Unit - 5

Sampling and Sampling Distributions

After studying this unit, the students will be able to

- Define sampling, sampling unit, sampling frame and sample design
- Differentiate between finite and infinite populations, sample survey and complete enumeration. Describe advantages and limitations of sampling.
- Distinguish between probability/random sampling and non-probability/non-random sampling, random sampling with and without replacement.
- Differentiate between sampling and non-sampling errors
- Describe simple random sampling stratified random sampling and systematic random sampling.
- Use the random digits/random numbers table to select a simple random sample from a given finite population.
- Define sampling distribution of statistic and standard error of statistic, sampling distribution of sample mean.
- Describe the properties of a sampling distribution of a sample mean.
- Construct the sampling distribution of sample mean to verify its properties about its mean and variance.
- Define the sampling distribution of difference between two sample means.
- Describe the properties of sampling distribution of difference between two sample means.
- Construct the sampling distribution of difference between two samples means to verify its properties about its mean and variance.
- Define sampling distribution of sample proportion.
- Describe the properties of a sampling distribution of sample proportion.
- Construct the sampling distribution of sample proportion to verify its properties about its mean and variance.
- Define sampling distribution of difference between two sample proportions.
- Describe the properties of sampling distribution of difference between two sample proportions .Construct sampling distribution of difference between two sample proportions and verify its properties about mean and variance.

5.1 Need of sampling survey

The origin of sampling is as old as our civilization and now it is a part of our day-to-day life. Human knowledge and actions are mainly based on samples, specifically, in scientific research. Sample data can either be obtained through sample surveys or experimentation which is the basic input in statistical analysis and inference.

5.1.1 Key terms and definitions

- **Population**

A group of individuals or objects about which we wish to know something is called population or universe.

- **Finite population**

If the number of individuals or items of a population are fixed and limited, it is known as finite population e.g. the population of government hospitals in Pakistan, students in a college, workers in a factory, etc. Finite population usually consists of existing items.

- **Infinite population**

If the population consists of an infinite number of items, it is called infinite population. For example, the population of all real numbers is lying between 10 and 20, the population of stars in the sky etc.

- **Sampling unit**

A single element or group of elements of a population from which required information can be obtained is called sampling unit or unit of the population. For example, if a population has 100 mango trees, then a single mango tree is the sampling unit. A unit may either be natural (e.g. a person, a family, a tree, an animal etc.) or artificial (e.g. a village, districts, a plot of specified size etc.).

- **Sample**

A part or fraction of a population is called sample. In everyday life decision about population is made on the basis of a sample, therefore, it is important that personal liking or disliking may not be involved during the selection of a sample, that is, sample should be random and must be a true representative of the concerned population.

♦ Survey

To ask a question or a series of questions form many people in order to gather information is called survey. When every individual of the population is examined, it is called population survey or census and if only a part of the population is examined, it is called sample survey or sampling survey. Nowadays sample surveys are being widely used by government departments like agriculture, industries, commerce etc. and private agencies for obtaining estimates of respective parameters. Sample survey has two types i.e. descriptive sample survey (simple information are obtained) and analytical sample survey (comparisons are made between subgroups of the population).

♦ Sampling frame

Before applying any sampling procedure, it is essential to have a list of all the sampling units or a map or other acceptable material which represent the population to be covered. Such a list or map is called sampling frame. For example, if we wish to estimate the wheat crop area in Khyber Pakhtunkhwa, the record of farms along with the names of the farmers, villages etc. are the frame. The frame should not contain inaccurate sampling units and be as up-to-date as possible at the time of use.

♦ Sample design

All principle steps including the methods which are taken in the selection of a sample is called sample design or sample plan or sampling plan. It is formulated before the actual collection of any data. For example, if we want to take a sample of students from Peshawar university, then a complete plan showing number of students to be included in the sample, which students are to be included, the proportion of students, method of sampling to be used, what characteristics are to be studied etc. will be called sample design.

♦ Survey design

The sample design along with some other aspects of the survey e.g. choice and training of interviewers, tabulation plans etc. is called survey design.

5.1.2 Difference between census and sampling

In census (complete enumeration), the information are obtained from each and every unit of the population. In our country census is done after every ten years. Census gives quite reliable information but practically we face the following problems:

i. When population is infinite or area of survey is wide, its study is impossible.
ii. Too much time is required to cover the whole population and often the study becomes out dated by the time it is completed.
iii. Too much resources i.e. money, trained persons etc. are required for survey of the whole population.
iv. When the item or unit is destroyed under investigation, the study of population serves no purpose e.g. testing the life of bulb, battery cell or any other electronic items.

In contrast to census, sampling studies only a selected number of units of the population. For instance (i) A housewife takes one or two grains of rice from the cooking pan and decides whether the rice is cooked or not (ii) A pathologist takes a few drops of blood and tests for any change in blood of the whole body than normal (iii) A quality controller takes a few items and decides whether the lot is in accordance with the desired specifications or not (iv) In a bulb manufacturing factory one tests the life of a few bulbs and then the conclusion about the average life of the whole bulbs produced by the factory is made. All these examples/uses clearly reveal that sampling has been an old practice. Sampling is less expensive and less time consuming.

5.1.3 Advantages of sampling

The most important advantages of sampling over census are;

i. Sampling saves a lot of time because the data are collected and analysed more quickly.
ii. The cost of sampling is very much low than 100 % enumeration because the cost per unit being the same and the number of units in the sample are always less than the number of units in the population.
iii. Sampling results can be more accurate because the sources of error, personal biases and analysis of data can be handled more easily.
iv. When the units are destroyed e.g. testing the life of any electronic instruments, then sampling is the only practical way to assess the average life of the whole lot.

5.1.4 Disadvantages of sampling

Sampling has some limitations given below:

i. When the information is required about every unit of the population, then no sampling method is suitable to give the desired information, only census can do it.
ii. Only large samples can indicate the true characteristics of the population.
iii. Sampling techniques require services of expert persons for better supervision, otherwise, results of the survey may not be reliable.

5.1.5 Parameter, statistic and estimator

Any characteristic of population is called parameter. Parameter is an arbitrary constant value and is usually denoted by Greek letter e.g. μ = mean, σ^2= variance, σ = standard deviation etc. Any function of sample data is called statistic. It is denoted by Latin letter e.g. $\overline{X}$ = mean, S^2 = variance, S = standard deviation etc. A statistic which is used for estimation of parameter is called an estimator. Note that (i) Statistic and estimator are random variables because they vary from sample to sample, (ii) every estimator is a statistic but every statistic is not an estimator.

5.1.6 Sampling and non-sampling errors

♦ Sampling error

The error which occurs due to the natural differences among the members of the population is called sampling error. For example, I.Q, educational level,

income, height, weight, age, etc. of human being are naturally different from each other. Mathematically, this error is written as $e = |\overline{X} - \mu|$ where μ denote parameter and $\overline{X}$ denote estimator. Sampling error can be reduced by (i) increasing the sample size (ii) improving the sample design. It is also called random error or compensating error.

♦ Non-sampling error

The error which occurs during the process of collection and processing the data is called non sampling error. It includes all kinds of human error. Non sampling error can be reduced by (i) employing trained personnel (ii) using modern computational aids.

5.1.7 Bias in sampling

The error which arises due to the personal interest of the investigator is called bias. Error can be minimized in the long run while bias cannot be. This error increases with the increase in sample size. It is also called cumulative error. Some of the factors which introduce bias are (i) deliberate selection of the sample units (ii) substitution of sample units by other units (iii) incomplete or inadequate interviewing (iv) haphazard or accidental selection. Bias cannot be reduced until it is detected.

