UNIT 1 MEASUREMENT

Structure

1.1 INTRODUCTION

The first principle of a scientific study is to describe objectively what a scientist observed, under what circumstances, and to communicate the same as precisely as possible. A scientist reports not only what has been observed, but states the circumstances and the methods used, too. This is a high priority condition because others must be given a chance to verify those observations. In fact, specification of the conditions of observation is the first step in the measurement of a given phenomenon.

Although arguments continue over the nature of measurement, measurement in some form or other has always been there even when social sciences were no more than a branch of speculative philosophy. Quantitative principles from physics and chemistry have given us very precise and accurate measurement in these fields. Biological sciences, of late, have established principles that are nearly at par with those found in physical sciences. Social sciences lag far behind as compared to physical and biological sciences. Measurement is the key to all sciences.

After studying this unit you will be able to:

- explain the meaning, concepts and importance of measurement in social science research.
- describe the levels of measurement that quantify social variables.
- distinguish between various levels of measurement that have been used in the social science research.

1.2 MEASUREMENT: MEANING AND CONCEPT

Measurement is an inseparable part of any science, natural or social. Any science aims to obtain a specific and accurate measurement of the events, of the characteristics of the different units of a phenomenon, and, of the inter-relationship

between the units. Measurement is assigning numbers to objects or events according to rules (Stevens, 1946). The purpose is to have information in a form in which variables can be related to each other. In social science research we have to deal with various social and psychological variables. Their measurement is one of the vital stages in the research process. Measurement of social and psychological variables is a complex and demanding task.

Measurement, simply speaking, is the assignment of numerals or other symbols or signs (male, female, occupational categories, for example) to objects or events according to a set of operational rules. Measurement always refers to some property of the object or event and not the object or event by itself. In this measuring process, the observer follows a scheme or procedure by which observed events can be classified into non-overlapping categories unambiguously, and the categories are given labels - numerals or symbols. The basic assumption in measurement theory is that every event or object in a set of events or an object possesses a specific quantity of the property under observation. This quantity of the property can be compared directly against a standard scale (as when we measure the length or mass of a material) or can be evaluated fairly accurately by trained observers (judges or experts assessing the level of performance of a student in a debate or in class examination). Often the measurement operations involve the use of mechanical devices such as thermometer, barometer, measuring tape, or weighing scale. The use of such mechanical devices is the least observer dependent, and, hence, the measurements are fairly precise, accurate, and objective. The thermometer, for example, when applied to a given object, gives a number, the temperature. This is a technical refinement of the precision of a crude judgment into categories such as very hot, hot, warm, cold, etc., that are obtained by the impression of the observer when he touches an object with his finger. In social sciences, too, we make use of a physical (not necessarily mechanical) component or stimulus.

Social science research follows a standardized procedure or mechanism, as is followed in the physical sciences. When a scale is applied on a person, it gives us a number (or symbol) on his attitude, IQ, interest, emotional stability, motivation, and so on. It follows, then, that a measurement operation is always a standardized way of proceeding, which may or may not make use of mechanical devices or stimuli, but which always results in classification of the objects measured into some non-overlapping categories labelled by numerals, or simply by symbols. From the viewpoint of mathematics, measurement operation is a standardized rule that maps each of a set of objects into one, and only one, of a set of several categories or numbers.

Although arguments continue over the nature of measurement, measurement of some form or other has always been there even when social sciences were no more than a branch of speculative philosophy. Quantitative principles from physics and chemistry have given us very highly precise and accurate measurement in these fields. Biological sciences, of late, have established principles that are nearly at par with those found in physical sciences while, social sciences lag much behind in this respect.

1.3 IMPORTANCE OF MEASUREMENT

"Measurement consists of identifying the values which may be assumed by some variable, and representing these values by some numerical notation. The numerical notation is systematically and consistently assigned, that is, it is assigned according to some set of rules."

Thus, the numerals assigned to the variables, indicate different condition of the variables or different values of the variables or different degrees or intensity of a quality possessed by units. From the above it is evident that:

i) measurement is a purely descriptive process.

ii) measurement implies that the attributes of persons or objects are possessed in varying degrees and the degree of variation can be measured and represented.

iii) measurement, in essence, is a numerical process.

Common objects of measurement in sociology are individual's characteristics, interactions, interrelations, consciousness, participation, socialization, motivation, etc. Some of the properties, of the individuals are visible and easily measurable, such as age, income, height, etc. Some characteristics are abstract and it is difficult to measure them.

These days, both in sociology and psychology, the measurement of beliefs and attitudes is common because of the greater emphasis on a democratic form of government which demands an assessment of peoples' attitude and opinion, from time to time. Besides this, public opinion studies are carried on regularly by various public and governmental agencies. Even the commercial organizations measure peoples' opinion and attitude to know the future market of their products. Polling agencies measure people's opinion and attitude to know the people's preferences for particular political parties and candidates. Thus, they want to predict, on the basis of such polling, the possibility of winning for any particular candidate. Attitudinal studies may also help in predicting an individual's future behaviour and their possible reactions towards different developmental programmes. Such studies might also help in making policies and specially in implementing them. For example, the study of beliefs and attitudes of Indian people towards illness and health measures or family size would help in instituting a social and educational programme to mobilize the people towards vaccination or adoption of family planning in urban areas.

Measurement and quantification of variables, beliefs, attitudes, etc., do help in statistical manipulation of them, in experiment, and in testing specific hypothesis related to them. Thus, ultimately, it helps in the development and advancement of a theory.

1.4 MEASUREMENT POSTULATES

There are three postulates that are basic to measurement. A postulate is an assumption that is an essential prerequisite for carrying out some operations or line of thinking. In this case, it is an assumption about the relations between the objects being measured.

The three postulates that are basic to measurement are as follows.

a) Either (a=b) or (a ≠ b) but not both. This postulate says a is either equal to b or not equal to b, but not to both. We must be able to assert, either that one object is the same in a characteristic as another, or, it is not the same. In measurement 'the same' does not necessarily mean complete identity. It can mean 'sufficiently the same' to be classed as members of the same set

Example: Duration of variety X is greater than variety Y:

Yield of variety X is greater than variety Y

Height of a person X is greater than person Y'

b) If (a=b) and (b=c) then (a=c). This postulate says, "If a equals b, and b equals c, then a equals c. If one member, of a universe is the same as another member, and the second member is the same as third member, then the first member is the same as the third member. This postulate enables a, researcher, to establish the quality of set members, on a characteristic by comparing objects.

Example: As farmers who have T.V. and radio have the same level of mass media exposure as that of the small farmer having T.V. and radio, and, this, in turn, is equal to the marginal farmer having T.V. and radio.

c) If (a>b) and (b>c) then (a>c). The third postulate is of more immediate and practical importance for our purpose. It says "If a is greater than b, and b is greater than c, then a is greater than c. Other symbols or words can be substituted for greater than (>) less than (<), such as, is at a greater distance, that is stronger than, and soon. Most measurement in psychology and education depends on this postulate. It must be possible to assert ordinal-a rank-order statements, such as a has more property than b, b has more of property than c, thus, a has more property than c.

Example: Yield of variety X is more than variety Y and yield of variety Y is more than variety Z.

In the above section you have studied about the meaning and postulates of measurement. Now, answer the questions given in *Check Your Progress 1*.

Check Your Progress 1

Note: a) Write your answer in about 50 words.

b) Check your answer with possible answers given at the end of the unit

1) What do you mean by measurement?

.....

.....

.....

.....

.....

.....

2) Name different postulates of measurement.

.....

.....

.....

.....

.....

1.5 LEVELS OF MEASUREMENT

The theory of measurement in social sciences really consists of a system of distinct theories each concerned with a distinct level of measurement. A set of data often will satisfy some of the levels, but not others. It is, therefore, necessary to understand the basic nature of data before applying a particular level of measurement. Choice of an appropriate statistical model for analysis is largely dependent on this level of measurement achieved. The higher the level of measurement, the more informed we are about a variable.

1.5.1 Nominal or Categorical Scale

Measurement, in its simplest, most primitive, and weakest form exists when we can substitute numbers (without meaning its numeral values) or symbols to real objects. That is, we use these symbols or numbers merely to classify or categorize objects, persons or even characteristics. A scientist, at the simplest level, must think of a classification scheme so that the different recorded events can be fitted into this scheme of classification. We give each category of event or object a name, or a number, or a symbol for convenience of identification. These symbols or numbers then constitute a nominal or classificatory scale. The categories making up the scale must be mutually exclusive (each observation can be put under only one category) and exhaustive (there are sufficient number of categories so that each observation can be put under some category) and unordered. The categories making up a nominal scale are usually called attributes. Thus, for mammals the attribute, sex, has only two categories: male and female. A population can be classified by the occupations they follow, or, by their rural-urban origin; STD dialling code 033 indicates all telephone subscribers in the Calcutta telephone zone; blue, brown, black, etc., are categories of human eye colour.

In a nominal scale, the scaling operation consists of partitioning a given class into a set of mutually exclusive sub classes. The member of any sub class must be equivalent in the characteristic or the attribute being scaled. The only relation used in this scaling is equivalence. *A* sign of equality (=) symbolizes the relationship. In addition, the relationship of equality is reflexive, symmetric, and transitive. By reflexive we mean x = x for all values of x; by symmetry we mean that, if the relationship of equality holds between x and y, it also holds between y and x, i.e., if x = y, then y = x; by transitivity we mean if x = y, and y = z, then x = z.

Such a classification scheme utilizes only a part of the information about the objects or events being classified. We said earlier, we never measure an object or

an event. An object may have many distinguishable properties or attributes. The scientist singles out those attributes only that have relevance to the objectives of his study. Therefore, the act of measurement, however simple or elementary, requires some degree of abstraction from the complex of properties possessed by the event or the object. The level of measurement in a nominal scale is so elementary and crude that many scientists would not like to consider it as measurement at all. Nevertheless, such classification into mutually exclusive categories is a necessary condition for all higher levels of measurement.

On a nominal scale, we are able to classify objects or events into non-overlapping categories purely on the basis of some qualitative character of the studied variable.

1.5.2 Ordinal Scale

The ordinal scale enables the researcher to categorize individuals, objects, or survey responses on the basis of a particular characteristic that they have in common. For example, sometimes, the objects in one class of a nominal scale are not merely different from objects in another class on the same scale; they stand in some kind of relationship. Typical relationships among the classes are that the members of one class have more of some property or characteristic than those in other classes. Such a relationship is often designated by the carat (>), which means 'greater than'. The symbol > is used to express all such relations between classes as 'preferred to', 'more than, 'greater than', 'higher than', etc. Conventionally, the number 1 is assigned to the class, which numerically has larger quantity of the property than all other classes. Number 2 is assigned to that class that has less of the property compared to that of the class given the number 1, but more of it than the rest of the classes, and so on. That is, the numbers in such a classification indicate the place of a category or class in an ordered series. In a sprint event, whoever touches the finishing tape first is given the number 1; the person coming second, the number 2, and so on, for other sprinters, no matter if the second sprinter comes 1 second or 0.04 seconds after the first, the third runner taking 4 seconds, or, more than the runner in the second spot. Such numbers, in an ordered series, constitute an ordinal scale.

The ordinal numbers actually state the relative position or the amount of the characteristic relative to others. The rank of a category depends only upon how many categories are ahead of it in respect to the quantity of the characteristic under comparison and not upon how many classes follow it. The differences of the ordinal numbers do not speak about the absolute differences in the amount of the characteristic the objects posses. A secures the first position by aggregating a total of 520 in the examination, B coming second with a total of 515, and C places in the third spot scoring 450. Although the rank differences between A and B, and between B and C are the same, the differences in absolute term of marks obtained in the examination are different. All we can say from an ordinal scale is that A is best in the group, B the second best (only one has a better record than him); and, C is the third best, coming after A and B, and, so on, with other orders. A large variety of measurements in psychology, sociology and other sciences yield only ordinal data. When a measuring instrument produces ordinal numbers, it is called ordinal scaling.

At the ordinal level of measurement, numbers are assigned to objects or events not only to categorize them, but also to indicate a 'greater than' or 'less than' relationship. The scale has no absolute zero point and there are unequal distances between the scale values. Numbers assigned at the ordinal level provide more

information than at the nominal level because they also establish an ordering of the objects or events. For example, various television programmes may be categorized according to popularity and assigned rank 1 to most popular and 5 to least popular ones. The programmes can then be ordered according to popularity, but it can not be ascertained how much more popular one programme is over another, since the numbers do not indicate equal distances among the scale values.

Ordinal level measurement allows objects or events to be ordered by degree on the basis of possession of some characteristic that can be abstracted and measured quantitatively.

1.5.3 Interval Scale

When a scale has all the properties of ordinal and ordered metric scale and when we have additional information about how large the distances (intervals) between any two stimuli are, we have achieved a more powerful measurement, stronger than ordinality. In such a device, a measurement has been achieved in the sense of an interval scale.

To understand the distance function it is necessary to assign 'real numbers' to all pairs of elements in the ordered set. That is, the position of an element or object in the scale is specified by a real number so that such numbers constitute points on an arithmetic scale with a common and constant unit of measurement. The ratio of any two intervals indicated by the real numbers, however, is independent of the unit of measurement and there is a lowest end point, the zero point. Thus, the ratio of two intervals 32cm and 40cm, and 1 00cm and 140cm is 1:5, which has no unit. If a constant, say 10cm is added to each of the interval points, i.e., if the new intervals are 42cm - 50cm and 110cm - 150cm respectively, the ratio between the two intervals remains the same.

Interval measurement should be used with due caution, especially when comparing differences between two or more attributes. Comparisons are meaningful when the origin, zero, for both the scales is the same and the units of measurements are identical. Measuring temperature with a thermometer, measuring the time from a selected starting moment, measuring the altitude from mean sea level are all done with interval scales.

In interval scale has all the properties of a nominal scale (equivalence relation), an ordinal scale (greater then or transitivity relation) and an ordered metric scale (transitively relation in respect to distance between classes). In addition, this scale is able to specify the ratio of any two intervals. It is, therefore, to be regarded as more powerful measurement compared to the three others already discussed.

