14

Survey Sampling and Sampling Distributions

14.1 INTRODUCTION

Sampling is a statistical technique which is used in almost every field in order to collect information and on the basis of this information inferences about the characteristics of a population are made. The values of the population characteristics are summarized by certain numerical descriptive measures, called *parameters*. The values of the population parameters, which are in most situations unknown, would have to be estimated and to get estimates, we resort to sampling. The observations composing a sample are used to calculate a corresponding numerical descriptive measure, called a *statistic*. Thus we use statistics to estimate parameters. Considerations of time and cost are other reasons for sampling. Prior to introducing some of the most commonly used sampling methods, we proceed to some definitions and to a brief description of the basic concepts involved in sampling.

14.1.1 Statistical Populations. A *statistical population* (or *universe*) is defined as the aggregate or totality of all individual members or objects, whether animate or inanimate, concrete or abstract, of some characteristics of interest. The individual members of the population are called *sampling units* or simply *units*. A sampling unit from which information is required, may be a college student, an animal, a tree, a household, a block, a town, a small area, a field, a business firm, etc. A set of n sampling units selected from a given population is called a *sample* of size n and the process of selecting a sample, is known as *sampling*. The numerical values assigned to units of interest are treated as values of a random variable X, and the distribution of X is called the *population distribution*.

truck-load, the sampler inspects the whole [illegible] selects according to his expert subjective judgement, those melons which he considers to be representative. Purposive sampling can also be appropriate when a population contains relatively few large units whose characteristics are known. Its main use, however, is in Economic and Business statistics.

14.3.2. Quota Sample. A quota sample is a type of judgement sample. It is a sample, usually of human being, in which the information is collected purposively from the segments of a population (*the quotas*), e.g. the quotas of men and women; urban and rural; upper, middle and lower income groups; etc. These factors are termed *quota controls*. They are intended to make the sample as representative as possible and to reduce sampling bias that creeps in because the selection of respondents within the quotas depends on the personal choice of the interviewers. Interviewers being human, are likely to look for persons who either share similar opinions or are personally known to them or are conveniently located.

Quota sampling may be considered as stratified sampling in which the selection of units within strata is non-random. The advantages of quota sampling are that it is cheaper, it is easy administratively and it is a very quick form of investigation. Quota sampling is widely used in *public opinion polls* and *market research surveys*.

14.4 SAMPLING DISTRIBUTIONS

A sampling distribution is defined as a probability distribution of the values of a *statistic* such as a mean, a standard deviation, a proportion, etc, computed from all possible samples of the same size, which might be selected with or without replacement from a population. As a sampling distribution of a statistic is a probability distribution, therefore the sum of all probabilities in it is always equal to one; and the distribution has its own mean and its own standard deviation. The values of the statistic computed from *one or more* samples actually selected from the population and the sampling distribution of the statistic provide all the information one needs in making decisions about the values of the population parameters. There are many types of sampling distributions but the most frequently used types in statistical inference are the binomial, the normal, the t-distribution, the chi-square distribution, and the F distribution. A sampling distribution should not be confused with a *sample distribution* which is the distribution of individual values of a single sample.

Standard Error. The standard deviation of a sampling distribution of a sample statistic is called the *standard error* (abbreviated to *S.E.*) *of the statistic*. The standard error thus measures the dispersion of the values of a statistic, that might be computed from all possible samples, whereas the standard deviation of a population (or sample) measures the dispersion of the values of the population (sample) units about the population (sample) mean.

14.4.1. Sampling Distribution of the Mean. The sampling distribution of the mean is the probability distribution or the relative frequency distribution of the means $\bar{X}$ of all possible random samples of the same size that *could* be selected from a given population. The mean of this distribution is represented by $\mu_{\bar{x}}$ and the standard deviation, which is called the *standard error of the mean*, by $\sigma_{\bar{x}}$ or *S.E.* $(\bar{X})$. The value $\sigma_{\bar{x}}$ indicates the spread in the distribution of all possible sample means.

The sampling distribution of $\bar{X}$ has the following properties.

(i) The mean of the sampling distribution of the mean (equivalently, the mean of all possible sample means) is equal to the population mean, that is $\mu_{\bar{x}} = \mu$, regardless of whether sampling is done with replacement or without replacement.

Proof. Let us first consider sampling *without replacement* from a finite population of size N. The number of distinct simple random samples of size n that can be selected without replacement from a population of size N is $\binom{N}{n} = k$, say. Let $\bar{X}_1, \bar{X}_2, \ldots, \bar{X}_i, \ldots, \bar{X}_k$ be the means of $k = \binom{N}{n}$ possible random samples of size n, where $\bar{X}_i$ is the mean of the *ith* sample. Then the mean of the sampling distribution of $\bar{X}$ (equivalently, the mean of all possible sample means), denoted by $\mu_{\bar{x}}$, is

$$\mu_{\bar{x}} = \frac{\bar{X}_1 + \bar{X}_2 + \ldots + \bar{X}_k}{k}$$

$$= \frac{1}{k}\left[\left(\frac{X_1+X_2+\ldots}{n}\right) + \left(\frac{X_1+X_3+\ldots}{n}\right) + \ldots + \left(\frac{X_2+X_3+\ldots}{n}\right) + \ldots\right]$$

In order to simplify the expression on the right, we find out the number of samples that contain any specified value X_i. The number of such samples is ([illegible]), that is, the number of ways in which the $(n-1)$ other units in the sample are to be selected from the remaining $(N-1)$ units.

Next, we determine the co-efficient of the value X_i by collecting all the terms in the expression containing X_i. Thus the co-efficient of X_i is

$$\frac{\binom{N-1}{n-1}}{k} \cdot \frac{1}{n} = \frac{\binom{N-1}{n-1}}{\binom{N}{n}} \frac{1}{n} = \frac{(N-1)!\,(N-n)!\,n!}{(n-1)!\,(N-n)!\,N!} \cdot \frac{1}{n} = \frac{1}{N}$$

$$\text{Hence } \mu_{\bar{x}} = \frac{X_1}{N} + \frac{X_2}{N} + \dots + \frac{X_i}{N} + \dots + \frac{X_N}{N}$$

$$= \frac{X_1 + X_2 + \dots + X_i + \dots + X_N}{N}$$

$$= \mu, \text{ mean of the population.}$$

Sampling with replacement. Let $X_1, X_2, \dots, X_n$ be the observations of a simple random sample of size n from a population having N observations. Then a specified X_i taken from the population, could be any one of the N values with an equal probability of $\frac{1}{N}$ as all the values are equally likely. Thus X_i is a random variable and therefore

$$E(X_i) = \frac{1}{N} \sum_{k=1}^{N} X_k = \mu.$$

For repeated sampling, the mean of a sample $\bar{X} = \frac{1}{n} \sum X_i$ varies from sample to sample, therefore

$$E(\bar{X}) = E\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right]$$

$$= \frac{1}{n}\left[E(X_1) + E(X_2) + \dots + E(X_n)\right]$$

$$= \frac{1}{n}\left[\mu + \mu + \dots + \mu\right]$$

$$= \frac{1}{n}\left[n\mu\right] = \mu, \text{ the population mean.}$$

(ii) The standard deviation of the sampling distribution of the mean is given by

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}}, \quad (\sigma = \text{population } s.d.)$$

when sampling is performed without replacement from a finite population of size N, *or*

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

when sampling is done with replacement from a finite population or sampling from an infinite population.

Proof. The variance of $\bar{X}$, denoted by $\sigma_{\bar{x}}^2$ or $\mathrm{Var}(\bar{X})$ is defined as

$$\begin{aligned}
\mathrm{Var}(\bar{X}) &= E\,[\bar{X} - E(\bar{X})]^2 = E\,[\bar{X} - \mu]^2 \\
&= E\left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)\right]^2 \\
&= \frac{1}{n^2}\cdot E\left[\sum_{i=1}^{n}(X_i - \mu)^2 + \sum_{i \neq j}^{n}(X_i - \mu)(X_j - \mu)\right] \\
&= \frac{1}{n^2}E\left[\sum_{i=1}^{n}(X_i - \mu)^2\right] + \frac{1}{n^2}E\left[\sum_{i \neq j}^{n}(X_i - \mu)(X_j - \mu)\right]
\end{aligned}$$

The simplification depends on whether the sampling is performed without replacement from a finite population of size N or sampling is done with replacement. The two cases are treated separately.

First case: *Sampling without replacement.*

Since the probability of obtaining $(X_i - \mu)^2$ on the *ith* draw is equal to the probability of obtaining X_i on the *ith* draw which is $\frac{1}{N}$, therefore the expected value of $(X_i - \mu)^2$ becomes σ^2, *i.e.*

$$E(X_i - \mu)^2 = \sum_{i=1}^{N}\frac{1}{N}(X_i - \mu)^2 = \sigma^2.$$

Again, since the sampling is without replacement, the probability of selecting $(X_i - \mu)(X_j - \mu)$ on the *ith* and *jth* draw is $\frac{1}{N}\cdot\frac{1}{N-1}$, because they are not independent on account of the reduction in size from N to $N-1$. Thus

$$E(X_i-\mu)(X_j-\mu) = \frac{1}{N}\cdot\frac{1}{N-1}\sum_{i\neq j}^{N}(X_i-\mu)(X_j-\mu)$$

$$= \frac{1}{N(N-1)}\left\{\left[\sum_{i=1}^{N}(X_i-\mu)\right]^2 - \sum_{i=1}^{N}(X_i-\mu)^2\right\}$$

$$= \frac{1}{N(N-1)}\left\{0 - \sum_{i=1}^{N}(X_i-\mu)^2\right\}$$

$$= \frac{-\sigma^2}{N-1}.$$

Substituting these values, we get

$$\mathrm{Var}(\bar{X}) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 + \frac{1}{n^2}\sum_{i\neq j}^{n}\frac{-\sigma^2}{N-1}$$

$$= \frac{1}{n^2}(n\sigma^2) - \frac{1}{n^2}\cdot\frac{1}{N-1}\sum_{i\neq j}^{n}\sigma^2$$

$$= \frac{\sigma^2}{n} - \frac{1}{n^2}\,\frac{1}{N-1}\cdot n(n-1)\,\sigma^2$$

$$= \frac{\sigma^2}{n}\cdot\frac{N-n}{N-1}.$$

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

The factor $\frac{N-n}{N-1}$ is usually called the *finite population correction* (*fpc*) or *finite correction factor* (*fcf*) for the variance because in sampling from finite population, the variance of the mean is reduced by this amount. It is important to note that in sampling without replacement from a finite population of size N, *fpc* is dropped from the formula whenever n, the sample size, is less than 5% of N; and *fpc* is used when n is 5% or greater than 5% of N.

