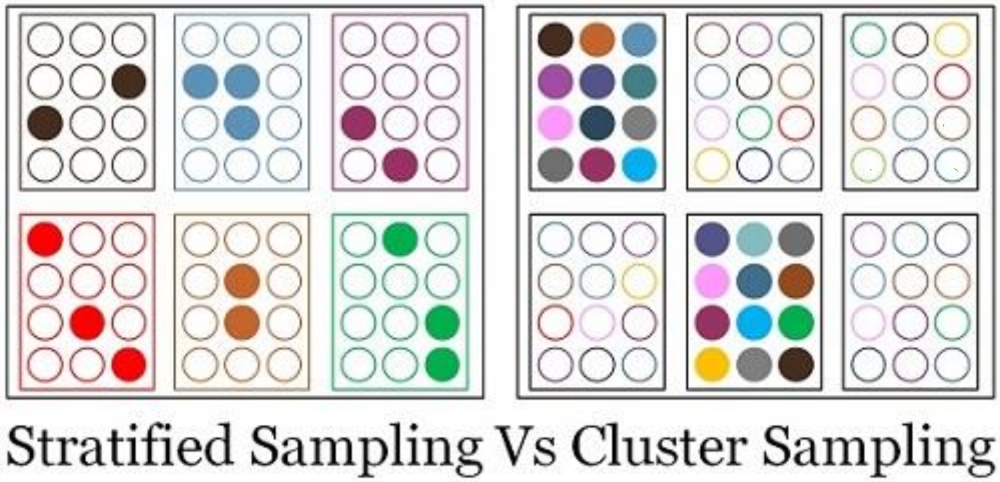
Exercises 4



Factors for Comparison	**Cluster Sampling**	**Stratified Sampling**
Definition	Members of this sample are chosen from naturally divided groups called clusters, by randomly selecting elements to be a part of the sample.	Members of this sample are randomly chosen from non-overlapping, homogeneous strata.
Purpose	Cost reduction and increased efficiency.	Enhanced precision and population depiction.
Sample selection	Selection of the sample is done by randomly selected clusters and including **all** the members from these clusters.	Selection of the sample is done by randomly selecting members from various formed strata.
Selection of elements that form a Sample	Conjointly	Distinctively
Division type	Naturally formed	Depends on the researcher
Heterogeneity	Internally, with the clusters	Externally, between various strata
Homogeneity	Externally, between various clusters	Internally, with the strata

#Cluster Sampling

Estimation of clusters Mean using simple random sampling **WOR**:

Unbiased estimator of population mean when M=$\sum_{i=1}^{N} M_i$ is **known**:

$$\bar{y}_{clu} = \frac{N}{n}\frac{\sum_{i=1}^{n} t_i}{M} = \frac{1}{\bar{M}n}\sum_{i=1}^{n} t_i$$

Variance of estimator $\bar{y}_{clu}$:

$$Var(\bar{y}_{clu}) = \frac{N(N-n)}{n\,M^2}\frac{1}{N-1}\sum_{i=1}^{N}(T_i - \frac{T}{N})^2 \qquad (10.2)$$

Estimator of Variance $Var(\bar{y}_{clu})$:

$$var(\bar{y}_{clu}) = \frac{N(N-n)}{n\,M^2}\frac{1}{n-1}\sum_{i=1}^{n}(t_i - \frac{t}{n})^2 \qquad (10.3)$$

Other equivalent formula

$$var(\bar{y}_{clu}) = \frac{(N-n)}{Nn\,\bar{M}^2}\frac{1}{n-1}\sum_{i=1}^{n}(t_i - \bar{M}\,\bar{y}_{clu})^2 \qquad (10.3)$$

Where,

N = number of clusters in the population

n = number of clusters in the sample

M_i = number of units in the i-th cluster of the population

M=$\sum_{i=1}^{N} M_i$ = total number of units in the population

Y_{ij}= value of the character under study for the j-th unit in the i-th cluster,
j = 1, 2, ... , M_i ; i = 1, 2, ... , N

$t_i = \sum_{j=1}^{M_i} y_{ij} =$ ith sample cluster total

$$T_i = \sum_{j=1}^{M_i} Y_{ij} = \text{i}^{th} \text{ cluster tota}$$

$$t = \sum_{i=1}^{n} t_i = \text{total of y (values for units in the Sample)}$$

$$T = \sum_{i=1}^{N} T_i = \text{ total of y (values for all units in the population)}$$

$$\bar{M} = \frac{\sum_{i=1}^{N} M_i}{N} = \text{ average number of units per cluster in the population}$$

If the clusters are selected using WR sampling, then fpc = (N-n)/(N-1) in relation (10.2) and the sampling fraction f = n/N in (10.3) are taken as **1** and **0** respectively to get the corresponding results for the with replacement case.

