

## UNIT II

### LESSON

# 2

## QUESTIONNAIRE AND SAMPLING

### LESSON OUTLINE

- ❖ **Meaning of questionnaire.**
- ❖ **Drafting of questionnaire.**
- ❖ **Size of questions**
- ❖ **Clarity of questions**
- ❖ **Logical sequence of questions**
- ❖ **Simple meaning questions**
- ❖ **Other requirements of a good questionnaire**
- ❖ **Meaning and essentials of sampling.**

### LEARNING OBJECTIVES:

- ❖ **After reading this lesson you**
- ❖ **should be able to**
- ❖ **Understand the meaning of questionnaire**
- ❖ **Different requirements and characteristics of a good questionnaire**
- ❖ **Meaning of sampling**
- ❖ **Essentials of sampling**



### **Introduction:**

Nowadays questionnaire is widely used for data collection in social research. It is a reasonably fair tool for gathering data from large, diverse, varied and scattered social groups. The questionnaire is the media of communication between the investigator and the respondents. According to Bogardus a questionnaire is a list of questions sent to a number of persons for their answers and which obtains standardized results that can be tabulated and treated statistically. The Dictionary of Statistical Terms defines it as a “ group of or sequence of questions designed to elicit information upon a subject or sequence of subjects from an information.” A questionnaire should be designed or drafted with utmost care and caution so that all the relevant and essential information for the enquiry may be collected without any difficulty, ambiguity and vagueness. Drafting of a good questionnaire is a highly specialized job and requires great care skill, wisdom, efficiency and experience. No hard and fast rule can be laid down for designing or framing a questionnaire. However, in this connection, the following general points may be borne in mind:

**1. Size of the questionnaire should be small:** A researcher should try his best to keep the number of the questions as small as possible, keeping in view the nature, objectives and scope of the enquiry. Respondent’s time should not be wasted by asking irrelevant and unimportant questions. A large number of questions would involve more work for the investigator and thus result in delay on his part in collecting and submitting the information. A large number of unnecessary questions may annoy the respondent and he may refuse to cooperate. A reasonable questionnaire should contain from 15 to 25 questions at large. If a still larger number of questions is a must in any enquiry, then the questionnaire should be divided into various sections or parts.

**2. The questions should be clear:** The questions should be easier, brief, unambiguous, non-offending, courteous in tone, corroborative in nature and to the point so that much scope of guessing is left on the part of the respondents.

**3. The questions should be arranged in a logical sequence:** Logical arrangement of questions reduces lot of unnecessary work on the part of the researcher because it not only facilitates the tabulation work but does not leave any chance for omissions or commissions. For example, to find if a person owns a television the logical order of questions would be: Do you own a television? When did you buy it? What is its make? How much did it cost you? Is its performance satisfactory? Have you ever got it serviced?

**4. Questions should be simple to understand:** The vague words like good, bad, efficient, sufficient, prosperity, rarely, frequently, reasonable, poor, rich etc., should not be used since these may be interpreted differently by different persons and as such might give unreliable and misleading information. Similarly the use of words having double meaning like price, assets, capital income etc., should also be avoided.

**5. Questions should be comprehensive and easily answerable:** Questions should be so designed that they are readily comprehensible and easy to answer for the respondents. They should not be tedious nor should they tax the respondents' memory. At the same time questions involving mathematical calculations like percentages, ratios etc., should not be asked.

**6. Questions of personal nature and sensitive should not be asked:** There are some questions which disturb the respondents and he may be shy or irritated by hearing such questions. Therefore, every effort should be made to avoid such questions. For example, do you cook yourself or your wife cooks? Or do you drink? Such questions will certainly irk the respondents and thus be avoided at any cost. If unavoidable then highest amount of politeness should be used

**7. Types of questions:** Under this head, the questions in the questionnaire may be classified as follows:

**(a) Shut questions:** Shut questions are those where possible answers are suggested by the framers of the questionnaire and the respondent is required to tick one of them. Shut questions can further be subdivided into the following forms:

**(i) Simple alternate questions:** In this type of questions the respondent has to choose from the two clear cut alternatives like ‘Yes’ or ‘No’ ‘Right or Wrong’ etc. Such questions are also called *dichotomous questions*. This technique can be applied with elegance to situations where two clear cut alternatives exist.

