Unit

5 Sample Survey

5.1 Concept of Population and Sample

Population: When we think of the term "population," we usually think of people in our town, region, state or country. However, in statistics, the term "population" takes a slightly different meaning. In statistics, "population or universe" is the aggregate of objects in any statistical investigation. That is, statistical population refers the totality or aggregate of all units or items under investigation or study. In other words, the group of items or units under the study (investigation or inquiry) is called population. There are two types of population. They are target population and sampling population.

The population or universe may be finite or infinite. A population containing a finite number of objects or items is known as finite population. For example, total no. of students of T.U in BCA, the total number of books in a library, total number of households in Kathmandu etc.

On the other hand, a population having an infinite number of objects is called infinite population. For example, the population of stars in the sky is an infinite population, total number of fishes in Pacific Ocean, total number of trees in a forest etc.

In the statistical investigation, the investigator usually deals with the general magnitude and the study of variation respect to one or more characteristics relating to individual belonging to a group. The group of individuals under study is called population or universe. Population is an aggregate or collections of objects, animates or inanimate defined according to characteristics under study. The population may be finite or infinite.

Complete enumeration of all units of the population is called **census**. In any statistical investigation complete enumeration is not practicable. In a statistical investigation the interest usually lies in the assessment of the general magnitude and the study of variation with respect to one or more characteristic relating to individuals under study is called population or universe. Thus, in statistics, population is an aggregate of objects under study.

It is obvious that for any statistical investigation complete enumeration of the population is rather impractical. For example, if we want to have an idea of the average per capita (monthly) income of the people of Nepal, we will have to enumerate all the earning individuals in the country, which is rather a very difficult task.

If population is infinite, complete enumeration is not possible. Also, if the units are destroyed in the course of inspection, 100% inspection is not taken because of multiplicity of cause, viz., administrative and financial implications, time factors etc. and we take the help of sampling.

Sample: A finite subset of statistical individuals in a population that is used to represent that population is called a sample (or the representative part of any population under study is a sample) and the number of individuals in a sample is called sample size. For the purpose of determining population characteristic, instead of enumerating entire population, the individual in the sample only are observed. Then the sample characteristics are utilized to approximately determine or estimate the population. For example, on examining the sample of a particular stuff we arrive at a decision of purchasing or reject the stuff.

Sampling is quite often used in our day-to-day practical life. For example, in a shop we assess the quality of sugar, wheat or any other commodity by taking a handful of it from the bag and then decided to

purchase it or not. A house wife normally tests the cooked products to find if they are properly cooked and contain the proper quantity of salt.

i) **Sampling frame:** A complete list of the sampling units, which represents the population to be covered is called the sampling frame popularly known as frame. In other words, a list of all the elements that is in the population.

ii) **Sampling units:** The sampling unit is the basic unit containing the elements of the population to be sampled. It may be the element itself or a unit in which the element is contained. The sampling unit selected is often dependent upon the sampling frame. The selection of sampling units is also partially dependent upon the overall design of the project.

iii) **Sample:** A part or portion of the population which is considered for study and analysis is called a sample. In other words, a finite subset of the population with the objective of investigating its properties is called a sample and the number of units in the sample is known as the sample size.

5.2 Needs of Sampling

Sampling is carried out because of following reasons

i) Sampling saves time and money. Sample study is less expensive than census and gives faster result.

ii) Only if test involves the destruction of items under investigation. Sample study prevents the items from destruction if items are of destructive nature.

iii) Only way for the infinite population. For the infinite population complete enumeration is not possible so that sample method is only the method of collecting information.

iv) Enable to estimate sampling error. The sample estimate gives the error from the population.

v) Enable more accurate result. Sampling conducted by well trained and experienced investigator gives accurate result.

5.3 Census and Sample Survey

Survey is the technique of investigation by direct observation of a phenomenon of collection of information through interview. The meaning of survey has been used in broader sense to include the observation of published documents. The survey is used describing current practice and events and analyzing the facts.

The main objective of the survey are as follows:

i) Supply of information on any problem ii) Description of the phenomenon

iii) Explanation of a phenomenon

Census

The method of enumerating each and every unit of the population is known as census. Some of the examples are Population census, manufacturing census, agricultural sample census etc.

