Practical : 01 Date : 17-06-2025

The height (in cm) of 6 students of M.Sc., majoring in statistics from RU during 2016. The data, so obtained, are given below:

Student	Name	Height (y)
1	Alom	168
2	Asad	175
3	Momin	185
4	Ali	173
5	Ripon	171
6	Kalam	172

(a) Calculate population mean ($\bar{y}$), variance (σ^2) and mean square error (s^2)

(b) Enumerate all possible samples of size two by without replacement method.

(i) Show that sample mean gives an unbiased estimate of the population mean and find its sampling variance.

(ii) Show that sample variance (s^2) is an unbiased estimate of the population variance (S^2).

(iii) Show that $E(v(\bar{y})) = V(\bar{y})$

Solution:

(a) According to our practical problem.

Population mean is defined as $= \bar{Y} = \frac{1}{N} \sum_{i=1}^{N} y_i$

$= \frac{1}{6} \sum_{i=1}^{6} y_i \quad [\because N = 6]$

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

$= \frac{1}{6} \sum_{i=1}^{6} (y_i - \bar{Y})^2$

$= \frac{1}{6} \left[\sum_{i=1}^{6} y_i^2 - \frac{(\sum_{i=1}^{6} y_i)^2}{6} \right]$

Mean square error $= s^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y})^2$

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

By using excel we get,

$$\sum_{i=1}^{6} y_i = 1044, \quad \sum_{i=1}^{6} y_i^2 = 181828$$

Hence, we obtain

$$\bar{Y} = \frac{1}{6} \times 1044 = 174$$

$$\sigma^2 = \frac{1}{6} \left[181828 - \frac{(1044)^2}{6} \right] = 28.667$$

$$s^2 = \frac{6}{6-1} \times 28.667 = 34.4$$

(b) No. of SRSWOR sample of size 2 will be

$$ {}^{N}C_{n} = {}^{6}C_{2} = \frac{6!}{2!(6-2)!} = 15 $$

For a sample of size $n=2$ which is drawn by using SRSWOR from a population of size $N=6$, the sample mean ($\bar{y}$) is defined as

$$ \bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{2}\sum_{i=1}^{2} y_i $$

And sample variance, $s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$

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

$$ = \sum_{i=1}^{2}(y_i - \bar{y})^2 $$

Now, the all possible samples of size 2, and other related statistic is given below (by using excel):

SL	Sample units	Height of the students (y)	$\bar{y}$	s^2
1	(1,2)	(168, 175)	171.5	24.5
2	(1,3)	(168, 165)	176.5	144.5
3	(1,4)	(168, 173)	170.5	12.5
4	(1,5)	(168, 171)	169.5	4.5
5	(1,6)	(168, 172)	170	8
6	(2,3)	(175, 185)	180	50
7	(2,4)	(175, 173)	174	2
8	(2,5)	(175, 171)	173	8
9	(2,6)	(175, 172)	173.5	4.5
10	(3,4)	(185, 173)	179	72
11	(3,5)	(185, 171)	178	98
12	(3,6)	(185, 172)	178.5	84.5

13	(4,5)	(173, 171)	172	2
14	(4,6)	(173, 172)	172.5	0.5
15	(5,6)	(171, 172)	171.5	0.5

(i) We know that sample mean gives an unbiased estimate of the population mean when expected value of sample mean is equal to the population mean.

i.e., we have to show that

$$E(\bar{y}) = \bar{Y}$$

Here, $E(\bar{y}) = \sum_{i=1}^{15} \bar{y}_i \, P(\bar{y}_i)$

$$= \frac{1}{15}(\bar{y}_1 + \bar{y}_2 + \cdots + \bar{y}_{15})$$

$$= \frac{1}{15}(171.5 + 176.5 + \cdots + 171.5)$$

$$= 174 \text{ cm}$$

$$= \bar{Y}$$

(Showed)

Now, sampling variance is given by

$$V(\bar{y}) = \frac{N-n}{Nn} S^2 \qquad [\text{From (i), } S^2 = 34.4]$$

$$= \frac{6-2}{6 \times 2} 34.4$$

$$= 11.4667 \text{ cm}^2$$

(ii) If expected value of sample variance is equal to population variance, then we can say sample variance (s^2) is an unbiased estimate of the population variance (S^2).

