13

Sampling and Sampling Distributions

INTRODUCTION

The need for adequate and reliable data is ever increasing for taking wise decisions in different fields of human activity and business is no exception to it. There are two ways in which the required information may be obtained :

1. Complete enumeration survey or census method, and
2. Sampling method.

Under complete enumeration survey method, data are collected for each and every unit (person, household, field, shop, factory, etc., as the case may be) belonging to the population or universe which is the complete set of items which are of interest in any particular situation. For example, we are interested in knowing consumers' reactions to a particular product. We may contact each and every person who uses that product or just take a sample of person. In the former case, we are using *census method* while in the latter *sample method.*

The main advantages of census method are :

1. *Information can be obtained for each and every unit.* If the information is required for each and every unit in the domain of study, a complete enumeration survey is necessary. Data for every individual or unit are necessary in cases where the action is taken separately for each one of them. Examples of such situations are recruitment of personnel in an establishment, preparation of voters' list for election purposes, income tax assessment, where the income of each individual is assessed and taxed.

2. *Greater Accuracy.* The results of a complete enumeration survey are expected to be more accurate than the sample method because information is obtained for each and every unit. However, it may be noted that a complete enumeration survey need not necessarily provide us with accurate information as evidenced by the census experience of a number of countries. The errors in a complete enumeration survey arise mainly from incomplete coverage, observational and tabulation errors due to the difficulties encountered in organising a survey on such a large scale and in getting adequate trained personnel to carry out the survey.

The effort, money and time required for carrying out complete enumeration will generally be extremely large, and, in many cases, cost may be so prohibitive that the very idea of collecting information by this method may have to be dropped. The choice, then, may be either no data or data through the sampling method. Unless the information is required for each and every unit in the domain of study, the sampling technique is generally used to obtain the information.

In the sampling method, instead of every unit of the population only a part of the population is studied and the conclusions are drawn on that basis for the entire population.

Although much of the development in the theory of sampling has taken place only in recent years, the idea of sampling is pretty old. A housewife examines only two or three grains of boiling rice to know, whether the pot of rice is ready or not. A doctor examines a few drops of blood and draws conclusion about the blood constitution of the whole body. A businessman places orders for materials by examining only a small smaple of the same. A teacher may put questions to one or two students and find out whether the class as a whole is following the lesson. In fact, there is hardly any field where the technique of sampling is not used either consciously or unconsciously.

Purpose of Sampling. A sample is not studied for its own sake. The basic objective of its study is to draw inference about the population. In other words, sampling is only a tool which helps to know the characteristics of the universe or population by examining only a small part of it. The values obtained from the study of sample, such as the average and variance, are known as *'statistic'*. On the other hand, such values for the population are called *'parameters'*.

Principles of Sampling

There are two important principles, on which the theory of sampling is based:

1. Principle of 'Statistical Regularity', and
2. Principle of 'Inertia of Large Numbers'.

Principle of Statistical Regularity

This principle is derived from the mathematical theory of probability. In the words of King, "*The law of statistical regularity lays down that a moderately large number of items chosen at random from a large group are almost sure on the average to possess the characteristics of the large group*". In other words, this principle points out that if a sample is taken at random from a population, it is likely to possess almost the same characteristics as that of the population. This principle directs our attention to one very important point, that is, the desirability of choosing the sample at random.

By random selection, we mean a selection where each and every item of the population has an **equal** chance of being selected in the sample. In other words, the selection must not be made by deliberate exercise of one's discretion. A sample selected in this manner would be representative of the population. If this condition is satisfied, it is possible for one to depict fairly, accurately the characteristics of the population by studying only a part of it. Hence, this principle is of great practical significance because it makes possible a considerable reduction of the work necessary before any conclusion is drawn regarding a large population. For example, if one intends to make a study of the cigarette buying habits of the students of Delhi University, it is not necessary to study each and every student, a few students may be selected at random from every college, and on that basis inference may be drawn for all the students of the University.

It should be noted that the results derived from sample data may be different from that of the population. This is for simple reason that the sample is only a part of the whole population.

Principle of Inertia of Large Numbers

This principle is a corollary of the principle of statistical regularity. It is of great significance in the theory of sampling. It states that, other things being equal, larger the size of the sample, more accurate the results are likely to be. This is because large numbers are more stable as compared to small ones. The difference in the aggregate result is likely to be significant when the number in the sample is large, because when large numbers are considered, the variations in the component parts tend to balance each other and, therefore, the variation in the aggregate is insignificant. For example, if a coin is tossed 10 times, we should expect an equal number of heads and tails, *i.e.*, 5 each. But since the experiment is tried a small number of times, it is likely that we may not get exactly 5 heads and 5 tails. The result may be combination of 9 heads and 1 tail, or 8 heads and 2 tails, or 7 heads and 3 tails etc. If the same experiment is carried out 1,000 times, the chance of getting 500 heads and 500 tails would be very high, *i.e.*, the results would be very near to 50% heads and 50% tails. The basic reason for such likelihood is that the experiment has been carried out a sufficiently large number of times and possibility of variations in one direction compensating for others in a different direction is greater. If at one time we get continuously 5 heads, it is likely that at other times we may get continuously 5 tails, and so on, and for the experiment as a whole, the number of heads and tails may be more or less equal. Similarly, if it is intended to study the variation in the production of rice over a number of years and data are collected from one or two States only, the result would reflect large variation in production due to the favourable or unfavourable factors in operation. If,on the other hand, figures of production are collected for all the States in India, it is quite likely that we find little variation in the aggregate. This does not mean that the production would remain constant for all years. It only implies that the changes in the production of the individual States will be counterbalanced and reflect small variation in production for the country as a whole.

Methods of Sampling

When a sample is required to be reflected from a population, it is necessary to decide which method should be applied. The various methods of sampling or sampling designs can be grouped under the heads as random sampling and non-random sampling. Random sampling is also referred to as probability sampling, since, if the sampling process is random, the laws of probability can be applied; thus the pattern of sampling distribution needed to interpret and evaluate a sample is provided. A non-random sample is selected on a basis other than probability considerations such as expert judgment, convenience or some other criteria. The most important aspect of non-random sampling worth noting is that it is subjected to sampling variability but there is no way of knowing the pattern of variability in the process.

We shall now discuss some of the various sampling methods under two separate headings as follows:

A. Random sampling methods :

(*i*) Simple Random Sampling

(*ii*) Stratified Sampling

(*iii*) Systematic Sampling

(*iv*) Multi-stage Sampling

B. Non-random sampling methods :

(*i*) Judgment Sampling

(*ii*) Quota Sampling

(*iii*) Convenience Sampling.

A. RANDOM SAMPLING METHODS

I. Simple Random Sampling*

Simple random sampling refers to the sampling technique in which each and every item of the population is given an equal chance of being included in the sample. The selection is thus free from personal bias because the investigator does not exercise his discretion of preference in the choice of items. Since selection of items in the sample depends entirely on chance, this method is also known as the method of chance selection. Some people believe that randomness of selection can be achieved by unsystematic and haphazard procedures. But this is quite wrong. However, the point to be emphasized is that unless precaution is taken to avoid bias and a conscious effort is made to ensure the operation of chance factors, the resulting sample shall not to be a random sample.

Random sampling is sometimes referred to as 'representative sampling'. If the sample is chosen at random and if the size of the sample is sufficiently large, it will represent all groups in the population. A

*Simple random samples are characterised by the way in which they are selected. Since, a simple random sample is drawn by chance selection, it must be differentiated from selection in a haphazard or hit-and-miss manner.

random sample is also known as a 'probability sample' because every item of the population has an equal opportunity of being selected in the sample.

Random sampling is not always used as primary sampling procedure. However, it is necessary to introduce an element of randomness in the final selection of items. For example, with each group the choice of cases to constitute the sample should be based on chance selection. If the element of randomness is not introduced, bias is likely to enter and make the sample unrepresentative.

Methods of Obtaining a Simple Random Sample

To ensure randomness of selection, one may adopt any of the following methods:

1. Lottery Method. This is a very popular method of taking a random sample. Under this method, all items of the population are numbered or named on separate slips of paper of identical size,colour and shape. These slips are then folded and mixed up in a container or drum. A blindfold selection is then made of the number of slips required to constitute the desired size of sample. The selection of items thus depends entirely on chance. The method would be quite clear with the help of an example. If we want to take a sample of 10 persons out of a population of 100, the producer is to write the names of all the 100 persons on separate slips of paper, fold these slips, mix them thoroughly and then make a blindfold selection of 10 slips.

The above method is very popular in lottery drawn, where a decision about prizes is to be made. However, while adopting lottery method, it is absolutely essential to see that the slips are of identical size, shape and colour, otherwise there is a lot of possibility of personal prejudice and bias affecting the results.

2. Table of Random Numbers*. The lottery method discussed above becomes quite cumbersome to use as the size of population increases. An alternative method of random selection is that of using the table of random numbers. A number of random number tables are available such as: (*i*) Tippett's table of random numbers, (*ii*) Fisher and Yates numbers, and (*iii*) Kendall and Babington Smith numbers. Tippett's numbers are most popular. They consist of 41,600 digits taken from census reports and combined by fours to give 10,400 four-figure numbers. We give here the first forty sets as an illustration of the general appearance of random numbers :

2952	6641	3992	9792	7969	5911	3170	5624	4167	9524
1542	1396	7203	5356	1300	2693	2370	7483	3408	2792
3563	1089	6913	7691	0560	5246	1112	6107	6008	8126
4233	8776	2754	9143	1405	9025	7002	6111	8816	6446

It is important that the starting point in the table of random numbers be selected in some random fashion so that every unit has an equal chance of being selected.

One may question, and quite rightly, as to how it is ensured that these digits are random. It may be pointed out that the digits in the table were chosen haphazardly, but the real guarantee of their randomness lies in practical complication. Tippett's table of random numbers may be used which is given below. Suppose we have to select 20 items out of 6,000. The procedure is to number all the 6,000 items from 1 to 6,000. A page from Tippett's table may then be consulted and the first twenty numbers up to 6,000 noted down. Items bearing these numbers will be included in the sample. Making use of the portion of the table, given above, the required numbers are :

2952	3992	5911	3170	5624	4167	1542	1396	5356	1300
2693	2370	3408	2762	3563	1089	0560	5246	1112	4233

The items which bear the above numbers constitute the sample.

*These days, Computerised Random Numbers are more popular in use.

Population size less than 1,000. If the size of population is less than 1,000, the procedure will be different, as Tippett's numbers are available only in four figures. Thus, for example, if it is desired to take a sample of 10 items out of 400, all items from 1 to 400 should be numbered as 0001 to 0400. We may now select 10 numbers from the table which are up to 0400.

