11 SAMPLING TECHNIQUES AND SAMPLING DISTRIBUTIONS

11.1 POPULATION (OR UNIVERSE)

A *population* is the totality of the observations made on all the objects (under investigation) possessing some common specific characteristics, which are of particular interest to researchers.

The population is the aggregate of the elements and these elements are the *basic units* that comprise and define a population. The population must be defined in terms of

(*i*) content, (*ii*) unit, (*iii*) extent, (*iv*) time

For instance, the students of first year class at a given college, the characteristic to be investigated may be the score received by each student in a college entrance examination, in a given year. Populations may be finite or infinite.

11.1.1 Finite Population. A population is said to be *finite* if it includes a limited number of elementary units (objects or observations).

Examples of a finite population are: the heights of all the students enrolled at a college in a given year, the wages of all employees of a steel mill in a given year, the amount of money spent by each student in an engineering university in a given academic year, or the grading of items as defective and non-defective that are produced by an industry on a given day.

11.1.2 Infinite Population. A population is said to be *infinite* if it consists of unlimited number of elementary units. At least hypothetically, there is no limit to the number of units it can include.

Examples of an infinite population are: the weights at birth of all human beings, the results obtained by rolling of a die, the lifetimes of all the bulbs produced in a production process that operates indefinitely under given manufacturing conditions.

11.2 SAMPLE

A *sample* is a part of the population which is selected with the expectation that it will represent the characteristics of the population.

11.2.1 Sampling. *Sampling* is a procedure of selecting a representative sample from a given population.

11.2.2 Sample Survey versus Complete Enumeration. The collection of information from a part of the population is called making a *sample survey.* The collection of information from all elements in a population is called taking a *census* or making a *complete enumeration.*

11.2.3 Purposes of Sampling. The two basic purposes of sampling are:

(*i*) To obtain maximum information about the characteristics of the population with minimum cost, time and effort.

(*ii*) To find the reliability of the estimates derived from the sample.

11.2.4 Advantages of Sampling. Following are the main advantages of sampling over a complete census.

(*i*) ***Time Saving*:** A sample survey involves lesser amount of time and energy than a complete enumeration both in the execution and the analysis of data. This is a vital consideration when the information is urgently needed as the results from a sample survey are more readily available.

(*ii*) ***Economic*:** A sample survey requires less expenses and labour as compared to a complete census because the cost of covering only a fraction will be lower than that of covering the whole population.

(*iii*) ***Accuracy*:** A sample survey provides the results which are almost as accurate as those obtained by complete census. A properly designed and carefully executed sample survey will provide even better results.

(*iv*) ***Feasibility*:** Sometimes the data are obtained by tests that are destructive. For example, to know the average life of certain type of electric bulb, we shall take a sample of these bulbs and keep them on until they burn out. We cannot think of testing the whole lot. In testing blood of a patient we do not drain the entire blood out of him but examine just a few drops. Sampling may be the only means available for obtaining the desired information when the population is infinite or inaccessible. In such cases complete enumeration would neither be physically possible nor practically feasible.

Whatever be the merits of sampling, it cannot totally replace a complete census. A census is a record of a nation's history and its importance has to be given due acknowledgement.

11.2.5 Limitations of Sampling. If the basic facts of each and every unit in the population are needed, census become indispensable. The sample will not meet such a requirement.

For example, the list of income tax payers is prepared very carefully, the list of voters is prepared to include the name of each and every voter, or an inventory of all goods and stocks is necessary to know the total amount of stocks of a firm.

11.3 SAMPLING DESIGN

A *sampling design* is a procedure or plan for obtaining a sample from a given population prior to collecting any data.

The collection of detailed information is known as *survey*. When a survey is carried out by a sampling design, it is called a *sample survey*. A sample survey should be properly planned and carefully executed in order to avoid inaccuracies.

11.3.1 Sampling Units. *Sampling units* are those basic units of the population in terms of which the sample design is planned.

The sampling units must be distinct and exhaustive, *i. e.*, they must make up the whole population and they must not be overlapping. Sometimes the sampling unit is obvious, as in a population of students or in a population of light bulbs. Sometimes there is a choice of sampling unit. In sampling an agricultural crop, the sampling unit might be a field, a farm or an area of land whose shape and area is at our disposal. In sampling a human population, the unit might be an individual person, the household or all the persons living in a block.

11.3.2 Sampling Frame. A *sampling frame* is a complete list of the sampling units.

For example, a complete list of all the students in a college on May 10, 1995, is the frame. A complete list of all households in a city is an other example of the frame.

11.3.3 Types of Sampling Designs. Meaningfulness of estimates, obtained from a sample, depends upon the methods of selecting a sample. Broadly speaking there are two different sampling schemes.

(*i*) Non-probability sampling

(*ii*) Probability sampling

11.4 NON-PROBABILITY (NON-RANDOM) SAMPLING

A *non-probability sampling* is a procedure in which we cannot assign to an element of population the probability of its being included in the sample.

We often make inferences about the population from arbitrary and informal samples. A wheat dealer forms his opinion about a sackful of wheat by examining just a few grains. To say about the quality of rice cooked in a big pot, the cook takes only a spoonful of rice to taste and decide on its quality of cooking. Such arbitrary selections are frequently made in research, in biological and physical sciences.

11.5 PROBABILITY (RANDOM) SAMPLING

A *probability sampling* is a process in which the sample is selected in such a way that every element of a population has a known non-zero (not necessarily equal) probability of being included in the sample.

The advantage of probability sampling is that it provides a measure of precision of the estimates. The underlying principle of a random sample is that personal factor is eliminated in the selection as the investigator does not exercise his discretion in the choice of items. No factor other than chance affects the likelihood of an item being included in or excluded from the sample. A random sample may be taken with or without replacement.

11.5.1 Random Sampling With Replacement. Sampling is said to be *with replacement* from a population (finite or infinite) when the unit selected at random is returned to the population before the next unit is selected. The formal description of the sampling method is as follows.

1. An object is selected from the population in a way that gives all objects in the population an equal chance of being selected.
2. The characteristics level of the object selected is observed, and the object is returned to the population prior to any subsequent selection.
3. For a sample of size n, steps (1) and (2) are performed n times

Thus the number of units available for future drawing is not affected. The population remains the same and a sampling unit might be selected more than once.

11.5.2 Random Sampling Without Replacement. Sampling is said to be *without replacement*, when the sampling unit selected at random is not returned to the population, before the next unit is selected. The formal description of the sampling method is as follows.

1. The first object is selected from the population in a way that gives all objects in the population an equal chance of being selected.

2. The characteristics level of the object selected is observed, but the object is not returned to the population.
3. An object is selected from the remaining objects in the population in a way that gives all the remaining objects an equal chance of being selected, and step (2) is repeated. For a sample of size n, step (3) is performed $(n - 1)$ times.

Thus the number of units remaining after each drawing will be reduced by one. In this case, a sampling unit selected once cannot be selected again for the sample because the selected unit is not replaced.

11.6 SIMPLE RANDOM SAMPLING

Simple random sampling is a procedure of selecting a sample of n units $(n = 1, 2, \cdots, N)$ from the population of N units in such a way that:

(*i*) Every unit available for sampling has an equal probability of being drawn.
(*ii*) Every sample of size n has the same probability of being selected.

A sample drawn by this procedure is called a *simple* or *unrestricted random sample.* A sample containing n elements selected from a population consisting of N elements is called a sample of size n. The simple random sampling is used in the population which is essentially homogeneous in terms of some characteristics relevant to the enquiry. For small populations where the elements are easily identifiable and accessible, simple random sampling may be easy to apply.

***Theorem* 11.1** *If a simple random sample of size n is selected from a finite population of size N, then the number of all possible samples is given as*

$$\text{No. of possible samples} = N^n, \quad \text{if sampling is done with replacement}$$

$$\text{No. of possible samples} = {}^N P_n, \quad \text{if sampling is done without replacement}$$

Proof. A sample of size n under simple random sampling (with or without replacement) consists of an ordered specification of n elements, namely

(the first chosen, the second chosen, $\cdots$, the n-th chosen.)

Sampling With Replacement. If we use sampling with replacement, the number of units available for each drawing are N.

The first unit of the sample can be selected in N different ways, the second unit of the sample can also be selected in N ways, the third unit of the sample can also be selected in N ways, and so on, the n-th unit of the sample can also be selected in N *ways.*

Using the multiplicative principle, the number of all possible samples of size n, that could be selected from a finite population of size N, is

$$\text{No. of possible samples} = N \times N \times N \times \cdots n \text{ times}$$
$$= N^n$$

A sample of n units constitutes only one arrangement, and there are N^n possible arrangements of n units from a finite population of N units. Each of the N^n possible samples is selected with the same probability

$$\frac{1}{N} \times \frac{1}{N} \times \frac{1}{N} \times \cdots n \text{ times} = \frac{1}{N \times N \times N \times \cdots n \text{ times}} = \frac{1}{N^n}$$

Sampling Without Replacement. If we use sampling without replacement, the number of units remaining after each drawing will be reduced by one

The first unit of the sample can be selected in N different ways, the second unit of the sample can be selected in $(N - 1)$ ways, the third unit of the sample can be selected in $(N - 2)$ ways, and so on, the n-th unit of the sample can be selected in $(N - n + 1)$ ways.

Using the multiplicative principle, the number of all possible samples of size n, that could be selected from a finite population of size N, is

$$\begin{aligned}\text{No. of possible samples} &= N(N-1)(N-2)\cdots\{N-(n-1)\}\\ &= N(N-1)(N-2)\cdots(N-n+1)\\ &= \frac{N(N-1)\cdots(N-n+1)(N-n)\cdots(3)(2)(1)}{(N-n)\cdots(3)(2)(1)}\\ &= \frac{N!}{(N-n)!} = {}^{N}P_n\end{aligned}$$

A sample of n units constitutes only one arrangement, and there are ${}^{N}\text{P}_n$ possible arrangements of n units from a finite population of N units. Each of the ${}^{N}\text{P}_n$ possible samples is selected with the same probability

$$\frac{1}{N} \times \frac{1}{N-1} \times \cdots \times \frac{1}{N-n+1} = \frac{1}{N(N-1)\cdots(N-n+1)} = \frac{1}{{}^{N}\text{P}_n}$$

11.6.1 Random Digits. A table of *random digits* consists of a sequence of digits designed to represent the result of a simple random sampling with replacement from a population of digits 0, 1, 2, $\cdots$, 9. In a table of random digits each digit from 0 to 9 is called a random digit, each having the probability of occurrence of $1/10$. *Here random implies that all of these digits have the same probability of occurrence and the occurrence and non-occurrence of any digit is independent of the occurrence and non-occurrence of all other digits.* Table 15 is such a table.

In a random digits table, random digits are normally combined to form numbers of more than one digit. For example, random digits taken in pairs will result in a set of 100 different numbers from 00 to 99, each having a probability of occurrence of $1/100$ and each being independent of other numbers similarly formed. Likewise random digits taken in triples will result in a set of 1000 different numbers from 000 to 999, each having a probability of occurrence of $1/1000$ and each being independent of other numbers similarly formed. Similarly, random digits taken in quadruples will result in 10000 different numbers from 0000 to 9999, each having a probability of occurrence of $1/10000$, and each being independent of other numbers similarly formed.

11.6.2 Selection of Simple Random Sample. A simple random sample can be selected by the following methods.

(*i*) ***Lottery Method.*** In this method, a distinct and different serial number from 1 to N is assigned to every unit of the population of N units and the number is recorded on a card or a slip of paper. All the numbered slips are then placed in a container, and they are thoroughly mixed. A blind selection is made of the number of slips required to constitute the desired size of the sample. The items corresponding to the slips drawn will constitute the random sample. The selection of items depends entirely on chance. Some lotteries use a rotating wheel in selecting tickets. The wheel has equal segments on its rim, one for each of the digits 0 through 9. N lottery tickets are numbered from 1 to N. Suppose the tickets have three-digit numbers. A ticket number then would be selected by spinning the wheel thrice and recording the digit which appears at the pointer each time the wheel stops. If the digit sequence is 534, then the ticket number 534 is selected. The lottery method becomes quite cumbersome to use as the size of population becomes large, then an alternative method of selection of a random sample is employed.

(*ii*) ***Using Random Digits***. In this method a distinct sampling number from 0 to $(N-1)$ is assigned to every unit of the population of N units. A table of random digits is consulted with a randomly selected starting point in the table. The table is read in single digits, in groups of two, three or more according to the number of digits in the sampling number $(N-1)$ assigned to the last unit in the population. Any number greater than $(N-1)$ is discarded. A number appearing second time is also discarded if the sampling is without replacement. Continue the process of selecting the random digits or numbers until the desired sample size is reached.

11.7 STRATIFIED SAMPLING

If the elements in the population are not homogenous, then the population is divided into non-overlapping homogeneous subgroups, called *strata*, and sample is drawn separately from each stratum by simple random sampling. This sample is called *stratified random sample*. The process of dividing a heterogeneous population into homogeneous subgroups is called *stratification*.

The benefit of this method is that if non-overlapping homogeneous subgroups of the population can be identified, then only a relatively small number of observations are needed to ascertain the characteristics of each subgroup. Stratification is used also to improve sample estimates of population characteristics. Stratification is used:

(*i*) to provide an adequate sample for each stratum,

(*ii*) because it can give more precise estimates of population characteristics than other types of samples.

11.8 ERRORS

11.8.1 True Value. By *true value* we mean the value that would be obtained if no errors were made in any way in obtaining the information or computing the characteristic of the population.

True value of the population is possibly obtained only if the exact procedures are used for collecting the correct data, each and every element of the population has been covered and no mistake or even the slightest negligence has happened during the process of data collection and its analysis. It is usually regarded as an unknown constant.