5.1.8 Sampling with and without replacement

♦ Sampling with replacement

When the item selected for a sample is returned to the population before drawing next item, it is called sampling with replacement (S.W.R) In this case;

i. Possible samples will be N^n
ii. Probability of selection of each unit remains equal.
iii. Sampling units will be independent.
iv. Sampling units can be selected more than once.
v. Population remains infinite.

♦ Sampling without replacement

When the item selected for a sample is not returned to the population before drawing the next item, it is called sampling without replacement (S.W.O.R). In this case;

i. Possible samples will be $\overset{N}{C}_{n}$

ii. Probability of selection of each unit changes from draw to draw.

iii. Sampling units will be dependent.

iv. Sampling units can be selected only once.

v. The sample size is less or equal to the population size, that is, $n \leq N$.

Example 5.1

Suppose a population consist of values 2, 4, 6, 8. Draw all possible samples of size two (i) with replacement and (ii) without replacement.

Solution:

Given 2, 4, 6, 8 (population), $N = 4$ (population size), $n = 2$ (sample size)

(i) Sampling with replacement case:

The number of samples in this case $= N^n = 4^2 = 16$. The procedure for drawing samples is that divide total number of samples by population size i.e. $\frac{16}{4} = 4$ and write every unit of the population 4-times in the first column, then divide the obtained result 4 by the population size i.e. $\frac{4}{4} = 1$ and write every element of the population once in the second column as shown below;

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

(ii) Sampling without replacement case

The number of samples in this case $= \overset{N}{C}_{n} = \overset{4}{C}_{2} = 6$. The procedure for drawing samples in this case is that consider the first unit 2 with all other units of the population and keep it aside, then consider 4 with remaining units and keep it aside and so on, we have

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

5.1.9 Probability and non-probability sampling methods

Sampling methods are broadly divided in to two categories (i) Probability sampling methods and (ii) non-probability sampling methods.

♦ Probability sampling methods

When each sampling unit of the population has non-zero probability of selection at each draw, the selection procedure is known as probability sampling or random sampling and the sample selected is called random sample. We will describe here the simple random sampling, stratified random sampling and systematic random sampling methods.

♦ Non-probability sampling methods

When selection of units for a sample on each draw is not based on probability but personal judgment plays a significant role, the selection procedure is called non-probability sampling or non-random sampling. Commonly used non-probability sampling methods are: purposive or judgment sampling, Quota sampling, convenience sampling etc.

5.1.10 Simple random sampling

In this method, each and every unit in the population has an equal chance of being selected for a sample. It is also known as random sampling. It is the simplest and widely used method among probability sampling methods. It is a base for other complicated sampling methods. It is the best one for homogenous population. The sample selected by this method is called simple random sample or simply random sample. A random sample may either be selected by with replacement process with probability $\frac{1}{N^n}$ or without replacement process with probability $\frac{1}{{}^{N}C_{n}}$.

♦ Methods used for selection of a simple random sample

i. Goldfish bowl method

According to this method, serial numbers are allotted from 1 to N to all units of the population. The serial numbers are then written on equal size pieces of paper and are placed in a basket or box or a bowl. After shuffling, a piece of paper is drawn. Its number is noted and is returned to the population in S.W.R case, and then another piece is drawn and so on. The process is continued till the desired number of units is obtained.

ii. Random numbers table method

This method has already been described in unit-3

iii. Computer method

This method is like random number table method. The only difference is that here random numbers are generated by computer packages or calculator.

Example 5.2

Select a simple random sample of size 5, without replacement from the population of 50 employees of a factory by using random number table.

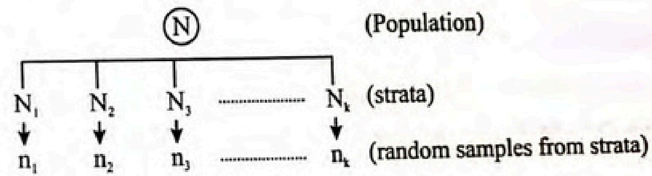
Solution:

Allot two digit serial numbers 00, 01, 02... 49 to employees. Select 5 two digit random number from the random number table, ignoring the two digit number which is greater than 49 or which is appearing again being S.W.O.R case. Let the random numbers be 04, 10, 37, 17, and 48. Thus we will include those employees of the factory in our sample whose serial numbers are 04, 10, 17, 37, and 48.

Note that if we proceed in the same way, we can draw different samples of size 5.

5.1.11 Stratified random sampling

When population is homogenous according to a characteristic e.g. population of first year students is homogenous according to education level, then simple random sampling is good method but when population is heterogeneous with respect to the characteristic in which we are interested e.g. population of students in a postgraduate college is heterogeneous according to the education standard because it contains first year, second year,..., 6th year students. Here simple random sampling does not give satisfactory results and in such situations, stratified random sampling is used to obtain representative sample. According to this method, heterogeneous population is divided in to homogeneous sub-populations, called strata. From each stratum a separate sample is selected by simple random sampling and the overall sample is obtained by combining the samples for all strata. This procedure is known as stratified random sampling and the sample selected by this method is called stratified random sample. For example,



$n_1 + n_2 + \ldots + n_k = n$ and n is called stratified random sample. It is important to note that stratum should be homogenous with respect to the characteristics under study. However, there should be heterogeneity among strata.

♦ Allocation of sample size to various strata

In stratified random sampling one of the major problem is that how much portion of the total sample size n is taken from each stratum? Four methods namely equal allocation, proportional allocation; Neyman's allocation and optimum allocation are generally used for allocation of n to various strata. The most commonly used one in practice is the proportional allocation method. According to this method, the number of units to be selected from a stratum, is proportional to the size of the stratum and are obtained by the formula:

$$n_h = \frac{n}{N} N_h$$

Where

N = population size

N_h = h^{th} stratum size

n_h = number of units from h^{th} stratum for n.

n = sample size

Example 5.3

Among the 250 employees of a local office 180 are matriculate, 50 are graduates and 20 are master degree holders. If we use proportional allocation to select a stratified random sample of 15 employees, how many employees must we take from each stratum?

Solution:

Given $N = 250, N_1 = 180, N_2 = 50, N_3 = 20, n = 15$

Now random samples from each stratum are obtained by the formula;

$$n_h = \frac{n}{N} N_h$$

$$n_1 = \frac{n}{N} N_1 = \frac{15}{250}(180) = 10.8 \cong 11$$

$$n_2 = \frac{n}{N} N_2 = \frac{15}{250}(50) = 3$$

$$n_3 = \frac{n}{N} N_3 = \frac{15}{250}(20) = 1.2 \cong 1$$

and the stratified random sample is $n = n_1 + n_2 + n_3 = 11 + 3 + 1 = 15$

5.1.12 Systematic random sampling

In systematic random sampling units are selected at equal interval i.e. every k^{th} unit, with a random start. The procedure is that items of the population are first numbered from 1 to N and are then divided into subgroups, each containing $\frac{N}{n} = k$ units, so that $N = nk$. A unit i is selected at random from the first k units, then every k^{th} unit starting with i is selected i.e. i, $(i+k)$, $(i+2k)$,...,$\{i+(n-1)k\}$. Let suppose one is interested in choosing 100 schools from a population of 1000 primary schools. All schools are first given a serial number as 001, 002, 003... 1000. Compute $\frac{N}{n} = \frac{1000}{100} = 10 = k$ .Choose a number at random from the first 10 numbers. Let it is 3, then our sample will include 3rd, 13th, 23rd, 33th...993rd number schools. Systematic random sampling is a type of random sampling. The sample selected by this method is called systematic random sample.