Population size less than 100. If the size of population is less than 100, the table is used as follows: Suppose ten numbers from out of 0 to 80 are required. We start anywhere in the table and write down the numbers in pairs. The table can be read horizontally, vertically, diagonally, or in any other methodical way. Starting with the first and reading horizontally (please see the part of table given on page 461), we obtain 29, 52, 66, 41, 39, 92, 97, 92, 79, 69, 59, 11, 31, 70, 56, 24, 41, 67, and so on. Ignoring the numbers greater than 80, we obtain for our purpose ten random numbers, namely, 29, 52, 66, 41, 39, 79, 69, 59, 11 and 31.

Fisher and Yates tables consist of 15,000 numbers. These have been arranged in two digits in 300 blocks, each block consisting of 5 rows and 5 columns. Kendall and Smith also constructed random numbers (10,000 in all) by using a randomising machine. However, this method of random selection cannot be followed in case of articles like ghee, oil, petrol, wheat, etc.

Merits. 1. Since the selection of items in the sample depends entirely on chance, there is no possibility of personal bias affecting the results.

2. As compared to judgment sampling, random sample represents the population in a better way. As the size of the sample increases, it becomes increasingly representative of the population.

3. The analyst can easily assess the accuracy of his estimate because sampling errors follow the principle of chance. The theory of random sampling is further developed than that of any other type of sampling which enables the surveyor to provide the most reliable information at the least cost.

Limitations. 1. The use of random sampling necessitates a completely catalogued population from which to draw the sample. But, it is often difficult for the investigator to have up-to-date lists of all the items of the population to be sampled. This restricts the use of any sampling method.

2. The task of preparing slips is time-consuming and expensive. However, this difficulty can at times be overcome by following regular interval sampling method which enable a random sample to be drawn without preparing slips.

3. The size of the sample required to ensure statistical reliability is usually large under random sampling than in stratified sampling.

4. From the point of view of field survey, it has been claimed that cases selected by random sampling tend to be too widely dispersed geographically and that the time and cost of collecting data become too large.

5. Random sampling may produce the most non-random looking results. For example, thirteen cards from a well-shuffled pack of playing cards may consist of one suit. But the probability of this type of incidence is very-very small.

II. Stratified Sampling

Stratified random sampling is one of the restricted random methods which, by using available information concerning the data attempts to design a more efficient sample than that obtained by the simple random procedure. The process of stratification requires that the population may be divided into homogeneous groups or classes called *strata*. Then a sample may be taken from each group by simple random method, and the resulting sample is called a stratified sample.

A stratified sample may be either proportional or disproportionate. In a proportional stratified sampling plan, the number of items drawn from each stratum is proportional to the size of the strata. For instance, if the population is divided into four strata, their respective sizes being 15, 10, 20, 55 per cent of the population and a sample of 1,000 is to be drawn, the desired proportional sample may be obtained in the following manner:

From stratum	one	1,000 (0.15) =	150 items
" "	two	1,000 (0.10) =	100 "
" "	three	1,000 (0.20) =	200 "
" "	four	1,000 (0.55) =	550 "
	Sample size	=	1,000

Proportional stratification yields a sample that represents the population with respect to the proportion in each stratum in the population. This procedure is satisfactory if there is no great difference in variation from stratum to stratum. But, it is certainly not the most efficient procedure, especially when there is considerable variation in different strata. This indicates that in order to obtain maximum efficiency in stratification, we should assign greater representation to a stratum with a large variation and smaller representation to one with small variation. For instance, in conducting an all-India survey of the market of a new product, all the States of India may be taken as strata. If the potential consumers of a given State are 10 per cent of all the consumers, but according to our information the product in that State is bound to have a market, we may take, say, 1 per cent or 2 per cent of our sample from that State. If, however, the outcome of another State is highly doubtful we may decide to give it a much greater representation in the sample from its relative size. A sample, thus obtained is a disproportionate stratified sample. Disproportionate stratified sampling also includes procedures of taking an equal number of items from each stratum irrespective of its size.

Merits. 1. Since the population is first divided into various strata and then a sample is drawn from each stratum there is little possibility of any essential group of the population being completely excluded. A more representative sample is thus secured. Stratified sampling is frequently regarded as the most efficient system of sampling.

2. Stratified sampling ensures greater accuracy. The accuracy is maximum if each stratum is so formed that it consists of uniform or homogeneous items.

3. As compared to random sample, stratified samples can be more concentrated geographically. Thus, the time and expense of interviewing may be considerably reduced.

Limitations. 1. Utmost care must be exercised in dividing the population into various strata. Each stratum must contain, as far as possible, homogeneous items as otherwise the results may not be reliable. However, this is a very difficult task and may involve considerable time and expense.

2. The items from each stratum should be selected at random. But, this may be difficult to achieve in the absence of skilled sampling supervisors and a random selection within each stratum may not be ensured.

III. Systematic Sampling

This method is popularly used in those cases where a complete list of the population from which sampling is to be drawn is available. The method is to select every *k*th* item from the list where '*k*' refers to the sampling interval. The starting point between the first and the *k*th item is selected at random.

$$*k = \frac{\text{Size of Population}}{\text{Sample size}} = \frac{N}{n}$$

For example, if a complete list of 1,000 students of a college is available and if we want to draw a sample of 200 students, this means we must take every fifth item (*i.e.*, $k = 5$). The first item between one and five shall be selected at random. Suppose it comes out to be three. Now, we shall go on adding five and obtain numbers of the desired sample. Thus, the second item would be the 8th student; the third, 13th student; the fourth, 18th student; and so on.

Systematic sampling is relatively a simple technique and may be more efficient than simple random sampling, provided the lists are arranged wholly as random. However, it is rare that this requirement is fulfilled. The nearest approach to randomness is provided by alphabetical lists such as are found in telephone directory, although even these may have certain non-random characteristics.

Merits. The systematic sampling is more convenient to adopt than the random sampling or the stratified sampling method. The time and work involved in sampling by this method are relatively smaller. The results obtained are also found to be generaly satisfactory provided care is taken to see that there are no periodic features associated with the sampling interval. If populations are sufficiently large, systematic sampling can often be expected to yield results that are similar to those obtained by proportional stratified sampling.

Limitations. Systematic sampling becomes a less representative design than simple random sampling if we are dealing with populations having hidden periodicities. For example, if the sales of every seventh day of the calendar year are included, the sample will contain, say, all Mondays or all Fridays. If there is a definite repetitive weekly pattern in sales (which is usually the case), our sample is not representative at all of sales for the whole year and consequently, the sample results may be seriously biased.

IV. Multi-stage Sampling

As the name implies, this method refers to a sampling procedure which is carried out in several stages. The material is regarded as made up of a number of first stage sampling units, each of which is made of a number of second stage units, etc. At first, the first stage units are sampled by some suitable method, such as random sampling. Then, a sample of second stage units is selected from each of the selected first stage units again by some suitable method which may be the same or different from the method employed for the first stage units. Further stages may be added as required. The procedure may be illustrated as follows:

Suppose, we want to take a sample of 5,000 households from the State of U.P. At the first stage, the State may be divided into a number of districts and a few districts selected at random. At the second stage, each district may be sub-divided into a number of villages and a sample of villages may be taken at random. At the third stage, a number of households may be selected from each of the villages selected at the second stage. In this way, at each stage, the sample size becomes smaller and smaller.

Merits. Multi-stage sampling introduces flexibility in the sampling method which is lacking in other methods. It enables existing divisions and sub-divisions of the population to be used as units at various stages, and permits the field work to be concentrated and yet large area to be covered. Another advantage of the method is that sub-division into second stage unit (*i.e.*, the construction of the second stage frame) need be carried out for only those first stage units which are included in the sample. It is, therefore, particularly valuable in surveys of underdeveloped areas where no frame is generally sufficiently detailed and accurate for sub-division of the material into reasonably small sampling units.

Limitations. However, a multi-stage sample is in general less accurate than a sample containing the same number of final stage units which have been selected by some suitable stage process.

B. NON-RANDOM SAMPLING METHODS

Random selection is generally recommended for large surveys, but certain types of non-random selection are sometimes justified. A few of the most important of these types are:

I. Judgment Sampling

In this method of sampling, the choice of sample items depends exclusively on the judgment of the investigator. In other words, the investigator exercises his judgment in the choice of sample items and includes those items in the sample which he thinks are most typical of the population with regard to the characteristics under investigation. For example, if a sample of ten students is to be selected from a class of sixty for analysing the spending habits of students, the investigator would select 10 students who, in his opinion, represent the class.

This method, though simple, is not scientific because the results may be considerably affected by the personal prejudice or bias of the investigator. Thus, judgment sampling involves the risk that the investigator may establish foregone conclusions by including those items in the sample which conform to his preconceived notions. For example, if an investigator holds the view that the wages of workers in a certain establishment are very low, and if he adopts the judgment sampling method, he may include only those workers in the sample whose wages are low and thereby establish his point of view which may be far from the truth.

Even though the principles of sampling theory are not applicable to judgment sampling, this method is often used in solving many types of economic and business problems such as:

(*i*) Judgment sampling is used when size of sample is small. In such a case, simple random sample may miss the more important elements, whereas judgment selection would certainly include them in the sample.

(*ii*) In solving everyday business problems and making public policy decisions, executives and public officials are often pressed for time and cannot wait for probability sample designs. Judgment sampling is then the only practical method, since estimates can be made available quickly that will enable businessmen and governmental officials to arrive at solutions to their urgent problems that are better than decisions made without any statistical data.

(*iii*) Judgment sampling may be used to conduct pilot studies. In any case, the reliability of sample results in judgment sampling depends on the quality of the sampler's expert knowledge or judgment. If it is good and is carefully and skilfully applied, judgment samples may be expected to be representative and to yield valuable results. On the other hand, when a sample is obtained with poor judgment, serious bias will be present.

The success of this method depends upon the excellence in judgment. If the individual making decisions is knowledgeable about the population and has good judgment, then the resulting sample may be representative, otherwise the inferences based on the sample may be erroneous. It may be noted that even if a judgment is reasonably representative, there is no objective method for determining the size or likelihood of sampling error. This is a big defect of the method.

II. Quota Sampling

Quota sampling is a type of judgment sampling. In a quota sample, quotas are set up according to given criteria, but, within the quotas the selection of sample items depends on personal judgment. For example, in a radio listening survey, the interviewers may be told to interview 500 people living in a certain area and that out of every 100 persons interviewed 60 are to be housewives, 25 farmers and 15 children under the age of 15. Within these quotas, the interviewer is free to select the people interviewed.

