ordinal scales measure non-numeric concepts like satisfaction, happiness, discomfort, beauty etc. By giving ranks. Median or mode are used as the measures of central tendency & spearman's rank correlation.

For example,

Order- How much happy are you with our services?

Very happy-1

Happy-2

Neutral-3

Unhappy-4

Very unhappy-5

- Another example, Ranks of students in an academic test, health status (excellent, average, poor)

Here, the order is represented but the difference between the variable is not indicated.

c) Interval scale

Interval scale is used for measuring variables which are quantitative in nature. It is the third level of measurement where labels are assigned to the variables in the form of numbers & they are arranged in a proper order with equal differences between the values. Along with the numbers & order, the difference between the values is also known. That's why it is called an Interval scale. It is an extension of ordinal scale. (i. e., it possesses the property of identity, order & equal intervals). Arithmetic operations like addition & subtraction can be performed on the variables but not multiplication & division and hence, ratios can't be calculated. Interval scales don't have a true zero meaning negative values also exist. Like -10-degree Celsius temperature. Mean, median, mode is used as the measure of central tendency. And standard deviation and range are used as the measures of dispersion. For example, a temperature scale where difference between 60 & 70 degree Celsius is same as that of the difference between 20 &30 degree Celsius

d) Ratio scale

Ratio scale is used for measuring variables which are quantitative in nature. It is the fourth level of measurement which possesses all the attributes of an interval scale along with the property of

absolute zero. Arithmetic operations like addition & subtraction can be performed on the variables along with multiplication & division. Here, the ratios can be calculated. That's why it is called a ratio scale. Like, weight of ram is double of that of Shyam. Ratio scales have a true zero meaning negative values don't exist. Like there cannot be a negative weight or negative length. It is the most powerful level of measurement. Mean, median, mode, harmonic mean, geometric mean are used as the measures of central tendency. And standard deviation and coefficient of variation are used as the measures of dispersion. For example, height, weight, length, distance etc.

Summary

Feature	Nominal	Ordinal	Interval	Ratio
Level of measurements	First	Second	Third	Fourth
Type of variable	Qualitative	Qualitative	Quantitative	Quantitative
Identity	Yes	Yes	Yes	Yes
Magnitude/order	No	Yes	Yes	Yes
Equal interval	No	No	Yes	Yes
Absolute zero	No	No	No	Yes
Central tendency	Mode	Median & mode	Mean, median & mode	Mean, median, mode, geometric & harmonic mean
Source of dispersion	…………..	………………	Standard deviation & range	Standard deviation & coefficient of variation
Arithmetic operation	…………..	……………..	Only addition & subtraction	Add, subtract, multiply & divide
Statistical tests	Non-parametric	Non-parametric	Parametric	Parametric

Characteristics of Good Measurement

Generally, measurement should be able to measure the things which a researcher intends. The tools, which are used, should be simple and able to increase the efficiency of a researcher. The main criteria for testing the goodness of measures are validity, reliability and practicability. These features of measurement are described below:

1. Reliability

Reliability refers to how consistently a method measures something. If the same result can be consistently achieved by using the same methods under the same circumstances, the measurement is considered reliable.

For example: you measure the temperature of a liquid sample several times under identical conditions. The thermometer displays the same temperature every time, so the results are reliable.

Test of reliability

a) Test-retest method

According to this method, the same scale is applied twice to the same population & the results obtained are compared by computing correlation between the first and second set of scores. If the correlation is high, it can be assumed that the measuring instrument used in the research is highly reliable but if there is low correlation then the instrument is less reliable so the modification in instrument is essential.

b) Alternative or parallel form method

According to this method, two forms of a scale are constructed & applied to the same population. If the results obtained by two methods show high degree of similarity, then the scale is considered reliable. In this method, both forms have similar items and the same response format with only the wording and ordering of the question are changed. These two forms are administered in the same time to the same sample; so, there is less chance of having errors due to [ace of time.

c) Split half method

The test is first divided into two equivalent halves and the correlations thus computed from two components and the test is divided into two components as odd and even numbers of sets. The reliability coefficient is called stepped up reliability coefficient and mathematically it is given by

$$R_s = \frac{nr_p}{1+(n-1)r_p}$$

where r_p is the correlation coefficient of two parts and n is the number of parts.

d) Rational- equivalence method

When the proportion of the group doing correctly & the proportion doing incorrectly are given, the reliability is calculated by using formula,

$$R_r = \frac{n}{n-1} \frac{(\sigma^2 - \sum pq)}{\sigma^2}$$

Where p is the proportion of test items correctly and q is the incorrect. σ is the standard deviation of the test score and n is the number of items in the test.

