This sampling method is seldom used and cannot be recommended for general use since it is often biased due to element of subjectiveness or the part of the investigator. However, if the investigator is experienced and skilled and this sampling is carefully applied, then judgment samples may yield valuable results.

7·8·2. Probability Sampling. Probability sampling is the scientific method of selecting samples according to some laws of chance in which *each unit in the population has some definite pre-assigned probability of being selected in the sample.* The different types of probability sampling are :

(*i*) Where each unit has an *equal chance* of being selected.

(*ii*) Sampling units have different probabilities of being selected.

(*iii*) Probability of selection of a unit is proportional to the sample size.

7·8·3. Mixed Sampling. If the samples are selected partly according to some laws of chance and partly according to a fixed sampling rule (no assignment of probabilities), they are termed as *mixed samples* and the technique of selecting such samples is known as *mixed sampling.*

The different types of sampling as given above have a number of variations, some of which may be listed below :

(*i*) *Simple Random Sampling*
(*ii*) *Stratified Random Sampling*
(*iii*) *Systematic Sampling*
(*iv*) *Multistage Sampling*
(*v*) *Quasi Random Sampling*
(*vi*) *Area Sampling*
(*vii*) *Simple Cluster Sampling*
(*viii*) *Multistage Cluster Sampling*
(*ix*) *Quota Sampling*

7·9. SIMPLE RANDOM SAMPLING (S.R.S.)

It is the technique of drawing a sample in such a way that *each unit of the population has an equal and independent chance of being included in the sample.*

In this method, an equal probability of selection is assigned to each unit of the population at the first draw. It also implies an equal probability of selecting any unit from the available units at subsequent draws.

Thus in S.R.S. from a population of N units, the probability of drawing any unit at the first draw is $1/N$, the probability of drawing any unit in the second draw from among the available $(N-1)$ units, is $1/(N-1)$, and so on.

Let E_r be the event that any specified unit is selected at the rth draw. Then

$P(E_r)$ = Prob. [that the specified unit is not selected in anyone of the previous $(r-1)$ draws and then selected at the rth draw]

$$\therefore \quad P(E_r) = \prod_{i=1}^{r-1} P\text{ (It is not selected at } i\text{th draw.)}$$

$\times$ P[It is selected at rth draw given that it is not selected at the previous $(r-1)$ draws]

(By compound probability theorem, since draws are independent)

$$\therefore \quad P(E_r) = \prod_{i=1}^{r-1} \left[1 - \frac{1}{N-(i-1)}\right] \times \frac{1}{N-(r-1)} = \prod_{i=1}^{r-1} \left(\frac{N-i}{N-i+1}\right) \times \frac{1}{N-r+1}$$

Scanned with CamScanner

$$= \frac{N-1}{N} \times \frac{N-2}{N-1} \times \frac{N-3}{N-2} \times \ldots \times \frac{N-r+1}{N-r+2} \times \frac{1}{N-r+1} = \frac{1}{N}$$

$$\therefore \quad P(E_r) = \frac{1}{N} = P(E_1) \qquad \ldots (7\cdot1)$$

This leads to a very interesting and important property of *Simple Random Sampling without Replacement* (srswor), *viz.*,

"The probability of selecting a specified unit of the population at any given draw is equal to the probability of its being selected at the first draw."

7·9·1. Probability of Selecting Any Specified Unit in the Sample. Since a specified unit can be included in the sample of size n in n mutually exclusive ways, *viz.*, it can be selected in the sample at the rth draw ($i = 1, 2, \ldots, n$) and since

$$P(E_r) = \frac{1}{N};\ r = 1, 2, \ldots, n$$

The probability that a specified unit is included in the sample

$$= \sum_{r=1}^{n} \left(\frac{1}{N}\right) = \frac{n}{N}, \text{ (by the addition theorem of probability)} \qquad \ldots (7\cdot2)$$

Remark. Simple Random Sampling can also be defined equivalently as follows :

Let us suppose that a sample of size n is drawn from a population of size N. There are ${}^NC_n$ possible samples. *S.R.S. is the technique of selecting the sample in such a way that each of ${}^NC_n$ samples has an equal chance or probability* $p = (1/{}^NC_n)$ of being selected, as explained below :

In S.R.S.,

Probability of selecting any unit at the first draw $= \dfrac{1}{N}$

Prob. of selecting any unit out of the remaining $(N-1)$ units in the second draw $= \dfrac{1}{N-1}$

and so on.

Probability of selecting any unit of the remaining $N-(i-1)$ units at the ith draw

$$= \frac{1}{N-(i-1)}, \ (i = 3, 4, \ldots, n)$$

Since all the draws are independent, by compound probability theorem, the probability of selecting a sample of size n in a fixed specified order, is $[1/N(N-1)(N-2)\ldots(N-n+1)]$.

Since this probability is independent of the order of the sample and since there are n ! permutations of the sample units, by addition theorem of probability, the required probability of obtaining a sample of size n (in any order) is given by : $\qquad \ldots (7\cdot3)$

$$p = \frac{n\,!}{N(N-1)\ldots(N-n+1)} = \frac{1}{{}^NC_n}$$

as required.

7·9·2. Selection of a Simple Random Sample. Random sample refers to that method of sample selection in which every item has an equal chance of being selected. But the random sample does not depend upon the method of selection only but also on the size and nature of the population. Some procedure which is simple and good for small population is not so for the large population. Generally, the method of selection should be independent of the properties of sampled population. Proper care has to be taken to ensure that selected

Scanned with CamScanner

sample is random. Human bias, which varies from individual to individual, is inherent in any sampling scheme administered by human beings. Random sample can be obtained by any of the following methods :

(a) *Lottery system Method*

(b) '*Mechanical Randomization*' or '*Random Numbers*' method.

We shall discuss these methods below.

(a) ***Lottery System.*** The simplest method of selecting a random sample is the *lottery system,* which is illustrated below by means of an example :

Suppose we want to select 'r' candidates out of n. We assign the numbers 1 to n; one number to each candidate and write these numbers (1 to n) on n slips which are made as homogeneous as possible in shape, size, colour, etc. These slips are then put in a bag and thoroughly shuffled and then 'r' slips are drawn one by one. The 'r' candidates corresponding to numbers on the slips drawn, will constitute a random sample.

This method of selection is quite independent of the properties of population. Generally in place of chits, cards are used. We make one card correspond to one of the units of the population by writing on it the number of the unit. The pack of cards is a kind of miniature of the population for sampling purposes. The cards are shuffled a number of times and then a card is drawn at random from them. This is one of the most reliable methods of selecting a random sample.

(b) ***'Mechanical Randomization' or 'Random Numbers' Method.*** The lottery method described above is quite time consuming and cumbersome to use if the population is sufficiently large. The most practical and inexpensive method of selecting a random sample consists in the use of '*Random Number Tables*', which have been so constructed that each of the digits 0, 1, 2, ..., 9 appears with approximately the same frequency and independently of each other. If we have to select a sample from a population of size $N(\leq 99)$ then the numbers can be combined two by two to give pairs from 00 to 99. Similarly if $N \leq 999$ or $N \leq 9999$, and so on, then combining the digits three by three (or four by four, and so on), we get numbers from 000 to 999 or (0000 to 9999), and so on. Since each of the digit 0, 1, 2, ..., 9 occurs with approximately the same frequency and independently of each other, so does each of the pairs 00 to 99 or triplets 000 to 999 or *quadruplets* 0000 to 9999, and so on.