Interval scale puts objects or events into a continuum with such units that measure intervals of equal distance. The starting point zero of the scale is arbitrarily chosen.

1.5.4 Ratio Scale

Ratio scale provides the most powerful measurement for it satisfies not only all the characteristics of an interval scale, but has also an additional and vital characteristic - that it has an invariant or absolute zero. This invariant zero point introduces a new dimension to mathematical operations. Not only is the ratio of intervals of two classes independent of origin and unit of measurement, the numbers associated with scale points can be expressed as ratios independent of

the unit of measurement. In an interval scale (the origin is arbitrary) we could not say that a score of 60 obtained by A is twice as large as the score of 30 obtained by B, for the simple reason that the zero point was chosen arbitrarily. If 5 is added to each score, the translated score of A will now read as 65 and is not twice as large as B's translated score 35, although the differences between the two sets of scores (60 - 30 and (65 - 35) remain the same. This shows that if we have an absolute zero in a scale, the scale values can no longer be translated, but can only be multiplied (or, divided) by a scalar. In the example cited in connection with measure of authority, *we* could not possibly associate a 'zero authority' with any regular staff member of a university.

Ratio scale is most commonly encountered in physical sciences. The weights of two objects whether measured in pounds or kilograms always yields the same ratio. The same is true of the length of two objects, or time taken by two individuals to complete a specific job.

A measurement will be said to be in ratio scale if it is possible to operationally attain the four relations: (i) equivalence (ii) greater than (iii) known ratio of any two intervals, and (iv) known ratio of any two real numbers associated with any two points on the scale.

Ratio scale places objects or events on a continuum, which has a rigidly defined starting point (real zero), such that the variable quantities (real numbers) can be expressed in terms of ratio of another number.

Table 1.1 : Examples of Questions in Each of the Five Basic Scale

S.No.	ScaleType	Examples
1.	Nominal Scale	Do you like the quality of health care services provided by your health centre? Yes-1 No-2
2.	Ordinal Scale	If you are asked to rate the quality of health care services provided by the primary health care centre, how will you rank it? Excellent-1 Very Good -2 Good-3 Average-4 Poor-5
3.	Interval Scale	In the past twelve months, how many times have you gone for check up to the primary health centre? <5 times – 1 6-10 times – 2 11-15 times – 3 >16 times – 4
4.	Ratio Scale	In the past twelve months, how much of money have you spent on health care? Amount of money spent……

1.6 ADMISSIBLE STATISTICAL TESTS FOR MEASUREMENT

The application of various statistical tests for different categories of measurement scales are discussed below.

Statistical Tests for Nominal Scale: Since the symbols or labels attached to any category are arbitrary and can be interchanged without altering essential information contained in the scale, the only kind of descriptive statistics that can be used are those, which would not be affected or altered by such interchange. They are crude mode, proportion and frequency. The nominal scale data, however, can be used for testing of hypothesis relating to distribution of events among the classes. Chi-square test, Contingency Coefficient, and certain other tests based on binomial expansion can be used for the purpose.

Statistical Tests for Ordinal Scale: Median is the most appropriate measure of central tendency of the scores that are in an ordinal scale. Obviously, quartile deviation is the measure of dispersion for such data. There are a number of non-parametric tests to test a hypothesis with scores in an ordinal scale - runs test, sign test, median test, Mann Whitney U- test, etc. These tests are often referred to as 'order statistics' or 'ranking statistics'. Interrelations can be computed from rankings of two sets of observations on the same group of individuals. Spearman's Rank Difference, or Kendall Rank Correlation coefficients are appropriate for such situations.

For applying tests to measurements on an ordinal scale, we make an assumption that the observations are drawn from a distribution, which is essentially continuous. Such assumptions are also made for all parametric tests. A continuous variate is one that is not restricted to having only isolated values. Given a certain limit (interval between two classes), we can have any number of values inserted in between. With an increase in the number of observations, more and more of these values are likely to be represented. It will suffice, at this point, to remind the readers that very often the crudeness of our measuring devices obscures the underlying continuity that may exist. The classification of respondents with respect to an attitude statement into categories strongly agree, agree, neutral, disagree, strongly disagree *essentially* presumes the presence of a continuum. If a variate is truly continuous and if the instrument for measuring the property in question is sensitive enough, then the probability of obtaining a tied observation is extremely small.

Statistical Tests for Interval Scale: The interval scale preserves both the ordering of objects and the relative differences between them, even though the numbers associated with the position of the object may be changed, following a regular system. A set of observations will be scalable by interval scale if the data permits a linear transformation. That is, if the equation $y = a + bx$, where a and b are two positive constants, satisfies a set of real numbers, the numbers are said to be in an interval scale.

All the common parametric tests - arithmetic mean, median, standard deviation, product-moment correlation, etc., are applicable to data that follow an interval scale. Non-parametric tests for statistical significance like Z, t, F are also applicable to data in interval scale.

Statistical Tests for Ratio Scale: Since the values in a ratio scale are real numbers with a true zero (no upper limit) and only the unit of measurement is arbitrary, the ratios between two numbers and intervals preserve all the information contained in the scale even if these true numbers are multiplied by a true positive constant. Any statistical test, parametric or non-parametric, is usable when a ratio scale is used, such statistical tools as geometric mean and coefficient of variation, which require knowledge of true scores, can be used with observations that are in ratio scale.

Table 1.2: Analysis Method in Various Scales

Scale	Basic Operation	Measure of Central Tendency	Other Appropriate Statistics
Nominal	Puts objects into classes, i.e., male/ female, marital status, occupation	Mode	Chi-square
Ordinal	Determination of greater or less, i.e., preference, level of education achieved	Median	Rank Order Correlation, Man Whitney U-test
Interval	Determination of equality of intervals or differences, i.e., temperature	Arithmetic Mean	Correlation Coefficient
Ratio	Determination of equality of ratios, i.e., weight, income, number of visits	Geometric Mean, Harmonic Mean	Coefficient of Variation

Source: John boyce, Marketing Research, Mc Graw Hill, Australia, 2005.

1.7 CRITERIA FOR JUDGING THE MEASURING INSTRUMENTS

A measurement, too, must satisfy certain criteria. The most important criteria to be used in evaluating a measurement tool are described below.

i) **Unidimensionality:** This means the scale should measure one characteristic at a time, e.g., the ruler should measure length**,** not temperature.

ii) **Linearity:** This means that a scale should follow the straight-line model. Some scoring system should be devised, preferably one based on inter-changeable units. In a ruler an inch is an inch whether it lies at one end of the ruler or at the other, but in altitude scales such interchangeability cannot be ensured. In such cases, ranking is preferable.

iii) **Validity:** This refers to the ability of a scale to measure what it is supposed to measure.

iv) **Reliability:** This is an attribute of consistency. A scale should give consistent results.

v) **Accuracy and Precision:** A tool should give an accurate and precise measure of what we want to measure.

vi) **Simplicity:** A scale should be as simple as possible; an elaborate, complicated, and over - refined scale may become unduly cumbersome, costly or even useless.

vii) **Practicability:** This is concerned with wide range of factors like cost effectiveness, convenience and interpretability. Some trade off is usually needed between an 'ideal' tool and, that which the budget can afford. The benefit to be derived should be commensurate with the cost incurred.

The tool should be easily administrable; it should contain proper instructions; it should be easily understandable and conveniently arranged for easy completion. In order to enable others to interpret the results of a test, there is need for such aids as a statement of its function, its construction procedure and guides for interpreting the result.

1.8 SOURCES OF ERRORS IN MEASUREMENT

Measurement should be precise and unambiguous in an ideal research study. This objective, however, is often not met with in entirety. As such, the researcher must be aware about the sources of error in measurement. The following are the possible sources of error in measurement.

a) **Respondent:** At times the respondent may be reluctant to express strong negative feelings, or, it is just possible that he may have very little knowledge but may not admit his ignorance. All this reluctance is likely to result in an interview of guesses. Transient factors like fatigue, boredom, anxiety, etc., may limit the ability of the respondent to respond accurately and fully.

b) **Situation:** Situational factors may also come in the way of correct measurement. Any condition which places a strain on interviews can have serious effects on the interviewer- respondent rapport. For instance, if someone else is present, he can distort responses by joining in or merely by being present. If the respondent feels that anonymity is not assured, he may be reluctant to express certain feelings.

c) **Measurer:** The interviewer can distort responses by rewording or reordering questions. His behaviour, style and looks may encourage or discourage certain replies from respondents. Careless mechanical processing may distort the findings. Errors may also creep in because of incorrect coding, faulty tabulation and /or statistical calculations, particularly in the data analysis stage.

d) **Instrument:** Error may arise because of the defective measuring instrument. The use of complex words, beyond the comprehension of the respondent, ambiguous meanings, poor printing, inadequate space for replies, response choice omissions, etc., are a few things that make the measuring instrument defective, and may result in measurement errors. Another type of instrument deficiency is the poor sampling of the universe of items of concern.

The researcher must know that correct measurement depends on successfully meeting all of the problems listed above. He must, to the extent possible, try to eliminate, neutralize, or, otherwise deal with all the possible sources of error so that the final results may not be contaminated.

In the above sections, you read about various kinds of measurement and statistical tests to be used in measurement. Now answer the questions given in *Check Your Progress 2.*

Check Your Progress 2

Note: a) Write your answer in about 50 words.

b) Check your answer with possible answers given at the end of the unit

1) What is interval scale?

.....

.....

.....

.....

.....

2) Write three important criteria for judging the measuring instruments.

.....

.....

.....

.....

.....

1.9 LET US SUM UP

In this unit, we discussed the meaning, concept, postulates and levels of measurement. Measurement is an undetachable part of any science, natural, or social. Any science aims to obtain a specific and accurate measurement of the events, of the characteristics of the different units of a phenomenon and of the inter-relationship between the units. Measurement is assigning numbers to objects or events according to rules.

The theory of measurement in social sciences really consists of a system of distinct theories each concerned with a distinct level of measurement. A set of data often will satisfy some of the levels, but not others. It is, therefore, necessary to understand the basic nature of data before applying a particular level of measurement. The choice of an appropriate statistical model for analysis is largely dependent on this level of measurement achieved.

Sometimes discrimination is made between scales on the basis of power of measurement. The scales like nominal, ordinal, and ordered metric scales are less powerful because they do not make use of all the information contained in

the data. As such, such measurements are referred to as scales by many of its users. The more powerful measurements like interval and ratio scales, on the other hand, make full use of all information contained in a set of observations, and, therefore, are labelled as measurements; whereas, the former scales refer mostly to qualitative aspects of measurement, the latter ones deal with quantitative measurements.

1.10 KEYWORDS

Measurement : The process of assigning symbols/numbers to dimensions of phenomena in order to characterize the status of a phenomena as precisely as possible.

Scale : A device to measure something. Scaling technique is used in ordering a series of items along some sort of continuum. In short, they are methods of turning a series of qualitative facts into a quantitative series.

Validity : Refers to the ability of a scale to measure what it is supposed to measure.

Reliability : An attribute of consistency. A scale should give consistent results.

1.11 REFERENCES AND SELECTED READINGS

Black, A.and Champion, J (1976). Methods and Issues in Social research. John Ditecy and Sons, New York.

Festinger L.and Katn D. (1953). Research Methods in Behavioural Sciences. Holt, Rinehart and Winston Inc., New York.

Goode W.J.and Hatt P.K. (1981). Methods in Social Research. McGraw- Hill Book Company, Singapore.

Hansraj (1990). Theory and Practice in Social Research. Asia Publishing House, Culcutta.

Kerlinger F.N. (1973). Foundstions of Behavioural Research. Holt, Rinebart and Winston Inc., New York.

Kothari C.R. (1996). Research Methodology: Methods and Techniques. Wishwa prakashan, New Delhi.

Mulay Sumati and Sabarathanam V.E. (1980). Research Methods in Extension Education. Manasayan, New Delhi.

Siegel S. (1956). Nonparametric Statistics for the Behavioural Sciences. McGraw-Hill Book Company Inc., New York.

Young P.V. (1996). Scientific Social Surveys and Research. Prentice –Hall of India Pvt. Ltd., New Delhi.

1.12 CHECK YOUR PROGRESS – POSSIBLE ANSWERS

Check Your Progress 1

1) Measurement consists of identifying the values which may be assumed by some variable, and representing these values by some numerical notation. The numerical notation is systematically and consistently assigned, that is, it is assigned according to some set of rules.

2) The three postulates basic to measurement are:

   a) Either (a=b) or (a =/= b), but not both. We must be able to assert either that one object is the same in a characteristic as another, or it is not the same.

   b) If (a=b) and (b=c) then (a=c). This postulate enables a researcher to establish the quality of set members, on a characteristic by comparing objects

   c) If (a>b) and (b>c) then (a>c). Most measurement in psychology and education depends on this postulate.

Check Your Progress 2

1) When a scale has all the properties of ordinal and ordered metric scale, and, when we have additional information about how large the distances (intervals) between any two stimuli are, we have achieved a more powerful measurement, stronger than ordinality. In such a device, a measurement has been achieved in the sense of an interval scale.

2) The most important criteria to be used in evaluating a measurement tool are unidimensionality, reliability and validity.

UNIT 2 SCALES AND TESTS

Structure

2.1 INTRODUCTION

Measurement plays an important role in any development research including urban development research. This is especially true when the measurement concepts are complex and when we do not possess the standardised measurement tools. To overcome this, social science researchers develop self reporting measuring instruments to assess people's knowledge, opinion, perceptions, attitudes etc., on urban development programmes. Technically speaking these reporting measurement instruments are popularly called as scales and tests. The scales and tests are the most popular methods of observation and data collection in behavioural sciences and more particularly in development studies.

After studying this unit, you should be able to:

- discuss the meaning and applicability of scales and tests.
- describe the important types of scales and tests.
- explain the test construction methodology.

2.2 SCALES: MEANING AND TECHNIQUES

Scales are also popularly called as rating scales. 'Rating' is a term applied to an expression of opinion or judgment regarding some situation, object or character. A rating scale is a psychological measuring instrument that requires the rater or observer to assign the rated object to categories or continuum that have numerical assigned to them.