11.8.2 Accuracy. By *accuracy* we refer to the difference between the sample result and the true value. The smaller the difference, the greater will be the accuracy. Accuracy can be increased:

(*i*) By elimination of technical errors.

(*ii*) By increasing the sample size.

11.8.3 Precision. By *precision* we refer to how closely we can reproduce, from a sample, the results which would be obtained if a complete count (census) was taken using the same method of measurement.

11.8.4 Error. The difference between an estimated value and the population true value is called an *error*. Since a sample estimate is used to describe a characteristic of a population. A sample being only a part of a population cannot provide a perfect representation of the population, no matter how carefully the sample is selected. We may think as to how close will the sample estimate be to the population true value. Generally it is seen that an estimate is rarely equal to the true value. There are two kinds of errors:

(*i*) Sampling (random) errors

(*ii*) Non-sampling (non-random) errors

11.8.5 Sampling Error. A *sampling error* is the difference between the value of a statistic obtained from an observed random sample and the value of corresponding population parameter being estimated.

A sample may not provide a true representation of the population under study, simply because samples represent only a part of a population and thus depend on " the luck of the draw ", even if the sample survey is properly designed and well-implemented. Generally, let T be sample statistic used to estimate the population parameter θ, then the sampling error, denoted by E, is defined as

$$E = T - \theta$$

The value of sampling error reveals the *precision* of the estimate. Smaller the sampling error, the greater will be the precision of the estimate. The sampling errors can be reduced:

(*i*) By increasing the sample size.

(*ii*) By improving the sampling design.

(*iii*) By using the supplementary information.

11.8.6 Non-sampling Errors. The errors that are caused by sampling the wrong population of interest and by response bias, as well as those made by an investigator in collecting analysing and reporting the data, are all classified as *non-sampling* or *non-random errors*. These errors are present in a complete census as well as in a sample survey.

11.8.7 Bias. *Bias* is the difference between the expected value of a statistic and the true value of the parameter being estimated. Let T be the sample statistic used to estimate the parameter θ, then the amount of bias is

$$Bias = E(T) - \theta$$

The bias is positive if $E(T) > \theta$, it is negative if $E(T) < \theta$ and it is zero if $E(T) = \theta$. Bias is a systematic component of error which refers to the long-run tendency of the sample statistic to differ from the parameter in a particular direction. Bias is cumulative and increases with the increase in size of the sample. If proper methods of selection of units in a sample are not followed, the sample results will not be free from bias.

Exercise 11.1

1. (*a*) Explain the terms: Population; Sample; Sampling frame; Sampling unit.

(*b*) Define suitable populations from which the following samples are selected:

(*i*) One thousand homes are called by telephone in the city of Karachi and asked to name the T.V programme that they are now watching

(*ii*) A coin is flipped 53 times and 32 heads are recorded.

(*iii*) Two hundred pairs of a new type of combat boots were tested for durability in Vietnam and, on the average, lasted two months.

{ (*i*) Homes in Karachi city having telephones and T.V.,

(*ii*) An infinite number of tosses of a coin,

(*iii*) Total production of a new type of combat boots during a particular period. }

(c) In each of the following situations, determine whether the sampling is done from a finite population or an infinite population and then define the population.

(i) A coin is tossed 20 times and 12 heads are recorded.

(ii) Ten employees of large manufacturing company are selected as representatives of labour to serve as a labour management committee.

(iii) A sample of bulbs is selected periodically to determine the number of defective bulbs produced by a production unit.

(iv) A coin is weighed 15 times to estimate its true weight.

{ (i) Infinite—population is all the potential tosses of the coin,
(ii) Finite—population is all the employees of the company,
(iii) Infinite—population is all the bulbs produced by the production unit,
(iv) Infinite—population is all the potential weights of the coin.}

2. (a) What is meant by sampling? Describe the advantages of sampling over complete enumeration.

(b) For each of the following reasons, give an example of a situation for which a census would be less desirable than a sample. In each case, explain why this is so,

(i)	Economy	(ii)	Timeliness
(iii)	Size of population	(iv)	Inaccessibility
(v)	Accuracy	(vi)	Destructive observations?

(c) Distinguish between the following:

(i) Population and sample,

(ii) Sampling with and without replacement.

3. (a) Distinguish between probability and non-probability sampling, giving examples. Describe the advantages of using a probability sample.

(b) What do you understand by a simple random sample? By taking some artificial example, explain the method of drawing a simple random sample.

(c) Distinguish between the following:

(i) Random sampling and simple random sampling.

(ii) Simple random sampling and stratified random sampling.

(iii) Sampling and non-sampling errors.

4. (a) Explain how would you select a random sample of 10 households from a list of 250 households, by using a table of random digits.

(b) A poll is to be conducted to determine the voting preference of the voters in a certain city. Design a sampling plan such that the sample would be representative of the population of all the voters.

(c) In a certain locality, there are 300 households. We wish to select a sample of 50 households. How would you select this sample using Random Numbers?

5. (a) What is the difference between precision and accuracy of a result? Explain with some examples.

(b) What are two broad categories of errors in data collected by sample surveys? What are the methods for reducing sampling error?

◆

11.9 SIMPLE RANDOM SAMPLING AND SAMPLING DISTRIBUTIONS

As we have already mentioned that a random sample must be chosen in such a way that it is representative of the population about which we want to make inferences. A random sample of observations can be chosen in either of the two ways: with replacement or without replacement.

11.10 SAMPLING DISTRIBUTION OF A STATISTIC

11.10.1 Parameters. The numerical quantities, that describe probability distributions, are called *parameters*. Parameters are fixed constants that characterise a population.

Parameters are usually denoted by Greek letters. Thus, π (the probability of success in a binomial experiment), and μ and σ (the mean and standard deviation of a normal distribution) are examples of parameters.

Let $x_1, x_2, \cdots, x_N$ be the N elements of a population. A population value summarizes the values of some characteristic (or characteristics) for all N units of an entire population. It describes some feature of the distribution of the random variable (or variables) in the defined population. Let x_j, $j = 1, 2, \cdots, N$, be the observed value of some random variable X for the j-th element in the population, then some of the examples of the population parameters are:

Population total: $$\tau = \sum_{j=1}^{N} \Sigma x_j$$

Population mean: $$\mu = \frac{\sum_{j=1}^{N} x_j}{N}$$

Population variance: $$\sigma^2 = \frac{\sum_{j=1}^{N} (x_j - \mu)^2}{N}$$

Population proportion: $$\pi = \frac{\textit{No. of elements with attribute A}}{\textit{Population size}} = \frac{k}{N}$$

11.10.2 Statistic. A *statistic* is a function, of the observations of a random sample, which does not contain any unknown parameter.

We know that a number of simple random samples can be drawn from the same population and each sample gives a different value of the statistic that is used as an estimator of the population parameter. The sample statistic is a random variable having its own probability distribution. We intend to use a statistic to make inferences about the distribution of the population. A statistic is usually denoted by a small Latin letter ($\bar{x}$, s, r) to represent its value obtained from an actually observed sample. A statistic is denoted by a capital Latin letter ($\bar{X}$, S, R) to represent its random nature.

Let $x_1, x_2, \cdots, x_n$ be the observed values of a random sample $X_1, X_2, \cdots, X_n$ of size n from a given population of N items. A sample value is an estimate calculated from the n

elements in the sample. Let x_i, $i = 1, 2, \cdots, n$ be the i-th element in the sample, then the observed values of some of the sample statistics are

Sample total: $\sum_{i=1}^{n} x_i$

Sample mean: $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$

Sample variance: $s^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}, \quad \hat{s}^2 = \dfrac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$

where $n s^2 = (n-1)\hat{s}^2$

Sample proportion: $p = \dfrac{\textit{No. of elements with attribute } A}{\textit{Sample size}} = \dfrac{x}{n}$

11.10.3 Sampling Distribution of a Statistic. The *sampling distribution of statistic* is the probability distribution of the statistic obtained from all possible samples of some specified size that can be drawn from a given population.

11.10.4 Standard Error of a Statistic. The standard deviation of the sampling distribution of a statistic is called the *standard error of the statistic*.

11.11 SAMPLING DISTRIBUTIONS FROM GENERAL POPULATIONS

We will now look at the most common situations where the Central Limit Theorem is used to specify approximate probability distributions for sample statistics where sampling is done from general (non-normal) populations.

The particular sampling distributions we are interested in are those for: (*i*) the mean, (*ii*) the difference between two means, (*iii*) the proportion of successes, (*iv*) the difference between two proportions.

11.12 SAMPLING DISTRIBUTION OF THE SAMPLE MEAN, $\overline{X}$

The *sampling distribution of the sample mean* $\overline{X}$ is the probability distribution of the means of all possible simple random samples of n observations that can be drawn from a given population with mean μ and variance σ^2.

11.12.1 Standard Error of $\overline{X}$. The standard deviation of the sampling distribution of the sample mean $\overline{X}$, denoted by $\sigma_{\overline{X}}$, is called the *standard error of* $\overline{X}$.

To discuss the relationships between the population and the sampling distribution of the sample mean, the following symbols will be used.

N = Population size	n = Sample size
μ = Population mean	$\mu_{\overline{X}}$ = Mean of the distribution of $\overline{X}$
σ^2 = Population variance	$\sigma^2_{\overline{X}}$ = Variance of the distribution of $\overline{X}$
σ = Population standard deviation	$\sigma_{\overline{X}}$ = Standard error of the distribution of $\overline{X}$

11.12.2 Properties of the Sampling Distribution of $\overline{X}$. The properties of the sampling distribution of the sample mean are given by the following theorems:

Theorem 11.2 *The mean of the sampling distribution of $\overline{X}$, denoted by $\mu_{\overline{X}}$, is equal to the mean of the sampled population, i. e.,*

$$\mu_{\overline{X}} = E(\overline{X}) = \mu$$

This theorem holds regardless of the sample size n or whether sampling is conducted with or without replacement.

Theorem 11.3 *The variance of the sampling distribution of $\overline{X}$ is equal to the variance of the sampled population divided by the sample size, i. e.,*

$$\sigma^2_{\overline{X}} = Var(\overline{X}) = \frac{\sigma^2}{n}$$

where $\overline{X}$ is the mean of a random sample of size n from an infinite population (or sampling with replacement) with mean μ and finite variance σ^2.

The standard error of $\overline{X}$ then becomes

$$\sigma_{\overline{X}} = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}}$$

However, if the value of σ is unknown, it is replaced by the sample standard deviation $\hat{S}$, the estimate of the standard error of $\overline{X}$ then becomes

$$S_{\overline{X}} = \hat{\sigma}_{\overline{X}} = \frac{\hat{S}}{\sqrt{n}}$$

where, $\hat{S} = \sqrt{\dfrac{\Sigma(X_i - \overline{X})^2}{n-1}}$

Theorem 11.4 *The variance of the sampling distribution of $\overline{X}$ is*

$$\sigma^2_{\overline{X}} = Var(\overline{X}) = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

where $\overline{X}$ is the mean of a random sample of size n drawn ***without replacement*** *from a finite population of size N with mean μ and variance σ^2. The factor $(N-n)/(N-1)$ is usually called as finite population correction ($f\,p\,c$) for variance.*

The standard error of $\overline{X}$ then becomes

$$\sigma_{\overline{X}} = \sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

However, if the value of σ is unknown, it is replaced by the sample standard deviation $\hat{S}$, the estimate of the standard error of $\overline{X}$ then becomes

$$S_{\overline{X}} = \hat{\sigma}_{\overline{X}} = \frac{\hat{S}}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

where, $\hat{S} = \sqrt{\dfrac{\Sigma(X_i - \overline{X})^2}{n-1}}$

***Theorem* 11.5** *If $\overline{X}$ is the mean of the random sample of size n drawn from a normal population with mean μ and variance σ^2 (known), the sampling distribution of $\overline{X}$ is a normal distribution with mean μ and variance σ^2/n regardless of the size of the sample (including sample size 1). The distribution of the standardized sampling errors*

$$Z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

will be standard normal distribution.

***Theorem* 11.6 (*Central Limit Theorem*).** *For a large sample size, the mean $\overline{X}$ of a random sample from a population with mean μ and finite variance σ^2 has a sampling distribution that is approximately normal with mean μ and variance σ^2/n regardless of the probability distribution (shape) of the sampled population. The lager the sample size, the better will be the normal approximation to the sampling distribution of $\overline{X}$. The distribution of the standardized sampling errors*

$$Z = \frac{\overline{X} - \mu_{\overline{X}}}{\sigma_{\overline{X}}} = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}$$

will approach the standard normal distribution as n tends to infinity.

***Example* 11.1** *A population consists of four children with ages 2, 4, 6 and 8. Take all possible simple random samples of size 2 with replacement. If X is the age of a child, find,*

(*i*) *the theoretical sampling distribution of $\overline{X}$, the mean age of two children in a sample;*

(*ii*) *the mean, variance and standard error of $\overline{X}$;*

(*iii*) *the mean, variance and standard deviation of the population.*

Verify the results

(*i*) $\mu_{\overline{X}} = \mu$ (*ii*) $\sigma^2_{\overline{X}} = \dfrac{\sigma^2}{n}$ (*iii*) $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$

Solution. Population: 2, 4, 6, 8; Population size: $N = 4$; Sample size: $n = 2$

Number of possible samples $= N \times N = 4 \times 4 = 16$

All possible samples that can be drawn with replacement from our population, and their means are shown in the following tree diagram.