Second case : *Sampling with replacement*

When sampling is done with replacement or sampling from an finite population, the X_i and X_j are statistically independent. Therefore

$$E(X_i - \mu)(X_j - \mu) = 0. \text{ Hence we get}$$

$$\text{Var}(\bar{X}) = \frac{1}{n^2}(n\sigma^2) = \frac{\sigma^2}{n},$$

or
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\text{population standard deviation}}{\text{square root of sample size}}.$$

It is to be noted that the standard error of the mean is always less than the standard deviation of the population. This means that the sampling distribution of the mean has less variability than the population from which the samples were taken. If the value of σ is not known and if the sample size is *large* (as a rule of thumb adopted by many authors, a sample containing 30 or more observations constitutes a *large* or *sufficiently large* sample), it is replaced by s, the standard deviation of the sample. The S.*E*. of the mean then becomes

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}.$$

(iii) **Shape of the distribution.** (a) If the population sampled is normally distributed, then the sampling distribution of the mean $\bar{X}$, will also be normal regardless of sample size.

To prove this, we proceed as follows:

By definition, the moment generating function of $\bar{X}$ is

$$M_{\bar{X}}(t) = E(e^{t\bar{X}}) = E(e^{\Sigma tX_i/n})$$

$$= E\left[\prod_{i=1}^{n} e^{tX_i/n}\right] = \prod_{i=1}^{n} E(e^{tX_i/n}).$$

But $\quad E(e^{tX_i/n}) = M_X\left(\frac{t}{n}\right).$

If X is $N(\mu, \sigma^2)$, then

$$M_X(t) = e^{\mu t + \sigma^2 t^2/2}, \text{ and}$$

$$M_X\left(\frac{t}{n}\right) = e^{\mu t/n + \frac{1}{2}\sigma^2(t/n)^2}.$$

Since $X_1, X_2, \ldots, X_n$ is a random sample, therefore

$$M_{\bar{X}}(t) = E\left(e^{t\bar{X}}\right) = \left(E\left(e^{tX_i/n}\right)\right)^n$$

$$= \left[M_X\left(\frac{t}{n}\right)\right]^n = e^{n\left(\frac{\mu t}{n} + \frac{1}{2}\sigma^2 t^2/n^2\right)}$$

$$= e^{\mu t + \sigma^2 t^2/2n}$$

But this is the *m.g.f.* of a normal distribution with mean = μ and variance = $\frac{\sigma^2}{n}$. Thus $\bar{X}$ is normally distributed variable with mean μ and variance σ^2/n where μ and σ^2 are the mean and variance of the population.

(b) If the population sampled is non-normal, then for sufficiently large sample size, the sampling distribution of $\bar{X}$ will approximate the normal distribution.

This is a special case of the most important statistical theorem, known as the *Central Limit Theorem*, which is stated and proved in the next section.

We know that the standardized form of a random variable is obtained by subtracting its mean from it and dividing the difference by its standard deviation, that is

$$Z = \frac{\text{value of random variable} - \text{mean of random variable}}{\text{standard deviation of random variable}}.$$

We have proved above that the sample mean $\bar{X}$ is normally distributed random variable with mean equal to population mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$. The standard normal variable then becomes

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}.$$

If sampling is without replacement and sample size n is 5% or greater than 5 per cent of the population size N, then Z values are obtained by the formula

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}}$$

The sampling distribution of $\bar{X}$ thus offers solutions to probability questions about the values of the sample means.

Example 14.7. Assume that a population consists of 7 similar containers having the following weights (kilograms):

9.8, 10.2, 10.4, 9.8, 10.0, 10.2 and 9.6.

(a) Find the mean μ and the standard deviation σ of the given population.

(b) Draw random samples of 2 containers without replacement and calculate the mean weight $\bar{X}$ of each sample.

(c) Form a frequency distribution of $\bar{X}$ and a sampling distribution of $\bar{X}$.

(d) Find the mean and the standard deviation of the sampling distribution of $\bar{X}$.

(a) The population mean μ and standard deviation σ are

$$\mu = \frac{\sum X}{N} = \frac{9.8 + 10.2 + \ldots + 9.6}{7} = \frac{70.0}{7} = 10.0 \text{ kg; and}$$

$$\sigma = \sqrt{\frac{\sum(X-\mu)^2}{N}} = \sqrt{\frac{(9.8-10)^2+(10.2-10)^2+\ldots+(9.6-10)^2}{7}}$$

$$= \sqrt{\frac{0.48}{7}} = \sqrt{0.0686} = 0.262 \text{ kg.}$$

(b) Let the containers be identified as A, B, C, D, E, F and G. Now the number of possible random samples of $n = 2$ containers without replacement is $\binom{7}{2} = 21$. The 21 possible random samples with the values of their mean weights are given on page 32:

$\frac{7!}{(7-2)!}$

(c) The frequency distribution of $\bar{X}$ and the sampling distribution of the mean $\bar{X}$, which is just the relative frequency distribution of $\bar{X}$, are obtained below:

(i) Frequency Distribution of $\bar{X}$

Sample Mean $\bar{x}$	Tally	f
9.7	II	2
9.8	II	2
9.9	IIII	4
10.0	~~IIII~~	5
10.1	IIII	4
10.2	II	2
10.3	II	2
Σ	--	21

(ii) Sampling Distribution of $\bar{X}$

Sample Mean $\bar{x}$	Probability $f(\bar{x})$
9.7	2/21
9.8	2/21
9.9	4/21
10.0	5/21
10.1	4/21
10.2	2/21
10.3	2/21
Σ	1

Sample No.	Sample Combination	Weights in Samples X_1	X_2	Sample Mean weight ($\bar{X}$)
1	*A, B*	9.8,	10.2	10.0
2	*A, C*	9.8,	10.4	10.1
3	*A, D*	9.8,	9.8	9.8
4	*A, E*	9.8,	10.0	9.9
5	*A, F*	9.8,	10.2	10.0
6	*A, G*	9.8,	9.6	9.7
7	*B, C*	10.2,	10.4	10.3
8	*B, D*	10.2,	9.8	10.0
9	*B, E*	10.2,	10.0	10.1
10	*B, F*	10.2,	10.2	10.2
11	*B, G*	10.2,	9.6	9.9
12	*C, D*	10.4,	9.8	10.1
13	*C, E*	10.4,	10.0	10.2
14	*C, F*	10.4,	10.2	10.3
15	*C, G*	10.4,	9.6	10.0
16	*D, E*	9.8,	10.0	9.9
17	*D, F*	9.8,	10.2	10.0
18	*D, G*	9.8,	9.6	9.7
19	*E, F*	10.0,	10.2	10.1
20	*E, G*	10.0,	9.6	9.8
21	*F, G*	10.2,	9.6	9.9

(d) The mean and standard deviation of sampling distribution of $\bar{X}$, are computed below:

Sample Mean $\bar{x}$	Probability $f(\bar{x})$	$\bar{x}\,f(\bar{x})$	$\bar{x}-\mu_{\bar{x}}$	$(\bar{x}-\mu_{\bar{x}})^2$	$(\bar{x}-\mu_{\bar{x}})^2\,f(\bar{x})$
9.7	2/21	19.4/21	−0.3	0.09	0.18/21
9.8	2/21	19.6/21	−0.2	0.04	0.08/21
9.9	4/21	39.6/21	−0.1	0.01	0.04/21
10.0	5/21	50.0/21	0	0	0
10.1	4/21	40.4/21	0.1	0.01	0.04/21
10.2	2/21	20.4/21	0.2	0.04	0.08/21
10.3	2/21	20.6/21	0.3	0.09	0.18/21
Σ	1	10.0	--	--	0.6/21

$$\mu_{\bar{x}} = \sum \bar{x}\, f(\bar{x}) = 10.0\ kg, \text{ and}$$

$$\sigma_{\bar{x}} = \sqrt{\sum (\bar{x} - \mu_{\bar{x}})^2 f(\bar{x})} = \sqrt{\frac{0.6}{21}} = \sqrt{0.0286} = 0.17\ kg,$$

which is a smaller value indicating that the sampling distribution of the mean is more concentrated about the population mean.

Example 14.8. A sample of size $n=3$ is to be randomly selected without replacement from a population that has $N=5$ items whose values are 0, 3, 6, 9 and 12.

(a) Find the sampling distribution of the sample mean, $\bar{X}$.

(b) Calculate the mean and the standard deviation of $\bar{X}$, and verify that

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}.$$

Let the items be designated by the letters A, B, C, D and E.

