Introduction to Sampling Theory

Lecture 6
Simple Random Sampling

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Slides can be downloaded from

http://home.iitk.ac.in/~shalab/sp

Estimation of Population Mean and Variance: Notations

$Y_1, Y_2, \ldots, Y_N$: **Population**

$y_1, y_2, \ldots, y_n$: **Sample**

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

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$: **Sample mean**

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(Y_i - \bar{Y}\right)^2 = \frac{1}{N-1}\left(\sum_{i=1}^{N} Y_i^2 - N\bar{Y}^2\right)$$

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

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

Estimation of Population Mean and Population Variance

One of the main objectives after the selection of a sample is to know about the tendency of the data to cluster around the central value and the scatterdness of the data around the central value in the population.

Popular choices are arithmetic mean and variance.

Population mean is generally measured by arithmetic mean (or weighted arithmetic mean).

Estimation of Population Mean:

Various estimators for estimating the population mean and population variance are available.

Sample arithmetic mean is more popular than other estimators because it possess nice statistical properties.

Estimation of Population Mean:

Let us consider the sample arithmetic mean $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ **as an estimator of population mean** $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$

Estimate population mean $\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} Y_i$ **by sample mean** $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$

$\bar{y}$ **is an unbiased estimator of** $\bar{Y}$ **under SRSWR and SRSWOR cases.**

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

Estimation of population mean:
Interpretation of unbiased estimator

Population: $X_1 = 1, X_2 = 3, X_3 = 5$

Population mean = 3

Number of Samples of size 2= $\binom{3}{2}$**=3**

Suppose the population mean is unknown.

Use sample arithmetic mean to estimate the population mean.

Estimation of population mean:

Interpretation of unbiased estimator

Sample arithmetic mean is an unbiased estimator of population mean.

Sample 1=(1,3) **Sample mean ($\bar{x}_1$) = 2**

Sample 2=(3,5) **Sample mean ($\bar{x}_2$) = 4**

Sample 3=(1,5) **Sample mean ($\bar{x}_3$) = 3**

$$\bar{x} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3}{3} = \frac{2+4+3}{3} = 3 =$$ **Population mean**

Proof: $\bar{y}$ is an unbiased estimator of $\bar{Y}$ in SRSWOR

Let $P_j(i)$ denotes the probability of selection of ith unit at j^{th} stage.

$$E(\bar{y}) = \frac{1}{n}\sum_{j=1}^{n} E(y_j)$$

$$= \frac{1}{n}\sum_{j=1}^{n}\left[\sum_{i=1}^{N} Y_i P_j(i)\right]$$

$$= \frac{1}{n}\sum_{j=1}^{n}\left[\sum_{i=1}^{N} Y_i . \frac{1}{N}\right]$$

$$= \frac{1}{n}\sum_{j=1}^{n} \bar{Y}$$

$$= \bar{Y}$$

Proof: $\bar{y}$ is an unbiased estimator of $\bar{Y}$ in SRSWR

Let $P_i = \frac{1}{N}, i = 1, 2, \ldots, N$ is the probability of selection of a unit.

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

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

$$= \frac{1}{n} \sum_{i=1}^{n} \left(Y_1 P_1 + \ldots + Y_N P_N\right)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \bar{Y}$$

$$= \bar{Y}$$

Sample Mean: Example

***Y*: Height of students in a class**

***N* = 10 : Number of students in the class (Population size)**

***n* = 3 : Number of students in the sample (Sample size)**

Name of Student	Y_i = Height of students (in Centimeters)
A	Y_1= 151
B	Y_2= 152
C	Y_3 = 153
D	Y_4= 154
E	Y_5 = 155
F	Y_6= 156
G	Y_7 = 157
H	Y_8= 158
I	Y_9 = 159
J	Y_{10}= 160

Sample Mean: Example

n = 3 : Number of students in the sample (Sample size)

y_i : Height of ith student in the sample

Sample 1: 3rd , 7th and 9th student

$y_1 = Y_3$ = 153 cms., $y_2 = Y_7$ = 157 cms., $y_3 = Y_9$ = 159 cms.

Sample mean 1 ($\bar{y}_1$) = (153 + 157 + 159)/3= 156.33 cms.

Sample 2: 2nd , 5th and 4th student

$y_1 = Y_2$ = 152 cms., $y_2 = Y_5$ = 155 cms., $y_3 = Y_4$ = 154 cms.

Sample mean 2 ($\bar{y}_2$) = (152 + 155 + 154)/3= 153.66 cms.

Sample Mean: Example

Sample 3: 1st , 6th and 10th student

$y_1 = Y_1$ = 151 cms., $y_2 = Y_6$ = 156 cms., $y_3 = Y_{10}$ = 160 cms.

Sample mean 3 ($\bar{y}_3$) = (151 + 156 + 160)/3= 155.66 cms.

Population mean = $\bar{Y} = \frac{1}{10} \sum_{i=1}^{10} Y_i$=155.5

Thus we have

$\bar{y}_1$ = 156.33 cms.

