Simple Random Sampling

(for III BA/BSc Statistics Students)

By

K. Ashok

Depart of Statistics

PR Govt. College (A)

Kakinada

Simple Random Sampling

Definition: If sample is drawn from a population such that each and every unit of the population has equal and independent chance of being included in the sample, then the sampling is called a **Simple Random Sampling** or simply **Random Sampling**. The sample so obtained is called a **Simple Random Sample**

It is TWO types

- Simple Random Sampling Without Replacement (SRSWOR)
- Simple Random Sampling With Replacement (SRSWR)

SRSWOR: In SRSWOR, units are drawn from the population one by one by without replacing the selected units in the previous draw in to the population before the next draw giving equal chance.

If Population size is 'N' and Sample size is 'n',

the number possible samples by SRSWOR is $\binom{N}{n}$

** The probability of each sample is $1/\binom{N}{n}$

1. In SRSWOR, the probability of selecting a specified unit in to the sample at any draw is equal to the probability of selecting it at first draw,
   i.e., 1/N.
   Proof:
   Let Ei = the event of selecting a specified unit in to the sample at ith draw
   => $P(E1) = 1/N$, $P(E2/\bar{E}1) = 1/(N-1)$,
   $P(E3/(\bar{E}1 \cap \bar{E}2) = 1/(N-2)$, ... $P(Ei/(\bar{E}1 \cap .. \bar{E}i\text{-}1)) = 1/(N-i)$,....
   Now. $P(Ei) = P(\bar{E}1 \cap \bar{E}2 \cap \bar{E}3 \cap ...... \bar{E}i-1 \cap Ei)$

By multiplication theorem of probability,

$$=P(\bar{E}1).P(\bar{E}2/\bar{E}1).P(\bar{E}3/(\bar{E}1\cap\bar{E}2))........P(Ei/(\bar{E}1\cap\bar{E}2\cap\bar{E}3\cap......\bar{E}i-1))$$

$$= (1-1/N).\ (1-1/(N-1)).\ (1-1/(N-2).........1/(N-i)$$

$$= (N-1)/N.\ (N-2)/(N-1).\ (N-3)/(N-2)......1/(N-i)$$

$$= 1/N = P(E1)$$

2. In SRSWOR, the probability of a specified unit being included in to the sample is n/N.

   Proof:

   P(a specified unit being included in the sample) = P(E1UE2UE3U.....UEn)

   = P(E1)+P(E2)+....+P(En)

   (Since E1,E2,...En are exclusive)

   = 1/N+1/N+.....+1/N (n times)

   = n/N

SRSWR: In SRSWR, units are drawn from the population one by one by replacing the selected units in the previous draw in to the population before the next draw giving equal chance.

If Population size is ‘N’ and Sample size is ‘n’,

the number possible samples by SRSWOR is N^n

** The probability of each sample is 1/ N^n

** The probability of a specified unit being included in the sample is n/N

Methods of selecting simple random sample:

- Lottery Method
- Random Numbers Method or Mechanical Randomization Method

Lottery Method: The simplest method of selecting a random sample is the Lottery method. The following are general steps in any lottery system to select 'n' units out of 'N' units

- Assigning numbers 1,2,....,N to the N units of sampled population.
- Preparing 'N' identical (in shape, size, colour, etc) slips/cards one per each number.
- Putting the slips in a bag or box and shuffling thoroughly.
- Now drawing 'n' slips from the bag one by one by WOR or WR
- The 'n' units corresponding to the drawn 'n' slips constitute a Random Sample.

** This is one the most reliable methods of selecting random sample and is independent of the properties of the population.

** This method is quite difficult and time consuming in case of large population. In this case an alternative method is Random Numbers method.

Random Numbers Method: This is most practical and inexpensive method for selecting a random sample, which consists in use of **Random Number Tables**. These tables have been constructed that each of the digits 0, 1, 2,...,9 appear with approximately the same frequency and independent of each other.

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

- Assigning numbers 1,2,....,N to the N units of sampled population.

- Selecting any page of the Random Number Tables at random and picking the numbers in a row or column or diagonal at random.
- The population units corresponding to the numbers selected in the above step constitute the random sample.

The following are different sets of random number tables commonly used in practice.

- Tippet's (1927) Random Number Tables – consisting of 10,400 four digit numbers (10,440 X 4 = 41,600 digits)
- Fisher and Yates (1938) Tables – 15,000 digits arranged in 7,500 two digit numbers.
- Kendall and Babington Smith (1939) – 1,00,000 digits grouped into 25,000 four digit numbers .
- Rand Corporation (1955) random number tables – 10,00,000 digits grouped into 2,00,000 five digit numbers.