5.2 Sampling distribution

Recall that frequency table or frequency distribution is constructed for summarization of raw data. Probability distribution is constructed for summarizing the values of a random variable. Now, here we are dealing with possible samples and the corresponding estimators like $\bar{X}, \hat{p}, S^2, S$ etc. The values of these estimators would also need to be summarized in tabular form. This table is technically called sampling distribution. Hence following the same pattern we can define the sampling distribution as; "The probability distribution of all possible values of an estimator is called sampling distribution of that estimator".

5.2.1 Standard error

The standard deviation of the sampling distribution of an estimator is called standard error (S.E). Both have the same concept. The standard error measures the dispersion of all values of an estimator from its average.

5.2.2 Sampling distribution of sample mean

The probability distribution of all possible values of the estimator $\overline{X}$ is called sampling distribution of $\overline{X}$ .i.e.

Sampling distribution of $\overline{X}$

$\overline{X}$	$f(\overline{x})$
$\overline{x}_1$	$f(\overline{x}_1)$
$\overline{x}_2$	$f(\overline{x}_2)$
$\vdots$	$\vdots$
$\overline{x}_k$	$f(\overline{x}_k)$
Total	1

The mean of the sampling distribution of $\overline{X}$ is denoted by $E(\overline{X})=\mu_{\overline{x}}$ (read as mu sub x bar) and its S.E by $\sigma_{\overline{x}}$ (read as sigma sub x bar).

5.2.3 Properties of the sampling distribution of $\overline{X}$

i. Mean of the sampling distribution of $\overline{X}$ is always equal to the population mean i.e. $\mu_{\overline{x}}=\mu$ (both in with and without replacement sampling)

ii. Standard error of the sampling distribution of $\overline{X}$ is

$\sigma_{\overline{x}}=\frac{\sigma}{\sqrt{n}}$ (in case of sampling with replacement)

$=\frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ (in case of sampling without replacement)

Note that the factor $\sqrt{\frac{N-n}{N-1}}$ is called finite population correction factor (fpc). It is dropped from the formula when $n<5\ \%$ N and is used when $n\geq 5\ \%\ N$.

iii. Shape of the sampling distribution of $\overline{X}$

There are two cases:

(a) If the sampled population is normal, then shape of the sampling distribution of $\overline{X}$ will also be normal irrespective of the sample size.

(b) If the sampled population is non-normal, then according to central limit theorem (CLT), shape of the sampling distribution of $\overline{X}$ will approximately be normal provided sample size n is large. Note that statisticians consider $n\geq 30$ as large sample.

5.2.4 Formulas for mean and S.E of the sampling distribution of $\overline{X}$

$\mu_{\overline{x}}=\Sigma\overline{x}f(\overline{x})$ (Mean)

$\sigma^2_{\overline{x}}=\Sigma\overline{x}^2f(\overline{x})-[\Sigma\overline{x}f(\overline{x})]^2$ (Variance)

$\sigma_{\overline{x}}=\sqrt{\Sigma\overline{x}^2f(\overline{x})-[\Sigma\overline{x}f(\overline{x})]^2}$ (S.E)

Note that the random variable $\overline{X}$ is standardized as $Z=\frac{\overline{X}-\mu_{\overline{x}}}{\sigma_{\overline{x}}}=\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}$ (In S.W.R. case)

Example 5.4

Consider a family (population) of four children having ages 2, 2, 6, 8 years.

i. Find mean and standard deviation of the population.

ii. Select random samples of two children with replacement and calculate the mean age $\overline{x}$ for each sample.

iii. Construct sampling distribution of $\overline{X}$

iv. Find the mean and standard error of the sampling distribution of mean.

v. Verify the results obtained in (iv) by properties of the sampling distribution of $\overline{X}$

Solution:

(i) Population mean is given by

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

Population Standard deviation is given by

$$\sigma = \sqrt{\frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2} = \sqrt{\frac{108}{4} - \left(\frac{18}{4}\right)^2} = \sqrt{6.75} = 2.598$$

(ii) Possible samples in S.W.R $= N^n = 4^2 = 16$

S.No.	Samples	$\bar{X}$	S.No.	Samples	$\bar{X}$
1	2, 2	2	9	6, 2	4
2	2, 2	2	10	6, 2	4
3	2, 6	4	11	6, 6	6
4	2, 8	5	12	6, 8	7
5	2, 2	2	13	8, 2	5
6	2, 2	2	14	8, 2	5
7	2, 6	4	15	8, 6	7
8	2, 8	5	16	8, 8	8

(iii) Sampling distribution of $\bar{X}$

$\bar{X}$	Tally bar	f	$f(\bar{x})$
2	IIII	4	4/16
4	IIII	4	4/16
5	IIII	4	4/16
6	I	1	1/16
7	II	2	2/16
8	I	1	1/16
Total		16	1

(iv) Calculation for the mean and S.E of the sampling distribution of $\bar{X}$

$\bar{X}$	$f(\bar{x})$	$\bar{X}f(\bar{x})$	$\bar{X}^2 f(\bar{x})$
2	4/16	8/16	16/16
4	4/16	16/16	64/16
5	4/16	20/16	100/16
6	1/16	6/16	36/16
7	2/16	14/16	98/16
8	1/16	8/16	64/16
Total	1	72/16	378/16

$$\text{Mean} = \mu_{\bar{x}} = \sum \bar{x} f(\bar{x}) = \frac{72}{16} = 4.5$$

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

$$= \sqrt{\frac{378}{16} - \left(\frac{72}{16}\right)^2}$$

$$= \sqrt{23.625 - 20.25} = \sqrt{3.375} = 1.84$$

(v) Verification:

(a) As $\mu = 4.5$ and $\mu_{\bar{x}} = 4.5$, hence the property $\mu_{\bar{x}} = \mu$ is satisfied i.e. mean of all sixteen sample means is equal to the population mean.

(b) By property S.E in S.W.R case is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{2.598}{\sqrt{2}} = \frac{2.598}{1.414} = 1.84$$. It is same as the computed S.E

We learnt that if population mean and standard deviation are known, then it is easy to compute $\mu_{\bar{x}}$ and $\sigma_{\bar{x}}$ by property. If σ is unknown and n is large then $\sigma_{\bar{x}}$ can be estimated by $s_{\bar{x}} = \frac{s}{\sqrt{n}}$, where s is sample standard deviation.

Example 5.5

A population consists of number 2, 4, 8, 8, 10, 10. Samples of size 2 are to be drawn without replacement from this population.

(i) Find mean and S.D of this population.

