CHAPTER - I

1.1 Introduction :-

In the conventional sampling methods such as simple random sampling, Stratified sampling, Systematic sampling, Cluster sampling, one selects 'n' samples from populations; studies each unit in the sample and draws conclusions about the population. But these procedures may not be appropriate when for example -

i) The polulation is very large (bulk),
ii) The charactristic under study is rare,
iii) The method of selecting sample and taking measurement on it is tedious, costly or time consuming.

In such cases another sampling procedure known as "**composite sampling** " may be used. Some situations where the use of composite sampling procedure are :

i) Testing blood for some disease which is rare.
ii) To obtain overall or average fertility of soil.
iii) Testing ferilizers to estimate percentage of nitrogen.
iv) Testing water for estimating percentage of a pollutant.
v) Testing fish for estimating the lavel of an toxicant.
vi) To estimate abundance of various plankton species in filtered water.

In order to describe the composite sampling procedure the following definations are needed.

i) Lot :-

A lot is the mass of material of which the mean charactristic is to be estimated, such as a truck-load of bags, drums of oil, bales of wool.

SHIVAJI UNIVERSITY LIBRARY KOLHAPUR

ii) **Segmented lot** : -

A lot which occurs in partitioned form such as packages is known as a segmented lot. For example bales of wool, drums of oil, bags of fertilizers.

iii) **Non-Segmentes lot** : -

A lot which does not occur in packages is known as a non-segmented lot. Such as a lake, coal, soil.

Some non-segmented lots can be made segmented for example, coal can be partitioned into bags, soil can be divided into plots.

iv) **Increment** :-

The initial portion (sample) drawn from a lot is known as sample increment or increment.

v) **Composite Sample** :-

A mixture of increments is known as a composite sample.

vi) **Blending** :-

A method of mixing increments to form a composite sample is known as blending.

vii) **Reduction** :-

Reduction is a method with the help of which a sub sample is obtained from a sample. It is used to obtain a laboratory sample from a composite sample.

viii) **Laboratory sample** :-

A sample that is sent to the laboratory for testing is known as the laboratory sample. It is obtained by reduction of composite sample, suitable for laboratory analysis.

ix) **Test unit** :-

The smallest sampling unit on which direct observations wiil be obtained is known as a test unit.

x) **Analysis** :-

The rule of measuring physical, biological or chemical charactristics of the lot is known as analysis.

xi) **Types of Variations** :-

The various variations in the procedure are as follows.

a) Variation due to selection of original increments.
b) Variation due to blending and reduction.
c) Variation due to laboratory sampling and test measurment.

The variation due to (a) above is known a sampling variance, due to (b) above is known as reduction variance and due to (c) above is known as measurment Variance.

xii) **Definition of mean** :-

If a lot were exhaustively divided to "M" similar test units, and if analysis were performed on all "M" test units, then the mean of results on all "M" test units will be defines as the mean of the lot.

The main object of composite sampling procedure is to estimate the mean charactristic of a population with minimum time

or cost. Having an estimate, its properties are to be studied. For this purpose appropriate model for composite sampling procedure is to be developed.

The general procedure of composite sampling can be explained as in the following steps.

Step - I : Populations contains "N" segments of material, select "n" segments from "N" segments by any appropriate conventional sampling method.

Step - II : Select "m" increments from each selected segment. So that there are in all "mn" increments.

Step - III : Form "r" composites randomly from "mn" increments each of size mn/r. This demands that "mn" is a multiple of "r" otherwise the composite sample may not be of equal size.

Step -IV : Select "s" laboratory samples from each composite sample. So that there are in all "rs" laboratory samples.

Step -V : Take t test units from each laboratory sample. So that there are "rst" test units; and take measurment on each test unit. Based on these "rst" measurments, an estimate is computed. The constants (N, n, m, r, s, t) are known as parameters of composite sampling procedure.