The cost per person interviewed may be relatively small for a quota sample but there are numerous opportunities for biases which may invalidate the results. For example, interviewers may miss farmers working in the fields or talk with those housewives who are at home. If a person refuses to respond, the interviewer simply selects someone else. Because of the risk of personal prejudice and bias entering the process of selection, the quota sampling is rarely used in practical work.

III. Convenience Sampling

The method of convenience sampling is also called the chunk. A chunk is a fraction of one population taken for investigation because of its convenient availability. Thus, a chunk is selected neither by probability nor by judgment but by convenience. A sample obtained from readily available lists, such as telephone directories or automobile registrations, is a convenience sample and not a random sample, even if the sample is drawn at random from the lists.

Convenience samples are sometimes called accidental samples because those entering into the sample enter by "accident"—they just happen to be at the right place and at the right time, that is, where and when the information for the study is being collected. The problem with convenience samples is that we have no way of knowing if those included in the sample are representatives of the target population.

A chunk—which is merely a convenient slice of the population—can hardly be representative of the population. Its results are generally biased and unsatisfactory. Formerly, the chunk was frequently used in public opinion surveys when interviewers stopped near the railway station or the bus stop or in front of office building to interview people. Today, accountants still use convenience sampling to analyse or audit accounts.

Convenience sampling is also useful in making pilot studies. Questions may be tested and preliminary information may be obtained by the chunk before the final sampling design is decided upon.

Size of Sample

An important decision that has to be taken while adopting a sampling technique is about the size of the sample. Different opinions have been expressed by experts on this point. For example, some have suggested that the sample size should be 5% of the size of population while others are of the opinion that sample size should be at least 10%. However, these views are of little use, as in practice, appropriate sample size depends on various factors relating to the subject under investigation like the time aspect, the cost aspect, the degree of accuracy desired, etc. Sampling theory is of little help in arriving at a good estimate of the sample size in any particular situation. However, the following two considerations may be kept in mind in determining the appropriate size of the sample.

1. The size of the sample should increase as the variation in the individual items increases.
2. The greater the degree of accuracy desired, the larger should be the sample size.

Merits of Sampling Method

The sampling method has the following merits over the complete enumeration survey method :

1. *Less time.* Since the sample is a study of part of the population, considerable time and labour are saved when a sample survey is carried out. Time is saved not only in collecting data but also in processing it. For these reasons, a sample provides more timely data in practice than a census.

2. *Less cost.* The amount of effort and expenses involved in collecting information is always greater per unit of the sample than a complete census, the total financial burden of a sample survey is generally less than that of a complete census. This is because of the fact that in sampling, we study only a part of the population and the total expense of collecting data is less than that required when the census method is adopted. This is a great advantage particularly in an underdeveloped economy where much of the information would be difficult to collect by the census method for lack of adequate resources.

3. *More reliable results.* Although the sampling technique involves certain inaccuracies owing to sampling errors, the result obtained is generally more reliable than that obtained from a complete count. There are several reasons for it. *First,* it is always possible to determine the extent of sampling errors. *Secondly,* other types of errors to which a survey is subjected, such as inaccuracy of information, incompleteness of returns, etc., are likely to be more serious in a complete census than in a sample survey. This is because more effective precautions can be taken in a sample survey to ensure that the information is accurate and complete. Moreover, it is possible to avail of the services of experts and to impart thorough training to the investigators in sample survey which further reduces the possibility of errors. Follow-up work can also be undertaken much effectively in the sampling method. Indeed, even a complete census can only be tested for accuracy by some type of sampling check.

4. *More detailed information.* Since the sampling technique saves time and money, it is possible to collect more detailed information in a sample survey. For example, if the population consists of 1,000 persons in a survey of the consumption pattern of the people, the two alternative techniques available are as follows :

(*a*) We may collect the necessary data from each one of the 1,000 persons through a questionnaire containing, say, 10 questions (census method), or

(*b*) We may take sample of 100 persons (*i.e.*, 10% of population) and prepare a questionnaire containing as many as 100 questions. The expenses involved in the latter case would almost be the same as in the former, but it will enable many times more information to be obtained.

5. *The destructive nature of certain tests.* Many tests are of destructive nature. Steel plates, wires and other similar products often must have a certain minimum tensile strength. To ensure that the product meets the minimum standard, a relatively small sample is selected. Each piece is stretched until it breaks, and the breaking point (usually measured in pounds) is recorded. Obviously, if all the wires or all the plates were tested for tensile strength, none would be available for sale or use. And for the same reason, only a sample of photographic film is selected to determine the quality of all the film produced, only a few seeds are tested for germination prior to planting season and only a few chalks out of a certain lot are tested for ascertaining the breaking strength.

Limitations of Sampling

Despite the various advantage of sampling, it is not altogether free from limitations. Some of the precautions involved in sampling are given below :

(*i*) A sample survey must be carefully planned and executed, otherwise, the results obtained may be inaccurate and misleading. Of course, even for a complete count care must be taken but serious errors may arise in sampling, if the sampling procedure is not perfect.

(*ii*) Sampling generally requires the services of experts, if only for consultation purposes. In the absence of qualified and experienced persons, the information obtained from sample surveys cannot be relied upon. A shortage of experts in the sampling field is a serious hurdle in the way of reliable statistics.

(*iii*) At times, the sampling plan may be so complicated that it requires more time, labour and money than a complete count. This is so if the size of the sample is a large proportion of the total population and if complicated weighted procedures are used. With each additional complication in the survey, the chances of errors multiply and greater care has to be taken which, in turn, means more time and labour.

(*iv*) If the information is required for each and every unit in the domain of study, a complete enumeration survey is necessary.

Sampling and Non-Sampling Errors

The term 'error' refers to the difference between the value of a 'sample statistic' and that of corresponding 'population parameter'. Various forces combine to produce deviations of sample statistic from population parameters, and errors, in accordance with the different causes, are classified into sampling and non-sampling errors.

The error arising due to drawing inferences about the population on the basis of few observations (sample) is termed as sampling error. Clearly, the sampling error in this sense is non-existent in a complete enumeration survey, since the whole population is surveyed. However, the errors mainly arising at the stages of ascertainment and processing of data which are termed non-sampling errors are common both in complete enumeration and sample surveys.

I. Sampling Errors

Even if utmost care has been taken in selecting a sample, the results derived from the sample may not be representative of the population from which it is drawn, because samples are seldom, if ever, perfect miniatures of the population. This gives rise to sampling errors. Sampling errors arise due to the fact that samples are used and to the particular method used in selecting the items from the population.

Sampling errors are of two types—biased and unbiased.

(1) *Biased errors.* These errors arise from any bias* in selection, estimation, etc. For example, if in place of simple random sampling, deliberate sampling has been used in a particular case, some bias is introduced in the result and hence such errors are called biased sampling errors.

(2) *Unbiased errors.* These errors arise due to chance differences between the members of population included in the sample and those not included.

Thus, the total sampling errors is made up of errors due to bias, if any, and the random sampling error. The essence of bias is that it forms a constant component of error that does not decrease in a large population as the number in the sample increases. Such error is, therefore, also known as *cumulative* or *non-compensating error*. The random sampling error, on the other hand, decreases on an average as the size of the sample increases. Such error is, therefore, also known as non-cumulative of compensating error.

Causes of Bias

Bias may arise due to:

(1) faulty process of selection;

(2) faulty work during the collection of information; and

(3) faulty methods of analysis.

(1) Faulty Selection. Faulty selection of the sample may give rise to bias in a number of ways, such as :

(*a*) *Deliberate selection* of a 'representative' sample.

(*b*) Conscious or unconscious *bias in the selection of a 'random' sample.* The randomness of selection may not really exist even though the investigator claims that he has a random sample if he allows his desire to obtain a certain result to influence his selection.

(*c*) *Substitution.* Substitution of an item in place of one chosen in a random sample sometimes leads to bias. Thus, if it were decided to interview every 50th householder in the street, it would be inappropriate to interview the 51st or any other number in his place as the characteristics possessed by them may differ from those who were originally to be included in the sample.

*Bias is said to exist when the value of a sample statistic shows a persistent tendency to deviate in one direction from the value of the parameter.

(*d*) *Non-response.* If all the items to be included in the sample are not covered, there will be bias even though no substitution has been attempted. This fault particularly occurs in mailed questionnaires, which are incompletely returned. Moreover, the information supplied by the informants may also be biased.

(*e*) *An appeal to the vanity* of the person questioned may give rise to yet another kind of bias. For example, the question 'Are you a good student ?' is such that most of the students would succumb to vanity and answer 'Yes'.

(2) Bias due to Faulty Collection of Data. Any consistent error in measurement will give rise to bias whether the measurements are carried out on a sample or on all the units of the population. The danger of error is, however, likely to be greater in sampling work, since the units measured are often smaller. Bias may arise due to improper formulation of the decision problems wrongly defining the population, specifying the wrong decision, securing and inadequate frame, and so on. Biased observations may result from a poorly designed questionnaire, an ill-trained interviewer, failure of a respondent's memory, etc. Bias in the flow of data may be due to unorganised collection procedure, faulty editing or coding of responses.

(3) Bias in Analysis. In addition to bias which arises from faulty process of selection and faulty collection of information, faulty method of analysis may also introduce bias. Such bias can be avoided by adopting the proper methods of analysis.

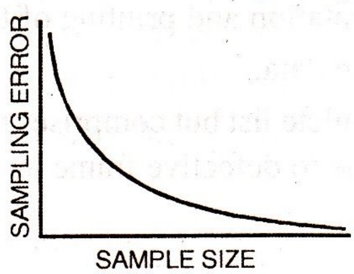
Avoidance of Bias

If possibilities of bias exist, fully objective conclusions cannot be drawn. The first essential of any sampling or census procedure must, therefore, be the elimination of all sources of bias. The simplest and the only certain way of avoiding bias in the selection process is for the sample to be drawn either entirely at random or at random subject to restrictions, which, while improving the accuracy, are of such a nature that they do not introduce bias in the results. In certain cases, systematic selection may also be permissible.

Method of Reducing Sampling Errors

Once the absence of bias has been ensured, attention should be given to the random sampling errors. Such errors must be reduced to the minimum so as to attain the desired accuracy.

Apart from reducing errors of bias, the simplest way of increasing the accuracy of a sample is to increase its size. The sampling error usually decreases with increase in sample size (number of units selected in the samples) and in fact in many situations the decrease is inversely proportional to the square root of the sample size as can be seen from the diagram given below.



From this diagram, it is clear that though the reduction in sampling error is substantial for initial increases in sample size, it becomes marginal after a certain stage. In other words, considerably greater effort is needed after a certain stage to decrease the sampling error than in the initial instances. Hence, after that stage sizable reduction in cost can be achieved by lowering even slightly the precision required. From this point of view, there is a strong case for resorting to a sample survey to provide estimates within permissible margins to error instead of a complete enumeration survey, as in the latter, the effort and the cost needed will be substantially higher due to the attempt to reduce the sampling error to zero.