(ii) Construct the sampling distribution of $\bar{X}$

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

Solution :

Given population is 2, 4, 8, 8, 10, 10

Here $N = 6, n = 2$

(i) $$\mu = \frac{\sum x}{N} = \frac{2+4+8+8+10+10}{6} = \frac{42}{6} = 7$$

$$\sigma = \sqrt{\frac{\sum x^2}{N} - \left(\frac{\sum x}{N}\right)^2} = \sqrt{\frac{348}{6} - \left(\frac{42}{6}\right)^2} = \sqrt{58-49} = \sqrt{9} = 3$$

(iii) Possible samples that could be drawn without replacement $= {}^6C_2 = 15$

S.No.	Samples	$\bar{X}$	S.No.	Samples	$\bar{X}$
1	2, 4	3	9	4, 10	7
2	2, 8	5	10	8, 8	8
3	2, 8	5	11	8, 10	9
4	2, 10	6	12	8, 10	9
5	2, 10	6	13	8, 10	9
6	4, 8	6	14	8, 10	9
7	4, 8	6	15	10, 10	10
8	4, 10	7			

The sampling distribution of $\overline{X}$ is given below:

$\overline{X}$	Tally bar	f	$f(\bar{x})$
3	I	1	1/15
5	II	2	2/15
6	IIII	4	4/15
7	II	2	2/15
8	I	1	1/15
9	IIII	4	4/15
10	I	1	1/15
Total		15	1

(iii) Calculation for mean and S.E of the sampling distribution of $\overline{X}$

$\overline{X}$	$f(\bar{x})$	$\overline{X} f(\bar{x})$	$\overline{X}^2 f(\bar{x})$
3	1/15	3/15	9/19
5	2/15	10/15	50/15
6	4/15	24/15	144/15
7	2/15	14/15	98/15
8	1/15	8/15	64/15
9	4/15	36/15	324/15
10	1/15	10/15	100/15
Total	1	105/15	789/15

$$E(\overline{X}) = \mu_{\bar{x}} = \Sigma \bar{x} f(\bar{x}) = \frac{105}{15} = 7$$

$$\sigma_{\bar{x}} = \sqrt{\Sigma \bar{x}^2 f(\bar{x}) - [\Sigma \bar{x} f(\bar{x})]^2} = \sqrt{\frac{789}{15} - \left(\frac{105}{15}\right)^2}$$

$$= \sqrt{52.6 - 49} = \sqrt{3.6} = 1.9$$

Now verification of the properties:

(i) As we see that $\mu = 7$, $E(\overline{X}) = 7$, hence proved that $E(\overline{X}) = \mu$

(ii) In S.W.O.R case, the S.E is

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N-n}} = \frac{3}{\sqrt{2}}\sqrt{\frac{6-2}{6-1}} = \frac{3}{\sqrt{2}}\sqrt{\frac{4}{5}} = 1.9$$

This is exactly the same as computed $\sigma_{\bar{x}}$.

Example 5.6

Draw all possible random samples of size 3 with replacement from a population of three values 3, 6, 9. Obtain the sampling distribution of sample mean and verify the theoretical results.

Solution:

Given population: 3, 6, 9, $N = 3$, $n = 3$

Here sampling is done with replacement, therefore, possible samples are $N^n = 3^3 = 27$. The possible samples and the sample means are listed in the following table.

S.No.	Samples	$\overline{X}$	S.No.	Samples	$\overline{X}$	S.No.	Samples	$\overline{X}$
1	3, 3, 3	3	10	6, 3, 3	4	19	9, 3, 3	5
2	3, 3, 6	4	11	6, 3, 6	5	20	9, 3, 6	6
3	3, 3, 9	5	12	6, 3, 9	6	21	9, 3, 9	7
4	3, 6, 3	4	13	6, 6, 3	5	22	9, 6, 3	6
5	3, 6, 6	5	14	6, 6, 6	6	23	9, 6, 6	7
6	3, 6, 9	6	15	6, 6, 9	7	24	9, 6, 9	8
7	3, 9, 3	5	16	6, 9, 3	6	25	9, 9, 3	7
8	3, 9, 6	6	17	6, 9, 6	7	26	9, 9, 6	8
9	3, 9, 9	7	18	6, 9, 9	8	27	9, 9, 9	9

The sampling distribution ...

$\bar{X}$	Tally bar	f	$f(\bar{x})$	$\bar{X}f(\bar{x})$	$\bar{X}^2 f(\bar{x})$
3	I	1	1/27	3/27	9/27
4	III	3	3/27	12/27	48/27
5	~~IIII~~ I	6	6/27	30/27	150/27
6	~~IIII~~ II	7	7/27	42/27	252/27
7	~~IIII~~ I	6	6/27	42/27	294/27
8	III	3	3/27	24/27	192/27
9	I	1	1/27	9/27	81/27
Total		27	1	162/27	1026/27

Now $E(\bar{X}) = \mu_{\bar{x}} = \Sigma \bar{x} f(\bar{x}) = \dfrac{162}{27} = 6$

And $\sigma_{\bar{x}} = \sqrt{\Sigma \bar{x}^2 f(\bar{x}) - [\Sigma \bar{x} f(\bar{x})]^2} = \sqrt{\dfrac{1026}{27} - (6)^2} = \sqrt{38-36} = 1.4142$

Population mean: $\mu = \dfrac{\sum x}{N} = \dfrac{3+6+9}{3} = \dfrac{18}{3} = 6$

Population S.D: $\sigma = \sqrt{\dfrac{\sum x^2}{N} - \left(\dfrac{\sum x}{N}\right)^2} = \sqrt{\dfrac{126}{3} - \left(\dfrac{18}{3}\right)^2} = \sqrt{42-36} = \sqrt{6} = 2.45$

Verification

(i) We see that $E(\bar{X}) = \mu = 6$

(ii) The S.E in S.W.R. case is $\dfrac{\sigma}{\sqrt{n}} = \dfrac{2.45}{\sqrt{3}} = 1.4142 = \sigma_{\bar{x}}$

5.3 Sampling distribution of the difference between two sample means

The probability distribution of all possible values of $\bar{X}_1 - \bar{X}_2$ is called sampling distributing of $\bar{X}_1 - \bar{X}_2$. Mean of the sampling distribution of $\bar{X}_1 - \bar{X}_2$ is denoted by $\mu_{\bar{x}_1 - \bar{x}_2}$ and S.E by $\sigma_{\bar{x}_1 - \bar{x}_2}$

5.3.1 Properties of the sampling distribution of $\bar{X}_1 - \bar{X}_2$

(i). Mean of the sampling distribution of $\bar{X}_1 - \bar{X}_2$ is equal to the difference in population means i.e. $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$ (both in S.W.R and S.W.O.R)

(ii). Standard error of the sampling distribution of $\bar{X}_1 - \bar{X}_2$ is

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} \quad \text{(in case of S.W.R)}$$

$$= \sqrt{\frac{\sigma_1^2}{n_1}\frac{N_1 - n_1}{n_1 - 1} + \frac{\sigma_2^2}{n_2}\frac{N_2 - n_2}{n_2 - 1}} \quad \text{(in case of S.W.O.R)}$$

(iii). Shape of the sampling distribution of $\bar{X}_1 - \bar{X}_2$

There are two cases:

a) If the sampled population is normal, then sampling distribution of $\bar{X}_1 - \bar{X}_2$ will also be normal irrespective of the samples size.

b) If the sampled population is non-normal, then according to central limit theorem, sampling distribution of $\bar{X}_1 - \bar{X}_2$ will approximately be normal provided n_1 and n_2 both are large.

The variable $\bar{X}_1 - \bar{X}_2$ in standard units is written as

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_{\bar{x}_1 - \bar{x}_2})}{\sigma_{\bar{x}_1 - \bar{x}_2}} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$ (in S.W.R case)

5.3.2 Formulas for $\mu_{\bar{x}_1 - \bar{x}_2}$ and $\sigma_{\bar{x}_1 - \bar{x}_2}$

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

$$\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\sum(\bar{x}_1 - \bar{x}_2)^2 f(\bar{x}_1 - \bar{x}_2) - [\sum(\bar{x}_1 - \bar{x}_2)\, f(\bar{x}_1 - \bar{x}_2)]^2}$$

Example 5.7

Draw all possible random samples of size $n_1 = 2$ with replacement from a finite population (3, 4, 5) and sample of size $n_2 = 2$ with replacement from another finite population (1, 1, 3).

