M.Sc. Information Technology

SEMESTER I

PAPER 1 - RESEARCH IN COMPUTING

UNIT IV

MEASUREMENT CONCEPTS, SAMPLING AND FIELD WORK

A. LEVELS OF SCALE MEASUREMENT

CONTENTS

1.1 OBJECTIVES

Following are the objectives of this unit:

- ✓ To understand the levels of measurement
- ✓ To differentiate between different scales of measurement
- ✓ To analyse scales
- ✓ To understand index measures
- ✓ To implement reliability
- ✓ To analyse validity
- ✓ To differentiate between reliability and validity
- ✓ To understand sensitivity

1.2 INTRODUCTION

Measurement is a procedure of allocating numerical value to some characteristics or variables or events according to scientific rules. It is the process observing and recording the observations which are collected as part of a research effort. Measurement means the description of data in terms of numbers – accuracy; objectivity and communication. The combined form of these three is the actual measurement.

In this unit, we will understand different levels of measurement and see their types.

Definition:

'Measurement is the process of observing and recording the observations that are collected as part of a research effort.'

'Measurement is a process of describing some property of a phenomenon of interest, usually by assigning numbers in a reliable and valid way.'

The decision statement, corresponding research questions, and research hypotheses can be used to decide what concepts need to be measured in a given project. Measurement is the process of describing some property of a phenomenon of interest, usually by assigning numbers in a reliable and valid way. The numbers convey information about the property being measured. When numbers are used, the researcher must have a rule for assigning a number to an observation in a way that provides an accurate description. Measurement can be illustrated by thinking about the way instructors assign students' grades.

Example:
A – (above 60% score)
B – (between 50 – 60% score)
C – (between 40 – 50 % score
Here A, B, C can also be termed as scales of measurement. Some scales may better classify the data and each scale has the potential of producing error or some lack of validity

1.3 LEVELS OF MEASUREMENT

Level of measurement refers to the relationship among the values that are assigned to the attributes for a variable. It is important because –

- ✓ Knowing the level of measurement helps you decide how to interpret the data from that variable
- ✓ Knowing that a measure is nominal, then you know that the numerical values are just short codes for the longer names.
- ✓ Knowing the level of measurement helps you decide what statistical analysis is appropriate on the values that were assigned.

If a measure is nominal, then you know that you would never average the data values or do a t-test on the data. There are four distinguish levels of measurement. The levels are –

- ✓ Nominal
- ✓ Ordinal

- ✓ Interval
- ✓ Ratio

Levels of measurement are important for two reasons.

i) First, they emphasize the generality of the concept of measurement. Although people do not normally think of categorizing or ranking individuals as measurement, in fact they are as long as they are done so that they represent some characteristic of the individuals.

ii) Second, the levels of measurement can serve as a rough guide to the statistical procedures that can be used with the data and the conclusions that can be drawn from them.

1.3.1 NOMINAL SCALE

The nominal scale (called as dummy coding) simply places people, events, perceptions, etc. into categories based on some common trait. Some data are naturally suited to the nominal scale such as males vs. females, white vs. black vs. blue, and American vs. Asian. The nominal scale forms the basis for such analyses as Analysis of Variance (ANOVA) because those analyses require that some category is compared to at least one other category.

The nominal scale is the lowest form of measurement because it doesn't capture information about the focal object other than whether the object belongs or doesn't belong to a category; either you are a smoker or not a smoker, you attended university or you didn't, a subject has some experience with computers, an average amount of experience with computers, or extensive experience with computers.

No data is captured that can place the measured object on any kind of scale say, for example, on a continuum from one to ten. Coding of nominal scale data can be accomplished using numbers, letters, labels, or any symbol that represents a category into which an object can either belong or not belong. In research activities a Yes/No scale is nominal. It has no order and there is no distance between Yes and No.

The statistics which can be used with nominal scales are in the non-parametric group. The most likely ones would be - mode; crosstabulation - with chi-square. There are also highly sophisticated modelling techniques available for nominal data.

1.3.2 ORDINAL SCALE

An ordinal level of measurement uses symbols to classify observations into categories that are not only mutually exclusive and exhaustive; in addition, the categories have some explicit relationship among them. For example, observations may be classified into categories such as taller and shorter, greater and lesser, faster and slower, harder and easier, and so forth.

However, each observation must still fall into one of the categories (the categories are exhaustive) but no more than one (the categories are mutually exclusive). Most of the commonly used questions which ask about job satisfaction use the ordinal level of measurement.

For example, asking whether one is very satisfied, satisfied, neutral, dissatisfied, or very dissatisfied with one's job is using an ordinal scale of measurement. The simplest ordinal scale is a ranking.

Ordinal data would use non-parametric statistics. These would include - median and mode; rank order correlation; non-parametric analysis of variance. Modelling techniques can also be used with ordinal data.

1.3.3 INTERVAL SCALE

An interval level of measurement classifies observations into categories that are not only mutually exclusive and exhaustive, and have some explicit relationship among them, but the relationship between the categories is known and exact. This is the first quantitative application of numbers. In the interval level, a common and constant unit of measurement has been established between the categories.

For example, the commonly used measures of temperature are interval level scales. We know that a temperature of 75 degrees is one degree warmer than a temperature of 74 degrees.

Numbers may be assigned to the observations because the relationship between the categories is assumed to be the same as the relationship between numbers in the number system.

For example, 74+1= 75 and 41+1= 42. The intervals between categories are equal, but they originate from some arbitrary origin, that is, there is no meaningful zero point on an interval scale. The standard survey rating scale is an interval scale.

When you are asked to rate your satisfaction with a piece of software on a 7 point scale, from Dissatisfied to Satisfied, you are using an interval scale. Interval scale data would use parametric statistical techniques - Mean and standard deviation; Correlation; Regression; Analysis of variance; Factor analysis; and whole range of advanced multivariate and modelling techniques.

1.3.4 RATIO

The ratio level of measurement is the same as the interval level, with the addition of a meaningful zero point. There is a meaningful and non-arbitrary zero point from which the equal intervals between categories originate.

For example, weight, area, speed, and velocity are measured on a ratio level scale.

In public policy and administration, budgets and the number of program participants are measured on ratio scales. In many cases, interval and ratio scales are treated alike in terms of the statistical tests that are applied. A ratio scale is the top level of measurement and is not often available in social research. The factor which clearly defines a ratio scale is that it has a true zero point.

The simple way to understand the levels of measurement or to select a measurement scale is as follows –

- ✓ If one object is different from another, then we use a nominal scale.
- ✓ If one object is bigger or better or more of anything than another, then we use an ordinal scale.
- ✓ If one object is so many units (degrees, inches, etc.) more than another, then we use an interval scale.
- ✓ If one object is certain times as big or bright or tall or heavy as another, then we use a ratio scale.

The following criteria should be considered in the selection of the measurement scale for variables in a study. Researcher should consider the scale that will be most suitable for each variable under study. Important points in the selection of measurement scale for a variable are:

- ✓ Scale selected should be appropriate for the variables one wishes to categorise.
- ✓ It should be of practical use.
- ✓ It should be clearly defined.
- ✓ The number of categories created (when necessary) should cover all possible values.
- ✓ The number of categories created (when necessary) should not overlap, i.e., it should be mutually exclusive.
- ✓ The scale should be sufficiently powerful. Variables measured at a higher level can always be converted to a lower level, but not vice versa.

For example, observations of actual age (ratio scale) can be converted to categories of older and younger (ordinal scale), but age measured as simply older or younger cannot be converted to measures of actual age.

The four levels of measurement discussed above have an important impact on how you collect data and how you analyze them later. Collect at the wrong level, and you will end of having to adjust your research, your design, and your analyzes. Make sure you consider carefully the level at which you collect your data, especially in light of what statistical procedures you intend to use once you have the data in hand.

1.4 ANALYSIS OF SCALES

Mathematical operations can be performed with numbers from nominal scales, the result may not have a great deal of meaning. Although you can put numbers into formulas and perform calculations with almost any numbers, the researcher has to know the meaning behind the numbers before meaningful conclusions can be drawn.

1.4.1 DISCRETE MEASURES

Discrete measures are those that take on only one of a finite number of values. A discrete scale is most often used to represent a classification variable. Therefore, discrete scales do not represent intensity of measures, only membership. Common discrete scales include any yes-or-no response, matching, colour choices, or practically any scale that involves selecting from among a small number of categories. Thus, when someone is asked to choose from the following responses:

- ✓ Disagree
- ✓ Neutral
- ✓ Agree

the result is a discrete value that can be coded 1, 2, or 3, respectively. This is also an ordinal scale to the extent that it represents an ordered arrangement of agreement. Nominal and ordinal scales are discrete measures.

1.4.2 CONTINUOUS MEASURES

Continuous measures are those assigning values anywhere along some scale range in a place that corresponds to the intensity of some concept. Ratio measures are continuous measures. Thus, when we measure sales for each salesperson using the money (every rupee) amount sold, he is assigning a continuous measure. A number line could be constructed ranging from the least amount sold to the most, and a spot on the line would correspond exactly to a salesperson's performance.

Table 1.1: Example of Continuous scales

Question/ Rating	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
I learned a lot from this study material.	5	4	3	2	1

This is a discrete scale because only the values 1, 2, 3, 4, or 5 can be assigned. Moreover, it is an ordinal scale because it only orders based on agreement. We really have no way of knowing that the difference in agreement of somebody marking a 5 instead of a 4 is the same as the difference in agreement of somebody marking a 2 instead of a 1.

(For calculation purpose: Only the mean is not an appropriate way of stating central tendency and, technically, we really shouldn't use many common statistics on these responses.)

A scaled response of this type (refer table 1.1) takes on more values, the error introduced by assuming that the differences between the discrete points are equal becomes smaller. This may be seen by imagining a ***Likert scale*** (the traditional business research agreement scale shown above) with a thousand levels of agreement rather than three.

The differences between the different levels become so small with a thousand levels that only tiny errors could be introduced by assuming each interval is the same. Therefore, business researchers generally treat interval scales containing five or more categories of response as interval. (They are commonly called 5-point Likert scale; 7-point Likert scale and so on)

When fewer than five categories are used, this assumption is inappropriate. The researcher should keep in mind, however, the distinction between ratio and interval measures. Errors in judgment can be made when interval measures are treated as ratio.