When mean of test score is given,

$$R_r = \frac{n\sigma^2 - \bar{X}(n-\bar{X})}{(n-1)\sigma^2}$$

Where $\bar{X}$ mean of the test score.

1. A test is divided into two parts where the correlations coefficient between two parts is 0.72. Find the stepped-up reliability coefficient of the test.
2. The stepped-up reliability coefficient is found to be 0.75. What will be the effect of the test reliability if the length of the test is tripled.
3. A test of 75 items is administered and the standard deviation of the test is 5.9, the sum of the product of the proportion of the test items correctly and incorrectly is 14.37. Find the reliability coefficient of the test.
4. A test of 50 structural multiple-choice tests is administered to the 100 college students where correct answer indicating 1 and incorrect answer indicating 0 of the tests. Mean score of the test is 27.75 and their standard deviation is 5.80. Find the reliability coefficient of the test.

2. Validity

Validity refers to how accurately a method measures what it is intend to measure. If research has high validity, that means it produces results that correspond to real properties, characteristics & variations in the physical or social world. High reliability is one indicator that a measurement is valid. If a method is not reliable, it probably isn't valid. There are three types of validity.

- **Content validity**
  It is also known as face validity. It refers to the adequate coverage of the concept. In other words, content validity ensures that the measuring tools include an adequate and representative set of items that would tap the concept. The more scale item represents the concept of the research topic; the instrument has the higher validity. If the instrument contains a representative sample of the universe, then the content validity is high. Its determination is judgmental and intuitive but it can also be determined by using a panel of person who shall judge how well the measuring instrument meets the standards, but there is no numerical way to express it.
  Content validity ratio (CVR) is measured as:

$$CVR = \frac{ne - \frac{N}{2}}{\frac{N}{2}}$$

Where ne is the number of expert panelists indicating "essential" and N is the total number of expert panelists. -This formula yields values of CVR range from +1 to -1; positive values indicate that at least half the experts rated the item as essential.

- **Criterion-Related validity**
  Criterion validity measures how well one measure predicts an outcome for another measure. A test has this type of validity if it is useful for predicting performance or behavior in another situation. Criterion validity evidence involves the correlation between the test and a criterion variable (or variables) taken as representative of the construct. For example, employee selection tests are often validated against measures of job performance (the criterion), and IQ tests are often validated against measures of academic performance (the criterion).
  a) If the test data and criterion data are collected at the same time, this is referred to as concurrent validity evidence.
  b) If the test data are collected first in order to predict and criterion data collected at a later point in time, then this is referred to as predictive validity evidence.

  It relates to our ability to predict some outcomes or estimated the existence of some current situation.

- **Construct validity**
  Construct validity is "the degree to which a test measures what it claims to be measuring". If a measure confirms the predicted correlation with other theoretical preposition, then such measure possesses construct validity. It wants the agreement between a theoretical concept and a specified measuring instrument. Construct validity is the appropriateness of inferences made on the basis of observations or measurements (often test scores), specifically whether a test measures the intended construct. Constructs are abstractions that are deliberately created by researchers in order to conceptualize the latent variable, which is correlated with scores on a given measure (although it is not directly observable). Construct validity examines the question: Does the measure behave like the theory says a measure of that construct should behave?

Relationship between Reliability and Validity

The relationship between reliability and validity can be assessed in various ways. For a measurement to be valid, it has to be reliable. While validity is associated with accuracy, reliability is all about consistency. Therefore, an unreliable measurement cannot be valid. However, a measurement can be reliable without being valid. It is often required for measurements to be both valid and reliable. The process of assessing psychological traits is called psychological measurement. In psychological measurements, reliability is used to determine the consistency of measurements and results. For example, when using test-retest reliability, an intelligent individual by certain metrics is expected to exhibit intelligence at all times. Validity measures the agreement

of test results with that which the test is intended to measure. For instance, somebody can decide to use the height of individuals as a measurement for confidence. The test may have several consistent results making it reliable, but that doesn't mean it is valid. Confidence has nothing to do with an individual's height. A psychological test's high validity and high reliability indicate that the test is effective and efficient. Therefore, the test can be used for research studies, evaluating job applicants, etc. In terms of coefficients, the expected values for reliability and validity should be 0.6 or greater.