First draw	Second draw	Sample values x_i	Sample sum $\sum x_i$	$\bar{x} = \frac{\sum x_i}{n}$
2	2	2, 2	4	2
	4	2, 4	6	3
	6	2, 6	8	4
	8	2, 8	10	5
4	2	4, 2	6	3
	4	4, 4	8	4
	6	4, 6	10	5
	8	4, 8	12	6
6	2	6, 2	8	4
	4	6, 4	10	5
	6	6, 6	12	6
	8	6, 8	14	7
8	2	8, 2	10	5
	4	8, 4	12	6
	6	8, 6	14	7
	8	8, 8	16	8

Fig. 11.1 A tree diagram showing all possible samples of size 2 drawn with replacement from a population of the 4 equiprobable values 2, 4, 6, 8

The sampling distribution of sample mean $\overline{X}$, its mean, variance and standard error are

Value of $\overline{X}$ $\bar{x}$	Number of occurrences f	Probability $p(\bar{x}) = f/\sum f$	$\bar{x}\, p(\bar{x})$	$\bar{x}^2\, p(\bar{x})$
2	1	1/16	2/16	4/16
3	2	2/16	6/16	18/16
4	3	3/16	12/16	48/16
5	4	4/16	20/16	100/16
6	3	3/16	18/16	108/16
7	2	2/16	14/16	98/16
8	1	1/16	8/16	64/16
Sums	$\sum f = 16$	1	80/16	440/16

$$\mu_{\overline{X}} = E(\overline{X}) = \sum \bar{x}\, p(\bar{x}) = \frac{80}{16} = 5$$

$$\sigma^2_{\overline{X}} = Var(\overline{X}) = \sum \bar{x}^2\, p(\bar{x}) - \mu^2_{\overline{X}} = \frac{440}{16} - (5)^2 = 2.5$$

$$\sigma_{\bar{X}} = \sqrt{Var(\bar{X})} = \sqrt{2.5} = 1.58$$

The mean, variance and standard deviation of the population

x_j	2	4	6	8	$\sum x_j = 20$
x_j^2	4	16	36	64	$\sum x_j^2 = 120$

$$\mu = \frac{\sum x_j}{N} = \frac{20}{4} = 5$$

$$\sigma^2 = \frac{\sum x_j^2}{N} - \mu^2 = \frac{120}{4} - (5)^2 = 5$$

$$\sigma = \sqrt{5} = 2.236$$

We are to verify that

(*i*) $\mu_{\bar{X}} = \mu$ (*ii*) $\sigma_{\bar{X}}^2 = \frac{\sigma^2}{n}$ (*iii*) $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$

$5 = 5$ $2.5 = \frac{5}{2}$ $1.58 = \frac{2.236}{\sqrt{2}}$

$2.5 = 2.5$ $1.58 = 1.58$

***Example* 11.2** *A population consists of values 3, 6, and 9. Take all possible simple random samples of size 3 with replacement. Form the sampling distribution of sample mean $\bar{X}$. Hence state and verify the relationship between*

(*i*) *the mean of $\bar{X}$ and the population mean,*

(*ii*) *the variance of $\bar{X}$ and the population variance,*

(*iii*) *the standard error of $\bar{X}$ and the population standard deviation.*

Solution. Population: 3, 6, 9; Population size: $N = 3$; Sample size: $n = 3$

Number of possible samples $= N \times N \times N = 3 \times 3 \times 3 = 27$

All possible samples that can be drawn with replacement from our population, and the sample means are shown in the following tree diagram.

First draw	Second draw	Third draw	Sample values x_i	$\sum x_i$	$\bar{x} = \frac{\sum x_i}{n}$
		3	3, 3, 3	9	3
	3	6	3, 3, 6	12	4
		9	3, 3, 9	15	5
		3	3, 6, 3	12	4
3	6	6	3, 6, 6	15	5
		9	3, 6, 9	18	6
		3	3, 9, 3	15	5
	9	6	3, 9, 6	18	6
		9	3, 9, 9	21	7

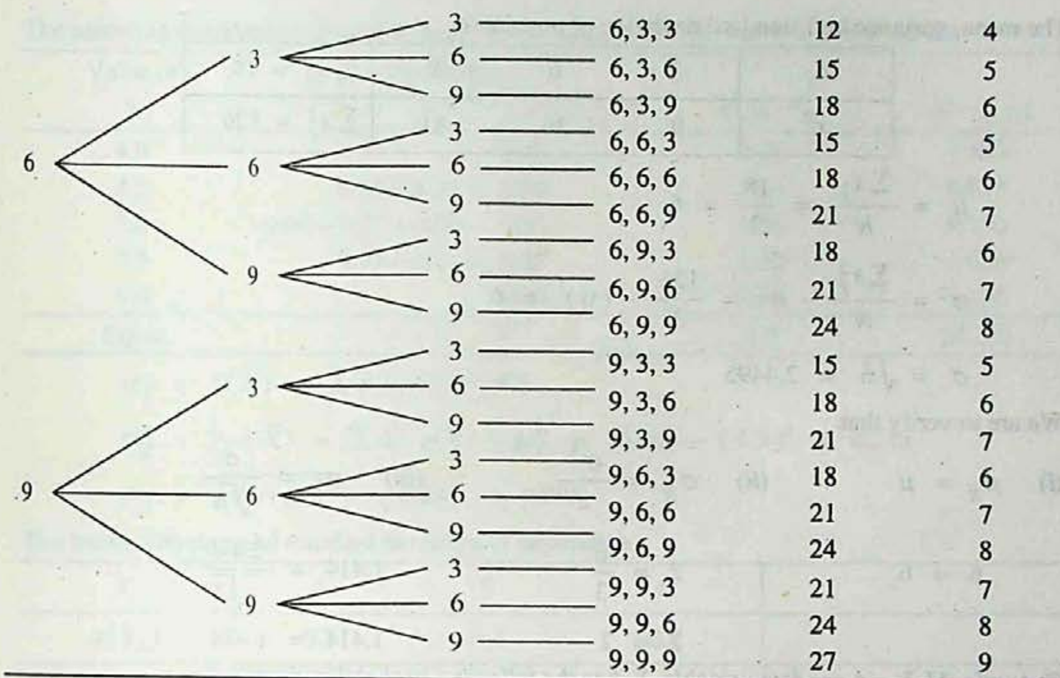


Fig. 11.2 A tree diagram showing all possible samples of size 3 drawn with replacement from a population of 3 equiprobable values 3, 6, 9.

The sampling distribution of sample mean $\overline{X}$, its mean, variance and standard error are

Value of $\overline{x}$	Number of occurrences f	Probability $p(\overline{x}) = f/\sum f$	$\overline{x}\ p(\overline{x})$	$\overline{x}^2\ p(\overline{x})$
3	1	1/27	3/27	9/27
4	3	3/27	12/27	48/27
5	6	6/27	30/27	150/27
6	7	7/27	42/27	252/27
7	6	6/27	42/27	294/27
8	3	3/27	24/27	192/27
9	1	1/27	9/27	81/27
Sum	$\sum f = 27$	1	162/27	1026/27

$$\mu_{\overline{X}} = E(\overline{X}) = \sum \overline{x}\ p(\overline{x}) = \frac{162}{27} = 6$$

$$\sigma_{\overline{X}}^2 = Var(\overline{X}) = \sum \overline{x}^2\ p(\overline{x}) - \mu_{\overline{X}}^2 = \frac{1026}{27} - (6)^2 = 2$$

$$\sigma_{\overline{X}} = \sqrt{Var(\overline{X})} = \sqrt{2} = 1.414$$

The mean, variance and standard deviation of population

x_j	3	6	9	$\sum x_j = 18$
x_j^2	9	36	81	$\sum x_j^2 = 126$

$$\mu = \frac{\sum x_j}{N} = \frac{18}{3} = 6$$

$$\sigma^2 = \frac{\sum x_j^2}{N} - \mu^2 = \frac{126}{3} - (6)^2 = 6$$

$$\sigma = \sqrt{6} = 2.4495$$

We are to verify that

(*i*) $\mu_{\overline{X}} = \mu$ (*ii*) $\sigma_{\overline{X}}^2 = \dfrac{\sigma^2}{n}$ (*iii*) $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$

$6 = 6$ $2 = \dfrac{6}{3}$ $1.414 = \dfrac{2.4495}{\sqrt{3}}$

$2 = 2$ $1.414 = 1.414$

***Example* 11.3** *A random variable X has the following probability distribution.*

x_j	4	5	6
$p(x_j)$	0.3	0.5	0.2

If a sample of size 2 is taken with replacement, obtain the sampling distribution of $\overline{X}$. Determine the mean and variance of the sampling distribution. Find the mean and variance of the population. Discuss the results.

Solution. We have an infinite population. Since the sample is drawn at random with replacement from the *infinite population*, the sample values are independent. Thus the distribution of possible samples of size $n = 2$ drawn with replacement is

Sample values x_i	Sample total $\sum x_i$	Sample mean $\bar{x} = \frac{\sum x_i}{n}$	Probability $p(\bar{x})$
4, 4	8	4.0	(0.3)(0.3) = 0.09
4, 5	9	4.5	(0.3)(0.5) = 0.15
4, 6	10	5.0	(0.3)(0.2) = 0.06
5, 4	9	4.5	(0.5)(0.3) = 0.15
5, 5	10	5.0	(0.5)(0.5) = 0.25
5, 6	11	5.5	(0.5)(0.2) = 0.10
6, 4	10	5.0	(0.2)(0.3) = 0.06
6, 5	11	5.5	(0.2)(0.5) = 0.10
6, 6	12	6.0	(0.2)(0.2) = 0.04
Sum			1

The sampling distribution of sample mean $\overline{X}$, its mean, variance and standard error are

Value of $\bar{x}$	Probability $p(\bar{x})$	$\bar{x}\ p(\bar{x})$	$\bar{x}^2\ p(\bar{x})$
4.0	0.09	0.36	1.440
4.5	0.15 + 0.15 = 0.30	1.35	6.075
5.0	0.06 + 0.25 + 0.06 = 0.37	1.85	9.250
5.5	0.10 + 0.10 = 0.20	1.10	6.050
6.0	0.04	0.24	1.440
Sums	1	4.9	24.255

$$\mu_{\overline{X}} = E(\overline{X}) = \sum \bar{x}\ p(\bar{x}) = 4.9$$

$$\sigma^2_{\overline{X}} = Var(\overline{X}) = \sum \bar{x}^2\ p(\bar{x}) - \mu^2_{\overline{X}} = 24.255 - (4.9)^2 = 0.245$$

$$\sigma_{\overline{X}} = \sqrt{Var(\overline{X})} = \sqrt{0.245} = 0.495$$

The mean, variance and standard deviation of population

x_j	4	5	6	
$p(x_j)$	0.3	0.5	0.2	
$x_j\ p(x_j)$	1.2	2.5	1.2	$\sum x_j\ p(x_j) = 4.9$
$x_j^2\ p(x_j)$	4.8	12.5	7.2	$\sum x_j^2\ p(x_j) = 24.5$

$$\mu = E(X) = \sum x_j\ p(x_j) = 4.9$$

$$\sigma^2 = Var(X) = \sum x_j^2\ p(x_j) - \mu^2 = 24.5 - (4.9)^2 = 0.49$$

$$\sigma = \sqrt{0.49} = 0.7$$

We are to verify that:

(*i*) $\mu_{\overline{X}} = \mu$ (*ii*) $\sigma^2_{\overline{X}} = \dfrac{\sigma^2}{n}$ (*iii*) $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}$

$4.9 = 4.9$ $0.245 = \dfrac{0.49}{2} = 0.245$ $0.495 = \dfrac{0.7}{\sqrt{2}} = 0.495$

***Example* 11.4** *A population consists of value 3, 5, 7 and 9. Take all possible simple random samples of size 2 without replacement. Form the sampling distribution of sample mean $\overline{X}$. Find the mean, variance and standard error of $\overline{X}$. Find the mean, variance and standard deviation of the population. Verify that:*

(*i*) $\mu_{\overline{X}} = \mu$ (*ii*) $\sigma^2_{\overline{X}} = \dfrac{\sigma^2}{n}\left(\dfrac{N-n}{N-1}\right)$ (*iii*) $\sigma_{\overline{X}} = \dfrac{\sigma}{\sqrt{n}}\sqrt{\dfrac{N-n}{N-1}}$

Solution. Population: 3, 5, 7, 9; Population size: $N = 4$; Sample size: $n = 2$

Number of possible samples $= N(N-1) = 4(4-1) = 12$

All possible samples that can be drawn without replacement from our population, and their means are shown in the following tree diagram.

First draw	Second draw	Sample values x_i	$\sum x_i$	$\bar{x} = \frac{\sum x_i}{n}$
3	5	3, 5	8	4
	7	3, 7	10	5
	9	3, 9	12	6
5	3	5, 3	8	4
	7	5, 7	12	6
	9	5, 9	14	7
7	3	7, 3	10	5
	5	7, 5	12	6
	9	7, 9	16	8
9	3	9, 3	12	6
	5	9, 5	14	7
	7	9, 7	16	8

Fig. 11.3 A tree diagram showing all possible samples of size 2 drawn without replacement from a population of 4 equiprobable values 3, 5, 7, 9.