Therefore, we have to show that

$$E(s^2) = S^2 \;,\; S^2 = 34.4$$

Here $E(s^2) = \sum_{i=1}^{15} s_i^2 \, P(s_i^2)$

$$= \frac{1}{15}(s_1^2 + s_2^2 + \cdots + s_{15}^2)$$

$$= \frac{1}{15}(24.5 + 144.5 + \cdots + 0.5)$$

$$= 34.4 \text{ cm}^2$$

$$= S^2$$

(showed)

(iii) We have, the unbiased estimator is $\bar{y}$

Variance of the unbiased estimator, $V(\bar{y}) = \frac{N-n}{Nn} S^2$

$$= 11.4667$$

Now, estimate the variance of the unbiased estimator is defined as

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

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

Now, we have to show that,

$$E(v(\bar{y})) = V(\bar{y})$$

Here $E(v(\bar{y})) = E\left[\frac{N-n}{Nn} s^2\right]$

$$= \frac{N-n}{Nn} E(s^2)$$

$$= \frac{6-2}{6\times 2} \times 34.4 \quad [\text{from 'ii', } E(s^2) = 34.4]$$

$$= 11.4667 \text{ cm}^2$$

$$= V(\bar{y}) \text{ (showed)}$$

Practical : 02 Date : 17-06-25

The following data represent the net area sown (in '000 hectares) in different financial year since 1950-51 to 1993-94. The area sown are presented serially.

Sl. No.	Area sown	Sl. No.	Area sown	Sl. No.	Area sown
01	118746	16	136198	31	140022
02	119400	17	137232	32	141928
03	123442	18	138876	33	140220
04	126806	19	137313	34	142841
05	127846	20	138772	35	140892
06	129156	21	140267	36	140901
07	130648	22	138721	37	139578
08	129060	23	137144	38	134085
09	131328	24	142911	39	141891
10	132937	25	141791	40	142339
11	133199	26	141252	41	142999
12	135397	27	139476	42	141632
13	136341	28	141953	43	142645
14	136433	29	142981	44	142095
15	133120	30	138903		

(a) Select a SRSWOR sample of size n = 10 years.
(b) Estimate the average net area sown per year from the selected sample.
(c) Estimate the variance of the estimate of the average net area sown.
(d) Estimate the total area sown during study period.
(e) Estimate the variance of the estimate of total area sown.

(f) Find the 95% confidence interval of the average net area sown.

(g) Comments on your overall findings.

Solution:

(a) We are given a population of size N = 44 years and let y denotes the study variable and $y_1, y_2, \ldots, y_{44}$ be the N = 44 values for 44 units of the population. According to our question a sample of n = 10 years are drawn from the given population i.e., net area sown (in '000 hectares) in different financial years by using SRSWOR method. Again let $y_1, y_2, \ldots, y_{10}$ denote the values of y for n = 10 units selected in the sample. By using excel, we get the following SRSWOR sample of size n = 10 years.

Sl. No.	Area sown (y)	Sl. No.	Area sown (y)
3	123442	5	127845
22	138721	38	134085
25	137791	32	141928
44	142095	20	138772
37	139578	24	142416

(b) For our given problem, The estimate of the average net area sown per year from the selected sample is

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{10}\sum_{i=1}^{10} y_i$$

By using excel, we get, $\sum_{i=1}^{10} y_i = 1366673$

$$\therefore \bar{y} = \frac{1}{10} \times 1366673$$

$$= 136667.3 \quad \text{(in '000 hectares)}$$

(c) Estimate the variance of the estimator ($\bar{y}$) of the average net area sown is given by,