(i) Find the possible differences between the sample means drawn from the populations.

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

(iii) Verify that $\mu_{\bar{x}_1 - \bar{x}_2} = \mu_1 - \mu_2$ and $\sigma_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

Solution:

Given	Population-1	Population-2
	[3, 4, 5]	[1, 1, 3]
	$N_1 = 3$	$N_2 = 3$
	$n_1 = 2$	$n_2 = 2$

As the case is that of sampling with replacement therefore possible samples from population-1 are $N_1^{n_1} = 3^2 = 9$ and population-2 are $N_2^{n_2} = 3^2 = 9$.

Samples from Population-1		
No	Samples	$\bar{X}_1$
1	3, 3	3
2	3, 4	3.5
3	3, 5	4
4	4, 3	3.5
5	4, 4	4
6	4, 5	4.5
7	5, 3	4
8	5, 4	4.5
9	5, 5	5

Samples from Population-2		
No	Samples	$\bar{X}_2$
1	1, 1	1
2	1, 1	1
3	1, 3	2
4	1, 1	1
5	1, 1	1
6	1, 3	2
7	3, 1	2
8	3, 1	2
9	3, 3	3

(i) The 9 × 9 = 81 possible differences $\bar{X}_1 - \bar{X}_2$ are given in the following table

$\bar{X}_2$ \ $\bar{X}_1$	3	3.5	4	3.5	4	4.5	4	4.5	5
1	2	2.5	3	2.5	3	3.5	3	3.5	4
1	2	2.5	3	2.5	3	3.5	3	3.5	4
2	1	1.5	2	1.5	2	2.5	2	2.5	3
1	2	2.5	3	2.5	3	3.5	3	3.5	4
1	2	2.5	3	2.5	3	3.5	3	3.5	4
2	1	1.5	2	1.5	2	2.5	2	2.5	3
2	1	1.5	2	1.5	2	2.5	2	2.5	3
2	1	1.5	2	1.5	2	2.5	2	2.5	3
3	0	0.5	1	0.5	1	1.5	1	1.5	2

(ii) The sampling distribution of $\bar{X}_1 - \bar{X}_2$

$(\bar{X}_1-\bar{X}_2)$	Tally	f	$f(\bar{x}_1-\bar{x}_2)$	$(\bar{X}_1-\bar{X}_2)f(\bar{x}_1-\bar{x}_2)$	$(\bar{X}_1-\bar{X}_2)^2 f(\bar{x}_1-\bar{x}_2)$
0	|	1	1/81	0	0
0.5	||	2	2/81	1/81	0.5/81
1	~~||||~~ ||	7	7/81	7/81	7/81
1.5	~~||||~~ ~~||||~~	10	10/81	15/81	22.5/81
2	~~||||~~ ~~||||~~ ~~||||~~ ||	17	17/81	34/81	68/81
2.5	~~||||~~ ~~||||~~ ~~||||~~ |	16	16/81	40/81	100/81
3	~~||||~~ ~~||||~~ ~~||||~~ |	16	16/81	48/81	144/81
3.5	~~||||~~ |||	8	8/81	28/81	98/81
4	||||	4	4/81	16/81	64/81
Total		81	1	189/81	504/81

Now $$\mu_{\bar{x}_1-\bar{x}_2} = \Sigma(\bar{x}_1-\bar{x}_2)f(\bar{x}_1-\bar{x}_2) = \frac{189}{81} = 2.33$$

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\Sigma(\bar{x}_1-\bar{x}_2)^2 f(\bar{x}_1-\bar{x}_2) - [\Sigma(\bar{x}_1-\bar{x}_2)f(\bar{x}_1-\bar{x}_2)]^2}$$

$$= \sqrt{\frac{504}{81} - \left(\frac{189}{81}\right)^2} = \sqrt{6.222-5.429} = \sqrt{0.7931} = 0.88$$

(iii) Verification:

Mean and variance of population 1:

$$\mu_1 = \frac{3+4+5}{3} = \frac{12}{3} = 4$$

$$\sigma_1^2 = \frac{\Sigma x_1^2}{N_1} - \left(\frac{\Sigma x_1}{N_1}\right)^2 = \frac{50}{3} - \left(\frac{12}{3}\right)^2 = 16.67 - 16 = 0.67$$

Mean and variance of population 2:

$$\mu_2 = \frac{1+1+3}{3} = \frac{5}{3} = 1.67$$

$$\sigma_2^2 = \frac{\Sigma x_2^2}{N_2} - \left(\frac{\Sigma x_2}{N_2}\right)^2 = \frac{11}{3} - \left(\frac{5}{3}\right)^2 = 3.67 - 2.789 = 0.88$$

By property:

$$\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2 = 4 - 1.67 = 2.33$$

$$\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} = \sqrt{\frac{0.67}{2} + \frac{0.88}{2}} = \sqrt{0.335+0.44} = 0.775 = 0.88$$

These are same as computed results.

5.4 Proportion

An expression which compares a characteristic with the total is called proportion i.e.

$$p = \frac{X}{N} \quad \text{(parameter)}$$

$$\hat{p} = \frac{X}{n} \quad \text{(estimator)}$$

Where X denotes the number of individuals having a specified characteristic and is a binomial random variable because each individual has two possibilities i.e. may or may not have the specified characteristic. Thus $X \sim B(n, p)$.

5.4.1 Sampling distribution of sample proportion $\hat{p}$

The probability distribution of all possible values of $\hat{p}$ is called sampling distribution of $\hat{p}$. Mean of the sampling distribution of $\hat{p}$ is denoted by $\mu_{\hat{p}}$ and it's S.E by $\sigma_{\hat{p}}$

5.4.2 Properties of the sampling distribution of $\hat{p}$

(i). Mean of the sampling distribution of sample proportion is always equal to the population proportion i.e. $\mu_{\hat{p}} = p$ (both in S.W.R and S.W.O.R cases)

(ii). S.E of the sampling distribution of $\hat{p}$ is:

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}} \quad , q=1-p \qquad \text{(in case of S.W.R)}$$

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}\frac{N-n}{N-1}} \qquad \text{(in case S.W.O.R)}$$

Note that if p is unknown and n is large, then $\sigma_{\hat{p}}$ can be estimated by

$$S_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n}}.$$

(iii). Shape of the sampling distribution of $\hat{p}$

For small n, its shape will be like binomial distribution but for sufficiently large sample sizes, the shape of the sampling distribution of $\hat{p}$ will be approximately normal.

The variable $\hat{p}$ can be standardized as $Z = \frac{\hat{p}-\mu_{\hat{p}}}{\sigma_{p}} = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}}$ (in S.W.R case)

5.4.3 Formulas for $\mu_{\hat{p}}$ and $\sigma_{\hat{p}}$.

Mean $\quad E(\hat{p}) = \mu_{\hat{p}} = \Sigma \hat{p} f(\hat{p})$

S.E $\quad \sigma_{\hat{p}} = \sqrt{\Sigma\, \hat{p}^2 f(\hat{p}) - [\Sigma\, \hat{p} f(\hat{p})]^2}$.