Sample survey

The survey carried out by selecting representative sample of the study population is known as sample survey. For example, Nepal living standard survey, family planning survey etc.

5.4 Basic Concept of Sampling

Generally, sampling has been employing not only in any statistical investigation and research works but also in the case of daily life of human beings. For example,

i) A housewife tests very small quantity of cooking foods to know about the taste of whole food that she is cooking.

ii) A doctor tests a drop of blood of a patient to know about the characteristic of the whole blood of the patient.

iii) A businessperson gives order for the commodities by examining a small unit of the same commodity.

In such practical decision making process associated with various fields of human activities, most of our conclusions, decisions, and findings depend upon the inspections or test or examination of few units of an aggregate or totality. This process of getting result or information of conclusion about the totality or universe or whole group by performing examination of only some parts of the universe under the study (investigation) is called sampling. In other words, sampling is the process by which inference is made to the whole by examining only a few parts. Sampling is a tool which enables us to draw conclusion about the characteristics of the population after studying the items in the sample.

5.5 Census versus Sampling

A census is a study of every unit, everyone or everything, in a population. That is, a census is the complete enumeration or count of all units of the population for certain character of the population. Hence, the complete enumeration of all units of the population is known as Census Survey.

The term census is used mostly in connection with National population census and Housing Censuses and other common censuses include agriculture census, business census, Industrial census survey etc. Census requires more money, manpower and time.

The method or process of selecting a sample from a population under study is called sampling.

A sample is a subset of units in a population selected to represent all units in a population. It is a partial enumeration because it is a count from part of the population. Therefore, the process (or survey) in which only part of the population is selected and examined to estimate the certain character of the population is known as sample survey. That is, the enumeration of the selected units is known as sample survey. Information from the sampled units is used to estimate the characteristics for the entire population. A sample survey will usually be less expensive than a census survey and the desired information will be obtained in less time.

When and Where Sampling/Census is Appropriate:

A sampling technique is appropriate

a. When the universe is very large

b. When the universe possesses homogeneous characteristics

c. When utmost accuracy is not required

d. Where census is impossible i.e., in destructive/explosive nature of testing.

A census is appropriate when

a. The universe is small

b. The population is heterogeneous

c. Hundred percent accuracy is required

d. The population frame is incomplete

Demerits of Sampling Technique:

i) Less accuracy ii) Misleading conclusions

iii) Need of specialized knowledge iv) When sampling is not possible

v) It cannot be used if the information of each and every unit of the population is required.

Demerits of Census:

i) Expensiveness ii) Excessive time and effort

iii) Not applicable for destructive testing iv) For infinite population, it is not possible.

5.6 Distinction between Census and Sampling

Census	Sampling
1. It takes each and every unit of the population.	1. It takes representative part of the population.
2. It is costly and time consuming i.e., It requires more money, manpower and time.	2. It is less expensive and less time consuming, i.e., It requires less money, manpower and time.
3. It is more accurate.	3. It is less accurate.
4. It has less scope.	4. It has greater scope.
5. It is appropriate when the universe is small.	5. It is appropriate when the universe is large.
6. It is appropriate when the population is heterogeneous.	6. It is appropriate when the universe possess homogeneous characteristics.

5.7 Organizational Aspect Sampling Survey

5.7.1 The Basic Organization Aspect Sampling Survey

i) Objective of the Survey

The first step in sampling is to define the objective of the survey in clear & lucid term. The sponsor of the survey should take care that these objectives are commensurate with the available resources in terms of money, manpower and the time limit required for availability of the results of the survey.

ii) Defining Population to be Sampled

The population i.e., the aggregate of objects from which sample is chosen should be defined in a clear and unambiguous terms. The population to be sampled (the sampled population) should be coincided with the population about which information is wanted (The target population). The definition of the population when sampling a batch of electric light bulb in order to estimate the average length of life of a bulb is clear but in sampling a population of farms, rules must be setup to define a farm regarding shape, size etc. keeping in mind the boarder-line cases so as to enable the investigator to decide in the field without much hesitation whether or not to include a given farm in the population.