The rating scale is very useful device in assessing quality – especially when quality is difficult to measure objectively in the programmes of development.

Example: How good is the Jawaharlal Nehru National Urban Renewal Mission Programme?

The above question can be hardly answered objectively. In this context, rating scales measure or order entities with respect to quantitative attributes or traits of the above programme. Certain rating scales permit estimation of magnitudes of the programme on a continuum, while other methods provide only for relative ordering of the entities.

2.2.1 Important Scaling Techniques

The two important scaling techniques are:

- Comparative scaling
- Non-comparative scaling

With comparative scaling, the items are directly compared with each other.

Example : Do you prefer weekly or daily payment of wages under National Urban Employment Programme?

Here, the respondent compares 'weekly' or 'daily' payments and gives his choice to confer his / her opinion or views.

In non-comparative scaling each item is scaled independently of the others.

Example : How do you feel about daily payment of wages under National Urban Employment Programme?

Unlike the above, in this case there is no comparison and the respondent has to give his / her opinion on 'daily' payment of wages under National Urban Employment Programme.

Activity 1

Visit a nearby any development department and enquire about scales and tests that they are using in measurement of outcomes of development programmes. Write your observations.

.....

.....

.....

.....

.....

.....

Check Your Progress 1

Note: a) Use the spaces given below for your answers.

b) Check your answer with those given at the end of the unit.

1) What do you mean by tests and scales in the context of development research?

.....

.....

.....

.....

2) What do you mean by a rating scale?

.....

.....

.....

.....

.....

.....

3) Write the difference between comparative and non-comparative scaling techniques.

.....

.....

.....

.....

.....

.....

2.3 TYPES OF RATING SCALES

The rating scale involves qualitative description of a limited number of aspects of a thing or of traits of a person. These ratings may be in such forms as :

1) like – dislike (2 categories)
2) above average – average – below average (3 categories)
3) like very much – like somewhat – neutral – dislike somewhat – dislike very much (5 categories)
   - excellent – good- average – below average - poor (5 categories)
   - always – often – occasionally – rarely – never (5 categories)
   - very strongly agree – strongly agree – agree – neutral – disagree- strongly disagree - very strongly disagree (7 categories)

There is no specific rule whether to use a two points scale, three point scale or with more points. In practice, odd number scales with three, five or seven points are popularly used for the simple reason that more points on a scale provide an opportunity for greater sensitivity of measurement.

Some of the important types of rating scales are:

i) The graphic rating scale

ii) The descriptive rating scale

iii) The numerical rating scale

iv) The itemized rating scale

2.3.1 Graphic Rating Scale

Graphic rating scale (also called continuous rating scale) is quite simple and is commonly used in practice. Graphic scale makes use of continuum along which the rater places a mark (v) on a line to indicate his / her rating with respect to certain characteristics. The line is usually labeled at each end. There are sometimes a series of numbers, called scale points under the line.

Example : Teachers often describes students personality during evaluation. The following is an example of graphic rating scale on personality rating of students.

Directions: Please give your frank opinion concerning the students' characteristics by putting an 'X' at the point along the scale that best describes the student.

a) Cooperation .....

(Consider willingness Obstructive Always willing to work with others)

b) Emotional Stability .....

(Consider reactions in stress Unstable Well balanced)

Advantages

- One of the major advantages of graphic rating scale is that they are relatively easy to use.
- Graphic scale provides opportunity for a given discrimination as that of which the mater is capable and the fineness of scoring can be as great as desired.

Limitations

- Respondents may check at almost any position along the continuum which increases the difficulty of analysis.
- Meanings of the terms like obstructive, always willing, etc., may depend upon respondent's frame of reference.

To overcome the limitations, several other rating variants (example: boxes replacing line) may be used.

2.3.2 Descriptive Scale

It is a variation of the graphic rating scale. It provides descriptive works or phrases that indicate the degree to which individual is believed to possess certain characteristic.

Example of a behavioural rating scale:

Direction:- For each of the items listed in this scale, place a 'X' in one of the columns to indicate the extent to which you feel that the student possesses the particular characteristic kind of behavior.

Item	Never	Seldom	Sometime	Usually	Always
Listens to others opinion					
Accepts constructive criticism					

Example: Suppose you would like to collect data on liking of information sources on development programmes from your study respondents. The following is an example of five- points graphic rating scale on liking of information sources.

How do you like the following information sources for obtaining information on development programmes?

Information source	Liking of information source				
	Like very much	Like Somewhat	Neutral	Dislike somewhat	Dislike very much
Institutional Sources					
BDO					
VDO					
Extension Personnel					
Any other (Please specify)					
Non Institutional Sources					
Other Beneficiaries					
Key Personnel					
Own Family Members					
Any other (Please specify)					
Mass Media Sources					
Radio					
TV					

2.3.3 Numerical Rating Scale

The numerical scale makes use of numbers to indicate the extent to which an individual is believed to possess certain characteristic or kinds of behaviour.

Example of a behaviour rating scale:

Directions : As you rate the student on each of the following items, circle 1 for inferior, 2 for below average 3 for average, 4 for above average and 5 for superior.

1) Cooperates with students 1 2 3 4 5

2) Cooperates with teachers 1 2 3 4 5

3) Maintains an attractive appearance 1 2 3 4 5

2.3.4 Itemized Rating Scale

It is also referred to as specific category scale. In this type of scale, the respondent selects or picks the one that best characterizes the behaviour or characteristic of the object being rated. Suppose a teacher's classroom behaviour is being rated. The characteristics rated say may be alertness or imaginativeness.

A category item might be 'how alert is he / she?' (Check one).

a) very alert

b) Alert

c) Not alert

d) Not at all alert

A slightly different category item might be 'how imaginative is he /she?' (check one)

a) Extremely imaginative

b) Very imaginative

c) Imaginative

d) Unimaginative

e) Very unimaginative

f) Extremely unimaginative

2.4 USES AND GUIDELINES FOR CONSTRUCTION OF RATING SCALES

2.4.1 Uses of Rating Scale

Rating scales are most commonly used instruments for making appraisals. Typically, they direct attention to a number of aspects or traits of the thing to be rated and provide a scale for assigning values to each of the aspects of characteristics of a person or phenomenon through the use of a series of numbers, qualitative terms, and named attributes of verbal descriptions.

Rating scales have been successfully used in :

- Teacher rating – for selection, evaluation and prediction.
- Personality rating – for various purposes.
- Testing the validity of many objective instruments like inventories of personality.
- School appraisal – including appraisal of courses, practices and programmes.

2.4.2 Guidelines in Construction of Rating Scales

i) Rating scales include three factors:

- The subjects or phenomena to be rated
- The continuum along which they will be rated
- The judges who will do the rating

The subjects or phenomena to be rated are usually a limited number of aspects of a thing, or of traits of a person. Therefore, only the most significant aspects for the purpose of the study should be chosen.

ii) A rating scale may have as many divisions as can be readily distinguished by the rates. Practically most scales have no more than 7 divisions. However, usually they contain five divisions. By numbering each division in sequence the description can be converted into arithmetic values for averaging and for further statistical application.

iii) The rating scale is composed of two parts:

- an instruction which names the subject and defines the continuum and
- a scale which defines the points to be used in rating

iv) Usually we can arrange the rating scales in four ways:

a) On a straight line eg:

_____

- Very good Good Average Poor Very poor

b) Ratings be marked in a column at the right with an instruction to encircle / underline the response.

- IGNOU course material was: Very good/ Good/Average/ Poor/Very poor

c) The scale can run down the page and look much like a checklist.

Example: For me the concepts presented in IGNOU course material;

- was very difficult to understand
- was difficult to understand
- was reasonably understandable
- was clearly understandable
- contained nothing new.

d) The scale may call for ranking like which course of the IGNOU's programme helped you most? Rank them by starting with one for the best unto least

Example: Suppose there are four courses in the programme viz., a, b, c, d and out of the four 'c' helped you most

c = 1 'c' is ranked highest.

b = 3

d = 2

a = 4

The other way is rank these courses in order of merit – starting with 1 for best

1c 3b 2d 4a

This type of ranking is a higher form of rating whereby individuals or phenomena are arranged in order of merit (i.e.) they are given position determined by their relation to the others in the group, not by certain predetermined standards. But is cannot be used when large numbers are concerned.

The investigator must arrange his items in any or all of the above forms according to the nature of the item and its purpose.

v) Anyone can serve as a rater where non-technical opinions, likes, dislikes and matters of easy observation are to be rated. But only well-informed and experienced persons should be selected for rating where technical competence is required.

vi) Pooled judgements increase the reliability of any rating scale. Employ several judges, depending on the rating situation to obtain desirable reliability. Individual ratings when combine into final rating give a safer assessment.

2.5 RATING ERRORS

Rating scales are subjected not only to errors inherent in their design but also to errors that are related to the way in which raters have marked the scale. Some of them are discussed below for your understanding.

Halo Effect : This is an error that occurs when a rater tends to rate an individual high or low on several characteristics because of a general impression that the rater has towards subject whom (s)he is rating. For example if a teacher assesses the quality of all essay test questions higher / lower than they should be based on the answer of the first question.

Personal Bias : It is an error that is made when a rater is prejudiced with regard to a certain group and tends to rate individual who belong to that group too higher or too low.

Logical Error : It results when the rater does not fully understand the term used in the rating scale.

Error of Central Tendency : It occurs when a rater does not have enough information about the individual to be rated and tends to rate the person as average. The rater feels that average ratings are safer to make than extreme ratings, because errors that are as a result of guessing will perhaps be smaller.

Generosity Error : It occurs because sometimes raters are very reluctant to give any ratings at the lower end of the scale. They tend, therefore, to rate every one as average or above average on all characteristics.

Error of Severity : This is an important source of constant error. It is a general tendency to rate all individuals too low on all characteristics.

Error of Leniency: This is the opposite to error of severity. The general tendency is to rate too high. A good fellow who likes every body and the likeness is reflected while rating.

It is impossible to eliminate these kinds of errors, but certain steps can be taken to minimize them. One suggested step is to inform the rater, either orally or in writing, about the possible source of error in rating and then urge him/her to be as objective as possible. Another suggestion is to construct the rating scale in such a way as to lower the possibility of error on the rater's part.

2.5.1 How to Overcome the Errors in Rating Scales?

There will be a certain amount of measurement error which results from the structural characteristics of rating scales. By adhering to the following rules of construction, however, it is possible to minimise these kinds of errors.

i) Provide direction to assist the rater in the use of the scale

ii) Use only enough items to obtain the information needed

iii) Clearly define each characteristic on which an individual is rated

iv) Use only those trails or characteristics in rating scales that can be observed readily

v) Define clearly the degree or different levels of gradation that are to be used in the scale (At least 4 or 5 gradations are recommended)

vi) Provide ample space between the items so that the descriptive phrases and the rating lines are not crowded.

vii) Position the ' average' or 'neutral' phrase at the centre of the rating line

viii) Try to avoid the use of phrases that are so extreme at the end positions of the scale that the raters will tend to avoid making them.

Check Your Progress 2

Note: a) Use the spaces given below for your answers.

b) Check your answer with those given at the end of the unit.

1) Write the examples for the following categorization of rating scales

3 divisions: .....

.....

5 divisions: .....

.....

2) Name the four important types of rating scales.

.....

.....

.....

.....

.....

.....

3) Name the common rating errors.

.....

.....

.....

.....

.....

.....

2.6 TESTS

The tests are frequently used in education and psychological researches and more recently in development studies to measure the achievement and personality tract of various categories of respondents.

According to the dictionary 'test' is defined as a series of questions on the basis of which some information is sought. According to Bean (1953) a test is "an organized succession of stimuli designed to measure quantitatively or to evaluate qualitatively some mental process, trait or characteristics".

The two types of tests popularly used are:

- Objective Tests
- Teacher-Made Tests

2.6.1 Objective Tests

There are various types of objective tests viz.,

i) Achievement Test

ii) Diagnostic Test

iii) Intelligent Test

iv) Aptitude Test

v) Personality Test

Achievement Test

Achievement or proficiency test is one, which measures the extent to which a person has acquired certain information or proficiency as a function of instruction or training. The achievement test is used in order to assess the achievement of a person in certain areas. For example a teacher can conduct a test to assess the student achievement in mathematics.

Diagnostic Test: This test intends to assess the strength and weakness of a person in one or more than one areas of his/her activities. It is conducted with a view to carry out interventions in weak areas. It also makes an enquiry about the weak areas of the respondent may be a student, employee or worker.

Intelligence Test: The intelligence test is prepared to measure the intelligence of a person. This test is used by the psychologists and educationists to measure the intelligence of students. Intelligence is measured in terms of intelligence quotient (IQ). The intelligence quotient is the ratio of the mental age to the chronological age. This ratio is multiplied by hundred for obtaining an integral value of quotient. The IQ gives an index of ability.

$$IQ = \frac{\text{Mental Age}}{\text{Chronological Age}} \times 100$$

Aptitude Test: An aptitude is a person's conditions, a pattern of traits and demand to be indicative of his potentialities. According to Freeman (1962) aptitude is a combination of characteristics indicative of an individual's capacity to acquire

some specific knowledge. The aptitude test is used in various competitive examinations such as banking, insurance and management organizations.

Personality Test: Personality test intends to measure the personality traits of the individuals. Some of the personality traits are cooperation, discipline, leadership, personal appearance, punctuality, patriotism, confidence, team spirit, etc.

2.6.2 Teacher Made Tests

The teacher made test are also called the non-standardized tests. Teachers have the obligation to provide the best possible instructions to the students. In order to judge the performance of the students, teachers assess the performances of the students from time to time. These are useful:

- To assess the extent and degree of student progress.
- To ascertain individual strength and weaknesses.
- To motivate the students.
- To provide immediate feedback.
- To provide continuous evaluation.

2.7 TYPES OF OBJECTIVE TEST QUESTIONS

Various types of objective test questions used in research studies are as follows:

i) True/ False

ii) Multiple choice

iii) Fill in the blanks

iv) Matching

v) Completion

True/ False Test: The true/false or yes/no or right/wrong type of tests are most commonly used. It is used to determine the respondent's ability to recall the facts.