The method of drawing the random sample consists in the following steps :

(*i*) Identify the N units in the population with the numbers from 1 to N.

(*ii*) Select at random, any page of the '*random number table*' and pick up the numbers in any row or column or diagonal at random.

(*iii*) The population units corresponding to the numbers selected in step (*ii*) constitute the random sample.

We give below different sets of random numbers commonly used in practice. The numbers in these tables have been subjected to various statistical tests for randomness of a series and their randomness has been well established for all practical purposes :

1. *Tippet's (1927) Random Numbers Tables.* (Tracts for computers No. 15, Cambridge University Press).

Tippet number tables consist of 10,400 four digited numbers, giving in all 10,400 × 4, *i.e.*, 41,600 digits selected at random from the British census reports.

2. *Fisher and Yates (1938) Tables (in Statistical tables for Biological, Agricultural and Medical Research)* comprise 15,000 digits arranged in twos. *Fisher* and *Yates* obtained these tables by drawing numbers at random from the 10th to 19th digits of *A.S. Thomson's* 20-figure logarithmic tables.

3. *Kendall and Babington Smith's (1939)* random tables consists of 1,00,000 digits grouped into 25,000 sets of 4 digited random numbers (Tracts for computers, No. 24, Cambridge University Press).

4. *Rand Corporation (1955) (Free Press, Illinois)* random number tables consist of one million random digits consisting of 2,00,000 random numbers of 5 digits each.

***Example 7·1.** Draw a random sample (without replacement) of size 15 from a population of size 500.*

Solution. First of all, we identify the 500 units in the population with the numbers from 1 to 500. Then we select at random one page of random numbers from any of the random number series discussed in § 7·9·2. Starting at random with any number on that page and moving row-wise, column-wise or diagonally, we select one by one the three digited numbers, discarding the numbers over 500, until 15 numbers below 500 are obtained. Since hare we have selected the random sample without replacement the numbers obtained previously (*i.e.*, in earlier selection) will also be discarded. Finally, the units in the population, corresponding to these 15 numbers will constitute our random sample without replacement.

The following is an extract from the first set of 40 four-digited numbers in Tippet's random number tables :

TABLE 7·2 : TABLE OF RANDOM NUMBERS

2952	6641	3992	9792	7969	5911	3170	5624
4167	9524	1545	1396	7203	5356	1300	2693
2370	7483	3408	2762	3563	1089	6913	7691
0560	5246	0112	6107	6008	8126	4233	8776
2754	9143	1405	9025	7002	6111	8816	6446

Starting with first number and moving column-wise, the units in the population with the numbers : 295, 416, 237, 056, 275, 266, 074, 052, 491, 413, 241, 460, 431, 408, 112, will be desired sample of size 15 without replacement.

Remark. If the extract given from the random number tables is small such that the sample of a given size cannot be drawn, then we assign more than one number to each sampling unit in the population as illustrated in the following example :

***Example 7·2.** The following Table 7·3 of ten random numbers of two digits each is provided to the field investigator :*

TABLE 7·3

34	*96*	*61*	*85*	*49*
78	*50*	*02*	*27*	*13*

Scanned with CamScanner

2. The standard error (S.E.) of the sampling distribution of $\bar{y}_n$ is given by :

$$\text{S.E.}\ (\bar{y}_n) = \sqrt{\frac{N-n}{N}} \cdot \frac{S}{\sqrt{n}} \qquad \ldots (7\cdot21)$$

Usually S is not known and in that case we replace S^2 by its unbiased estimate s^2 and get

$$\text{Est. (S.E. } \bar{y}_n) = \sqrt{\frac{N-n}{N}} \cdot \frac{s}{\sqrt{n}} = \sqrt{(1-f)}\ \frac{s}{\sqrt{n}} \qquad \ldots (7\cdot21a)$$

3. ***Srswor vs. Srswr.*** If we consider simple random sampling with replacement (*srswr*) from a population with variance σ^2, then

$$\text{Var}\ (\bar{y}_n) = \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} \text{Var}\ (y_i),$$

the covariance terms vanish since in srswr all the draws are independent and consequently y_i, 's; $(i = 1, 2, \ldots, n)$ are independently and identically distributed ($i.\ i.\ d.$) with the same variance σ^2.

$$\therefore \quad \text{Var}\ (\bar{y}_n) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n} \qquad \left[\because\ \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y}_N)^2 \ \Rightarrow\ N\sigma^2 = (N-1)S^2\right]$$

$$\therefore \quad \text{Var}\ (\bar{y}_n) = \frac{N-1}{nN}\, S^2 \qquad \ldots (7\cdot22)$$

Comparing this with the expression in (7·16), we see that the variance of the sample mean is more in srswr as compared with its variance in the case of srswor. In other words, *srswor provides a more efficient estimator of* $\bar{Y}_N$ *relative to srswr.*

7·9·5. Merits and Drawbacks of Simple Random Sampling.

Merits **1.** Since the sample units are selected at random giving each unit an equal chance of being selected, the element of subjectivity or personal bias is completely eliminated. As such a simple random sample is more representative of the population as compared to the judgement or purposive sampling.

2. The statistician can ascertain the efficiency of the estimates of the parameters by considering the sampling distribution of the statistics (estimates), *e.g.*, $\bar{y}_n$ as an estimate of $\bar{Y}_N$ becomes more efficient as sample size n increases.

Drawbacks **1.** The selection of a simple random sample requires an up-to-date frame, *i.e.*, a completely catalogued population from which samples are to be drawn. Frequently, it is virtually impossible to identify the units in the population before the sample is drawn and this restricts the use of simple random sampling technique.

2. *Administrative Inconvenience.* A simple random sample may result in the selection of the sampling units which are widely spread geographically and in such a case the cost of collecting the data may be much in terms of time and money.

3. At times, a simple random sample might give most *non-random* looking results. For example, if we draw a random sample of size 13 from a pack of cards, we may get all the cards of the same suit. However, the probability of such an outcome is extremely small.

4. For a given precision, simple random sampling usually requires larger sample size as compared to stratified random sampling discussed in § 7·10.

Example 7·7. *Consider a population of 6 units with values 1, 2, 3, 4, 5, 6. Write down all possible samples of 2 (without replacement) from this population and verify that sample mean is an unbiased estimate of the population mean.*

Scanned with CamScanner

Substituting these values in (3), we get

$$\text{Var}(\bar{y}) = \frac{S^2}{N(n+n_1)^2}\left[4n_1(N-n_1)+(n-n_1)(N-n+n_1)-4n_1(n-n_1)\right]$$

$$= \frac{S^2}{N(n+n_1)^2}\left[Nn\left(1+\frac{3n_1}{n}\right)-(n+n_1)^2\right] \quad \text{(on simplification)}$$

For original sample of size n, we have

$$\text{Var}(\bar{y}_n) = \frac{N-n}{Nn}\cdot S^2$$

$$\therefore \quad \frac{\text{Var}(\bar{y})}{\text{Var}(\bar{y}_n)} = \frac{n}{(N-n)(n+n_1)^2}\left[Nn\left(1+\frac{3n_1}{n}\right)-(n+n_1)^2\right]$$

$$= \left[\frac{\{n/(n+n_1)\}^2\{1+(3n_1/n)\}}{\{1-(n/N)\}} - \frac{(n/N)}{\{1-(n/N)\}}\right]$$