1.5 INDEX MEASURES

Multi-item instruments for measuring a construct are called index measures, or composite measures. An index measure assigns a value based on how much of the concept being measured is associated with an observation. Indexes often are formed by putting several variables together.

For example, a social class index might be based on three weighted variables: occupation, education, and area of residence. Usually, occupation is seen as the single best indicator and would be weighted highest. With an index, the different attributes may not be strongly correlated with each other.

A person's education does not always relate strongly to their area of residence. The Consumer Satisfaction Index shows how satisfied consumers are based on an index of satisfaction scores. Readers are likely not surprised to know that certain consumers appear more satisfied with soft drinks than they are with cable TV companies based on this index.

Composite measures also assign a value based on a mathematical derivation of multiple variables.

For example, salesperson satisfaction may be measured by combining questions such as "How satisfied are you with your job? How satisfied are you with your territory? How satisfied are you with the opportunity your job offers?" For most practical applications, composite measures and indexes are computed in the same way.

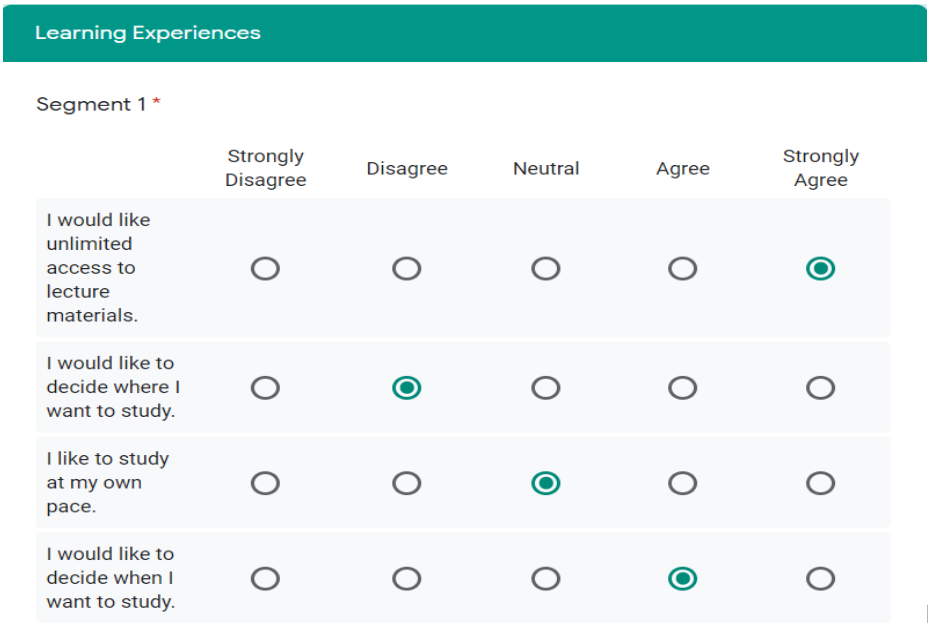
Definitions:
Index Measure: An index assigns a value based on how much of the concept being measured is associated with an observation. Indexes often are formed by putting several variables together.

Attribute: A single characteristic or fundamental feature of an object, person, situation, or issue.

Composite Measures: Assign a value to an observation based on a mathematical derivation of multiple variables.

1.5.1 COMPUTING SCALE VALUES

The below stated example is a computation of the data collected using Likert Scale. For this scale, the value of Strongly Agree (SA) is 5, Agree (A) is 4, Neutral (N) is 3, Disagree (D) is 2 and Strongly Disagree (SD) is 1. For the total score obtained for these segments of questions is 5 + 2 + 3 + 4 = 14


Learning Experiences

Segment 1 *

	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
I would like unlimited access to lecture materials.	○	○	○	○	◉
I would like to decide where I want to study.	○	◉	○	○	○
I like to study at my own pace.	○	○	◉	○	○
I would like to decide when I want to study.	○	○	○	◉	○

Figure 1.1: Sample of Likert Scale

Such scales are also called as Summated scales.

<u>Definition:</u>

Summated Scale: A scale created by simply summing (adding together) the response to each item making up the composite measure.

Sometimes, a response may need to be reverse-coded before computing a summated or averaged scale value. Reverse coding means that the value assigned for a response is treated oppositely from the other items. Thus, on a 5-point scale, the values are reversed as follows:

- 5 becomes 1
- 4 becomes 2
- 3 stays 3
- 2 becomes 4
- 1 becomes 5

This happens for questions which are negative in nature. An ideal scale must have 60-70 % questions positive in nature and 30-40 % questions negative in nature. This is done to ensure that the person filling the questions is not selecting options randomly.

Example of a negative question: I would not like to decide when I want to study (based on the questions given in figure 1.1)

1.6 CRITERIA FOR GOOD MEASUREMENT

1.6.1 RELIABILITY

Reliability refers to the consistency or repeatability of an operationalized measure. A reliable measure will yield the same results over and over again when applied to the same thing. It is the degree to which a test consistently measures whatever it measures. If you have a survey question that can be interpreted several different ways, it is going to be unreliable. One person may interpret it one way and another may interpret it another way. You do not know which interpretation people are taking.

Even answers to questions that are clear may be unreliable, depending on how they are interpreted. Reliability refers to the consistency of scores obtained by the same persons when they are re-examined with the same tests on different occasions, or with different sets of equivalent items, or under other variable examining conditions.

Research requires dependable measurement. Measurements are reliable to the extent that they are repeatable and that any random influence which tends to make measurements different from occasion to occasion or circumstance to circumstance is a source of measurement error. Errors of measurement that affect reliability are random errors and errors of measurement that affect validity are systematic or constant errors. Reliability of any research is the degree to which it gives an accurate score across a range of measurement. It can thus be viewed as being 'repeatability' or 'consistency'.

Internal consistency: Different questions, same construct. Test-retest, equivalent forms and split-half reliability are all determined through correlation. There are a number of ways of determining the reliability of an instrument. The procedure can be classified into two groups –

External Consistency Procedures: It compare findings from two independent processes of data collection with each other as a means of verifying the reliability of the measure. *For example*, test-retest reliability, parallel forms of the same test, etc.

Internal Consistency Procedures: The idea behind this procedure is that items measuring the same phenomenon should produce similar results. For example, split-half technique.

1.6.1.1 TYPES OF RELIABILITY

1) Test-Retest Reliability

The most obvious method for finding the reliability of test scores is by repeating the identical test on a second occasion. Test-retest reliability is a measure of reliability obtained by administering the same test twice over a period of time to a group of individuals.

For Example- A test designed to assess student learning in psychology could be given to a group of students twice, with the second administration perhaps coming a week after the first. The obtained correlation coefficient would indicate the stability of the scores.

2) Split-Half Reliability

Split-half reliability is a subtype of internal consistency reliability. In split half reliability we randomly divide all items that purport to measure the same construct into two sets. We administer the entire instrument to a sample of people and calculate the total score for each randomly divided half. The most commonly used method to split the test into two is using the odd-even strategy.

3) Inter-Rater Reliability

Inter-rater reliability is a measure of reliability used to assess the degree to which different judges or raters agree in their assessment decisions. Inter-rater reliability is also known as inter-observer reliability or inter-coder reliability. Inter-rater reliability is useful because human observers will not necessarily interpret answers the same way; raters may disagree as to how well certain responses or material demonstrate knowledge of the construct or skill being assessed. Inter-rater reliability might be employed when different judges are evaluating the degree to which art portfolios meet certain standards. Inter-rater reliability is especially useful when judgments can be considered relatively subjective.

4) Parallel-Forms Reliability

Parallel forms reliability is a measure of reliability obtained by administering different versions of an assessment tool to the same group of individuals. The scores from the two versions can then be correlated in order to evaluate the consistency of results across alternate versions.

5) Coefficient alpha (α):

It is the most commonly applied estimate of a multiple-item scale's reliability. Coefficient α represents internal consistency by computing the average of all possible split-half reliabilities for a multiple-item scale. The coefficient demonstrates whether or not the different items converge. Although coefficient α does not address validity, many researchers use α as the sole indicator of a scale's quality. Coefficient alpha ranges in value from 0, meaning no consistency, to 1, meaning complete consistency

1.6.2 VALIDITY

Validity refers to whether the measure actually measures what it is supposed to measure. If a measure is unreliable, it is also invalid. That is, if you do not know what it is measuring, it certainly cannot be said to be measuring what it is supposed to be measuring. On the other hand, you can have a consistently unreliable measure.

For example, if we measure income level by asking someone how many years of formal education they have completed, we will get consistent results, but education is not income (although they are positively related).

In general, validity is an indication of how sound your research is. More specifically, validity applies to both the design and the methods of your research. Validity in data collection means that your findings truly represent the phenomenon you are claiming to measure. Valid claims are solid claims.

There are two main types of validity, internal and external. Internal validity refers to the validity of the measurement and test itself, whereas external validity refers to the ability to generalize the findings to the target population.

1.6.2.1 TYPES OF VALIDITY

1) Face Validity

Face validity refers to the degree to which a test appears to measure what it purports to measure. The stakeholders can easily assess face validity. Although this is not a very 'scientific' type of validity, it may be an essential component in enlisting motivation of stakeholders. If the stakeholders do not believe the measure is an accurate assessment of the ability, they may become disengaged with the task.

For example, if a measure of art appreciation is created all of the items should be related to the different components and types of art. If the questions are regarding historical time periods, with no reference to any artistic movement, stakeholders may not be motivated to give their best effort or invest in this measure because they do not believe it is a true assessment of art appreciation.

2) Predictive Validity

Predictive validity refers to whether a new measure of something has the same predictive relationship with something else that the old measure had. In predictive validity, we assess the operationalization's ability to predict something it should theoretically be able to predict.

For example, we might theorize that a measure of math ability should be able to predict how well a person will do in an engineering-based profession. We could give our measure to experienced engineers and see if there is a high correlation between scores on the measure and their salaries as engineers. A high correlation would provide evidence for predictive validity - it would show that our measure can correctly predict something that we theoretically think it should be able to predict.