Example: India follows mixed economy system True ☐ False ☐

Multiple Choices Test: In the multiple choice test, the respondents are given multiple option of a question. Here the choices or the alternatives should be written in such a way that it may not create ambiguity in the mind of respondents. The multiple choices may contain more than one valid choice.

Example: Which is not the indicator of Human Development Index?

i) Life Expectancy ☐

ii) Literacy ☐

iii) Per capita income ☐

iv) Poverty level ☐

Fill in the Blanks: In the fill in the blanks question, the respondent is asked to supply correct answer to the blank left in the statement. However, while formulation of fill-in-the-blanks test, too many blanks should not be provided which will create confusion in the minds of respondents. One example of fill in the blanks is given below:

Example: The JNNURM started in the year ————————

Matching: In the matching test, there are two columns right and left. The items on the left column are to be paired with items on the right column. Items on the left which constitute a set of related streams called premises and items on the right are is called cassette options or responses of the items.

Example: Match column A with B

	Column A	**Column B**
1.	Mouse	Bicycle
2.	Wheel	Computer
3.	Remote	Television

Completion: Completion test is one which is presented in the form of an incomplete statement. This is also called supply type of test.

Example: There is a famous saying 'all is well that ends well'

2.8 TEST CONSTRUCTION

Construction of a test is very important in order to arrive at accuracy in the result. The various steps adopted in test construction are as follows:

i) Planning of the test

ii) Writing items of the test

iii) Preliminary administration of the test

iv) Item Analysis

v) Establishing Reliability of the test

vi) Establishing Validity of the test

vii) Preparation of norms for the final test

viii) Preparation of manual and reproduction of the test

Planning of the Test: At the outset, the tests are to be carefully planned. While planning, the test constructor has to take into consideration, the general and specific objectives of the test in clear terms and the nature of the content or items to be included.

Writing Items of the Test: Item writing is one of the very important aspects of test construction. Although there is no set rule for item writing, yet lot depends on the ingenuity, intuition, experience, knowledge, practice and imagination of the test constructor. It can be said that writing test item is essentially an art.

Preliminary Administration: After the items are written, it is better to tryout them. It will help the test constructor to find out the weaknesses and inadequacies

in test items. It helps in finding out the appropriate length and time limit of the tests administration.

Item Analysis

After the items are being written, they are carefully analysed and reviewed. In item analysis, items are validated and suited for the purpose. The objectives of item analysis are as follow:

- Helps to indicate the difficulty level of the item such as which is more difficult, moderately difficult or easy.
- Help to provide indication regarding the ability of the item to discriminate between inferior and superior item.

Two common indices used in item analysis are:

- Difficulty Index
- Discrimination Index

Difficulty Index: The difficulty index indicates how difficult an item is? The difficulty value of an item indicates the proportion or percentage of candidates who have given correct answer. This proportion or percentage is called Item Difficulty Index. The formula used for the calculation of item difficulty index of each item is given below.

$$IDI = \frac{R}{N}$$

IDI = Item difficulty index

R = Number of right responses

N = Total number of candidates attempting that item.

Besides this method which takes into consideration all the examiness, there is also another method which can determine the index on the basis of only a portion of the examinee. The formula is:

$$IDI = \frac{R_U + R_L}{N_U + N_L}$$

Where

IDI = Item Difficultly Index

R_U = Right responses in the Upper group

R_L = Right responses in the Lower group.

N_U = Number of examinees in Upper group

N_L = Number of examinees in the lower group

For example if there are 200 examinees of a test, N_U=50 and N_C=50. Out of these groups R_U=25 and R_L=25

Then:

$$IDI = \frac{25+25}{50+50} = \frac{50}{100} = 0.50$$

Discrimination Index: The discrimination index distinguishes between the well-informed examinees to that of the less-informed examinee. It is the degree to which the single item separates the superior from the inferior individuals in the trait or group of trait being measured.

$$DI = \frac{R_H - R_L}{N}$$

Where:

DI = Discrimination Index

R_H = Number of rights in the higher ability group

R_L = Total Number of rights in the lower ability group.

N = Total number of examinees in either of the group.

Let us explain this with the help of an example. After getting the responses from 100 examinees they were divided into upper group (25%) and lower group(25%).Suppose in a particular item, right responses in the upper group is 80 and right responses in the lower group is 60, then the item discrimination index is:

$$DI = \frac{80-60}{100} = 0.20$$

This value of 0.20 clearly states that item has negligible discriminatory power. Such items are usually dropped or suitably modified.

The factor which influences item difficulty and item discrimination index are:

- The ambiguity and complexity of items in a test item may lower the difficulty index value of the item.
- Previous learning experiences may be helpful in deciding the item difficulty index or discrimination index.
- It depends on the ability of the test constructor to effectively frame the distracters. They must be appealing to those who do not know the correct answer.

Establishing Test Reliability: Finally test is administered to find out their reliability. Reliability is the degree to which a test measure whatever it actually measures.

Validity of the Test: Validity means what the test measures. There are various kinds of validity viz., criterion- referenced validity and construct validity.

Norms of the Test: Norms are set to meaningfully interpret the scores obtained on the test. The common types of norms are the age norms, the grade norms, the percentile norms, etc. For example a test constructed for class-v student should not be administered over the class-viii student.

Preparation of the Manual and Reproduction Test: Finally, the constructors has to produce a manual which will give a clear cut instruction regarding the procedures of the test administration, the scoring methods, time limit, etc.

Check Your Progress 3

Note: a) Use the spaces given below for your answers.

b) Check your answer with those given at the end of the unit.

1) What do you mean by test ?

.....

.....

.....

.....

.....

2) Name the various types of objective tests.

.....

.....

.....

.....

.....

3) Name the various types of objective test questions.

.....

.....

.....

.....

.....

4) Write the objectives of item analysis.

.....

.....

.....

.....

.....

2.9 LET US SUM UP

In this unit we started by discussing the meaning of rating scales and understood that rating scale is very useful device in assessing quality – especially when quality is difficult to measure objectively in the programmes of development. Then we examined the two important scaling techniques viz., comparative and non-comparative scaling techniques. We also described the four types of rating scales viz., graphic, descriptive, numerical and itemized rating scales with

examples. Later we discussed the important rating errors. In the second part of the unit we have discussed the concept and two types of tests viz., objective and teacher -made tests. Later various types of objective tests viz., achievement, diagnostic, intelligent and aptitude tests were discussed. At the end we discussed the test construction methodology.

2.10 KEYWORDS

Scales and Tests : They are the self reporting measuring instruments to assess people's knowledge, opinion, perceptions, attitudes etc.

Rating Scale : Is a psychological measuring instrument that requires the rater or observer to assign the rated object to categories or continuum that have numerical assigned to them.

Index : Indexes are similar to scales except multiple indicators of a variable are combined into a single measure.

Graphic Rating Scale : Graphic scale makes use of continuum along which the rater places a mark (v) on a line to indicate his / her rating with respect to certain characteristics.

Descriptive Scale : It provides descriptive works or phrases that indicate the degree to which and individual is believed to possess certain characteristic.

Numerical Rating Scale : It makes use of numbers to indicate the extent to which an individual is believed to possess certain characteristic or kinds of behaviour.

Itemized Rating Scale : In this type of scale, the respondent selects or picks the one that best characterizes the behaviour or characteristic of the object being rated.

Halo Effect : An error that occurs when a rater tends to rate an individual high or low because of a general impression that the rater has towards subject.

Error of Leniency : The general tendency to rate too high.

Test : Test is an an organized succession of stimuli designed to measure quantitatively or to evaluate qualitatively some mental process, trait or characteristics.

Achievement Test : Is one which measures the extent to which a person has acquired certain information or proficiency as a function of instruction or training.

Diagnostic Test : This test intends to assess the strength and weakness of a person in one or more than one areas of his/her activities.

Intelligence Quotient : It is the ratio of the mental age to the chronological age multiplied by hundred.

Difficulty Index : The difficulty index indicates how difficult an item is in the test.

Discrimination Index : The discrimination index distinguishes between the well- informed examinees to that of the less-informed examinee. It is the degree to which the single item separates the superior from the inferior individuals in the trait or group of trait being measured.

2.11 REFERENCES AND SELECTED READINGS

Kothari, C.R. (2004). Research Methodology – Methods and Techniques, Second Revised Edition, New Age International Publishers, New Delhi.

Patten M.L. (2005). Understanding Research Methods – An Overview of the Essentials. Pyrczak Publishing, USA.

Reddy, S.V., and Praveena, C. (1985). An Invitation to Research Methodology in Behavioural Sciences, EEI, Hyderabad.

2.12 CHECK YOUR PROGRESS – POSSIBLE ANSWERS

Check Your Progress 1

1) In the context of development research, scales and tests are the self reporting measuring instruments to assess people's knowledge, opinion, perceptions, attitudes etc.

2) A rating scale is a psychological measuring instrument that requires the rater or observer to assign the rated object to categories or continuum that have numerical assigned to them.

3) With comparative scaling, the items are directly compared with each other and in non-comparative scaling, each item is scaled independently of the others.

Check Your Progress 2

1) Three divisions : above average – average – below average

   Five divisions : like very much – like somewhat – neutral – dislike somewhat – dislike very much

2) The four important types of rating scales are: graphic; descriptive; numerical and itemized rating scales.

3) The common rating errors are: halo effect; personal bias; logical error; error of central tendency; generosity error; error of severity and error of leniency.

Check Your Progress 3

1) Test is an organized succession of stimuli designed to measure quantitatively or to evaluate qualitatively some mental process, trait or characteristics.

2) Various types of objective tests are : achievement test; diagnostic test; intelligent test; aptitude test and ; personality test.

3) Various types of objective test questions are : true/ false; multiple choice; fill in the blanks; matching and; completion.

4) The objectives of item analysis are : to indicate the difficulty level of the item such as which is more difficult, moderately difficult or easy and ; to provide indication regarding the ability of the item to discriminate between inferior and superior item.

UNIT 3 RELIABILITY AND VALIDITY

Structure

3.1 INTRODUCTION

Dear learners, in the first unit of this block, we discussed that measurement of social and psychological variables is a complex and demanding task. In urban development research, the common term for any type of measurement devise is 'instrument'. Thus the instrument could be a test, scale, questionnaire, interview schedule etc. An important question that is often addressed is what is the reliability and validity of the measuring instrument? Therefore, the purpose of this unit is to make you understand the concept of reliability and validity and their interrelationship in urban development research.

After studying this unit you should be able to:

- discuss the meaning of reliability and methods of determining the reliability of measuring instruments.
- describe the meaning of validity, approaches and types of validating measuring instruments.
- differentiate the interrelationship between reliability and validity of measuring instruments.

3.2 RELIABILITY

In the context of development research, one of the most important criterions for the quality of measurement is reliability of the measuring instrument. A reliable person for instance, is one whose behavior is consistent, dependable and predictable – what (s)he will do tomorrow and next week will be consistent with what (s)he does today and what (s)he has done last week. An unreliable person is one whose behavior is much more variable and one can say (s)he is inconsistent.

The inherent aspects and synonyms of reliability are:

- dependability
- stability

- consistency
- predictability
- accuracy
- equivalence

3.2.1 What is Reliability of Measuring Instrument?

Reliability means consistency with which the instrument yields similar results. Reliability concerns the ability of different researchers to make the same observations of a given phenomenon if and when the observation is conducted using the same method(s) and procedure(s).

Stability and Equivalence Aspects of Reliability

Stability and equivalence deserves special attention among different aspects of reliability,

- The *stability* aspect is concerned with securing consistent results with repeated measurements of the same researcher and with the same instrument. We usually determine the degree of stability by comparing the results of repeated measurements.
- The *equivalence* aspect considers how much error may get introduced by different investigators or different samples of the items being studied. A good way to test for the equivalence of measurements by two investigators is to compare their observations of the same events.

3.2.2 How to Improve Reliability?

The reliability of measuring instruments can be improved by two ways.

i) By standardizing the conditions under which the measurement takes place i.e. we must ensure that external sources of variation such as boredom, fatigue etc., are minimized to the extent possible to improve the *stability* aspect.

ii) By carefully designing directions for measurement with no variation from group to group, by using trained and motivated persons to conduct the research and also by broadening the sample of items used to improve *equivalence* aspect.

Check Your Progress 1

Note: a) Use the spaces given below for your answers.

b) Check your answers with those given at the end of the unit.

1) What is the common name for any type of measurement device?

.....

.....

.....

.....

.....

2) What do you mean by reliability?

.....

.....

.....

.....

.....

3) Write the synonyms for reliability.

.....

.....

.....

.....

.....

4) How can you improve the reliability of measuring instruments?

.....

.....

.....

.....

.....

3.3 METHODS OF DETERMINING THE RELIABILITY

The three basic methods for establishing the reliability of empirical measurements are:

i) Test - Retest Method

ii) Alternative Form Method / Equivalent Form / Parallel Form

iii) Split-Half Method

3.3.1 Test - Retest Method

One of the easiest ways to estimate the reliability of empirical measurements is by the test - retest method in which the same test is given to the same people after a period of time. Two weeks to one month is commonly considered to be a suitable interval for many psychological tests. The reliability is equal to the correlation between the scores on the same test obtained at two points in time. If one obtains the same results on the two administrations of the test, then the test – retest reliability coefficient will be 1.00. But, invariably, the correlation of measurements across time will be less than perfect. This occurs because of the instability of measures taken at multiple points in time. For example, anxiety, motivation and interest may be lower during the second administration of the test simply because the individual is already familiar with it.

Advantages

- This method can be used when only one form of test is available.
- Test – retest correlation represent a naturally appealing procedure.

Limitations

- Researchers are often able to obtain only a measure of a phenomenon at a single point in time.
- Expensive to conduct test and retest and some time impractical as well.
- Memory effects lead to magnified reliability estimates. If the time interval between two measurements is short, the respondents will remember their early responses and will appear more consistent than they actually are.
- Require a great deal of participation by the respondents and sincerity, devotion by the research worker. Because, behaviour changes and personal characteristics may likely to influence the re-test as they are changing from day to day.
- The validity process of re-measurement may intensify difference in momentary factors such as anxiety, motivation etc.
- The interpretation of test-retest correlation is not necessary straightforward. A low correlation may not indicate low reliability, may instead signify that the underlying theoretical concept itself has changed.