Example 5.8

Draw all possible samples of size n = 3 without replacement from the population consists of six numbers 3, 4, 5, 6, 7, 8 and find the proportion of even numbers in the samples. Construct the sampling distribution of sample proportion and verify that;

(a) $\mu_{\hat{p}} = p$

(b) $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}\frac{N-n}{N-1}}$

Solution:

The number of possible samples of size 3 in without replacement case are $^{6}C_{3}$ = 20. The samples and the proportion of even numbers in the samples are:

S.No.	Samples	$\hat{p}$	S.No.	Samples	$\hat{p}$	S.No.	Samples	$\hat{p}$
1	3, 4, 5	1/3	8	3, 6, 7	1/3	15	4, 6, 8	1
2	3, 4, 6	2/3	9	3, 6, 8	2/3	16	4, 7, 8	2/3
3	3, 4, 7	1/3	10	3, 7, 8	1/3	17	5, 6, 7	1/3
4	3, 4, 8	2/3	11	4, 5, 6	2/3	18	5, 6, 8	2/3
5	3, 5, 6	1/3	12	4, 5, 7	1/3	19	5, 7, 8	1/3
6	3, 5, 7	0	13	4, 5, 8	2/3	20	6, 7, 8	2/3
7	3, 5, 8	1/3	14	4, 6, 7	2/3			

The sampling distribution of $\hat{p}$

$\hat{p}$	f	$f(\hat{p})$	$\hat{p}f(\hat{p})$	$\hat{p}^2 f(\hat{p})$
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
Total	20	1	10/20	6/20

$$E(\hat{p}) = \mu_{\hat{p}} = \Sigma \hat{p} f(\hat{p}) = \frac{10}{20} = 0.5$$

$$\sigma_{\hat{p}} = \sqrt{\Sigma \hat{p}^2 f(\hat{p}) - [\Sigma \hat{p} f(\hat{p})]^2}$$

$$= \sqrt{\frac{6}{20} - \left(\frac{10}{20}\right)^2} = \sqrt{0.3 - 0.25} = \sqrt{0.05} = 0.22$$

Verification:

(i) As even numbers in the population are 3 therefore,

$$p = \frac{X}{N} = \frac{3}{6} = \frac{1}{2} = 0.5, \text{ hence proved that } \mu_{\hat{p}} = p$$

(ii) S.E in without replacement case by property is

$$\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}\frac{N-n}{N-1}} = \sqrt{\frac{(0.5)(0.5)}{3} \cdot \frac{6-3}{6-1}} = \sqrt{\frac{0.75}{15}} = \sqrt{0.05} = 0.22$$

This is same as computed value of $\sigma_{\hat{p}}$

5.5 Sampling distribution of difference between two sample proportions

The probability distribution of all possible values of $\hat{p}_1 - \hat{p}_2$ is called sampling distribution of $\hat{p}_1 - \hat{p}_2$. Its mean is denoted by $E(\hat{p}_1 - \hat{p}_2) = \mu_{\hat{p}_1 - \hat{p}_2}$ and S.E by $\sigma_{\hat{p}_1 - \hat{p}_2}$

5.5.1 Properties of the sampling distribution of difference between two sample proportions

(i). Mean of the sampling distribution of $\hat{p}_1 - \hat{p}_2$ is equal to the difference between the population proportions i.e.

$\mu_{\hat{p}_1 - \hat{p}_2} = p_1 - p_2$ (both in S.W.R and S.W.O.R)

(ii). S.E of the sampling distribution of $\hat{p}_1 - \hat{p}_2$ is

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}} \quad \text{(S.W.R case)}$$

$$= \sqrt{\frac{p_1 q_1}{n_1}\frac{N_1 - n_1}{N_1 - 1} + \frac{p_2 q_2}{n_2}\frac{N_2 - n_2}{N_2 - 1}} \quad \text{(S.W.O.R case)}$$

If $p_1 \neq p_2$ and also unknown, then for large n_1, n_2 they are replaced with the sample proportions $\hat{p}_1$ and $\hat{p}_2$ respectively. The $\sigma_{\hat{p}_1 - \hat{p}_2}$ is then estimated by

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

(iii). Shape of the sampling distribution of $\hat{p}_1 - \hat{p}_2$.

For sufficiently large n_1 and n_2 the sampling distribution of $\hat{p}_1 - \hat{p}_2$ is approximately normal.

The variable $\hat{p}_1 - \hat{p}_2$ in standard form is written as

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - \mu_{\hat{p}_1 - \hat{p}_2}}{\sigma_{\hat{p}_1 - \hat{p}_2}} = \frac{(\hat{p}_1 - \hat{p}_2) - (p_1 - p_2)}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}} \quad \text{(in S.W.R case)}$$

5.5.2 Formulas for mean and S.E of the sampling distribution of $\hat{p}_1 - \hat{p}_2$

$$\mu_{\hat{p}_1 - \hat{p}_2} = \Sigma(\hat{p}_1 - \hat{p}_2) f(\hat{p}_1 - \hat{p}_2)$$

$$\sigma_{\hat{p}_1 - \hat{p}_2} = \sqrt{\Sigma(\hat{p}_1 - \hat{p}_2)^2 f(\hat{p}_1 - \hat{p}_2) - [\Sigma(\hat{p}_1 - \hat{p}_2) f(\hat{p}_1 - \hat{p}_2)]^2}$$

Example 5.9

Let $\hat{p}_1$ represent the proportion of even numbers in samples of size $n_1 = 2$ drawn with replacement from a finite population consisting units 7, 8, 9. Let $\hat{p}_2$ represent the proportion of odd numbers in samples of size $n_2 = 2$ selected with replacement from another finite population having units 4, 5, and 5. Construct the sampling distribution of the difference between the sample proportions $\hat{p}_1 - \hat{p}_2$ and verify that $\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2$ and $\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$

Solution:

Given	population-1	population-2
	[7, 8, 9]	[4, 5, 5]
	$N_1 = 3$	$N_2 = 3$
	$n_1 = 2$	$n_2 = 2$

Since sampling is with replacement, therefore, possible samples are

$N_1^{n_1} = 3^2 = 9, \quad N_2^{n_2} = 3^2 = 9$

Population-1		
No	Samples	$\hat{p}_1$
1	(7, 7)	0
2	(7, 8)	1/2
3	(7, 9)	0
4	(8, 7)	1/2
5	(8, 8)	1
6	(8, 9)	1/2
7	(9, 7)	0
8	(9, 8)	1/2
9	(9, 9)	0

Population-2		
No	Samples	$\hat{p}_2$
1	(4, 4)	0
2	(4, 5)	1/2
3	(4, 5)	1/2
4	(5, 4)	1/2
5	(5, 5)	1
6	(5, 5)	1
7	(5, 4)	1/2
8	(5, 5)	1
9	(5, 5)	1

Now $9 \times 9 = 81$ possible differences are given in the following table.

$\hat{p}_2$ \ $\hat{p}_1$	0	$\frac{1}{2}$	0	$\frac{1}{2}$	1	$\frac{1}{2}$	0	$\frac{1}{2}$	0
0	0	$\frac{1}{2}$	0	$\frac{1}{2}$	1	$\frac{1}{2}$	0	$\frac{1}{2}$	0
$\frac{1}{2}$	$-\frac{1}{2}$	0	$-\frac{1}{2}$	0	$\frac{1}{2}$	0	$-\frac{1}{2}$	0	$-\frac{1}{2}$
$\frac{1}{2}$	$-\frac{1}{2}$	0	$-\frac{1}{2}$	0	$\frac{1}{2}$	0	$-\frac{1}{2}$	0	$-\frac{1}{2}$
$\frac{1}{2}$	$-\frac{1}{2}$	0	$-\frac{1}{2}$	0	$\frac{1}{2}$	0	$-\frac{1}{2}$	0	$-\frac{1}{2}$
1	−1	$-\frac{1}{2}$	−1	$-\frac{1}{2}$	0	$-\frac{1}{2}$	−1	$-\frac{1}{2}$	−1
1	−1	$-\frac{1}{2}$	−1	$-\frac{1}{2}$	0	$-\frac{1}{2}$	−1	$-\frac{1}{2}$	−1
$\frac{1}{2}$	$-\frac{1}{2}$	0	$-\frac{1}{2}$	0	$\frac{1}{2}$	0	$-\frac{1}{2}$	0	$-\frac{1}{2}$
1	−1	$-\frac{1}{2}$	−1	$-\frac{1}{2}$	0	$-\frac{1}{2}$	−1	$-\frac{1}{2}$	−1
1	−1	$-\frac{1}{2}$	−1	$-\frac{1}{2}$	0	$-\frac{1}{2}$	−1	$-\frac{1}{2}$	−1