3) Criterion-Related Validity

Criterion validity is a test of a measure when the measure has several different parts or indicators in it - compound measures. Each part or criterion of the measure should have a relationship with all the parts in the measure for the variable to which the first measure is related in a hypothesis. When you are expecting a future performance based on the scores obtained currently by the measure, correlate the scores obtained with the performance. The later performance is called the criterion and the current score is the prediction. It is used to predict future or current performance - it correlates test results with another criterion of interest.

For example, if a physics program designed a measure to assess cumulative student learning throughout the major. The new measure could be correlated with a standardized measure of ability in this discipline, such as GRE subject test. The higher the correlation between the established measure and new measure, the more faith stakeholders can have in the new assessment tool.

4) Content Validity

In content validity, you essentially check the operationalization against the relevant content domain for the construct. This approach assumes that you have a good detailed description of the content domain, something that's not always true. In content validity, the criteria are the construct definition itself - it is a direct comparison. In criterion-related validity, we usually make a prediction about how the operationalization will perform based on our theory of the construct. When we want to find out if the entire content of the behavior/ construct/ area is represented in the test we compare the test task with the content of the behavior. This is a logical method, not an empirical one.

For Example, if we want to test knowledge on Bangladesh Geography it is not fair to have most questions limited to the geography of Australia.

5) Convergent Validity

Convergent validity refers to whether two different measures of presumably the same thing are consistent with each other - whether they converge to give the same measurement. In convergent validity, we examine the degree to which the operationalization is similar to (converges on) other operationalizations that it theoretically should be similar to.

For example, to show the convergent validity of a test of arithmetic skills, we might correlate the scores on test with scores on other tests that purport to measure basic math ability, where high correlations would be evidence of convergent validity. Or, if SAT scores and GRE scores are convergent, then someone who scores high on one test should also score high on the other. Different measures of ideology should classify the same people the same way. If they do not, then they lack convergent validity.

6) Concurrent Validity:

Concurrent validity is the degree to which the scores on a test are related to the scores on another already established, test administered at the same time or to some other valid criterion available at the same time. This compares the results from a new measurement technique to those of a more established technique that claims to measure the same variable to see if they are related. In concurrent validity, we assess the operationalization's ability to distinguish between groups that it should theoretically be able to distinguish between.

For example, if we come up with a way of assessing manic-depression, our measure should be able to distinguish between people who are diagnosed manic-depression and those diagnosed paranoid schizophrenic. If we want to assess the concurrent validity of a new measure of empowerment, we might give the measure to both migrant farm workers and to the farm owners, theorizing that our measure should show that the farm owners are higher in empowerment. As in any discriminating test, the results are more powerful if you are able to show that you can discriminate between two groups that are very similar.

7) Construct Validity

Construct validity is used to ensure that the measure is actually measure what it is intended to measure (i.e. the construct), and not other variables. Using a panel of 'experts' familiar with the construct is a way in which this type of validity can be assessed. The experts can examine the items and decide what that specific item is intended to measure. This is whether the measurements of a variable in a study behave in exactly the same way as the variable itself. This involves examining past research regarding different aspects of the same variable. It is also the degree to which a test measures an intended hypothetical construct.

For example, if we want to validate a measure of anxiety. We have a hypothesis that anxiety increases when subjects are under the threat of an electric shock, then the threat of an electric shock should increase anxiety scores.

8) Formative Validity

When applied to outcomes assessment it is used to assess how well a measure is able to provide information to help improve the program under study.

For example - when designing a rubric for history one could assess student's knowledge across the discipline. If the measure can provide information that students are lacking knowledge in a certain area, for instance the Civil Rights Movement, then that assessment tool is providing meaningful information that can be used to improve the course or program requirements.

9) Sampling Validity

Sampling validity ensures that the measure covers the broad range of areas within the concept under study. Not everything can be covered, so items need to be sampled from all of the domains. This may need to be completed using a panel of 'experts' to ensure that the content area is adequately sampled. Additionally, a panel can help limit 'expert' bias.

For example - when designing an assessment of learning in the theatre department, it would not be sufficient to only cover issues related to acting. Other areas of theatre such as lighting, sound, functions of stage managers should all be included. The assessment should reflect the content area in its entirety.

10) Discriminant Validity

In discriminant validity, we examine the degree to which the operationalization is not similar to (diverges from) other operationalizations that it theoretically should be not be similar to.

For example, to show the discriminant validity of a Head Start program, we might gather evidence that shows that the program is not similar to other early childhood programs that don't label themselves as Head Start programs.

1.6.2.2 RELIABILITY VERSUS VALIDITY

Reliability is a necessary but not sufficient condition for validity. A reliable scale may not be valid. For example, a purchase intention measurement technique may consistently indicate that 20 percent of those sampled are willing to purchase a new product. Whether the measure is valid depends on whether 20 percent of the population indeed purchases the product. A reliable but invalid instrument will yield consistently inaccurate results.

1.6.3 SENSITIVITY

The sensitivity of a scale is an important measurement concept, particularly when changes in attitudes or other hypothetical constructs are under investigation. Sensitivity refers to an instrument's ability to accurately measure variability in a concept. A dichotomous response category, such as "agree or disagree," does not allow the recording of subtle attitude changes. A more sensitive measure with numerous categories on the scale may be needed.

For example, adding "strongly agree," "mildly agree," "neither agree nor disagree," "mildly disagree," and "strongly disagree" will increase the scale's sensitivity. The sensitivity of a scale based on a single question or single item can also be increased by adding questions or items. In other words, because composite measures allow for a greater range of possible scores, they are more sensitive than single-item scales. Thus, sensitivity is generally increased by adding more response points or adding scale items.

1.7 SUMMARY

In this unit, we understood what is meant by Measurement when considered in research. Some key definitions we saw were:

Measurement: Measurement is the process of observing and recording the observations that are collected as part of a research effort.

Index Measure: An index assigns a value based on how much of the concept being measured is associated with an observation. Indexes often are formed by putting several variables together.

Attribute: A single characteristic or fundamental feature of an object, person, situation, or issue.

Composite Measures: Assign a value to an observation based on a mathematical derivation of multiple variables.

Summated Scale: A scale created by simply summing (adding together) the response to each item making up the composite measure.

We also understood the important concepts like reliability and validity and understood the difference between them.

1.8 QUESTIONS

1. Discuss different levels of measurement.
2. Using appropriate examples, write notes on:
    a. Nominal Scale
    b. Ordinal Scale
    c. Interval Scale

3. What is meant by analysis of scales?
4. Elaborate on the criteria for good measurement.
5. Write a note on sensitivity as a criteria for measurement.

1.9 REFERENCES

1. Albright Winson, 2015, Business Analytics, 5th Edition,
2. Hair, 2014, Multivariate data Analysis, 7th Edition, Pearson Publication
3. Kabir, Syed Muhammad. (2016). Measurement Concepts: Variable, Reliability, Validity, and Norm.
4. Mark Saunders, 2011, Research Methods for Business Students, 5th Edition
5. Shefali Pandya, 2012, Research Methodology, APH Publishing Corporation, ISBN: 9788131316054, 813131605X
6. William G.Zikmund, B.J Babin, J.C. Carr, Atanu Adhikari, M.Griffin, 2016, Business Research Methods, Edition 8, Cengage Publications

M.Sc. Information Technology

SEMESTER I

PAPER 1 - RESEARCH IN COMPUTING

UNIT IV

MEASUREMENT CONCEPTS, SAMPLING AND FIELD WORK

B. ATTITUDE MEASUREMENT & QUESTIONNAIRE DESIGN

CONTENT

2.1 OBJECTIVES

Following are the objectives of this unit:

- ✓ To understand the attitude rating scale
- ✓ To categorise different attitude rating scales
- ✓ To understand the development of questionnaire procedure

2.2 INTRODUCTION

Attitude can be defined as a tendency to react favourably, neutrally, or unfavourably toward a particular class of stimuli, such as a custom, institutional practice, or national group. There are two challenges a researcher faces when measuring an individual's attitude. First, an individual's attitude toward an object cannot be observed directly but must be inferred from observed behaviour, such as responses to a questionnaire. And second, there is no inherent scale associated with the observed behaviour. Techniques for measuring attitude are:

1. *Ranking*: A measurement task that requires respondents to rank order a small number of stores, brands, or objects on the basis of overall preference or some characteristic of the stimulus.
2. *Rating*: A measurement task that requires respondents to estimate the magnitude of a characteristic or quality that a brand, store, or object possesses.
3. *Sorting*: A measurement task that presents a respondent with several objects or product concepts and requires the respondent to arrange the objects into piles or classify the product concepts.

Along with attitude, another critical part of the survey is the creation of questions that must be framed in such a way that it results in obtaining the desired information from the respondents. There are no scientific principles that assure an ideal questionnaire and in fact, the questionnaire design is the skill which is learned through experience. In this unit, we will see Attitude rating scales and Questionnaire designing

2.3 ATTITUDE RATING SCALES

Simple attitude scaling may be used when questionnaires are extremely long, when respondents have little education, or for other specific reasons. A number of simplified scales are merely checklists: A respondent indicates past experience, preference, and the like merely by checking an item.

In many cases the items are adjectives that describe a particular object. In a survey of small-business owners and managers, respondents indicated whether they found working in a small firm more rewarding than working in a large firm, as well as whether they agreed with a series of attitude statements about small businesses. For example, 77 percent said small and mid-sized businesses "have less bureaucracy," and 76 percent said smaller companies "have more flexibility" than large ones.

2.3.1 CATEGORY SCALE

A rating scale that consists of several response categories, often providing respondents with alternatives to indicate positions on a continuum.

Table 2.1: Sample of category scale

How frequently do you use online payment modes?				
Never	Rarely	Sometimes	Often	Very Often

The simplest rating scale contains only two response categories: agree/disagree. Expanding the response categories provides the respondent with more flexibility in the rating task. Even more information is provided if the categories are ordered according to a particular descriptive or evaluative dimension.

This category scale is a more sensitive measure than a scale that has only two response categories. By having more choices for a respondent, the potential exists to provide more information. However, if the researcher tries to represent something that is truly bipolar (yes/no, female/male, member/non-member, and so on) with more than two categories, error may be introduced.