Theorems:

Notations:

Population	Sample
N = Population Size	n = Sample Size
Yi = The measurement of ith unit of population	yi = The measurement of ith unit of sample
$\bar{Y}_N$ = Population Mean = $\frac{\sum Yi}{N}$	$\bar{y}_n$ = Sample Mean = $\frac{\sum yi}{n}$
S^2 = Population Mean Square	s^2 = Sample Mean Square
= $\frac{\sum(Yi-\bar{Y})^2}{N-1}$ or $\frac{1}{N-1}\left(\sum Yi^2 - N\,(\bar{Y})^2\right)$	= $\frac{\sum(yi-\bar{y})^2}{n-1}$ or $\frac{1}{n-1}\left(\sum yi^2 - n\,(\bar{y})^2\right)$

σ2 = Population Variance

f = n/N = sampling fraction

= $\frac{\sum(Yi-\bar{Y})^2}{N}$ or $\frac{1}{N}\sum Yi^2 - (\bar{Y})^2$

1. **In SRSWOR, the sample mean is an unbiased estimate of the population mean.**

   Proof:

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

   Consider, $E(\bar{y}_n) = E\left(\frac{\sum yi}{n}\right) = \frac{E(\sum_{i=1}^{n} yi)}{n}$ .... (1)

   Let ai = $\begin{cases} 1, \text{ if ith unit is included in the sample (i= 1,2,... N)} \\ 0, \text{ if ith unit is not included in the sample (by WOR)} \end{cases}$

   Now we can write, $\sum_{i=1}^{n} yi = \sum_{i=1}^{N} aiYi$ ....(2)

   and we know that ai ~ Bernoulli dist with p=n/N

   => E(ai) = n/N ....(3)

Substituting (2) and (3) in (1),

$E(\bar{y}_n) = \frac{E(\sum_{i=1}^{n} yi)}{n} = \frac{E(\sum_{i=1}^{N} aiYi)}{n} = \frac{\sum_{i=1}^{N} E(aiYi)}{n} = \frac{\sum_{i=1}^{N} E(ai)Yi}{n}$

$= \frac{\sum_{i=1}^{N}(\frac{n}{N})Yi)}{n} = \frac{\sum_{i=1}^{N} Yi)}{N} = \bar{Y}_N$

Hence, the sample mean is an unbiased estimate of population mean.

Note: The unbiased estimate of Population total $(Y_N) = \hat{Y} = N\,\bar{y}_n$

2. **In SRSWOR, the sample mean square is an unbiased estimate of the population mean square.**

Proof:

To prove, $E(s^2) = S^2$

Consider, $E(s^2) = E(\frac{1}{n-1}(\sum yi^2 - n\ (\bar{y})^2))$

$= \frac{1}{n-1} E(\sum yi^2 - n\ (\frac{\sum yi}{n})^2)$

$= \frac{1}{n-1} E(\sum yi^2 - \frac{(\sum yi)^2}{n})$

$= \frac{1}{n-1} E(\sum yi^2 - \frac{\sum yi^2 + 2\sum\sum yiyj}{n})$

$= \frac{1}{n-1} E((n-1)/n \sum yi^2 - \frac{2\sum\sum yiyj}{n})$

$E(s^2) = \frac{1}{n} E(\sum yi^2) - \frac{E(2\sum\sum yiyj)}{n(n-1)}$ .... (1)

Let ai = { 1, if ith unit is included in the sample
0, if ith unit is not included in the sample (i= 1,2,... N)

and

Let aiaj = { 1, if both ith and jth units are included in the sample
0, otherwise (i≠j= 1,2,... N)

Now we can write, $\sum yi^2 = \sum aiYi^2$ and

$\sum\sum yiyj = \sum\sum aiajYiYj$ .... (2)

and we know that ai ~ Bernoulli dist with p=n/N

=> E(ai) = n/N and E(aiaj) = n(n-1)/N(N-1).......(3)