  Example: The attitude of a person towards functioning of a public hospital may be very different before and after the person's visit. The true change in this example is interpreted as instability of attitude scale measurement.
- The longer the time interval between measurements, the more likely that the concept has changed.
- The process of measuring a phenomenon can induce change in the phenomenon itself. This process is called reactivity. In measuring a person's attitude at test, the person can be sensitized to the subject under investigation and demonstrate change during retest. Thus the test - retest correlation will be low.

3.3.2 Alternative Form Method/Equivalent Form/Parallel Form

The alternative form method which is also known as equivalent / parallel form is used extensively in education, extension and development research to estimate the reliability of all types of measuring instruments. It also requires two testing situations with the same people like test- retest method. But it differs from test – retest method on one very important regard i.e., the same test is not administered on the second testing, but an alternate form of the same test is administered. Thus two equivalent reading tests should contain reading passages and questions of the same difficulty. But the specific passages and questions should be different i.e., approach is different. It is recommended that the two forms be administered about two weeks apart, thus allowing for day –to- day fluctuations in the person to occur. The correlation between two forms will provide an appropriate reliability coefficient.

Advantages

- The use of two parallel tests forms provides a very sound basis for estimating the precision of a psychological or educational test
- Superior to test- retest method, because it reduces the memory related inflated reliability.

Limitations

- Basic limitation is the practical difficulty of constructing alternate forms of two tests that are parallel.
- Requires each person's time twice.
- To administer a secondary separate test is often likely to represent a somewhat burdensome demand upon available resources.

3.3.3 Split-Half Method

Split - half method is also a widely used method of testing reliability of measuring instrument for its internal consistency. In split-half method, a test is given and divided into halves and are scored separately, then the score of one half of test are compared to the score of the remaining half to test the reliability.

In split-half method, 1st-divide test into halves. The most commonly used way to do this would be to assign odd numbered items to one half of the test and even numbered items to the other, this is called, Odd-Even reliability. 2nd- Find the correlation of scores between the two halves by using the Pearson *r* formula. 3rd- Adjust or revaluate correlation using Spearman-Brown formula which increases the estimate reliability even more.

Spearman-Brown formula

$$r = \frac{2\,r}{1+\,r}$$

r = estimated correlation between two halves (Pearson r).

Advantages

- Both, the test – retest and alternative form methods require two test administrations with the same group of people. In contrast the split –half method can be conducted on one occasion.
- Split-half reliability is a useful measure when impractical or undesirable to assess reliability with two tests or to have two test administrations because of limited time or money.

Limitations

- Alternate ways of splitting the items results in different reliability estimates even though the same items are administered to the same individuals at the same time.

  Example: The correlation between the first and second halves of the test would be different from the correlation between odd and even items.

Major Limitations of Reliability Estimating Methods

Test-retest method: Experience in the first testing usually will influence responses in the second testing.

Alternative form method: It can be quite difficult to construct alternative forms of a test that are parallel.

Split-half method: The correlation between the halves will differ depending on how the total number of items is divided into halves.

Alternate form method provide excellent estimate of reliability in spite of its limitation of constructing two forms of a test. To over come this limitation, it is recommended that, randomly divide a large collection of items in half to have two test administrations.

Check Your Progress 2

Note: a) Use the spaces given below for your answers.

b) Check your answers with those given at the end of the unit.

1) Write the three basic methods of determining the reliability?

.....

.....

.....

.....

.....

2) Write the major limitations in reliability determining methods?

.....

.....

.....

.....

.....

3.4 VALIDITY

According to Goode and Hatt, a measuring instrument (scale, test etc) possesses validity when it actually measures what it claims to measure. The subject of validity is complex and very important in development research because it is in this more than anywhere else, that the nature of reality is questioned. It is possible to study reliability without inquiring into the nature and meaning of one's variable. While measuring certain physical characteristics and relatively simpler attributes of persons, validity is no great problem. For example, the anthropometrics measurements of a pre-school child i.e., head and chest circumference can be measured by a measuring instrument having standard number of centimeters or inches. The weight of the child can be measured in pounds and kilograms. On the other hand, if a child development extension professional wish to study the

relation between malnutrition and intellectual development of pre-school children, there are neither any rule to measure the degree of malnutrition nor there any scales or clear cut physical attributes to measure intellectual development. It is necessary in such cases to invent indirect means to measure these characteristics. These means are often so indirect that the validity of the measurement and its product is doubtful.

Validity of Measuring Instrument or Measuring Phenomenon?

We defined validity as the extent to which any measuring instrument measures what it is intended to measure. But, strictly speaking, one validates not a measuring instrument, but an interpretation of data arising from a specified procedure. This distinction is central to validation, because it is quite possible for a measuring instrument to be relatively valid for measuring one kind of phenomenon but entirely invalid for assessing other phenomenon. Thus, one validates not the measuring instrument itself, but the measuring instrument in relation to the purpose for which it is being used.

3.4.1 Approaches to Validation of Measuring Instrument

Every measuring instrument, to be useful, must have some indication of validity. There are four approaches to validation of measuring instruments:

i) Logical validity / Face validity

ii) Jury opinion

iii) Known-group

iv) Independent criteria

i) Logical Validity

This is one of the most commonly used methods. It refers to either theoretical or commonsense analysis, which concludes simply that, the items, being what they, the nature of the continuum cannot be other than it is stated to be. Logical validation or face validity as it is sometimes called is almost always used because it automatically springs from the careful definition of the continuum and the selection of items. Such measure, which focuses directly on behavior of the kind in which the tester is interested, is said to have logic / face validity.

Example: The reading speed is measured by computing how much of a passage person reads with comprehension in a given time and the ability to solve arithmetic problems by success in solving a sample of such problems.

Limitation

- It is not wise to rely on logical and commonsense validation alone. Such claims for validity can at best be merely plausible and never definite. More than logical validity, it is required to render satisfactory use of a measuring instrument.

ii) Jury Opinion

This is an extension of the method of logical validation, except that in this case the confirmation of the logic is secured from a group of persons who

would be considered experts in the field in which the measuring instrument is being used.

Example: If a scale to measure mental retardation of pre-school children is constructed, psychologists, psychiatrists, pediatrician, clinical psychologists, social worker and teachers might constitute the jury to determine the validity of the scale.

Limitation

- Experts too are human and nothing but logical validity can result from this approach. Therefore, jury validation can be considered only slightly superior to logical validation.

iii) Known-Group

This technique is a variant of the jury procedure. In this case, the validity is implied from the known attitudes and other characteristics of analytical groups, however, rather than from their specific expertness. Thus, if a scale were being devised for the purpose of measuring the attitudes of people towards the Church, the questions could be tested by administering them to one group known to attend Church, to be active in Church activities and otherwise to give evidence of a favorable attitude towards this institution. These answers would be compared with those from a group known not to attend Church and also known to oppose the Church. If the scale failed to discriminate between the two groups it could not be considered to measure this attitude with validity. The known group technique of validation is frequently used and should not be discarded for falling somewhat short of perfection.

Limitation

- There might be other differences between the groups in addition to their known behavior with regard to religion, which might account for the differences in the scale scores.

Example: Differences in age, socioeconomic status, ethnic background etc.

- Further perhaps the known behavior under the study might be associated with a differential inclination to agree or disagree on a question in general. Hence careful use of the known group technique should be made.

iv) Independent Criteria

This is an ideal technique abstractly speaking but its application is usually difficult. There are four qualities desired in a criterion measure. In order of their importance they are :

a) *Relevance:* We judge a criterion to be relevant the extent to that standing on the criterion measure corresponds to the scores on scale.

b) *Freedom from bias :* By this we mean that the measure should be one on which each person has the same opportunity to make a good score. Example of biasing factors are such things as variation in the quality of equipment or conditions of work for a factory worker, a variation in the quality of teaching received by studying in different classes.

c) *Reliability :* If the criterion score is one that jumps around from day to day, so that the person who shows high job performance one week may show low job performance the next or who receives a high rating from one supervisor gets a low rating from another, then there is no possibility of finding a test that will predict that score. A measure that is completely unstable by itself cannot be predicted by anything else.

d) *Availability:* Finally, in the choice of a criterion measure we always encounter practical problems of convenience and availability. How long will we have to wait to get a criterion score for each individual? How much is it going to cost? Any choice of a criterion measure must make a practical limit to account.

However, when the independent criteria are good validation, it becomes a powerful tool and is perhaps the most effective of all techniques of validation.

Check Your Progress 3

Note: a) Use the spaces given below for your answers.

b) Check your answers with those given at the end of the unit.

1) Do you agree that 'one validates not the measuring instrument, but the purpose for which it is being used'? Write your agreement or disagreement.

.....

.....

.....

.....

.....

2) Name the four approaches to validation of measuring instrument.

.....

.....

.....

.....

.....

3.5 TYPES OF VALIDITY

The most important classification of types of validity is that prepared by a Joint Committee of American Psychological Association, the American Educational Research Association and the National Council on measurements used in education. There are three types of validity:

i) Content validity

ii) Criterion validity (Predictive validity and Concurrent validity)

iii) Construct validity.

3.5.1 Content Validity

The term content validity is used, since the analysis is largely in terms of the content.

Content validity is the representative ness or sampling adequacy of the content. Consider a test that has been designed to measure competence in using the English language. How can we tell how well the test in fact measures that achievement? First we must reach at some agreement as to the skills and knowledge that comprise correct and effective use of English, and that have been the objectives of language instruction. Then we must examine the test to see what skills, knowledge and understanding it calls for. Finally, we must match the analysis the test content against of course content and instrumental objectives, and see how well the former represents the latter. If the test represents the objectives, which are the accepted goals for the course, then the test is valid for use.

3.5.2 Criterion Validity

The two types of criterion validities are predictive validity and concurrent validity. They are much alike and with some exceptions, they can be considered the same, because they differ only in the time dimension. They are characterized by prediction and by checking the measuring instrument either now or in future against some outcome.

Example: A test that help researcher / teacher to distinguish between students who can study by themselves after attending the class and those who are in need of extra and special coaching, is said to have concurrent validity. The test distinguishes individually who differ in their present status. On the other hand, the investigator may wish to predict the percentage of passes during the final examination for that particular period. The adequacy of the test for distinguishing individuals who differ in the future may be called as predictive validity.

Predictive Validity Vs. Concurrent Validity

Predictive validity concerns a future criterion which is correlated with the relevant measure.

Example: Tests used for selection purposes in different occupations are, by nature, concerned with predictive validity. Thus a test used to screen applications for the post of 'health extension and development workers' could be validated by correlating their test scores with future performance in fulfilling the duties associated with health extension work.

Concurrent criterion is assessed by correlating a measure and the criterion at the same point in time.

Example: A verbal report of voting behaviour could be correlated with participation in an election, as revealed by official voting records.

3.5.3 Construct Validity

Both content and criterion validities have limited usefulness for assessing the validity of empirical measures of theoretical concepts employed in extension and development studies. In this context, construct validity must be investigated whenever no criterion or universe of content is accepted as entirely adequate to

define the quality to be measured. Examination of construct validity involves validation not only of the measuring instrument but of the theory underlying it. If the predictions are not supported, the investigator may have no clear guide as to whether the shortcoming is in the measuring instrument or in the theory.

Construct validation involves three distinct steps.

a) specify the theoretical relationship between the concepts themselves

b) examine the empirical relationship between the measures of the concepts

c) interpret the empirical evidence in terms of how it clarifies the construct validity of the particular measure.

Indeed strictly speaking, it is impossible to validate a measure of a concept in this sense unless there is a theoretical network that surrounds the concept.

Check Your Progress 4

Note: a) Use the spaces given below for your answers.

b) Check your answers with those given at the end of the unit.

1) Name the three types of validity.

.....

.....

.....

.....

.....

2) Write the major difference between predictive and concurrent validities.

.....

.....

.....

.....

.....

3.6 RELIABILITY OR VALIDITY - WHICH IS MORE IMPORTANT?

The real difference between reliability and validity is mostly a matter of definition. Reliability estimates the consistency of your measurement, or more simply the degree to which an instrument measures the same way each time it is used in under the same conditions with the same subjects. Validity, on the other hand, involves the degree to which you are measuring what you are supposed to, more simply, the accuracy of your measurement. Reliability refers to the consistency or stability of the test scores; validity refers to the accuracy of the inferences or interpretations you make from the test scores. Note also that reliability is a necessary but not sufficient condition for validity (i.e., you can have reliability without validity, but in order to obtain validity you must have reliability). In this

context, validity is more important than reliability because if an instrument does not accurately measure what it is supposed to, there is no reason to use it even if it measures consistently (reliably).

Let us examine the following three principles to understand the relationship between reliability and validity and to answer the question which is more important.

a) A test with high reliability may have low validity.

b) In the evaluation of measuring instruments, validity is more important than reliability.

c) To be useful, a measuring instrument must be both reasonably valid and reasonably reliable.

Consider the following four figures to understand easily the complex relationship between reliability and validity (Source: Patten, 2005).

In Fig. 3.1, the gun is aimed in a valid direction towards the target, and all the shots are consistently directed, indicating that they are reliable.

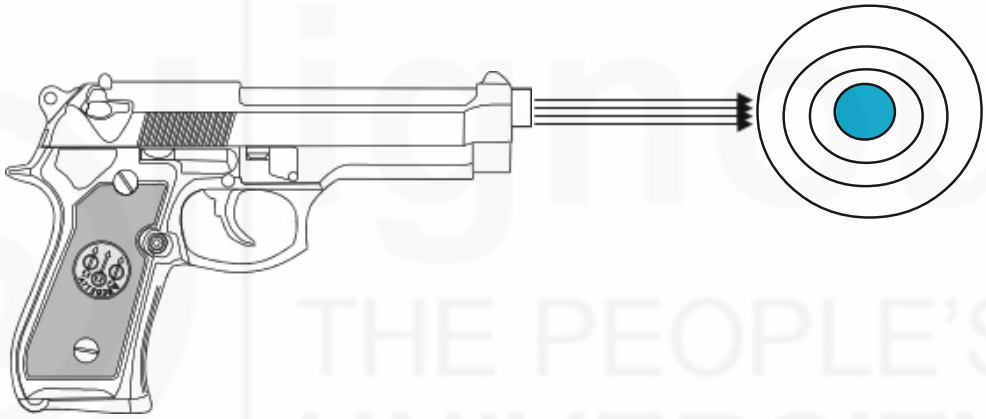


Fig. 3.1: Reliable and valid

In Fig. 3.2, the gun is also aimed in the direction of the target, but the shots are widely scattered, indicating low consistency or reliability. Thus the poor reliability undermines an attempt to achieve validity.