2.3.2 LIKERT SCALE

A Likert scale is a psychometric scale commonly used in questionnaires, and is the most widely used scale in survey research. When responding to a Likert questionnaire item, respondents specify their level of agreement to a statement. The scale is named after its inventor, psychologist Rensis Likert. The Likert scale can also be used to measure attitudes of people. When responding to a Likert questionnaire item, respondents specify their level of agreement or disagreement on a symmetric agree-disagree scale for a series of statements. Thus, the range captures the intensity of their feelings for a given item.

Definition:

A measure of attitudes designed to allow respondents to rate how strongly they agree or disagree with carefully constructed statements, ranging from very positive to very negative attitudes toward some object.

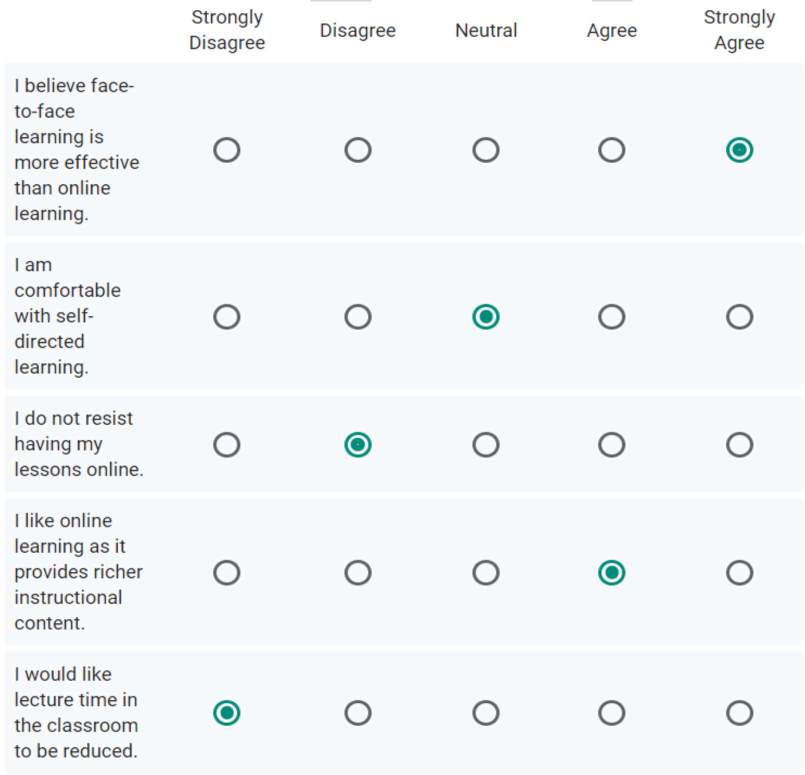


Figure 2.1: Sample of 5-point Likert Scale

Reverse Recording

The statement given in this example (figure 2.1) is positively framed. If a statement is framed negatively (such as “I dislike online learning as it provides richer instructional content”), the numerical scores would need to be reversed. This is done by reverse recoding the negative item so that a strong agreement really indicates an unfavourable response rather than a favourable attitude. In the case of a five-point scale, the recoding is done as follows:

Table 2.2: Reverse Scoring of Likert Scale

Old Value	New Value
1	**5**
2	**4**
3	**3**
4	**2**
5	**1**

2.3.3 COMPOSITE SCALING

A Likert scale may include several scale items to form a composite scale. Each statement is assumed to represent an aspect of a common attitudinal domain. The total score is the summation of the numerical scores assigned to an individual's responses. Based on the example given in figure 1.1, twethe maximum possible score for the composite would be 20 if a 5 were assigned to "strongly agree" responses for each of the positively worded statements and a 5 to "strongly disagree" responses for the negative statement. Item 3 is negatively worded and therefore it is reverse coded, prior to being used to create the composite scale.

Definition:

Composite Scaling: A way of representing a latent construct by summing or averaging respondents' reactions to multiple items each assumed to indicate the latent construct.

2.3.4 SEMANTIC DIFFERENTIAL

The semantic differential is actually a series of attitude scales. This popular attitude measurement technique consists of getting respondents to react to some concept using a series of seven-point bipolar rating scales. Bipolar adjectives—such as "good" and "bad," "modern" and "old fashioned," or "clean" and "dirty"—anchor the beginning and the end (or poles) of the scale. The subject makes repeated judgments about the concept under investigation on each of the scales.

Definition

Semantic Scaling: It's a measure of attitudes that consists of a series of seven point rating scales that use bipolar adjectives to anchor the beginning and end of each scale.

The scoring of the semantic differential can be illustrated using the scale bounded by the anchors "modern" and "old-fashioned." Respondents are instructed to check the place that indicates the nearest appropriate adjective. From left to right, the scale intervals are interpreted as "extremely modern," "very modern," "slightly modern," "both modern and old-fashioned," "slightly old-fashioned," "very old-fashioned," and "extremely old-fashioned":

Modern--- Old-fashioned

2.3.5 NUMERIC SCALING

A numerical scale simply provides numbers rather than a semantic space or verbal descriptions to identify response options or categories (response positions). For example, a scale using five response positions is called a five-point numerical scale. A six-point scale has six positions and a seven-point scale seven positions, and so on. Consider the following numerical scale:

Now that you've had your laptop for about one year, please tell us how satisfied you are.

Extremely Dissatisfied 1 2 3 4 5 6 7 Extremely Satisfied

This numerical scale uses bipolar adjectives in the same manner as the semantic differential. In practice, researchers have found that a scale with numerical labels for intermediate points on the scale is as effective a measure as the true semantic differential.

Definition
Numeric Scaling: An attitude rating scale similar to a semantic differential except that it uses numbers, instead of verbal descriptions, as response options to identify response positions.

2.3.6 STAPEL SCALE

The Stapel scale, is used to measure simultaneously the direction and intensity of an attitude. Modern versions of the scale, with a single adjective, are used as a substitute for the semantic differential when it is difficult to create pairs of bipolar adjectives. The modified Stapel scale places a single adjective in the center of an even number of numerical values (ranging, perhaps, from +3 to –3). The scale measures how close to or distant from the adjective a given stimulus is perceived to be.

The advantages and disadvantages of the Stapel scale are very similar to those of the semantic differential. However, the Stapel scale is markedly easier to administer, especially over the telephone. Because the Stapel scale does not require bipolar adjectives, it is easier to construct than the semantic differential.

Definition
Staple Scale: A measure of attitudes that consists of a single adjective in the center of an even number of numerical values.

2.3.7 CONSTANT SUM SCALE

A constant sum scale is a type of question used in a market research survey in which respondents are required to divide a specific number of points or percent as part of a total **sum**. The allocation of points is divided to detail the variance and weight of each category. A constant-sum scale requires respondents to divide a fixed number of points among several attributes corresponding to their relative importance or weight.

Definition

Constant Sum Scale: A measure of attitudes in which respondents are asked to divide a constant sum to indicate the relative importance of attributes; respondents often sort cards, but the task may also be a rating task.

For Example:

Question: Using 100 points, please apply a number of points to each factor based on how important each is to you when buying a home. You must total 100 points divided among the factors.

Answer: Price, Location, School District, Inside Features, etc.

The respondent is given 100 points. They may choose to apply 80 to price, 15 to location, and spread out the remaining 5 points among other factors. When you analyze this data set, the differentiation between factors becomes evident. Most survey software will automatically tally and sum the point values to ensure they add to a constant sum of 100.

This constant sum scale adds another layer of analytical thinking for the respondent rather than just selecting one, running through a checklist of choices, or selecting from a grid or scaling question. It forces respondents to slow down and understand the relative value of each factor and compare the importance of one over another. It maximizes the chances of creating differentiation between your choices.

2.3.8 GRAPHIC RATING SCALE

A graphic rating scale lists the traits each employee should have and rates workers on a numbered scale for each trait. The scores are meant to separate employees into tiers of performers, which can play a role in determining promotions and salary adjustments. A graphic rating scale presents respondents with a graphic continuum. The respondents are allowed to choose any point on the continuum to indicate their attitude.

Definition

Graphic Rating Scale: A measure of attitude that allows respondents to rate an object by choosing any point along a graphic continuum.

For Example:

Table 2.3: Example of Analyses performance of employees working on Project A from April to June 2020

	Extremely Poor	**Bad**	**Average**	**Good**	**Excellent**
Attention to detail					
Knowledge					
Teamwork					
Initiative					
Creative					

A variation of the graphic ratings scale is the ladder scale. This scale also includes numerical options.

Example: This ladder scale represents the "ladder of life." As you see, it is a ladder with eleven rungs numbered 0 to 10. Let's suppose the top of the ladder represents the best possible life for you as you describe it, and the bottom rung represents the worst possible life for you as you describe it.

On which rung of the ladder do you feel your life is today?

0 1 2 3 4 5 6 7 8 9 10

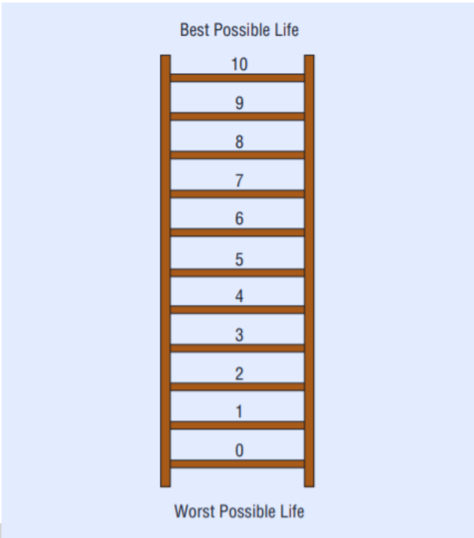


Figure 2.2 Ladder Scale example

2.3.9 THURSTONE INTERVAL SCALE

It is an attitude scale in which judges assign scale values to attitudinal statements and subjects are asked to respond to these statements. This was developed because the attitudes vary along continua and should be measured accordingly. A Thurstone scale has a number of "agree" or "disagree" statements. It is a unidimensional scale to measure attitudes towards people. The construction of a Thurstone scale is a fairly complex process that requires two stages. The first stage is a ranking operation, performed by judges who assign scale values to attitudinal statements. The second stage consists of asking subjects to respond to the attitudinal statements.

2.4 QUESTIONNAIRE DESIGN

Questionnaire is a systematic, data collection technique consists of a series of questions required to be answered by the respondents to identify their attitude, experience, and behavior towards the subject of research.