(a) The number of samples of size $n=3$ that could be drawn without replacement from a population of size $N=5$ is

$$\binom{5}{3} = \frac{5!}{2!\,3!} = 10.$$

The 10 possible samples and their means are given below:

Sample No.	Sample Combinations	Sample Values	Sample Mean ($\bar{X}$)
1	*A, B, C*	0, 3, 6	3
2	*A, B, D*	0, 3, 9	4
3	*A, B, E*	0, 3, 12	5
4	*A, C, D*	0, 6, 9	5
5	*A, C, E*	0, 6, 12	6
6	*A, D, E*	0, 9, 12	7
7	*B, C, D*	3, 6, 9	6
8	*B, C, E*	3, 6, 12	7
9	*B, D, E*	3, 9, 12	8
10	*C, D, E*	6, 9, 12	9

The sampling distribution is obtained by listing all possible means and their probabilities (relative frequencies) as below:

Sampling Distribution of $\bar{X}$

Sample Mean $\bar{X}$	Number of sample means (f)	Probability $f(\bar{x})$
3	1	1/10
4	1	1/10
5	2	2/10
6	2	2/10
7	2	2/10
8	1	1/10
9	1	1/10
Σ	10	1

(b) Next, we calculate the mean and the standard deviation (the *standard error*) of the sampling distribution of the mean as follows:

Calculation of Mean and S.D. of Sampling Distribution of $\bar{X}$.

Sample Mean $\bar{x}$	Probability $f(\bar{x})$	$\bar{x}\,f(\bar{x})$	$\bar{x}^2\,f(\bar{x})$
3	1/10	3/10	9/10
4	1/10	4/10	16/10
5	2/10	10/10	50/10
6	2/10	12/10	72/10
7	2/10	14/10	98/10
8	1/10	8/10	64/10
9	1/10	9/10	81/10
Σ	1	60/10	390/10

Now $\mu_{\bar{x}} = \sum \bar{x}\,f(\bar{x}) = \frac{60}{10} = 6$, and

$$\sigma_{\bar{x}} = \sqrt{[\sum \bar{x}^2 f(\bar{x})] - [\sum \bar{x} f(\bar{x})]^2}$$

$$= \sqrt{\frac{390}{10} - \left(\frac{60}{10}\right)^2} = \sqrt{39 - 36} = \sqrt{3} = 1.732$$

In order to verify the given result, we first calculate the mean μ and the variance σ^2 of the given population. Thus

$$\mu = \frac{1}{5}[0 + 3 + 6 + 9 + 12] = \frac{1}{5}[30] = 6, \text{ and}$$

$$\sigma^2 = \frac{1}{5}[(0-6)^2+(3-6)^2+(6-6)^2+(9-6)^2+(12-6)^2] = 18$$

Verification: Now $\frac{\sigma^2}{n} \cdot \frac{N-n}{N-1} = \frac{18}{3} \cdot \frac{5-3}{5-1} = \frac{18\times 2}{3\times 4} = 3 = \sigma^2_{\bar{x}}$

Hence the result.

Example 14.9. Suppose that a random variable X has the following population distribution:

x	3	6	9
$f(x)$	1/3	1/3	1/3

If a sample of three numbers is taken with replacement, obtain the sampling distribution of the sample mean and verify that $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

opulation distributi... of the r.v.X may be written as

$$\cdot) = P(X=x) = \begin{cases} 1/3 & \text{for } x = 3 \\ 1/3 & \text{for } x = 6 \\ 1/3 & \text{for } x = 9 \end{cases}$$

ing that the members of the population have the numerical 3, 6 and 9.

umber of possible samples of size $n=3$, which could be selected cement from this population is $3^3 = 27$. The 27 random samples means are given below:

Sample No.	Sample Values X_1	X_2	X_3	Sample Mean $\bar{X}$
1	3	3	3	3
2	3	3	6	4
3	3	3	9	5
4	3	6	3	4
5	3	6	6	5
6	3	6	9	6
7	3	9	3	5
8	3	9	6	6
9	3	9	9	7
10	6	3	3	4
11	6	3	6	5
12	6	3	9	6
13	6	6	3	5
14	6	6	6	6
15	6	6	9	7
16	6	9	3	6
17	6	9	6	7
18	6	9	9	8
19	9	3	3	5
20	9	3	6	6
21	9	3	9	7
22	9	6	3	6
23	9	6	6	7
24	9	6	9	8
25	9	9	3	7
26	9	9	6	8
27	9	9	9	9

The sampling distribution of the sample mean $\bar{X}$ is obtained below, together with two columns needed for the calculation of the S.E. of the mean:

Sampling Distribution of $\bar{X}$ and Calculation of S.E. ($\bar{X}$)

$\bar{X}$	No. of sample means	Probability $f(\bar{x})$	$\bar{x}\,f(\bar{x})$	$\bar{x}^2 f(\bar{x})$
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
Σ	27	1	162/27	1026/27

Now, $\mu_{\bar{x}} = \sum \bar{x}\, f(\bar{x}) = \dfrac{162}{27} = 6$

$$\sigma^2_{\bar{x}} = \sum \bar{x}^2 f(\bar{x}) - [\sum \bar{x}\, f(\bar{x})]^2$$

$$= \frac{1026}{27} - (6)^2 = 38 - 36 = 2.$$

And, $\mu = \sum x\, f(x) = 3 \times \frac{1}{3} + 6 \times \frac{1}{3} + 9 \times \frac{1}{3} = 6,$

$$\sigma^2 = \sum x^2 f(x) - \mu^2 = [9 \times \tfrac{1}{3} + 36 \times \tfrac{1}{3} + 81 \times \tfrac{1}{3}] - (6)^2$$

$$= (3 + 12 + 27) - 36 = 6.$$

Verification: $\dfrac{\sigma^2}{n} = \dfrac{6}{3} = 2 = \sigma^2_{\bar{x}}$

Hence, $\sigma_{\bar{x}} = \dfrac{\sigma}{\sqrt{n}}$.

Example 14.10. The weights of 1500 ball bearings are normally distributed with a mean of 22.40 ounces and a standard deviation of 0.048 ounces. If 300 random samples of size 36 are drawn from this population, (a) determine the expected mean and standard deviation of the sampling distribution of mean if sampling is done (i) with replacement, (ii) without replacement; (P.U. B.A./B.Sc., 1971)

(b) How many of the random samples would have their means between 22.39 and 22.42 oz?

(a) There would be $(1500)^{36}$ and $\binom{1500}{36}$ possible samples of size 36, that could be obtained theoretically from a population of weights of 1500 ball bearings with and without replacement respectively. Obviously the number of theoretically possible samples is much larger than 300. Therefore the sampling distribution of the mean will not be a true sampling distribution. (Such a distribution is called *experimental* sampling distribution). But 300 being a large number, there should be a close agreement between the experimental sampling distribution of 300 sample means and the true sampling distribution of mean. Hence the expected mean and standard deviation are found to be as:

(i) *Sampling with replacement:*

$$\mu_{\bar{x}} = \mu = 22.40\ oz.,$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.048}{\sqrt{36}} = 0.008\ oz.$$

(ii) *Sampling without replacement:*

$$\mu_{\bar{x}} = \mu = 22.40\ oz.,$$

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

Since sample size $n=36$ is less than 5% of the population size $N=1500$, therefore according to the generally accepted rule for the use of *fpc*, the factor $\sqrt{\frac{N-n}{N-1}}$ is dropped. Thus

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{0.048}{\sqrt{36}} = 0.008\ oz.$$

(b) The sampling distribution of the mean $\bar{X}$ is normal because the population sampled is normally distributed. Thus

$$Z = \frac{\bar{X} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - 22.40}{0.008}$$ is a standard normal variable.

To find the expected number of samples that would have their means between 22.39 and 22.42 oz, we will transform these values to the corresponding z-values. Thus

at $\bar{x} = 22.39$, we find $z_1 = \dfrac{22.39 - 22.40}{0.008} = -1.25$, and

at $\bar{x} = 22.42$, we find $z_2 = \dfrac{22.42 - 22.40}{0.008} = 2.50.$

Using the Table of areas under the normal curve, we find

$$P(22.39 \le \bar{X} \le 22.42) = P(-1.25 \le Z \le 2.50)$$
$$= P(-1.25 \le Z \le 0) + P(0 \le Z \le 2.50)$$
$$= 0.3944 + 0.4938 = 0.8882.$$

Hence the expected number of samples = (300) (0.8882) = 267.

Example 14.11. A construction company has 310 employees who have an average annual salary of Rs. 24,000. The standard deviation of annual salaries is Rs. 5,000. In a random sample of 100 employees, what is the probability that the average salary will exceed Rs. 24,500?

The sample size ($n = 100$) is large enough to assume that the sampling distribution of $\bar{X}$ is approximately normally distributed with mean

$$\mu_{\bar{x}} = \mu = \text{Rs. } 24{,}000.$$

and standard deviation

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-1}} = \frac{5000}{\sqrt{100}} \sqrt{\frac{310-100}{310-1}}$$
$$= \text{Rs. } 412.20,$$

where we have used *fpc*, because the sample size $n = 100$ is greater than 5 per cent of the population size $N = 310$.

Equivalently, $Z = \dfrac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \dfrac{\bar{X} - 24000}{412.20}$ is approximately $N(0, 1)$.

We are required to evaluate $P(\bar{X} > 24{,}500)$.