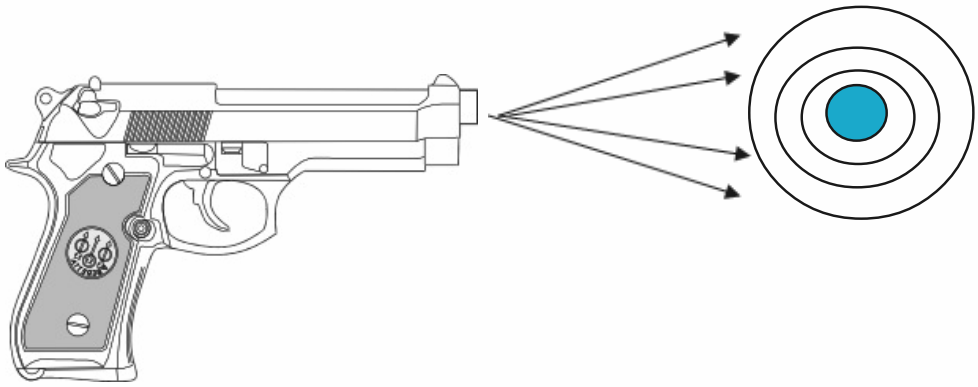


Fig. 3.2: Unreliable which undermines the valid aim of the gun – Not usefull

In Fig. 3.3, the gun is not pointed at the target, making it invalid, but there is great consistency in the shots in one direction, indicating that it is reliable (In a sense, it is very reliably invalid).

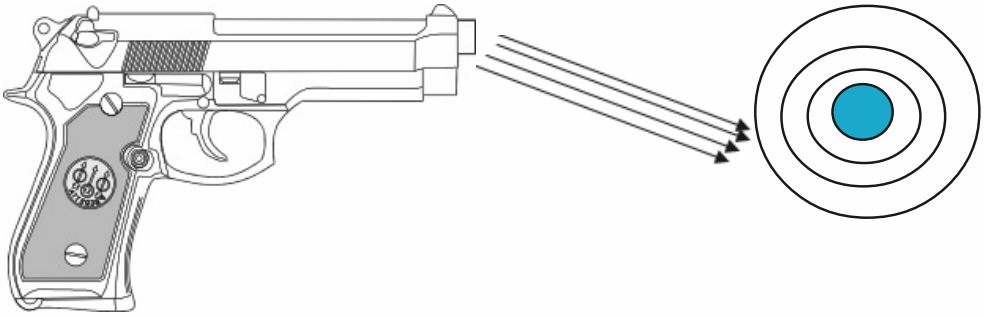


Fig. 3.3: Reliable but invalid – Not useful

In Fig. 3.4, the gun is not pointed at the target making it invalid, and the lack of consistency in the direction of the shots indicates its poor reliability.

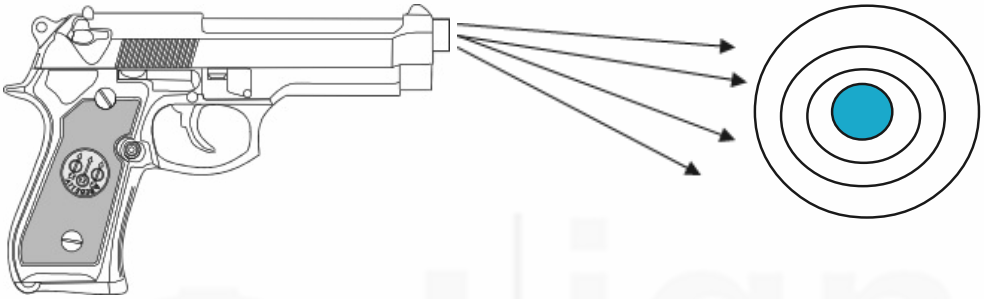


Fig. 3.4: Unreliable and invalid – Not useful

We may arrive at a conclusion that Fig. 3.1 represents the ideal in measurement. However, due to the limitations of measuring instruments in extension and development studies / social and behavioural sciences, we should not expect perfect reliability and validities. The direction of gun should be off at least a small amount - indicating a less than perfect validity. We also should expect some scatter in the shots, indicating less- than - perfect reliability. Clearly, our first priority should be to point the gun in the correct general direction, which promotes validity and then work on increasing reliability. This indicates that both reliability and validity are important in measurement, but among them validity is more important.

Check Your Progress 5

Note: a) Use the spaces given below for your answers.

b) Check your answers with those given at the end of the unit.

1) Among reliability and validity, which is more important and why?

.....

.....

.....

.....

.....

.....

.....

3.7 LET US SUM UP

In this unit we started by discussing the meaning of reliability and understood that reliability means consistency with which the instrument yields similar results. Later we highlighted that, among different aspects of reliability, two aspects i.e. stability and equivalence deserves special attention. We discussed the three important methods for assessing the reliability of measuring instruments. For the limitations mentioned in the discussion, neither test- retest method nor split-half method is recommended for estimating reliability. In contrast, the alternative form method is excellent for estimating reliability.

In the second part of the unit we have discussed the concept of validity and understood a measuring instrument possesses validity when it actually measures what it claims to measure. We examined the four approaches of validation of measuring instruments: logical validity / face validity, jury opinion, known-group and independent criteria. We also discussed the three types of validities and found that both content and criterion validities have limited usefulness in assessing the quality of development measures. In contrast, construct validation has generalized applicability in the extension and development research by placing the measure in theoretical context.

In the third and final part of the unit, we discussed, the relationship between reliability and validity and concluded that both reliability and validity are important in measurement, but among them validity is more important.

3.8 KEYWORDS

Reliability : Reliability means consistency with which the instrument yields similar results.

Validity : Validity is the ability of a measuring instrument to actually measure what it claims to measure.

Logical Validity : It refers to either theoretical or commonsense analysis, which concludes simply that, the items, being what they, the nature of the continuum cannot be other than it is stated to be.

Jury Opinion : The confirmation of the logic is secured from a group of persons who would be considered experts in the field in which the measuring instrument is being used.

Known-Group : The validity is implied from the known attitudes and other characteristics of analytical groups, however, rather than from their specific expertness.

Content Validity : Content validity is the representativeness or sampling adequacy of the content.

Predictive Validity : It concerns a future criterion which is correlated with the relevant measure.

Concurrent Validity : It is assessed by correlating a measure and the criterion at the same point in time.

Construct Validity : Construct validity involves validation of not only the measuring instrument but of the theory underlying it.

3.9 REFERENCES AND SELECTED READINGS

Carmines, E.G., and Zeller, R.A., (1979). Reliability and Validity Assessment. Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 17. Beverly Hills and London: Sage Publications.

Kothari, C.R. (2004). Research Methodology – Methods and Techniques, Second Revised Edition, New Age International Publishers, New Delhi.

Patten M.L. (2005). Understanding Research Methods – An Overview of the Essentials. Pyrczak Publishing, USA.

3.10 CHECK YOUR PROGRESS – POSSIBLE ANSWERS

Check Your Progress 1

1) The common name for any type of measurement device is 'instrument'.

2) Reliability estimates the consistency of our measurement, or more simply the degree to which an instrument measures the same way each time it is used in under the same conditions with the same subjects.

3) The synonyms for reliability are : dependability; stability; consistency; predictability; accuracy and equivalence .

4) The reliability of measuring instruments can be improved by (i) by standardizing the conditions under which the measurement takes place and (ii) by carefully designing directions for measurement with no variation from group to group, by using trained and motivated persons to conduct the research and also by broadening the sample of items.

Check Your Progress 2

1) The three basic methods of determining the reliability are : test – retest method; alternative form method and split-half method.

2) The major defect of test-retest method is that experience in the first testing usually will influence responses in the second testing. The practical limitation of alternative form method is that it can be quite difficult to construct alternative forms of a test that are parallel. The major problem with the split-half method approach is that the correlation between the halves will differ depending on how the total number of items is divided into halves.

Check Your Progress 3

1) Yes. I agree with the statement 'one validates not the measuring instrument, but the purpose for which it is being used' because it is quite possible for a measuring instrument to be relatively valid for measuring one kind of phenomenon, but entirely invalid for assessing other phenomenon.

2) The four approaches to validation of measuring instrument are: logical validity / face validity; jury opinion; known-group and; independent criteria.

Check Your Progress 4

1) The three types of validity are : Content validity; Criterion validity (Predictive validity and Concurrent validity) and Construct validity.

2) Predictive validity concerns a future criterion which is correlated with the relevant measure. Concurrent criterion is assessed by correlating a measure and the criterion at the same point in time.

Check Your Progress 5

1) Validity is more important than reliability because if an instrument does not accurately measure what it is supposed to, there is no reason to use it even if it measures consistently (reliably).

UNIT 4 SAMPLING

Structure

4.1 INTRODUCTION

Sampling has been an age old practice in everyday life. Whenever we want to buy a huge quantity of a commodity, we decide about the total lot by simply examining a small fraction of it. It has been established that the sample survey if planned properly, can give very precise information. Since in surveys a part of the population is only surveyed and inference is drawn about the whole population, the results likely to be different from the population values. But the advantage with the sample survey is that this type of error can be measured and controlled and it can be eliminated to great extent by employing properly trained persons in surveys. The other advantage of sample surveys are that it is less time consuming and involves less cost. Usually, the population is too large for the researcher to attempt to survey all of its members. A small, but carefully chosen sample can be used to represent the population. The sample reflects the characteristics of the population from which it is drawn.

After studying this unit, you should be able to

- discuss the meaning and importance of sampling
- describe the steps and criteria involved in selecting a sampling procedure
- distinguish different types of sampling
- explain the process of determination of sampling size

4.2 SAMPLING: MEANING AND CONCEPTS

4.2.1 Meaning of Sampling

According to Levin and Rubin, statisticians use the word, population, to refer not only to people, but, to all items that have been chosen for study. They use the word, sample, to describe a portion chosen from the population.

According to Croach and Housden, a sample is a limited number taken from a large group for testing and analysis, on the assumption that the sample can be taken as representative for the whole group.

According to Boyce, sampling makes an estimate about some of the characteristics of a population. To sample is to make a judgment or a decision about something after experiencing just part of it.

4.2.2 Concepts in Sampling

For clarity and brevity, some concepts and preliminaries of sampling theory, which are used in the study material, are discussed below.

- *Sampling Units and Population:* a unit may be taken as a well defined and identifiable element or a group of elements on which observations can be made. The aggregate of these units is termed as population and the population is said to be finite, if the units are countable. The population is sub-divided into suitable small units known as sampling units for the purpose of sampling. Sampling units may consist of one or more elementary units and each elementary unit belongs to one and one sampling unit.
- *Sampling Frame:* a sampling frame is a list of sampling units with identification particulars indicating the location of the sampling units. A sampling frame represents the population under investigation, and it is the base of drawing a sample. As far as possible, it should be up-to–date, i.e., free from omissions and duplications.
- *Sample:* a fraction of the population is said to constitute a sample. The number of units included in the sample is known as the size of the sample.
- *Sampling Fraction:* the ratio of the sample size, n, to the population size, N, is known as sampling fraction and it is denoted by (n / N).
- *Sampling Procedure/Method:* this is the method of selecting a sample from a population.
- *Census:* this denotes all the elements or units of a population which are used to explain the features of population. It usually refers to complete enumeration of all persons in the population.
- *Population Parameter and Sample Estimator:* any function of the values of units in the population, such as population mean or population variance, is termed a population parameter. There can only be one set of values for a population and the population values are treated as constant. However, the function of the values of the units in the sample, such as sample mean and sample variance, is known as a statistic. The value of the mean and variance differ from sample to sample and, therefore, it is a random variable.

4.2.3 Advantages of Sampling

Some of the key advantages of sampling are:

i) it costs less

ii) takes less time

iii) data are acquired quickly

iv) fewer mistakes are likely

v) a more detailed study can be done.

Now that you have read about the meaning and concept of sampling, answer the following questions in *Check Your Progress 1.*

Check Your Progress 1

Note: a) Write your answer in about 50 words.

b) Check your answer with possible answers given at the end of the unit

1) What do you mean by sampling? What are the advantages of sampling?

.....

.....

.....

.....

.....

2) What is the difference between a parameter and an estimator?

.....

.....

.....

.....

.....

4.3 TYPES OF SAMPLING

There are broadly two types of sampling:

i) Probability sampling

ii) Non-probability sampling

4.3.1 Probability Sampling

A probability sample is one in which each element of the population has a known, non zero chance of being included in the sample. Probability methods include simple random sampling, systematic sampling, and stratified sampling.

1) Simple Random Sample

The random sample entails that each and every individual in a population has an equal chance of being included in the sample and that the selection of one individual is in no way dependent upon the selection of another person. The two popularly used methods in random sampling are

i) draw of lottery

ii) using a random number table.

i) In lottery draw, for example, if we have to select a sample of 25 students from a total of 600 students in a college, then we make separate slips of paper for 600 students and put them in a box and thoroughly mix them.

After that, a person is asked to pick up one slip. Here, the probability of each of the student being selected in the sample is 1/600. This procedure is continued till the sample size is acquired.

ii) Another method of simple random sampling is to use a random number table for drawing 25 students from a total of 600 students. The procedure for using a random number table follows.

1) Number each element in the sample frame from 001 to 600.
2) Decide a random starting point in the table. Any point will do. Say second row in the second column (Appendix 1).
3) Look at the first digits at that point, because there are three digit in 600.
4) Then, if the number is less than 600, include it in the sample; if not then look for a number where the first three digits are less than 600.
5) From that point you can move in any direction. Select only three digit numbers that are less than 600, until you have 25 such numbers.