The following steps are involved in the questionnaire design process:

1. **Specify the Information Needed:** The first and the foremost step in designing the questionnaire is to specify the information needed from the respondents such that the objective

of the survey is fulfilled. The researcher must completely review the components of the problem, particularly the hypothesis, research questions, and the information needed.

2. **Define the Target Respondent:** At the very outset, the researcher must identify the target respondent from whom the information is to be collected. The questions must be designed keeping in mind the type of respondents under study. Such as, the questions that are appropriate for serviceman might not be appropriate for a businessman. The less diversified respondent group shall be selected because the more diversified the group is, the more difficult it will be to design a single questionnaire that is appropriate for the entire group.

3. **Specify the type of Interviewing Method:** The next step is to identify the way in which the respondents are reached. In personal interviews, the respondent is presented with a questionnaire and interacts face-to-face with the interviewer. Thus, lengthy, complex and varied questions can be asked using the personal interview method. In telephone interviews, the respondent is required to give answers to the questions over the telephone. Here the respondent cannot see the questionnaire and hence this method restricts the use of small, simple and precise questions.

   The questionnaire can be sent through mail or post. It should be self-explanatory and contain all the important information such that the respondent is able to understand every question and gives a complete response. The electronic questionnaires are sent directly to the mail ids of the respondents and are required to give answers online.

4. **Determine the Content of Individual Questions:** Once the information needed is specified and the interviewing methods are determined, the next step is to decide the content of the question. The researcher must decide on what should be included in the question such that it contributes to the information needed or serve some specific purpose.

   In some situations, the indirect questions which are not directly related to the information needed may be asked. It is useful to ask neutral questions at the beginning of a questionnaire with intent to establish respondent's involvement and rapport. This is mainly done when the subject of a questionnaire is sensitive or controversial. The researcher must try to avoid the use of double-barrelled questions. A question that talks about two issues simultaneously, such as Is the Real juice tasty and a refreshing health drink?

5. **Overcome Respondent's Inability and Unwillingness to Answer:** The researcher should not presume that the respondent can provide accurate responses to all the questions. He must attempt to overcome the respondent's inability to answer. The questions must be designed in a simple and easy language such that it is easily understood by each respondent. In situations, where the respondent is not at all informed about the topic of interest, then the researcher may ask the filter questions, an initial question asked in the questionnaire to identify the prospective respondents to ensure that they fulfil the requirements of the sample.

   Despite being able to answer the question, the respondent is unwilling to devote time in providing information. The researcher must attempt to understand the reason behind such unwillingness and design the questionnaire in such a way that it helps in retaining the respondent's attention.

6. **Decide on the Question Structure:** The researcher must decide on the structure of questions to be included in the questionnaire. The question can be structured or unstructured. The unstructured questions are the open-ended questions which are answered by the

respondents in their own words. These questions are also called as a free-response or free-answer questions.

While, the structured questions are called as closed-ended questions that pre-specify the response alternatives. These questions could be a multiple-choice question, dichotomous (yes or no) or a scale.

7. **Determine the Question Wording:** The desired question content and structure must be translated into words which are easily understood by the respondents. At this step, the researcher must translate the questions in easy words such that the information received from the respondents is similar to what was intended.

   In case the question is written poorly, then the respondent might refuse to answer it or might give a wrong answer. In case, the respondent is reluctant to give answers, then "nonresponse" arises which increases the complexity of data analysis. On the other hand, if the wrong information is given, then "response error" arises due to which the result is biased.

8. **Determine the Order of Questions:** At this step, the researcher must decide the sequence in which the questions are to be asked. The opening questions are crucial in establishing respondent's involvement and rapport, and therefore, these questions must be interesting, non-threatening and easy. Usually, the open-ended questions which ask respondents for their opinions are considered as good opening questions, because people like to express their opinions.

9. **Identify the Form and Layout:** The format, positioning and spacing of questions has a significant effect on the results. The layout of a questionnaire is specifically important for the self-administered questionnaires. The questionnaires must be divided into several parts, and each part shall be numbered accurately to clearly define the branches of a question.

10. **Reproduction of Questionnaire:** Here, we talk about the appearance of the questionnaire, i.e. the quality of paper on which the questionnaire is either written or printed. In case, the questionnaire is reproduced on a poor-quality paper; then the respondent might feel the research is unimportant due to which the quality of response gets adversely affected.

    Thus, it is recommended to reproduce the questionnaire on a good-quality paper having a professional appearance. In case, the questionnaire has several pages, then it should be presented in the form of a booklet rather than the sheets clipped or stapled together.

11. **Pretesting:** Pretesting means testing the questionnaires on a few selected respondents or a small sample of actual respondents with a purpose of improving the questionnaire by identifying and eliminating the potential problems. All the aspects of the questionnaire must be tested such as question content, structure, wording, sequence, form and layout, instructions, and question difficulty. The researcher must ensure that the respondents in the pre-test should be similar to those who are to be finally surveyed.

    Thus, the questionnaire design is a multistage process that requires the researcher's attention to many details.

2.5 SUMMARY

In this unit, we saw the attitude of measurement and process of designing a questionnaire. Attitude can be defined as a tendency to react favourably, neutrally, or unfavourably toward a particular class of stimuli, such as a custom, institutional practice, or national group. A questionnaire is a research instrument consisting of a series of questions for the purpose of gathering information from respondents. Questionnaires can be thought of as a kind of written interview. ... Often a questionnaire uses both open and closed questions to collect data.

Some of the key definitions in this section include:

Likert Scale: A measure of attitudes designed to allow respondents to rate how strongly they agree or disagree with carefully constructed statements, ranging from very positive to very negative attitudes toward some object.

Composite Scaling: A way of representing a latent construct by summing or averaging respondents' reactions to multiple items each assumed to indicate the latent construct.

Numeric Scaling: An attitude rating scale similar to a semantic differential except that it uses numbers, instead of verbal descriptions, as response options to identify response positions.

Graphic Rating Scale: A measure of attitude that allows respondents to rate an object by choosing any point along a graphic continuum.

2.6 QUESTIONS

1 What is attitude rating scale? Discuss any two scales using appropriate examples.
2 Discuss the structure of Likert scale.
3 What is understood by Numeric scaling?
4 Discuss in details the steps of Questionnaire Design.
5 Write a note on Graphic rating scale.

2.7 REFERENCES

1 Albright Winson, 2015, Business Analytics, 5th Edition,
2 Hair, 2014, Multivariate data Analysis, 7th Edition, Pearson Publication
3 Kabir, Syed Muhammad. (2016). Measurement Concepts: Variable, Reliability, Validity, and Norm.
4 Mark Saunders, 2011, Research Methods for Business Students, 5th Edition
5 Shefali Pandya, 2012, Research Methodology, APH Publishing Corporation, ISBN: 9788131316054, 813131605X
6 William G.Zikmund, B.J Babin, J.C. Carr, Atanu Adhikari, M.Griffin, 2016, Business Research Methods, Edition 8, Cengage Publications

M.Sc. Information Technology

SEMESTER I

PAPER 1 - RESEARCH IN COMPUTING

UNIT IV

MEASUREMENT CONCEPTS, SAMPLING AND FIELD WORK

C. SAMPLING DESIGNS AND PROCEDURES

D. DETERMINATION OF SAMPLE SIZE

CONTENT

3.1 OBJECTIVES

Following are the objectives of this unit:

- ✓ To get well versed with Sampling terminologies
- ✓ To understand the purpose of sampling
- ✓ To analyse the stages of sampling
- ✓ To understand the techniques of sampling
- ✓ To differentiate between different types of sampling techniques
- ✓ To calculate the sample size

3.2 INTRODUCTION

Sampling is a familiar part of daily life. A customer in a bookstore picks up a book, looks at the cover, and skims a few pages to get a sense of the writing style and content before deciding whether to buy. A high school student visits a college classroom to listen to a professor's lecture. Selecting a university on the basis of one classroom visit may not be scientific sampling, but in a personal situation, it may be a practical sampling experience. When measuring every item in a population is impossible, inconvenient, or too expensive, we intuitively take a sample. Although sampling is commonplace in daily activities, these familiar samples are seldom scientific. For researchers, the process of sampling can be quite complex. Sampling is a central aspect of business research, requiring in-depth examination.

Formally defining sampling: It is a process used in statistical analysis in which a predetermined number of observations are taken from a larger population. The methodology used to sample from a larger population depends on the type of analysis being performed, but it may include simple random sampling or systematic sampling.

To choose the right size of sample is the next section of this unit titled as determination of sample size. Sample size determination is the act of choosing the number of observations or replicates to include in a statistical sample. The sample size is an important feature of any empirical study in which the goal is to make inferences about a population from a sample.

This unit explains the nature of sampling and ways to determine the appropriate sample design.

3.3 SAMPLING TERMINOLOGIES

- ✓ **Population**: Total of items about which information is desired. It can be classified into two categories- finite and infinite. The population is said to be finite if it consists of a fixed number of elements so that it is possible to enumerate in its totality.
  Examples of finite population are the populations of a city, the number of workers in a factory, etc.

  An infinite population is that population in which it is theoretically impossible to observe all the elements. In an infinite population the number of items is infinite.

  Example of infinite population is the number of stars in sky. From practical consideration, we use the term infinite population for a population that cannot be enumerated in a reasonable period of time.

- ✓ **Sample**: It is part of the population that represents the characteristics of the population.

- ✓ **Population Sample Sampling:** It is the process of selecting the sample for estimating the population characteristics. In other words, it is the process of obtaining information about an entire population by examining only a part of it.

- ✓ **Sampling Unit:** Elementary units or group of such units which besides being clearly defined, identifiable and observable, are convenient for purpose of sampling are called sampling units. For instance, in a family budget enquiry, usually a family is considered as the sampling unit since it is found to be convenient for sampling and for ascertaining the required information. In a crop survey, a farm or a group of farms owned or operated by a household may be considered as the sampling unit.

- ✓ **Sampling Frame:** A list containing all sampling units is known as sampling frame. Sampling frame consists of a list of items from which the sample is to be drawn. Sample Survey: An investigation in which elaborate information is collected on a sample basis is known as sample survey.

