STAT 403 Tutorial · Week 12 .

— Stratified random sampling . & SRS .

Example . Consider a population of 6 students . Suppose we know the test scores of the students to be :

Student	1	2	3	4	5	6
Score	66	59	70	83	82	71

a) . Find the mean $\bar{Y}$ and variance S^2 of the population ;

Solution : Denote the population by $\mathcal{U} = \{1,2,3,4,5,6\}$ (So $N=6$) . and the score of each student (unit) by y_i $(i=1,2,3,4,5,6)$. Then the population mean is the average of all y_i's $(i \in \mathcal{U})$:

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{1}{6}(66+59+70+83+82+71) = 71.83$$

the population variance :

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{Y})^2 = \frac{1}{5}\left[(66-71.83)^2+(59-71.83)^2+\cdots+(71-71.83)^2\right] = 86.17 .$$

b) . How many SRS's of size 4 are possible ?

Solution : By definition , a simple random sample of size 4 is subset of $\mathcal{U}$ with 4 elements . Therefore, there are $\binom{6}{4} = \binom{6}{2} = 15$ possible SRS of size 4 .

c) . List all possible SRS's . For each, find the sample mean $\bar{y}$. Calculate the variance of $\bar{y}$.

※ (i) Recall for an SRS of size n , the prob. that u is selected is $\frac{1}{\binom{N}{n}}$.

(ii) The variance of the sample mean $\bar{y}$ under simple random sampling is .

$$Var(\bar{y}) = \left(1-\frac{n}{N}\right)\cdot\frac{S^2}{n} ,$$ where S^2 is population variance .

Solution :

sample	corresponding y_i's.	probability	sample mean $\bar{y} = \frac{1}{n}\sum_{i \in s} y_i$
{1, 2, 3, 4}	{66, 59, 70, 83}	1/15	69.5
{1, 2, 3, 5}	{66, 59, 70, 82}	1/15	69.25
{1, 2, 3, 6}	{66, 59, 70, 71}	1/15	66.5
{1, 2, 4, 5}	{66, 59, 83, 82}	1/15	72.5
{1, 2, 4, 6}	{66, 59, 83, 71}	1/15	69.75
{1, 2, 5, 6}	{66, 59, 82, 71}	1/15	69.5
{1, 3, 4, 5}	{66, 70, 83, 82}	1/15	75.25
{1, 3, 4, 6}	{66, 70, 83, 71}	1/15	72.5
{1, 3, 5, 6}	{66, 70, 82, 71}	1/15	72.25
{1, 4, 5, 6}	{66, 83, 82, 71}	1/15	75.5
{2, 3, 4, 5}	{59, 70, 83, 82}	1/15	73.5
{2, 3, 4, 6}	{59, 70, 83, 71}	1/15	70.75
{2, 3, 5, 6}	{59, 70, 82, 71}	1/15	70.5
{2, 4, 5, 6}	{59, 83, 82, 71}	1/15	73.75
{3, 4, 5, 6}	{70, 83, 82, 71}	1/15	76.5.

The variance of $\bar{y}$ can be calculated in two ways.

Way 1: $\frac{1}{15}(69.5 - 71.83)^2 + \frac{1}{15}(69.25 - 71.83)^2 + \dots + \frac{1}{15}(76.5 - 71.83)^2 = 7.18$

Way 2: $Var(\bar{y}) = (1 - \frac{n}{N}) \cdot \frac{S^2}{n} = (1 - \frac{4}{6}) \times \frac{86.17}{4} = 7.18.$

d). Now let stratum 1 consist of students 1-3, and stratum 2 consist of students 4-6. How many stratified random samples of size 4 are possible in which 2 students are selected from each stratum?

Solution: First select two students from the first stratum — $\binom{3}{2} = 3$ ways.

Then select two from the second stratum — $\binom{3}{2} = 3$ ways.

In total there are $\binom{3}{1} \times \binom{3}{1} = 3 \times 3 = 9$ possible samples.

e) List the possible stratified samples. Which of the samples from (c) cannot occur with the stratified design?

(Generally, Note this is not simply the average of y_i's corresponding to the 3rd column!).

Solution:

Sample from stratum 1	Sample from stratum 2	sample	sample mean $\hat{\bar{Y}}$
{1, 2}	{4, 5}	{1, 2, 4, 5}	72.5
{1, 2}	{4, 6}	{1, 2, 4, 6}	69.75
{1, 2}	{5, 6}	{1, 2, 5, 6}	69.5
{1, 3}	{4, 5}	{1, 3, 4, 5}	75.25
{1, 3}	{4, 6}	{1, 3, 4, 6}	72.5
{1, 3}	{5, 6}	{1, 3, 5, 6}	72.25
{2, 3}	{4, 5}	{2, 3, 4, 5}	73.5
{2, 3}	{4, 6}	{2, 3, 4, 6}	70.75
{2, 3}	{5, 6}	{2, 3, 5, 6}.	70.5

Compared with the table (c), we know that samples {1, 2, 3, 4}, {1, 2, 3, 5}, {1, 2, 3, 6}, {1, 4, 5, 6}, {2, 4, 5, 6}, {3, 4, 5, 6} do not occur here.

f). Find the sample mean $\hat{\bar{Y}}$ based on the stratified sampling for each possible sample listed in e). Find its variance $Var(\hat{\bar{Y}})$ and compare it with $Var(\bar{y})$ calculated in c).

Solution: The sample mean is calculated in this way: First calculate the sample mean for 1st stratum $\bar{y}_1$, and sample mean for 2nd stratum $\bar{y}_2$; then $\hat{\bar{Y}} = \frac{N_1}{N}\bar{y}_1 + \frac{N_2}{N}\bar{y}_2$, where N_h is the number of units in stratum h. The numbers are filled in the last column of the table in (e).

Again $Var(\hat{\bar{Y}})$ can be calculated in two ways:

(i). Use last column of the table in (e), since each sample occurs with probability $\frac{1}{3}\times\frac{1}{3}=\frac{1}{9}$,

$$\mathbb{E}(\hat{\bar{Y}}) = \frac{1}{9}(72.5+69.75+\dots+70.5) = 71.83.$$

$$Var(\hat{\bar{Y}}) = \frac{1}{9}(72.5-71.83)^2+\frac{1}{9}(69.75-71.83)^2+\dots+\frac{1}{9}(70.5-71.83)^2 = 3.14.$$

$< Var(\bar{y})$

(ii) Use formula. $Var(\hat{\bar{Y}}) = \frac{1}{N^2}\sum_{h=1}^{H} N_h^2\left(\frac{1}{n_h}-\frac{1}{N_h}\right)S_h^2 = \frac{9}{36}\left\{\left(\frac{1}{2}-\frac{1}{3}\right)\times\frac{1}{2}\left[(66-65)^2+(59-65)^2+(70-65)^2\right]+\left(\frac{1}{2}-\frac{1}{3}\right)\times\frac{1}{2}\left[(83-78.67)^2+(82-78.67)^2+(71-78.67)^2\right]\right\} = 3.14.$