SIMPLE RANDOM SAMPLING

PRESENTED
BY
MR. PRADIP PANDA
ASSTT. PROF., DEPTT. OF STATISTICS
SERAMPORE COLLEGE

DEFINITION: Simple random sampling is a method of drawing a sample such that each and every unit in the population has an equal chance of being included in the sample.

SIMPLE RANDOM SAMPLING(SRS)

- **SRS WITH REPLACEMENT(SRSWR)**
- **SRS WITHOUT REPLACEMENT(SRSWOR)**

SRSWR Each unit is replaced before the next drawing.

SRSWOR Each unit is not replaced before the next drawing.

Mr. Pradip Panda, Asstt. Prof., Department of Statistics, Serampore College

RESULTS RELATED TO SRSWR

- In SRSWR ,the probability that i^{th} unit is selected at any draw

$$P(u_i) = \frac{1}{N}, \forall i$$

- π_i is the probability that i^{th} unit is selected in the sample of size n.

$$\pi_i = 1 - P(i^{th} \ unit \ is \ not \ selected)$$

$$= 1 - [1 - \frac{1}{N}]^n$$

- π_{ij} is the probability that both the i^{th} and j^{th} unit is selected in the sample of size n.

$$\pi_{ij} = P(i^{th}\ unit\ and\ j^{th}\ unit\ is\ selected\ in\ the\ sample)$$

$$= P(i^{th}\ unit\ is\ selected).P(j^{th}\ unit\ is\ selected)$$

$$= \left[1-\{1-\frac{1}{N}\}^n\right].\left[1-\{1-\frac{1}{N}\}^n\right]$$

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

Theorem I: In case of SRSWR , sample mean is an unbiased estimator of the population mean.

Proof: Sample size = n and population size = N

Let Y_i be the value of the character y for the ith unit of the population.

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i = Population mean \text{ and } \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = Sample mean.$$

$$E(y_i) = \sum_{i=1}^{N} Y_i . \frac{1}{N} = \bar{Y}, \forall i$$

$$E(\bar{y}) = E(\frac{1}{n}\sum_{i=1}^{n} y_i) = \frac{1}{n}\sum_{i=1}^{n} E(y_i) = \frac{1}{n}.n.\bar{Y}$$

$\therefore$ *Sample mean is the unbiased estimator of the population mean.*

$$So, \hat{\bar{Y}} = \bar{y}$$

Corollary:

$$E(N.\bar{y}) = N.E(\bar{y}) = N.\bar{Y} = \sum_{i=1}^{N} Y_i = Y, \textit{ the population total.}$$

$$\therefore \hat{Y} \textit{ is an unbiased estimator of population total} = N.\bar{y}$$

Theorem II: The sampling variance of sample mean is given by

$$V\left(\hat{\bar{Y}}\right) = V(\bar{y}) = \frac{\sigma^2}{n};$$

$$\textit{where, } \sigma^2 = \frac{1}{N}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2 \textit{ is the population variance.}$$

Proof:

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

$$= E[\{\bar{y} - \bar{Y}\}^2] = E[\{\frac{1}{n}\sum_{i=1}^{n} y_i - \bar{Y}\}^2]$$

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

$$= \frac{1}{n^2} E[\sum_{i=1}^{n}(y_i - \bar{Y})^2 + \sum\sum_{i \neq j}(y_i - \bar{Y})(y_j - \bar{Y})]$$

$$= \frac{1}{n^2} [\sum_{i=1}^{n} E(y_i - \bar{Y})^2 + \sum\sum_{i \neq j} E(y_i - \bar{Y})(y_j - \bar{Y})]$$

$$= \frac{1}{n^2}\sum_{i=1}^{n} E(y_i - \bar{Y})^2 = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{i'=1}^{N}(Y_{i'} - \bar{Y})^2 . \frac{1}{N}$$

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

Corollary:

$$i)\ V(\hat{Y}) = V(N.\bar{y}) = N^2.V(\bar{y}) = \frac{N^2.\sigma^2}{n}$$

$$ii)\ The\ r.s.e.\ of\ \bar{y} = \frac{\sqrt{V(\bar{y})}}{E(\bar{y})} = \frac{1}{\sqrt{n}}.\frac{\sigma}{\bar{Y}} = \frac{C_y}{\sqrt{n}}.$$

C_y *is the population coefficient of variation.*

Theorem III: An unbiased estimator of sampling variance of sample mean is given by

$$\hat{V}(\bar{y}) = \frac{\hat{\sigma}^2}{n} = \frac{s^2}{n};$$

$$where\ s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2, is\ the\ sample\ mean\ square.$$