- ✓ **Statistic**: Characteristics of the sample. For example, sample Mean, proportion, etc.

- ✓ **Parameter**: Characteristics of the population. For example, population Mean, proportion, etc.

- ✓ **Target Population:** A target population is the entire group about which information is desired and conclusion is made.

- ✓ **Sampled Population**: The population, which we actually sample, is the sampled population. It is also called survey population.

- ✓ **Sampling With and Without Replacement:** Sampling schemes may be without replacement ('WOR' - no element can be selected more than once in the same sample) or with replacement ('WR' - an element may appear multiple times in the one sample). For example, if we catch fish, measure them, and immediately return them to the water before continuing with the sample, this is a WR design, because we might end up catching and measuring the same fish more than once. However, if we do not return the fish to the water (e.g. if we eat the fish), this becomes a WOR design.

- ✓ **Sample Design:** Sample design refers to the plans and methods to be followed in selecting sample from the target population and the estimation technique formula for computing the sample statistics. These statistics are the estimates used to infer the population parameters.

3.4 PURPOSE OF SAMPLING

The basic purpose of sampling is to provide an estimate of the population parameter and to test the hypothesis. Advantages of sampling are –

- ✓ Save time and money
- ✓ Enable collection of comprehensive data
- ✓ Enable more accurate measurement as it conducted by trained and experienced investigators
- ✓ Sampling remains the only way when population contains infinitely many members
- ✓ In certain situation, sampling is the only way of data collection. For example, in testing the pathological status of blood, boiling status of rice, etc
- ✓ It provides a valid estimation of sampling error

3.5 STAGES OF SAMPLING PROCESS

The sampling process comprises several stages

1. Define the population.
2. Specifying the sampling frame.
3. Specifying the sampling unit.
4. Selection of the sampling method.
5. Determination of sample size.
6. Specifying the sampling plan.
7. Selecting the sample.

Define the Population: Population must be defined in terms of elements, sampling units, extent and time. Because there is very rarely enough time or money to gather information from everyone or everything in a population, the goal becomes finding a representative sample (or subset) of that population.

Sampling Frame: As a remedy, we seek a sampling frame which has the property that we can identify every single element and include any in our sample. The most straightforward type of frame is a list of elements of the population (preferably the entire population) with appropriate contact information. A sampling frame may be a telephone book, a city directory, an employee roster, a listing of all students attending a university, or a list of all possible phone numbers.

Sampling Unit: A sampling unit is a basic unit that contains a single element or a group of elements of the population to be sampled. The sampling unit selected is often dependent upon the sampling frame. If a relatively complete and accurate listing of elements is available (e.g. register of purchasing agents) one may well want to sample them directly. If no such register is available, one may need to sample companies as the basic sampling unit.

Sampling Method: The sampling method outlines the way in which the sample units are to be selected. The choice of the sampling method is influenced by the objectives of the research, availability of financial resources, time constraints, and the nature of the problem to be investigated. All sampling methods can be grouped under two distinct heads, that is, probability and non-probability sampling.

Sample Size: The sample size calculation depends primarily on the type of sampling designs used. However, for all sampling designs, the estimates for the expected sample characteristics (e.g. mean, proportion or total) desired level of certainty, and the level of precision must be clearly specified in advanced. The statement of the precision desired might be made by giving the amount of error that we are willing to tolerate in the resulting estimates. Common levels of precisions are 5% and 10%.

Sampling Plan: In this step, the specifications and decisions regarding the implementation of the research process are outlined. As the interviewers and their co-workers will be on field duty of most of the time, a proper specification of the sampling plans would make their work easy and they would not have to reverting operational problems.

Select the Sample: The final step in the sampling process is the actual selection of the sample elements. This requires a substantial amount of office and fieldwork, particularly if personal interviews are involved.

3.6 TECHNIQUES OF SAMPLING

There are two basic approaches to sampling: Probability Sampling and Non-probability Sampling.

3.6.1 PROBABILITY SAMPLING

Probability sampling is also known as random sampling or chance sampling. In this, sample is taken in such a manner that each and every unit of the population has an equal and positive chance of being selected. In this way, it is ensured that the sample would truly represent the overall population. Probability sampling can be achieved by random selection of the sample among all the units of the population.

Major random sampling procedures are –

- ✓ Simple Random Sample
- ✓ Systematic Random Sample
- ✓ Stratified Random Sample
- ✓ Cluster/ Multistage Sample

3.6.1.1 SIMPLE RANDOM SAMPLE

For this, each member of the population is numbered. Then, a given size of the sample is drawn with the help of a random number chart. The other way is to do a lottery. Write all the numbers on small, uniform pieces of paper, fold the papers, put them in a container and take out the required lot in a random manner from the container as is done in the kitty parties. It is relatively simple to implement but the final sample may miss out small sub groups.

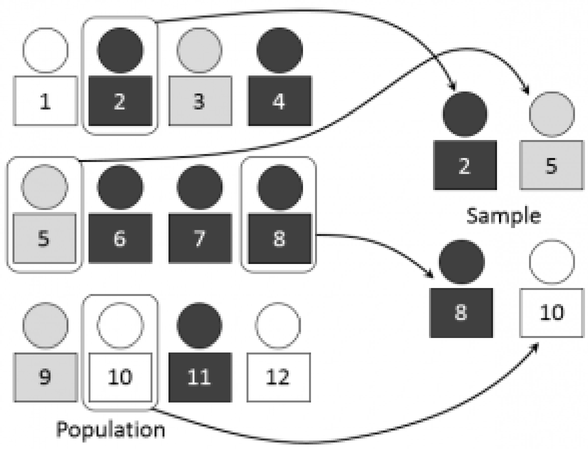


Figure 3.1: Example of Simple Random Sampling

A simple random sample is chosen in such a way that every set of individuals has an equal chance to be in the selected sample.
For example, if you wanted to study all the adults in the India who had high cholesterol, the list would be practically impossible to get unless you surveyed every person in the country. Therefore, other sampling methods would probably be better suited to that particular experiment.

Advantages:
The sample will be free from Bias (i.e. it's random!).

Disadvantages:
Difficult to obtain

Due to its very randomness, "freak" results can sometimes be obtained that are not representative of the population. In addition, these freak results may be difficult to spot. Increasing the sample size is the best way to eradicate this problem.

3.6.1.2. SYSTEMATIC RANDOM SAMPLE

Systematic sampling is one method in the broader category of random sampling (for this reason, it requires precise control of the sampling frame of selectable individuals and of the probability that they will be selected). It involves choosing a first individual at random from the population, then selecting every following nth individual within the sampling frame to make up the sample.

Systematic sampling is a very simple process that requires choosing only one individual at random. The rest of the process is fast and easy. As with simple random sampling, the results that we obtain are representative of the population, provided that there is no factor intrinsic to the individuals selected that regularly repeats certain characteristics of the population every certain number of individuals—which is very rarely the case.

The *process* for conducting systematic sampling is as follows:

1. We prepare an ordered list of N individuals in the population; this will be our sampling frame.

2. We divide the sampling frame into n fragments, where n is our desired sample size. The size of these fragments will be

K=N/n

where K is called the interval or lift coefficient.

3. The initial number: we randomly obtain a whole number A, which is less than or equal to the interval. The number corresponds to the first subject who we select for the sample within the first fragment into which we have divided the population.

4. Selection of the remaining n-1 individuals: We select the subsequent individuals based on where they fall, in simple arithmetic succession, after the randomly selected individual, selecting individuals who occupy the same position as the initial subject in the rest of the fragment into which we have divided the sample. This is the equivalent of saying that we will select the individuals

A, A + K, A + 2K, A + 3K, ..., A + (n-1)K

Example
Let’s say that our sampling frame includes 5,000 individuals and we want a 100-individual sample. First, we divide the sampling frame into 100 fragments of 50 individuals. Then, we randomly select one number between 1 and 50 in order to randomly select the first individual for the first fragment. Let's say we select 24. The sample is defined by this individual; we select the remaining individuals from the list at intervals of 50 units: 24, 74, 124, 174, ..., 4.974

It also requires numbering the entire population. Then every nth number (say every 5th or 10th number, as the case may be) is selected to constitute the sample. It is easier and more likely to represent different subgroups.

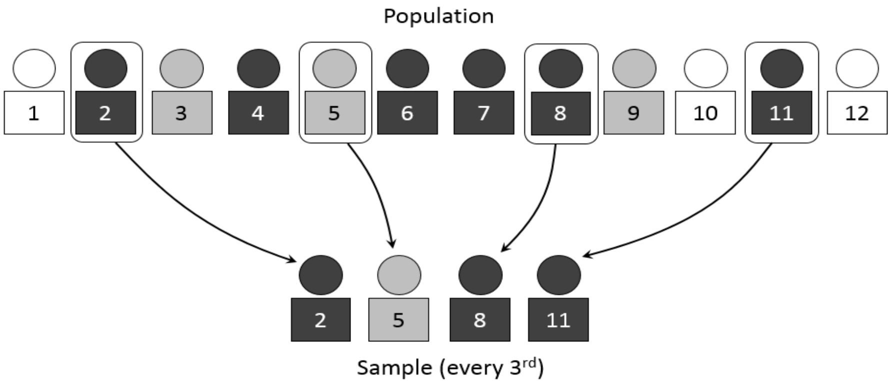


Figure 3.2: Example of Systematic Random Sampling

Advantages
Can eliminate other sources of bias.

Disadvantages
Can introduce bias where the pattern used for the samples coincides with a pattern in the population.

3.6.1.3 STRATIFIED RANDOM SAMPLE
At first, the population is first divided into groups or strata each of which is homogeneous with respect to the given characteristic feature. From each strata, then, samples are drawn at random. This is called stratified random sampling.

For example, with respect to the level of socio-economic status, the population may first be grouped in such strata as high, middle, low and very low socio-economic levels as per pre-determined criteria, and random sample drawn from each group.

The sample size for each sub-group can be fixed to get representative sample. This way, it is possible that different categories in the population are fairly represented in the sample, which could have been left out otherwise in simple random sample.

As with stratified samples, the population is broken down into different categories. However, the size of the sample of each category does not reflect the population as a whole. The Quota sampling technique can be used where an unrepresentative sample is desirable (e.g. you might want to interview more children than adults for a survey on computer games), or where it would be too difficult to undertake a stratified sample.