**(ii) Multiple choice questions:** Many a times it becomes difficult to define a clear cut alternative and accordingly in such a situation either the first method is not used or additional answers between Yes and No like Do not know, No opinion, Occasionally, Casually, Seldom etc. are added. For example, in order to find if a person smokes or drinks, the following multiple choice answers may be used:

Do you smoke?

- (a) Yes regularly    [   ]    (b) No never        [   ]  
(c) Occasionally    [   ]    (d) Seldom         [   ]

Multiple choice questions are very easy and convenient for the respondents to answer. Such questions save time and also facilitate tabulation. This method should be used if only a selected few alternative answers exist to a particular question.

**8. Leading questions should be avoided:** Questions like ‘Why do you use a particular type of car, say Maruti car’ should preferably be framed into two questions-

(i) Which car do you use?

(ii) Why do you prefer it?

It gives smooth ride [ ]

It gives more mileage [ ]

It is cheaper [ ]

It is maintenance free [ ]

**9 Cross Checks:** The questionnaire should be so designed as to provide internal checks on the accuracy of the information supplied by the respondents by including some connected questions at least with respect to matters which are fundamental to the enquiry.

**10 Pre testing the questionnaire:** It would be practical in every sense to try out the questionnaire on a small scale before using it for the given enquiry on a large scale. This has been found extremely useful in practice. The given questionnaire can be improved or modified in the light of the drawbacks, shortcomings and problems faced by the investigator in the pre test.

**11 A covering letter:** A covering letter from the organizers of the enquiry should be enclosed along with the questionnaire for the purposes of – regarding definitions, units, concepts used in the questionnaire, for taking the respondent's confidence, self addressed envelop in case of mailed questionnaire, mention about award or incentives for the quick response, a promise to send a copy of the survey report etc.

## **SAMPLING**

Though sampling is not new but the sampling theory has been developed recently. People knew or not but they have been using the sampling technique in their day to day life. For example a house wife tests a small quantity of rice to see whether it has been well-cooked and give the generalized result about the whole rice boiling in the vessel. The result arrived at is most of the times 100%

correct. In another example, when a doctor wants to examine the blood for any deficiency, takes only a few drops of blood of the patient and examines. The result arrived at is most of the times correct and represent the whole amount of blood available in the body of the patient. In all these cases, by inspecting a few, they simply believe that the samples give a correct idea about the population. Most of our decision are based on the examination of a few items only i.e. sample studies. In the words of Croxton and Cowdon, "It may be too expensive or too time consuming to attempt either a complete or a nearly complete coverage in a statistical study. Further to arrive at valid conclusions, it may not be necessary to enumerate all or nearly all of a population. We may study a sample drawn from the large population and, if that sample is adequately representative of the population, we should be able to arrive at valid conclusions."

According to Rosander, "The sample has many advantages over a census or complete enumeration. If carefully designed, the sample is not only considerably cheaper; but may give results which are just accurate and sometimes more accurate than those of a census. Hence a carefully designed sample may actually be better than a poorly planned and executed census."

**Merits:**

- 1. It saves time:** Sampling method of data collection saves time because fewer items are collected and processed. When the results are urgently required, this method is very helpful.
- 2. It reduces cost:** Since only a few and selected items are studied in sampling, so there is reduction in cost of money and reduction in terms of man hours.
- 3. More reliable results can be obtained:** Through sampling more reliable results can be obtained because (a) there are fewer chances of sampling

statistical errors. If there is sampling error, it possible to estimate and control the results.(b) Highly experienced and trained persons can be employed for scientific processing and analyzing of relatively limited data and they can use their high technical knowledge and get more accurate and reliable results.

4. **It provides more detailed information:** As it saves time, money and labor, more detail information can be collected in a sample survey.
5. **Some times only method to depend upon:** Some times it so happens that one has to depend upon sampling method alone because if the population under study is finite, sampling method is the only method to be used. For example, if some ones blood has to be examined, it will become fatal to take all the blood out from the body and study depending upon the total enumeration method.
6. **Administrative convenience:** The organization and administration of sample survey are easy for the same time, money and labor reasons which have been discussed earlier.
7. **More scientific:** Since the methods used to collect data are based on scientific theory and results obtained can be tested, sampling is more scientific method to collect data.

It is not that sampling is free from demerits or shortcomings. There are certain **shortcomings of this method** which are discussed below:

1. **Illusory conclusion:** If a sample enquiry is not carefully planned and executed, the conclusions may be inaccurate and misleading.
2. **Sample not representative:** To make the sample representative is a difficult task. If a representative sample is taken from the universe, the result is applicable to the whole population. If the sample is not representative of the universe the result may be false and misleading.