Note: You can move in any direction in the random number table because every digit has been placed in the table at random.

For example, here if we start from the second row in the third column, then, the random numbers are: 31684; 09865; 14491; 34691, continuing till 25 samples are selected.

2) Systematic Random Sample

Designing a Systematic Random Sample is sometimes quite difficult and time consuming and therefore, Systematic Random Sample, like Simple Random Sample, also uses a list of all members of the population in its sampling frame. However, instead of using random numbers to select the sample elements, the researcher applies a skip interval to the list to produce a sample of the required size.

$$\text{Skip interval} = \frac{\text{number of elements in the population}}{\text{the required sample size}}$$

$$K = \frac{N}{n}$$

K = skip interval

N = Universe size

n = Sample size

For example if we have to select a sample of 100 persons from a universe of 1000 population, then the skip is 10. In this case one number between 1 and 10 has to be selected. Suppose 5 is selected, then the first sample would be 5^{th} and the next one 15^{th}, 25^{th}, 35^{th}, 45^{th}, and so on. One of the advantages of this method is that it is more convenient than other methods and simple to design. Again, it is used with very large populations.

3) Stratified Random Sample

In Stratified Random Sampling, the target population of N units is first divided into k subpopulations of N_1, N_2, ..., N_k units. These populations are non-overlapping and together they comprise the whole population. So that $N_1 + N_2 + ... + N_k = N$

The sub-populations are called strata. The number in each stratum should be known. A sample is drawn from each stratum independently. The sample sizes within 'k' strata are denoted by n_1, n_2, ..., n_k respectively. If the total sample size 'n' is to be drawn from the target population than $n_1 + n_2 + ... + n_k = n$

If a simple random sample is drawn in each stratum, the whole procedure is described as stratified random sampling.

Stratified random sampling requires more than making a list of elements (and estimating the number of elements on the list). It also involves ordering that list by sub groups (or strata) and then, to do sampling randomly or systematically within those sub groups. This method of sampling is used for the following reasons.

- It can reduce the errors in the statistical estimates calculated from the sample.
- It allows you to create a sample that is exactly representative of the various sub groups in the population that you find to be of special interest.

For example, the selected village may have households of SC, ST, OBCs, Others, Minority. The village population first may be divided in to smaller sub groups of different sections of population (stratum) and, thus, the village sample may consist of households from each stratum so that sample may contain all the important characteristics of the village population. In the case of SRS, the sample of all strata/ sub groups sometimes may not be included or covered adequately.

- This method helps in conducting and managing a large scale survey to be conducted in a country like India. The agency conducting the survey may have field offices in different locations; each one can supervise the survey for a part of the population.
- The basic idea is that it sub-divides the heterogeneous population into homogeneous sub-populations. If each stratum is homogenous in itself, a precise estimate of any stratum mean can be obtained from a small sample, thus, saving a lot of time and cost.

There are two types of stratified samples.

A **proportionate stratified sample** selects the number of elements from each stratum so that the stratum sample size (n_1, n_2,……., n_k) is proportional to their respective stratum population size (N_1, N_2, ……., N_k).

Consider the following examples:

- A selected village may have households of SC(10%), ST (5%), OBCs (45%), Others (30%), Minority (10%). A village sample of 100 may constitute the households of various casts in the above proportion/ percentage so that the sample may contain all important characteristics of village population.

- Hospital patients are stratified according to age, dividing the population into those who are aged 50years or above, and, those who are under 50. If there are twice as many people aged 50 or above admitted to the hospital as those under 50, a proportionate stratified sample will include twice as many people aged 50 or above.

A **disproportionate stratified sample** selects the number of elements from each stratum so that the stratum sample size is not proportional to the stratum population size. The most common reason for selecting this type of sample is when you want to study a relatively rare but important subpopulation, such as younger patients suffering from heart disease. Proportionate stratification may result in too few elements being selected so that little, if any, statistical analysis can be done. Consequently, even if these patients represent only 1% of the population, you might decide to make them 10% of the final sample. However, once we combine values of all stratums, the size of the higher selected proportion needs to be readjusted which is called weighted estimate.

4) Probability Proportion to Size (PPS) Sample

It has been observed that the elementary units of the population vary in size. Such ancillary information about the size of the unit can be utilized in selecting the sample so as to get better and efficient estimates of the population parameter. For example villages with larger geographical area are likely to have larger area under food crops; therefore, in estimating the production, it would be desirable to adopt a sampling scheme in which villages are selected with probability proportional to geographical area. When units vary in their size and the variable under study is directly related with the size of the unit, the probabilities may be assigned proportional to the size of the unit.

Probability Proportion to Size (PPS) Sampling assures higher probability of selection to sampling unit which are larger in size. This technique was initially used in estimation of crop production, fruits production etc because productivity is directly related with the size of field. In social science surveys also characteristics of village population is influenced by the size of population. The **procedure of selecting the sample** is described below.

Suppose you have to select 5 villages from the list of 10 using PPS sampling. First arrange all villages in ascending or descending order of population size as may be seen in column 2 of the table 1. Then, in the third column, find the cumulative sum of population size and in the fourth column, assign them range of serial numbers as shown below in the table.

Table 4.1: Village population Size

Sl.No.	Village Population Size	Cumulative Sum of Population Size	Cumulative Population Size Interval
1	**2**	**3**	**4**
1	200	200	0001 - 0200
2	250	450	0201 - 0450
3	300	750	0451 - 0750
4	350	1100	0751 - 1100
5	400	1500	1101- 1500
6	450	1950	1501 - 1950
7	500	2450	1951 - 2450
8	550	3000	2451 - 3000
9	600	3600	3001 - 3600
10	650	4250	3601 - 4250
Total	**4250**		

Please notice that the total population of all villages in the target population is a four digit number (4250). Therefore, initially, a random number in four digits, which is less than or equal to the total population of all villages (4250), is selected from the random number table. For example, it is 0331 which will correspond to serial number 2. Next random number is 4320; therefore, it may be discarded. The next number selected is 1296; therefore, it will correspond to serial number 5. The next random numbers may be 1553, 2402 and 3640 which will correspond to serial numbers 6, 8, and 10 respectively. In this way, selected villages will be serial numbers 2, 5, 6, 8, 10.

5) Cluster Sample

Cluster sampling is a sampling technique used when natural groupings are evident in a statistical population. It is often used in marketing research. In this technique, the total population is divided into these known groups (or clusters) and a sample of the groups is selected. Then the required information is collected from the elements within each selected group. This may be done for every element in these groups, or a sub sample of elements may be selected within each of these groups. The technique works best when most of the variation in the population is within the groups, not between them.

Briefly, the procedure for selecting a cluster sample is given below.

- The population is divided into N groups, called clusters.
- The researcher randomly selects n clusters to include in the sample.
- The number of observations within each cluster is known:

  M = M1 + M2 + M3 + ...…… + MN
- Each element of the population can be assigned to one, and only one, cluster.

Cluster sampling should be used only when it is economically justified - when reduced costs can be used to overcome losses in precision. This is most likely to occur in the following situations.

- Constructing a complete list of population elements is difficult, costly, or impossible. For example, it may not be possible to list all elementary units of the populations, for example all households in village, block, etc. However, it would be possible to randomly select a subset of villages, blocks (stage 1 of cluster sampling) and, then, interview the head of family in a house of the selected cluster (stage 2).
- The population is concentrated in natural clusters (city blocks, schools, hospitals, etc.). For example, to conduct personal interviews of operating room nurses, it might make sense to randomly select a sample of hospitals (stage 1 of cluster sampling) and then interview all of the operating room nurses at that hospital. Using cluster sampling, the interviewer could conduct many interviews in a single day at a single hospital. Simple random sampling, in contrast, might require the interviewer to spend all day travelling to conduct a single interview at a single hospital.

As discussed above, in the cluster sampling method, the primary selecting unit is not a household, rather a natural cluster of households, viz., hamlets in villages, or, created clusters, viz., schools, malls, etc., may be decided. The first list of clusters may be selected using the SRS or the PPS sampling techniques. Then, from each selected cluster, all units, or, some of the units, may be selected as per the required sample size using Stratified Random Sampling or the Systematic Random Sampling techniques.

This sampling technique is quite popular in evaluation surveys in health – it is also called the 30 Cluster Sampling Technique. This is also a rapid method of data collection as the researcher can collect more data in less time due to the decrease in transportation time as compared with other sampling techniques.

4.3.2 Non-Probability Sampling

A **non-probability sample** is one in which a case in a sample is chosen in such a manner that it gives you information for the sample itself and makes it possible to generalize the findings for the population with certain degree of precision. Such a sample is also called a purposive sample. This kind of sampling is primarily used to collect information on market surveys to know the attitude, opinion, behaviour, reactions of individuals. There are many types of non-probability samples, including snowball sampling, convenience, purposive/ judgment, quota sampling, etc.

1) Convenience Sample

The convenience sample is so called because it is relatively easy to obtain and contact. In this method the investigators are usually asked to select the people for the interview in accordance to the instructions from the researcher. The benefit of a convenience sample is that the interviewer can usually get interviews done quickly and cheaply. Convenience sampling is appropriate for exploratory research.

2) Judgments Sample

A judgment sample is similar to that of convenience sample. In a judgment sample, the researcher selects samples that are believed to represent the population. The selection of samples is based on the knowledge of the population and the characteristics which the sample is to represent. It is less costly and very useful for forecasting.

3) Quota Sample

Quota sampling is like stratified sampling. In quota sampling, the population is categorized into several strata which consist of an expected size, and the samples are considered to be important for the population they represent. The advantages of quota sample are that it involves a short time duration, is less costly, and gives moderate representation to a heterogeneous population.

4) Snowball Sample

This is one of the important types of non-probability sampling. In snowball sampling, the investigator encourages the respondents to give the names of other acquaintances and it continues growing in size and chains until the research purpose is achieved. It is also, therefore, known as networking, chain, or referred sampling method. It is very useful in the study of networking and is less costly.

A comprehensive overview of the various types of sampling can be seen in figure 4.1

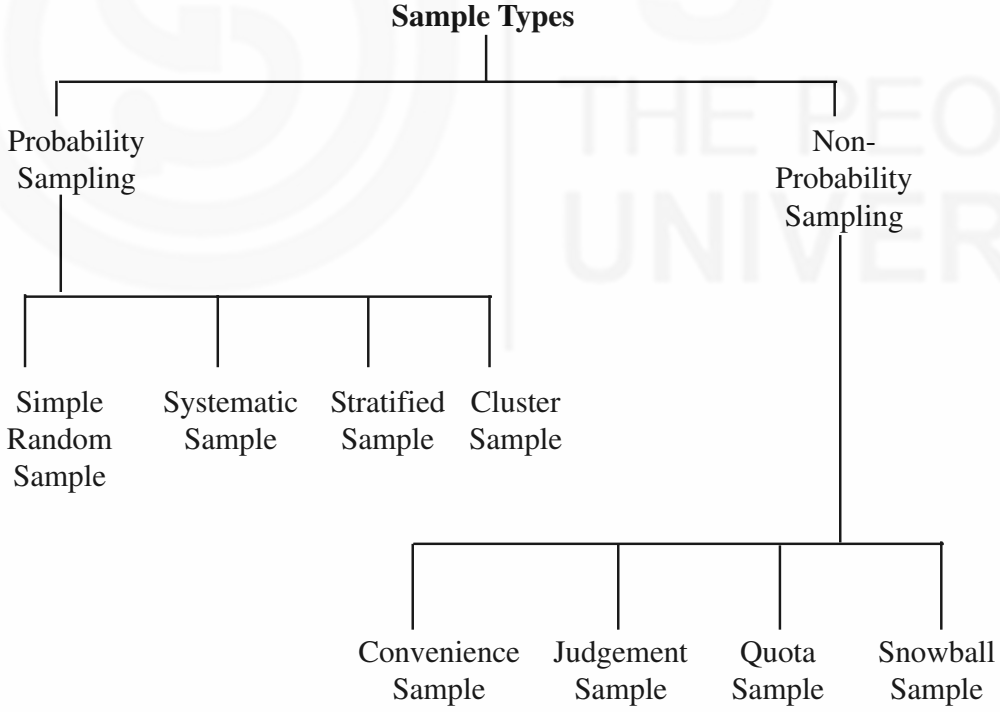
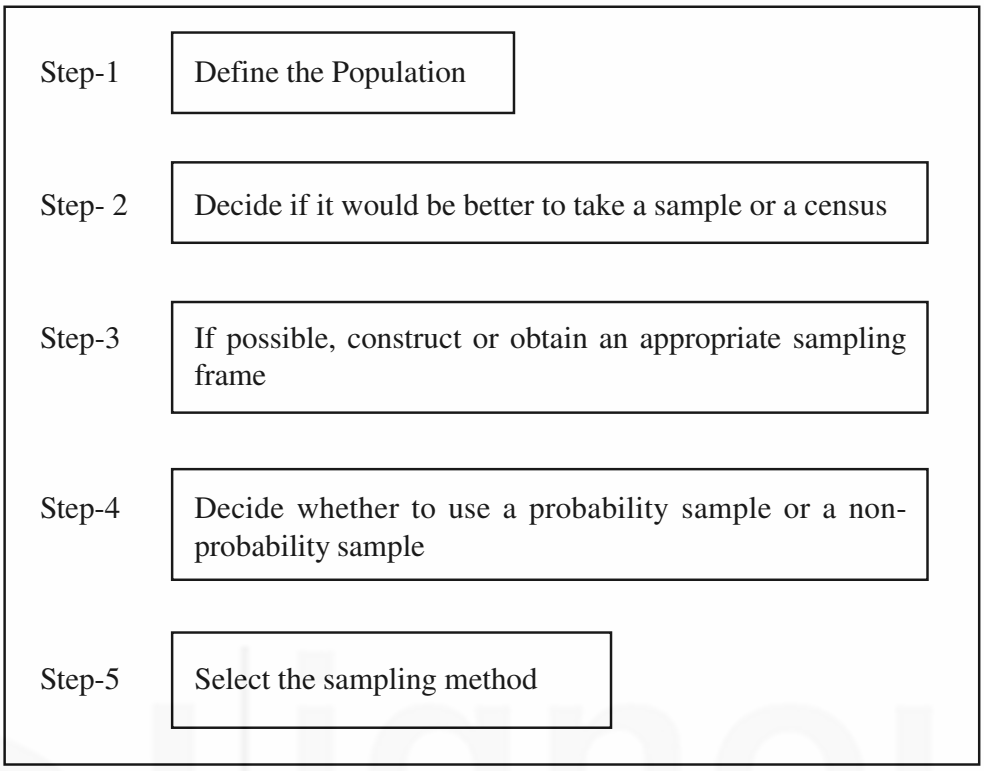


Fig. 4.1: Types of Sample

4.4 SAMPLE DESIGN PROCESS

The sample design process follows five steps as given in Box-1

Box 4.1: Sample Design Process



Source: John Boyce (www.mhhe.com/av/boyceze)

Step-1: Define the Population

We use the word, population, frequently in our day-to-day conversations, for example, 'The population of India', or, 'The population of Punjab', or, 'The population of Kerela'. However, the meaning of the word, population, in research is different from what we use in day-to-day conversation. A research population may be defined as 'a clearly defined group of entities that have some characteristics in common'. This means the kind of people on whom we wish to base our research project. Sometimes, in research, we use the word, universe, instead of population.