iii) Data to be collected

It is well to verify that all the data are relevant to the purpose of the survey and that no essential data are omitted. There is frequently a tendency, particularly with human population to ask too many questions, some of which are never subsequently analyzed. An overlong questionnaire lowers the quality of the answer to important as well as unimportant questions.

iv) Degree of precision desired

The result of sample surveys are always subject to some uncertainty because only part of population has been measured and because of errors of measurement. This uncertainty can be reduced by taking larger samples and by using superior instruments of measurement. But this usually cost time and money. Consequently, the specification of degree of precision wanted in the results is an important step. This step is the responsibility of the person who is going to use the data.

v) Method of Measurement

There may be choice of measuring instrument & method of approach to the population. The survey may employ a self-administered questionnaire, an interviewer who reads a standard set of questions with no discretion or an interview process that allows much latitude in the form & ordering of the questions (Schedule). The approach may be by mail, by telephone, by personal visit or its combinations.

vi) Frame

Before selecting the sample, the population must be divided into parts that are called sampling units or units. These units must cover the whole of the population and they must not overlap; in the sense that every element in the population belongs to one and only one unit. For example in socio-economic survey for selecting people in a town the sampling unit might be individual person, a family, a household or a block of locality. In order to cover the population decided upon, there should be some list, map or other acceptable materials, called frame, which serves as a guide to the population to be covered. The construction of the frame is often one of the major practical problem, since it is a frame which determines the structure of the sample survey. The list which have been routinely collected for some purpose, are usually found to be incomplete or partially eligible or often contains an unknown amount of duplication such list should be carefully scrutinized & examining to ensure that they are free from those defects and are up to date.

vii) Selection of proper sampling design

There are varieties of plans by which the sample may be selected. For each plan that is considered rough estimates of the size of sample can be made from knowledge of the degree of precision desired. The relative cost and time involved for each plan are also compared before making decision.

viii) Organization of field work

It is essential that the personnel should be thoroughly trained in locating the sample units, recording the measurements, the methods of collection of required data before starting the field work. The success of a survey to a great extent depends upon the reliable field work. It is very necessary to make provisions form adequate supervisory staff for inspection after field work.

It has been found useful to try out the questionnaire and the field method on small scale called pretest. It is always helping to decide upon effective method of asking questions and results in the improvement of the questionnaire. Moreover, it might disclose certain problems and troubles that will otherwise be quite serious on a large-scale survey.

5.7.2 Step for Sample Survey

Step 1. Define the population

Step 2. Specify the sampling frame

Step 3. Specify sampling unit

Step 4. Selection of sampling method

Step 5. Determine of the sample size

Step 6. Specify the sampling plan

Step 7. Select the sample

Step 8. Analysis the representative sample with appropriate statistical tools (measures).

5.8 Questionnaire Design

Introduction

It is popular a method of data collection used in case of big inquiries. It is adopted by individuals, researchers, private organizations, public organizations, governmental organizations etc.

The questionnaire is a set of questions, which is mailed or sent to respondents via a link person or an agency or an e-mail or a post office and which should be filled by respondents themselves. It is more or less the same to schedule but different in sending and filling procedure. George A. Landberg says, "Fundamentally, the questionnaire is a set of questions to which literate people are exposed in order to observe their verbal behavior under their stimuli".

In other words, it is a research instrument consisting of a series of questions for the purpose of gathering information from respondents. It is sent to the concerned persons with request to answer the questions and to return the questionnaire. It is to be filled up by respondents. It is a valuable method of collecting a wide range of information from a large number of individuals.