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

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

$$= \frac{1}{9}\sum_{i=1}^{10} (y_i - \bar{y})^2$$

By using excel, we get

$$\sum_{i=1}^{10} (y_i - \bar{y})^2 = 36797457.6$$

$$\therefore s^2 = \frac{1}{9} \times 36797457.6 = 4088606.4 \text{ (in '000 hectare)}^2$$

Therefore, $\hat{V}(\bar{y}) = \frac{44-10}{44\times10} \times 4088606.4$

$$= 3159377.7 \text{ (in '000 hectares)}^2$$

(d) From 'b', $\bar{y} = 136667.3$ (in '000 hectares)

We know, estimate the total area sown is given by

$$\hat{Y} = N\bar{y} = 44 \times 136667.3$$
$$= 6013361 \quad \text{(in '000 hectares)}$$

(e) From 'c', $\hat{V}(\bar{y}) = 3159377.7$ (in '000 hectares)^2

Estimate the variance of the estimate of total area sown is given by

$$\hat{V}(\hat{Y}) = \hat{V}(N\bar{y})$$
$$= N^2 \hat{V}(\bar{y})$$
$$= (44)^2 \times 3159377.7$$
$$= 6116555176 \quad \text{(in '000 hectares)}^2$$

(f) We know, 95% confidence interval for the the population mean is

$$C.I = \bar{y} \pm Z_{\alpha/2}\sqrt{\hat{V}(\bar{y})}$$

$$= 136667.3 \pm 1.96 \times \sqrt{3159377.7}$$

$$= (133183.5, 140151.1)$$
$$= (a, b) \quad \text{(say)}$$

This interval a and b covers the average net area sown with probability 0.95. In other words, we are 95% confident that the population mean (net average sown) will lie in this interval (a, b).

(g) Comments on overall findings:

At first, we select a SRSWOR of size n=11 years from the population i.e., the net area sown (in '000 hectares) in different financial year of size N=44 years with the help of excel. After that, in 'a' and 'b' we estimate the average net area sown per year and estimate the variance of the estimate of the average net area sown. And they are $\bar{y}$ = 136667.3 (in '000 hectares) $\hat{V}(\bar{y})$ = 3153377.7 (in '000 hectares)^2.

On questions 'd' and 'e' we estimate the total area sown and estimate the variance of the estimate of total area sown respectively. They are, $\hat{Y}$ = 6013836 (in '000 hectares) and $\hat{V}(\hat{Y})$ = 6116555196 (in '000 hectares)^2.

At last, we calculate the 95% CI of the net average area sown i.e. (133183.5, 140151.1)

Practical : 03

Date : 19-05-25

All the 80 farms in a population are stratified by farm size. The expenditure on the insecticides used during last year by each farmer is presented in the table below:

Table: Expenditure (in '00 rupees) on insecticides used

Larger farmers		Medium farmers			Small farmers	
75	62	55	33	54	35	42
65	92	45	43	36	28	33
86	50	35	53	44	36	29
57	48	30	37	47	40	25
45	77	42	52	39	25	35
69	60	38	39	41	18	26
48	64	40	46	28	28	30
60	58	36	42	47	32	37
55		48	51	61	13	26
66		46	55	35	19	32
76		40	41	31	31	18
79		38	48	23	38	16

(a) Compute the overall population mean ($\bar{Y}$) and the population mean square (S^2).

(b) Calculate a stratified sample of 24 farmers by using (i) equal allocation (ii) proportional allocation (iii) Neyman allocation. Work out the relative efficiency of stratified sample mean ($\bar{y}_{st}$) based on each of the above mentioned allocation, with respect to the simple random sample mean ($\bar{y}_{SRS}$) for the total sample size. Assume that the sampling is WOR.