II. Non-sampling Errors

When a complete enumeration of units in the universe is made, one would expect that it would give rise to data free from errors. However, in practice, it is not so. For example, it is difficult to completely avoid errors of observation or ascertainment. So, also in the processing of data tabulation errors may be committed affecting the final results. Errors arising in this manner are termed *non-sampling errors*, as they are due to factors other than the inductive process of inferring about the population from a sample. Thus, the data obtained in an investigation by a complete enumeration, although free from sampling error, would still be subjected to non-sampling error, whereas the result of a sample survey would be subjected to sampling error as well as non-sampling error.

As regards non-sampling errors, they are likely to be more in case of complete enumeration survey than in case of a sample survey, since it is possible to reduce the non-sampling errors to a greater extent by using better organisation and suitably trained personnel at the field and tabulation stages. The behaviour of the non-sampling error with increase in sample size is likely to be opposite of that of sampling error, that is, the non-sampling error is likely to increase with increase in sample size. In many situations, it is quite possible that the non-sampling error in a complete enumeration survey is greater than both the sampling and non-sampling errors taken together in a sample survey, and naturally in such a situation, the latter is to be preferred to the former.

Non-sampling errors can occur at every stage of planning and execution of the census or survey. Such errors can arise due to a number of causes such as defective methods of data collection and tabulation, faulty definition, incomplete coverage of the population or sample, etc. More specifically, non-sampling errors may arise from one or more of the following factors :

1. Data specification being inadequate and inconsistent with respect to the objectives of the census or survey.

2. Omission or duplication of units due to imprecise definition or boundaries of area units, incomplete or wrong identification of units or faulty methods of enumeration.

3. Inaccurate or inappropriate methods of interview, observation or measurement with inadequate or ambiguous schedules, definitions or instructions.

4. Lack of trained and experienced investigators.

5. Lack of adequate inspection and supervision of primary staff.

6. Errors due to non-response, *i.e.*, incomplete coverage in respect of units.

7. Errors in data processing operations such as coding, punching, certification, tabulation, etc.

8. Errors committed during presentation and printing of tabulated results.

9. Inadequate scrutiny of the basic data.

This should not be taken as a complete list but comprises major sources of error. In a sample survey, non-sampling errors may also arise due to defective frame and faulty selection of sampling units.

Control of Non-sampling Errors

In some situations, the non-sampling errors may be large and deserve greater attention than the sampling errors. While in general, sampling error decreases with increase in sample size, non-sampling error tends to increase with the sample size. In the case of complete enumeration, non-sampling errors and in the case of sample surveys, both sampling and non-sampling errors require to be controlled and reduced to a level at which their presence does not vitiate the use of final results.

In recent years, there has been a growing need for assessing and controlling the non-sampling errors that are likely to arise at the various stages of collection and tabulation of statistical data in large-scale census and surveys. The increasing awareness of the existence of such errors is due to the fairly widespread use of the sampling method, one of the main advantages of which is that it provides an opportunity for greater control of non-sampling errors as well.

SAMPLING DISTRIBUTIONS

Much of the information used in business and industry is gathered by means of sampling. It has been pointed out earlier that not only it is often impossible either physically or because of limitations imposed by time or pecuniary considerations, to take a census of all the items in the population, but it is also usually unnecessary. The results of a properly taken sample, if subjected to rigorous analysis, will ordinarily enable the investigator to arrive at generalisations that are valid for the entire population.

The process of generalising these sample results of the population is referred to as statistical inference. In this chapter, along with the knowledge of certain probability distributions, we shall use certain sample statistics (such as the sample mean, the sample proportion, etc.) in order to estimate and draw inferences about the true population parameters.

For example, in order to be able to use the sample mean to estimate the population mean, we should examine every possible sample (and its mean) that could have occurred in the process of selecting one sample of a certain size. If this selection of all possible samples actually were to be done, the distribution of the results would be referred to as a sampling distribution. Although, in practice, only one such sample is actually selected, the concept of sampling distributions must be examined so that probability theory and its distribution can be used in making inferences about the population parameter values.

Sampling theory has made it possible to deal effectively with these problems. However, before we discuss in detail about them from the standpoint of sampling theory, it is necessary to understand the Central Limit Theorem and the following three probability distributions, their characteristics and relations :

(1) The population (universe) distribution,

(2) The sample distribution, and

(3) The sampling distribution.

Central Limit Theorem. The Central Limit Theorem, first introduced by De Moivre during the early eighteenth century, happens to be the most important theorem in statistics. According to this theorem, if we select a large number of simple random samples, say, from any population distribution and determine the mean of each sample, the distribution of these sample means will tend to be described by the normal probability distribution with a mean μ and variance σ^2/n. This is true even if the population distribution itself is not normal. Or, in other words, we can say that the sampling distribution of sample means approaches to a normal distribution, irrespective of the distribution of population from where sample is taken and approximation to the normal distribution becomes increasingly close with increase in sample size. Symbolically, the theorem can be explained as follows :

When given n independent random variables $X_1, X_2, X_3, \ldots\ldots X_n$, which have the same distribution (no matter what the distribution), then :

$$X = X_1 + X_2 + X_3 + \ldots + X_n$$

is a normal variate. The mean μ and variance σ^2 of X are

$$\mu = \mu_1 + \mu_2 + \mu_3 + \ldots\ldots + \mu_n = n\mu_i$$

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 + \sigma_3^2 + \ldots\ldots + \sigma_n^2 = n\sigma_i^2$$

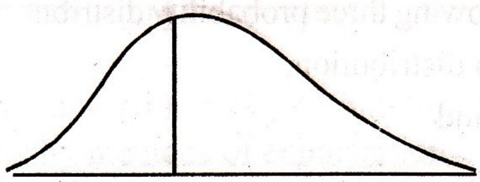
where μ_1 and σ_1^2 are the mean and variance of X_i.

The utility of this theorem is that it requires virtually no conditions on distribution patterns of the individual random variable being summed. As a result, it furnishes a practical method of computing approximate probability values associated with sums of arbitrarily distributed independent random variables. This theorem helps to explain why a vast number of phenomena show approximately a normal distribution. Consider a case when the population is skewed, skewness of the sampling distribution of means is inversely proportional to the square root of the sample size. Consider the case when $n = 16$ that means the sampling distribution of means will exhibit only one-fourth as much skewness as the population has. Consider the case when $n = 100$, skewness becomes one-tenth as much, *i.e.*, as the sample size increases, the skewness will decrease. As a practical consequence, the normal curve will serve as a satisfactory model when samples are small and population is close to a normal distribution, or when samples are large and population is markedly skewed. Because of its theoretical and practical significance, this theorem is considered as most remarkable theoretical formulation of all probability laws.

The Population (Universe) Distribution

When we talk of population distribution, we assume that we have investigated the population and have full knowledge of its mean and standard deviation. For example, a company might have manufactured 1,00,000 tyres of cars in the year 2004. Suppose, it contacts all those who had bought these tyres and gathers information about the life of these tyres. On the basis of the information obtained, the mean of the population which is also called true mean symbolised by μ and its standard deviation symbolised by σ can be worked out. These Greek letters μ and σ are used for these measures to emphasise their difference from corresponding measures taken from a sample. It may be noted such measures characterising a population are called population *parameters*.

The shape of the distribution of the life of tyres may be as follows :



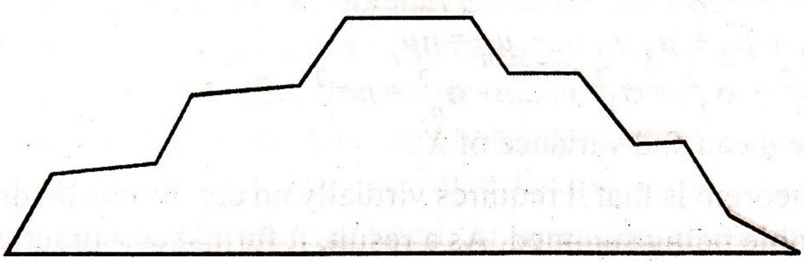
DISTRIBUTION OF THE LIFE OF TYRES

It is clear from above that, though, the distribution shows slight skewness, it does not depart radically from a normal distribution. However, this should not lead one to the conclusion that for sampling theory to apply, it is necessary that the distribution must be normally distributed.

The Sample Distribution

When we talk of a sample distribution, we take a sample from the population. A sample distribution may take any shape. The mean and standard deviation of the sample distribution are symbolised by $\bar{x}$ and s respectively. A measure characterising a sample such as $\bar{x}$ or s is called a sample *statistic*. It may be noted that several sample distributions are possible from a given population.

Suppose, in the above illustration, the manufacturer takes a sample of 500 tyres. He contacts th buyers and enquires about the life of tyres. The shape of the distribution of these tyres may be as follows



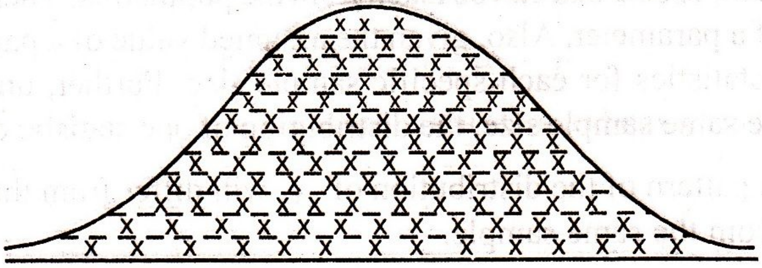
SAMPLE DISTRIBUTION OF 500 TYRES

The mean values of these tyres can be expected to differ somewhat from one sample to another. These sample means constitute the raw material out of which a sampling distribution is constructed.

The Sampling Distribution

Sampling distributions constitute the theoretical basis of statistical inference and are of considerable importance in business decision-making. If we take numerous different samples of equal size from the same population, the probability distribution of all the possible values of a given statistic from all the distinct possible samples of equal size is called a sampling distribution.

It is interesting to note that sampling distributions closely approximate a normal distribution. It can be seen that the mean of a sampling distribution of sample means is the same as the mean of the population distribution from which the samples were taken.* The following diagram would make it clear :



SAMPLING DISTRIBUTION OF SAMPLE MEANS

The mean of the sampling distribution is designated by the same symbol as the mean of the population, namely μ. However, the standard deviation of the sampling distribution of means given a special name, *standard error of mean,* and is symbolised by $\sigma_{\bar{x}}$. The subscript indicates that in this case, we are dealing with a sampling distribution of means.