The sampling distribution of sample mean $\overline{X}$, its mean, variance and standard error are

Value of $\overline{X}$ $\bar{x}$	Number of occurrences f	Probability $p(\bar{x}) = f/\sum f$	$\bar{x}\ p(\bar{x})$	$\bar{x}^2\ p(\bar{x})$
4	2	2/12	8/12	32/12
5	2	2/12	10/12	50/12
6	4	4/12	24/12	144/12
7	2	2/12	14/12	98/12
8	2	2/12	16/12	128/12
Sum	$\sum f = 12$	1	72/12	452/12

$$\mu_{\overline{X}} = E(\overline{X}) = \sum \bar{x}\ p(\bar{x}) = \frac{72}{12} = 6$$

$$\sigma^2_{\overline{X}} = Var(\overline{X}) = \sum \bar{x}^2\ p(\bar{x}) - \mu^2_{\overline{X}} = \frac{452}{12} - (6)^2 = 1.667$$

$$\sigma_{\overline{X}} = \sqrt{Var(\overline{X})} = \sqrt{1.667} = 1.291$$

The mean, variance and standard deviation of population

x_j	3	5	7	9	$\sum x_j = 24$
x_j^2	9	25	49	81	$\sum x_j^2 = 164$

$$\mu = \frac{\sum x_j}{N} = \frac{24}{4} = 6$$

$$\sigma^2 = \frac{\sum x_j^2}{N} - \mu^2 = \frac{164}{4} - (6)^2 = 5$$

$$\sigma = \sqrt{5} = 2.236$$

We are to verify that

(*i*) $\mu_{\overline{X}} = \mu$ (*ii*) $\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$ (*iii*) $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

$6 = 6$ $1.667 = \frac{5}{2}\left(\frac{4-2}{4-1}\right)$ $1.291 = \frac{2.236}{\sqrt{2}}\sqrt{\frac{4-2}{4-1}}$

$1.667 = 1.667$ $1.291 = 1.291$

***Example* 11.5** *A population consists of values* 0, 3, 6 *and* 9. *Take all possible simple random samples of size* 3 *without replacement. Form the sampling distribution of sample mean* $\overline{X}$. *Hence state and verify the relationship between*

(*i*) *the mean of* $\overline{X}$ *and the population mean,*

(*ii*) *the variance of* $\overline{X}$ *and the population variance,*

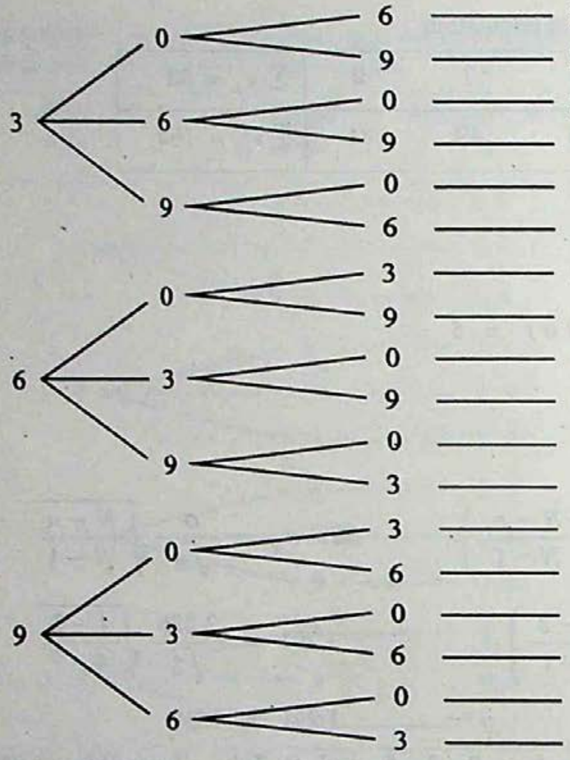
(*iii*) *the standard error of* $\overline{X}$ *and the population standard deviation.*

Solution. Population: 0, 3, 6, 9; Population size: $N = 4$; Sample size: $n = 3$

Number of possible samples $= N(N-1)(N-2) = 4(4-1)(4-2) = 24$

All possible samples that can be drawn without replacement from our population, and the sample means are shown in the following tree diagram.

First draw	Second draw	Third draw	Sample values x_i	$\sum x_i$	$\bar{x} = \frac{\sum x_i}{n}$
0	3	6	0, 3, 6	9	3
		9	0, 3, 9	12	4
	6	3	0, 6, 3	9	3
		9	0, 6, 9	15	5
	9	3	0, 9, 3	12	4
		6	0, 9, 6	15	5



3, 0, 6	9	3
3, 0, 9	12	4
3, 6, 0	9	3
3, 6, 9	18	6
3, 9, 0	12	4
3, 9, 6	18	6
6, 0, 3	9	3
6, 0, 9	15	5
6, 3, 0	9	3
6, 3, 9	18	6
6, 9, 0	15	5
6, 9, 3	18	6
9, 0, 3	12	4
9, 0, 6	15	5
9, 3, 0	12	4
9, 3, 6	18	6
9, 6, 0	15	5
9, 6, 3	18	6

Fig. 11.4 A tree diagram showing all possible samples of size 3 drawn without replacement from a population of 4 equiprobable values 0, 3, 6, 9.

The sampling distribution of sample mean $\overline{X}$, its mean, variance and standard error are

Value of $\overline{X}$ $\bar{x}$	Number of occurrences f	Probability $p(\bar{x}) = f/\Sigma f$	$\bar{x}\ p(\bar{x})$	$\bar{x}^2\ p(\bar{x})$
3	6	6/24	18/24	54/24
4	6	6/24	24/24	96/24
5	6	6/24	30/24	150/24
6	6	6/24	36/24	216/24
Sums	$\Sigma f = 24$	1	108/24	516/24

$$\mu_{\overline{X}} = E(\overline{X}) = \Sigma \bar{x}\ p(\bar{x}) = \frac{108}{24} = 4.5$$

$$\sigma^2_{\overline{X}} = Var(\overline{X}) = \Sigma \bar{x}^2\ p(\bar{x}) - \mu^2_{\overline{X}} = \frac{516}{24} - (4.5)^2 = 1.25$$

$$\sigma_{\overline{X}} = \sqrt{Var(\overline{X})} = \sqrt{1.25} = 1.118$$

The mean, variance and standard deviation of population

x_j	0	3	6	9	$\sum x_j = 18$
x_j^2	0	9	36	81	$\sum x_j^2 = 126$

$$\mu = \frac{\sum x_j}{N} = \frac{18}{4} = 4.5$$

$$\sigma^2 = \frac{\sum x_j^2}{N} - \mu^2 = \frac{126}{4} - (4.5)^2 = 11.25$$

$$\sigma = \sqrt{11.25} = 3.3541$$

We are to verify that

(i) $\mu_{\overline{X}} = \mu$ (ii) $\sigma^2_{\overline{X}} = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$ (iii) $\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

$4.5 = 4.5$ $1.25 = \frac{11.25}{3}\left(\frac{4-3}{4-1}\right)$, $1.118 = \frac{3.3541}{\sqrt{3}}\sqrt{\frac{4-3}{4-1}}$

$1.25 = 1.25$ $1.118 = 1.118$

***Example* 11.6** *A random variable X has the following probability distribution.*

x_j	3	4	5
$p(x_j)$	0.2	0.4	0.4

If a simple random sample of 3 numbers is taken without replacement, obtain a sampling distribution of the sample mean $\overline{X}$. Find the mean, variance and standard error of $\overline{X}$.

Solution. We have an infinite population. The actual sampling distribution of $\overline{X}$, the sample mean of three numbers taken without replacement, is impracticable because the population is infinite. Since the sample is drawn at random without replacement from the infinite population, the sample values become independent. Then the actual sampling distribution of $\overline{X}$, the sample mean of three numbers taken without replacement, is impossible but it virtually becomes the sampling distribution of $\overline{X}$ the sample mean of three numbers taken with replacement.
The population size N is infinite and $n = 3$. Then the finite population correction

$$\frac{N-n}{N-1} \to 1 \text{ as } N \to \infty \quad \text{and} \quad \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} \to \frac{\sigma}{\sqrt{n}}$$

The mean, variance and standard deviation of population

x_j	3	4	5	Sum
$p(x_j)$	0.2	0.4	0.4	$\sum p(x_j) = 1$
$x_j\, p(x_j)$	0.6	1.6	2.0	$\sum x_j\, p(x_j) = 4.2$
$x_j^2\, p(x_j)$	1.8	6.4	10.0	$\sum x_j^2\, p(x_j) = 18.2$

$$\mu = E(X) = \sum x\,p(x) = 4.2$$

$$\sigma^2 = Var(X) = \sum x^2 p(x) - \mu^2 = 18.2 - (4.2)^2 = 0.56$$

The mean, variance and standard error of $\overline{X}$

$$\mu_{\overline{X}} = \mu = 4.2$$

$$\sigma^2_{\overline{X}} = \frac{\sigma^2}{n} = \frac{0.56}{3} = 0.187$$

$$\sigma_{\overline{X}} = \sqrt{0.187} = 0.432$$

Example 11.7 *The weights of 1000 students of a college are normally distributed with mean 68.5 kg and standard deviation 2.7 kg. If a simple random sample of 25 students is obtained from this population, find the expected mean and standard deviation of the sampling distribution of means if sampling were done (i) with replacement and (ii) without replacement.*

Solution. We have

Population mean: $\mu = 68.5$, Population standard deviation: $\sigma = 2.7$

Population size: $N = 1000$, Sample size: $n = 25$

(i) ***Sampling with replacement:***

$$\mu_{\overline{X}} = \mu = 68.5\ kg.$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{2.7}{\sqrt{25}} = 0.54\ kg$$

(ii) ***Sampling without replacement:***

$$\mu_{\overline{X}} = \mu = 68.5\ kg.$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{2.7}{\sqrt{25}}\sqrt{\frac{1000-25}{1000-1}} = 0.53\ kg$$

Example 11.8 *Given the population 1, 1, 1, 3, 4, 5, 6, 6, 6, and 7.*

(a) *Find the mean and standard deviation for the sampling distribution of mean for a sample of size 36 selected at random with replacement.*

(b) *Find the mean and standard deviation for the sampling distribution of mean for a sample of size 4 selected at random without replacement.*

Solution. The mean and standard deviation of the population are:

x_j	1	1	1	3	4	5	6	6	6	7	$\sum x_j = 40$
x_j^2	1	1	1	9	16	25	36	36	36	49	$\sum x_j^2 = 210$

$$\mu = \frac{\sum x_j}{N} = \frac{40}{10} = 4$$

$$\sigma = \sqrt{\frac{\sum x_j^2}{N} - \mu^2} = \sqrt{\frac{210}{10} - (4)^2} = 2.236$$

(a) ***Sampling With Replacement.*** We have, sample size $n = 36$. The mean and standard error of $\overline{X}$ are:

$$\mu_{\overline{X}} = \mu = 4$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{2.236}{\sqrt{36}} = 0.373$$

(b) ***Sampling Without Replacement.*** We have, sample size $n = 4$. The mean and standard error of $\overline{X}$ are:

$$\mu_{\overline{X}} = \mu = 4$$

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}} = \frac{2.236}{\sqrt{4}}\sqrt{\frac{10-4}{10-1}} = 0.913$$

Exercise 11.2

1. (a) How do you define a population and a sample? Differentiate between parameter and statistic. Why a parameter is said to be a constant and statistic a variable?

(b) A labour union has 1000 members. A random sample of 50 members of the union gave an average age of 40 years. The average age of the members of the labour union was, therefore, estimated to be 40 years. A complete enumeration of all the members indicated that the true mean age was 43 years. Answer the following:

(i) Which figure is a parameter?

(ii) Which figure is a statistic?

{ *(i)* Population size $N = 1000$, and population mean age $\mu = 43$ years;
(ii) Sample size n, and sample mean $\overline{x} = 40$ years.}

2. (a) What is meant by a sampling distribution and a standard error? Describe the properties of the sampling distribution of sample mean.

(b) What is meant by standard error and what are its practical uses?

(c) What is the finite population correction factor? When is it appropriately used in sampling applications and when can it, without too great an undesirable consequence, be ignored?

3. (a) A finite population consists of the numbers 2, 4, 6, 8, 10 and 12. Calculate the sample means for all possible random samples of size $n = 2$, that can be drawn from this population, with replacement. Assuming the 36 possible samples equally likely, make the sampling distribution of sample means and find the mean and variance of this distribution. Calculate mean and variance of the population and verify that

(i) $\mu_{\overline{X}} = \mu$ *(ii)* $\sigma^2_{\overline{X}} = \dfrac{\sigma^2}{n}$

{ $\mu = 7$, $\sigma^2 = 11.667$, $\mu_{\overline{X}} = 7$, $\sigma^2_{\overline{X}} = 5.833$ }

(b) A finite population consists of the numbers 2, 4, 6, 6, 8 and 10. Calculate the sample means for all possible random samples of size $n = 2$, that can be drawn from this population, with replacement. Assuming the 36 possible samples equally likely, form the sampling distribution of sample means and find the mean and variance of this distribution. Calculate mean and variance of the population and verify that

(*i*) $\mu_{\bar{X}} = \mu$ (*ii*) $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}}$

$\{ \mu = 6,\ \mu_{\bar{X}} = 6,\ \sigma = 2.582,\ \sigma_{\bar{X}} = 1.826 \}$

(c) Draw all possible samples of size $n = 3$ with replacement from the population 3, 6, 9 and 12. Assuming the 64 possible samples equally likely, form a sampling distribution of the sample means. Hence state and verify the relation between -

(*i*) the mean of the sampling distribution of the sample mean and the population mean;

(*ii*) the variance of the sampling distribution of the sample mean and the population variance.