At $\bar{x} = 24{,}500$, we find that $z = \dfrac{24500 - 24000}{412.20} = 1.21$

Hence using *Table* of areas under normal curve, we get

$$P(\bar{X} > 24{,}500) = P(Z > 1.21)$$
$$= 0.5 - P(0 \le Z \le 1.21)$$
$$= 0.5 - 0.3869 = 0.1131.$$

Example 14.12. Calculate the standard error of the mean from the following data collected in one of the many random sample inquiries to find average earning of a particular class.

Earning (Rs.)	1–10	11–20	21–30	31–40	41–50	51–60	61–70	71–80
Number of persons	20	98	150	218	200	164	110	40

Since the population standard deviation σ is not known and the sample size ($n = 1000$) is large enough to replace it with the sample standard deviation s, we therefore first calculate the sample standard deviation as below:

Earning (Rs.)	f	u	fu	fu^2
1–10	20	−3	−60	180
11–20	98	−2	−196	392
21–30	150	−1	−150	150
31–40	218	0	0	0
41–50	200	1	200	200
51–60	164	2	328	656
61–70	110	3	330	990
71–80	40	4	160	640
Σ	1000	--	612	3208

The sample mean and the sample standard deviation are;

$$\bar{x} = a + \frac{\Sigma fu}{n} \times h = 35.5 + \frac{612}{1000} \times 10 = \text{Rs. } 41.62, \text{ and}$$

$$s = h\sqrt{\frac{\Sigma fu^2}{n} - \left(\frac{\Sigma fu}{n}\right)^2}$$

$$= 10\sqrt{\frac{3208}{1000} - \left(\frac{612}{1000}\right)^2} = 10\sqrt{2.833456} = \text{Rs. } 16.83.$$

Hence the standard error of the sample mean is

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} = \frac{16.83}{\sqrt{1000}} = \frac{16.83}{31.62} = \text{Rs. } 0.53.$$

14.4.2 Central Limit Theorem. The *central limit theorem* is perhaps the most important theorem in all of statistical inference. It is concerned with the means of large samples and provides solutions when the shape of the population distribution is unknown or highly skewed.

The theorem states that "If a variable X from a population has mean μ and *finite* variance σ^2, then the sampling distribution of the sample mean $\bar{X}$ approaches a normal distribution with mean μ and variance $\frac{\sigma^2}{n}$ as the sample size n approaches infinity."

To prove this theorem, we use the moment generating function about the mean. By definition, we have

$$M_x(t) = E\left[e^{(X-\mu)t}\right]$$

$$= 1 + \mu_1 t + \mu_2 \frac{t^2}{2!} + \mu_3 \frac{t^3}{3!} + \ldots$$

$$= 1 + 0 + \frac{\sigma^2 t^2}{2!} + \frac{\mu_3 t^3}{3!} + \ldots$$

Let us define a random variable Y as $Y = \frac{X-\mu}{\sigma\sqrt{n}}$.

Then the *m.g.f.* of Y is

$$M_Y(t) = E\left[e^{t(X-\mu)/\sigma\sqrt{n}}\right] = M_x(t/\sigma\sqrt{n})$$

$$= 1 + \frac{\sigma^2}{2!}\left(\frac{t}{\sigma\sqrt{n}}\right)^2 + \frac{\mu_3}{3!}\left(\frac{t}{\sigma\sqrt{n}}\right)^3 + \frac{\mu_4}{4!}\left(\frac{t}{\sigma\sqrt{n}}\right)^4 + \ldots$$

[Replacing t by $\frac{t}{\sigma\sqrt{n}}$ in $M_x(t)$]

$$= 1 + \frac{t^2}{2n} + \frac{\mu_3 t^3}{3!\,\sigma^3 n\sqrt{n}} + \frac{\mu_4 t^4}{4!\,\sigma^4 n^2} + \ldots$$

Let us define another variable Z as a linear function of $\bar{X}$ as

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{1}{n}\sum_{i=1}^{n}\frac{X_i-\mu}{\sigma/\sqrt{n}} = \sum_{i=1}^{n}\frac{X_i-\mu}{\sigma\sqrt{n}} = \sum_{i=1}^{n} Y_i$$

We know that the *m.g.f.* of a sum of identically distributed random variables is the *nth* power of their own *m.g.f.* Thus

$$M_z(t) = [M_y(t)]^n$$

$$= \left[1 + \frac{t^2}{2n} + \frac{\mu_3 t^3}{3!\,\sigma^3 n\sqrt{n}} + \ldots\right]^n = e^{t^2/2} \text{ as } n \to \infty.$$

But this is the *m.g.f.* of a normal random variable with zero mean and unit variance. Hence Z has in the limit a *standard normal distribution.*

$$\text{Now} \quad \bar{X} = \mu + \frac{Z\sigma}{\sqrt{n}}.$$

Since a linear function (here $\bar{X}$) of a normal random variable (Z) is a normal random variable, therefore $\bar{X}$ is in the limit normally distributed with mean μ and variance σ^2/n.

It is interesting to note that we have neither assumed that the distribution of X is continuous, nor we have said anything about the shape of the distribution of X, whereas the limiting distribution of $\bar{X}$ is continuous and normal. Thus the distribution of the sample mean regardless of the shape of the population distribution but having a finite variance, is approximately normal with mean μ and variance σ^2/n.

Example 14.13 Given the population 1, 1, 1, 3, 4, 5, 6, 6, 6 and 7.

(a) Find the probability that a random sample of size 36 selected with replacement will yield a sample mean between 3.26 and 4.74.

(b) Find the mean and standard deviation for the sampling distribution of means for a sample of size 4 selected at random without replacement. Between what two values would you expect at least $\frac{3}{4}$ of the sample means to fall? (P.U., B.A./B.Sc. 1986)

The mean and standard deviation of the population are:

$$\mu = \frac{\Sigma X}{N} = \frac{40}{10} = 4, \text{ and}$$

$$\sigma = \sqrt{\frac{\Sigma(X-\mu)^2}{N}} = \sqrt{\frac{\Sigma X^2}{N} - \left(\frac{\Sigma X}{N}\right)^2}$$

$$= \sqrt{\frac{210}{10} - \left(\frac{40}{10}\right)^2} = \sqrt{5} = 2.236$$

To calculate mean and standard deviation, we may describe the population by the following probability distribution:

x	1	3	4	5	6	7
$P(X=x)$	3/10	1/10	1/10	1/10	3/10	1/10

(a) As the sampling is performed with replacement, therefore a sample of any size can be selected. A sample of size $n = 36$ is large enough for the *central limit theorem* to apply. The sampling distribution of $\bar{X}$ is therefore approximately normal with mean $\mu_{\bar{X}} = \mu = 4$ and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{2.236}{6} = 0.373$, that is

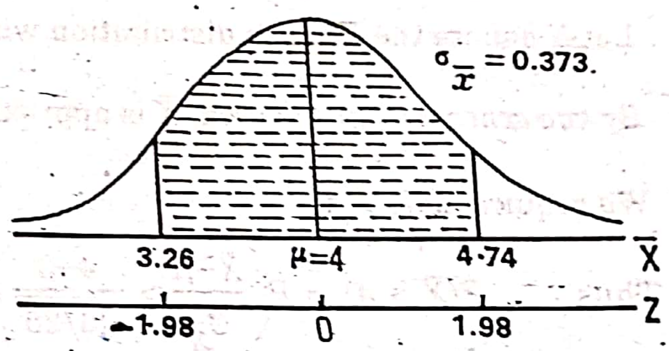
$$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} = \frac{\bar{X} - 4}{0.373} \text{ is approximately } N(0, 1).$$

To find the probability that the mean of a random sample of size $n = 36$ will fall between 3.26 and 4.74, we transform 3.26 and 4.74 to z values. Thus at $\bar{x} = 3.26$, we find

$$z = \frac{3.26-4}{0.373} = -1.98,$$

and at $\bar{x} = 4.74$, we find

$$z = \frac{4.74-4}{0.373} = 1.98.$$



Hence using Table of areas under normal curve, we find

$$P(3.26 \le \bar{X} \le 4.74) = P(-1.98 \le Z \le 1.98)$$
$$= P(-1.98 \le Z \le 0) + P(0 \le Z \le 1.98)$$
$$= 0.4762 + 0.4762 = 0.9524.$$

(b) As the sample is without replacement and sample size $n = 4$ is greater than 5% of the population size $N = 10$, therefore the mean and standard deviation of the sampling distribution of $\bar{X}$, are

$$\mu_{\bar{X}} = \mu = 4, \text{ and}$$

$$\sigma_{\bar{X}} = \frac{4}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

$$= \frac{2.236}{\sqrt{4}} \sqrt{\frac{10-4}{10-1}} = (1.118)\,(0.816) = 0.912$$

The *Chebyshev's inequality* says "*at least* $\left(1 - \frac{1}{k^2}\right)$ fraction of the data lies in the interval mean $\pm\, k(s.d.)$" and the problem says "*at least* $\frac{3}{4}$

of the sample means should fall in the same interval," so $\frac{3}{4}$ is $1 - \frac{1}{k^2}$

that is

$$1 - \frac{1}{k^2} = \frac{3}{4} \text{ or } \frac{1}{k^2} = \frac{1}{4}$$

or $\quad k^2 = 4$ or $k = 2$.

Hence we would expect *at least* $\frac{3}{4}$ of the sample means to fall in the interval $\mu_{\bar{X}} \pm 2\sigma_{\bar{X}}$, that is between $4-2(0.912)$ and $4+2(0.912)$ or between 2.2 and 5.8.

Example 14.14 A random sample of size 25 is selected from a Poisson distribution with $\mu = 3$. Find, using the *central limit theorem*, the probability that the sample mean will be greater than 4.

Let X denote the Poisson distribution with $\mu = 3$. Then $Var(X) = 3$.