If N is large as compared to n such that *f. p.c.* n/N is ignored, then

$$\frac{\text{Var}(\bar{y})}{\text{Var}(\bar{y}_n)} = \frac{\{1+\{3n_1/n\}}{[1+(n_1/n)]^2}$$

Thus, the relative loss in efficiency resulting from duplicating of a sub-set of elements is given by [f.p.c. ignored).

$$\frac{\text{Var}(\bar{y})-\text{Var}(\bar{y}_n)}{\text{Var}(\bar{y}_n)} = \frac{\text{Var}(\bar{y})}{\text{Var}(\bar{y}_n)}-1 = \frac{(n_1/n)[1-(n_1/n)]}{[1+(n_1/n)]^2} \quad \text{(On simplification)}$$

In particular, if we take $\frac{n_1}{n}=\frac{1}{3}$, then: Loss in efficiency $= \frac{(1/3)(2/3)}{(16/9)} = 0{\cdot}125.$

7·9·6. Simple Random Sampling of Attributes. An attribute is a qualitative characteristic which cannot be measured quantitatively, *e.g.*, honesty, beauty, intellegence, etc. Quite often we come across situations where it may not be possible to measure the characteristic under study but it may be possible to classify the whole population into various classes *w.r.t.* the attribute under study. We consider the simplest of the cases where the population is divided into two classes only, say, C and C' with respect to an attribute. Such a classification is termed as dichotomous classification. Hence any sampling unit in the population may be placed in class C or C' respectively according as it possesses or does not possess the given attribute. In the study of attributes we may be interested in estimating the *total number of proportion* of

(*i*) defective items in a large consignment of such items,

(*ii*) the literates or the bread winners in a town,

(*iii*) the educated (university graduates) unemployed persons in a city, and so on.

Notations and Terminology. Let us suppose that a population with N units $U_1, U_2, \ldots U_N$ is classified into two mutually disjoint and exhaustive classes C and C' respectively *w.r.t.* a given attribute. Let the number of individuals in classes C and C' be A and A' respectively such that $A + A' = N$.

Then

P = The proportion of units processing the given attribute = A/N.

Q = The proportion of units which do not possess the given attribute = $\frac{A'}{N} = 1 - P$.

Scanned with CamScanner

In statistical language we call it a success if ith sampling unit possesses the given attribute otherwise a failure. Accordingly P and Q represent the proportion of successes and failures respectively in the population.

Let us consider a *srswor* of size n from this population. If 'a' is the number of units in a sample possesing the given attribute then

$$\left.\begin{aligned} p &= \text{Proportion of sampled units possessing the given attribute} = a/n \\ \text{and} \quad q &= \text{Proportion of sampled units which do not possess the given attribute} \\ &= 1 - p \end{aligned}\right\} \quad \ldots(7\cdot23)$$

With the ith sampling unit, let us associate a variate Y_i ($i = 1, 2, \ldots, N$) defined as follows :

$Y_i = 1$, if it belongs to class C, *i.e.*, if it possesses the given attribute, and $Y_i = 0$, if it belongs to class C', *i.e.*, if it does not possess the given attribute.

Similarly, let us associate a variable y_i, ($i = 1, 2, \ldots, n$) with the ith sampled unit defined as follows :

$$\begin{aligned} y_i &= 1, \text{ if } i\text{th sampled unit possesses the given attribute,} \\ &= 0, \text{ if } i\text{th sampled unit does not possess the given attribute.} \end{aligned}$$

Then $\sum_{i=1}^{N} Y_i = A$, the number of units in the population possessing the given attribute.

and $\sum_{i=1}^{n} y_i = a$, the number of sample units possessing the given attribute.

Thus

$$\overline{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i = \frac{A}{N} = P \quad \text{and} \quad \overline{y} = \frac{1}{n} \sum_{i=1}^{N} y_i = \frac{a}{n} = p \qquad \ldots(7\cdot24)$$

Similarly, we have

$$\sum_{i=1}^{N} Y_i^2 = A = NP \quad \text{and} \quad \sum_{i=1}^{n} y_i^2 = a = np \qquad \ldots(7\cdot25)$$

$$\therefore \quad S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (Y_i - \overline{Y})^2 = \frac{1}{N-1} \left[\sum_{i=1}^{N} Y_i^2 - N\overline{Y}^2 \right]$$

$$= \frac{1}{N-1} [NP - NP^2] = \frac{NPQ}{N-1} \qquad \ldots(7\cdot26)$$

Similarly, we get

$$s^2 = \frac{1}{n-1} \left[\sum_{i=1}^{N} y_i^2 - n\overline{y}^2 \right] = \frac{1}{n-1} [np - np^2] = \frac{npq}{n-1}. \qquad \ldots(7\cdot26a)$$

Theorem 7·4. *Sample proportion 'p' is an unbiased estimate of population proportion P, i.e.,*

$$E(p) = P \qquad \ldots(7\cdot27)$$

Proof. We know that in simple random sampling the sample mean provides an unbiased estimate of the population mean.

$$\therefore \quad E(\overline{y}) = \overline{Y} \quad \Rightarrow \quad E(p) = P \qquad \text{[On using (7·24)]}$$

Cor. We have $\quad E(p) = P \quad \Rightarrow \quad E(Np) = NP = A$

Scanned with CamScanner

SIMPLE RANDOM SAMPLING (SRS)

It is the technique of drawing a sample in such a way that each unit of population has an equal and independent chance of being included in the sample.

In SRS, from a population of N units, the probability of drawing any unit at the first draw is $\frac{1}{N}$, the probability of drawing any unit in the second draw from among the available (N-1) units, is $\frac{1}{N-1}$ and so on.

Theorem: In SRSWOR,
Show that the probability of selecting a specified unit of the population at any given draw is equal to the probability of its being selected at the first draw.

Proof:

Let 'N' be the population size and let E_r be the event that any specified unit is selected at 'r'th draw. Then

$P(E_r)$ = Prob.{ that the specified unit is not selected in anyone of the previous $(r-1)$ draws and then selected at the r^{th} draw }

$= \prod_{i=1}^{r-1}$ P{ It is not selected at i^{th} draw }

$\times$ P{ It is selected at r^{th} draw given that it is not selected at the previous $(r-1)$ draws }

(By compound prob. theorem, since draws are independent)

Scanned with CamScanner

Prob. of selecting at the 1^{st} draw unit, is $\frac{1}{N}$ and prob. of selecting at the 2^{nd} draw from among remaining $(N-1)$ units is $\frac{1}{N-1}$ ... (ie. $\frac{1}{N}, \frac{1}{N-1}, \frac{1}{N-2}, \dots$)

$$\therefore P(E_r) = \prod_{i=1}^{r-1} \left[1 - \frac{1}{N-(i-1)}\right] \times \frac{1}{N-(r-1)}$$

$$= \prod_{i=1}^{r-1} \left[\frac{N-i+\cancel{1}-\cancel{1}}{N-i+1}\right] \times \frac{1}{N-r+1}$$

$$= \frac{\cancel{N-1}}{N} \times \frac{\cancel{N-2}}{\cancel{N-1}} \times \dots \times \frac{\cancel{N-r+1}}{\cancel{N-r+2}} \times \frac{1}{\cancel{N-r+1}}$$

$$= \frac{1}{N}$$

$$\therefore P(E_r) = \frac{1}{N} = P(E_1)$$

Remark:

The prob. that the unit is included in the sample $= \sum_{i=1}^{n} \frac{1}{N} = \frac{n}{N}$

Selection of a SRS

Random sample can be obtained by any of the following methods

a) By Lottery system

b) 'Mechanical Randomization' or 'Random Numbers' method

(Give the notes from page 7.14 to 7.15 in Fundamentals of Applied Statistics - S. C. Gupta and V.K. Kapoor)

Scanned with CamScanner

Notations and Terminology

Let us Consider a (finite) population of 'N' units and let 'y' be the character under consideration. Let Y_i, $i=1,2,\dots,N$ be the quality characteristic of the population and y_i, $i=1,2,\dots,n$ be the character which is selected in the sample units.