3. **Lack of experts:** As there is lack of experts to plan and conduct a sample survey, its execution and analyze, the results of the sample survey are not satisfactory and trustworthy.
4. **Some times more difficult than census method:** Some times the sampling plan may be complicated and requires more money, labor, time than a census method.
5. **Personal bias:** There may be personal biases and prejudices with regard to the choice of technique and drawing of sampling units.
6. **Choice of sample size:** If the size of the sample is not appropriate then it may lead to untrue characteristics of the population.
7. **Conditions of complete coverage:** If the information is required for each and every item of the universe, then a complete enumeration survey is better.

**Essentials of sampling:** In order to reach to a clear conclusion, the sampling should possess the following essentials:

1. **It must be representative:** The sample selected should possess the similar characteristics of the original universe from which it has been drawn.
2. **Homogeneity:** Selected samples from the universe should have similar nature and should not have any difference when compared with the universe.
3. **Adequate samples:** In order to have a more reliable and representative result, a good number of items are to be included in the sample.
4. **Optimization:** All efforts should be made to get maximum results both in terms of cost as well as efficiency. If size of the sample is larger, there is better efficiency and at the same time the cost is more. A proper size

of sample is maintained in order to have optimized results in terms of cost and efficiency.

## **STATISTICAL LAWS**

One of the basic reasons for undertaking a sample survey is to predict and generalize the results for the population as a whole. The logical process of drawing general conclusions from a study of representative items is called induction. In statistics induction is a generalization of facts on the assumption that the results provided by an adequate sample may be taken as applicable to the whole. The fact that the characteristics of the sample provide a fairly good idea about the population characteristics is borne out by the theory of probability. Sampling is based on two fundamental principles of statistics theory viz, (i) the Law of Statistical Regularity and (ii) the Law of Inertia of Large Numbers.

### **THE LAW OF STATISTICAL REGULARITY**

The Law of Statistical Regularity is derived from the mathematical theory of probability. According to W.I.King, “The Law of Statistical Regularity formulated in the mathematical theory of probability lays down that a moderately large number of items chosen at random from a very large group are almost sure to have the characteristics of the large group.” For example, if we want to find out the average income of 10,000 people, we take a sample of 100 people and find the average. Suppose that another person takes another sample of 100 people from the same population and finds the average. The average income found out by both the persons will have the least difference. On

the other hand if the average income of the same 10,000 people is found out by the census method, the result will be more or less same.

### **Characteristics**

1. The item selected will represent the universe and the result is generalized to universe as a whole.
2. Since sample size is large, it is representative of the universe.
3. There is a very remote chance of bias.

### **LAW OF INERTIA OF LARGE NUMBERS**

The Law of inertia of Large Numbers is an immediate deduction from the Principle of Statistical Regularity .Law of Inertia of Large Numbers states, ” Other things being equal, as the sample size increases, the results tend to be more reliable and accurate.” This is based on the fact that the behavior or a phenomenon en masse. i.e., on a large scale is generally stable. It implies that the total change is likely to be very small, when a large number or items are taken in a sample .The law will be true on an average. If sufficient large samples are taken from the patent population, the reverse movements of different parts in the same will offset by the corresponding movements of some other parts.

**Sampling Errors:** In a sample survey, since only a small portion of the population is studied its results are bound to differ from the census results and thus, have a certain amount of error. In statistics the word error is used to denote the difference between the true value and the estimated or approximated value. This error would always be there no matter that the sample is drawn at random and that it is highly representative. This error is attributed to fluctuations of sampling and is called sampling error. Sampling error is due to the fact that only a sub set of the population has been used to estimate the population parameters

and draw inferences about the population. Thus, sampling error is present only in a sample survey and is completely absent in census method.

Sampling errors occur primarily due to the following reasons:

- 1. Faulty selection of the sample:** Some of the bias is introduced by the use of defective sampling technique for the selection of a sample e.g. purposive or judgment sampling in which the investigator deliberately selects a representative sample to obtain certain results. This bias can be easily overcome by adopting the technique of simple random sampling.
- 2. Substitution:** When difficulties arise in enumerating a particular sampling unit included in the random sample, the investigators usually substitute a convenient member of the population. This obviously leads to some bias since the characteristics possessed by the substituted unit will usually be different from those possessed by the unit originally included in the sample.
- 3. Faulty demarcation of sampling units:** Bias due to defective demarcation of sampling units is particularly significant in area surveys such as agricultural experiments in the field of crop cutting surveys etc. In such surveys, while dealing with border line cases, it depends more or less on the discretion of the investigator whether to include them in the sample or not.
- 4. Error due to bias in the estimation method:** Sampling method consists in estimating the parameters of the population by appropriate statistics computed from the sample. Improper choice of the estimation techniques might introduce the error.
- 5. Variability of the population:** Sampling error also depends of the variability or heterogeneity of the population to be sampled.