Types of Questionnaires

There are different kinds of questionnaires, some of them are

a) Structured questionnaire
b) Unstructured questionnaire
c) Open ended questionnaire
d) Close ended questionnaire
e) Mixed questionnaire
f) Pictorial questionnaire.

a) Structured Questionnaire

It is a questionnaire used in formal inquiry having definite questions. The Structured questionnaire is that one, in which there are definite, unambiguous concrete and preordered questions with additional questions limited to those necessary to clarify inadequate questions or to elicit (bring out) a more detailed response. It is well prepared and used in studies of economics, social problems, administrative policies etc.

b) Unstructured questionnaire

In this type of questionnaire, only the direct questions or answer-required questions are arranged in some array. It is similar to those set of questions of oral or written interview. It is the questionnaire used at the time of interview and acts as guide for interviewer. It is flexible in working and used in the studies related to families or personal experience, beliefs etc.

c) Closed questionnaire

It is the questionnaire, in which responses are limited to the stated alternatives. It is the set of questions, in which there are limited options to answer the question and no chance to express the respondents own views except given options.

d) Open questionnaire

In this type of questionnaire, the views of the respondents are welcome. In open questionnaire, some spaces are given to express the views of the respondents beside of the set answer options. It is the questionnaire, in which respondent is free to express his/her view and the ideas. It is used in making

intensive studies of the limited number of the case and do not provide any structure for the respondent's reply. The questions and their orders are pre-determined in the nature.

e) Pictorial questionnaire

It is the set of questions, in which pictures are given and from those pictures, the respondents have to answer the questions. It is very rare and used in the studies related to social attitude in children.

f) Mixed questionnaire

It is the questionnaires in which various form of the questions are set to answer. Particularly this type of questionnaire is the mix of both close and open-ended questions. It is used in field of social research.

Requisites of a Good Questionnaire

There is certain assumption that the questionnaire should behave. The common and acceptable forms of the requisites of a questionnaire are given below.

i) A questionnaire should be written within the informational scope of respondent
ii) It should put up with the objectives of the study
iii) The questions should be interconnected
iv) The questions should be short, clear and not so many.
v) It should contain choice answer type or dichotomous (say Yes/No) questions.
vi) The questions should be simple to understand (technical terms be given with meaning)
vii) It should contain format of cross-checking questions.
viii) It should include as far as possible impersonal (remote) questions
ix) The questions should be free from ambiguity.

Importance of Questionnaire Method

The significance of the questionnaire method of data collection are:

i) It is very useful in the study of large population.
ii) It is practical method of data collection.
iii) This method of data collection can be implemented with minimum expenses.
iv) It is the way of receiving the information quickly.
v) This method is free of surveyor biases and gives valid information.
vi) It is very convenience way of data gathering for the study.
vii) The data collection procedure by this method is self administered.
viii) Data collected by questionnaire methods can be analyzed more scientifically.

Limitations of Questionnaire Method

The following are the limitations of the questionnaire method of data collection.

i) It has low rate of return of the filled-up questionnaire.
ii) It can be used for educated respondents only.
iii) It has no control over questionnaire after sent.
iv) In this method, it is always possible to get representative samples.
v) There are possibilities of ambiguous replies or omission of replies to certain questions.

vi) It is difficult to know whether the willing respondents are truly representatives.

vii) As in schedule method it has no assistance in answering.

viii) It is slowest method of data collection.

ix) Problem of non-response is great limitation of this method.

x) The filled questionnaires re also difficult to read out because of respondents' bad hand-writing.

5.9. Principles of Sample Survey

The theory of sampling is based on the following important principles:

1. Principle of statistical regularity

This principle has the origin in the mathematical theory of probability. According to King, "the law of statistical regularity lays down that a moderately large number of items chosen at random for a large group are almost sure on the average to possess the characteristics of the large group." This implies that each and every unit in the population should have equal chance of being selected.

Principle of inertia of large is derived from this concept, it states that, "other things being equal as the sample size increases, the results tend to be more reliable and accurate."

2. Principle of validity

By the validity of a sample design, we mean that it should enable us to obtain valid tests and estimate about the parameters of the population. The samples obtained by the technique of probability sampling satisfy this principle.

3. Principle of Optimization

This principle impress upon obtaining optimum results in terms of efficiency and cost of the design with least sacrifice of the resources. The reciprocal of sampling variance of the estimate provides a measure of its efficiency while a measure of the cost of the design is provided by the total expenses incurred in terms of money and labour. The principle of optimization consists of

i) Achieving a given level of efficiency at minimum cost

ii) Obtaining maximum possible efficiency with given level of cost.