$\{ \mu = 7.5,\ \mu_{\bar{X}} = 7.5,\ \mu_{\bar{X}} = \mu;\ \sigma^2 = 11.25,\ \sigma^2_{\bar{X}} = 3.75,\ \sigma^2_{\bar{X}} = \sigma^2/n \}$

4. (a) A finite population consists of the numbers 2, 4, 6, 6, 8 and 10. Calculate the sample means for all possible random samples of size $n = 2$, that can be drawn from this population, without replacement. Assuming the 30 possible samples equally likely, make the sampling distribution of sample mean. Find the mean and variance of this distribution. Calculate mean and variance of the population and verify that

(*i*) $\mu_{\bar{X}} = \mu$. (*ii*) $\sigma_{\bar{X}} = \dfrac{\sigma}{\sqrt{n}} \sqrt{\dfrac{N-n}{N-1}}$

$\{ \mu = 6,\ \mu_{\bar{X}} = 6,\ \sigma = 2.582,\ \sigma_{\bar{X}} = 1.633 \}$

(b) A finite population consists of the values 6, 6, 9, 15 and 18. Calculate the sample means for all possible random samples of size $n = 3$, that can be drawn from this population, without replacement. Assuming the 60 possible samples equally likely, make the sampling distribution of sample mean and find the mean and variance of this distribution. Calculate mean and variance of the population and show that

(*i*) $\mu_{\bar{X}} = \mu$ (*ii*) $\sigma^2_{\bar{X}} = \dfrac{\sigma^2}{n}\left(\dfrac{N-n}{N-1}\right)$

$(\mu = 10.8,\quad \sigma^2 = 23.76,\quad \mu_{\bar{X}} = 10.8,\quad \sigma^2_{\bar{X}} = 3.96)$

(c) Find the mean μ and variance σ^2 of the finite population 1, 4, 7 and 8. Take all possible samples of size 2, that can be drawn at random without replacement from this population. Assuming the 12 possible samples equally likely, make the sampling distribution of sample mean and find the mean and variance of this distribution. Verify that

(*i*) $E(\bar{X}) = \mu$ (*ii*) $Var(\bar{X}) = \dfrac{\sigma^2}{n}\left(\dfrac{N-n}{N-1}\right)$

where $\overline{X}$ is the random variable 'the sample mean', N is the population size and n is the sample size. What happens as $N \to \infty$?

$\{ \mu = 5,\ \sigma^2 = 7.5,\ E(\overline{X}) = 5,\ Var(\overline{X}) = 2.5,\ \dfrac{N-n}{N-1} \to 1 \}$

5. (a) In an infinite population $\mu = 50$ and $\sigma^2 = 250$, find the mean and variance for the distribution of $\overline{X}$ if :

(i) $n = 25$, *(ii)* $n = 100$, *(iii)* $n = 1250$

{ *(i)* $\mu_{\overline{X}} = 50,\ \sigma^2_{\overline{X}} = 10$ *(ii)* $\mu_{\overline{X}} = 50,\ \sigma^2_{\overline{X}} = 2.5$ *(iii)* $\mu_{\overline{X}} = 50,\ \sigma^2_{\overline{X}} = 0.2$ }

(b) A large number of samples of size 50 were selected at random from a normal population with mean μ and variance σ^2. The mean and standard error of the sampling distribution of the sample mean were obtained 2500 and 4 respectively. Find the mean and variance of the population.
(2500, 800)

6. (a) If the size of the simple random sample from an infinite population is 55, the variance of sample mean is 27, what must be the standard error of sample mean if $n = 165$?
$(\sigma_{\overline{X}} = 3)$

(b) If the size of the simple random sample from an infinite population is 36 and the standard error of the mean is 2, what must the size of the sample become if the standard error is to be reduced to 1.2?
$(n = 100)$

7. (a) The random variable X has the following probability distribution:

x_j	4	5	6	7
$p(x_j)$	0.2	0.4	0.3	0.1

Find the mean $\mu_{\overline{X}}$, variance $\sigma^2_{\overline{X}}$ and standard error $\sigma_{\overline{X}}$ of the mean $\overline{X}$ for a random sample of size 36.
$(\mu_{\overline{X}} = 5.3,\ \sigma^2_{\overline{X}} = 0.0225,\ \sigma_{\overline{X}} = 0.15)$

(b) A random sample of 36 cases is drawn from a negatively skewed probability distribution with a mean of 2 and a standard deviation of 3. Find the mean and standard error of the of the sampling distribution of $\overline{X}$.
$(\mu_{\overline{X}} = 2,\ \sigma_{\overline{X}} = 0.5)$

(c) A random sample of 100 is taken from a population with mean 30 and standard deviation 5. The probability distribution of the parent population is unknown, find the mean and standard error of the of the sampling distribution of $\overline{X}$.
$(\mu_{\overline{X}} = 30,\ \sigma_{\overline{X}} = 0.5)$

◆

11.13 SAMPLING DISTRIBUTION OF THE DIFFERENCE BETWEEN TWO SAMPLE MEANS, $\bar{X}_1 - \bar{X}_2$

The *sampling distribution of the difference between two sample means* $\bar{X}_1 - \bar{X}_2$ is the probability distribution of all possible differences between means $\bar{X}_1$ and $\bar{X}_2$ obtained from all possible independent simple random samples of n_1 and n_2 observations that can be drawn from two given populations with means μ_1, μ_2 and variances σ_1^2, σ_2^2 respectively.

Often we wish to compare the means of two random variables. The comparison is made on the basis of two independent random samples drawn from given populations.

Suppose that two independent random samples of sizes n_1 and n_2 are drawn from populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively. Let $\bar{X}_1$ be the mean of sample of size n_1 from the population with mean μ_1 and variance σ_1^2, then $\bar{X}_1$ is a random variable that has its own probability distribution with mean μ_1 and variance σ_1^2/n_1. Let $\bar{X}_2$ be the mean of sample of size n_2 from the population with mean μ_2 and variance σ_2^2, then $\bar{X}_2$ is a random variable that has its own probability distribution with mean μ_2 and variance σ_2^2/n_2.

Then the differences $\bar{X}_1 - \bar{X}_2$ can be obtained from all possible pairs of $\bar{X}_1$ and $\bar{X}_2$. Consequently, the difference $\bar{X}_1 - \bar{X}_2$ between two sample means is a random variable that has its own probability distribution which is called the sampling distribution of the difference between two sample means.

11.13.1 Properties of the Sampling Distribution of the Difference between Two Sample Means. The properties of the sampling distribution of the difference $\bar{X}_1 - \bar{X}_2$ between two sample means are given by the following theorems:

***Theorem* 11.7** *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.,*

$$\mu_{\bar{X}_1 - \bar{X}_2} = E(\bar{X}_1 - \bar{X}_2) = \mu_1 - \mu_2$$

This theorem holds regardless of the sample sizes n_1 and n_2 or whether sampling is done with or without replacement.

***Theorem* 11.8** *The variance of the sampling distribution of* $(\bar{X}_1 - \bar{X}_2)$, *denoted by* $\sigma^2_{\bar{X}_1 - \bar{X}_2}$, *is equal to sum of the variances of the sampled populations divided by the respective sample sizes, i. e.,*

$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = Var(\bar{X}_1 - \bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

where $\bar{X}_1$ *and* $\bar{X}_2$ *are means of two independent random samples of sizes* n_1 *and* n_2 *from infinite populations (or sampling with replacement) with means* μ_1 *and* μ_2 *and finite variances* σ_1^2 *and* σ_2^2 *respectively.*

The standard error of $(\overline{X}_1 - \overline{X}_2)$ then becomes

$$\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{Var(\overline{X}_1 - \overline{X}_2)} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

However, if σ_1^2 and σ_2^2 are unknown, these are replaced by the sample variances $\hat{S}_1^2$ and $\hat{S}_2^2$, the estimate of the standard error of $(\overline{X}_1 - \overline{X}_2)$ then becomes

$$S_{\overline{X}_1 - \overline{X}_2} = \hat{\sigma}_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{\hat{S}_1^2}{n_1} + \frac{\hat{S}_2^2}{n_2}}$$

where $\hat{S}_1^2 = \dfrac{\Sigma(X_{i1} - \overline{X}_1)^2}{n_1 - 1}$ and $\hat{S}_2^2 = \dfrac{\Sigma(X_{i2} - \overline{X}_2)^2}{n_2 - 1}$

Theorem 11.9 *The variance of the sampling distribution of* $(\overline{X}_1 - \overline{X}_2)$ *is.*

$$\sigma^2_{\overline{X}_1 - \overline{X}_2} = Var(\overline{X}_1 - \overline{X}_2) = \frac{\sigma_1^2}{n_1}\left(\frac{N_1 - n_1}{N_1 - 1}\right) + \frac{\sigma_2^2}{n_2}\left(\frac{N_2 - n_2}{N_2 - 1}\right)$$

where $\overline{X}_1$ *and* $\overline{X}_2$ *are the means of random samples of sizes* n_1 *and* n_2 *drawn without replacement from finite populations of sizes* N_1 *and* N_2 *with means* μ_1 *and* μ_2 *and variances* σ_1^2 *and* σ_2^2, *respectively.*

The standard error of $(\overline{X}_1 - \overline{X}_2)$ then becomes

$$\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{Var(\overline{X}_1 - \overline{X}_2)} = \sqrt{\frac{\sigma_1^2}{n_1}\left(\frac{N_1 - n_1}{N_1 - 1}\right) + \frac{\sigma_2^2}{n_2}\left(\frac{N_2 - n_2}{N_2 - 1}\right)}$$

Theorem 11.10 *If* $\overline{X}_1$ *and* $\overline{X}_2$ *are the means of random samples of* n_1 *and* n_2 *observations from two independent normal populations with means* μ_1, μ_2 *and variances* σ_1^2, σ_2^2 *respectively, then the sampling distribution of the difference between sample means* $\overline{X}_1 - \overline{X}_2$ *is normal with mean and variance*

$$\mu_{\overline{X}_1 - \overline{X}_2} = \mu_1 - \mu_2$$

$$\sigma^2_{\overline{X}_1 - \overline{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$$

That is, the distribution of the random variable

$$Z = \frac{(\overline{X}_1 - \overline{X}_2) - \mu_{\overline{X}_1 - \overline{X}_2}}{\sigma_{\overline{X}_1 - \overline{X}_2}} = \frac{(\overline{X}_1 - \overline{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$$

is a standard normal distribution.

***Example* 11.9** *Let $\overline{X}_1$ represent the mean of a sample of size $n_1 = 2$ selected at random with replacement from a finite population consisting of values 7 and 9. Similarly, let $\overline{X}_2$ represent the mean of a sample of size $n_2 = 3$ selected at random with replacement from another finite population consisting of values 3 and 6. Form a sampling distribution of the random variable $(\overline{X}_1 - \overline{X}_2)$. Verify that*

(i) $\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$ (ii) $\sigma^2_{\bar{X}_1 - \bar{X}_2} = \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$

Solution. We have

***Population* I:** 7, 9; $N_1 = 2$; $n_1 = 2$

Number of possible samples $= N_1 \times N_1 = 2 \times 2 = 4$

Possible samples (7, 7), (7, 9), (9, 7), (9, 9)

Sample means $\bar{x}_1$ 7, 8, 8, 9

***Population* II:** 3, 6; $N_2 = 2$; $n_2 = 3$

Number of possible samples $= N_2 \times N_2 \times N_2 = 2 \times 2 \times 2 = 8$

Possible samples (3, 3, 3), (3, 3, 6), (3, 6, 3), (6, 3, 3), (3, 6, 6), (6, 3, 6), (6, 6, 3), (6, 6, 6)

Sample means $\bar{x}_2$ 3, 4, 4, 4, 5, 5, 5, 6

All possible differences between sample means $(\overline{X}_1 - \overline{X}_2)$ are

	$\bar{x}_2$							
$\bar{x}_1$	3	4	4	4	5	5	5	6
7	4	3	3	3	2	2	2	1
8	5	4	4	4	3	3	3	2
8	5	4	4	4	3	3	3	2
9	6	5	5	5	4	4	4	3

The sampling distribution of $\overline{X}_1 - \overline{X}_2$, its mean and variance are

Value of $\overline{X}_1 - \overline{X}_2$ $\bar{x}_1 - \bar{x}_2$	Number of occurrences f	Probability $p(\bar{x}_1 - \bar{x}_2) = f/\Sigma f$	$(\bar{x}_1 - \bar{x}_2)\, p(\bar{x}_1 - \bar{x}_2)$	$(\bar{x}_1 - \bar{x}_2)^2\, p(\bar{x}_1 - \bar{x}_2)$
1	1	1/32	1/32	1/32
2	5	5/32	10/32	20/32
3	10	10/32	30/32	90/32
4	10	10/32	40/32	160/32
5	5	5/32	25/32	125/32
6	1	1/32	6/32	36/32
Sums	$\Sigma f = 32$	1	112/32	432/32

$$\mu_{\bar{X}_1 - \bar{X}_2} = E(\bar{X}_1 - \bar{X}_2) = \sum (\bar{x}_1 - \bar{x}_2)\, p(\bar{x}_1 - \bar{x}_2) = \frac{112}{32} = 3.5$$

$$\sigma^2_{\bar{X}_1 - \bar{X}_2} = Var(\bar{X}_1 - \bar{X}_2) = \sum (\bar{x}_1 - \bar{x}_2)^2\, p(\bar{x}_1 - \bar{x}_2) - \mu^2_{\bar{X}_1 - \bar{X}_2}$$

$$= \frac{432}{32} - (3.5)^2 = 1.25$$

The mean and variance of population I are

x_1	7	9	$\sum x_1 = 16$
x_1^2	49	81	$\sum x_1^2 = 130$

$$\mu_1 = \frac{\sum x_1}{N_1} = \frac{16}{2} = 8$$

$$\sigma_1^2 = \frac{\sum x_1^2}{N_1} - \mu_1^2 = \frac{130}{2} - (8)^2 = 1$$

The mean and variance of population II are

x_2	3	6	$\sum x_2 = 9$
x_2^2	9	36	$\sum x_2^2 = 45$

$$\mu_2 = \frac{\sum x_2}{N_2} = \frac{9}{2} = 4.5$$

$$\sigma_2^2 = \frac{\sum x_2^2}{N_2} - \mu_2^2 = \frac{45}{2} - (4.5)^2 = 2.25$$

We are to verify that

(*i*). $\mu_{\bar{X}_1 - \bar{X}_2} = \mu_1 - \mu_2$ (*ii*) $\sigma^2_{\bar{X}_1 - \bar{X}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$

$3.5 = 8 - 4.5$ $1.25 = \frac{1}{2} + \frac{2.25}{3}$

$3.5 = 3.5$ $1.25 = 1.25$

Example 11.10 *Two independent random samples of sizes $n_1 = 30$ and $n_2 = 50$ are taken from two populations having means $\mu_1 = 78$ and $\mu_2 = 75$ and variances $\sigma_1^2 = 150$ and $\sigma_2^2 = 200$. Let $\bar{X}_1$ be the mean of the first random sample and $\bar{X}_2$ be the mean of the second random sample. Find the mean and standard error of $\bar{X}_1 - \bar{X}_2$.*

Solution. We have $\mu_1 = 78,\quad \sigma_1^2 = 150,\quad n_1 = 30$

$\mu_2 = 75,\quad \sigma_2^2 = 200,\quad n_2 = 50$

Then mean and standard error of $\overline{X}_1 - \overline{X}_2$ are

$$\mu_{\overline{X}_1 - \overline{X}_2} = \mu_1 - \mu_2 = 78 - 75 = 3$$

$$\sigma_{\overline{X}_1 - \overline{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{150}{30} + \frac{200}{50}} = 3$$

Exercise 11.3

1. (a) What is meant by the sampling distribution of the difference between two sample means. Describe the properties of the sampling distribution of the differences between two sample means.