Then

i) Population mean $= \bar{Y}_N = \frac{1}{N}\sum_{i=1}^{N} Y_i$

ii) Sample mean $= \bar{y}_n = \frac{1}{n}\sum_{i=1}^{n} y_i$

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

where $a_i = \begin{cases} 1, & \text{if } i^{th} \text{ unit is included in the sample} \\ 0, & \text{if } i^{th} \text{ unit is not } \quad " \quad\quad " \end{cases}$

iii) Population mean square $= S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y}_N)^2$

$$= \frac{1}{N-1}\left[\sum_{i=1}^{N} Y_i^2 - N\bar{Y}_N^2\right]$$

iv) Sample mean square $= s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y}_n)^2$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\bar{y}_n^2\right]$$

Scanned with CamScanner

SIMPLE RANDOM SAMPLING WITHOUT REPLACEMENT

(SRSWOR)

Important results

1. $E(\bar{y}_n) = \bar{Y}_N$
2. $E(s^2) = S^2$
3. $V(\bar{y}_n) = \frac{N-n}{N} \frac{S^2}{n}$

Theorem 1 :

Prove that, in SRSWOR, the sample mean is an unbiased estimate of the population mean.

ie, $E(\bar{y}_n) = \bar{Y}_N$

Proof :

We Know that $\bar{y}_n = \frac{1}{n} \sum_{i=1}^{n} y_i$

$$\bar{y}_n = \frac{1}{n} \sum_{i=1}^{N} a_i y_i \quad \text{——— (1)}$$

where $a_i = \begin{cases} 1 & \text{if } i^{th} \text{ unit is included in the sample} \\ 0 & \text{if } i^{th} \text{ unit is not} \quad " \quad\quad " \quad\quad " \end{cases}$

Taking expectation on both sides in (1), we get

$$E(\bar{y}_n) = E\left[\frac{1}{n} \sum_{i=1}^{N} a_i y_i\right]$$

$$= \frac{1}{n} \sum_{i=1}^{N} E(a_i) y_i \quad \text{——— (2)}$$

Since a_i takes only two values 1 and 0

$$E(a_i) = 1 \cdot P(a_i = 1) + 0 \cdot P(a_i = 0)$$

Scanned with CamScanner

$E(a_i) = 1 . P$ (i^{th} unit is included in a sample of size n)

$+ 0 . P$ (i^{th} unit is not included in a sample of size n)

$$= 1 . \frac{n}{N} + 0 \left(1 - \frac{n}{N}\right)$$

$$= \frac{n}{N} \quad \text{——— (3)}$$

Put (3) in (2), we get

$$E(\bar{y}_n) = \frac{1}{\cancel{n}} \sum_{i=1}^{N} \frac{\cancel{n}}{N} y_i$$

$$= \frac{1}{N} \sum_{i=1}^{N} y_i$$

$$= \bar{Y}_N$$

$$\Rightarrow E(\bar{y}_n) = \bar{Y}_N$$

Theorem 2:

P.T. in SRSWOR, the sample mean square is an unbiased estimate of the population mean square

i.e., $E(s^2) = S^2$

Proof:

We know that $s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\bar{y}_n^2\right]$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - n\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \cancel{n} . \frac{1}{n^2}\left(\sum_{i=1}^{n} y_i\right)^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2\right]$$

Scanned with CamScanner

Scanned with CamScanner

$(y_1+y_2)^2 = y_1^2 + y_2^2 + 2y_1y_2$

$= \sum_{i=1}^{2} y_i^2 + \sum_{i \neq j=1}^{2} y_i y_j$

i.e., $= y_1^2+y_2^2+y_1y_2+y_2y_1$

$= y_1^2+y_2^2+2y_1y_2$

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i^2 + \sum_{i\neq j=1}^{n} y_i y_j\right)\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\sum_{i\neq j=1}^{n} y_i y_j\right]$$

$$= \frac{1}{n-1}\left[\left(1-\frac{1}{n}\right)\sum_{i=1}^{n} y_i^2 - \frac{1}{n}\sum_{i\neq j=1}^{n} y_i y_j\right]$$

$$= \frac{1}{\cancel{n-1}}\left(\frac{\cancel{n-1}}{n}\right)\sum_{i=1}^{n} y_i^2 - \frac{1}{n(n-1)}\sum_{i\neq j=1}^{n} y_i y_j$$

$$= \frac{1}{n}\sum_{i=1}^{n} y_i^2 - \frac{1}{n(n-1)}\sum_{i\neq j=1}^{n} y_i y_j$$

Taking expectation on both sides, we get

$$E(s^2) = E\left[\frac{1}{n}\sum_{i=1}^{n} y_i^2\right] - E\left[\frac{1}{n(n-1)}\sum_{i\neq j=1}^{n} y_i y_j\right] \quad \text{(1)}$$

Consider $E\left[\frac{1}{n}\sum_{i=1}^{n} y_i^2\right] = \frac{1}{n} E\left[\sum_{i=1}^{n} y_i^2\right]$

$$= \frac{1}{n} E\left[\sum_{i=1}^{N} a_i Y_i^2\right]$$

$$= \frac{1}{n}\sum_{i=1}^{N} E(a_i)\, Y_i^2$$

where $E(a_i) = 1 \cdot P(a_i=1) + 0 \cdot P(a_i=0)$

$$= 1 \cdot \frac{n}{N} + 0\left(1-\frac{n}{N}\right)$$

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

$$\therefore E\left[\frac{1}{n}\sum_{i=1}^{n} y_i^2\right] = \frac{1}{\cancel{n}}\sum_{i=1}^{N} \frac{\cancel{n}}{N} Y_i^2$$

$$= \frac{1}{N}\sum_{i=1}^{N} Y_i^2 \quad \text{(2)}$$

Scanned with CamScanner

Consider $E\left[\frac{1}{n(n-1)}\sum_{i\neq j=1}^{n} y_i y_j\right] = E\left[\frac{1}{n(n-1)}\sum_{i\neq j=1}^{N} a_i a_j Y_i Y_j\right]$

$$= \frac{1}{n(n-1)}\sum_{i\neq j=1}^{N} E(a_i a_j) Y_i Y_j$$

Where $E(a_i a_j) = 1.P(a_i a_j = 1) + 0.P(a_i a_j = 0)$

$$= 1.P[(a_i = 1)\cap(a_j = 1)]$$

$$= P(a_i = 1).P(a_j = 1/a_i = 1)$$ $\because$ the events are independent

$$= \frac{n}{N}\cdot\frac{n-1}{N-1}$$ ie. $P(A\cap B) = P(A).P(B/A)$

$$= \frac{1}{\cancel{n}(\cancel{n-1})}\sum_{i\neq j=1}^{N}\frac{\cancel{n}}{N}\frac{\cancel{n-1}}{N-1} Y_i Y_j$$

$$= \frac{1}{N(N-1)}\sum_{i\neq j=1}^{N} Y_i Y_j \quad \text{——— (3)}$$

Substitute (2) & (3) in (1), we get

$$E(s^2) = \frac{1}{N}\sum_{i=1}^{N} Y_i^2 - \frac{1}{N(N-1)}\sum_{i\neq j=1}^{N} Y_i Y_j \quad \text{——— (4)}$$