Duncan (1964) first introduced a linear model for composite sampling of bulk material such as bags of fertilizers, based on which an estimate of the mean of a charactristic is derived. He Considered composite sampling procedure with parameters (N, n, 1, r, 1, 1) and proposed linear model for analysis of the various variances.

Brown and Fisher (1972) and Rohde (1976) considered different models for both segmented and non-segmented lots. Which are based on random weights of increments. Brown and Fisher considered the composite sampling procedure with

parameters (N, n, k, l, r, l) and proposed a product model for segmented material. Rohde (1976) considered the product model for composite sampling procedure with parameters (ø ,n, l, r, s, k) for non-segmented material.

These models are discussed in detail in later chapters.

1.2 Use of Composite Sampling :

Here are some situations where the composite sampling is appropriate.

(1) Composite sampling for the detection of certain rare trait :-

Composite sampling procedure is used to detect a certain rare trait present in population. For example testing of blood for the presence of a rare disease. Testing water for the presence of a rare substance like Mercury.

Suppose "n" increments from different places are selected, mix these increments to form one composite sample and make a test on the composite sample.

If a trait is present in the composite sample, then it is present in any of individual increment and hence individual test on each increments is required to find in which it is present.

If a trait is not present in the composite sample (Which is more likely since the trait is rare), then it is not present in any of the increments and hence individual tests are not required.

Remarks :-

If a trait is present in the composite sample, then make two composite samples and take tests on each. If it is present in one or both, repeat the above procedure for each composite in

which the trait is present.

Let Ii (i= 1,2,3,....n) be results on n increments and

Let

$$Ii = \begin{cases} 1 & \text{if trait is present in } i^{th} \text{ increment} \\ \emptyset & \text{otherwise} \end{cases}$$

Suppose Ii are IID random variables with P(Ii = 1) = P

Then result on composite sample can be written as

$$I = 1 - \prod_{i=1}^{n} (1 - Ii)$$

Hence, Probability that the composite contains a trait is

$$P(I=1) = 1 - P(I_1=\emptyset,\ I_2=\emptyset, \ldots\ldots I_n = \emptyset)$$

$$= 1 - (1 - P)^n$$

<u>**for example**</u> :

If 1Ø increments are to be combined to form a composite sample and the composite sample indicates the presence of a trait, Then it remains to identify which of the 1Ø increments has a trait. The simplest way is to test all 1Ø increments. If N^* denotes the number of tests to be performed is given by

$$E(N^*) = 11\ (\ 1 - (1 - P)^{1\emptyset}) + (1-P)^{1\emptyset}$$

if p = Ø.ØØØ1 then

$$E(N^*) = 1.\emptyset 89$$

Therefore, expected numbers of tests required to determine the sample which has a trait is 1.Ø9 as against the 1Ø to be performed if the composite sampling is not used.

In general for n increment in the composite, the expected number of tests required for procedure to identify which of "n" increments have trait is

$$E(N^*) = (n+1)(1 - (1 - P)^n) + (1-P)^n$$

$$= (n+1) - n(1-P)^n$$

The relative cost factor (R.C.F.) is defined as the ratio of expected number of tests required for composite sampling to the number of tests required without compositing

Therefore,

$$\text{Relative cost factor(R.C.F.)} = \frac{E(N^*)}{N^*} = \frac{(n+1)-n(1-P)^n}{n}$$

$$= 1 + (1/n) - (1-P)^n$$

A program in BASIC to find the minimum value of "P" that gives the corroseponding optimum value of "n" which gives minimum R.C.F. is given below. The output in tabular form is given in table 1.1. If the value of "P" is increased from one of these values, the optimum value of "n" will be unchanged until the next table value of "P". The R.C.F. values for given value of "P" is also given in table. For example,