Substituting (2) and (3) in (1),

$$E(s^2) = \frac{1}{n}E(\sum yi^2) - \frac{E(2\sum\sum yiyj)}{n(n-1)}$$

$$= \frac{1}{n}E(\sum aiYi^2) - \frac{E(2\sum\sum aiajYiYj)}{n(n-1)}$$

$$= \frac{1}{n}\sum E(ai)Yi^2 - \frac{2\sum\sum E(aiaj)YiYj}{n(n-1)}$$

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

$$= \frac{1}{N}\sum Yi^2 - \frac{2\sum\sum YiYj}{N(N-1)}$$

$$= \frac{1}{N}\sum Yi^2 - \frac{(\sum Yi)^2 - \sum Yi^2}{N(N-1)}$$

$$= \frac{1}{N-1}\sum Yi^2 - \frac{(\sum Yi)^2}{N(N-1)}$$

$$= \frac{1}{N-1}\left(\sum Yi^2 - \frac{(\sum Yi)^2}{N}\right)$$

$$= \frac{1}{N-1}\left(\sum yi^2 - N\left(\frac{\sum Yi}{N}\right)^2\right)$$

$$= \frac{1}{N-1}\left(\sum yi^2 - N(\bar{Y})^2\right)$$

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

Hence, s^2 is an unbiased estimate of S^2

3. In SRSWOR, the variance of the estimate of the population mean is,

$$V(\bar{y}_n) = \frac{N-n}{nN} S^2 \text{ or } \left(\frac{1}{n} - \frac{1}{N}\right) S^2 \text{ or } (1-f)\, S^2/n$$

Proof:

Consider, $V(\bar{y}) = E(\bar{y}^2) - (E(\bar{y}))^2$

$$= E\left(\frac{\sum yi}{n}\right)^2 - \bar{Y}_N^2$$

$$= E\,\frac{(\sum yi)^2}{n^2} - \bar{Y}_N^2$$

$$= \frac{1}{n^2}E\left(\sum yi^2 + 2\sum\sum yiyj\right) - \bar{Y}_N^2 \quad \text{........ (1)}$$

Let ai = $\begin{cases} 1, \text{ if ith unit is included in the sample} \\ 0, \text{ if ith unit is not included in the sample } (i = 1,2,\ldots N) \end{cases}$

and

Let aiaj = $\begin{cases} 1, \text{ if both ith and jth units are included in the sample} \\ 0, \text{ otherwise } (i \neq j = 1,2,\ldots N) \end{cases}$

Now we can write, $\sum yi^2 = \sum aiYi^2$ and

$$\sum\sum yiyj = \sum\sum aiajYiYj \quad \ldots\ldots.. (2)$$

and we know that ai ~ Bernoulli dist with p=n/N

=> E(ai) = n/N and E(aiaj) = n(n-1)/N(N-1).......(3)

Substituting (2) and (3) in (1),

$$V(\bar{y}) = \frac{1}{n^2}E\left(\sum yi^2 + 2\sum\sum yiyj\right) - \bar{Y}_N^2$$

$$= \frac{1}{n^2}E\left(\sum aiYi^2 + 2\sum\sum aiajYiYj\right) - \bar{Y}^2$$

$$= \frac{1}{n^2}\left(\sum E(ai)Yi^2 + 2\sum\sum E(aiaj)YiYj\right) - \bar{Y}^2$$

$$= \frac{1}{n^2}\left(\sum\left(\frac{n}{N}\right)Yi^2 + 2\sum\sum \frac{n(n-1)}{N(N-1)}YiYj\right) - \bar{Y}^2$$

$$= \frac{1}{nN}\sum Yi^2 + \frac{(n-1)2\sum\sum YiYj}{nN(N-1)} - \bar{Y}^2$$

$$= \frac{1}{nN}\sum Yi^2 + \frac{(n-1)((\sum Yi)^2 - \sum Yi^2)}{nN(N-1)} - \bar{Y}^2$$

$$= \frac{N-n}{nN(N-1)}\sum Yi^2 + \frac{(n-1)(N\bar{Y})^2}{nN(N-1)} - \bar{Y}^2$$

$$= \frac{N-n}{nN(N-1)}\sum Yi^2 + \frac{(n-1)N\bar{Y}^2}{n(N-1)} - \bar{Y}^2$$

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

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

$$= \frac{N-n}{nN(N-1)}\left[\sum Yi^2 - N\bar{Y}^2\right]$$

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

Note: The variance of the estimate of Population total = $V(\hat{Y})$

$= V(N\bar{y}_n) = N^2 V(\bar{y}_n)$

4. In SRSWR, the sample mean is an unbiased estimate of the population mean.

Proof:

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

Consider, $E(\bar{y}_n) = E(\frac{\sum yi}{n}) = \frac{E(\sum_{i=1}^{n} yi)}{n}$

$= \frac{\sum_{i=1}^{n} E(yi)}{n}$ ..... (1)

Since each y1, y2, y3,....,yn are drawn by WR from population with mean $\bar{Y}$ and variance σ^2 , y1,y2,...,yn are independent and E(yi)= $\bar{Y}$ and $V(yi) = \sigma^2$

From (1), $E(\bar{y}_n) = \frac{\sum_{i=1}^{n} E(yi)}{n} = \frac{\sum_{i=1}^{n} \bar{Y}}{n} = n\,\bar{Y}/n = \bar{Y}$

Hence, the sample mean is an unbiased estimate of population mean.

5. In SRSWR, the variance of the estimate of the population mean is,

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

Proof:

Since each y1, y2, y3,....,yn are drawn by WR from population with mean $\bar{Y}$ and variance σ^2 , y1,y2,...,yn are independent and E(yi)= $\bar{Y}$ and $V(yi) = \sigma^2$

$V(\bar{y}_n) = V(\frac{\sum yi}{n}) = \frac{\sum V(yi)}{n^2} = \frac{\sum \sigma^2}{n^2} = n\,\sigma^2/n^2$

$= \sigma^2/n$

$= \frac{N-1}{nN} S^2$ (since $N\sigma^2 = (N-1)S^2$)

6. SRSWOR is more efficient than SRSWR. i.e., $V(\bar{y})_{WOR} \leq V(\bar{y})_{WR}$

Proof:

$V(\bar{y})_{WOR} = \frac{N-n}{nN} S^2$

$V(\bar{y})_{WR} = \frac{N-1}{nN} S^2$

Clearly , $N-n \leq N-1$

=> $\frac{N-n}{nN} S^2 \leq \frac{N-1}{nN} S^2$ (since all coefficients are non-negative)

=> $V(\bar{y})_{WOR} \leq V(\bar{y})_{WR}$

Hence SRSWOR is efficient than SRSWR.

Merits and Limitations of Simple Random Sampling:

Merits:

1. Since the sample is selected at random giving each unit an equal chance of being selected, personal bias is completely eliminated.
2. Simple Random Sample is more presentative of the population compared with non-random sample
3. This is very simple and efficient compared to other random sampling methods when the population is small and homogeneous.
4. As sample size ‘n’ increases, it provides more efficient estimates for the population parameters.

Limitations:

1. The selection simple random sample requires a complete up-to-date frame. Practically it may not possible to identity the all units before sample is drawn.
2. It requires more time and money for selecting the units that spread widely.
3. It usually requires large sample size for efficient estimates compared to the stratified random sampling.
4. It is inefficient if the population is heterogeneous.

Problem:

Suppose that a population consists of 6 units with measurements 2, 4, 6, 8, 10 and 12. Write all the possible samples of size 2 by without replacement from the population and verify

a) The sample mean is unbiased estimate of the population
b) The sample mean square is unbiased estimate if the population mean square

Also calculate the sampling variance of the estimate, sample mean and verify

i) It with the formula of variance
ii) SRSWOR is efficient than SRSWR and find the gain in efficiency.

Solution:

Given, N = 6

Yi : 2, 4, 6, 8, 10, 12

=> Population mean, $\overline{Y} = \frac{\sum Yi}{N} = \frac{2+4+6+8+10+12}{6} = 7$

and Population mean square, $S^2 = \frac{1}{N-1} (\sum Yi^2 - N\, (\bar{Y})^2)$

$= \frac{1}{6-1} (2^2+4^2+6^2+8^2+10^2+12^2 - 6(7)^2)$

$= \frac{1}{5} (70) = 14$

also given n = 2

=> the no. of possible samples of size 2 from 6 units by WOR = $6c_2$ = 15

The list of samples and verification of the bits (a), (b) and (i), (ii) are given below

Sample No.	Sample values (n=2) (y1, y2)	Sample mean $\bar{y}$= (y1+y2)/2	$(\bar{y})^2$	Sample mean square $s^2 = \frac{1}{n-1}(\sum yi^2 - n(\bar{y})^2)$ $s^2 = y1^2+y2^2 - 2(\bar{y})^2$
1	2, 4	3	9	2
2	2, 6	4	16	8
3	2, 8	5	25	18
4	2, 10	6	36	32
5	2, 12	7	49	50
6	4, 6	5	25	2
7	4, 8	6	36	8
8	4, 10	7	49	18
9	4, 12	8	64	32
10	6, 8	7	49	2
11	6, 10	8	64	8
12	6, 12	9	81	18
13	8, 10	9	81	2
14	8, 12	10	100	8
15	10. 12	11	121	2
Total		**105**	**805**	**210**

a) Sampling mean, $E(\bar{y}) = \frac{\sum \bar{y}}{15}$ = 105/15 = 7 = $\bar{Y}$