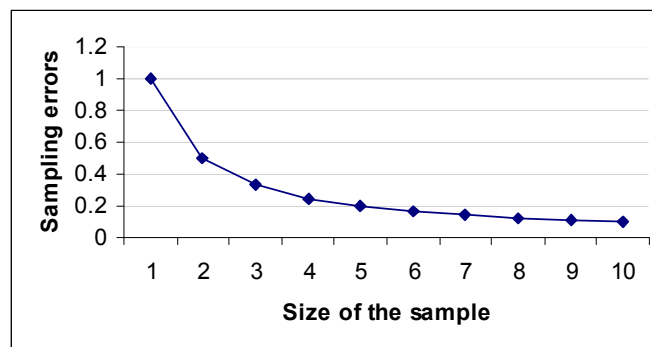
**Sampling errors are of two types- Biased Errors and Unbiased Errors**

**Biased Errors:** The errors that occur due to a bias of prejudice on the part of the informant or enumerator in selecting, estimating measuring instruments are called biased errors. Suppose for example, the enumerator used the deliberate sampling method in the place of simple random sampling method; then it is called biased errors. These errors are cumulative in nature and increase when the sample size also increases. These errors arise due to defect in the methods of collection of data, defect in the method of organization of data and defect in the method of analysis of data.

**Unbiased errors:** Errors which occur in the normal course of investigation or enumeration on account of chance are called unbiased errors. They may arise accidentally without any bias or prejudice. These errors occur due to faulty planning of statistical investigation.

To avoid these errors, the statistician must take proper precaution and care in using the correct measuring instrument. He must see that the enumerators are also not biased. Unbiased errors can be removed with the proper planning of statistical investigations. Both of these errors should be avoided by the statisticians.

**Reducing Sampling Errors:** Errors in sampling can be reduced, if the size of sample is increased. This is shown in the following diagram.



From the above diagram it is clear that when the size of the sample increases, sampling error decreases. And by this process samples can be made more representatives to the population.

**Testing of Hypothesis:** As a part of investigation, samples are drawn from the population and results are drawn which helps take the decision. But such decisions involve an element of uncertainty causing wrong decisions. Hypothesis is an assumption which may or may not be true about a population parameter. For example, if we toss a coin 200 times, we may get 110 heads and 90 tails. At this instance we are interested in testing whether the coin is unbiased or not.

Therefore, we may conduct a test to judge significance whether the difference is due to sampling or otherwise. To carry out a test of significance following procedure has to be followed:

**1. Framing the Hypothesis:** To verify the assumption, which is based on sample study, we collect data and find out the difference between the sample value and the population value. If there is no difference found or the difference is very small then the hypothetical value is correct. Generally two hypotheses are constructed, and if one is found correct the other is rejected.

**(a) Null Hypothesis:** The random selection of the samples from the given population makes the tests of significance valid for us. For applying any test of significance we first set up a hypothesis- a definite statement about the population parameter/s. Such a statistical hypothesis, which is under test, is usually a hypothesis of no difference and hence is called *Null hypothesis*. It is usually denoted by  $H_0$ . In the words of Prof. R.A.Fisher “**Null hypothesis is the hypothesis which**

is tested for possible rejection under the assumption that it is true.”

**(b) Alternative Hypothesis.** Any hypothesis which is complementary to the null hypothesis is called an alternative hypothesis. It is usually denoted by  $H_1$ . It is very important to explicitly state the alternative hypothesis in respect of any null hypothesis  $H_0$  because the acceptance or rejection of  $H_0$  is meaningful only if it is being tested against a rival hypothesis. For example, if we want to test the null hypothesis that the population has a specified mean  $\mu_0$  (say), i.e.,

$$H_0: \mu = \mu_0$$

Then the alternative hypothesis could be:

**(i)  $H_1: \mu \neq \mu_0$  (i.e.  $\mu > \mu_0$  or  $\mu < \mu_0$ )**

**(ii)  $H_1: \mu > \mu_0$**

**(iii)  $H_1: \mu < \mu_0$**

The alternative hypothesis (i) is known as a two – tailed alternative and the alternatives in (ii) and (iii) are known as right – tailed and left- tailed alternatives. Accordingly, the corresponding tests of significance are called two-tailed, right-tailed and left-tailed tests respectively.