The sampling distribution of $\hat{p}_1 - \hat{p}_2$

$\hat{p}_1 - \hat{p}_2$	f	$f(\hat{p}_1 - \hat{p}_2)$	$(\hat{p}_1 - \hat{p}_2) f(\hat{p}_1 - \hat{p}_2)$	$(\hat{p}_1 - \hat{p}_2)^2 f(\hat{p}_1 - \hat{p}_2)$
-1	16	$\frac{16}{81}$	$-\frac{16}{81}$	$\frac{16}{81}$
$-\frac{1}{2}$	32	$\frac{32}{81}$	$-\frac{16}{81}$	$\frac{8}{81}$
0	24	$\frac{24}{81}$	0	0
$\frac{1}{2}$	08	$\frac{8}{81}$	$\frac{4}{81}$	$\frac{2}{81}$
1	01	$\frac{1}{81}$	$\frac{1}{81}$	$\frac{1}{81}$
Total	81	1	$\frac{-27}{81} = -\frac{1}{3}$	$\frac{27}{81} = \frac{1}{3}$

Mean of sampling distribution is;

$$\mu_{\hat{p}_1-\hat{p}_2} = \Sigma(\hat{p}_1 - \hat{p}_2) f(\hat{p}_1 - \hat{p}_2) = -\frac{27}{81} = -\frac{1}{3}$$

S.E of sampling distribution is;

$$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\Sigma(\hat{p}_1 - \hat{p}_2)^2 f(\hat{p}_1 - \hat{p}_2) - [\Sigma(\hat{p}_1 - \hat{p}_2) f(\hat{p}_1 - \hat{p}_2)]^2}$$

$$= \sqrt{\frac{1}{3} - \left(-\frac{1}{3}\right)^2} = \sqrt{\frac{1}{3} - \frac{1}{9}} = \sqrt{\frac{2}{9}} = 0.47$$

Verification:

There is one even digit in population-1 so $p_1 = \frac{X_1}{N_1} = \frac{1}{3}$

There are two odd digits in population-2 so $p_2 = \frac{X_2}{N_2} = \frac{2}{3}$

Now $p_1 - p_2 = \frac{1}{3} - \frac{2}{3} = -\frac{1}{3} = \mu_{\hat{p}_1-\hat{p}_2}$.

Now by property, the S.E in S.W.R case is

$$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}} = \sqrt{\frac{\left(\frac{1}{3}\right)\left(\frac{2}{3}\right)}{2} + \frac{\left(\frac{2}{3}\right)\left(\frac{1}{3}\right)}{2}} = \sqrt{\frac{1}{9} + \frac{1}{9}} = \sqrt{\frac{2}{9}} = 0.47.$$

Example 5.10

Suppose that 43% of men and 32% of women of a city are in favor of a proposed recreational facility. Describe the sampling distribution of the difference between sample proportions from samples of sizes 350 men and 270 women.

Solution:

Let proportion of men is denoted by p_1 and proportion of women by p_2, then given information is symbolized as;

$p_1 = 43\% = 0.43$ $\quad$ $p_2 = 32\% = 0.32$

$q_1 = 1 - P_1 = 1 - 0.43 = 0.57$ $\quad$ $q_2 = 1 - P_2 = 1 - 0.32 = 0.68$

$n_1 = 350$ $\quad$ $n_2 = 270.$

Describing the sampling distribution of the difference between sample proportions means to find its mean, S.E and the shape.

By property we know that:

$$\mu_{\hat{p}_1-\hat{p}_2} = p_1 - p_2 = 0.43 - 0.32 = 0.11$$

$$\sigma_{\hat{p}_1-\hat{p}_2} = \sqrt{\frac{p_1q_1}{n_1} + \frac{p_2q_2}{n_2}}$$

$$= \sqrt{\frac{(0.43)(0.57)}{350} + \frac{(0.32)(0.68)}{270}} = \sqrt{0.0007 + 0.0008} = \sqrt{0.00151} = 0.04$$

Since the samples are sufficiently large, it means that sampling distribution of $\hat{p}_1 - \hat{p}_2$ has approximately normal distribution with mean 0.11 and standard deviation 0.04.

Key points

- A group of individuals or objects about which we wish to know something is called population or universe.
- If the number of individuals or items of a population are fixed and limited, it is known as finite population.
- If the population consists of an infinite number of items it is called infinite population.
- A single element or group of elements of a population from which required information can be obtained is called sampling unit or unit of the population.
- A part or fraction of a population is called sample.
- To ask a question or a series of questions from many people in order to gather information is called survey.
- A list of all the sampling units or a map or other acceptable material which represent the population to be covered. Such a list or map is called sampling frame.
- The sample design along with some other aspects of the survey e.g. choice and training of interviewers, tabulation plans etc. is called survey design.
- Any function of population data is called parameter.
- Any function of sample data is called statistic.
- A statistic which is used for estimation of parameter is called an estimator.
- Every estimator is a statistic but every statistic is not an estimator.
- The error which occurs due to the natural differences among the members of the population is called sampling error.
- The error which occurs during the process of collection and processing the data is called non sampling error.
- The error which arises due to the personal interest of the investigator is called bias.
- The probability distribution of all possible values of an estimator is called sampling distribution of that estimator.
- The standard deviation of the sampling distribution of an estimator is called standard error
- An expression which compares a characteristic with the total is called proportion

Exercise

5.1 Read the fallowing statements carefully and indicate which statement is true or false.

i. To perform a census, one would need to examine every item in a population under consideration.
ii. A primary objective of sampling is to choose a sample that is representative of the population under consideration.
iii. There is no difference between random error and sampling bias.
iv. An estimator or statistic is characteristic of population.
v. The sample mean and sample standard deviation are not statistics.
vi. As the sample size increases, the S.E decreases.
vii. A sampling procedure that selects items for a sample at uniform intervals is called stratified random sampling.
viii. The probability distribution of all possible means of samples is called sampling distribution of sample mean.
ix. The S.E of the mean is as the S.D of the sampling distribution of the sample mean.
x. S.E can be negative.

5.2 Fill in the blanks.

i. If the number of units in a population is limited, it is known as ______ population.
ii. A population consisting of an unlimited number of units is called a ______ population.
iii. If all units of a population are surveyed, it is called ________.
iv. The discrepancy between a parameter and its estimator due to sampling process is known as ________.
v. Standard deviation of all possible estimates from samples of fixed size is called _______.
vi. The list of all the items of a population is known as ________.

vii. Under simple random sampling with replacement the same unit can occur ________ in the sample.

viii. The expression $\frac{n}{N}$ is known as ________.

ix. The quantity $\sqrt{\frac{N-n}{N-1}}$ is called ________.

x. fpc is dropped from the S.E formula if ________.