Solution:

(a) According to given problem,

population mean $= \bar{Y} = \frac{\sum_{i=1}^{N} y_i}{N} = \frac{\sum_{i=1}^{80} y_i}{80}$

population mean square $= S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{y})^2$

$= \frac{1}{79} \left[\sum_{i=1}^{80} y_i^2 - \frac{\left(\sum_{i=1}^{80} y_i\right)^2}{80} \right]$

By using excel we get

$\sum_{i=1}^{80} y_i = 3503$, $\sum_{i=1}^{80} y_i^2 = 174613$

$\therefore \bar{y} = \frac{1}{80} \times 3503 = 43.7875$ (in '00 rupees)

$S^2 = \frac{1}{79} \left[174613 - \frac{(3503)^2}{80} \right]$

$= 268.6758$ (in '00 rupees)^2

(b) For the given problem, we have

$N = 80$, $n = 24$, $N_1 = 20$, $N_2 = 36$, $N_3 = 24$

Population proportion for the hth stratum

$$W_h = \frac{N_h}{N}$$

$\therefore W_1 = \frac{N_1}{N} = \frac{20}{80} = 0.25$, $W_2 = \frac{N_2}{N} = \frac{36}{80} = 0.45$

$$W_3 = \frac{N_3}{N} = \frac{24}{80} = 0.3$$

population variance for the h-th stratum

$$S_h^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y_{hi} - \bar{Y}_h)^2$$

$$\therefore S_1^2 = \frac{1}{N_1 - 1}\left[\sum_{h=1}^{N_1} y_{1i}^2 - N_1 \bar{Y}_1^2\right]$$

$$S_2^2 = \frac{1}{N_2 - 1}\left[\sum_{h=1}^{N_2} y_{2i}^2 - N_2 \bar{Y}_2^2\right]$$

$$S_3^2 = \frac{1}{N_3 - 1}\left[\sum_{h=1}^{N_3} y_{3i}^2 - N_3 \bar{Y}_3^2\right]$$

By using excel, we get

$S_1^2 = 169.5158$, $S_2^2 = 70.56111$, $S_3^2 = 61.44928$

(i) Equal Allocation :

For equal allocation, $n_{h=} \frac{n_h}{h} = \frac{24}{3} = 8$

$\therefore n_1 = n_2 = n_3 = 8$

Variance of the sample mean in equal allocation

$$V(\bar{y}_{st})_{EQ} = \sum_{i=1}^{3} W_h^2 \left(\frac{N_h - n_h}{N_h n_h}\right) S_h^2$$

$$= W_1^2 \frac{N_1 - n_1}{N_1 n_1} S_1^2 + W_2^2 \frac{N_2 - n_2}{N_2 n_2} S_2^2 + W_3^2 \frac{N_3 - n_3}{N_3 n_3} S_3^2$$

$= 2.644647$ (by using excel)

(in 100 rupees)^2

(ii) Proportional Allocation:

According to proportional allocation

$$n_h \propto N_h$$

$$\Rightarrow n_h = k N_h \quad \text{---(1)}$$

$$\Rightarrow \Sigma n_h = \Sigma k \cdot N_h$$

$$\Rightarrow n = k \Sigma N_h$$

$$\therefore k = \frac{n}{N}$$

putting this value in (1), we get

$$n_h = \frac{n}{N} N_h$$

$$n_1 = \frac{24}{80} N_1 \simeq 6, \quad n_2 = \frac{24}{80} N_2 \approx 11, \quad n_3 = \frac{24}{80} N_3 \approx 7$$

Variance of the sample mean in proportional allo

$$V(\bar{y}_{st})_{prop} = \sum_{h=1}^{3} W_h^2 \left(\frac{N_h - n_h}{N_h n_h}\right) S_h^2$$

$$= 2.69774 \text{ (in '00 rupees)}^2$$

(iii) Neyman allocation:

According to the Neyman allocation

$$n_h \propto S_h N_h$$

$$\Rightarrow n_h = k S_h N_h \quad \text{---(i)}$$

$$\Rightarrow \sum n_h = k \sum S_h N_h$$

$$\Rightarrow n = k \sum N_h S_h$$

$$\therefore k = \frac{n}{\sum N_h S_h}$$

putting this value in (i), we get

$$n_h = \frac{n}{\sum N_h S_h} N_h S_h$$

$$n_1 = \frac{24}{\sum_{h=1}^{3} N_h S_h} N_1 S_1, \quad n_2 = \frac{24}{\sum_{i=1}^{3} N_h S_h} N_2 S_2, \quad n_2 = \frac{24}{\sum_{i=1}^{3} N_h S_h} N_h S_h$$