4. The questionnaire should contain questions which can be answered in minimum of writing.

5. Why questions should be included as supplementary questions with choice answer.

6. The question's arrangement should be simple to tabulate.

7. Some of the bad questions should be avoided. Some examples of bad questions are too long, complex, person, ambiguous, leading, non-relevant, embarrassing etc.

8. Some of the word should be avoided and some precautions should be taken in using such words. The following are some of the words (which are difficult to understand by the respondents) that should be avoided as far as studies.

i) Technical terms for non-technical studies.

ii) Native and unused words

iii) Words having emotional feelings.

iv) Qualitative, subjective, offending words.

v) Ambiguous words etc.

5.10 Sample Selection and Determination of Sample Size

In sample technique the size of sample should be large enough to give a confidence interval of preferred width. To determine the size of the sample researcher should keep in mind the following points:

a) Nature of universe
b) Number of classes purposed
c) Nature of study
d) Types of sampling
e) Standard of accuracy and acceptance confidence level
f) Availability of finance
g) Other consideration $n = \left(\frac{Z_\alpha\, \sigma}{E}\right)^2$

where, n = Sample size, Z_α = Standard normal variate (normal table value), E = Permissible error (Different between sample statistic and population parameter).

$$n = \left(\frac{Z_\alpha}{E}\right)^2 PQ$$

where, P = Probability of certain characteristics, Q = Non happening of certain characteristics for referred data.

Hence, selection of sample dependent upon the variation of statistical data level of significance (α) referred to standard normal variable, and permissible error is different between sample statistics and population parameter.

Solved Examples

Example 5.1 The mean and standard deviation of a random sample of 50 were found to be 100 and 10. If the investigator wants to be 95% confidence level that the error in estimate mean. Should not exceed ± 2. How many additional observations are required?

Solution: Here, sample size (n) = 50 (First taken)

Sample mean ($\bar{X}$) = 100

Standard deviation (σ) = 10

Permissible error (E) = ±2

For 95% confidence level,

Level of significance (α) = (100 – 95) % = 5% = 0.05

Standard normal variate by table,

$$z_\alpha = z_{0.05} = 1.96$$

Now, sample size $(n) = \left(\frac{z_\alpha\, \sigma}{E}\right)^2 = \left(\frac{1.96 \times 10}{2}\right)^2 = 96.05$

$$\therefore\ n = 96 \text{ nearly.}$$

Hence, the number of additional observations required = 96 – 50 = 46.

Example 5.2 Find the sample size when probability 0.35 for all most same and 5% permissible error.

Solution: $E = 5\% = 0.05$

$P = 0.35$

$Q = 0.65\ (1 - 0.35)$

$z_\alpha = -3$ for all most same.

By formula $n = \left(\frac{z_\alpha}{E}\right)^2 PQ = \left(\frac{3}{0.05}\right)^2 \times 0.35 \times 0.65 = 819$

Required sample size is 819.

Example 5.3 A manufacturing concern wants to estimate the average amount of purchase of its production a month by the customers. If the standard deviation is Rs. 10, find the sample size if the maximum error is not be exceed Rs. 3 with a 99% confidence.

Solution: Standard deviation $\sigma = 10$

Level of significance = 1%

99% level of confidence

$\alpha = 0.01$

Then, $z_{0.01} = 2.58$

$E = 3$

$\therefore \quad n = \left(\frac{z_\alpha\, \sigma}{E}\right)^2 = \left(\frac{10 \times 2.58}{3}\right)^2 = 74.$

Example 5.4 We would like to know the average time that child spends watching television over the weekend. We want our estimate to be within ± 1 hour of the true population average. Previous studies have shown the population standard deviation to be 3 hours. What sample size should be taken for this purpose, if we want to be 95% confident that the error in our estimate will not exceed the maximum allowable error?

Solution: Given,

Maximum allowable error $(E) = |\pm 1| = 1$

Population S.D. $(\sigma) = 3$

Sample $(n) = ?$

Confidence level $(1 - \alpha) = 0.95$

Risk $(\alpha) = 0.05$

$Z_\alpha = Z_{0.05} = 1.96$

$\therefore$ Required sample size $(n) = \left(\frac{Z_\alpha\, \sigma}{E}\right)^2 = \left(\frac{3 \times 1.96}{1}\right)^2 = 34.57 \approx 35$

$\therefore$ Sample size should be 35, if we want to 95%. Confident that the error in our estimate will not exceed the maximum allowed error.