The null hypothesis consists of only a single parameter value and is usually simple while alternative hypothesis is usually composite.

**Types of Errors in Testing of Hypothesis:** As stated earlier, the inductive inference consists in arriving at a decision to accept or reject a null hypothesis ( $H_0$ ) after inspecting only a sample from it. As such an element of risk – the risk of taking wrong decision is involved. In any test procedure, the four possible mutually disjoint and exhaustive decisions are:

**(i) Reject  $H_0$  when actually it is not true, i.e., when  $H_0$  is false.**

**(ii) Accept  $H_0$  when it is true.**

**(iii) Reject  $H_0$  when it is true.**

**(iv) Accept  $H_0$  when it is false.**

The decision in (i) and (ii) are correct decisions while the decisions (iii) and (iv) are wrong decisions. These decisions may be expressed in the following dichotomous table:

		Decision from sample	
		Reject $H_0$	Accept $H_0$
True State	$H_0$ True	Wrong Type I Error	Correct
	$H_0$ False ( $H_1$ True)	Correct	Wrong Type II Error.

Thus, in testing of hypothesis we are likely to commit two types of errors. The error of rejecting  $H_0$  when  $H_0$  is true is known as Type I error and the error of accepting  $H_0$  when  $H_0$  is false is known as Type II Error.

For example, in the Industrial Quality Control, while inspecting the quality of a manufactured lot, the Inspector commits Type I Error when he rejects a good lot and he commits Type II Error when he accepts a bad lot.

**SUMMARY**

Nowadays questionnaire method of data collection has become very popular. It is a very powerful tool to collect required data in shortest period of time and with little expense. It is scientific too. But drafting of questionnaire is a very skilled and careful work. Therefore, there are certain requirements and essentials which should be followed at the time of framing the questionnaire. They include- size of the questionnaire should be small, questions should be very



clear in understanding, questions should be put in a logical order, questions should have simple meaning etc. Apart from this, multiple choice questions should be asked. Questionnaire should be pre tested before going for final data collection. Information supplied should be cross checked for any false or insufficient information. After all these formalities have been completed, a covering note should accompany the questionnaire explaining various purposes, designs, units and incentives.

There are two ways of survey- Census survey and Sample survey through which data can be collected. Census survey means total enumeration i.e. collecting data from each and every unit of the universe whereas sample survey concentrates on collecting data from few units of the universe selected scientifically for the purpose. Since census method is more time taking, expensive and labor intensive, it becomes impractical to depend on it. Therefore, sample survey is preferred which is scientific, less expensive, less time taking and less labor intensive too.

But there are merits and demerits of this method which are detailed below:

Merits- it reduces cost, it is more reliable, it saves time; it provides more detailed information, some times only method to depend upon, administrative convenience, more scientific etc.

Demerits- it may give illusory conclusions, sometimes samples may not be representative, there is lack of experts, some times it is more difficult than census method, personal bias, determining the size of the sample very difficult etc.

Apart from these, there are some essentials of sampling which must be followed. They are – Samples must be representative, samples must be homogeneous and the number of samples must be adequate. When the researcher resorts to sampling, he intends to collect some data which help him to

draw results and finally take a decision. When he takes a decision on the basis of hypothesis which is precisely assumption and is prone to two types of errors- Type I Error and Type II Error. When a researcher rejects a correct hypothesis, he commits type I error and when he accepts a wrong hypothesis he commits type II error. The researcher should try to avoid both types of errors but committing type II error is more harmful than type I error.

#### **SELF ASSESMENT QUESTIONS (SEQs)**

1. Explain questionnaire and examine its main characteristics.  
(Refer to the introduction part of the questionnaire section)
2. Explain main requirements of a good questionnaire.  
(Refer to the sub points from 1 to 11)
3. What is sampling? Explain its main merits and demerits.  
(Refer to the introduction and the following part of the lesson)
4. What are null and alternative hypothesis? Explain.  
(Refer the point Framing the Hypothesis)
6. What are Type I error and Type II error? (Refer to types of error in hypothesis)

\*\*\*