By using excel,

$$n_1 \approx 8, \quad n_2 \approx 10, \quad n_3 \approx 6$$

variance of the sample mean in Neyman allocation

$$V(\bar{y}_{st})_{Ney} = \sum_{h=1}^{3} W_h^2 \left(\frac{N_h - n_h}{N_h n_h}\right) S_h^2$$

$$= 2.517866 \text{ (in "00 rupees)}^2$$

Percent relative efficiency of the mentioned methods compared to SRS mean.

$$RE_{EQ} = \frac{V(\bar{y})_{SRS}}{V(\bar{y}_{st})_{EQ}} = 2.96 > 1$$

$$RE_{prop} = \frac{V(\bar{y})_{SRS}}{V(\bar{y}_{st})_{prop}} = 2.90 > 1$$

$$RE_{Ney} = \frac{V(\bar{y})_{SRS}}{V(\bar{y}_{st})_{Ney}} = 3.11 > 1$$

The gain in efficiency of the mentioned methods compared to SRS mean

$$G_{Eq} = \frac{V(\bar{y})_{SRS} - V(\bar{y}_{st})_{EQ}}{V(\bar{y}_{st})_{EQ}} \times 100\%$$

$$= \left(\frac{V(\bar{y})_{SRS}}{V(\bar{y})_{EQ}} - 1\right) \times 100\%$$

$$= 196\%$$

$$G_{Prop} = \left(\frac{V(\bar{y})_{SRS}}{V(\bar{y}_{st})_{Prop}} - 1\right) \times 100\%$$

$$= 190\%$$

$$G_{Ney} = \left(\frac{V(\bar{y})_{SRS}}{V(\bar{y}_{st})_{Ney}} - 1\right) \times 100\%$$

$$= 211\%$$

The sample mean under stratified sampling in equal allocation, proportional allocation and Neyman allocation are 2.96, 2.90 and 3.11 times more efficient than SRS mean respectively. So, we can say that Neyman allocation is the most efficient among them. Equal allocation also performs well and even slightly better than proportional allocation for the given problem. And in each case, there is gain compared to SRS mean. The values of the gain are 196%, 190% and 211% for the respective methods.

Practical : 04 Date : 19-06-25

There are 400 villages in a sub-division of a district. Thirty villages out of 400 villages are randomly selected to estimate the total cultivated land for jute in the district. The number of jute grower farmers (x) and amount of land cultivated (y in acres) for jute by these farmers are recorded by an inspection. It is noted that there are 2450 jute grower farmers in the study area. The information are given below:

Sl. No. of villages	y	x	Sl. No. of villages	y	x
1	5.4	3	16	20.0	8
2	10.6	5	17	18.8	6
3	15.2	10	18	14.2	7
4	12.7	8	19	11.3	12
5	8.5	4	20	14.4	8
6	10.0	4	21	20.2	5
7	16.2	8	22	18.5	6
8	15.5	6	23	12.2	4
9	12.2	7	24	15.0	7
10	10.5	3	25	8.2	2
11	6.2	3	26	10.5	5
12	4.4	2	27	12.6	8
13	8.5	5	28	17.2	6
14	20.8	10	29	5.6	2
15	24.0	10	30	8.5	4

(a) Estimate total land area cultivated for jute using ratio and regression methods of estimat

(b) Estimate the variance of your estimators usi above two methods.

(c) Calculate the gain in efficiency of ratio and regression estimators compared to simple estimator ($\bar{y}$).

(d) Comment on your findings.