Example 5.5 A research workers wishes to estimate the mean of a population by using sufficiently large sample. The probability is 0.95 that the sample mean will not differ from the true mean by more than 25% of the standard deviation. How large a sample should be taken?

Solution: Given,

$$\text{Probability} = 0.95 = \text{Confidence level} = 1 - \alpha$$

$$\Rightarrow \quad \alpha = (1 - 0.95) = 0.05$$

$$Z_\alpha = Z_{0.05} = 1.96$$

$$\text{Error } (E) = 25\% \text{ of s.d.} = \left(\frac{25}{100}\right)\sigma = 0.25\,\sigma$$

$$n = \left(\frac{Z_\alpha\,\sigma}{E}\right)^2 = \left(\frac{1.96 \times \sigma}{0.25 \times 6}\right)^2 = 61.47$$

$$\therefore \quad n = 61$$

Example 5.6 The average outstanding balance of loans issued by a bank varies from month to month. From past experience it is known that the amounts are normally distributed with a standard deviation of Rs. 5,000. The bank wishes to estimate the average by drawing a random sample such that the probability is 0.95 that the mean of the sample will not deviate by more than Rs. 600 from the universe mean. What should be the sample size?

Solution: Given,

$$\text{Population standard deviation } (\sigma) = 5000$$

$$\text{Probability} = 0.95$$

$$\Rightarrow \quad \alpha = (1 - \text{prob.}) = (1 - 0.95) = 0.05$$

$$Z_\alpha = Z_{0.05} = 1.96$$

$$\text{Error } (E) = 600$$

Now,
$$\text{Sample size } (n) = \left(\frac{Z_\alpha\,\sigma}{E}\right)^2$$

$$\Rightarrow \quad n = \left(\frac{1.96 \times 5000}{600}\right)^2 = 266.78$$

$$\therefore \quad n = 267$$

Example 5.7 In measuring reaction time, a psychologist estimates that the standard deviation is 0.05 sec. How large a sample of measurements must be taken in order to be 95% confident that the error of his estimate of mean will not exceed 0.0 sec?

Solution: Given,

$$\text{Population, S.D. } (\sigma) = 0.05$$

$$\text{Confidence level, } (1 - \alpha) = 95\% = 0.95$$

$$\therefore \quad \alpha = 0.05,\ Z_\alpha = Z_{0.05} = 1.96$$

$$\text{Error } (E) = 0.01$$

We have,
$$n = \left(\frac{Z_\alpha\,\sigma}{E}\right)^2 = \left(\frac{1.96 \times 0.05}{0.01}\right)^2 = 96.04$$

$$\therefore \quad n = 96$$

Example 5.8 It is desired to estimate the proportion of junior executive who change their first job within the first five years. This proportion is to be estimate within 3 percent of error and 0.95 degree of confidence is to be used. A study conducted several years ago revealed than 30% of such junior executives changed their first job within 5 years.

i) How large a sample is required to update the study?

ii) How large should the sample be if no such previous estimates are available?

Solution: Given,

$$\text{Confidence level, } 1 - \alpha = 0.95;\ \alpha = 0.05$$

$$Z_\alpha = Z_{0.05} = 1.96$$

$$P = 30\% = 0.3$$

$$Q = 1 - P = 0.7$$

$$\text{Error } (E) = 3\% = 0.03$$

We have,

$$\text{Sample size } (n) = \left(\frac{Z_\alpha}{E}\right)^2 PQ = \left(\frac{1.96}{0.03}\right)^2 \times 0.3 \times 0.7 = 896.37$$

$$\therefore \quad n = 896$$

(ii) Since, no previous estimate (i.e., P and Q) are available so we take $P = Q = 0.5$

Now,

$$\text{Simple size } (n) = \left(\frac{Z_\alpha}{E}\right)^2 PQ = \left(\frac{1.96}{0.03}\right)^2 0.5 \times 0.5 = 1067.11$$

$$\therefore \quad n = 1067$$

Example 5.9 A political pollster wants to estimate the proportion of voters who will vote for the democratic candidate in a presidential campaign. The pollster wishes to have 90% confidence that her prediction is correct to within ±0.04 of the population proportion.

i) What sample size is needed?

ii) If the pollster wants to have 95% confidence. What sample size is needed?

iii) If she wants to have 95% confidence and a sampling error of +0.03, what sample is needed?