The greatest importance of sampling distributions is the assistance that they give us in revealing the patterns of sampling errors and their magnitude in terms of standard error. In sampling with replacement, we can observe a good deal of fluctuations in the sample mean as compared to fluctuations in the actual population. The fact that the sample means are less variable than the population data follows logically from an understanding of the averaging process. A particular sample mean averages together all the values in the sample. A population (universe) may consist of individual outcomes that can take on a wide range of values from extremely small to extremely large. However, if an extreme value falls into the sample, although it will have an effect on the mean, the effect will be reduced since it is being averaged in with all the other values in the sample. Moreover, as the sample size increases, the effect of a single extreme value gets even smaller, since it is being averaged with more observations. This phenomenon is expressed statistically in the value of the standard deviation of the sample mean. This is

---

*Let $x_1, x_2, \ldots\ldots x_n$ represent independent random variables corresponding to the n observations in a sample from a population having the same mean μ.

$$\bar{x} = \frac{x_1 + x_2 + \ldots. + x_n}{n}$$

$$\bar{x} = E(\bar{x}) = E\left[\frac{x_1 + x_2 + \ldots. + x_n}{n}\right]$$

$$= \frac{1}{n}\{E[x_1] + E[x_2] + \ldots + E[x_n]\}$$

$$= \frac{1}{n}\{\mu + \mu + \ldots + \mu\} = \frac{1}{n} . n\mu = \mu$$

the measure of variability of the mean from sample to sample and is referrd to as the standard deviation of the sampling distribution of sample mean or *the standard error of the mean* denoted by $\sigma_{\bar{x}}$ and is calculated by*

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

This formula holds only when population is infinite or samples are from finite population with replacement.

It may be noted that in deducing a sampling distribution, we must first make an assumption about the appropriate population parameter. In as much as any value can be assumed for a parameter, depending upon our knowledge or guess of the population, there is no theoretical limit to the number of sampling distribution for the same sample size that can be taken from the population. There is a sampling distribution for each assumed value of a parameter. Also, given the assumed value of a parameter, there is a different sampling distribution of statistics for each specific sample size. Further, under the same assumptions about a population and the same sample size, the distribution of one statistic differs from that of another statistic. For example, the pattern of the distribution of $\overline{X}$ will differ from that of s^2, even though both measures are computed from the same sample.

Relationship between Population, Sample and Sampling Distributions

It will be interesting to note that the mean of the sampling distribution is the same as the mean of the population. It is possible that many sample means may differ from the population mean. However, the sample information can be used as an estimate of population values.

It has also been established that the observed standard deviation of a sample is close to the standard deviation of the population values.

In fact, the standard deviation of the samples is usually so good an approximation that it can safely be used as an estimate of the corresponding population measure. In order to use s of the sample to estimate σ of the population, we make a slight adjustment which has been found to contribute to greater accuracy of the estimate. The adjustment consists of using $(n - 1)$ instead of n in the formula for the standard deviation of a sample, *i.e.*, we use

$$s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} \text{ instead of } \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

The adjustment decreases the denominator and, therefore, gives a larger result. Thus, the estimated standard deviation of the population is slightly larger than the observed standard deviation of the sample.

Sampling Distribution of the Mean

If a population distribution is normal, the sampling distribution of the mean $(\bar{x})$ is also normal for samples of all sizes as can be seen from the following diagram :

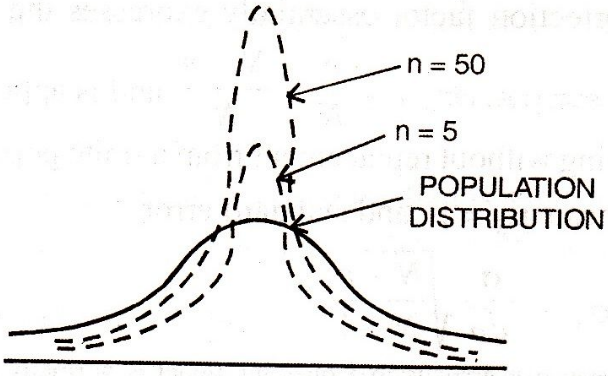
---

*Let $x_1, x_2, \ldots, x_n$ be independent random variables, each having the same variance σ^2.

$$\sigma_{\bar{x}}^2 = \text{Var}\ (\bar{x}) = \text{Var}\ \left[\frac{x_1+x_2+\ldots+x_n}{n}\right]$$

$$= \frac{1}{n^2}\ [\text{Var}\ (x_1) + \text{Var}\ (x_2) + \ldots + \text{Var}\ (x_n)]$$

$$= \frac{1}{n^2}\, n\sigma^2 = \frac{\sigma^2}{n} \quad \text{or} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$



Relationship between a normal population distribution and sampling distribution of the mean for $n = 5$ and $n = 50$.

The following are the important properties of the sampling distribution of mean :

(1) It has a mean equal to the population mean, *i.e.*, $\mu_{\bar{x}} = \mu$.

(2) It has a standard deviation equal to the population standard deviation divided by the square root of the sample size. That is :

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

where $\sigma_{\bar{x}}$ is a measure of the spread of $\bar{x}$ values around μ or a measure of average sampling error or simply stated *standard error of the mean.*

σ = standard deviation of the population

n = size of the sample.

(3) It is normally distributed. The distribution of sample means for large samples is distributed normally whatever the shape of the population distribution, provided σ is finite. Samples of 30 or more items are frequently considered large for statistical purposes. It may be pointed out that, if a population is normal, the distribution of sample means is normal, even if the sample size is small.

It should be noted that $\sigma_{\bar{x}}$ is a measure of the precision with which the sample mean can be used to estimate the true population mean, μ, the standard error, $\sigma_{\bar{x}}$ varies directly with the variation in the original population, σ, and inversely with the square root of the sample size n. Thus, as might be expected, the greater the variation among the items in the original population, the greater is the expected sampling error in using $\bar{x}$ as an estimate of μ. Also the larger the sample size, the smaller the standard error and the smaller the sample size, the larger the standard error.

In practice, the standard deviation of the population is rarely known, and therefore, the standard deviation of the samples which closely approximates the standard deviation of the population is used in place of σ. Hence, the formula for standard error takes the following form :

$$\sigma_{\bar{x}} = \frac{s}{\sqrt{n}}$$

where s refers to the standard deviation of the sample.

The central limit theorem and the standard error of a sample statistic were based upon the premise that the samples selected were chosen with replacement. However, in survey research and in business, sampling is conducted without replacement from populations that are of a finite size N. In these cases, particularly, when the sample size n is not small as compared to the population size N, a *finite population correction factor* should be used in developing the particular sampling distribution.

The finite population correction factor essentially expresses the proportion of observations th have notbeen included in the sample, *viz.*, $1-\frac{n}{N}=\frac{N-n}{N}$ and is approximately equal to $\frac{N-n}{N-1}$ wh N is large. Therefore, in sampling without replacement from a finite population, the sampling distributio of a sample mean will have mean $\mu_{\bar{x}}=\mu$ and standard error,

$$\sigma_{\bar{x}}=\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}.$$

Illustration 1. The time between two arrivals in a queuing model is normally distributed with a mean 2 minutes a standard deviation 0.25 minute. If a random sample of size 36 is drawn, what is the probability that the sample mean will greater than 2.1 minutes ?

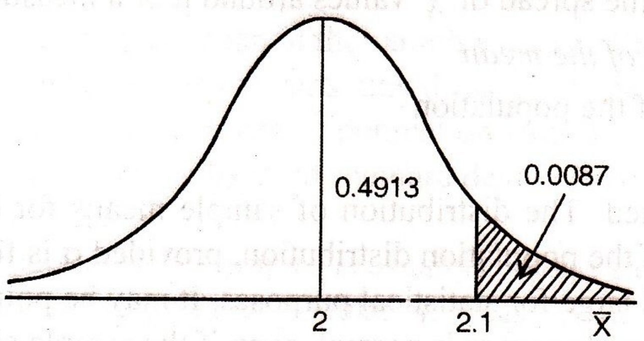
Solution. Since the population is normally distributed, therefore, the sampling distribution of the sample mean will al follow a normal distribution with mean

$$\mu_{\bar{x}}=\mu=2$$

and standard error (s.d. of the sampling distribution)

$$\sigma_{\bar{x}}=\frac{\sigma}{\sqrt{n}}=\frac{0.25}{\sqrt{36}}=0.042$$

Therefore, the probability that the sample mean will be greater than 2.1 minutes is given by $P_r\ [\bar{x}\geq 2.1]$



To get the values from the standard normal distribution, this normal variate $\bar{x}$ must be converted into a standard normal variate by the transformation

$$z=\frac{\bar{x}-\mu}{\sigma_{\bar{x}}}$$

Therefore, the above probability statement becomes

$$P_r\left[\frac{\bar{x}-\mu}{\sigma_{\bar{x}}}\geq\frac{2.1-\mu}{\sigma_{\bar{x}}}\right]$$

or $$P_r\left[z\geq\frac{2.1-2}{0.042}\right]$$

or $$P_r\ [z\geq 2.38].$$

From the table, this value of $z=2.38$ corresponds to the area of 0.4913 to the left of the value $z=2.38$. To get the required probability, this area of 0.4913 must be subtracted from the total area, *i.e.*,

$$P_r\ [\bar{x}\geq 2.1]=0.5-0.4913=0.0087$$

Therefore, in only 0.87% of all possible sample of size $n=36$, the sample mean will be greater than 2.1 minutes.

Distribution of Sample Medians

If a universe is large and can be approximated closely by a normal distribution with a mean μ and a standard deviation σ, the medians of random samples of size n are distributed with a mean μ and a standard deviation $1.2533\ \sigma/\sqrt{n}$, and the distribution of sample medians is nearly normal if n is large.

The standard deviation of the distribution of sample medians is called the *standard error of the sample median* and is denoted by :

$$\sigma_{Med} = 1.2533\ \sigma / \sqrt{n}$$

It should be noted that while the expectation of the median is same as expectation of the mean, the standard error of the median is greater than the standard error of the mean by a multiplier of 1.2533.

Distribution of Sample Standard Deviations

In the analysis of random variables relevant to business problems, it is common that the standard deviation of the population is unknown. In such a case, σ must be estimated by the sample standard deviation. This distribution is defined by a theorem that states :

If a population is large and normally distributed with a standard deviation σ, the standard deviation of the population is unknown. In such a case, σ must be estimated by the sample standard deviation. This distribution is defined by a theorem that states :

If a population is large and normally distributed with a standard deviation σ, the standard deviations of random samples of size n (where n is large), are closely approximated by a normal distribution with a standard deviation $\sigma / \sqrt{2n}$.

The standard deviation of the distribution of standard deviations of samples drawn from a normal population is called the *standard error of the standard deviation* and is denoted by :

$$S = \sigma / \sqrt{2n}$$

where S = Standard error of the standard deviations.