Proof:

$$V(\bar{y}) = \frac{\sigma^2}{n}$$

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

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

$$=\frac{1}{n-1}\sum_{i=1}^{n}\;[\sum_{i'=1}^{N}Y_{i'}^{2}.\frac{1}{N}]-\frac{n}{n-1}[\bar{Y}^{2}+\frac{\sigma^{2}}{n}]$$

$$=\frac{n}{n-1}\sum_{i'=1}^{N}Y_{i'}^{2}.\frac{1}{N}-\frac{n}{n-1}\bar{Y}^{2}-\frac{\sigma^{2}}{n-1}$$

$$=\frac{n}{n-1}[\frac{1}{N}\sum_{i'=1}^{N}Y_{i'}^{2}-\bar{Y}^{2}]-\frac{\sigma^{2}}{n-1}$$

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

$$\therefore \hat{\sigma}^{2}=s^{2}\ or,\hat{V}(\bar{y})=\frac{\hat{\sigma}^{2}}{n}=\frac{s^{2}}{n}.$$

$$=\sigma^{2}$$

Corollary: $Unbiased\,estimate\,of\ r.s.e.\ will\ be\ \hat{C}(\bar{y})=\frac{1}{\sqrt{n}}.\frac{s}{\bar{y}}.$

RESULTS RELATED TO SRSWOR

❖ In SRSWOR ,the probability that i^{th} unit is selected at any draw is 1/N.

Proof:

Let U_{i1} be the event that the i^{th} unit is drawn at the 1^{st} draw.

$P(U_{i1}) = 1/N$, if there are N units in the population.

$P(U_{i2}) = P(i^{th}$ unit appears at the 2^{nd} draw.)

$= P(i^{th}$ unit does not appear at the 1^{st} draw.)$P(i^{th}$ unit appears at the 2^{nd} draw /i^{th} unit does not appear at the 1^{st} draw.)

Therefore, $P(U_{i2}) = (1 - 1/N).\ 1/(N-1) = \{(N-1)/N\}\{1/(N-1)\}$
$= 1/N$

$P(U_{ir})$ = P(i^{th} unit appears at the r^{th} draw.)

= P(i^{th} unit does not appear upto $(r-1)^{th}$ draw.)
P(i^{th} unit appears at the r^{th} draw /i^{th} unit does not appear upto $(r-1)^{th}$ draw.)

$$=\left(1-\frac{1}{N}\right)\left(1-\frac{1}{N-1}\right)\left(1-\frac{1}{N-2}\right)\cdots\cdots\left(1-\frac{1}{N-r+2}\right)\frac{1}{N-r+1}$$

$$=\frac{N-1}{N}\cdot\frac{N-2}{N-1}\cdot\frac{N-3}{N-2}\cdots\cdots\frac{N-r+1}{N-r+2}\cdot\frac{1}{N-r+1}=\frac{1}{N}$$

Here, $P(U_{ir})$ is independent of r.

So, in SRSWOR ,the probability of drawing i^{th} unit at any draw is 1/N.

- π_i is the probability that i^{th} unit is selected in the sample of size n.

$$\pi_i = P(A_1 \cup A_2 \cup \ldots\ldots \cup A_n)$$

A_j be the event that the i^{th} unit is selected at the j^{th} draw. j = 1(1)n.

In SRSWOR , A_j's are mutually exclusive events, j = 1(1)n.

$$So,\ \pi_i = \sum_{j=1}^{n} P(A_j) = \sum_{j=1}^{n} \frac{1}{N} = \frac{n}{N}$$

- π_{ij} is the probability that both the i^{th} and j^{th} unit is selected in the sample of size n.

$$\pi_{ij} = P(i^{th}\ unit\ and\ \ j^{th}\ unit\ is\ selected\ in\ the\ sample)$$

$$= P(i^{th}\ unit\ is\ selected).$$

$$P(j^{th}\ unit\ is\ selected\ /\ i^{th}\ unit\ is\ already\ selected)$$

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

Theorem I: Sample mean is the unbiased estimator of population mean.

Proof: Same as in case of SRSWR.

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

$Where, V(\bar{y})\ is\ the\ sampling\ variance\ of\ sample\ mean.$

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

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

$$\sigma^2 = Population\ variance = \frac{1}{N}\sum_{i=1}^{N}(Y_i - \bar{Y})^2.$$

Proof:

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

$$= E[\{\bar{y} - \bar{Y}\}^2] = E[\{\frac{1}{n}\sum_{i=1}^{n} y_i - \bar{Y}\}^2]$$