Solution: Given,

$$\text{Confidence level, } (1 - \alpha) = 0.90$$

$$\text{Risk } (\alpha) = 0.10$$

$$Z_\alpha = Z_{0.1} = 1.645$$

$$\text{Maximum allowable error } (E) = |\pm 0.04| = 0.04$$

i) Sample size $(n) = ?$

If previous estimate (i.e., P and Q) are not equally available.

We assume, $P = Q = 0.50$

$$\therefore \text{ Required sample size } (n) = PQ\left(\frac{Z_\alpha}{E}\right)^2 = 0.5 \times 0.5 \left(\frac{1.645}{0.04}\right)^2 = 422.82 \approx 423$$

ii) Confidence level, $(1-\alpha) = 0.95$

Risk $(\alpha) = 0.05$

$Z_\alpha = Z_{0.05} = 1.96$

If previous estimates (i.e., P & Q) are not given,

We assume, $P = Q = 0.50$

$\therefore$ Required sample size $(n) = PQ\left(\frac{Z_\alpha}{E}\right)^2 = 0.5 \times 0.5 \times \left(\frac{1.96}{0.03}\right)^2 = 1067.11 \approx 1067$

Parameter and Statistic

Parameter: Parameters are the functions of population values. In other words, the values that describe the characteristics of a population are called parameters. The statistical constants of the population like population mean, population variance, skewness etc. are parameters. Hence, a parameter is a characteristic of a population and it is denoted by θ.

Statistic: Statistics are the functions of the sample observations. In other words, the statistical measures or constants which are calculated from the sample data are called statistics. The statistical constants of the samples like sample mean, sample variance, sample skewness etc. are statistics. Hence, a statistic is a characteristic of a sample and it is used to estimate the value of population.

Parameter	**Statistic**
Population size (N)	Sample size (n)
Population mean (μ)	Sample mean $(\bar{x})$
Population standard deviation (σ)	Sample standard deviation (s)
Population proportion (P)	Sample proportion (p)
Population correlation coefficient (P)	Sample correlation coefficient (r)
Population coefficient of skewness β_1	Sample coefficient of skewness (b_1)
Population coefficient of Kurtosis (β_2) etc.	Sample coefficient of kurtosis (b_2) etc.

Standard Error (S.E.)

The standard deviation of the sampling distribution of sample statistic is known as its standard error (S.E.) of the statistic. Thus, the standard error of statistic t is given by

$$\text{S.E. } (t) = \sqrt{\text{Variance } (t)} = \sqrt{\frac{1}{n}\Sigma(t - \bar{t})^2}$$

The standard deviation of the distribution of sample mean is called the standard error of the sample mean. It is denoted by $\sigma_{\bar{x}}$ or, S.E. $(\bar{X})$. Standard error of the mean is a measure of dispersion of the distribution of sample mean. Similarly, the standard deviation of the sampling distribution of proportion is called the standard error of the proportion. The standard error (S.E.) of some well-known statistics are given in the following table. If n is the sample size, σ is the population standard deviation, P is the population proportion.

Statistic	**Standard error (S.E.)**
Sample mean $(\bar{X})$	$\frac{\sigma}{\sqrt{n}}$ or, $\sqrt{\frac{\sigma^2}{n}}$

sample proportion (p)	$\sqrt{\frac{PQ}{n}}$, $Q = 1 - P$
Sample standard deviation (s)	$\frac{s}{\sqrt{2n}}$
Sample correlation coefficient (r)	$\frac{(1 - \rho^2)}{\sqrt{n}}$

5.12 Sampling and Non-Sampling Errors

Sampling Error

A sampling error is the error, which is made in selecting samples that are not representative of population. It arises due to the fact that only a part of the population has been used to estimate population parameter and draw inferences about the population. As such sampling errors are absent in a complete enumeration.