(b) Let $\overline{X}_1$ represent the mean of a sample of size $n_1 = 2$, selected with replacement from a finite population -2, 0, 2, and 4. Similarly, let $\overline{X}_2$ represent the mean of a sample of size $n_2 = 2$, selected with replacement from the population -1 and 1.

(i) Assuming that the 64 possible differences $\overline{X}_1 - \overline{X}_2$ are equally likely to occur, construct the sampling distribution of $\overline{X}_1 - \overline{X}_2$.

(ii) Verify that $\sigma^2_{\overline{X}_1 - \overline{X}_2} = \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$

$\{ \mu_1 = 1, \mu_2 = 0, \mu_{\overline{X}_1 - \overline{X}_2} = 1, \sigma_1^2 = 5, \sigma_2^2 = 1, \sigma^2_{\overline{X}_1 - \overline{X}_2} = 3 \}$

2. (a) Let the variable $\overline{X}_1$ represent the mean of random samples of size $n_1 = 2$, with replacement drawn from the finite population 3, 4, 5. Similarly, let $\overline{X}_2$ represent the means of random samples of size $n_2 = 3$, with replacement, drawn from the population 0, 3. Assuming that the 72 possible differences $\overline{X}_1 - \overline{X}_2$ are equally likely to occur, construct the sampling distribution of $\overline{X}_1 - \overline{X}_2$. Show that

(i) $\mu_{\overline{X}_1 - \overline{X}_2} = \mu_1 - \mu_2$ (ii) $\sigma^2_{\overline{X}_1 - \overline{X}_2} = \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$

$\{ \mu_1 = 4, \mu_2 = 1.5, \mu_{\overline{X}_1 - \overline{X}_2} = 2.5, \sigma_1^2 = 0.667, \sigma_2^2 = 2.25, \sigma^2_{\overline{X}_1 - \overline{X}_2} = 1.083 \}$

(b) Let the variable $\overline{X}_1$ represent the means of random samples of size 2 without replacement, drawn from the finite population 5, 7, 9. Similarly, let $\overline{X}_2$ represent the means of random sample of size 2, without replacement from another finite population 4, 6, 8. Assuming that the 36 possible differences $\overline{X}_1 - \overline{X}_2$ are equally likely to occur, construct the sampling distribution of $\overline{X}_1 - \overline{X}_2$ and verify that

(i) $\mu_{\overline{X}_1 - \overline{X}_2} = \mu_1 - \mu_2$

(ii) $\sigma^2_{\overline{X}_1 - \overline{X}_2} = \dfrac{\sigma_1^2}{n_1}\left(\dfrac{N_1 - n_1}{N_1 - 1}\right) + \dfrac{\sigma_2^2}{n_2}\left(\dfrac{N_2 - n_2}{N_2 - 1}\right)$

$\{ \mu_1 = 7, \mu_2 = 6, \mu_{\bar{X}_1 - \bar{X}_2} = 1, \sigma_1^2 = 2.667, \sigma_2^2 = 2.667\ \sigma^2_{\bar{X}_1 - \bar{X}_2} = 1.333 \}$

3. **(*a*)** The television picture tubes of manufacturer A have a mean lifetime of 6.5 years and a standard deviation of 0.9 years, while those of manufacturer B have a mean lifetime of 6.0 years and a standard deviation of 0.8 years. A random sample of size 36 tubes is selected from manufacturer A and its mean $\bar{X}_1$ is calculated. An other random sample of size 49 tubes is selected from manufacturer B and its mean $\bar{X}_2$ is calculated. Find the mean and standard error of the sampling distribution of the difference $\bar{X}_1 - \bar{X}_2$.
$(\mu_{\bar{X}_1 - \bar{X}_2} = 0.5, \sigma_{\bar{X}_1 - \bar{X}_2} = 0.1886)$

(*b*) Random samples of each size 100 are drawn from two independent probability distributions and their means $\bar{X}_1$ and $\bar{X}_2$ computed. If the means and standard deviations of the two populations are $\mu_1 = 10, \sigma_1 = 2, \mu_2 = 8, \sigma_2 = 1$, find the mean and standard error of the sampling distribution of the difference $\bar{X}_1 - \bar{X}_2$.
$(\mu_{\bar{X}_1 - \bar{X}_2} = 2, \sigma_{\bar{X}_1 - \bar{X}_2} = 0.2236)$

11.14 SAMPLING DISTRIBUTION OF SAMPLE PROPORTION, *P*

The *sampling distribution of sample proportion* P is the probability distribution of the proportions of successes obtained from all possible simple random samples of n observations that can be drawn from a Bernoulli population with proportion of successes π.

11.14.1 Population Proportion. The *population proportion* is defined as

$$\pi = \frac{\textit{No. of elements with attribute } A}{\textit{Population size}} = \frac{k}{N}$$

where k is the number of elements in the population of size N that possess a certain characteristic. In many applications of sampling the characteristic of interest in the population elements is qualitative with two possible outcomes. Quite often, however, we are interested not in the number of successes but rather in the proportion of successes.

11.14.2 Sample Statistics X and P. When the characteristic of interest is qualitative with two possible outcomes, a sample statistic of interest is the *number of occurrences* among the n sample observations consisting of the particular outcome reflected in the population proportion. This number of occurrence is denoted by X. Another sample statistic is the *sample proportion*, denoted by P, which is defined as

$$P = \frac{\textit{No. of elements with attribute } A}{\textit{Sample size}} = \frac{X}{n}$$

The observed value $p = x/n$ of sample proportion P will serve as an estimate of π. Obviously, the actual value we obtain for p will vary from sample to sample. So we ask, how good the estimate obtained will be. Are the values of P likely to be close to the true proportion π in the population. To what extent will they vary from one sample to another. Now for our theoretical model, we define a population in which a given proportion π have a specific attribute

A. We suppose that every unit in the population falls into one of the two categories A and $\overline{A}$. The notation is as follows

Number of units in A in	
Population	Sample
k	x

Proportion of units in A in	
Population	Sample
$\pi = \dfrac{k}{N}$	$p = \dfrac{x}{n}$

The estimate of proportion of successes π in the population is the sample proportion p and the estimate of the total number of successes k in population is thus Np or Nx/n.

11.14.3 Binomial Distribution as Sampling Distribution: ***Sampling Infinite Populations.*** If a simple random sample of size n is selected from an infinite population (or with replacement from a finite population) whose elements are characterised by some attribute to belong to one of the two mutually exclusive and exhaustive categories where one of these will be designated a 'success' and the other will be designated a 'failure', then the exact sampling distribution of the proportion of successes P is a binomial distribution.

11.14.4 Properties of Sampling Distribution of P. The properties of the sampling distribution of the sample proportion P are as follows:

Mean and Variance. The mean and variance of the binomial sampling distribution of P for a simple random sample of size n from an infinite Bernoulli population (or with replacement from a finite Bernoulli population) are given in the following theorem.

***Theorem* 11.11** *If the population is infinite or the sampling is done with replacement, the sample proportion P has its mean and variance as*

$$\mu_P = E(P) = \pi$$

$$\sigma_P^2 = Var(P) = \frac{\pi(1-\pi)}{n}$$

where π is the probability of success and $(1 - \pi)$ is the probability of failure. The standard deviation (often called the standard error or sampling variability) is

$$\sigma_P = \sqrt{Var(P)} = \sqrt{\frac{\pi(1-\pi)}{n}}$$

However, if the value of π is unknown, it is replaced by sample proportion P, the estimate of the standard error of P then becomes

$$\hat{\sigma}_P = \sqrt{\frac{P(1-P)}{n}}$$

Shape of Distribution. The sampling distribution of P is skewed to the right if $\pi < 0.5$, skewed to the left if $\pi > 0.5$ and symmetrical if $\pi = 0.5$.

Normal approximation. As n tends to infinity, the distribution of P becomes approximately normal with mean π and variance $\pi(1-\pi)/n$. That is, the distribution of the random variable

$$Z = \frac{P - \mu_P}{\sigma_P} = \frac{P - \pi}{\sqrt{\pi(1-\pi)/n}}$$

approach the standard normal distribution as n approaches infinity.

***Example* 11.11** *A population consists of 5 members. The marital status of each member is given below*

Member	1	2	3	4	5
Marital status	*S*	*M*	*S*	*M*	*S*

where M and S stands for married and single respectively. Determine the proportion of married members in the population. Take all possible samples of two members with replacement from this population and find the proportion of married members in each sample. Form the sampling distribution of the sample proportion P and verify that

(*i*) $\mu_P = \pi$ (*ii*) $\sigma_P^2 = \dfrac{\pi(1-\pi)}{n}$

Solution. Population: 1, 2, 3, 4, 5; Population size: $N = 5$; Sample size: $n = 2$

The members with even serial numbers 2 and 4 are married while those with odd serial numbers 1, 3 and 5 are single.

Number of married members in the population: $k = 2$

Proportion of married members in the population: $\pi = \dfrac{k}{N} = \dfrac{2}{5} = 0.4$

Number of possible samples $= N \times N = 5 \times 5 = 25$

All possible samples, the number of married members and the proportion of married members in each sample are given below.

Members in sample	Number of married members x	Proportion of married members $p = x/n$	Members in sample	Number of married members x	Proportion of married members $p = x/n$
1, 1	0	0	3, 3	0	0
1, 2	1	1/2	3, 4	1	1/2
1, 3	0	0	3, 5	0	0
1, 4	1	1/2	4, 1	1	1/2
1, 5	0	0	4, 2	2	1
2, 1	1	1/2	4, 3	1	1/2
2, 2	2	1	4, 4	2	1
2, 3	1	1/2	4, 5	1	1/2
2, 4	2	1	5, 1	0	0
2, 5	1	1/2	5, 2	1	1/2
3, 1	0	0	5, 3	0	0
3, 2	1	1/2	5, 4	1	1/2
		Continued	5, 5	0	0

The sampling distribution of sample proportion P, its mean and variance are

Value of P P	Number of occurrences f	Probability $f(p) = f/\Sigma f$	$p\,f(p)$	$p^2 f(p)$
0	9	9/25	0	0
1/2	12	12/25	6/25	3/25
1	4	4/25	4/25	4/25
Sum	$\Sigma f = 25$	1	10/25	7/25

$$\mu_P = E(P) = \sum p\, f(p) = \frac{10}{25} = 0.4$$

$$\sigma_P^2 = Var(P) = \sum p^2 f(p) - \mu_P^2 = \frac{7}{25} - (0.4)^2 = 0.12$$

We are to verify that (*i*) $\mu_P = \pi$ (*ii*) $\sigma_P^2 = \dfrac{\pi(1-\pi)}{n}$

$$0.4 = 0.4 \qquad 0.12 = \frac{0.4(1-0.4)}{2}$$

$$0.12 = 0.12$$

***Example* 11.12** *It is known that 3 % of the persons living in Gujranwala city are known to have a certain disease. Find the mean and standard error of sampling distribution of proportion of diseased persons in a random sample of 500 persons.*

Solution. We have proportion in the population $\pi = 0.03$ and the sample size $n = 500$. Let P be the random variable 'the proportion of persons in the sample which are diseased'. Then, the mean and standard error of P are

$$\mu_P = \pi = 0.03$$

$$\sigma_P = \sqrt{\frac{\pi(1-\pi)}{n}} = \sqrt{\frac{0.03(1-0.03)}{500}} = 0.00763$$

11.14.5 Hypergeometric Distribution as Sampling Distribution: *Sampling Finite Populations.* When a simple random sample of size n is selected without replacement from a finite population whose elements are characterised by some attribute to belong to one of the two mutually exclusive and exhaustive categories where one of these will be designated a 'success' and the other will be designated a 'failure', then the exact sampling distribution of the proportion of successes P is a hypergeometric distribution.

11.14.6 Properties of Sampling Distribution of *P*. The properties of the sampling distribution of the sample proportion P are as follows:

Mean and Variance. The mean and variance of the hypergeometric sampling distribution of P for simple random sampling without replacement from a finite Bernoulli population are given in the following theorem.