We know that

$$S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N} Y_i^2 - N\overline{Y_N}^2\right]$$

$$(N-1)S^2 = \sum_{i=1}^{N} Y_i^2 - N\overline{Y_N}^2$$

$$(N-1)S^2 - \sum_{i=1}^{N} Y_i^2 = -N\overline{Y_N}^2$$

$$\sum_{i=1}^{N} Y_i^2 - (N-1)S^2 = N\overline{Y_N}^2$$

Scanned with CamScanner

$$\frac{\sum_{i=1}^{N} y_i^2}{N} - \frac{N-1}{N} S^2 = \bar{Y}_N^2$$

$$\frac{1}{N}\sum_{i=1}^{N} y_i^2 - \frac{N-1}{N} S^2 = \left[\frac{1}{N}\sum_{i=1}^{N} y_i\right]^2$$

$$= \frac{1}{N^2}\left[\sum_{i=1}^{N} y_i^2 + \sum_{i\neq j=1} y_i y_j\right]$$

$$N^2\left[\frac{1}{N}\sum_{i=1}^{N} y_i^2 - \frac{N-1}{N} S^2\right] = \sum_{i=1}^{N} y_i^2 + \sum_{i\neq j=1} y_i y_j$$

$$\Rightarrow \sum_{i\neq j=1} y_i y_j = N\sum_{i=1}^{N} y_i^2 - N(N-1)S^2 - \sum_{i=1}^{N} y_i^2 \quad \text{— (5)}$$

Put (5) in (4), we get

$$E(s^2) = \frac{1}{N}\sum_{i=1}^{N} y_i^2 - \frac{1}{N(N-1)}\left[N\sum_{i=1}^{N} y_i^2 - N(N-1)S^2 - \sum_{i=1}^{N} y_i^2\right]$$

$$= \frac{1}{N}\sum_{i=1}^{N} y_i^2 - \frac{1}{N-1}\sum_{i=1}^{N} y_i^2 + S^2 + \frac{1}{N(N-1)}\sum_{i=1}^{N} y_i^2$$

$$= \sum_{i=1}^{N} y_i^2\left(\frac{1}{N} - \frac{1}{N-1}\right) + S^2 + \frac{1}{N(N-1)}\sum_{i=1}^{N} y_i^2$$

$$= \sum_{i=1}^{N} y_i^2\left[\frac{\cancel{N}-1-\cancel{N}}{N(N-1)}\right] + S^2 + \frac{1}{N(N-1)}\sum_{i=1}^{N} y_i^2$$

$$= \cancel{\frac{-1}{N(N-1)}\sum_{i=1}^{N} y_i^2} + S^2 + \cancel{\frac{1}{N(N-1)}\sum_{i=1}^{N} y_i^2}$$

$$= S^2$$

$$\therefore E(s^2) = S^2$$

Scanned with CamScanner

Theorem 3 :

P.T. in SRSWOR $V(\bar{y}_n) = \frac{N-n}{N} \frac{S^2}{n}$

Proof

We know that $V(\bar{y}_n) = E(\bar{y}_n^2) - [E(\bar{y}_n)]^2$

$= E(\bar{y}_n^2) - \bar{Y}_N^2$

$\because$ Sample mean is an unbiased estimate of the population mean

$E(\bar{y}_n) = \bar{Y}_N$

$$\therefore V(\bar{y}_n) = E\left[\frac{1}{n}\sum_{i=1}^{n} y_i\right]^2 - \bar{Y}_N^2$$

$$= \frac{1}{n^2} E\left(\sum_{i=1}^{n} y_i\right)^2 - \bar{Y}_N^2$$

$$= \frac{1}{n^2} E\left[\sum_{i=1}^{n} y_i^2 + \sum_{i \neq j=1}^{n} y_i y_j\right] - \bar{Y}_N^2$$

$$= \frac{1}{n^2}\left[E\left(\sum_{i=1}^{n} y_i^2\right) + E\left(\sum_{i \neq j=1}^{n} y_i y_j\right)\right] - \bar{Y}_N^2 \quad \text{—— (1)}$$

Consider $E\left(\sum_{i=1}^{n} y_i^2\right) = E\left(\sum_{i=1}^{N} a_i y_i^2\right)$

$$= \sum_{i=1}^{N} E(a_i) y_i^2$$

$\because E(a_i) = 1.P(a_i=1) + 0.P(a_i=0)$

$= 1.\frac{n}{N} - 0.\left(1-\frac{n}{N}\right)$

$= \frac{n}{N}$

$$= \sum_{i=1}^{N} \frac{n}{N} y_i^2$$

$$= \frac{n}{N}\sum_{i=1}^{N} y_i^2 \quad \text{—— (2)}$$

Scanned with CamScanner

But we know that

$$S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N} y_i^2 - N\overline{Y}_N^2\right]$$

$$(N-1)S^2 = \sum_{i=1}^{N} y_i^2 - N\overline{Y}_N^2$$

$$(N-1)S^2 + N\overline{Y}_N^2 = \sum_{i=1}^{N} y_i^2 \quad \text{——— (3)}$$

Substitute (3) in (2), we get

$$E\left[\sum_{i=1}^{n} y_i^2\right] = \frac{n}{N}\left[(N-1)s^2 + N\overline{Y}_N^2\right]$$

$$= \frac{n(N-1)}{N} s^2 + n\overline{Y}_N^2$$

$$= n\left(1-\frac{1}{N}\right)s^2 + n\overline{Y}_N^2 \quad \text{——— (4)}$$

Consider $E\left[\sum_{i\neq j=1}^{n} y_i y_j\right] = E\left[\sum_{i\neq j=1}^{N} a_i a_j y_i y_j\right]$

$$= \sum_{i\neq j=1}^{N} E(a_i a_j)\, y_i y_j$$

where $E(a_i a_j) = 1 \cdot P(a_i a_j = 1) + 0 \cdot P(a_i a_j = 0)$

$$= 1\, P\left[(a_i=1) \cap (a_j=1)\right]$$

$$= P(a_i=1)\, P(a_j=1/a_i=1)$$

$$= \frac{n}{N}\, \frac{n-1}{N-1}$$

$$= \sum_{i\neq j=1}^{N} \frac{n}{N}\, \frac{n-1}{N-1}\, y_i y_j$$

$$= \frac{n(n-1)}{N(N-1)} \sum_{i\neq j=1}^{N} y_i y_j \quad \text{——— (5)}$$

Scanned with CamScanner

But $\left(\sum_{i=1}^{N} y_i\right)^2 = \sum_{i=1}^{N} y_i^2 + \sum_{i\neq j=1}^{N} y_i y_j$

$$\left(\sum_{i=1}^{N} y_i\right)^2 - \sum_{i=1}^{N} y_i^2 = \sum_{i\neq j=1}^{N} y_i y_j$$