By the *central limit theorem*, $\bar{X}$ is approximately $N\left(3, \frac{3}{25}\right)$,

We require $P(\bar{X} > 4)$

Thus $\quad P(\bar{X} > 4) = P\left(\frac{\bar{X}-\mu}{\sigma_{\bar{X}}} > \frac{4-3}{\sqrt{3/25}}\right)$

$$= P(Z > 2.89) = 0.0019$$

14.4.3. Sampling Distribution of Differences between Means. Suppose we have two large or infinite populations with means μ_1 and μ_2, and variances σ_1^2 and σ_2^2 respectively. Let independent random samples of sizes n_1 and n_2 be selected from the respective populations, and the differences $\bar{x}_1 - \bar{x}_2$ between the means of all possible pairs of samples be computed. Then a probability distribution of the differences $\bar{X}_1 - \bar{X}_2$ can be obtained. Such a distribution is called the *sampling distribution of the differences of sample mean* $\bar{X}_1 - \bar{X}_2$. The sampling distribution of the differences $\bar{X}_1 - \bar{X}_2$ has the following properties:

(i) 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 population means, that is

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

(ii) The standard deviation of the sampling distribution of $\bar{X}_1-\bar{X}_2$ (*standard error of* $\bar{X}_1-\bar{X}_2$), denoted by $\sigma_{\bar{X}_1-\bar{X}_2}$, is given by

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}. \quad \left[\because \text{Var}(\bar{X}_1-\bar{X}_2) = \text{Var}(\bar{X}_1)+\text{Var}(\bar{X}_2) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}\right] \text{ (Samples are independent)}$$

This expression for the *S.E.* of $\bar{X}_1-\bar{X}_2$ also holds for finite populations when sampling is performed with replacement. When population standard deviations are equal or both the samples come from the same population, the expression for the *S.E.* becomes

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

$$= \sqrt{\frac{2\sigma^2}{n}}, \text{ when } n_1 = n_2 = n.$$

If the values of σ_1 and σ_2 are not known and if both sample sizes are large, they are replaced by S_1 and S_2, the standard deviations of the respective samples. The *S.E.* becomes

$$S_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}.$$

If, on the other hand, the populations are finite, sampling is done without replacement and the sample sizes are larger than 5 per cent of the population sizes, then the *S.E.* is

$$\sigma_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sigma_1^2}{n_1}\cdot\frac{N_1-n_1}{N_1-1} + \frac{\sigma_2^2}{n_2}\cdot\frac{N_2-n_2}{N_2-1}}.$$

(iii) *Shape of the distribution.* If the populations are normally distributed, the sampling distribution of $\bar{X}_1-\bar{X}_2$, regardless of sample sizes, will be normal with mean $\mu_1 - \mu_2$ and variance $\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$. In other words, the variable

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

is normally distributed with zero mean and unit variance. If the populations are non-normal and if both sample sizes are large, (≥ 30), then the sampling distribution of differences between means is approximately a normal by the central limit theorem.

Example 14.15. Draw all possible random samples of size $n_1=2$ with replacement from a finite population consisting of 4, 6, 8. Similarly draw all possible random samples of size $n=2$ with replacement from another finite population consisting of 1, 2, 3.

(a) Find the possible differences between the sample means of the two populations.

(b) Construct the sampling distribution of $\bar{X}_1-\bar{X}_2$ and compute its mean and variance.

(c) Verify that $\mu_{\bar{x}_1-\bar{x}_2} = \mu_1-\mu_2$ and $\sigma^2_{\bar{x}_1-\bar{x}_2} = \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_1}$.

There are $(3)^2 = 9$ possible samples which can be drawn with replacement from each population. These two sets of samples and their means are given below:

From population 1			From population 2		
Sample No.	Sample values	$\bar{x}_1$	Sample No.	Sample values	$\bar{x}_2$
1	4,4	4	1	1,1	1.0
2	4,6	5	2	1,2	1.5
3	4,8	6	3	1,3	2.0
4	6,4	5	4	2,1	1.5
5	6,6	6	5	2,2	2.0
6	6,8	7	6	2,3	2.5
7	8,4	6	7	3,1	2.0
8	8,6	7	8	3,2	2.5
9	8,8	8	9	3,3	3.0

(a) The 81 possible differences $\bar{x}_1 - \bar{x}_2$ are presented in the following table.

Differences of Independent Means

$\bar{x}_2$	$\bar{x}_1$								
	4	5	6	5	6	7	6	7	8
1.0	3.0	4.0	5.0	4.0	5.0	6.0	5.0	6.0	7.0
1.5	2.5	3.5	4.5	3.5	4.5	5.5	4.5	5.5	6.5
2.0	2.0	3.0	4.0	3.0	4.0	5.0	4.0	5.0	6.0
1.5	2.5	3.5	4.5	3.5	4.5	5.5	4.5	5.5	6.5
2.0	2.0	3.0	4.0	3.0	4.0	5.0	4.0	5.0	6.0
2.5	1.5	2.5	3.5	2.5	3.5	4.5	3.5	4.5	5.5
2.0	2.0	3.0	4.0	3.0	4.0	5.0	4.0	5.0	6.0
2.5	1.5	2.5	3.5	2.5	3.5	4.5	3.5	4.5	5.5
3.0	1.0	2.0	3.0	2.0	3.0	4.0	3.0	4.0	5.0

(b) The sampling distribution of $\bar{X}_1-\bar{X}_2$ (*i.e.*, the relative frequency distribution of the possible differences $\bar{x}_1-\bar{x}_2$) is constructed below and the mean and variance of this distribution are also computed below.

$\bar{x}_1-\bar{x}_2$ $(=d)$	Tally	f	Probability $f(\bar{x}_1-\bar{x}_2)$	$df(d)$	$d^2 f(d)$
1.0	I	1	1/81	1/81	1.0/81
1.5	II	2	2/81	3/81	4.5/81
2.0	卌	5	5/81	10/81	20.0/81
2.5	卌 I	6	6/81	15/81	37.5/81
3.0	卌 卌	10	10/81	30/81	90.0/81
3.5	卌 卌	10	10/81	35/81	122.5/81
4.0	卌 卌 III	13	13/81	52/81	208.0/81
4.5	卌 卌	10	10/81	45/81	202.5/81
5.0	卌 卌	10	10/81	50/81	250.0/81
5.5	卌 I	6	6/81	33/81	181.5/81
6.0	卌	5	5/81	30/81	180.0/81
6.5	II	2	2/81	13/81	84.5/81
7.0	I	1	1/81	7/81	49.0/81
Total	---	81	1	324/81	1431/81

Thus the mean and the variance are

$$\mu_{\bar{x}_1-\bar{x}_2} = \sum (\bar{x}_1 - \bar{x}_2)\, f(\bar{x}_1 - \bar{x}_2)$$

$$= \sum d f(d) = \frac{324}{81} = 4, \text{ and}$$

$$\sigma^2_{\bar{x}_1-\bar{x}_2} = \sum (d - \mu_{\bar{x}_1-\bar{x}_2})^2 f(d) = \sum d^2 f(d) - [\sum d f(d)]^2$$

$$= \frac{1431}{81} - \left(\frac{324}{81}\right)^2 = \frac{53}{3} - 16 = \frac{5}{3}$$

(c) The mean and variance of the first population are

$$\mu_1 = \frac{4+6+8}{3} = 6, \text{ and}$$

$$\sigma_1^2 = \frac{(4-6)^2 + (6-6)^2 + (8-6)^2}{3} = \frac{8}{3}.$$

The mean and variance of the second population are

$$\mu_2 = \frac{1+2+3}{3} = 2, \text{ and}$$

$$\sigma_2^2 = \frac{(1-2)^2 + (2-2)^2 + (3-2)^2}{3} = \frac{2}{3}.$$

Now $\mu_{\bar{x}_1-\bar{x}_2} = 4 = 6 - 2 = \mu_1 - \mu_2$, and

$$\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} = \frac{8}{3}\cdot\frac{1}{2} + \frac{2}{3}\cdot\frac{1}{2}$$

$$= \frac{4}{3} + \frac{1}{3} = \frac{5}{3} = \sigma^2_{\bar{x}_1-\bar{x}_2}$$

Hence the result.

Example 14.16. Car batteries produced by company A have a mean life of 4.3 years with a standard deviation of 0.6 years. A similar battery produced by company B has a mean life of 4.0 years and a standard deviation of 0.4 years. What is the probability that a random sample of 49 batteries from company A will have a mean life of *at least* 0.5 years more than the mean life of a sample of 36 batteries from company B?

We are given the following data:

Population A: $\mu_1 = 4.3$ years, $\sigma_1 = 0.6$ years, $n_1 = 49$

Population B: $\mu_2 = 4.0$ years, $\sigma_2 = 0.4$ years, $n_2 = 36$

Both sample sizes ($n_1 = 49$, $n_2 = 36$) are large enough to assume that the sampling distribution of the differences $\bar{X}_1 - \bar{X}_2$ is approximately a normal with mean

$$\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2 = 4.3 - 4.0 = 0.3 \text{ years and standard deviation}$$

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{0.36}{49} + \frac{0.16}{36}} = 0.1086 \text{ years.}$$

Thus the variable

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

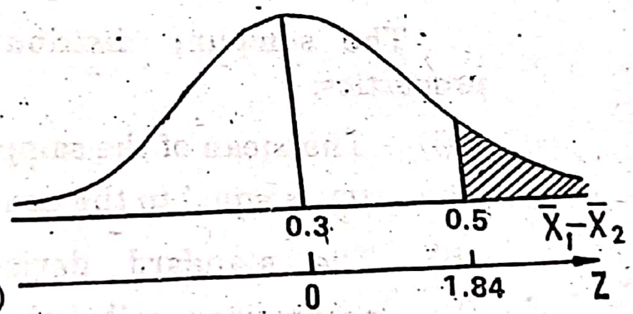
$$= \frac{(\bar{X}_1 - \bar{X}_2) - 0.3}{0.1086} \text{ is approximately } N(0, 1).$$

We are required to find the probability that the mean life of 49 batteries produced by company A will have a mean life of *at least* 0.5 years longer than the mean life of 36 batteries produced by company B, that is we want $P(\bar{X}_1 - \bar{X}_2 \geq 0.5)$.