Sampling error occurs because of the particular sapling technique which is being used for selection of sampling. Like simple random sampling requires larger size, cluster sampling have the problem of clustered population etc.

Following are some of the reasons for sampling error:

1. **Faulty Selection of a Sample**

   Some of the sampling error occurs by the use of defective sampling technique for the selection of a sample.

2. **Substitution**

   If difficulties arise in enumerating a particular sapling unit include in random sample, the investigators usually substitute convenient member of the population. This is obviously leads to sampling error. Since the characteristic possessed by the unit originally included in the sample.

3. **Faulty Demarcation of Sampling Unit**

   Sampling error observed due to defective demarcation of sampling unit is particularly significant in area survey. In such survey, while dealing with boarder line cases it depends more or less on discretion of the investigator whether to include them in sample or not.

4. **Improper Choice of Statistic to Estimate Population Parameter**

   Sometimes the biased estimator may create sampling error while estimating population parameter. Increase in sample size (i.e., numbers of units in the sample) usually results in the decrease in sampling error. In fact, in many situations this decrease in sampling

Non-Sampling Error

Non–sampling errors arise at the stages of planning, observation, ascertainment and processing of the data and are thus present in both the complete enumeration survey (and census) and sample survey. Thus, the data obtained from a complete census are free from sampling errors but may not be free from non-sampling errors whereas data obtained from the sample survey would be both sampling and non-sampling errors. So, non-sampling errors are more serious in census survey in compared to a sample survey.

Non-sampling errors arise due to the following factors:

i) Faulty planning or definitions
ii) Response errors
iii) Non –response errors
iv) Compiling errors
v) Publication errors
vi) Coverage error

Exercise 5.1

1. What is sampling? Describe the various methods of sampling with their merits and demerits.

2. The business manager of a large company wants to check the inventory record against the physical inventory record against the physical inventories by a sample survey. He wants to almost assure that the maximum sampling error should not be more than 5% above or below the true proportion of inaccurate records is estimated at 35% from past experience. Determine the sample size. [Ans. 350]

3. Write the sort note on organization aspect of sample survey.

4. How do you determine the sample size in any investigation?

5. What do mean by questionnaire design?

6. Write the different type of questionnaire design.

7. What are the criteria for good questionnaire design?

8. Write the different between sampling and non-sampling error.

Exercise 5.2

Multiple Choice Questions circle (O) the correct answer.

1. A population consisting of all the items which are physically present is called
(a) hypothetical population (b) Real population
(c) Infinite population (d) None of the above

2. A population consisting of the results of the conceptually repeated trials is known as
(a) hypothetical population (b) finite population
(c) infinite population (d) real population

3. Mistakes in sample refers to
(a) Wrong columns or cells (b) use of the wrong formulae
(c) Bias of sample (d) None sampling errors

4. Sampling error in sample may be reduced by
(a) Increasing the sample size (b) decreasing in the sampling size
(c) No charge in sample size (d) All of these

5. Determination of sample size is done by formula
(a) $\left(\frac{Z_\alpha \sigma}{E}\right)^2$ (b) $\left(\frac{a\sigma}{E}\right)^2$ (c) $\left(\frac{\sigma}{E}\right)^2$ (d) $\left(\frac{1}{2}\right)^2$

6. Determination of sample size is done by
(a) $\left(\frac{z_\alpha}{E}\right)^2$ (b) $\left(\frac{z_\alpha}{E}\right)^2 PQ$ (c) PQ (d) $1 - P$

7. There are more chances of non-sampling errors than sampling errors in case of
(a) studies of large samples (b) complete enumeration

(c) in efficient investigations (d) all the above

8. Increasing in reliability and accuracy of results from a sampling study with the increase in sample size is known as the principle of

(a) optimization (b) statistical regulations

(c) law of increasing returns (d) inertia of large numbers

9. The magnitude of the standard error of an estimate is an index of its

(a) accuracy (b) precism (c) efficiency (d) all the above

Answer Key

1. (b)	2. (a)	3. (a)	4. (a)	5. (a)	6. (b)	7. (d)	8. (d)	9. (c)

***