

Standard Error in Simple Random Sampling

S. Rithvik Reddy

IIT H (CSE)

April 12, 2021

Let's understand some terms.

What is simple random sampling ?

- It is a sampling model used in Inferential statistics.
- **Inferential statistics** is one of the two statistical methods employed to analyse data. It is used make inferences about a population by examining one or more random samples drawn from that population.

Eg:- We can arrive at a good estimation of average CGPA of 3000 IIT H students just by observing few 100 students CGPA and using inferential statistics.

- Descriptive statistics allow us to describe a data set, while inferential statistics allow us to make inferences based on a data set.

Simple Random Sampling

- Simple random sampling (SRS) is a method of selection of a sample comprising of n number of sampling units out of the population having N number of sampling units such that every sampling unit has an **equal chance** of being chosen.

Example

If you want to know the rating that IIT H students give to their faculty. You have two options

- ① Ask all the students to give a review.
- ② Ask few 100 and use SRS concept to find the range of average rating.

Obviously second one is more convenient and less time taking at the cost of precision.

Simple random sampling Vs Random sampling

- The above example is branch dependent as each branch will not have same amount of students or same faculty.
- If CSE has the highest no of students among all branches then the chance of a person being from CSE in the 100 you who gave review will be high and the data will be biased.
- If all branches has the same no of students then the chance of a person being from a given branch in the 100 you who gave review will be same and the data will be unbiased.
- The former case represents the Random sampling while the later one represents Simple Random Sampling

Simple Random Sampling can be done in two ways

SRSWOR

Simple Random sampling With out Replacement is a method of selection of n units out of the N units one by one such that at any stage of selection, any one of the remaining units have the same chance of being selected.

- A selection in this way can be done in $\binom{N}{n}$ ways.

SRSWR

Simple Random Sampling With Replacement is a method of selection of n units out of the N units one by one such that at each stage of selection, each unit has an equal chance of being selected, i.e., $1/N$.

- A selection in this way can be done in N^n ways.

Notations

n : Sample class size.

N : Population class size.

y : Sample under consideration.

y_i : Value of the characteristic i^{th} sample in the sample class.

Y : Element of population class.

Y_i : Value of the characteristic i^{th} element in the population class.

$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$: Average of sample class.

$\bar{Y} = \frac{1}{N} \sum_{i=1}^N Y_i$: Average of population class.

$S^2 = \frac{1}{N-1} \sum_{i=1}^N (Y_i - \bar{Y})^2$: S =Std dev of the population class.

$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (Y_i - \bar{Y})^2$: Variance of the population class.

$s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$: S =Std dev of the sample class.

Standard Error of Mean (SE)

- Knowing the mean of the sample (\bar{y}) we can estimate the mean of population (\bar{Y}).

$$\bar{Y} = \bar{y} \pm SE \quad (1)$$

$$SE = \frac{s}{\sqrt{n}} \quad (2)$$

Where s is the standard deviation of means of all possible sample classes of size n .

$$\Rightarrow SE = \sqrt{\frac{Var(\bar{y})}{n}} \quad (3)$$

Variance of the \bar{y}

Assume variance of each term is σ^2 .

$$\begin{aligned}\text{Var}(\bar{y}) &= E(\bar{y} - \bar{Y})^2 \\&= E \left[\frac{1}{n} \sum_{i=1}^n (y_i - \bar{Y}) \right]^2 \\&= E \left[\frac{1}{n^2} \sum_{i=1}^n (y_i - \bar{Y})^2 + \frac{1}{n^2} \sum_{1 \leq i \neq j \leq n} (y_i - \bar{Y})(y_j - \bar{Y}) \right] \\&= \frac{1}{n^2} \sum_{i=1}^n E(y_i - \bar{Y})^2 + \frac{1}{n^2} \sum_{1 \leq i \neq j \leq n} E(y_i - \bar{Y})(y_j - \bar{Y})\end{aligned}$$

$$\text{Let } K = \sum_{1 \leq i \neq j \leq n} E(y_i - \bar{Y})(y_j - \bar{Y})$$

$$= \frac{1}{n^2} \sum_{i=1}^n \sigma^2 + \frac{K}{n^2} \quad (4)$$

$$= \frac{1}{n^2} n \sigma^2 + \frac{K}{n^2} \quad (5)$$

$$= \frac{N-1}{Nn} S^2 + \frac{K}{n^2} \quad (6)$$

- Finding the value of K in case of SRSWR and SRSWOR allows us to calculate the variance of mean.

K value in case of SRSWOR

$$K = \sum_{1 \leq i \neq j \leq n} E(y_i - \bar{Y})(y_j - \bar{Y})$$

Consider

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

Since

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

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

$$\Rightarrow E(y_i - \bar{Y})(y_j - \bar{Y}) = \frac{1}{N(N-1)}(N-1)(-S^2)$$

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

Putting this value in (6) gives us

$$Var(\bar{y})_{WOR} = \frac{N-1}{Nn} S^2 + \frac{n-1(-S^2)}{Nn} \quad (7)$$

$$= \frac{N-n}{Nn} S^2 \quad (8)$$

K value in case of SRSWR

$$K = \sum_{1 \leq i \neq j \leq n} \sum E(y_i - \bar{Y})(y_j - \bar{Y})$$

Since we are selecting the samples with replacements choosing i^{th} and j^{th} sample is independent of each other. So,

$$K = \sum_{1 \leq i \neq j \leq n} \sum E(y_i - \bar{Y})E(y_j - \bar{Y})$$

$$= 0$$

(Since deviation about mean is 0)

Putting $K=0$ in (6) we get

$$V(\bar{y})_{WR} = \frac{N-1}{Nn} S^2 \quad (9)$$

Standard Error of mean

From equation (8) standard error of mean of sample class chosen without repetition is

$$SE_{WOR} = \frac{s}{\sqrt{n}} \quad (10)$$

$$= \sqrt{\frac{V(\bar{y})_{WOR}}{n}} \quad (11)$$

$$= \sqrt{\frac{N-n}{Nn^2}} S \quad (12)$$

Standard Error of Mean

From equation (9) standard error of mean of sample class chosen with repetition

$$SE_{WR} = \sqrt{\frac{V(\bar{y})_{WR}}{n}} \quad (13)$$

$$= \sqrt{\frac{N-1}{Nn^2}} S \quad (14)$$

UGC/MATH 2018(June maths set-a), Q.58

- ① A simple random variable of size n will be drawn from a class of 125 students, and the mean mathematics score of the sample will be computed, If the standard error of the sample mean for "with replacement sampling" is twice as much as the standard error of the sample mean for "without replacement sampling", the value of n is ?
- ① 32
 - ② 63
 - ③ 79
 - ④ 94

Solution

Given to find the value of n if $2 \times SE_{WOR} = SE_{WR}$. From (12) and (14) we can write

$$\begin{aligned}2\sqrt{\frac{N-n}{Nn^2}}S &= \sqrt{\frac{N-1}{Nn^2}}S \\ \implies 4(N-n) &= N-1 \\ \implies 4N+1-N &= 4n \\ \implies 4n &= 3(125)+1 \\ \implies n &= 94\end{aligned}$$

Therefore the sample size for the given condition to be met is $n=94$. (**Option 4**)