Solution:

We are given,

$N = 400$, $n = 30$

Auxiliary information $= X$

$$= \sum_{i=1}^{N} x_i$$

$$= 2450$$

$\therefore$ population mean of auxiliary variable, $\bar{x} = \frac{2450}{400}$

$$= 6.125$$

(a) Ratio Method:

The estimate of the mean land area cultivated for jute is

$$\bar{y}_R = \frac{\bar{y}}{\bar{x}} \bar{x}$$

where,

sample mean of land cultivated for jute is

$$\bar{y} = \frac{\sum_{i=1}^{30} y_i}{30} = 12.93 \text{ (acres)}$$

sample mean of jute growers, $\bar{x} = \frac{\sum_{i=1}^{30} x_i}{30} = 5.933$

$$\therefore \bar{y}_R = \frac{12.93}{5.933} \times 5.125$$

$$= 13.34768 \text{ (acres)}$$

The estimate total land area cultivated for jute

$$\hat{Y}_R = N \bar{y}_R = 400 \times 13.34768$$

$$= 5339.073 \text{ (acres)}$$

Regression Method:

Estimate the mean lard area cultivated for jute is

$$\bar{y}_{lr} = \bar{y} + b(\bar{x} - \bar{x})$$

where, $b = \frac{s_{xy}}{s_x^2}$

By using excel, $s_{xy} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{n-1}$

$$= 8.950345$$

$$s_x^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}\right]$$

$$= 6.96092$$

$$\bar{y}_{lr} = \bar{y} + \frac{s_{yx}}{s_x^2}(\bar{X} - \bar{x})$$

$$= 12.93 + \frac{8.950345}{6.96092}(6.125 - 5.03$$

$$= 13.17644 \quad \text{(acres)}$$

∴ the estimate of total land area cultivate for jute

$$\hat{Y}_{lr} = N \bar{y}_{lr}$$

$$= 400 \times 13.17644$$

$$= 5270.5781 \text{ (acres)}$$

(b) Ratio method:

Estimate the variance of the estimator $(\hat{Y}_R)$ i

$$\hat{V}(\hat{Y}_R) = N^2 \hat{V}(\bar{y}_R)$$

$$= N^2 \cdot \frac{N-n}{Nn}\left[s_y^2 + \hat{R}^2 s_x^2 - 2\hat{R} s_{yx}\right]$$

where, $s_y^2 = \frac{1}{n-1}\sum_{i=1}^{30}(y_i - \bar{y})^2$

$$= 25.55114 \quad \text{[by using excel]}$$

$$\hat{R} = \frac{\bar{y}}{\bar{x}}$$

$$\therefore \hat{V}(\hat{Y}_R) = 96688.01 \quad (\text{acres})^2$$

Regression Method:

Estimate the variance of $\hat{Y}_{LR}$ is

$$\hat{V}(\hat{Y}_{LR}) = N^2 \hat{V}(\bar{y}_{LR})$$

$$= N^2 \cdot \frac{N-n}{Nn}(1-\hat{\rho}^2)$$

where, $\hat{\rho} = \frac{s_{yx}}{s_x s_y}$

$$\therefore \hat{V}(\hat{Y}_{LR}) = N \cdot \frac{N-n}{n}(1-\hat{\rho}^2)$$

$$= 2711.338 \text{ (acres)}^2$$

(c) Estimate the variance of simple estimator ($\bar{y}$)

$$\hat{V}(\hat{Y})_{SRS} = N^2 V(\bar{y})_{SRS}$$

$$= N^2 \cdot \frac{N-n}{Nn} s_y^2$$

$$= 12652.28 \text{ (acres)}^2$$

the gain in efficiency of ratio and regression estimators compared to simple estimator are

$$G_R = \left(\frac{\hat{V}(\hat{Y})_{SRS}}{\hat{V}(\hat{Y}_R)} - 1\right) \times 100\%$$

$$= 30.3701\%$$

$$G_{LR} = \left(\frac{\hat{V}(\hat{Y})_{SRS}}{\hat{V}(\hat{Y}_{LR})} - 1\right) \times 100\%$$

$$= 4549.08\%$$

The gain in efficiency of ratio and regression estimators compared to simple simple estimator are 30.3701 % and 4549.08 % respectively. So, we can say, there is gain for each method compared to simple estimator. Between the two methods regression method performs well and it is the most efficient method for the given problem. Also ratio method performs better than simple method.