Sampling Distribution of the Difference of the Two Means

Suppose we have two populations, the first of size N_1, with mean μ_1 and standard deviation σ_1, and the second of size N_2, with mean μ_2 and standard deviation σ_2. The comparison is made on the basis of two independent random samples, with one of size n_1 drawn from the first population and the other of size n_2 drawn from the second population. If $\bar{x}_1$ and $\bar{x}_2$ are the two sample means, we can evaluate the possible difference between μ_1 and μ_2, by the difference of the sample means $\bar{x}_1 - \bar{x}_2$. The problem is one of determining the properties of a sampling distribution of $\bar{x}_1 - \bar{x}_2$.

The important properties of the sampling distribution of $\bar{x}_1 - \bar{x}_2$ are :

1. With simple random sampling from two independent populations, the mean of the sampling distribution of $\bar{x}_1 - \bar{x}_2$, denoted by $\mu_{\bar{x}_1 - \bar{x}_2}$ is equal to the difference between the population means, *i.e.*,

$$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_{\bar{x}_1} - \mu_{\bar{x}_2} = \mu_1 - \mu_2$$

2. The standard deviation of the sampling distribution of $\bar{x}_1 - \bar{x}_2$ (also known as standard error of $\bar{x}_1 - \bar{x}_2$) is given by

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\sigma^2_{\bar{x}_1} + \sigma^2_{\bar{x}_2}} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

(since $\bar{x}_1$ and $\bar{x}_2$ are independent random variables, the variance of their difference is the sum of their variances).

3. If $\bar{x}_1$ and $\bar{x}_2$ are the means of two independent samples drawn from two large or infinite populations, the sampling distribution of $\bar{x}_1 - \bar{x}_2$ will be normal if the samples are of sufficiently large size.

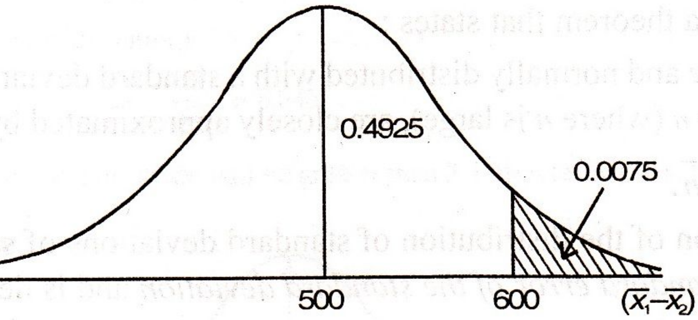
Illustration 2. Strength of the wire produced by company A has a mean of 4,500 kg and a standard deviation of 200 kg. Company B has a mean of 4,000 kg and a standard deviation of 300 kg. If 50 wires of company A and 100 wires of company B are selected at random and tested for strength, what is the probability that the sample mean strength of A will be (i) at least 600 kg more (ii) at least 400 kg more than that of company B. *(MBA, Delhi Univ. 2007)*

Solution. For the sampling distribution of the difference of two means, we have

$$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2 = 4500 - 4000 = 500$$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{40000}{50} + \frac{90000}{100}} = \sqrt{800 + 900} = \sqrt{1700} = 41.23$$

The desired probability is given by $P_r\,(\bar{x}_1 - \bar{x}_2) \geq 600$ and is shown as shaded region below :



To convert this into standard normal variate, we get

$$P\left[\frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}\right] \geq \left[\frac{600 - (\mu_1 - \mu_2)}{\sigma_{\bar{x}_1 - \bar{x}_2}}\right]$$

or

$$P\left[Z \geq \frac{600 - 500}{41.23}\right] = P\,[Z \geq 2.43]$$

From the standard normal table, the corresponding value for Z = 2.43 is 0.4925. Hence, the required probability is given by

$$P\,(\bar{x}_1 - \bar{x}_2) \geq 600 = 0.5 - 0.4925 = 0.0075$$

Therefore, the probability that the sample mean strength of the wire produced by company A is greater than or equal to 600 kg than that of company B is given by 0.0075.

Sampling Distribution of the Number of Successes

If a random sample of size n is taken from a population whose elements belong to two mutually exclusive categories—one containing elements which possess a certain trait and the other containing elements which do not possess the trait—then the sampling distribution of the number of successes is the binomial distribution if sampling is made with replacement; and it is the hypergeometric distribution if sampling is made without replacement.

The sampling distribution of the number of successes being a binomial probability model will have its mean $\mu = np$ and standard error denoted by $\sigma = \sqrt{npq}$.

Illustration 3. If a coin is tossed 20 times and the coin falls on head after any toss, it is a success. Suppose the probability of success is 0.5. What is the probability that the number of successes is less than or equal to 12 ?

Solution : Given $\mu = np = 20 \times 0.5 = 10$, $\sigma = \sqrt{npq} = \sqrt{20 \times 0.5 \times 0.5} = \sqrt{5} = 2.24$

$$Z = \frac{X - \mu}{\sigma} = \frac{X - np}{\sqrt{npq}} = \frac{12 - 10}{2.24} = 0.89$$

From the table, the corresponding value of $Z = 0.89$ is 0.8133. Hence, the probability that the number of successes is less than or equal to 12 is 0.8133.

Note : Since we are dealing with a proportion, the binomial distribution tends to normal distribution provided n is large enough to make both np and nq at least 5.

Sampling Distribution of Proportions

A population proportion is defined as $\pi = X/N$, where X is the number of elements which possess a certain trait and N is the total number of items in the population. A sample proportion is defined as $p = x/n$, where x is the number of items in the sample which possess a certain trait and n is the sample size. A proportion may be considered as a proportion of successes and is obtained by dividing the number of successes by sample size n. If a random sample of n is obtained with replacement, then the sampling distribution of p follows the binomial probability law.

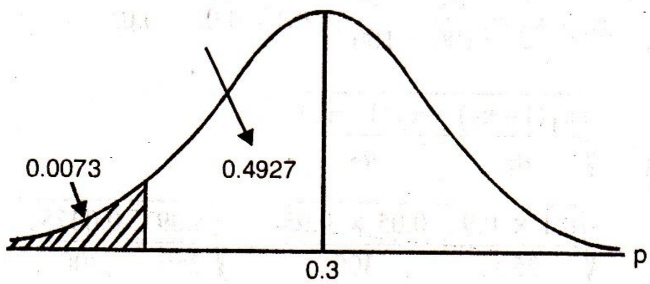
Suppose that a population is infinite and that the probability of occurrence of an event (called its success) is π while the probability of non-occurrence of the event is $(1 - \pi)$. Now, consider all possible samples of size n drawn from this population and for each sample determine the proportion 'p' of successes. Then we obtain a *sampling distribution of sample proportion* whose mean is μ_p, and standard deviation σ_p are given by :

$$\mu_p = \pi \text{ ; and } \sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}}$$

where σ_p = standard error of sample proportion. It measures chance variations of sample proportions from sample to sample. For large values of n ($n \geq 30$) the sampling distribution is very closely approximated as normally distributed.

Illustration 4. In a quality department of manufacturing paints, at the time of despatch of decorative paints, 30% of the containers are found to be defective. If a random sample of 500 is drawn with replacement from the population, what is the probability that the sample proportion will be less than or equal to 25% defective ?

Solution. We have
$$\mu_p = \pi = 0.3$$
and
$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.3 \times 0.7}{500}} = 0.0205$$



$$P[p \leq 0.25] = \left[Z \leq \frac{0.3 - 0.25}{0.0205}\right]; Z \leq 2.44$$

From the table, the corresponding value of Z is 0.4927, and therefore, the required probability is 0.0073 that sample proportion will be less than or equal to 0.25.

Illustration 5. In the year 2001, a policy is introduced to give loan to unemployed engineers to start their own business. Out of 1,00,000 unemployed engineers, 60,000 accept the policy and got the loan. A sample of 100 unemployed engineers is taken at the time of allotment of loan. What is the probability that sample proportion would have exceeded 50% acceptance ?

Solution. Here,
$$\mu_p = \pi = 0.60$$

$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} \sqrt{\frac{N-n}{N-1}}$$

$$= \sqrt{\frac{0.6 \times 0.4}{100}} \sqrt{\frac{1{,}00{,}000 - 100}{1{,}00{,}000 - 1}} = \sqrt{0.0024}\ \sqrt{0.999} = 0.0489$$

$$z = \frac{p - \pi}{\sigma_p} = \frac{0.50 - 0.60}{0.0489} = -2.04$$

The value of z can be found from the normal table and the corresponding value is 0.9793.

Therefore, the probability that sample proportion would have exceeded 50% acceptance is 0.9793. If all possible samples of 100 unemployed engineers are taken from the population of 1,00,000 then in 97.93% of these samples, the proportion of engineers who are in favour, is greater than 0.50.

Sampling Distribution of the Difference of Two Proportions

Earlier, in this chapter we referred to the sampling distribution of the difference of two means. Corresponding results can be obtained for the sampling distributions of difference of two proportions from two binomially distributed populations with parameters π_1 and π_2 respectively, when two random samples are drawn from two binomial populations and then compared. Unless both samples are of the same size, we cannot work with the number of successes, one must work only with the proportion of successes. Consider an example, a sample of 100 salesmen is taken from a chemical company, 50 are found to be in favour of new advertising policy. Another example of a textile company shows that 60 out of 150 are found to be in favour of policy. These two cases cannot be calculated unless they are reduced to proportions. The mean and standard deviation of this sampling distribution is given below :

$$\mu_{p_1 - p_2} = \mu_{p_1} - \mu_{p_2} = \pi_1 - \pi_2$$

and $$\sigma_{p_1 - p_2} = \sqrt{\sigma^2_{p_1} + \sigma^2_{p_2}} = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}$$

If n_1 and n_2 are large (n_1, $n_2 \geq 30$), the sampling distributions of difference of two proportions are very closely normally distributed.

Illustration 6. Ten per cent of machines produced by company *A* are defective and five per cent of those produced by company *B* are defective. *A* random sample of 250 machines is taken from company *A* and a random sample of 300 machines from company *B*. What is the probability that the difference in sample proportion is less than or equal to 0.02 ?