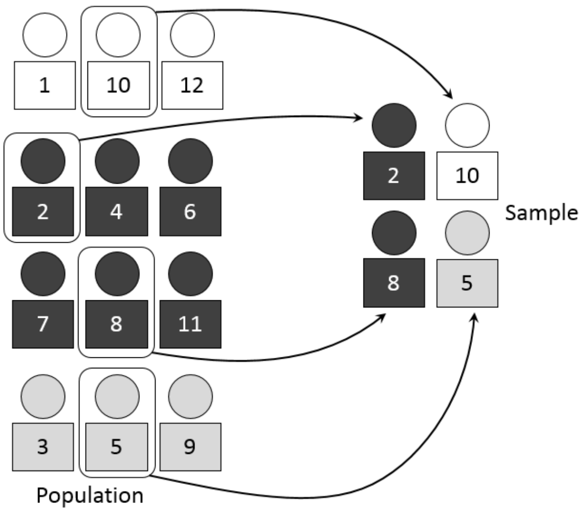


Figure 3.3: Example of Stratified Random Sampling

Advantages
Yields more accurate results than simple random sampling
Can show different tendencies within each category (e.g. men and women)

Disadvantage
Nil

3.6.1.4 CLUSTER/ MULTISTAGE SAMPLING

In some cases, the selection of units may pass through various stages, before you finally reach your sample of study. For this, a State, for example, may be divided into districts, districts into blocks, blocks into villages, and villages into identifiable groups of people, and then taking the random or quota sample from each group.

For example, taking a random selection of 3 out of 15 districts of a State, 6 blocks from each selected district, 10 villages from each selected block and 20 households from each selected village, totalling 3600 respondents. This design is used for large-scale surveys spread over large areas.

The advantage is that it needs detailed sampling frame for selected clusters only rather than for the entire target area. There are savings in travel costs and time as well. However, there is a risk of missing on important sub-groups and not having complete representation of the target population.

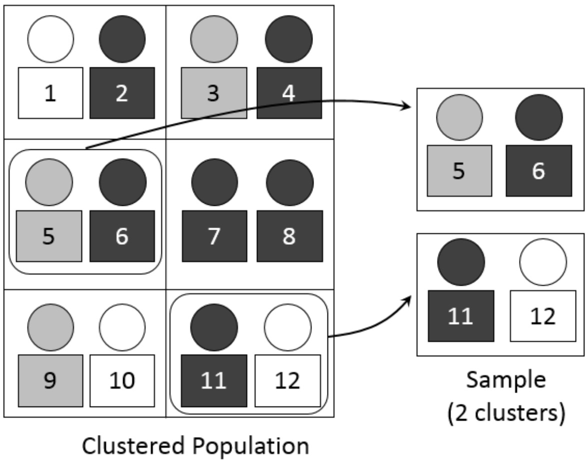


Figure 3.4: Example of Cluster Sampling

Advantages
Less expensive and time consuming than a fully random sample.
Can show 'regional' variations.

Disadvantages
Not a genuine random sample.
Likely to yield a biased result (especially if only a few clusters are sampled).

3.6.2 NON-PROBABILITY SAMPLING

Non-probability sampling is any sampling method where some elements of the population have no chance of selection (these are sometimes referred to as 'out of coverage'/'under covered'), or where the probability of selection can't be accurately determined. It involves the selection of elements based on assumptions regarding the population of interest, which forms the criteria for selection.

Hence, because the selection of elements is non-random, non-probability sampling does not allow the estimation of sampling errors. Non-probability sampling is a non-random and subjective method of sampling where the selection of the population elements comprising the sample depends on the personal judgment or the discretion of the sampler.

Non-probability sampling includes:

- ✓ Accidental/ Convenience Sampling
- ✓ Quota Sampling

- ✓ Judgment/ Subjective/ Purposive Sampling
- ✓ Snowball Sampling

3.6.2.1 CONVENIENCE/ ACCIDENTAL SAMPLING

Accidental sampling is a type of non-probability sampling which involves the sample being drawn from that part of the population which is close to hand. That is, a sample population selected because it is readily available and convenient.

The researcher using such a sample cannot scientifically make generalizations about the total population from this sample because it would not be representative enough.

For example, if the interviewer was to conduct such a survey at a shopping center early in the morning on a given day, the people that s/he could interview would be limited to those given there at that given time, which would not represent the views of other members of society in such an area, if the survey was to be conducted at different times of day and several times per week. This type of sampling is most useful for pilot testing.

The primary problem with availability sampling is that you can never be certain what population the participants in the study represent. The population is unknown, the method for selecting cases is haphazard, and the cases studied probably don't represent any population you could come up with. However, there are some situations in which this kind of design has advantages.

For example, survey designers often want to have some people respond to their survey before it is given out in the 'real' research setting as a way of making certain the questions make sense to respondents. For this purpose, availability sampling is not a bad way to get a group to take a survey, though in this case researchers care less about the specific responses given than whether the instrument is confusing or makes people feel bad.

Advantages

Expedited data collection: When time is of the essence, many researchers turn to convenience sampling for data collection, as they can swiftly gather data and begin their calculations.

Ease of research: For researchers who are not looking for an accurate sampling, they can simply collect their information and move on to other aspects of their study. This type of sampling can be done by simply creating a questionnaire and distributing it to their targeted group.

Ready availability: Since most convenience sampling is collected with the populations on hand, the data is readily available for the researcher to collect.

Cost effectiveness: One of the most important aspects of convenience sampling is its cost effectiveness. This method allows for funds to be distributed to other aspects of the project.

Disadvantages

Bias: The results of the convenience sampling cannot be generalized to the target population because of the potential bias of the sampling technique due to under-representation of subgroups in the sample in comparison to the population of interest.

Power: Convenience sampling is characterized with insufficient power to identify differences of population subgroups

3.6.2.2 QUOTA SAMPLING

In quota sampling, the population is first segmented into mutually exclusive sub-groups, just as in stratified sampling. Then judgment is used to select the subjects or units from each segment based on a specified proportion. For example, an interviewer may be told to sample 200 females and 300 males between the age of 45 and 60. In quota sampling the selection of the sample is non-random.

For example: interviewers might be tempted to interview those who look most helpful. The problem is that these samples may be biased because not everyone gets a chance of selection. This random element is its greatest weakness and quota versus probability has been a matter of controversy for many years.

Process

In quota sampling, a population is first segmented into mutually exclusive sub-groups, just as in stratified sampling. Then judgment is used to select the subjects or units from each segment based on a specified proportion. For example, an interviewer may be told to sample 200 females and 300 males between the age of 45 and 60. This means that individuals can put a demand on who they want to sample (targeting).

This second step makes the technique non-probability sampling. In quota sampling, there is non-random sample selection and this can be unreliable. For example, interviewers might be tempted to interview those people in the street who look most helpful, or may choose to use accidental sampling to question those closest to them, to save time. The problem is that these samples may be biased because not everyone gets a chance of selection, whereas in stratified sampling (its probabilistic version), the chance of any unit of the population is the same as 1/n (n= number of units in the population). This non-random element is a source of uncertainty about the nature of the actual sample and quota versus probability has been a matter of controversy for many years.

3.6.2.3 SUBJECTIVE OR PURPOSIVE OR JUDGMENT SAMPLING

In this sampling, the sample is selected with definite purpose in view and the choice of the sampling units depends entirely on the discretion and judgment of the investigator. This sampling suffers from drawbacks of favouritism and nepotism depending upon the beliefs and prejudices of the investigator and thus does not give a representative sample of the population. This sampling method is seldom used and cannot be recommended for general use since it is often biased due to element of subjectivity on the part of the investigator. However, if the investigator is experienced and skilled and this sampling is carefully applied, then judgment samples may yield valuable results. Some purposive sampling strategies that can be used in qualitative studies are given below. Each strategy serves a particular data gathering and analysis purpose. Extreme Case Sampling: It focuses on cases that are rich in information because they are unusual or special in some way. e.g. the only community in a region that prohibits felling of trees.

3.6.2.4 SNOWBALL SAMPLING

Snowball sampling is a method in which a researcher identifies one member of some population of interest, speaks to him/her, and then asks that person to identify others in the population that the researcher might speak to. This person is then asked to refer the researcher to yet another person, and so on.

This sampling technique is used against low incidence or rare populations. Sampling is a big problem in this case, as the defined population from which the sample can be drawn is not available. Therefore, the process sampling depends on the chain system of referrals. Although small sample sizes and low costs are the clear advantages of snowball sampling, bias is one of its disadvantages.

The referral names obtained from those sampled in the initial stages may be similar to those initially sampled. Therefore, the sample may not represent a cross-section of the total population. It may also happen that visitors to the site or interviewers may refuse to disclose the names of those whom they know.

Snowball sampling uses a small pool of initial informants to nominate, through their social networks, other participants who meet the eligibility criteria and could potentially contribute to a specific study. The term "snowball sampling" reflects an analogy to a snowball increasing in size as it rolls downhill.

Method

- ✓ Draft a participation program (likely to be subject to change, but indicative).
- ✓ Approach stakeholders and ask for contacts.
- ✓ Gain contacts and ask them to participate.
- ✓ Community issues groups may emerge that can be included in the participation program.
- ✓ Continue the snowballing with contacts to gain more stakeholders if necessary.
- ✓ Ensure a diversity of contacts by widening the profile of persons involved in the snowballing exercise.

Advantages

- ✓ Locate hidden populations: It is possible for the surveyors to include people in the survey that they would not have known but, through the use of social network.
- ✓ Locating people of a specific population: There are no lists or other obvious sources for locating members of the population.

- ✓ Methodology: As subjects are used to locate the hidden population, the researcher invests less money and time in sampling. Snowball sampling method does not require complex planning and the staffing required is considerably smaller in comparison to other sampling methods.

Disadvantages

- ✓ Community bias: The first participants will have a strong impact on the sample.
- ✓ Non-random: Snowball sampling contravenes many of the assumptions supporting conventional notions of random selection and representativeness.
- ✓ Unknown sampling population size: There is no way to know the total size of the overall population.
- ✓ Anchoring: Lack of definite knowledge as to whether or not the sample is an accurate reading of the target population.
- ✓ Lack of control over sampling method: As the subjects locate the hidden population, the research has very little control over the sampling method, which becomes mainly dependent on the original and subsequent subjects, who may add to the known sampling pool using a method outside of the researcher's control.