P	n	rcf	n+1	rcf
Ø.125	3	Ø.6634	4	Ø.6638
Ø.1241	3	Ø.6611	4	Ø.6614
Ø.124Ø	3	Ø.661Ø7	4	Ø.6611
Ø.1239	3	Ø.66Ø881	4	Ø.66Ø86

Table 1.1

n	P	R.C.F.	n	P	R.C.F.
3	Ø.124Ø	66	25	Ø.ØØ17	8
4	Ø.Ø656	48	26	Ø.ØØ16	8
5	Ø.Ø412	39	26	Ø.ØØ14	7
6	Ø.Ø283	33	28	Ø.ØØ13	7
7	Ø.Ø2Ø7	28	29	Ø.ØØ12	7
8	Ø.Ø158	25	3Ø	Ø.ØØ116	7
9	Ø.Ø125	22	31	Ø.ØØ11	7
1Ø	Ø.Ø11Ø	2Ø	32	Ø.ØØ1Ø1	6
11	Ø.ØØ84	18	33	Ø.ØØØ92	6
12	Ø.ØØ7Ø	16	34	Ø.ØØØ87	6
13	Ø.ØØ6Ø	15	35	Ø.ØØØ82	6
14	Ø.ØØ52	14	36	Ø.ØØØ78	6
15	Ø.ØØ45	13	37	Ø.ØØØ74	5
16	Ø.ØØ4Ø	13	38	Ø.ØØØ7Ø	5
17	Ø.ØØ35	12	39	Ø.ØØØ66	5
18	Ø.ØØ31	11	4Ø	Ø.ØØØ63	5
19	Ø.ØØ28	11	5Ø	Ø.ØØØ41	4
2Ø	Ø.ØØ26	1Ø	6Ø	Ø.ØØØ28	3
21	Ø.ØØ23	1Ø	7Ø	Ø.ØØØ21	3
22	Ø.ØØ21	9	8Ø	Ø.ØØØ16	3
23	Ø.ØØ19	9	9Ø	Ø.ØØØ125	2
24	Ø.ØØ18	8	1ØØ	Ø.ØØØ11	2

BARR. BALASAHEB KHARDEKAR LIBRARY
SHIVAJI UNIVERSITY, KOLHAPUR

2) Estimation for Incidence of a trait with composite sampling for confidentiality :

When testing for a trait is likely to cause embarrassment to an individual, then confidentiality can be assured by composite sampling procedure.

Let "**P**" be the probability that an increment contains a trait. If a composite sample is formed from "n" increments then probability that the composite contains the trait is $1-(1-P)^n$. If m composites are tested and if "x" composites are found containing the trait, then

$$P_m = x/m$$

is an estimate of "**P**" has binomial distribution with parameters m and $1-(1-P)^n$.

Hence an estimate of P is given by

$$\hat{P} = 1 - (1 - P_M)^{1/n}$$

Using Jensen's inequality for expectation,
it follows that

$$\begin{aligned} E(\hat{P}) &> 1 - (1 - E(P_m))^{1/n} \\ &= 1 - (1 - (1-(1-P_m))^n)^{1/n} \\ &= 1 - (1 - P) \\ &= P \end{aligned}$$

$$E(\hat{P}) \geq P$$

Hence estimate of P has a positive bias.

3) Composite sampling to test for compliance : -

If we have to find items that does not contain a particular rare trait more than at a fixed level. For example level of intoxicant in fish, level of pollutant in water etc. then we use composite sampling procedure.

Let "c" be the criterion level that should not be exceeded. If level exceeds "c" then the product can not be used. If the composite is formed from equal sized increments, then measure on the composite is an average of measure on individual samples. In this case if any sample exceeds "c" then a composite sample of n individual samples will or will not exceeds c/n. But if none of the individual samples exceeds c then the composite sample will not exceeds c/n .Hence if the composite sample exceess c/n then it is necessary to test each individual sample. On the other hand, if the composite sample does not exceed c/n then none of the individual samples needs to be tested.