(MBA, South Gujrat Univ.; MBA, Delhi Univ., 1999)

Solution. Under the assumption, the sampling distribution of $p_1 - p_2$ would have mean

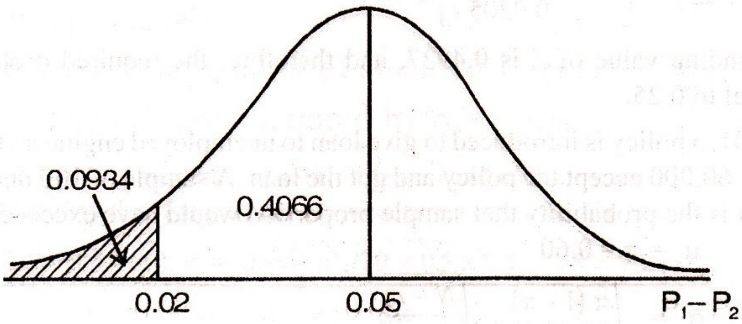
$$\mu_{p_1 - p_2} = \pi_1 - \pi_2 = \frac{10}{100} - \frac{5}{100} = 0.1 - 0.05 = 0.05$$

and standard deviation

$$\sigma_{p_1 - p_2} = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}$$

$$= \sqrt{\frac{0.1 \times 0.9}{250} + \frac{0.05 \times 0.95}{300}} = \sqrt{\frac{0.09}{250} + \frac{0.0475}{300}}$$

$$= \sqrt{0.00036 + 0.00016} = \sqrt{0.00052} = 0.0228$$



The probability that the difference in sample proportion is less than or equal to 0.02 is given by

$$P(p_1 - p_2) \leq 0.02$$

Hence, the required probability is obtained by transforming into a standard normal variate as

$$P\left[\frac{(p_1-p_2)-(\pi_1-\pi_2)}{\sqrt{\frac{\pi_1(1-\pi_1)}{n_1}+\frac{\pi_2(1-\pi_2)}{n_2}}} \le \frac{0.02-0.05}{0.0228}\right] \quad \text{or } P\,[z \le -1.32]$$

From the table, its corresponding value of z is 0.4066 and, therefore, the required probability is 0.0934 that the difference in sample proportion is less than or equal to 0.02.

We have discussed above few important types of sampling distributions. Just as we have discussed the sampling distributions of means, proportions, etc. It is also possible to discuss about the sampling distributions of first and third quartiles, quartile deviation, coefficient of skewness, coefficient of correlation, etc. However, they have not been discussed and only the formulae for standard error (S. E.) are given :

$$\text{S.E. of quartiles or } \sigma_{Q_1} = \sigma_{Q_3} = \frac{1.3632\sigma}{\sqrt{n}}$$

$$\text{S.E. of Q.D. of } \sigma_{QD} = \frac{0.7867\sigma}{\sqrt{n}}$$

$$\text{S.E. of coefficient of skewness } \sigma_{sk.} = \sqrt{\frac{3}{2n}}$$

$$\text{S.E. of coefficient of correlation } \sigma_r = \frac{1-r^2}{\sqrt{n}}$$

Sampling distributions occupy a place of great prominence in statistical theory. A sampling distribution shows that a statistic of a random sample may take on any set of values; but these values do not have the same probability of occurrence. Sampling distribution constitute the basis of testing hypothesis and enable us to evaluate the validity of statistical inferences.

MISCELLANEOUS ILLUSTRATIONS

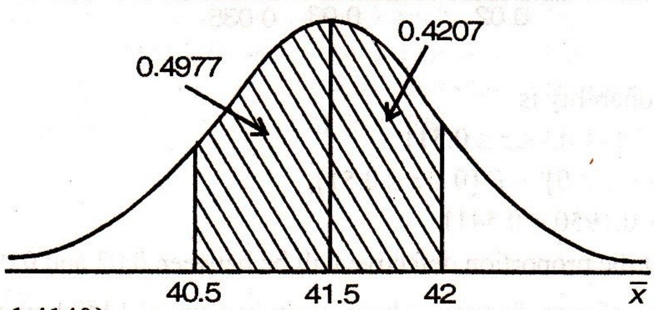
Illustration 7. The mean length of life of a certain cutting tool is 41.5 hours with a standard deviation of 2.5 hours. What is the probability that a simple random sample of size 50 drawn from this population will have a mean between 40.5 hours and 42 hours ? *(MBA, DU, 2005)*

Solution. Given, $\mu = 41.5, \sigma = 2.5, n = 50$

$$\sigma_{\bar{x}} = \frac{2.5}{\sqrt{50}} = \frac{2.5}{7.0711} = 0.3536$$

Therefore, the required probability is given by

$$P\{40.5 \le \bar{x} \le 42\} = P\left\{\frac{40.5-41.5}{0.3536} \le z \le \frac{42-41.5}{0.3536}\right\}$$



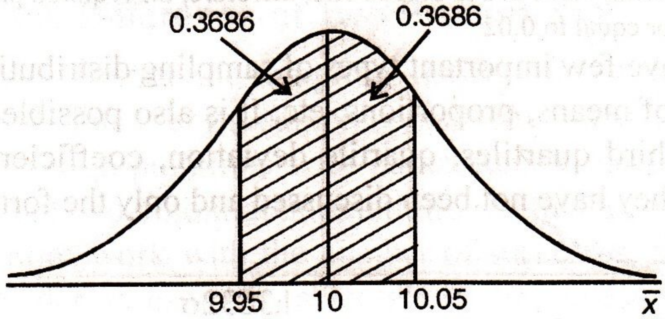
$$= P\{-2.8281 \le z \le 1.4140\}$$
$$= P\{-2.8281 \le z \le 0\} + P\{0 \le z \le 1.4140\}$$
$$= 0.4977 + 0.4207 = 0.9184.$$

Therefore, the probability is 0.9184 for the cutting tool to have a mean life between 40.5 hours and 42 hours.

Illustration 8. A diameter of a component produced on a semi-automatic machine is known to be distributed normally with a mean of 10 mm and a standard deviation of 0.1 mm. If we pick up a random sample of size 5, what is the probability that the sample mean will be between 9.95 mm and 10.05 mm ?

Solution. Given : $\mu = 10$, $\sigma = 0.1$ and $n = 5$

$$z_1 = \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} = \frac{10.05-10}{0.1/\sqrt{5}} = \frac{0.05}{0.0447} = 1.12$$



$$z_2 = \frac{9.95-10}{0.1/\sqrt{5}} = -\frac{0.05}{0.0447} - 1.12$$

The area corresponding to Z = 1.12 is 0.3686. Hence, the required probability is 0.3686 + 0.3686 = 0.7372.

Illustration 9. A manufacturer of watches has determined from experience that 3% of the watches he produces are defective. If a random sample of 300 watches is examined, what is the probability that the proportion defective is between 0.02 and 0.035 ?

(MBA, Delhi Univ., 2000)

Solution. Here $\pi = 0.03$

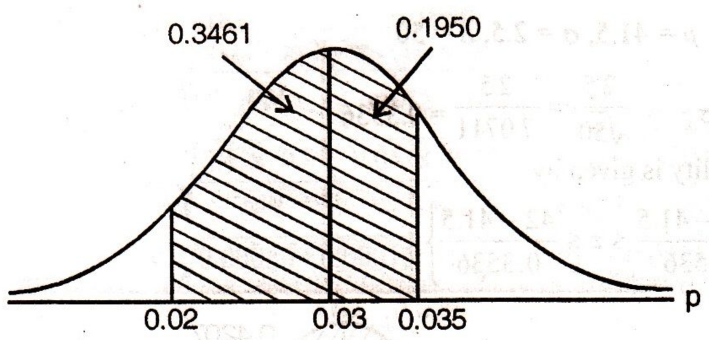
$$\sigma_p = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.03 \times 0.97}{300}} = \sqrt{\frac{0.0291}{300}} = \sqrt{0.000097} = 0.0098$$

When $p = 0.02$, then

$$z_1 = \frac{0.02-0.03}{0.0098} = \frac{-0.01}{0.0098} = 1.02$$

When $p = 0.035$, then

$$z_1 = \frac{0.035-0.03}{0.0098} = \frac{0.005}{0.0098} = 0.51$$



Therefore, the required probability is

$$P\{-1.02 \le z \le 0.51\}$$
$$= P\{-1.02 \le z \le 0\} + P\{0 \le z \le 0.51\}$$
$$= 0.3461 + 0.1950 = 0.5411$$

Hence, the probability that the proportion defective will be between 0.02 and 0.035 is given by 0.5411.

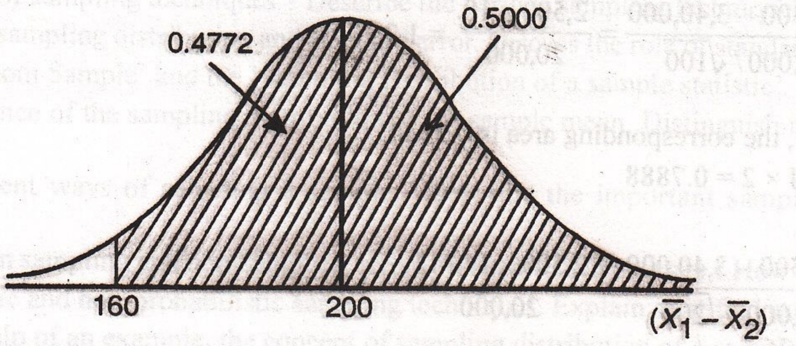
Illustration 10. Car Stereo of manufacturer *A* have mean lifetime of 1400 hours with a standard deviation of 200 hours, while those of manufacturer *B* have a mean lifetime of 1200 hours with a standard deviation of 100 hours. If random sample of 125 stereos of each manufacturer are tested, what is the probability that the manufacturer *A* stereos will have a mean lifetime which is at least (*i*) 160 hours more than the, manufacturer *B* stereos and (*ii*) 250 hours more than the manufacturer *B* stereos ?

(MBA, Delhi Univ., 1999)

Solution. Let $\bar{x}_1$ and $\bar{x}_2$ denote the mean lifetime of samples *A* and *B* respectively. Then

$$\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2 = 1400 - 1200 = 200$$

and
$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{(200)^2}{125} + \frac{(100)^2}{125}} = \sqrt{320+80} = \sqrt{400} = 20$$



(i) The required probability is given by

$= P(\bar{x}_1 - \bar{x}_2) \geq 160$

To convert this into a standard normal variate, we get

$$P\left[\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sigma_{\bar{x}_1-\bar{x}_2}} \geq \frac{160-200}{20}\right] = P\,[z \geq -2]$$

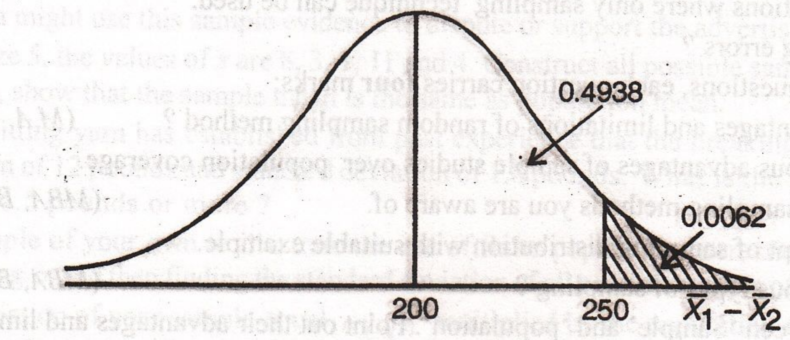
Required probability = area under normal curve to right of $z = -2$

$= 0.5000 + 0.4772 = 0.9772$

(ii) The required probability is given by

$P\,[\,\bar{x}_1 - \bar{x}_2\,) \geq 250]$

$$P\left[\frac{(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)}{\sigma_{\bar{x}_1-\bar{x}_2}} \geq \frac{250-200}{20}\right] = P\,[z \geq +2.5]$$



Required probability = area under normal curve to right of $z = 2.5$.