$\because \bar{Y}_N = \frac{1}{N}\sum_{i=1}^{N} y_i$

$N\bar{Y}_N = \sum_{i=1}^{N} y_i$

$(N\bar{Y}_N)^2 = \left(\sum_{i=1}^{N} y_i\right)^2$

$$(N\bar{Y}_N)^2 - (N-1)S^2 - N\bar{Y}_N^2 = \sum_{i\neq j=1}^{N} y_i y_j$$

$$\Rightarrow \sum_{i\neq j=1}^{N} y_i y_j = N^2\bar{Y}_N^2 - N\bar{Y}_N^2 - (N-1)S^2$$

$$= \bar{Y}_N^2 N(N-1) - (N-1)S^2 \quad \text{—— (6)}$$

& $\because S^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N} y_i^2 - N\bar{Y}_N^2\right]$

$(N-1)S^2 = \sum_{i=1}^{N} y_i^2 - N\bar{Y}_N^2$

$\sum_{i=1}^{N} y_i^2 = (N-1)S^2 + N\bar{Y}_N^2$

Substitute (6) in (5) we get

$$E\left[\sum_{i\neq j=1}^{n} y_i y_j\right] = \frac{n(n-1)}{N(N-1)}\left[N(N-1)\bar{Y}_N^2 - (N-1)S^2\right]$$

$$= n(n-1)\bar{Y}_N^2 - \frac{n(n-1)}{N}S^2 \quad \text{—— (7)}$$

Substitute (4) & (7) in (1), we get

$$V(\bar{y}_n) = \frac{1}{n^2}\left[\left\{n\left(1-\frac{1}{N}\right)S^2 + n\bar{Y}_N^2\right\} + \left\{n(n-1)\bar{Y}_N^2 - \frac{n(n-1)}{N}S^2\right\}\right] - \bar{Y}_N^2$$

$$= \frac{1}{n^2}\left[nS^2 - \frac{n}{N}S^2 + n\bar{Y}_N^2 + n^2\left(1-\frac{1}{n}\right)\bar{Y}_N^2 - n^2\left(1-\frac{1}{n}\right)\frac{S^2}{N}\right] - \bar{Y}_N^2$$

$$= \frac{S^2}{n} - \frac{S^2}{nN} + \frac{\bar{Y}_N^2}{n} + \left(1-\frac{1}{n}\right)\bar{Y}_N^2 - \left(1-\frac{1}{n}\right)\frac{S^2}{N} - \bar{Y}_N^2$$

$$= \frac{S^2}{n} - \cancel{\frac{S^2}{nN}} + \cancel{\frac{\bar{Y}_N^2}{n}} + \cancel{\bar{Y}_N^2} - \cancel{\frac{\bar{Y}_N^2}{n}} - \frac{S^2}{N} + \cancel{\frac{S^2}{nN}} - \cancel{\bar{Y}_N^2}$$

$$= \frac{S^2}{n} - \frac{S^2}{N} = S^2\left(\frac{1}{n} - \frac{1}{N}\right) = S^2\left(\frac{N-n}{Nn}\right)$$

$$\therefore V(\bar{y}_n) = \frac{N-n}{N}\cdot\frac{S^2}{n}$$

Scanned with CamScanner

Remark

1. We Know that $V(\bar{y}_n) = \frac{N-n}{N} \frac{S^2}{n}$

$$= \left(1-\frac{n}{N}\right)\frac{S^2}{n}$$

$$= (1-f)\frac{S^2}{n}$$

where $f = \frac{n}{N}$ is called sampling fraction and the factor $(1-f)$ is called the finite population Correction (f.P.C).

If the population size 'N' is very large or if 'n' is small compared with 'N' then $f = \frac{n}{N} \to 0$ and $f.P.C \to 1$.

If the $f.P.C \to 1$ when 'N' is large and 'n' is small then $V(\bar{y}_n) = \frac{S^2}{n}$.

2. Standard error of the sampling distrn. of $\bar{y}_n$ is given by

$$S.E(\bar{y}_n) = \sqrt{\text{Variance}} = \sqrt{\frac{N-n}{N}\frac{S^2}{n}}$$

$$= \sqrt{\left(1-\frac{n}{N}\right)}\frac{S}{\sqrt{n}}$$

$$= \sqrt{1-f}\,\frac{S}{\sqrt{n}}$$

Usually 'S' is not known and in that case we replace 'S' by its unbiased estimate 's' then

$$S.E(\bar{y}_n) = \sqrt{1-f}\,\frac{s}{\sqrt{n}}.$$

Scanned with CamScanner

SIMPLE RANDOM SAMPLING WITH REPLACEMENT

(SRSWR)

Important results

1. $E(\bar{y}_n) = \bar{Y}_N$

3. $E(s^2) = \sigma^2$

2. $V(\bar{y}_n) = \frac{\sigma^2}{n}$ (or) $\frac{N-1}{N} \frac{S^2}{n}$

4. Relationship between σ^2 and S^2 is $\sigma^2 = \frac{N-1}{N} S^2$

Notations and Terminology

i) Population mean $= \bar{Y}_N = \frac{1}{N} \sum_{i=1}^{N} y_i$

ii) Sample mean $= \bar{y}_n = \frac{1}{n} \sum_{i=1}^{n} y_i$

iii) Population variance ~~mean Square~~ $= \sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (y_i - \bar{Y}_N)^2$

iv) Sample mean Square $= s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y}_n)^2$

Theorem 1:

In SRSWR, P.T. $E(\bar{y}_n) = \bar{Y}_N$.

Proof:

The proof in SRSWOR can be reproduced here

Scanned with CamScanner

Theorem : 2

In SRSWR, P.T. $V(\bar{y}_n) = \frac{\sigma^2}{n}$ (or) $\frac{N-1}{N}\frac{S^2}{n}$.

Proof :

We know that $V(\bar{y}_n) = V\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)$

$$= \frac{1}{n^2} V\left(\sum_{i=1}^{n} y_i\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} V(y_i)$$

Here each draws are identically independently distributed, we get the Common ~~values~~ Variance

ie., $V(y_i) = \sigma^2$

$$\therefore V(\bar{y}_n) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma_i^2$$

$$= \frac{1}{n^2}\, n\sigma^2$$

$$= \frac{\sigma^2}{n} \quad \text{——} \quad (1)$$

But $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y}_N)^2 \Rightarrow N\sigma^2 = \sum_{i=1}^{N}(Y_i - \bar{Y}_N)^2$ —— (2)

and $S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(Y_i - \bar{Y}_N)^2 \Rightarrow (N-1)S^2 = \sum_{i=1}^{N}(Y_i - \bar{Y}_N)^2$ —— (3)

Comparing (2) & (3), we get

$$N\sigma^2 = (N-1)S^2$$

$$\Rightarrow \sigma^2 = \frac{N-1}{N} S^2 \quad \text{——} \quad (4)$$

Put (4) in (1), we get

$$V(\bar{y}_n) = \frac{N-1}{N}\cdot\frac{S^2}{n}$$

Scanned with CamScanner

Theorem 3 :

In SRSWR, P.T. $E(s^2) = \sigma^2$

Proof :