(d) Comment on overall findings:

At first, we estimate the total land area cultivated for jute using ratio and regression methods and the estimators are 5339.073 and 5270.578 (acres). After that we estimate the variance of the estimators using above two methods. Our finding estimators are 96885.01 and 2711.338 $(acres)^2$. Finally we calculat the gain in efficiency of ratio and regression estimators compared to simple estimator. In each case, we see that there is gain and the methods perform better than simple methods.

Practical: 05 Date: 23.06.25

A pilot sample survey for study of cultivation practices and yeild of guava was conducted by [ASRI] in India. From Umerpur Neerna village, out of a total of 412 bearing trees, 15 clusters of size 4 trees each were selected and yield (in kg) recorded as given below:

Cluster	1st tree	2nd tree	3rd tree	4th tree
1	5.53	4.84	0.69	15.79
2	26.11	10.93	19.08	11.18
3	11.08	0.65	4.21	7.56
4	12.66	32.52	16.92	37.02
5	0.87	3.56	4.81	57.54
6	6.40	11.68	40.05	5.15
7	54.21	34.63	52.55	37.96
8	1.94	35.97	29.54	25.98
9	37.94	47.07	16.94	26.11
10	56.92	17.69	26.24	6.77
11	27.59	38.10	24.76	6.53
12	45.98	5.17	1.17	6.53
13	7.13	34.35	12.18	9.86
14	14.23	16.89	28.93	21.70
15	3.53	40.76	5.15	1.25

(a) Estimate the average yield (in kg) per tree of guava in the Umerpur Neerna village along with standard errors.

(b) Estimate the intra cluster correlation coefficient between trees within clusters and efficiency of cluster sampling as compared to SRS sampling.

(c) Comment on your findings.

Solution:

(a) For our given problem, we have

Total cluster, $N = \frac{412}{4} = 103$, $M = 4$, $n = 15$

Let, y_{ij} be the yield of jth tree in to ith cluster $[i = 1, 2, \ldots, n;\ j = 1, 2, 3, 4)$.

An estimate of the average yield is given by

$$\bar{y}_{c1} = \frac{1}{nM} \sum_{i=1}^{n} \sum_{j=1}^{M} y_{ij}$$

$$= \frac{1}{n} \sum_{i=1}^{n} \bar{y}_i \quad ; \quad \bar{y}_i = \frac{1}{M} \sum_{i=1}^{4} y_{ij}$$

$$= \frac{1}{15} \sum_{i=1}^{15} \bar{y}_i$$

$$= 20.151333 \text{ (kg)} \quad [\text{by using excel}]$$

Estimate the variance of the estimator is

$$\hat{V}(\bar{y}_{c1}) = \frac{N-n}{Nn} s_b^2$$

where, $s_b^2 = \frac{1}{n-1} \sum_{i=1}^{n} (\bar{y}_i - \bar{y}_{c1})^2$

$$= 99.02632 \quad (\text{kg})^2$$

$\therefore \hat{V}(\bar{y}_{c1}) = 5.640334$

$\therefore$ Estimate the standard error of the estimator

$$SE(\bar{y}_{c1}) = \sqrt{\hat{V}(\bar{y}_{c1})}$$

$$= 2.374939$$

(b) Estimate the efficiency of cluster sampling as compared to SRS sampling is given by

$$\hat{E} = \frac{\hat{S}^2}{M \hat{S}_b^2}$$

where, $S_b^2 = \hat{S}_b^2 = 99.02632 \ (kg)^2$

The estimate of S^2 is given by

$$\hat{S}^2 = \frac{1}{MN-1}\left[N(M-1)S_w^2 + (N-1)M S_b^2\right]$$

where, $S_w^2 = \frac{1}{n(M-1)} \sum_{i=1}^{n} \sum_{j=1}^{M} (y_{ij} - \bar{y}_i)^2$

$$= \frac{1}{n(M-1)} \sum_{i=1}^{n} \left[\sum_{j=1}^{M} y_{ij}^2 - M \bar{y}_i^2\right]$$