Transforming $\bar{X}_1 - \bar{X}_2 = 0.5$ to z value, we find that

$$z = \frac{0.5 - 0.3}{0.1086} = 1.84$$

Hence using *Table* of areas under normal curve, we find



$$P(\bar{X}_1 - \bar{X}_2 \geq 0.5) = P(Z \geq 1.84)$$
$$= 0.5 - P(0 < Z < 1.84)$$
$$= 0.5 - 0.4671$$
$$= 0.0329$$

14.4.4. Sampling Distribution of Sample Proportion. A population proportion p may be identified with the population mean, where the mean is obtained from the units whose possible values are either 0's or 1's. In other words, let

$Y_i = 1$, if the *ith* unit possesses the characteristic of interest,

$= 0$, if the *ith* unit does not possess the characteristic of interest,

Then the mean is

$$\mu = \frac{1}{N}\sum_{i=1}^{N} Y_i$$

$$= \frac{\text{Number of units having the characteristic of interest}}{\text{Total number of units in the population}}$$

$$= \frac{X}{N},$$ where X represents the number of units having the characteristics of interest.

Thus the mean is simply the proportion of 1's in the population and we write p for μ, meaning *proportion* (usually called the proportion of success).

Similarly, the sample proportion $\hat{P}$ is defined as

$$\hat{P} = \bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i = \frac{X}{n}.$$

It is interesting to note that $X = \sum Y_i$ is a binomial random variable and the binomial parameter p is being called a proportion of success here. The sample proportion $\hat{P}$ has different values in different samples. It is obviously a random variable and has a probability distribution. This probability distribution of the proportions of all possible random samples of size n, is called the *sampling distribution of* $\hat{P}$.

The sampling distribution of $\hat{P}$ has the following important properties:

(i) The mean of the sampling distribution of proportions, denoted by $\mu_{\hat{p}}$, is equal to the population proportion p, that is $\mu_{\hat{p}} = p$.

(ii) The standard deviation of the sampling distribution of proportions, called the *standard error of* $\hat{P}$ and denoted by $\sigma_{\hat{p}}$, is given as $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$, when the sampling is performed with replacement or

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n} \cdot \frac{N-n}{N-1}},$$ when sampling is done without replacement

from a finite population, and where $q = 1-p$. It is of importance to remember that $\frac{N-n}{N-1}$ is to be used when the sample size n is 5% or more than 5% of the population size N.

When the population proportion p is not known and both the population and the sample sizes are large, then the sample proportion $\hat{p}$

obtained from sample data is used in place of p in the expression for the *S.E.* of P, getting

$$S_{\hat{p}} = \sqrt{\frac{\hat{p}\,\hat{q}}{n}}, \text{ where } \hat{q} = 1 - \hat{p}.$$

When the sample is selected without replacement from a finite population of size N, the *S.E.* becomes

$$S_{\hat{p}} = \sqrt{\frac{\hat{p}\,\hat{q}}{n} \cdot \frac{N-n}{N-1}}.$$

(iii) *Shape of the distribution.* The sampling distribution of $\hat{P}$ is the binomial distribution. However, for sufficiently large sample sizes, the sampling distribution of P is approximately normal. As a rule of thumb, the sampling distribution of P will be approximately normal whenever both np and nq are equal to or greater than 5.

It helps to remember that we use a continuity correction of $\pm \frac{1}{2}$, whenever we consider the normal approximation to the binomial distribution. Now, we need to use a continuity correction of $\pm\frac{1}{2n}$ as $\hat{P} = \frac{X}{n}$.

Example 14.17. A population consists of $N=6$ values 1, 3, 6, 8, 9 and 12. Draw all possible samples of size $n=3$ without replacement from the population and find the proportion of even numbers in the samples. Construct the sampling distribution of sample proportions and verify that

(i) $\mu_{\hat{p}} = p$, (ii) $\text{Var}(\hat{P}) = \frac{pq}{n} \cdot \frac{N-n}{N-1}$,

where $q = 1 - p$; $\hat{P}$ and p are sample and population proportions respectively.

The number of possible samples of size $n=3$ that could be selected without replacement is $\binom{6}{3} = 20$. Let $\hat{P}$ represent the proportion of even numbers in the sample. then the 20 possible samples and the proportion of even numbers are given as follows:

To verify the given relations, we first calculate the population proportion p and the population variance pq. Thus

$$p = \frac{X}{N}, \text{ where } X \text{ represents the number of even numbers.}$$

$$= \frac{3}{6} = 0.5, \text{ and}$$

$$\sigma^2 = pq = (0.5)(0.5) = 0.25$$

Therefore $\mu_{\hat{p}} = 0.5 = p$, and

$$\frac{pq}{n} \cdot \frac{N-n}{N-1} = \frac{0.25}{3} \cdot \frac{6-3}{6-1} = \frac{0.25}{5} = 0.05 = \text{Var}(\hat{P})$$

Hence the result.

Example 14.18. Ten percent of the 1-kilogram boxes of sugar in a large warehouse are underweight. Suppose a retailer buys a random sample of 144 of these boxes. What is the probability that at least 5 per cent of the sample boxes will be underweight?

Here the statistic is the sample proportion $\hat{P}$.

The sample size ($n = 144$) is large enough to assume that the sample proportion $\hat{P}$ is approximately normally distributed with mean

$$\mu_{\hat{p}} = p = 0.10, \text{ and standard error}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} = \sqrt{\frac{(0.10)(0.9)}{144}} = \frac{0.3}{12} = 0.025.$$

Therefore $Z = \dfrac{\hat{P} - \mu_{\hat{p}}}{\sigma_{\hat{p}}} = \dfrac{\hat{P} - p}{\sqrt{pq/n}} = \dfrac{\hat{P} - p}{0.025}$ is approximately $N(0, 1)$.

We are asked to find the probability that the sample proportion of the underweight boxes is equal to or greater than 5% *i.e.*, we require $P(\hat{p} \geq 0.05)$.

Thus $P(\hat{p} \geq 0.05) \Rightarrow P\left(\hat{p} \geq 0.05 - \dfrac{1}{(2)(144)}\right)$ [Continuity correction]

$$= P\left(\frac{\hat{p} - 0.10}{0.025} \geq \frac{(0.05 - 1/288) - 0.10}{0.025}\right)$$

$$= P(Z \geq -2.14)$$

$$= P(-2.14 \leq Z \leq 0) + P(0 \leq Z \leq \infty)$$

$$= 0.4838 + 0.5 = 0.9838 \text{ (From area table)}$$

14.4.5. Sampling Distribution of Differences between Proportions. Suppose there are two *binomial* populations with proportions of successes p_1 and p_2 respectively. Let independent random even Numb

Sample No.	Sample Data	Sample Proportion ($\hat{P}$)
1	1, 3, 6	1/3
2	1, 3, 8	1/3
3	1, 3, 9	0
4	1, 3, 12	1/3
5	1, 6, 8	2/3
6	1, 6, 9	1/3
7	1, 6, 12	2/3
8	1, 8, 9	1/3
9	1, 8, 12	2/3
10	1, 9, 12	1/3
11	3, 6, 8	2/3
12	3, 6, 9	1/3
13	3, 6, 12	2/3
14	3, 8, 9	1/3
15	3, 8, 12	2/3
16	3, 9, 12	1/3
17	6, 8, 9	2/3
18	6, 8, 12	1
19	6, 9, 12	2/3
20	8, 9, 12	2/3

The sampling distribution of sample proportion is given below:

$\hat{P}$	No. of samples	Probability $f(P)$	$\hat{p}\, f(\hat{p})$	$\hat{p}^2 f(\hat{p})$
0	1	1/20	0	0
1/3	9	9/20	3/20	1/20
2/3	9	9/20	6/20	4/20
1	1	1/20	1/20	1/20
Σ	20	1	10/20	6/20

Now $\mu_{\hat{p}} = \sum \hat{p}\, f(\hat{p}) = \dfrac{10}{20} = 0.5$, and

$$\sigma^2_{\hat{p}} = \sum \hat{p}^2 f(\hat{p}) - [\sum \hat{p}\, f(\hat{p})]^2 = \frac{6}{20} - \left(\frac{10}{20}\right)^2 = \frac{1}{20} = 0.05.$$

Example 14.19. Two random samples of sizes $n_1=40$ and $n_2=45$ are drawn from a binomial population with $p=0.60$. What is the probability that $-0.15<\hat{p}_1-\hat{p}_2<+0.15$?