5.3 Choose the correct answer.

i. A characteristic of population is called

(a) parameter (b) statistic

(c) sample (d) constant

ii. As the sample size increases the standard error of the sampling distribution of mean:

(a) increases (b) decreases

(c) remains the same (d) becomes negative

iii. Probability of selection varies at each subsequent draw in:

(a) S.W.R (b) S.W.O.R

(c) both (a) and (b) (d) neither (a) and (b)

iv. Which of the following statement is true?

(a) S.E is always zero (b) S.E is always unity

(c) more the S.E, better it is (d) less the S.E, better it is

v. fpc can be dropped from the S.E formula if

(a) $n < 5\% N$ (b) $n < 0.05 N$

(c) $\frac{n}{N} < 5\%$ (d) all of the above

vi. If $\mu = 40, \sigma = 10, n = 25$, then value of $\sigma_{\bar{x}}$ is

(a) 40 (b) 10 (c) 2 (d) 0.4

vii. The number of random samples of size 2 that can be selected W.O.R from a population having 7 items are equal to

(a) 12 (b) 49 (c) 21 (d) 14

viii. A characteristic of sample is called

(a) parameter (b) sample (c) statistic (d) constant

ix. The finite population correction factor is

(a) $\sqrt{\frac{n-1}{n-N}}$ (b) $\sqrt{\frac{N-1}{N}}$ (c) $\sqrt{\frac{N-n}{n-1}}$ (d) $\sqrt{\frac{N-n}{N-1}}$

x. An estimator is a

(a) variable (b) constant

(c) fixed value (d) random variable

5.4 What is meant by population, sample and sampling?

5.5 What is the need of sampling as compared to complete enumeration?

5.6 Write short notes on:

(i) sampling frame (ii) sample design (iii) survey

5.7 Differentiate between:

(i) Finite and infinite population.

(ii) Sampling with and without replacement

(iii) Sampling and non-sampling errors

5.8 Match the symbols on the left with the phrase on the right.

symbol	phrase
μ	Sample mean
$\bar{X}$	Sample variance
σ^2	Population proportion
S^2	Population variance
p	Sample proportion
$\hat{p}$	population mean

5.9 (a) What are the main objectives of sampling?

(b) Define parameter, statistic and estimator.

5.10 Explain sampling and non-sampling errors. What are the methods of reducing these errors?

5.11 Differentiate between probability and non- probability sampling.

5.12 Explain the following methods of selecting a sample:

(i) Simple random sampling

(ii) Stratified random sampling

(iii) Systematic random sampling

5.13 Define a simple random sample. How a random sample could be selected by:

(i) Goldfish bowl method

(ii) Random number table method

5.14 Define the following types of samples:

(i) Random sample (ii) Non-random sample

(iii) Stratified random sample. (iv) Systematic random sample.

5.15 Select a random sample of 10 cities from a list of 200 cities by using a random numbers table.

5.16 (a) Describe stratified random sampling and proportional allocation methods.

(b) The grades in an inter examination of a college were as follows:

Grade	A	B	C	D
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?

5.17 A stratified random sample of size $n = 200$ is to be taken from a population of size $N = 40{,}000$ divided in to five strata of sizes $N_1 = 15{,}000$, $N_2 = 10{,}000$, $N_3 = 5000$,

$N_4 = 8000$ and $N_5 = 2000$. If the allocation is to be proportional, how large a sample must be taken from each stratum?

5.18 Describe the concept of

(i) Sampling distribution

(ii) Standard error

(iii) Finite population correction factor

5.19 (a) Define the sampling distribution of sample mean and describe its properties.

(b) A population has five elements 4, 5, 6, 7, 8. Draw all possible samples of size 2 with replacement and compute mean for each sample. Construct the sampling distribution of $\bar{X}$ and calculate its mean and S.E also verify that $\mu_{\bar{x}} = \mu$ and

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

5.20 A finite population consists of the numbers 2, 2, 4, 6, 5. Select possible random samples of size 2 without replacement and find their means. Construct the sampling distribution of the sample mean and verify that $\mu_{\bar{x}} = \mu$ and

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

5.21 A population has a mean of $\mu = 50$ and a standard deviation of $\sigma = 12$. For samples of size $n = 4$, what is the mean (expected value) and the standard deviation (standard error) for the distribution of sample mean?

5.22 For a population with a mean of $\mu = 40$ and a standard deviation of $\sigma = 8$, find the Z-score corresponding to a sample mean of $\bar{x} = 44$ for each of the following sample sizes:

(i) $n = 4$(ii) $n = 16$

5.23 Given the five-element population 4, 5, 7, 9, 10.

(i) Compute population mean and variance.

(ii) Suppose samples of size $n = 3$ are selected without replacement but you compute $\sigma_{\bar{x}}^2$ directly.

(iii) Suppose samples of size $n = 3$ are selected with replacement but you compute $\sigma_{\bar{x}}^2$ directly.

5.24 Draw all possible samples of size $n = 3$ without replacement from the population 0, 3, 6, 12, 15, 18. Construct the sampling distribution of the sample mean and verify the relation between:

(i) Mean of the sampling distribution and the population mean.

(ii) Standard deviation of the sampling distribution of the mean and the population standard deviation.

5.25 Define sampling distribution of difference between means of two samples. Describe its important properties.

5.26 Draw all possible random samples of size $n_1 = 2$ with replacement from the finite population consisting of –2, 0, 2/and 4. Similarly draw all possible random samples of size $n_2 = 2$ with replacement from the population –1 and +1.

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

(ii) Construct the sampling distribution of $\bar{X}_1 - \bar{X}_2$.

(iii) Verify that $\mu_{\bar{x}_1-\bar{x}_2} = \mu_1 - \mu_2$ and $\sigma_{\bar{x}_1-\bar{x}_2} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$

5.27 Draw all possible random samples of size $n_1 = 2$ without replacement from a finite population consisting of 3, 6, 9. Similarly draw all possible random samples of size $n_2 = 2$ without replacement from another finite population consisting of 2, 4, 6.

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

(ii) Construct the sampling distribution of $\bar{X}_1 - \bar{X}_2$ and compute its mean and variance. Also verify the theoretical results.

5.28 What is meant by proportion and sampling distribution of sample proportion $\hat{p}$? Describe its important properties.

5.29 Draw all possible samples of size $n = 3$ without replacement from the population 2, 3, 3, 4, 5, 6 and find the sample proportion $\hat{p}$ of odd numbers in the samples. Construct the sampling distribution of sample proportion and verify that $\mu_{\hat{p}} = p$ and $Var(\hat{p}) = \frac{pq}{n}\left[\frac{N-n}{N-1}\right]$

5.30 The marital status of a population of seven friends is U, M, M, U, M, U, U where U and M stand for unmarried and married respectively. Find the proportion of married friends in the population. Take all possible samples of two friends without replacement from this population and find the proportion of married friends in each sample. Make the sampling distribution of the sample proportion and verify that $\mu_{\hat{p}} = p$ and $\sigma_{\hat{p}}^2 = \frac{pq}{n}\left(\frac{N-n}{N-1}\right)$.

5.31 Let $\hat{p}_1$ 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, 9. Similarly, let $\hat{p}_2$ 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, 2, 5. Form a sampling distribution of $\hat{p}_1 - \hat{p}_2$ and verify that:

(i) $E(\hat{p}_1 - \hat{p}_2) = p_1 - p_2$

(ii) $Var(\hat{p}_1 - \hat{p}_2) = \frac{p_1 q_1}{n_1}\frac{N_1 - n_1}{N_1 - 1} + \frac{p_2 q_2}{n_2}\frac{N_2 - n_2}{N_2 - 1}$