Scaling

Scaling is the assignment of numerals to the object items or variables. In scaling, individuals are sorted according to some known and specified characteristics or attributes which helps in comparative study of individuals. Good scaling must hold some of the special characteristics.

Characteristics of scaling

- Simple
- Reliable
- Valid
- Economic
- Exact
- Comprehensive
- Practical

Scaling consists of psychophysical and Attitudinal scaling.

a) **Psychological scaling**- like honesty, aggressiveness, pleasantness, quality of beauty, sense of human behavior, etc.
b) **Attitudinal scaling**- feeling, perception and behaviors towards other things -people, places, times, etc. are called attributes.

Techniques for developing attitude scale

There are four techniques for developing attitude scales. They are given below,

a) **Choice:** Respondents were given number of alternatives and ask to choose the preferred alternative. If a respondent chooses one alternative, then it is assumed that respondent prefers that alternative than other alternatives.
b) **Ranking:** Various alternatives are provided to the respondents and asked to give rank as per their priority. Respondent provides rank to the given alternatives that helps to know their attitude.
c) **Rating:** Researcher asks the respondent to estimate the magnitude of quality or a characteristic that a product or event possess. The position is found/seen on the basis of their rating.

d) **Sorting:** Several concepts are provided to the respondents and asked to arrange them in order on the basis of their priority/ value. Such arrangement reflects the attitude of the respondents.

Attitude Measurement

Every person is unique in their experiences, behaviors, and culture. Human attitudes are shaped by their experiences, behavior, and culture. An attitude is a person's response to any situation, item, or other object. Research in social science examines how individuals and societies behave and think. In order to investigate these aspects, a researcher has to gauge participants' attitudes. The primary goal of studying human attitudes is to identify the factors that influence human behavior. Researchers in social science have created a number of scales to evaluate people's attitudes. We refer to this type of measurement as attitude measurement.

Issues in attitude measurement

Measurement refers to the process of gathering data that may be examined. The technique of determining a person's attitude toward an object is known as attitude measurement. Prior to measuring attitudes or any other parameter, one has to clearly sort out the following:

- "what" has to be measured?
- "who" is to be measured?
- the accuracy desired in the measurement
- the costs permissible
- the choices available in the measurement/data collection techniques.

When measuring attitudes, the researcher's main focus is on determining the respondent's "state of mind". It might involve elements such as decision-making, attitudes, and awareness. The challenge of verifying these measures is an interesting feature. The "truth" about a respondent's level of liking for a new product, like ice cream mix, cannot be determined. Unless the researcher is a "telepathist," he is unable to analyze mental states like preferences, likes, dislikes, and so on. These things can only be assumed.

Scale construction for Attitude measurement

Attitude is a qualitative subject. Thus, numbers or symbols are offered for measuring such qualitative subjects. People's attitudes vary from one another. Thus, a variety of scales can be employed to quantify human attitudes. Some of the important method of attitude scale are given below:

a) Arbitrary scale

Sometimes, a researcher develops a new sensible and appropriate method. If he or she believes that the subject of research is completely novel and that a new approach or methodology is required for measuring attitude, he or she creates a new scale known as an arbitrary scale. A researcher must be able to demonstrate the reliability and validity of such a scale. Because the nature of social events varies, the researcher creates alternative scales from time to time and uses them as needed.

b) Summated rating scale or Likert scale

An assessment tool where respondents rate their level of agreement or disagreement with a sequence of statements assessing the same construct or variable. For every question, there are usually five to seven response alternatives (strongly agree to strongly disagree, for example). A person's overall attitude toward the construct of interest can be inferred by adding the response values for each question to get an average score. Summated rating scales that are most frequently employed are Likert scales. Summated scale is another name for the summated rating system.