We know that

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2$$

Add & Subtract $\bar{Y}$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}\left[(y_i-\bar{Y})-(\bar{y}-\bar{Y})\right]^2\right]$$

$$= \frac{1}{n-1}\left[\Sigma(y_i-\bar{Y})^2 - 2\sum_{i=1}^{n}(y_i-\bar{Y})(\bar{y}-\bar{Y}) + \sum_{i=1}^{n}(\bar{y}-\bar{Y})\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(y_i-\bar{Y})^2 - 2n(\bar{y}-\bar{Y})(\bar{y}-\bar{Y}) + n(\bar{y}-\bar{Y})^2\right]$$

$$\because \Sigma(y_i-\bar{Y}) = \Sigma y_i - n\bar{Y}$$
$$= n\bar{y} - n\bar{Y}$$
$$= n(\bar{y}-\bar{Y})$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(y_i-\bar{Y})^2 - 2n(\bar{y}-\bar{Y})^2 + n(\bar{y}-\bar{Y})^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}(y_i-\bar{Y})^2 - n(\bar{y}-\bar{Y})^2\right]$$

Taking Expectations on both sides,

$$E(s^2) = \frac{1}{n-1}\left[\sum_{i=1}^{n}E(y_i-\bar{Y})^2 - nE(\bar{y}-\bar{Y})^2\right]$$

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}\sigma^2 - \cancel{n}\cdot\frac{\sigma^2}{\cancel{n}}\right] \qquad \because E(\bar{y}-\bar{Y})^2 = V(\bar{y}) = \sigma^2/n$$

$$= \frac{1}{n-1}\left[n\sigma^2 - \sigma^2\right]$$

$$= \frac{1}{\cancel{n-1}}\sigma^2\cancel{(n-1)}$$

$$= \sigma^2$$

$$\therefore E(s^2) = \sigma^2$$

Scanned with CamScanner

SRSWOR Vs SRSWR

In SRS WOR, we know that

$$V(\bar{y}_n) = \frac{N-n}{N}\frac{S^2}{n} \quad \text{——— (1)}$$

If we consider SRSWR from a population with variance σ^2, then

$$V(\bar{y}_n) = V\left(\frac{1}{n}\sum_{i=1}^{n} y_i\right)$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} V(y_i),$$

the co-variance terms vanish since in SRSWR all the draws are independent and consequently y_i's $(i=1,2,\dots,n)$ are i.i.d with same variance σ^2.

$$\therefore V(\bar{y}_n) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2$$

$$= \frac{\cancel{n}\sigma^2}{n^{\cancel{2}}} = \frac{\sigma^2}{n}$$

But $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y}_N)^2$

$$\Rightarrow N\sigma^2 = (N-1)S^2$$

$$\therefore V(\bar{y}_n) = \frac{N-1}{N}\frac{S^2}{n} \quad \text{——— (2)}$$

Comparing (1) & (2), we see that the variance of the sample mean is more in SRSWR as compared with its variance in the case of SRSWOR.

$$\text{i.e., } V(\bar{y}_n)_{SRSWR} > V(\bar{y}_n)_{SRSWOR}$$

Hence SRSWOR provides a more efficient estimator of $\bar{Y}_N$ relative to SRSWR.

Scanned with CamScanner

(.) Consider a population of 4 units with values 2, 7, 9, 12. Write down all possible samples of 2 (without replacement) from this population and verify that sample mean is an unbiased estimate of the population mean.

Also calculate its sampling variance and verify that

(i) it agrees with the formula for the variance of the sample mean, and

(ii) this variance is less than the variance obtained from sampling with replacement

Soln: Here $N = 4$ and $n = 2$. There are $^{N}C_{n} = {}^{4}C_{2} = 6$ samples

Sample No.	Sample Values y_i	Sample mean $\bar{y}$	$\bar{y} - \bar{Y}$	$(\bar{y} - \bar{Y})^2$	$s^2 = \frac{1}{n-1}\Sigma(y_i - \bar{y})^2$
1	(2,7)	4.5	−3	9	$\frac{1}{2-1}[(2-4.5)^2+(7-4.5)^2]$ = 12.5
2	(2,9)	5.5	−2	4	24.5
3	(2,12)	7.0	−0.5	0.25	50.0
4	(7,9)	8.0	0.5	0.25	2.0
5	(7,12)	9.5	2.0	4	12.5
6	(9,12)	10.5	3.0	9	4.5
		45.0		26.50	106.0

With usual notations

$$\text{Population mean } = \bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i = \frac{2+7+9+12}{4} = \frac{30}{4} = 7.5$$

$$\left.\begin{array}{l}\text{Population}\\\text{Variance}\end{array}\right\} = \sigma^2 = \frac{\sum_{i=1}^{N} Y_i^2}{N} - \bar{Y}_N^2$$

$$= \frac{2^2 + 7^2 + 9^2 + 12^2}{4} - (7.5)^2$$

$$= \frac{278}{4} - 56.25 = 69.5 - 56.25 = 13.25$$

Scanned with CamScanner

18

$$S^2 = \frac{N}{N-1}\sigma^2 \quad (or) \quad S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y}_N)^2$$

$$= \frac{4}{4-1}(13.25) \qquad = \frac{1}{4-1}\left[(2-7.5)^2+(7-7.5)^2+(9-7.5)^2+(12-7.5)^2\right]$$

$$= 17.66 \qquad = \frac{53}{3} = 17.66$$

Under SRSWOR

i) $E(\bar{y}) = \frac{1}{6}(4.5+5.5+7.0+8.0+9.5+10.5)$

$$= \frac{45}{6} = 7.5 = \bar{Y} \qquad \therefore E(\bar{y}) = \bar{Y}$$

ii) $E(s^2) = \frac{106}{6} = 17.66 = S^2 \qquad \therefore E(s^2) = S^2$

iii) $V(\bar{y}_n) = \frac{N-n}{N}\frac{S^2}{n} \quad (or) \quad V(\bar{y}_n) = \frac{1}{6}\sum(\bar{y}_i - \bar{Y})^2$

$$= \frac{4-2}{4} \times \frac{17.66}{2} \qquad = \frac{26.5}{6}$$

$$= 4.415 \quad \text{—— (1)} \qquad = 4.415$$

Under SRSWR

$$V(\bar{y}_n) = \frac{\sigma^2}{n} = \frac{13.25}{2} = 6.625 \quad \text{—— (2)}$$

Hence $V(\bar{y})_{SRSWR} = 6.625 > V(\bar{y})_{SRSWOR} = 4.415$

$$\therefore V(\bar{y})_{SRSWR} > V(\bar{y})_{SRSWOR}$$

Scanned with CamScanner

SIMPLE RANDOM SAMPLING OF ATTRIBUTES

An attribute is a qualitative characteristic which cannot be measured quantitatively, eg., honesty, beauty, intelligence etc. Quite often we come across situations where it may not be possible to measure the characteristic under study but it may be possible to classify the whole population into various classes w.r.t. the attribute under study. We consider the simplest of the cases where the population is divided into two classes only say c and c' w.r.t. an attribute.

Notations and Terminology

Let us suppose that a population with 'N' units is classified into two mutually disjoint and exhaustive classes 'c' and 'c'' respectively w.r.t. a given attribute. Let the number of individuals in classes 'c' and 'c'' be 'A' and 'A'' respectively such that $A + A' = N$.