In a research project, our intention is to learn or infer something about the population. Whether we would use a sample or a population has to be clearly defined. For example, if we want to conduct a study on road safety, then the task of defining the population for a survey would be whether we should

- interview only the people who drive two wheelers
- interview only the people who drive four wheelers
- interview the pedestrians
- interview only who are hand rickshaw pullers or ride bi-cycle.

Therefore, judging a population is the starting of the sampling process.

Step-2: Decide whether to take a sample or a census

After judging the population, the next step in the sampling process is to decide whether to take a sample or a population in your research project. In a census, usually every member of the population is interviewed.

While in a sample method only selected members of the population are included. From the census we obtain data that are called population parameter, and from the sample we obtain statistics in a parameter. A parameter is a measurement of a characteristic of a population, while a sample statistic is used as an estimate of a population parameter.

Usually, in individual research, we use census when the population is small, and sample when the population is large.

Step-4: Decide whether to use a probability sample or a non-probability sample

The fourth step in the sampling process is whether to use probability sampling or non-probability sampling. In a probability sample, the sample elements are chosen by random selection, while in non-probability sampling, each sample element is chosen according to whether the researcher decides that it should be included or not.

Step-5: Select the sampling method

Last, but not least in the sampling process, is the selection of the sampling method. In the probability sampling method, the following four principal kinds of probability sampling are used: the simple random sample, the systematic sample, the stratified sample, and the cluster sample. The main non-probability sampling methods are the convenience sample, the judgment sample, the quota sample, and the snowball sample.

Sampling Frame

A sampling frame is a listing of all the elements from which you will draw the sample. In the ideal situation, the sampling frame will include all elementary units in the target population. A list of employees in an organization can create a sampling frame that exactly matches the population of interest. You should try to ensure that the sampling frame has the following characteristics.

- It is actually created from the target population.
- It is as complete a list as possible of the elements in the population.

In this section, you have read about the various types of sampling and the sample design process. Now, answer the following questions in *Check Your Progress 2.*

Check Your Progress 2

Note: a) Write your answer in about 50 words.

b) Check your answer with possible answers given at the end of the unit

1) What is a stratified sample?

.....

.....

.....

.....

.....

2) What is a cluster sample and when is it best used?

.....

.....

.....

.....

.....

.....

3) What do you understand by a quota sample?

.....

.....

.....

.....

.....

.....

4.5 ERRORS IN SAMPLING

Many mistakes and errors in social science research happen because of misleading and biased sampling. A sample which does not represent the population is called a biased sample. According to Yule and Kendal, "Bias may be due to imperfect instruments, the personal qualities of the observer, defective techniques and other cases. Like experimental error, it is difficult to eliminate entirely, but usually may be reduced to relatively small dimensions by taking proper care." There are two types of errors such as sampling errors and non-sampling errors. These are discussed below:

i) Sampling Error

By definition, when you have collected a sample from a population, you have less than complete information about the population. This, in turn, means that there is a chance that the sample statistics you calculate, (for example, the mean of a variable, a frequency distribution, etc.) may not be an unbiased estimate of the population parameter.

The error in the sample estimate is not an intrinsic impediment to analysis. For probability samples, sampling theory allows you to calculate the expected amount of error given a particular sample size, sampling method, and the specific statistic of interest. In general terms, the sampling error for a statistic can be defined as:

$$\text{Standard error} = \sqrt{\frac{Variance}{n}} = \frac{\text{sd}}{\sqrt{\text{n}}}$$

Where n refers to the number of respondents (sample size).

As the sample size increases, the standard error of a statistic decreases; as the variance, or dispersion, of a statistic increases, so does its sampling error.

Sampling error decreases rapidly as the sample size increases from a few hundred to about 1000 respondents. However, there is rarely any reason to select larger samples while comparing the increased cost of survey with reduction in sampling error (see 'Calculating the Sample Size', in next section).

The formula for the standard error of a proportion is simple and easy to apply:

Standard error = $\sqrt{\frac{p \times (1-p)}{n}}$

Here, p represents the proportion of successes (favourable response, those who received the benefits), {q = (1-p)} represents the proportion of failures (those who did not receive the benefits), and n is the total number of respondents. The standard error of a statistic is greatest when p and (1-p) are equal, which occurs when each is 0.50, or 50%, of the sample.

ii) Non-Sampling Error

Before discussing how to determine sample size, we will briefly review other sources of error in surveys. When you read a a news article that reports the results of a national poll, the error in the estimates is always listed, derived, generally speaking, from Equation 6.2. However, experienced survey researchers know that errors due to other sources are typically greater than the error due to sampling alone. Following are some other types of errors.

- Measurement errors, caused by poorly written questions, poorly designed questionnaires, respondent errors in completing questionnaires, and so on.
- Non-response errors, caused because the respondents are not a representative subset of the population.
- Data coding errors, caused, by errors in coding and entering the data.

Of these error sources, the first two are typically more severe. In mail surveys, non-response error is often the most serious problem.

There are two critical characteristics of these non sampling errors. First, as mentioned above, their sum is often greater than the sampling error. Second, and more insidious, these errors are often impossible to estimate for any one survey, especially measurement and non-response errors. Consequently, using Equation 6.1 and Equation 6.2 to estimate the error in a statistics often provides a false sense of security.

Experienced survey researchers take this fact into account by being more cautions in discussing survey results than the sampling error alone would indicate, and you should do the same. Ideally, the other sources of error would balance themselves out so that errors in one direction negate errors in the other directions, but you cannot assume that this is the case.

4.6 DETERMINATION OF SAMPLE SIZE

The sample size can be determined by:

i) Using a formula

ii) Using a table

4.6.1 Determining Sample Size Using a Formula

(when population is greater than 10,000)

$$nf = \frac{n}{1+(n/N)}$$ (when population is less than 10,000)

n, nf = desired sample size

Z = the standard normal deviate

p = the portion in the target population estimated to have a particular characteristic. If there is no reasonable estimate, then use 50 percent (.50).

q = 1-p

d= degree of accuracy desired, usually set at .05 or occasionally at .02.

n = the estimate of the population size

Z at 99% confidence level i.e. at 1% level of significance = 2.58

Z at 95% confidence level i.e. at 5% level of significance = 1.96

Z at 90% confidence level i.e. at 10% level of significance = 1.65

Exaple: (when population is more than 10,000)

If the proportion of target population with a certain characteristic is .50, the Z statistic 1.96 and we desire accuracy at the 0.05 level, then the sample size is

$$n = \frac{(1.96)^2\ (.50\times.50)}{(0.05)^2}$$

$$= \frac{3.84\times0.25}{(.0025)}$$

$$= \frac{0.96}{0.0025}$$

$$= 384$$

If we use the more convenient 2.0 for the Z statistic, then the sample size will be smaller.

=

= 286

4.6.2 Determining Sample Size by Using a Table

Another way to determine sample size is to rely on published tables which provide the sample size for a given set of criteria. Table 1 presents sample size values that will be appropriate for many common sampling problems. The table includes sample sizes for both continuous and categorical data assuming alpha levels of .10, .05, or .01.

Table 4.1: Table for Determining Minimum Returned Sample Size for a Given Population Size for Continuous and Categorical Data

	Sample size					
	Continuous data (margin of error= .03)			Categorical data (margin of error= .05)		
Population size	alpha=.10 t=1.65	alpha=.05 t=1.96	alpha=.01 t=2.58	p=.50 t=1.65	p=.50 t=1.96	p=.50 t=2.58
100	46	55	68	74	80	87
200	59	75	102	116	132	154
300	65	85	123	143	169	207
400	69	92	137	162	196	250
500	72	96	147	176	218	286
600	73	100	155	187	235	316
700	75	102	161	196	249	341
800	76	104	166	203	260	363
900	76	105	170	209	270	382
1,000	77	106	173	213	278	399
1,500	79	110	183	230	306	461
2,000	83	112	189	239	323	499
4,000	83	119	198	254	351	570
6,000	83	119	209	259	362	598
8,000	83	119	209	262	367	613
10,000	83	119	209	264	370	623

In this session you studied about errors in sampling and determination of sample size. Now, answer the questions given in *Check Your Progress 3*.

Check Your Progress 3

Note: a) Write your answer in about 50 words.

b) Check your answer with possible answers given at the end of the unit

1) What is sampling error?

.....

.....

.....

.....

.....

2) How the sample size is determined using the formula?

.....

.....

.....

.....

.....

4.7 LET US SUM UP

In this unit, we discussed the meaning and various concepts in sampling particularly of sample and population. There is also a detailed discussion on the sample types and sample design process. There are two types of sampling such as probability and non-probability sampling. The types of probability sampling are the Simple Random Sample, the Systematic Sample, the Stratified Sample, and the Cluster Sample, while different types of non-probability sample are the Convenience Sample, the Quota Sample, the Judgment Sample, and the Snowball Sample. The unit also discusses various steps of the sampling design process. This is followed by two of the very important concepts of sampling: the determination of sample size and errors in sampling.

4.8 KEYWORDS

Sample : A sample is simply a subset of a larger aggregation, i.e., typically a population and it contains all the characteristics of a population,

Sampling : The process of selection of subjects/study elements to create a sample for collecting information about a population.

Standard Error : This is the expected amount of error while estimating the specific statistic of interest, using a particular sample size and sampling method with respect to actual population value.

Sampling Error : While collecting information from a sample, there is a chance that the sampling statistics may not be equal to the same values in the population. The error is that the sample does not contain complete information about the population.

Confidence Interval : This gives the probability of the sample estimate falling within the interval.

Sample Size : The number of elementary units in a sample is called a sample size.

4.9 REFERENCES AND SELECTED READINGS

Daroga S. and F.S. Chaudhary (2002), *Theory & Analysis of Sample Survey Designs,* New Age International Publishers Ltd.

Goldberg H. I. (1991), *Survey Sampling: In Epidemiological Approach to Reproductive Health*, WHO.

Henderson R. H and Sundaresan T (1982), *Cluster Sampling to Assess Immunisation Coverage: A Review of Experience with a Simplified Sampling Method*, WHO Bulletin.

Pandurang V. S. (1984), *Sampling Theory of Surveys with Applications,* Iowa State Press.

Cochran W. G. (1977), *Sampling Technique*, John Willey & Sons.

4.10 CHECK YOUR PROGRESS – POSSIBLE ANSWERS

Check Your Progress 1

1) What do you mean by sampling? What are the advantages of sampling?

   Generally, a sample implies a small representative of a large whole. This sampling method is frequently used in social science research to save time. Some of the key advantages of sampling are: (i) it costs less; (ii) takes less time; (iii) data are sometimes wanted quickly; (iv) fewer mistakes are likely; (v) a more detailed study can be done.

2) What is the difference between a parameter and an estimator?

   Any function of the values of units in the population, such as the population mean or population variance, is termed, a population parameter. There can only be one set of values for a population, there population values are treated as constant. However, the function of the values of the units in the sample, such as the sample mean and sample variance is known as a statistic. The value of the mean and variance differs from sample to sample and, therefore, it is a random variable.

Check Your Progress 2

1) What is stratified sampling?

In *stratified sampling*, the target population of N units is first divided into k subpopulations of units. These populations are non-overlapping and together they comprise the whole population, so that

The sub-populations are called strata. The number in each stratum should be known. A sample is drawn from each stratum independently. The sample sizes within 'k' strata are denoted by respectively If the total sample size n is to be drawn from the target population then

If a simple random sample is drawn in each stratum, the whole procedure is described as *stratified random sampling*.

2) What is cluster sampling and when is it best used?

Cluster sampling is a sampling technique used when natural groupings are evident in a statistical population. It is often used in marketing research. In this technique, the total population is divided into these known groups (or clusters) and a sample of the groups is selected. Then, the required information is collected from the elements within each selected group. This may be done for every element in these groups or a sub sample of elements may be selected within each of these groups. The technique works best when most of the variation in the population is within the groups, not between them.

3) What do you understand by quota sample?

Quota sampling is like stratified sampling. In quota sampling, the population is categorized into several strata which consist of an expected size and they are considered to be important for the population they are supposed to represent. The advantages of the quota sample are: shorter time duration, less costly, and gives moderate representation to a heterogeneous population.

Check Your Progress 3

1) What is sampling error?

By definition, when you have collected a sample from a population, you have less than complete information about the population. This, in turn, means that there is a chance that the sample statistics you calculate, (for example, the mean of a variable, a frequency distribution, etc.) may not be unbiased estimate of the population parameter. This error is called sampling error.

2) How is the sample size determined using the formula?

The calculation of the sample size is concerned with the number of respondents required. To determine the number to select for the sample drawn from the sampling frame, you must estimate the non-response rate. The actual sample size to be drawn is:

So, if any survey organization decides that they need 700 respondents, and the expected response rate from the population is 50%, then 700/0.50, or 1400, customers must be drawn from the sampling frame.