***(Some more types of Non- Probability Sampling Methods are listed after the references of this chapter to get a through understanding of this topic)**

3.7 DETERMINATION OF SAMPLE SIZE

Determination of sample size is probably one of the most important phases in the sampling process. Generally the larger the sample size, the better is the estimation. But always larger sample sizes cannot be used in view of time and budget constraints. Moreover, when a probability sample reaches a certain size the precision of an estimator cannot be significantly increased by increasing the sample size any further. Indeed, for a large population the precision of an estimator depends on the sample size, not on what proportion of the population has been sampled. It can be stated that whenever a sample study is made, there arises some sampling error which can be controlled by selecting a sample of adequate size.

Sample size is a frequently-used term in statistics and market research, and one that inevitably comes up whenever you're surveying a large population of respondents. It relates to the way research is conducted on large populations.

When you survey a large population of respondents, you're interested in the entire group, but it's not realistically possible to get answers or results from absolutely everyone. So you take a random sample of individuals which represents the population as a whole.

The size of the sample is very important for getting accurate, statistically significant results and running your study successfully.

If your sample is too small, you may include a disproportionate number of individuals which are outliers and anomalies. These skew the results and you don't get a fair picture of the whole population.

If the sample is too big, the whole study becomes complex, expensive and time-consuming to run, and although the results are more accurate, the benefits don't outweigh the costs.
If you've already worked out your variables you can get to the right sample size quickly with the online sample size calculator below:

Confidence Level:

95%
Population Size:
10000
Margin of Error:

5%
Ideal Sample Size:
370
If you want to start from scratch in determining the right sample size for your market research, let us walk you through the steps.

Learn how to determine sample size
To choose the correct sample size, you need to consider a few different factors that affect your research, and gain a basic understanding of the statistics involved. You'll then be able to use a sample size formula to bring everything together and sample confidently, knowing that there is a high probability that your survey is statistically accurate.

The steps that follow are suitable for finding a sample size for continuous data – i.e. data that is counted numerically. It doesn't apply to categorical data – i.e. put into categories like green, blue, male, female etc.

STAGE 1: Consider your sample size variables
Before you can calculate a sample size, you need to determine a few things about the target population and the level of accuracy you need:

1. Population size
How many people are you talking about in total? To find this out, you need to be clear about who does and doesn't fit into your group. For example, if you want to know about dog owners, you'll include everyone who has at some point owned at least one dog. (You may include or exclude those who owned a dog in the past, depending on your research goals.) Don't worry if you're unable to calculate the exact number. It's common to have an unknown number or an estimated range.

2. Margin of error (confidence interval)
Errors are inevitable – the question is how much error you'll allow. The margin of error, AKA confidence interval, is expressed in terms of mean numbers. You can set how much difference you'll allow between the mean number of your sample and the mean number of your

population. If you've ever seen a political poll on the news, you've seen a confidence interval and how it's expressed. It will look something like this: "68% of voters said yes to Proposition Z, with a margin of error of +/- 5%."

3. Confidence level
This is a separate step to the similarly-named confidence interval in step 2. It deals with how confident you want to be that the actual mean falls within your margin of error. The most common confidence intervals are 90% confident, 95% confident, and 99% confident.

4. Standard deviation
This step asks you to estimate how much the responses you receive will vary from each other and from the mean number. A low standard deviation means that all the values will be clustered around the mean number, whereas a high standard deviation means they are spread out across a much wider range with very small and very large outlying figures. Since you haven't yet run your survey, a safe choice is a standard deviation of .5 which will help make sure your sample size is large enough.

STAGE 2: Calculate sample size
Now that you've got answers for steps 1 – 4, you're ready to calculate the sample size you need. This can be done using an online sample size calculator or with paper and pencil.

5. Find your Z-score
Next, you need to turn your confidence level into a Z-score. Here are the Z-scores for the most common confidence levels:
90% – Z Score = 1.645
95% – Z Score = 1.96
99% – Z Score = 2.576

6. Use the sample size formula
Plug in your Z-score, standard of deviation, and confidence interval into the sample size calculator or use this sample size formula to work it out yourself:

Necessary Sample Size =

$$\frac{(Z - score)^2 \times StdDev \times (1 - StdDev)}{(Margin\,of\;error)^2}$$

This equation is for an unknown population size or a very large population size. If your population is smaller and known, just use the sample size calculator.

Solved example:
assuming you chose a 95% confidence level, 0.5 standard deviation, and a margin of error (confidence interval) of +/- 5%.

$= ((1.96)^2 \text{ x } .5(.5)) / (.05)^2$

= (3.8416 x .25) / .0025

= 0.9604 / .0025

= 384.16

385 respondents are needed

3.8 SUMMARY

In this unit we saw sampling which is a process used in statistical analysis in which a predetermined number of observations are taken from a larger population. The methodology used to sample from a larger population depends on the type of analysis being performed, but it may include simple random sampling or systematic sampling.

We also saw the terminologies used along with sampling which includes population, sample, parameter, target population, sampled population etc. We further analysed the purpose of sampling along with understating the stages of sampling which included 7 steps. Later, in this unit, we saw the techniques of sampling (Probability Sampling and non-Probability Sampling techniques)

Later we understood the concept of determining the size of sample. To choose the right size of sample is the next section of this unit titled as determination of sample size. Sample size determination is the act of choosing the number of observations or replicates to include in a statistical sample. The sample size is an important feature of any empirical study in which the goal is to make inferences about a population from a sample.

3.9 QUESTIONS

3 Write a note on purpose of sampling.
4 Discuss the seven stages of sampling.
5 What are sampling techniques? Why do we need them?
6 Elaborate on Probability sampling and state its 3 types.
7 Elaborate on Non-Probability sampling and state its 3 types.
8 Why do we need to determine a sample size? Justify your answer.
9 Explain with appropriate example, snowball sampling.
10 Explain with appropriate figure and example, simple random sampling.
11 Discuss the steps for determining of the sample size.

3.10 REFERENCES

1 Albright Winson, 2015, Business Analytics, 5th Edition,
2 Hair, 2014, Multivariate data Analysis, 7th Edition, Pearson Publication

3 Kabir, Syed Muhammad. (2016). Measurement Concepts: Variable, Reliability, Validity, and Norm.
4 Mark Saunders, 2011, Research Methods for Business Students, 5th Edition
5 Shefali Pandya, 2012, Research Methodology, APH Publishing Corporation, ISBN: 9788131316054, 813131605X
6 William G.Zikmund, B.J Babin, J.C. Carr, Atanu Adhikari, M.Griffin, 2016, Business Research Methods, Edition 8, Cengage Publications
7 Images: https://www.wikiwand.com/en/Sampling_(statistics) creative commons

---

*** Some more types of Non-Probability Sampling Techniques**

3.6.2.5 MAXIMUM VARIATION SAMPLING
Aims at capturing the central themes that cut across participant variations. e.g. persons of different age, gender, religion and marital status in an area protesting against child marriage. Homogeneous Sampling: Picks up a small sample with similar characteristics to describe some particular sub-group in depth. e.g. firewood cutters or snake charmers or bonded laborers.

3.6.2.6 TYPICAL CASE SAMPLING
Uses one or more typical cases (individuals, families / households) to provide a local profile. The typical cases are carefully selected with the co-operation of the local people/ extension workers.

3.6.2.7 CRITICAL CASE SAMPLING
Looks for critical cases that can make a point quite dramatically. e.g. farmers who have set up an unusually high yield record of a crop. Chain Sampling: Begins by asking people, 'who knows a lot about ________'. By asking a number of people, you can identify specific kinds of cases e.g. critical, typical, extreme etc.

3.6.2.8 CRITERION SAMPLING
Reviews and studies cases that meet some pre-set criterion of importance e.g. farming households where women take the decisions. In short, purposive sampling is best used with small numbers of individuals/groups which may well be sufficient for understanding human perceptions, problems, needs, behaviours and contexts, which are the main justification for a qualitative audience research.

3.6.2.9 MATCHED RANDOM SAMPLING
A method of assigning participants to groups in which pairs of participants are first matched on some characteristic and then individually assigned randomly to groups. The Procedure for Matched random sampling can be briefed with the following contexts- (a) Two samples in which the members are clearly paired, or are matched explicitly by the researcher. For example, IQ measurements or pairs of identical twins. (b) Those samples in which the same attribute, or variable, is measured twice on each subject, under different circumstances. Commonly called repeated measures.

3.6.2.10 MECHANICAL SAMPLING

Mechanical sampling is typically used in sampling solids, liquids and gases, using devices such as grabs, scoops; thief probes etc. Care is needed in ensuring that the sample is representative of the frame.

3.6.2.11 LINE-INTERCEPT SAMPLING

Line-intercept sampling is a method of sampling elements in a region whereby an element is sampled if a chosen line segment, called a 'transect', intersects the element.

3.6.2.12 PANEL SAMPLING

Panel sampling is the method of first selecting a group of participants through a random sampling method and then asking that group for the same information again several times over a period of time. Therefore, each participant is given the same survey or interview at two or more time points; each period of data collection is called a 'wave'. This sampling methodology is often chosen for large scale or nation-wide studies in order to gauge changes in the population with regard to any number of variables from chronic illness to job stress to weekly food expenditures. Panel sampling can also be used to inform researchers about within-person health changes due to age or help explain changes in continuous dependent variables such as spousal interaction.

3.6.2.13 RANK SAMPLING

A non-probability sample is drawn and ranked. The highest value is chosen as the first value of the targeted sample. Another sample is drawn and ranked; the second highest value is chosen for the targeted sample. The process is repeated until the lowest value of the targeted sample is chosen. This sampling method can be used in forestry to measure the average diameter of the trees.

3.6.2.14 VOLUNTARY SAMPLE

A voluntary sample is made up of people who self-select into the survey. Often, these folks have a strong interest in the main topic of the survey. Suppose, for example, that a news show asks viewers to participate in an on-line poll. This would be a volunteer sample. The sample is chosen by the viewers, not by the survey administrator.