$= 0.5000 - 0.4938 = 0.0062.$

Illustration 11. The average annual starting salary for MBA (Marketing majors) is Rs. 3,40,000. Assume that for the population of MBA (Marketing majors), the average annual starting salary is $\mu = 3{,}40{,}000$ and the standard deviation is $\sigma = 20{,}000$. What is the probability that a simple random sample of MBA (Marketing majors) will have a sample mean within ± Rs. 2,500 of the population mean for each of the sample sizes : 50, 100 and 200 ? What is your conclusion ? (*MBA, Delhi Univ., 2003*)

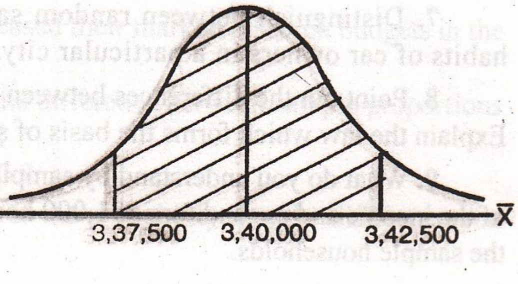
Solution : Given $\mu = 3{,}40{,}000, \sigma = 20{,}000, n_1 = 50, n_2 = 100, n_3 = 200$

For $n_1 = 50$

$$Z_1 = \frac{3{,}42{,}500 - 3{,}40{,}000}{20{,}000/\sqrt{50}}$$

$$= \frac{2{,}500 \times 7.071067}{20{,}000}$$

$$= \frac{17{,}677.67}{20{,}000} = 0.88$$



$$Z_2 = \frac{3,37,500 - 3,40,000}{20,000/\sqrt{50}} = -0.88$$

For $Z = \pm 0.88$, the corresponding area is 0.3106.

Required prob. $= 0.3106 \times 2 = 0.6212$

For $n_2 = 100$

$$Z_1 = \frac{3,42,500 - 3,40,000}{20,000/\sqrt{100}} = \frac{2,500 \times 10}{20,000} = 1.25$$

$$Z_2 = -1.25$$

For $Z = \pm 1.25$, the corresponding area is 0.3944.

Required prob. $= 0.3944 \times 2 = 0.7888$

For $n_3 = 200$

$$Z = \frac{3,42,500 - 3,40,000}{20,000/\sqrt{200}} = \frac{2,500 \times 14.142}{20,000} = 1.7676$$

$$Z = -1.7676$$

For $Z = \pm 1.7676$ corresponding area is 0.4616.

Required prob. $= 0.4616 \times 2 = 0.9232$.

PROBLEMS

1-A : Answer the following questions, each question carries **one** mark:

(*i*) What are the advantages of sampling ?
(*ii*) Define 'stratified random sampling'.
(*iii*) Explain probability sampling.
(*iv*) What is systematic random sampling ? (*M.A. Eco., M.K. Univ., 2003*)
(*v*) What is random sampling ?
(*vi*) How a stratified sample is selected ?
(*vii*) What is multi-stage sampling ?
(*viii*) What are the limitations of sampling ?
(*ix*) Mention few situations where only sampling technique can be used.
(*x*) What are sampling errors ?

1-B : Answer the following questions, each question carries **four** marks:

(*i*) What are the advantages and limitations of random sampling method ? (*M.A. Eco., Madras Univ., 2003*)
(*ii*) What are the various advantages of sample studies over population coverage.
(*iii*) Explain any four sampling methods you are aware of. (*MBA, Bharathidasan Univ., 2003*)
(*iv*) Explain the concept of sampling distribution with suitable example.
(*v*) What are the various types of sampling ? (*MBA, Bharathidasan Univ., 2001*)
(*vi*) Differentiate between 'Sample' and 'population'. Point out their advantages and limitations.

2. Point out the importance of sampling in solving business problems. What are the principles on which sampling theory rests ?

3. "Sampling is necessary under certain conditions". Explain this with illustrative examples.

4. Describe the various methods of sampling and the requisites of a good sample.

5. What is sampling ? Explain the importance of sampling in solving business problems. Critically examine the well known methods of probability sampling and non-probability sampling.

6. Define judgment sampling, quota sampling, and convenience sampling. Under what conditions, can each of these designs be used to advantage ?

7. Distinguish between random sampling and stratified sampling. Suppose it is desired to survey petrol buying habits of car owners in a particular city. How would you proceed about it ?

8. Point out the differences between a sample survey and a census survey. Under what conditions, are these undertaken ? Explain the law which forms the basis of sampling.

9. What do you understand by sampling ? In order to determine a new cost of living index, it is proposed to make a survey of the income and expenditure of 1,000 households in a large city. Describe carefully two methods which might be used to select the sample households.

10. Suppose you are asked to conduct a survey on the smoking habits of the Delhi University teachers. How will you proceed ?
11. "In any sample survey there are many sources of errors. A perfect survey is a myth." Discuss the statement.
12. "Data collected in census are automatically free of errors." Discuss the validity of the statement.
13. Enumerate the various methods of sampling and describe two of them mentioning the situations where each one is to be used.
14. What is the importance of sampling techniques ? Describe the various sampling techniques.
15. Explain the concepts of sampling distribution and standard error. Discuss the role of standard errors in large sample theory.
16. Explain the terms 'Random Sample' and the 'Sampling Distribution of a sample statistic'.
17. Find the mean and variance of the sampling distribution of the sample mean. Distinguish between standard deviation and standard error.
18. "There are many different ways of selecting a sample." Describe the important sampling methods pointing out the characteristics of each.
19. (*a*) Distinguish between sampling and non-sampling errors. What are their sources ? How these errors can be controlled ?
    (*b*) List the probabilistic and non-probabilistic sampling techniques. Explain stratified random sampling technique.
    (*c*) Explain with the help of an example, the concept of sampling distribution of a sample statistic and point out its role in managerial decision-making. *(MBA, Delhi Univ., 2003)*
20. The weight of certain type of a car tyre is normally distributed with a mean of 25 pounds and variance of 3 pounds. A random sample of 50 tyres is selected. What is the probability that the mean of this sample lies between 24.5 and 25.5 pounds ?
    [0.9586]
21. For a particular brand of T.V. picture tube, it is known that the mean operating life of the tubes is 1,000 hours with a standard deviation of 250 hours. What is the probability that the mean for a random sample of size 25 will be (*i*) greater than 1,000 hours, (*ii*) less than 1,000 hours, (*iii*) between 950 and 1,050 hours ? *(MBA, DU, 2005, 2006)*
    [0.5, 0.5, 0.6826]
22. An auditor takes a sample of size 36 from a population of 1,000 accounts receivable. The standard deviation of the population is unknown, but the standard deviation of the sample is Rs. 100. If the true mean value of the accounts receivable from the population is Rs. 3000, what is the probability that the sample mean will be less than or equal to Rs. 2800 ?
23. A manufacturer of razor blades claims that his product will, on the average, give 15 good shaves. Suppose you have five friends who try using one of these razor blades each. The number of shaves reported by your friends are 12, 16, 8, 14 and 10.
    (*a*) Find the mean and standard deviation of this sample.
    (*b*) Suggest how you might use this sample evidence to dispute or support the advertiser's claim.
24. For a population of size 5, the values of x are 8, 3, 1, 11 and 4. Construct all possible sample of size two and calculate their sample means. Hence, show that the sample mean is the same as population mean.
25. A manufacturer of knitting yarn has established from past experience that the breaking strength of this yarn is normally distributed with a mean of 12 pounds and standard deviation of 1.8 pounds. What is the probability that a sample size of 49 yield a mean of 14.5 pounds or more ?
26. Design a simple example of your own to illustrate the use of finite population correction factor by listing your values of some population, finding σ, and then finding the standard deviation of all possible sample of size 3 drawn without replacement. Does the standard deviation of your sample equal $\sigma/\sqrt{n}$ multiplied by the population correction factor ?
27. Two methods of performing a certain task in a manufacturing plant, method A and method B, are under study. The variable of interest is length of time required to perform the task. It is known that the variance of method A is 9 minutes squared and variance of method B is 12 minutes squared. A simple random sample of 35 employees performed the task by method A and independent simple random sample of 35 employees performed the task by method B. The average time required by the first group to complete the task was 25 minutes and the average time for the second group was 23 minutes. What is the probability of observing difference this large, if there is no difference in the true average length of time required to perform the task ?
28. An accountant has determined from prior experience that 60 per cent of his client's customers respond to initial requests for confirmation of their account balances. If a simple random sample of 64 customers is sent requests for confirmation, what is the probability that 50 per cent or more will respond ?
29. A research group stated that 16 per cent of the firms of a particular type, A increased their marked research budgets in the five years preceding the study. For type B firms the figure was 9 per cent.
    (*a*) What are the mean and standard deviation of the sampling distribution of the difference between sample proportions based on independent simple random samples of 100 firms from each type ?
    (*b*) What proportion of the sample differences would be between 0.05 and 0.10 ?
    (*c*) If you took a simple random sample of size 100 from each industry, what is the probability that the difference you would observe would be equal to or less than 0.02 ?

30. Suppose it is known that 5 per cent of forms processed by a clerical pool contain at least one error. If a simple random sample of 475 forms is examined, what is the probability that the proportion containing at least one error will be between 0.03 and 0.075 ?

31. A manufacturer of pens has determined from experience that 4 per cent of the pens he produces are defective. If a random sample of 400 pens is examined, what is the probability that the proportion defective is between 0.025 and 0.048 ?

32. Marks obtained by a number of students are assumed to be normally distributed with mean 65 and variance 25. If 3 students are taken at random, what is the probability that exactly two of them will have marks over 70 ?

33. A firm produces light bulbs that are known to have a mean life time of 1,200 hours with a standard deviation of 210 hours. What is the probability that a simple random sample of 100 bulbs will yield a mean that falls between 1,140 and 1,260 hours ?
[0.9956]

*****