$\bar{y}_2$ = 153.66 cms.

$\bar{y}_3$ = 155.66 cms.

The total number of samples = $\binom{10}{3} = 120$, find their sample means.

Variance of Sample Mean

Population variability is generally measured by variance.

Several sample can be drawn by SRSWR as well as SRSWOR from a population.

Each sample will have different sample mean.

Sample mean is a statistic, i.e., a function of random variables.

So sample mean will also have variance.

Consider an example.

Variance of Sample Mean: Example

***Y*: Height of students in a class**

***N* = 10 : Number of students in the class (Population size)**

***n* = 3 : Number of students in the sample (Sample size)**

Name of Student	Y_i = Height of students (in Centimeters)
A	Y_1= 151
B	Y_2= 152
C	Y_3 = 153
D	Y_4= 154
E	Y_5 = 155
F	Y_6= 156
G	Y_7 = 157
H	Y_8= 158
I	Y_9 = 159
J	Y_{10}= 160

Variance of Sample Mean: Example

n = 3 : Number of students in the sample (Sample size)

y_i : Height of ith student in the sample

Sample 1: 3rd , 7th and 9th student

$y_1 = Y_3$ = 153 cms., $y_2 = Y_7$ = 157 cms., $y_3 = Y_9$ = 159 cms.

Sample mean 1 ($\bar{y}_1$) = (153 + 157 + 159)/3= 156.33 cms.

Sample 2: 2nd , 5th and 4th student

$y_1 = Y_2$ = 152 cms., $y_2 = Y_5$ = 155 cms., $y_3 = Y_4$ = 154 cms.

Sample mean 2 ($\bar{y}_2$) = (152 + 155 + 154)/3= 153.66 cms.

Variance of Sample Mean: Example

Sample 3: **1st , 6th and 10th student**

$y_1 = Y_1$ = 151 cms., $y_2 = Y_6$ = 156 cms., $y_3 = Y_{10}$ = 160 cms.

Sample mean 3 ($\bar{y}_3$) = (151 + 156 + 160)/3= 155.66 cms.

Population mean = $\bar{Y} = \frac{1}{10} \sum_{i=1}^{10} Y_i$=155.5

Thus we have

$\bar{y}_1$ = 156.33 cms.

$\bar{y}_2$ = 153.66 cms.

$\bar{y}_3$ = 155.66 cms.

The total number of samples = $\binom{10}{3} = 120$, find their variances.

Variance of Sample Mean

Variance of sample mean under SRSWOR

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

Variance of sample mean under SRSWR

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

Proof: Variance of Sample Mean

Assume that each observation has same variance var(Y_i)= σ^2

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

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

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

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

$$= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 + \frac{K}{n^2}$$

$$= \frac{N-1}{Nn}S^2 + \frac{K}{n^2}$$

where $K = \sum_{i}^{n}\sum_{\neq j}^{n}E(y_i - \bar{Y})(y_j - \bar{Y}).$

Proof: Variance of Sample Mean: SRSWOR

$$K = \sum_{i}^{n} \sum_{\neq j}^{n} E(y_i - \bar{Y})(y_j - \bar{Y})$$

Consider

$$E(y_i - \bar{Y})(y_j - \bar{Y}) = \frac{1}{N(N-1)} \sum_{k}^{N} \sum_{\neq l}^{N} (y_k - \bar{Y})(y_l - \bar{Y}).$$

Since

$$\left[\sum_{k=1}^{N} (y_k - \bar{Y}) \right]^2 = \sum_{k=1}^{N} (y_k - \bar{Y})^2 + \sum_{k}^{N} \sum_{\neq l}^{N} (y_k - \bar{Y})(y_l - \bar{Y})$$

$$0 = (N-1)S^2 + \sum_{k}^{N} \sum_{\neq l}^{N} (y_k - \bar{Y})(y_l - \bar{Y})$$

$$E(y_i - \bar{Y})(y_j - \bar{Y}) = \frac{1}{N(N-1)} \left[-(N-1)S^2 \right] = -\frac{S^2}{N}.$$

Thus $K = -n(n-1)\dfrac{S^2}{N}$

Proof: Variance of Sample Mean: SRSWOR

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

$$= \frac{N-1}{Nn} S^2 + \frac{K}{n^2}$$

where $K = -n(n-1)\frac{S^2}{N}$

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

$$= \frac{N-n}{Nn} S^2.$$

Proof: Variance of Sample Mean: SRSWR

$$K = \sum_{i}^{n} \sum_{\neq j}^{n} E(y_i - \overline{Y})(y_j - \overline{Y})$$

$$= \sum_{i}^{n} \sum_{\neq j}^{n} E(y_i - \overline{Y}) E(y_j - \overline{Y})$$

= 0 **because** i^{th} **and** j^{th} **draws are independent**

$$V(\overline{y}) = E(\overline{y} - \overline{Y})^2$$

$$= \frac{N-1}{Nn} S^2 + \frac{K}{n^2}$$

$$V(\overline{y}_{WR}) = \frac{N-1}{Nn} S^2.$$