Therefore, sample mean is an unbiased estimate of population mean

b) $E(s^2) = \frac{\sum s2}{15}$ = 210/15 = 14 = S^2

Therefore, sample mean square is an unbiased estimate of population mean square

and Sampling variance, $V(\bar{y}) = \frac{\sum \bar{y}2}{15} - \bar{Y}^2$ = 805/15 – $(7)^2$ = 4.6667

i) by formula, $V(\bar{y})_{WOR} = \frac{N-n}{nN} S^2$ = (6-2)x14/(2x6) = 56/12 = 4.6667

Therefore, sampling variance agrees with formula

ii) $V(\bar{y})_{WR} = \frac{N-1}{nN} S^2$ = (6-1)x14/(2x6) = 70/12 = 5.8333

$V(\bar{y})_{WOR}$ = 4.6667 < $V(\bar{y})_{WR}$ = 5.8333

Therefore, SRSWOR is efficient than SRSWR

The Efficiency = 5.8333/4.6667 = 1.25

The gain in efficiency = 25 %

i.e., for the given figures, SRSWOR is 25% more efficient than SRSWOR

Self Assessment Questions:

Multiple Choice Questions (MCQ):

1. If the sample is selected by giving equal importance to each and every unit of the population then the method is known as ( )
   a) simple random sampling b) systematic sampling
   c) judgement sampling d) quota sampling
2. The no. of possible samples of size 3 that can be drawn from a population of size 10 by SRSWOR is ( )
   a) 10 x 3 b) ${}^{10}C_3$ c) 10^3 d) 3^{10}
3. The no. of possible samples of size 3 that can be drawn from a population of size 10 by SRSWR is ( )
   a) 10 x 3 b) ${}^{10}C_3$ c) 10^3 d) 3^{10}
4. In SRSWOR, the probability of selecting a specified unit at i^{th} (i=1,2,..n) draw is equal to ( )
   a) 1/N b) n/N c) 1/(N-i+1) d) N/n
5. In SRSWOR, the probability of including a specified unit in to the sample is ( )
   a) 1/N b) n/N c) 1/(N-1) d) N/n
6. Which of the following method is used to draw a random sample ( )
   a) lottery method b) random numbers method
   c) roulette wheel d) All of these
7. If N=population size , n=sample size , $\bar{y}$=sample mean, $\bar{Y}$=population mean, then by simple random sampling the estimate of population total $\hat{Y}$ is ( )
   a) $n\bar{y}$ b) $N\bar{y}$ c) $\bar{y}$ d) $\bar{Y}$
8. If S^2 is population mean square, in SRSWOR, the variance of sample mean is ( )
   a) $V(\bar{y})= \frac{N-n}{Nn} S^2$ b) $V(\bar{y})= \frac{n-N}{Nn} S^2$ c) $V(\bar{y})= \frac{N-n}{(N-1)n} S^2$ d) $V(\bar{y})= \frac{n-N}{(N-1)n} S^2$
9. If S^2 is population mean square, in SRSWR, the variance of sample mean is ( )
   a) $V(\bar{y})= \frac{N-n}{Nn} S^2$ b) $V(\bar{y})= \frac{n-N}{Nn} S^2$ c) $V(\bar{y})= \frac{N-1}{Nn} S^2$ d) $V(\bar{y})= \frac{N}{(N-1)n} S^2$
10. Sampling fraction is ( )
   a) N/n b) n/N c) 1/N d) 1/n
11. Which of the following is always true ( )

a) $V(\bar{y}_{WOR}) = V(\bar{y}_{WR})$ b) $V(\bar{y}_{WOR}) \geq V(\bar{y}_{WR})$

c) $V(\bar{y}_{WOR}) \leq V(\bar{y}_{WR})$ d) None

Descriptive Questions:

1. Explain SRSWR and SRSWOR.
2. In SRSWOR, Show that $E(s^2)=S^2$
3. In SRSWOR, Derive the formula of the variance of the estimate of the population mean.