***Theorem* 11.12** *If the population is finite and the sampling is done without replacement, the sample proportion P has its mean and variance as*

$$\mu_P = E(P) = \pi$$

$$\sigma_P^2 = Var(P) = \frac{\pi(1-\pi)}{n}\left(\frac{N-n}{N-1}\right)$$

where π is the probability of success and $(1 - \pi)$ is the probability of failure. The standard deviation (often called the standard error or sampling variability) is

$$\sigma_P = \sqrt{Var(P)} = \sqrt{\frac{\pi(1-\pi)}{n}}\sqrt{\frac{N-n}{N-1}}$$

However, if the value of π is unknown, it is replaced by sample proportion P, the estimate of the standard error of P then becomes

$$\hat{\sigma}_P = \sqrt{\frac{P(1-P)}{n}}\sqrt{\frac{N-n}{N-1}}$$

***Example* 11.13** *Draw all possible samples of size* 2 *at random without replacement from the population* 1, 2, 3, 4, 5. *Find the proportion of even numbers in the* samples. *Form the sampling distribution of the sample proportion* P *and verify that*

(*i*) $\mu_P = \pi$ (*ii*) $\sigma_P^2 = \frac{\pi(1-\pi)}{n}\left(\frac{N-n}{N-1}\right)$

Solution. Population: 1, 2, 3, 4, 5; Population size: $N = 5$; Sample size: $n = 2$

Number of even numbers in the population: $k = 2$

Proportion of even numbers in the population: $\pi = \frac{k}{N} = \frac{2}{5} = 0.4$

Number of possible samples $= N(N-1) = 5(5-1) = 20$

All possible samples, the number of even numbers and the proportion of even numbers in each sample are given below.

Sample values	Number of even numbers x	Proportion of even numbers $p = x/n$	Sample values	Number of even numbers x	Proportion of even numbers $p = x/n$
1, 2	1	1/2	3, 4	1	1/2
1, 3	0	0	3, 5	0	0
1, 4	1	1/2	4, 1	1	1/2
1, 5	0	0	4, 2	2	1
2, 1	1	1/2	4, 3	1	1/2
2, 3	1	1/2	4, 5	1	1/2
2, 4	2	1	5, 1	0	0
2, 5	1	1/2	5, 2	1	1/2
3, 1	0	0	5, 3	0	0
3, 2	1	1/2	5, 4	1	1/2
		Continued			

The sampling distribution of sample proportion P, its mean and variance are

Value of P	Number of occurrences	Probability		
p	f	$f(p) = f/\Sigma f$	$p\,f(p)$	$p^2 f(p)$
0	6	6/20	0	0
1/2	12	12/20	6/20	3/20
1	2	2/20	2/20	2/20
Sum	$\Sigma f = 20$	1	8/20	5/25

$$\mu_P = E(P) = \sum p\,f(p) = \frac{8}{20} = 0.4$$

$$\sigma_P^2 = Var(P) = \sum p^2 f(p) - \mu_P^2 = \frac{5}{20} - (0.4)^2 = 0.09$$

We are to verify that

(*i*) $\mu_P = \pi$ (*ii*) $\sigma_P^2 = \dfrac{\pi(1-\pi)}{n}\left(\dfrac{N-n}{N-1}\right)$

$0.4 = 0.4$ $\quad 0.12 = \dfrac{0.4(1-0.4)}{2}\left(\dfrac{5-2}{5-1}\right)$

$$0.09 = 0.09$$

11.15 SAMPLING DISTRIBUTION OF THE DIFFERENCE BETWEEN TWO SAMPLE PROPORTIONS, $P_1 - P_2$

The *sampling distribution of the difference between two sample proportions* $P_1 - P_2$ is the probability distribution of all possible differences between proportions P_1 and P_2 obtained from all possible independent simple random samples of n_1 and n_2 observations that can be drawn from two Bernoulli populations with population proportions of π_1 and π_2, respectively. Often we wish to compare the proportions of successes in two Bernoulli populations. We must use the sample proportions of successes as our basis of comparison. Obviously, the number of successes in both samples cannot be used alone as a means of evaluation. Specifically we require a probability model of the difference between two sample proportions.

Suppose that two independent random samples of sizes n_1 and n_2 are drawn from Bernoulli populations with population proportions of π_1 and π_2, respectively. Let P_1 be the proportion of successes in sample of size n_1 from the population with population proportion π_1, then P_1 is a random variable that has its own probability distribution with mean π_1 and variance $\pi_1(1-\pi_1)/n_1$. Let P_2 be the proportion of successes in sample of size n_2 from the population with population proportion π_2, then P_2 is a random variable that has its own probability distribution with mean π_2 and variance $\pi_2(1-\pi_2)/n_2$. Then the difference $P_1 - P_2$ can be obtained from all possible pairs of P_1 and P_2. Consequently, the difference $P_1 - P_2$ between the two sample proportions is a random variable that has its own probability distribution which is called the sampling distribution of the difference between two sample proportions.

11.15.1 Properties of the Sampling Distribution of the Difference between Two Sample Proportions. The properties of the sampling distribution of the difference $P_1 - P_2$ between two sample proportions are given by the following theorems.

***Theorem* 11.13** *The mean of the sampling distribution of* $P_1 - P_2$, *denoted by* $\mu_{P_1 - P_2}$, *is equal to the difference between the population proportions, i. e.,*

$$\mu_{P_1 - P_2} = E(P_1 - P_2) = \pi_1 - \pi_2$$

This theorem holds regardless of the sample sizes n_1 and n_2 or whether the sampling is done with or without replacement.

***Theorem* 11.14** *If the populations are infinite or the sampling is done with replacement, the difference between sample proportions* $P_1 - P_2$ *has its variance as*

$$\sigma^2_{P_1 - P_2} = Var(P_1 - P_2) = \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}$$

The standard error of $P_1 - P_2$ becomes

$$\sigma_{P_1 - P_2} = \sqrt{Var(P_1 - P_2)} = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}$$

However, if the values of π_1 and π_2 are unknown, these are replaced by sample proportions P_1 and P_2, the estimate of the standard error of $P_1 - P_2$ then becomes

$$\hat{\sigma}_{P_1 - P_2} = \sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}}$$

***Theorem* 11.15** *If the populations are finite and the sampling is done without replacement, the difference between sample proportions* $P_1 - P_2$ *has its variance as*

$$\sigma^2_{P_1 - P_2} = Var(P_1 - P_2)$$

$$= \frac{\pi_1(1-\pi_1)}{n_1}\left(\frac{N_1 - n_1}{N_1 - 1}\right) + \frac{\pi_2(1-\pi_2)}{n_2}\left(\frac{N_2 - n_2}{N_2 - 1}\right)$$

The standard error of $P_1 - P_2$ is

$$\sigma_{P_1 - P_2} = \sqrt{Var(P_1 - P_2)}$$

$$= \sqrt{\frac{\pi_1(1-\pi_1)}{n_1}\left(\frac{N_2 - n_2}{N_2 - 1}\right) + \frac{\pi_2(1-\pi_2)}{n_2}\left(\frac{N_2 - n_2}{N_2 - 1}\right)}$$

***Example* 11.14** *Let P_1 represent the proportion of odd numbers in a sample of size $n_1 = 2$ selected at random with replacement from a finite population consisting of values 4 and 5. Similarly, let P_2 represent the proportion of odd numbers in a sample of size $n_2 = 2$ selected at random with replacement from another finite population consisting of values 2, 3 and 6. From a sampling distribution of the random variable $(P_1 - P_2)$. Verify that*

(i) $\mu_{P_1 - P_2} = \pi_1 - \pi_2$ *(ii)* $\sigma^2_{P_1 - P_2} = \dfrac{\pi_1(1-\pi_1)}{n_1} + \dfrac{\pi_2(1-\pi_2)}{n_2}$

Solution. We have

***Population* I:** 4, 5; $N_1 = 2$; $n_1 = 2$

Number of odd numbers: $k_1 = 2$

Proportion of odd numbers: $\pi_1 = \dfrac{k_1}{N_1} = \dfrac{1}{2}$

Number of possible samples: $N_1 \times N_1 = 2 \times 2 = 4$

Possible samples:

(4, 4) (4, 5) (5, 4) (5, 5)

Sample proportion of odd numbers: p_1

0 1/2 1/2 1

***Population* II:** 2, 3, 6; $N_2 = 3$; $n_2 = 2$

Number of odd numbers: $k_2 = 1$

Proportion of odd numbers: $\pi_2 = \dfrac{k_2}{N_2} = \dfrac{1}{3}$

Number of possible samples: $N_2 \times N_2 = 3 \times 3 = 9$

Possible samples:

(2, 2) (2, 3) (2, 6) (3, 2) (3, 3) (3, 6) (6, 2) (6, 3) (6, 6)

Sample proportion of odd numbers: p_2

0 1/2 0 1/2 1 1/2 0 1/2 0

All possible differences between sample proportions $(P_1 - P_2)$ are

	p_2								
p_1	0	0	0	0	1/2	1/2	1/2	1/2	1
0	0	0	0	0	−1/2	−1/2	−1/2	−1/2	−1
1/2	1/2	1/2	1/2	1/2	0	0	0	0	−1/2
1/2	1/2	1/2	1/2	1/2	0	0	0	0	−1/2
1	1	1	1	1	1/2	1/2	1/2	1/2	0

The sampling distribution of $P_1 - P_2$, its mean and variance are

Value of $P_1 - P_2$ $p_1 - p_2$	Number of occurrences f	Probability $f(p_1-p_2) = f/\sum f$	$(p_1-p_2)f(p_1-p_2)$	$(p_1-p_2)^2 f(p_1-p_2)$
-1	1	1/36	$-1/36$	1/36
$-1/2$	6	6/36	$-3/36$	3/72
0	13	13/36	0	0
1/2	12	12/36	6/36	3/36
1	4	4/36	4/36	4/36
Sum	$\sum f = 36$	1	6/36	19/72

$$\mu_{P_1 - P_2} = E(P_1 - P_2) = \sum (p_1 - p_2)(p_1 - p_2) = \frac{6}{36} = \frac{1}{6}$$

$$\sigma^2_{P_1 - P_2} = Var(P_1 - P_2) = \sum (p_1 - p_2)^2 f(p_1 - p_2) - \mu^2_{P_1 - P_2}$$

$$= \frac{19}{72} - \left(\frac{1}{6}\right)^2 = \frac{17}{72}$$

We are to verify that

(i) $\mu_{P_1 - P_2} = \pi_1 - \pi_2$ (ii) $\sigma^2_{P_1 - P_2} = \frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}$

$$\frac{1}{6} = \frac{1}{2} - \frac{1}{3} \qquad \frac{17}{72} = \frac{\frac{1}{2}\left(1-\frac{1}{2}\right)}{2} + \frac{\frac{1}{3}\left(1-\frac{1}{3}\right)}{2}$$

$$\frac{1}{6} = \frac{1}{6} \qquad \frac{17}{72} = \frac{17}{72}$$

***Example* 11.15** *The actual proportion of men who like a certain TV programme is 0.30 and the corresponding proportion for women is 0.25. A questionnaire about this program is given to 500 men and 500 women, and the individual responses are looked upon as the values of independent random variables having Bernoulli distributions with parameters $\pi_1 = 0.30$ and $\pi_2 = 0.25$, respectively. Find the mean and standard error of $P_1 - P_2$, the difference between the sample proportions of successes.*

Solution. We have $\pi_1 = 0.30$, $\pi_2 = 0.25$; $n_1 = 500$, $n_2 = 500$.

The mean and standard error of $P_1 - P_2$ are

$$\mu_{P_1 - P_2} = \pi_1 - \pi_2 = 0.30 - 0.25 = 0.05$$

$$\sigma_{P_1 - P_2} = \sqrt{\frac{\pi_1(1-\pi_1)}{n_1} + \frac{\pi_2(1-\pi_2)}{n_2}}$$

$$= \sqrt{\frac{0.30(1-0.30)}{500} + \frac{0.25(1-0.25)}{500}} = 0.028$$

11.16 OTHER SAMPLING DISTRIBUTIONS

We have considered the sampling distributions of sample mean, difference between sample means, sample proportions and difference between sample proportions. Other statistics such as sample median, sample variance and sample standard deviation have their own sampling distributions. There is different sampling distribution for each different statistic even though the statistics may be computed from the same sample. For a given statistic, the sampling distribution will vary for samples of different sizes. Thus, in a sampling distribution it is necessary to specify the population, the statistic and the size of the sample. A change in any of these specifications will result a different sampling distribution.

11.17 SAMPLING DISTRIBUTION OF THE SAMPLE VARIANCE, S^2

The *sampling distribution of sample variance* S^2 is the probability distribution of the variances obtained from all possible simple random samples of n observations that can be drawn from a population with variance σ^2.