4) To reduce variance in the presence of inexact analytical procedure : -

Suppose composite sample consists of "n" individual increments. If we test individual increments then the result of each test has an error with mean zero and variance 6^2t. So that variance of the sample mean ($\bar{x}$) is $(6^2 + 6^2t)/n$. Where 6^2 is population variance. But the variance of the composite sample is $(6^2/n) + 6^2t$. Hence variance of composite sample is larger than the variance of an average of n increments. But if we form "K" equal sized composite samples from "n" increments then variance of average of K composite samples is $(6^2/n)+(6^2t)/k$. Which is less than the variance of an average of "n" increments for $n>k$. Besides a composite sampling reduces the cost of estimation.

5) To reduce the cost of estimating the mean

The usual method of estimating the a stocastic process is to take n observations and calculate the sample mean. Nearly the same information can be obtained by mixing "n" observations and taking single measurment on the composite sample. Some times laboratory procedures are time consuming and costly. In such cases composite sampling procedure is useful.

let $X_1, X_2, \ldots\ldots X_N$ be measurments on "n" increments which are iid random variables with mean u and variance 6^2. Let g_1 g_2 measurment on composite sample then

$$Z = \sum_{i=1}^{n} WiXi \quad \text{Where} \quad Wi = gi/\sum gi$$

$$\begin{aligned} E(Z) &= \sum_{i=1}^{n} WiE(Xi) \\ &= \sum Wiu \\ &= u \end{aligned}$$

Hence Z is an unbiased estimate of u and

$$\begin{aligned} Var(Z) &= Var(\sum WiXi) \\ &= \sum_{i=1}^{n} Wi^2 \, Var(Xi) \\ &= \sum_{i=1}^{n} Wi^2 6^2 \end{aligned}$$

Now suppose that each composite consists of the same amount of material then $Wi = 1/n$ and $Z = \bar{X}$

Therefore, estimate based on composite is same as that based on "n" samples.

1.3 Chapter wise Summary

In chapter -I we give some definitions conserned with the composite sampling procedure and uses of composite sampling procedure.

In chapter -II we study composite sampling procedure for "Bulk material" introduced by Ducan (1962) with parameters (N,n,1,r,1,1). Here observation on test unit is represented by the statistical model

$$Y_i = Y + c + \Sigma d_i + \Sigma r_i + \Sigma t_i + \Sigma e_i \qquad i = 1,2,3,4...r$$

Where,

- Y - Population mean
- c - Deviation of mean of a lot from mean of population
- d_i - Deviation of mean of ith composite from mean of a lot
- r_i - Deviation of mean of ith laboratory sample from mean of ith composite sample.
- t_i - Deviation of ith test unit from mean of ith laboratory sample
- e_i - random error.

By using this model we obtain an unbiased estimate of population mean and its variance. Further the work is extended to obtain optimum values of n, r, k which gives maxmimum net gain, where k is proportion of composite sample in laboratory sample.

In chapter -III we study composite sampling procedure for segmented and nonsegmented population. In section I of this chapter we consider composite sampling procedure for segmented population given by Brown and Fisher (1972) with parameters (N,J,K,1,r,1). The observation on random subsample is given by

$$Yi = \sum_{j}^{J} \sum_{k}^{K} a_{ijk} X_{jk}$$

Where a_{ijk} is proporation of kth increments of jth segment incorporated in i^{th} subsample and X_{jk} is observation on k^{th} increments of jth segment. Here we obtain unbiased estimate of population mean and its variance.

In section II of this chapter we consider composite sampling procedure for non-segmented population given by Rohde (1976) with parameters (∞, n, l, r, s, k) the observation on composite sample is given by

$$Zi = \mathbf{ai'y}$$

Where **ai** is proporation vector of ith composite sample. Here also we obtain unbiased estimate of population mean and its variance. The results verified by using Dirichiet distribution. Finally we study the assymptotic distribution of mean of composite sample.