Construction of Likert scale

- Write a significant number of statements on the particular attitudinal item under investigation. For example, one might consider the role of volunteer organizations in providing health services. in rural areas. The majority of these statements should be either somewhat positive or negative. Neutral objects are usually avoided on these scales. The elements should be evenly distributed between positive and negative remarks.
- Administer the pool of statements to a sample of respondents who are similar to the demographic for which the scale will be utilized. For example, if we wish to research housewives' attitudes, we should administer the pool to a group of housewives from a comparable background to our final sample.
- Assign scale values to the degrees of agreement or disagreement with each item. The particular values may differ from one researcher to another. Sometimes one may adopt the values 1, 2, 3, 4, 5 and sometimes +2, +1, 0, -1, -2. For negative items the directions should be reversed.
- Calculate a total attitude score for each respondent using the same scaling procedure. The distribution of total, scores is then used to refine the list of items. This step is called item analysis.
- The next stage is to identify the statements with the highest discriminatory power. A few extreme groups are viewed as representing the most and least positive attitudes.
- Only statements that correlate with the overall test should be included in the final instrument, with the rest being discarded/removed.

Scales used in Business Research / Social Science Research

Experts create several scales to assess an individual's or group's attitudes. The researcher uses various scales to measure the attitude of people or groups dependent on the nature of the event or the objectives of the study effort. Business research has its own unique nature and goals. As a result, when doing business research, multiple scales are used. Some of the most frequent scales used in business research are listed below:

1) Likert scale

It is the most extensively used scale in business research for assessing individual or group attitudes, norms, values, and behavior. In this scale, respondents are asked to express their opinions using the given scales, and they indicate whether they agree or disagree with the statement provided in

the questionnaire. It displays not just agreement and disagreement, but also the level of agreement or disagreement.

Example: employees are committed towards the organization.

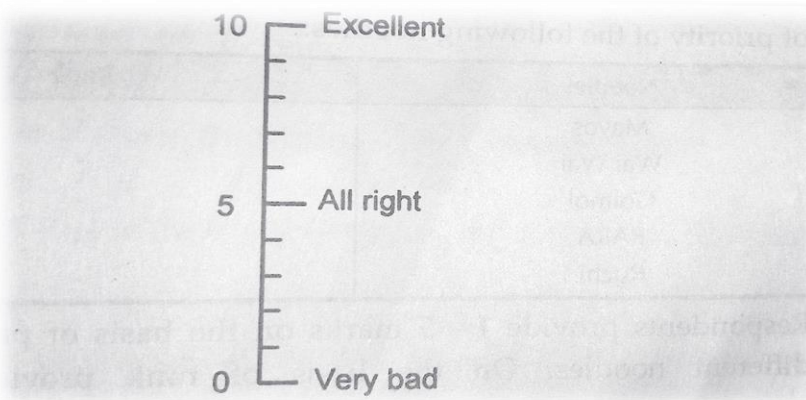1- Strongly disagree   2- Disagree   3- Undecided

4- Agree   5- Strongly agree

2. Graphic rating scale:

In this scale, participants are asked to mark the relevant spot on the chart to indicate how they feel about a certain issue. This type of labeling aids in comprehending respondents' opinions.

Example: On scale of 0 to 10 how do you rate your departmental head?

Request a single marking from the respondent on a scale of 0 to 10. This choice expresses the respondent's viewpoint. A researcher determines whether an employee views his or her department head favorably or unfavorably based on employee opinions.

3. Itemized rating scale:

Utilizing this scale, the researcher offers a range of possible answers from which the participant chooses the most pertinent response to address the research questions. Since it may be adjusted to many different scenarios where variables need to be monitored, it is well-liked in business research. For example:

a) How do you rate your interest in changing organizational policies?

Extremely poor   *Not at all*

Some what well   *Very well*

b) How well is the new distribution channel working?

Not at all   *somewhat good*   *very much good*

4. Rank order rating scale:

Respondents are requested to rank the provided products or things according to their priority using this scale. They rank them in order of highest priority to lowest. The methodology is comparative. It serves as a gauge for the organization's priority with regard to any topic. According to priority, it forces the responders to divide the product or problems into subsequent groups. Only ordinal data is developed by it.

Example: a researcher asked to the customers to identify the rank on the basis of priority of the following noodles:

Noodles	Rank
Wai Wai	-
Golmol	-
Rara	-
Ramba	-
Rumpum	-

Respondent provide 1-5 marks on the basis of priority to the different noodles. On the basis of rank provided by the respondents, a researcher finds out the priority of the customers.