Both sample sizes ($n_1=40$ and $n_2=45$) are large enough to assume that the sampling distribution of $\hat{P}_1-\hat{P}_2$ is approximately a normal with mean

$$\mu_{\hat{P}_1-\hat{P}_2} = p_1-p_2 = 0, \text{ and standard deviation}$$

$$\sigma_{\hat{P}_1-\hat{P}_2} = \sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)} = \sqrt{(0.60)\,(0.40)\left(\frac{1}{40}+\frac{1}{45}\right)}$$

$$= \sqrt{(0.24)\,(0.0472)} = 0.106.$$

Thus the variable

$$Z = \frac{(\hat{P}_1-\hat{P}_2)-0}{\sqrt{pq\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

$$= \frac{\hat{P}_1-\hat{P}_2}{0.106} \text{ is approximately } N(0,\,1).$$

Now, at $\hat{p}_1-\hat{p}_2=-0.15$, we find that $z=\dfrac{-0.15}{0.106}=-1.42$,

Hence using **Table** of areas under normal curve, we find

$$P(-0.15<\hat{p}_1-\hat{p}_2<0.15) = P(-1.42<Z<1.42)$$

$$= P(-1.42<Z<0)+P(0<Z<1.42)$$

$$= 0.4222+0.4222 = 0.8444.$$

The desired probability is therefore 0.8444.

14.4.6. Sampling Distribution of Variances. The sampling distribution of the sample variances calculated from all possible random samples of size n from a normal population with variance σ^2, is the so called *Chi-Square Distribution*, which is discussed in chapter 17. The sampling distribution followed by the ratio of two sample variances is called the *F-distribution* to be introduced in chapter 19.

EXERCISES

14.1 (a) Explain the following terms:

samples of sizes n_1 and n_2 be drawn from the respective populations, and the differences $\hat{p}_1 - \hat{p}_2$ between the proportions of all possible pairs of samples be computed. Then a probability distribution of the differences $\hat{P}_1 - \hat{P}_2$ can be obtained. Such a probability distribution is called the *sampling distribution of the differences between the proportions* $\hat{P}_1 - \hat{P}_2$, which has the following important properties:

(i) The mean of the sampling distribution of $\hat{P}_1 - \hat{P}_2$, denoted by $\mu_{\hat{P}_1 - \hat{P}_2}$, is equal to the difference between the population proportions, that is $\mu_{\hat{P}_1 - \hat{P}_2} = p_1 - p_2$.

(ii) The standard deviation of the sampling distribution of $\hat{P}_1 - \hat{P}_2$, (i.e. *the standard error of* $\hat{P}_1 - \hat{P}_2$) denoted by $\sigma_{\hat{P}_1 - \hat{P}_2}$ is given by

$$\sigma_{\hat{P}_1 - \hat{P}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}, \text{ where } q = 1 - p.$$

If both populations have the same proportion of successes, i.e. $p_1 = p_2 = p$ or if both the samples have been drawn from a common binomial population, then

$$\sigma_{\hat{P}_1 - \hat{P}_2} = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

Whenever the value of the common proportion p is not known, then for sufficiently large sample sizes, it is replaced with its estimate $\hat{p}_c$, which is computed by taking a weighted mean of the two observed sample proportions $\hat{p}_1$ and $\hat{p}_2$ as follows:

$$\hat{p}_c = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2} = \frac{\text{Sum of successes in the two samples}}{\text{Total sample size}}$$

The standard error of $\hat{P}_1 - \hat{P}_2$ then becomes

$$S_{\hat{P}_1 - \hat{P}_2} = \sqrt{\hat{p}_c \hat{q}_c \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}, \text{ where } \hat{q}_c = 1 - \hat{p}_c.$$

If, on the other hand, $p_1 \neq p_2$ and also are not known, then for large sample sizes, they are replaced with the sample proportions $\hat{p}_1$ and $\hat{p}_2$ respectively. The *S.E.* of $\hat{P}_1 - \hat{P}_2$ then becomes

$$S_{\hat{P}_1 - \hat{P}_2} = \sqrt{\frac{\hat{p}_1 \hat{q}_1}{n_1} + \frac{\hat{p}_2 \hat{q}_2}{n_2}}.$$

(iii) *Shape of the distribution.* The sampling distribution of $\hat{P}_1 - \hat{P}_2$ is approximately normal for sufficiently large sample sizes.

(ii) The random number table.

14.10 (a) What is a stratified random sample? In what way does it differ from a simple random sample, and what are the advantages and disadvantages of using this sampling technique?

(b) What is meant by allocation of sample size? Explain how a sample is allocated in stratified sampling.

14.11 Describe, in detail, various probability and non-probability sampling techniques. (P.U., B.A./B.Sc. 1981)

14.12 (a) Distinguish between

(i) Probability and Non-probability Sampling

(ii) Sampling and Non-sampling Errors

(iii) Multistage and Multiphase Sampling. (P.U., B.A/B.Sc. 1974, 92)

(b) What is a cluster sample? Why are cluster samples used?

14.13 (a) Describe the necessity of sampling and sample surveys.

(b) What is a Systematic sample? Describe the procedure of drawing a systematic sample of n units from a population of N units.

14.14 Describe the following types of Samples:

(a) Random Sample (b) Quota Sample

(c) Systematic Random Sample (d) Multi-stage Random Sample

State the conditions under which each would be used and the advantages to be gained. (P.U., M.A. Stats: 1964)

14.15 Explain each of the following:

(i) Area Sampling (ii) Optimum Allocation

(iii) Sequential Sampling (iv) Purposive Sampling

(v) Quota Sampling.

14.16 Draw all possible distinct samples of size two from the following population: 2, 4, 6, 8, 10.

Calculate the means and the variances of the samples and of the population. Discuss the results. (P.U., B.A/B.Sc. 1972)

14.17 Explain how you would select a random sample of 10 households from a list of 250 households, by using a table of random numbers. (P.U.M.A. Stats. 1967)

14.18 Using a random number table, select 30 samples of size 3 each with replacement from the following population distribution of heights. Find the mean of sample means.

Height (inches)	No. of students f
60 – 62	5
63 – 65	18
66 – 68	42
69 – 71	27
72 – 74	8

(P.U. B.Sc. Hons. Part I; 1971)

14.19 Draw, with the help of random numbers, a random sample of size 10 from a

(i) Binomial distribution with parameters $p = 0.4$ and $n = 5$;

(ii) Poisson distribution with the parameter $\mu = 4$.

14.20 Using a random number table, draw a sample of size 30 from a Normal distribution with $\mu = 100$ and $\sigma^2 = 64$.

14.21 (a) Describe stratified random sampling, explaining in detail the following types of allocation of sample sizes:

(i) Proportional Allocation (ii) Optimum Allocation.

(P.U., B.A./B.Sc., 1991)

(b) Select a stratified random sample of size $n=8$ by proportional allocation from the following population and find the sample mean and the estimate of the population mean.

Stratum I	$X_{11} = 3, X_{12} = 6, X_{13} = 4, X_{14} = 7,$
Stratum II	$X_{21}=10, X_{22}=12, X_{23}=15, X_{24}=16, X_{25}=16, X_{26} = 20,$
Stratum III	$X_{31}=16, X_{32}=18, X_{33}=21, X_{34}=22, X_{35}=26, X_{36}=23$

(P.U., B.A/B.Sc. 1975)

14.22 (a) At a small private college, the students are classified as follows:

Classification	B.Sc.	B.A.	F.Sc.	F.A.
No. of students	150	163	195	220

If we wish to select a stratified random sample of size $n=40$ by proportional allocation, how large a sample must we take from each stratum?

(P.U., B.A/B.Sc. 1987)

(b) A large company has 300,000 employees, the age distribution of whom is shown as follows:

Age (years)	Percentages
25 or younger	15
26 – 35	30
36 – 45	25
46 – 55	20
56 or older	10

A sample of 2 per cent of all the employees is desired. Design a sampling plan such that each age-group is proportionally represented. (P.U., B.A/B.Sc., 1988)

14.23 (a) What is a sampling distribution? Describe the properties of the sampling distribution of the means.

(b) What is the finite-correction factor? When is it appropriately used in sampling applications and when can it, without too great an undesirable consequence, be ignored? (P.U., B.A./B.Sc., 1996)

14.24 (a) Explain the difference between the population distribution, the sample distribution and sampling distribution. (P.U., B.A./B.Sc. 1989, 93)

(b) What is meant by standard error and what are its practical uses? Derive a formula for the standard error of the mean. (P.U., B.A. Hons. Part I; 1970)

14.25 (a) Explain the difference between (i) a sample distribution and a sampling distribution (ii) a standard deviation and a standard error.

(b) Suppose a friend says, "I know the formula for computing the standard error of the mean, but I do not understand what the standard error really is." Write a note to your friend explaining what the standard error really is.

14.26 (a) Distinguish between a parameter and a statistic. What is meant by Standard Error and what are its practical uses? (P.U. B.A./B.Sc. 1993)

(b) Assume that simple random samples of two children are selected with replacement from a population of five children with ages 4, 5, 6, 7 and 8. Let X be the age of any child; find the following:

(i) The theoretical sampling distribution of $\bar{X}$, the mean age of two children in any sample.

(ii) The mean and the standard error of $\bar{X}$.

(B.Z.U. B.A/B.Sc. 1976)

14.27 Given the population 2, 4, 8, 8, 10, 10.

(i) How many samples of size $n=2$ can be drawn without replacement from this population?

(ii) Compute and tabulate the sampling distribution of the mean for samples of size $n=2$.

14.28 A finite population consists of the numbers 2, 4 and 6.

(a) Form the sampling distribution of $\bar{X}$, when random samples of size 4 are drawn, with replacement.

(b) Verify that $\mu_{\bar{X}} = \mu$ and $\sigma^2_{\bar{X}} = \frac{\sigma^2}{n}$. (P.U., B.A/B.Sc. 1980)

14.29 A population consists of 2, 2, 4, 4, 6, 8 and 10.

(i) Calculate the sample means for all possible random samples of size $n=2$, that can be drawn from this population, without replacement.