The sampling distribution of sample variance has the property

$$\mu_{S^2} = E(S^2) = \frac{n-1}{n}\sigma^2$$

***Example* 11.16** *A population consists of five numbers* 2, 4, 6, 8, *and* 10. *Consider all possible samples of size* 2 *which can be drawn with replacement from this population. Form the sampling distribution of sample variance and verify that*

$$\mu_{S^2} = \frac{n-1}{n}\sigma^2$$

Solution. Population: 2, 4, 6, 8, 10; Population size: $N = 5$; Sample size : $n = 2$

Number of possible samples $= N \times N = 5 \times 5 = 25$

All possible samples that can be drawn with replacement from our population are

(2, 2)	(2, 4)	(2, 6)	(2, 8)	(2, 10)
(4, 2)	(4, 4)	(4, 6)	(4, 8)	(4, 10)
(6, 2)	(6, 4)	(6, 6)	(6, 8)	(6, 10)
(8, 2)	(8, 4)	(8, 6)	(8, 8)	(8, 10)
(10, 2)	(10, 4)	(10, 6)	(10, 8)	(10, 10)

(i) All possible sample variances: $s^2 = \frac{\Sigma(x_i - \bar{x})^2}{n} = \frac{(x_1 - x_2)^2}{4}$ when $n = 2$

0	1	4	9	16
1	0	1	4	9
4	1	0	1	4
9	4	1	0	1
16	9	4	1	0

The sampling distribution of sample variance S^2 and its mean are

Value of S^2 s^2	Number of occurrences: f	Probability $p(s^2) = f/\sum f$	$s^2\, p(s^2)$
0	5	5/25	0
1	8	8/25	8/25
4	6	6/25	24/25
9	4	4/25	36/25
16	2	2/25	32/25
	$\sum f = 25$	1	$\sum s^2 p(s^2) = 100/25$

$$\mu_{S^2} = E(S^2) = \sum s^2\, p(s^2) = \frac{100}{25} = 4$$

The mean and variance of the population are

x_j	2	4	6	8	10	$\sum x_j = 30$
x_j^2	4	16	36	64	100	$\sum x_j^2 = 220$

$$\mu = \frac{\sum x_j}{N} = \frac{30}{5} = 6$$

$$\sigma^2 = \frac{\sum x_j^2}{N} - \mu^2 = \frac{220}{5} - (6)^2 = 8$$

We are to verify that $\quad \mu_{S^2} = \frac{n-1}{n}\sigma^2$

$$4 = \frac{2-1}{2}(8)$$

$$4 = 4$$

Exercise 11.4

1. (*a*) A fair coin is tossed 50 times and the number of heads recorded are 27. The proportion of heads was, therefore, estimated to be 0.54. Answer the following.
   (*i*) Which figure is a parameter?
   (*ii*) Which figure is a statistic?
   { (*i*) The probability of head in a single trial $\pi = 0.5$;
   (*ii*) Sample size $n = 50$, number of heads in the sample $x = 30$ and the proportion of heads in the sample $p = x/n = 0.54$. }

   (*b*) What is meant by the sampling distribution of sample proportion? Describe the properties of the sampling distribution of sample proportion.

2. (a) A finite population consists of the numbers 2, 3, 4, 5, 6 and 8. Find the proportion P of even numbers in all possible random samples of size $n = 2$ that can be drawn with replacement from this population. Assuming the 36 possible samples equally likely, make the sampling distribution of sample proportions and find the mean and variance of this distribution. Verify that

(*i*) $E(P) = \pi$ (*ii*) $Var(P) = \dfrac{\pi(1-\pi)}{n}$

where P and π are sample and population proportions respectively.
$\{\pi = 2/3,\ \mu_P = 2/3,\ \sigma_P^2 = 1/9\}$

(b) A population consists of $N = 4$ numbers 1, 3, 4 and 5. Find the proportion P of odd numbers in all possible samples of size $n = 3$ that can be drawn without replacement from this population. Assuming the 24 possible samples equally likely, construct the sampling distribution of sample proportions and find the mean and variance of this distribution. Verify that

(*i*) $\mu_P = \pi$, (*ii*) $\sigma_P^2 = \dfrac{\pi(1-\pi)}{n}\left(\dfrac{N-n}{N-1}\right)$

where P and π are sample and population proportions respectively.
$\{\pi = 3/4,\ \mu_P = 3/4,\ \sigma_P^2 = 1/48\}$

3. (a) Suppose that 60 % of a city population favours public finding for a proposed recreational facility. If 150 persons are to be randomly selected and interviewed, what is the mean and standard error of the sample proportion favouring this issue.
$\{\mu_P = 0.60,\ \sigma_P = 0.04\}$

(b) A small, professional society has $N = 4500$ members. The president has mailed $n = 400$ questionnaires to a random sample of members asking whether they wish to affiliate with a large group. Assuming that the proportion of the entire membership favouring consolidation is $\pi = 0.7$, find the mean and standard error of the sample proportion P.
$\{\mu_P = 0.7,\ \sigma_P = 0.022\}$

4. (a) What is meant by the sampling distribution of the difference between two sample proportions? Describe the properties of the sampling distribution of difference between two sample proportions. Explain its usefulness in statistical inference.

(b) Let P_1 represent the proportion of odd numbers in a random sample of size $n_1 = 3$ with replacement from a finite population consisting of values 4 and 5. Similarly, let P_2 represent the proportion of odd numbers in a random sample of size $n_2 = 2$ with replacement from another finite population consisting of values 2, 3 and 6. Assuming that the 72 possible differences $P_1 - P_2$ are equally likely to occur, construct the sampling distribution of $P_1 - P_2$. Verify that

(*i*) $\mu_{P_1-P_2} = \pi_1 - \pi_2$ (*ii*) $\sigma^2_{P_1-P_2} = \dfrac{\pi_1(1-\pi_1)}{n_1} + \dfrac{\pi_2(1-\pi_2)}{n_2}$

$\{\pi_1 = 1/2,\ \pi_2 = 1/3,\ \mu_{P_1-P_2} = 1/6,\ \sigma^2_{P_1-P_2} = 7/36\}$

5. (*a*) Let P_1 represent the proportion of even numbers in a random sample of size $n_1 = 2$ without replacement from a finite population consisting of values 4, 6 and 9. Similarly, let P_2 represent the proportion of even numbers in a random sample of size $n_2 = 2$ without replacement from another finite population consisting of values 2, 3 and 5. Assuming that the 36 possible differences $P_1 - P_2$ are equally likely to occur, construct the sampling distribution of $P_1 - P_2$. Verify that

(*i*) $\mu_{P_1 - P_2} = \pi_1 - \pi_2$

(*ii*) $\sigma^2_{P_1 - P_2} = \dfrac{\pi_1(1-\pi_1)}{n_1}\left(\dfrac{N_1 - n_1}{N_1 - 1}\right) + \dfrac{\pi_2(1-\pi_2)}{n_2}\left(\dfrac{N_2 - n_2}{N_2 - 1}\right)$

$\{\pi_1 = 2/3,\ \pi_2 = 1/3,\ \mu_{P_1 - P_2} = 1/3,\ \sigma^2_{P_1 - P_2} = 1/9\}$

(*b*) The percentage of families with a monthly income of Rs. 1,000 or more in city *A* and city *B* is 25 % and 20 % respectively. If a random sample of 100 families is selected from each of these two cities and the proportions of families earning Rs. 1,000 or more in the two samples are compared, what is the mean and standard error of $P_1 - P_2$, the difference between the sample proportions?
$\{\mu_{P_1 - P_2} = 0.05,\ \sigma_{P_1 - P_2} = 0.059\}$

6. (*a*) A finite population consists of five values 2, 4, 6, 8 and 10. Take all possible samples of size 2 which can be drawn with replacement from this population. Assuming the 25 possible samples equally likely, construct the sampling distributions of sample means and sample variances and find the mean of these distributions. Calculate the mean and variance of the population and verify that

(*i*) $\mu_{\bar{X}} = \mu$ (*ii*) $\mu_{S^2} = \dfrac{n-1}{n}\sigma^2$

where $\bar{X} = \dfrac{\Sigma X_i}{n}$ and $S^2 = \dfrac{\Sigma(X_i - \bar{X})^2}{n}$

$\{\mu = 6,\ \sigma^2 = 8,\ \mu_{\bar{X}} = 6,\ \mu_{S^2} = 4\}$

(*b*) A finite population consists of five values 1, 3, 5, 7 and 9. Take all possible samples of size 2 which can be drawn with replacement from this population. Assuming the 25 possible samples equally likely, construct the sampling distributions of sample means and sample variances and find the mean of these distributions. Calculate the mean and variance of the population. Discuss the results.

$\left(\mu = 5,\ \sigma^2 = 8,\ \mu_{\bar{X}} = 5,\ \mu_{S^2} = 4,\ \mu_{\bar{X}} = \mu,\ \mu_{S^2} = \dfrac{n-1}{n}\sigma^2\right)$

7. (*a*) A finite population consists of 5 values 1, 3, 5, 7 and 9. Take all possible samples of size 2 which can be drawn without replacement from this population. Assuming the 20 possible samples equally likely, construct the sampling distributions of sample means and sample variances and find the mean of these distributions. Calculate the mean

and variance of the population and verify that

(i) $\mu_{\bar{x}} = \mu$ (ii) $\mu_{S^2} = \dfrac{N}{N-1} \dfrac{n-1}{n} \sigma^2$

where $\bar{X} = \dfrac{\Sigma X_i}{n}$ and $S^2 = \dfrac{\Sigma(X_i - \bar{X})^2}{n}$

$\{\mu = 5, \sigma^2 = 8, \mu_{\bar{x}} = 5, \mu_{S^2} = 5\}$

(b) Take all possible samples of 2 distinct values from the population 2, 4, 6, 8 and 10. Assuming the 20 possible samples equally likely, construct the sampling distributions of sample means and sample variances and find the mean of these distributions. Calculate the mean and variance of the population. Discuss the results.

$\left(\mu = 6, \sigma^2 = 8, \mu_{\bar{x}} = 6, \mu_{S^2} = 5, \mu_{\bar{x}} = \mu, \mu_{S^2} = \dfrac{N}{N-1} \dfrac{n-1}{n} \sigma^2\right)$

Exercise 11.5

Objective Questions

1. Fill in the blanks.

(i) A ——— is the totality of the observations made on all the objects possessing some common specific characteristics. *(population)*

(ii) A ——— is a part of the population which is selected with the expectation that it will represent the characteristics of the population. *(sample)*

(iii) ——— is a procedure of selecting a representative sample from a given population. *(Sampling)*

(iv) The descriptive measures of a population are called ———. *(parameters)*

(v) A descriptive measure on the sample observations is called ———. *(statistic)*

(vi) A population is called ——— if it includes a limited number of sampling units. *(finite)*

(vii) A population is called ——— if it includes an unlimited number of sampling units. *(infinite)*

(viii) Sampling ——— is a complete list of the sampling units. *(frame)*

(ix) A ——— sampling is a procedure in which we cannot assign to an element of the population the probability of its being included in the sample. *(non-probability)*

(x) A ——— sampling is a process in which the sample is selected in such a way that every element of a population has a known nonzero probability of being included in the sample. *(probability)*

(xi) Another name of a probability sampling is ——— sampling. *(random)* (*xii*) Random sampling provides reliable ———. (*estimates*)

(*xiii*) The sampling is said to be ——— replacement when the unit selected at random is returned to the population before the next unit is selected. (*with*)

(*xiv*) The sampling is said to be ——— replacement when the unit selected at random is returned to the population before the next unit is selected. (*without*)

(*xv*) A sample is usually selected by ——— replacement. (*without*)

(*xvi*) In sampling ——— replacement, a sampling unit can be selected more than once. (*with*)

2. Fill in the blanks.

(*i*) In sampling ——— replacement, a sampling unit cannot be selected more than once. (*without*)

(*ii*) In sampling with replacement, a finite population becomes ———. (*infinite*)

(*iii*) ——— random sampling is a procedure of selecting a sample from the population in such a way that every unit available for sampling has an equal probability of being selected. (*simple*)

(*iv*) The sampling error decreases by increasing the sample ———. (*size*)

(*v*) The ——— errors may be present both in sample survey and census. (*non-sampling*)

(*vi*) The bias increases by increasing the sample ———. (*size*)

(*vii*) A sample which is free from bias is called an ——— sample. (*unbiased*)

(*viii*) ——— errors may arise due to faulty sampling frames, non-responses and processing of data. (*Non-sampling*)

(*ix*) ——— errors can be controlled by the proper training of the investigators and following up the non-responses (*Non-sampling*)

(*x*) The probability distribution of a sample statistic is called ——— distribution of that statistic. (*sampling*)

(*xi*) The standard deviation of sampling distribution of a sample statistic is called the ——— of that statistic. (*standard error*)

(*xii*) The standard error can be reduced by increasing the ———. (*sample size*)

(*xiii*) The number of all possible samples of size n taken with replacement from a population of size N is ———. $(N)^n$

(*xiv*) The number of all possible samples of size n taken without replacement from a population of size N is ———. $({}^{N}P_{n})$

3. Mark off the following statements as true or false.

(*i*) A descriptive measure on the sample observations is called parameter and a descriptive measure of a population is called statistic. (*false*)

(*ii*) A sample statistic is a random variable whereas the parameter being estimated is constant. (*true*) (*iii*) A sample survey provides the results which are more accurate than those obtained from a census. (*false*)

(*iv*) A sample design is a procedure for obtaining a sample from a given population prior to collecting any data. (*true*)

(*v*) More detailed information can be obtained in a sample survey as compared to a census. (*true*)

(*vi*) Sampling may be the only means available for obtaining the desired information if the population is infinite. (*true*)

(*vii*) If the data are obtained by tests that are destructive, then complete enumeration becomes essential. (*false*)

(*viii*) Every random sample is a simple random sample. (*false*)

(*ix*) In sampling with replacement, the sample size may be greater than the population size. (*true*)

(*x*) In sampling without replacement, the sample size can be greater than population size. (*false*)

(*xi*) The number of units available for the next drawing does not change in a random sampling with replacement. (*true*)

(*xii*) In sampling without replacement, the number of units remaining after each drawing will be reduced by one. (*true*)

4. Mark off the following statements as true or false.

(*i*) The number of all possible samples of size n taken without replacement from a population of size N is ${}^{N}C_{n}$. (*false*)

(*ii*) In sampling without replacement, a sampling unit can be selected more than once. (*false*)

(*iii*) In sampling with replacement, the sample size may be greater than the population size. (*true*)

(*iv*) In sampling with replacement, a finite population becomes infinite. (*true*)

(*v*) Non-sampling errors may be present both in sample survey and census. (*true*)

(*vi*) The sampling error increases by increasing the sample size. (*false*)

(*vii*) Sampling and non-sampling errors are both controllable. (*true*)

(*viii*) The standard deviation of a sampling distribution of a statistic is called the standard error of that statistic. (*true*)

(*ix*) Standard error is the difference of a statistic from the parameter being estimated. (*false*)

(*x*) We can decrease both sampling error and standard error by increasing the sample size. (*true*)

(*xi*) The reliability of an estimate can be determined by its standard error. (*true*)