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

$$= \frac{1}{n^2} E[\sum_{i=1}^{n}(y_i - \bar{Y})^2 + \sum\sum_{i \neq j}(y_i - \bar{Y})(y_j - \bar{Y})]$$

$$= \frac{1}{n^2}[\sum_{i=1}^{n} E(y_i - \bar{Y})^2 + \sum\sum_{i \neq j} E(y_i - \bar{Y})(y_j - \bar{Y})]$$

$$= \frac{1}{n^2}[\sum_{i=1}^{n}\{\sum_{i'=1}^{N} (Y_{i'} - \bar{Y})^2 . \frac{1}{N}\} + \sum\sum_{i \neq j}\{\sum\sum_{i' \neq j'} (Y_{i'} - \bar{Y})(Y_{j'} - \bar{Y}). \frac{1}{N(N-1)}\}]$$

$$= \frac{1}{n^2}[n\sigma^2 + \frac{n(n-1)}{N(N-1)} \sum\sum_{i' \neq j'}(Y_{i'} - \bar{Y})(Y_{j'} - \bar{Y})]$$

Mr. Pradip Panda, Asstt. Prof., Department of Statistics, Serampore College

$$\left[\sum_{i=1}^{N}\left(Y_i-\bar{Y}\right)\right]^2=\sum_{i=1}^{N}\left(Y_i-\bar{Y}\right)^2+\sum\sum_{i\neq j=1}^{N}\left(Y_i-\bar{Y}\right)\left(Y_j-\bar{Y}\right)$$

$$or,\ 0=N.\sigma^2+\sum\sum_{i\neq j=1}^{N}\left(Y_i-\bar{Y}\right)\left(Y_j-\bar{Y}\right)$$

$$or,\ \sum\sum_{i\neq j=1}^{N}\left(Y_i-\bar{Y}\right)\left(Y_j-\bar{Y}\right)=-N\sigma^2$$

$$\therefore V(\bar{y})=\frac{1}{n^2}\left[n\sigma^2-\frac{n(n-1)}{N(N-1)}.N\sigma^2\right]$$

$$=\frac{1}{n}\left[\frac{N-1-n+1}{(N-1)}.\sigma^2\right]\quad=\frac{N-n}{(N-1)}.\frac{\sigma^2}{n}$$

Theorem III: Show that sample mean square is the unbiased estimator of population mean square in SRSWOR.

Proof:

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

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

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

$$= \frac{1}{n-1}\left[\sum_{i=1}^{n}\sum_{i'=1}^{N}Y_{i'}^2.\frac{1}{N} - n\left\{V(\bar{y})+E^2(\bar{y})\right\}\right]$$

Mr. Pradip Panda, Asstt. Prof., Department of Statistics, Serampore College

$$=\frac{1}{n-1}\left[\sum_{i=1}^{n}\sum_{i'=1}^{N}Y_{i'}^{2}.\frac{1}{N}-n\left\{\frac{N-n}{N-1}.\frac{\sigma^{2}}{n}+\bar{Y}^{2}\right\}\right]$$

$$=\frac{n}{n-1}\left[\frac{1}{N}\sum_{i'=1}^{N}Y_{i'}^{2}-\bar{Y}^{2}-\frac{N-n}{N-1}.\frac{\sigma^{2}}{n}\right]$$

$$=\frac{n}{n-1}\left[1-\frac{N-n}{N-1}.\frac{1}{n}\right]\sigma^{2}=\frac{n}{n-1}\left[\frac{Nn-n-N+n}{n(N-1)}\right]\sigma^{2}$$

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

$$\therefore E(s^{2})=S^{2}\ and\ \hat{V}(\bar{y})=\frac{N-n}{N}.\frac{s^{2}}{n}$$

Efficiency of sample mean under SRSWOR over SRSWR :

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

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

$$= V(\bar{y})_{WOR} + a \ positive \ quantity.$$

$\therefore V(\bar{y})_{WR} > V(\bar{y})_{WOR}$. *Equality holds if N is very large compared to n.*

So, SRSWOR is more efficient than SRSWR.

Variance of the estimate of the population total and its unbiased estimate :

In SRSWR

$N\bar{y}$ *is the unbiased estimator of the population total.*

$$V(N\bar{y}) = N^2 V(\bar{y}) = N^2 . \frac{\sigma^2}{n}$$

$$\hat{V}(N\bar{y}) = N^2 \hat{V}(\bar{y}) = N^2 . \frac{\hat{\sigma}^2}{n} = N^2 . \frac{s^2}{n}$$

$$\because s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2 \text{ is the unbiased estimator of the}$$

population variannce in SRSWR.

In SRSWOR

$N\bar{y}$ *is the unbiased estimator of the population total.*

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

$$\hat{V}(N\bar{y}) = N^2 \hat{V}(\bar{y}) = N^2 . \frac{N-n}{N} . \frac{\hat{S}^2}{n} = N^2 . \frac{N-n}{N} . \frac{s^2}{n}$$

$\because s^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2$ *is the unbiased estimator of the population mean square in SRSWOR.*

Learn more.......