Then

P = The proportion of units possessing the given attribute

$$= \frac{A}{N}$$

Q = The proportion of units which do not possess the given attribute

$$= \frac{A'}{N} = 1 - P$$

Scanned with CamScanner

Let us consider a SRSWOR of size 'n' from this population. If 'a' is the number of units in a sample possessing the given attribute then

p = Proportion of sampled units possessing the given attribute

$= \frac{a}{n}$

and q = Proportion of sampled units which do not possess the given attribute

$= 1-p$

Let us associate a variate Y_i, $i=1,2,\ldots,N$ with the i^{th} sampling unit defined as follows:

$Y_i = 1$, if it belongs to class C

$= 0$, if it belongs to class C'

Similarly let us associate a variable y_i, $i=1,2,\ldots,n$ with the i^{th} sampled unit defined as follows:

$y_i = 1$, if i^{th} sampled unit possesses the given attribute

$= 0$, if " " does not possess "

Then $\sum_{i=1}^{N} Y_i = A$

and $\sum_{i=1}^{n} y_i = a$

Thus
$$\left.\begin{aligned} \bar{Y} &= \frac{1}{N}\sum_{i=1}^{N} Y_i = \frac{A}{N} = P \\ \bar{y} &= \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{a}{n} = p \end{aligned}\right\} \quad \text{--- (1)}$$

(and)

IIIly we have

$$\left.\begin{aligned} \sum_{i=1}^{N} Y_i^2 &= A = NP \\ \sum_{i=1}^{n} y_i^2 &= a = np \end{aligned}\right\} \quad \text{--- (2)}$$

(and)

$$\therefore S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y})^2$$

$$= \frac{1}{N-1} \left[\sum_{i=1}^{N} y_i^2 - N\bar{Y}^2 \right]$$

$$= \frac{1}{N-1} \left[NP - NP^2 \right]$$

$$= \frac{NP(1-P)}{N-1}$$

$$= \frac{NPQ}{N-1} \quad \text{——— (3)}$$

III^{ly} we get

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$= \frac{1}{n-1} \left(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2 \right)$$

$$= \frac{1}{n-1} (np - np^2)$$

$$= \frac{np(1-p)}{n-1}$$

$$= \frac{npq}{n-1} \quad \text{——— (4)}$$

Theorem 1 :

Sample proportion 'p' is an unbiased estimate of population proportion 'P'

i.e., $E(p) = P$.

Proof :

We know that in SRS, the sample mean provides an unbiased estimate of the popln.. mean.

i.e.. $E(\bar{y}) = \bar{Y}$

$\Rightarrow \quad E(p) = P$ by using (1).

Scanned with CamScanner

Theorem 2 :

In SRSWOR $V(p) = \frac{N-n}{N-1} \frac{PQ}{n}$

Proof :

$$V(p) = Var(\bar{y})$$

But $V(\bar{y}) = \frac{N-n}{Nn} S^2$

$$= \frac{N-n}{\cancel{N}n} \frac{\cancel{N}PQ}{N-1} \quad \text{using } (3)$$

$$= \frac{N-n}{N-1} \cdot \frac{PQ}{n}$$

r

Scanned with CamScanner

ESTIMATION OF SAMPLE SIZE

Let $\bar{Y}_N$ be the mean of a population of size N.
We know that $\bar{y}_n$ is an unbiased estimate of $\bar{Y}_N$.
If the permissible error in estimating $\bar{Y}_N$ is 'd' and Confidence Co-eff t is $1-\alpha$, then the sample size 'n' is determined by the equation:

$$P\left[|\bar{y}_n - \bar{Y}_N| < d\right] = 1-\alpha \quad \text{——} (1)$$

or $$P\left[|\bar{y}_n - \bar{Y}_N| \geq d\right] = \alpha \quad \text{——} (2)$$

Where α is very small pre-assigned probability and is known as L.O.S.

and $d = t \times \sqrt{\text{variance}}$

$$\left.\begin{array}{l}\text{Precision of} \\ \text{the estimate}\end{array}\right\} = \begin{array}{c}\text{Reliability} \\ \text{Co-eff t}\end{array} \times \begin{array}{c}\text{S.E. of the} \\ \text{estimate}\end{array}$$

<u>If 'n' is sufficiently large</u>

$$t = \frac{\bar{y}_n - E(\bar{y}_n)}{S.E(\bar{y}_n)} = \frac{\bar{y}_n - \bar{Y}_N}{S\sqrt{\left(\frac{1}{n} - \frac{1}{N}\right)}}$$ is a Std. Normal Variate

if we take $\alpha = 0.05$, we have

$$P\left[|\bar{y}_n - \bar{Y}_N| \geq 1.96\ S.E(\bar{y}_n)\right] = 0.05$$

$$= P\left[\frac{|\bar{y}_n - \bar{Y}_N|}{S\sqrt{\left(\frac{1}{n} - \frac{1}{N}\right)}} \geq 1.96\right] = 0.05 \quad \text{——} (3)$$

Scanned with CamScanner

Comparing ② and ③, we get

$$d = 1.96\, s\sqrt{\left(\frac{1}{n} - \frac{1}{N}\right)}$$

$$\Rightarrow \frac{d^2}{s^2(1.96)^2} = \frac{1}{n} - \frac{1}{N}$$

$$\Rightarrow n = \frac{Ns^2(1.96)^2}{Nd^2 + s^2(1.96)^2}$$

$$= \frac{3.84\, Ns^2}{3.84\, s^2 + Nd^2}$$

If 'n' is small

$$P\left[|\bar{y}_n - \bar{Y}_N| \geq s\sqrt{\left(\frac{1}{n} - \frac{1}{N}\right)}\, t_\alpha\right] = \alpha$$

$$\Rightarrow d = t_\alpha\, s\sqrt{\left(\frac{1}{n} - \frac{1}{N}\right)}$$

$$\Rightarrow n = \frac{NS^2 t_\alpha^2}{Nd^2 + S^2 t_\alpha^2} = \frac{S^2 t_\alpha^2}{d^2 + (S^2 t_\alpha^2 / N)}$$

Where t_α is the significant value of 't' for $(n-1)$ d.f for the given α (L.O.S).

Scanned with CamScanner

Example :

From a population of 1000 units, the population mean square S^2 is 100. What should be the size of a sample taken from it, so that 95% of sample means may differ from the population mean by not more than 0.5?

Soln :

It is given that $d = 0.5$, $S^2 = 100$ & $N = 1000$

For $\alpha = 5\%$, $$n = \frac{NS^2 \times (1.96)^2}{(1.96)^2 S^2 + Nd^2}$$

$$= \frac{(1000 \times 100) \times 3.84}{(3.84 \times 100) + [1000 \times (0.5^2)]}$$

$$= \frac{3,84,000}{384 + 250} = \frac{3,84,000}{634} = 605.67 \simeq 606.$$

Theory Questions

1. Define Simple random Sampling. Describe any one method of selecting a simple random sample (NOV'05)
2. What is random sampling? State the merits and limitations of simple random sampling. (NOV'04)
3. Discuss any two methods of drawing SRSWOR (NOV'08)
4. Differentiate between SRSWR and SRSWOR (NOV'07)
5. What is finite population correction and Sampling fraction. (APT'08) & (NOV'07)

Scanned with CamScanner

Source :

1. S.C. Gupta and V.K. Kapoor: Fundamental of Applied Statistics –Sultan Chand & Sons, Fourth Edition, 2015.