(ii) Verify that $\mu_{\bar{X}} = \mu$ and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$

(iii) Between what two values would you expect *at least* $\frac{8}{9}$ of the sample means to fall? (P.U., B.A./B.Sc. 1993)

14.30 Given the six-element population 0, 3, 6, 12, 15, and 18. How many samples of size $n=3$ can be drawn, without replacement from this population? Compute the sampling distribution of the mean for samples of size 3. Compute the mean and standard deviation of this distribution.

14.31 Draw all possible samples of size $n=3$ with replacement from the population 3, 6, 9 and 12. Form a sampling distribution of the sample means. Hence state and verify the relation between

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

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

(P.U., B.A./B.Sc. 1982)

14.32 A population of $N=5$ has the following values:

4, 5, 7, 9, 10.

(a) Find the population mean and variance.

(b) Suppose samples of size $n=3$ are selected. Find $\sigma^2_{\bar{x}}$ when sampling is done. (i) without replacement, (ii) with replacement.

(c) Select all possible samples of size $n=3$ without replacement and calculate $\sigma^2_{\bar{x}}$ directly.

(d) Select all possible samples of size $n=3$ with replacement and calculate $\sigma^2_{\bar{x}}$ directly.

14.33 A population consists of four numbers 2, 4, 6, 8. Draw all possible sample of size $n=3$ with replacement. Find the mean and the median for each sample. Form the sampling distribution of means and the sampling distribution of medians. Which of these distributions has the smaller variance? How did the means of these two distributions compare with the population mean?

Mean only.

14.34 Given the following population distribution:

x	1	2	3	4
$f(x)$	$\frac{1}{7}$	$\frac{3}{7}$	$\frac{2}{7}$	$\frac{1}{7}$

Find the sampling distribution of the mean if a sample of three numbers is taken without replacement. How does the variance of the sampling distribution compare with the population variance?

14.35 A random sample of size $n=100$ is taken from a population having a mean of 20 and a standard deviation of 5. The shape of the population distribution is unknown.

(a) What can you say about the sampling distribution of the sample mean $\bar{X}$?

(b) Find the probability that $\bar{X}$ will exceed 20.75.

(P.U., B.A/B.Sc. 1996)

14.36 In a local agriculture reporting area, the average wheat yield is known to be 60 bushels per acre with a standard deviation of 10 bushels. If a random sample of 64 acres is selected and the wheat yield recorded, what is the probability that the sample mean will lie between 59 and 61 bushels?

14.37 The heights of 1000 students are approximately normally distributed with a mean of 68.5 inches and a standard deviation of 2.7 inches. If 200 random samples of size 25 are drawn from this population and the means recorded to the nearest tenth of an inch, determine

(a) The expected mean and standard deviation of the sampling distribution of the mean.

(b) The number of sample means that fall between 67.9 and 69.2 inclusive. (P.U., B.A./B.Sc. 1977-S)

14.38 The heights of a large number of shrubs of the same kind produced for sale by a horticultural nursery are normally distributed with mean 1.14m and standard deviation 0.25m. Fifty samples, each consisting of 100 shrubs, are selected. In how many of these samples would you expect to find the mean sample being to be (i) greater than 1.16m; (ii) between 1.13m and 1.18m?

14.39 (a) The following table shows the distribution of 14-year-old schoolboy intelligence test markings:

I.Q.	80-89	90-99	100-109	110-119	120-129	130-139	140-149
Number	30	52	75	109	65	42	27

On the assumption that this group is a random sample, estimate the standard error of the mean and explain its usefulness.

(b) The random variable X has the following probability distribution:

x	4	5	6	7
$P(X=x)$	0.2	0.4	0.3	0.1

(i) Find the mean $\mu_{\bar{x}}$ and variance $\sigma^2_{\bar{x}}$ of the mean $\bar{X}$ for a random sample of 36.

(ii) Find the probability that the mean of 36 items will be less than 5.5. (P.U., B.A./B.Sc. 1987)

14.40 (a) The mean of a certain normal distribution is equal to $S.E.$ of the mean of samples of 100 from that distribution. Find the probability that the mean of a sample of 25 from the distribution will be negative.

(b) A normal population has a mean of 0.1 and a standard deviation of 2.1. Find the probability that the mean of a simple random sample of 900 members will be negative.

Solution. (a) The variable X is $N\left(\frac{\sigma}{10}, \sigma\right)$ as

$\mu = \frac{\sigma}{10}$, the S.E. of the mean of samples of size $n = 100$.

And the S.E. of the mean of samples of size $25 = \frac{\sigma}{5}$.

From the standard normal variable $Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$, we get

$$\bar{X} = \mu + Z\frac{\sigma}{\sqrt{n}} = \frac{\sigma}{10} + Z \cdot \frac{\sigma}{5}.$$

Now $\bar{X}$ will be negative if $\left(\frac{\sigma}{10} + Z\frac{\sigma}{5}\right) < 0$,

i.e. if $Z < -\frac{\sigma}{10} \times \frac{5}{\sigma}$ or if $Z < -0.5$.

Hence using the *Table* of areas under normal curve, we get

$$P(Z < -0.5) = 0.5 - P(-0.5 < Z < 0)$$
$$= 0.5 - 0.1915 = 0.3085$$

14.41 (a) A random sample of size 100 is taken from a Binomial distribution with parameters $p = 0.5$ and $n = 40$. Find, using the central limit theorem, the approximate probability that $\bar{X}$ is (i) greater than 20.5; (ii) less than 19.3; and hence (iii) between 19.3 and 20.5.

(b) A sample of 36 cases is drawn from a negatively skewed population with a mean of 2 and a standard deviation of 3. What is the probability that the sample mean obtained will be negative? How many points must we go from the mean to include 50 percent of all sample means? (P.U., B.A/B.Sc., 198

14.42 (a) Describe the properties of the sampling distribution of th differences between two means. (P.U., B.A./B.Sc. 1983,8

(b) Random samples of size 100 are drawn, with replacemer from two populations and their means $\bar{X}_1$ and $\bar{X}_2$ comput If $\mu_1 = 10$, $\sigma_1 = 2$, $\mu_2 = 8$ and $\sigma_2 = 1$, find the probabil that the difference between a given pair of sample mean (i) less than 1.5, and (ii) greater than 1.75 but less than 2.

14.43 Let $\bar{X}_1$ represent the mean of a sample of size $n_1=2$, with replacement, from the finite population 3, 4, 5. Similarly, let $\bar{X}_2$ represent the mean of a sample of size $n_2 = 2$, with replacement, from the population 1, 1, 3.

(a) Find the possible differences between the sample means of the two populations.

(b) Construct the sampling distribution of $\bar{X}_1 - \bar{X}_2$ and compute its mean and variance.

(c) Verify that $\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2$ and $\sigma^2_{\bar{x}_1-\bar{x}_2} = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}$.

(P.U., B.A./B.Sc. 1985)

14.44 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. What is the probability that a random sample of 36 tubes from manufacturer A will have a mean lifetime that is at least 1 year more than the mean lifetime of a sample of 49 tubes from manufacturer B? (P.U., B.A./B.Sc. 1980)

14.45 A random sample of size 25 is taken from a normal population having a mean of 80 and a standard deviation of 5. A second random sample of size 36 is taken from a different normal population having a mean of 75 and a standard deviation of 3. Find the probability that the sample mean computed from the 25 measurements will exceed the sample mean computed from the 36 measurements by at least 3.4 but less than 5.9. Assume the means to be measured to the nearest tenth.(P.U. B.A./B.Sc. 1986-S)

14.46 What is meant by the term sampling distribution of sample proportion P? Describe its important properties and explain its usefulness in statistical inference.

14.47 A population consists of $N=6$ numbers 0, 3, 4, 6, 9 and 15. Draw all possible samples of size $n=3$, without replacement, from the population and find the sample proportion of even numbers in the samples. Construct the sampling distribution of sample proportions and verify that

$$\mu_{\hat{p}} = p \text{ and } \mathrm{Var}(\hat{P}) = \frac{pq}{n} \cdot \frac{N-n}{N-1}.$$

14.48 A population consists of $N=7$ numbers 1, 1, 2, 3, 4, 4, 5. Draw all possible samples of size $n=3$ without replacement from this population and find the sample proportion of odd numbers in the samples. Construct the sampling distribution of sample proportion and verify

(i) $\mu_{\hat{p}} = p$, (ii) $\sigma^2_{\hat{p}} = \frac{pq}{n} \cdot \left(\frac{N-n}{N-1}\right)$. (P.U., B.A./B.Sc., 1989)

14.49 (a) Two per cent of the trees in a plantation are known to have a certain disease. What is the probability that, in a sample of 250 trees, (i) less than 1%, (ii) more than 4% are diseased?

(b) 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 probability that the sample proportion favouring this issue will be less than 0.52?

14.50 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 larger group. Assuming that the proportion of the *entire* membership favouring consolidation is $p=0.7$, find the probability that the sample proportion P differs from this by no more than 0.05.

14.51 (a) Describe the sampling distribution of differences between proportions and explain its usefulness in statistical inference.

(b) Two random samples of sizes $n_1 = 40$ and $n_2 = 45$ are drawn from a binomial population with $p=0.70$. What is the probability that $-0.1 < \hat{P}_1 - \hat{P}_2 < 0.1$?

14.52 A population consists of five observations 1, 2, 3, 4, 5. Draw all possible samples of size 2 with replacement. Find the mean of the sampling distribution of the variances. Compare it with the variance of the population. (P.U., B.A./B.Sc. 1990)