$$= 225.4269$$

$$\therefore \hat{S}^2 = 267.785$$

Therefore, $\hat{E} = \frac{\hat{S}^2}{M \hat{S}_b^2} = 0.676015$

Estimate the intra cluster correlation coefficient is given by

$$\hat{\rho} = \frac{1 - \hat{E}}{(M-1)\hat{E}}$$

$$= 0.15973$$

The estimate of the intra cluster correlation coefficient is 0.15973, it suggests that trees within the cluster are not very

similar, so there's a reasonable amount of within-cluster variation. The estimate of the efficiency of cluster sampling compared to SRS is 0.676045. This means cluster sampling is 67.6% more efficient than SRS in this situation.

(c) Comments on overall findings:

In question 'a', we estimate the average yield (in kg) per tree of guava i.e. 20.15133 (kg) along with its standard error is 2.374939.

In question 'b', we estimate the intra-cluster correlation coefficient between trees within clusters and efficiency of cluster sampling as compared to SRS sampling. They are 0.15973 and 0.676045.

Practical : 06 Date : 23-06-25

A survey on pepper was conducted to estimate the number of pepper standards and production. For this, 3 clusters from 95 clusters were selected by SRSWOR method. The information on the number of pepper standards recorded is given below.

Cluster No.	Cluster size	No. of pepper standards
1	12	41, 16, 19, 15, 144, 454, 212, 57, 28, 76, 119, 110
2	10	39, 70, 38, 37, 161, 38, 27, 219, 36, 128
3	07	252, 386, 92, 293, 115, 59, 120

(a) Estimate total number of pepper standards along with its standard error, given $\bar{M}$ the average cluster size for the population to be 10.

(b) Find 95 % confidence interval of population total.

(c) Comments on your results.

Solution:

(a) We have

$N = 95, \quad n = 3$

$M_1 = 12, \quad M_2 = 10, \quad M_3 = 7, \quad \bar{M} = 10$

An unbiased estimate of the $\bar{Y}$ is given by

$$\bar{y}_{cl}^{*} = \frac{\sum_{i=1}^{3} M_i \bar{y}_i}{n\bar{M}}$$

$$= \frac{1}{n\bar{M}} [M_1\bar{y}_1 + M_2\bar{y}_2 + M_3\bar{y}_3]$$

$$= 133.3667$$

Estimate of Y is $\hat{Y}_{cl}^{*} = M_0 \bar{y}_{cl}^{*}$

where, $\bar{M} = \frac{\sum_{i=1}^{3} M_i}{N} = \frac{M_0}{N}$

$\Rightarrow M_0 = N\bar{M}$

$\therefore \hat{Y}_{cl}^{*} = 950 \times 133.3667$

$= 107698.3$

The estimate of the variance of the total number of pepper standards is

$$\hat{V}(\hat{Y}_{cl}^{*}) = M_0^2 \frac{N-n}{Nn} s_b^{*2}$$

where, $s_b^{*2} = \frac{1}{n-1} \sum_{i=1}^{3} \left(\frac{M_i \bar{y}_i}{\bar{M}} - \bar{y}_{cl}^{*} \right)^2$

$= 872.0933$

Therefore $\hat{V}(\hat{Y}_{cl}^{*}) = 254069858$

$\therefore SE(\hat{Y}_{cl}^{*}) = 15939.569$

(b) The 95 % confidence interval of population total is given

$$C.I = \hat{Y}_{ct}^{*} \pm Z_{\alpha/2} \times SE(\hat{Y}_{ct}^{*})$$

$$= 107698.3 \pm 1.96 \times 15939.569$$

$$= (76456.78, 138939.9)$$

$$= (a, b) \quad [say]$$

This interval a and b covers the population total with probability 0.95. In other words we are 95 % confident that the population total (total number of pepper standards) will lie in this interval (a, b).

(c) Comments on overall findings:

Firstly, we estimate the total number of pepper standards and that is 107698.3 along with its standard error 15939.569. Finally we comput 95 % confidence interval of population total. The result is (76456.78, 138939.9).