Note: *(fpc) Finite population correction.*

When $M = \sum_{i=1}^{N} M_i$ is not known and the values of M_i are known only for the sample clusters, then $\bar{Y}$ can be estimated by using the following estimator .

Estimator of population mean when $M = \sum_{i=1}^{N} M_i$ **unknown**:

$$\bar{y}_{clu} = \frac{1}{n}\sum_{i=1}^{n} \bar{y}_i$$

Variance of estimator $\bar{y}_{clu}$:

$$Var(\bar{y}_{clu}) = \frac{N-n}{nN}\frac{1}{N-1}\sum_{i=1}^{N}(\bar{Y}_i - \bar{y}_{clu})^2$$

Estimator of variance $Var(\bar{y}_{clu})$:

$$var(\bar{y}_{clu}) = \frac{N-n}{nN}\frac{1}{n-1}\sum_{i=1}^{n}(\bar{y}_i - \bar{y}_{clu})^2$$

Where,

$\bar{y}_i = \frac{t_i}{M_i} =$ per unit ith sample cluster mean

$\bar{Y}_i = \frac{T_i}{M_i}$ per unit ith cluster mean

Example 1 :

A random sample of 8 clusters is drawn from a population of 200 clusters. Each of the cluster has 10 units. The means of sample clusters are respectively 32, 25, 41,42, 36, 47, 39 and 43. Estimate population mean with a 95% confidence interval.

Solution: we have n= 8 , N=200 and M= 10. The estimator of population mean in cluster sampling its variance estimator are

$$\bar{y}_{clu} = \frac{1}{n}\sum_{i=1}^{n}\bar{y}_i$$

$$var(\bar{y}_{clu}) = \frac{N-n}{nN}\frac{1}{n-1}\sum_{i=1}^{n}(\bar{y}_i - \bar{y}_{clu})^2$$

Further, a 100 (1 -α)% confidence interval for true population mean is

$$\bar{y}_{clu} \pm Z_{1-\frac{\alpha}{2}}\sqrt{var(\bar{y}_{clu})}$$

Now estimate of population mean is:

$$\bar{y}_{clu} = \left(\frac{1}{8}\right)(32+25+41+42+36+47+39+43) = 38.125$$

$$var(\bar{y}_{clu}) = \frac{200-8}{8*200}\frac{1}{8-1}[(32-38.125)^2 + (25-38.125)^2 + \cdots + (43-38.125)^2\,]$$

$$var(\bar{y}_{clu}) = \ 5.844$$

For a 95% confidence interval we have $Z_{0.975} = \ 1.96$ The interval for population mean is therefore

$38.125 \pm 1.96 * \sqrt{5.844}$

38.125 ± 4.7382

[33.3868 ,42.8632]

Example 2:

The recommended dose of nitrogen for wheat crop is 120 kg per hectare. A survey project was undertaken by the Department of Agriculture with a view to estimate the amount of nitrogen actually applied by the farmers. For this purpose, 12 villages from a population of 170 villages of a development block were selected using equal probabilities WOR sampling, and the information regarding the nitrogen use was collected from all the farmers in the selected villages. The data collected are presented in table 10.1. The total number of farmers in these 170 villages is available from the patwari's record as 2890.

Estimate the average amount of nitrogen used in practice by a farmer. Also, obtain standard error of the estimate, and place 95% confidence limits on the population mean.

Table 10.1 Per hectare nitrogen (in kg) applied to wheat crop by farmers

Village	M_i	Nitrogen applied (in kg) by a farmer									t_i
1	15	105	128	130	108	135	122	120	138	126	1843
		117	125	126	123	118	122				
2	18	135	128	105	130	120	125	114	128	121	2206
		109	128	122	129	112	133	117	119	131	
3	25	124	118	128	106	132	121	126	108	136	3085
		121	128	125	136	128	121	127	122	113	
		117	132	128	125	130	109	124			
4	21	108	116	111	129	119	137	129	121	118	2582
		126	131	128	134	125	112	121	116	114	
		129	127	131							
5	11	114	105	126	132	116	125	104	121	132	1292
		106	111								
6	13	128	116	132	136	121	122	129	123	127	1627
		118	134	126	115						
7	22	103	118	107	128	132	136	124	129	130	2686
		134	108	106	117	129	113	118	126	127	
		129	119	125	128						
8	12	109	121	114	128	133	135	114	128	107	1471
		125	126	131							
9	10	119	128	117	131	105	128	136	113	127	1234
		130									
10	20	130	127	116	128	114	120	127	123	134	2449
		122	126	121	117	125	129	122	113	111	
		126	118								
11	10	126	117	124	121	131	133	126	120	128	1242
		116									
12	16	124	121	127	119	120	123	128	117	121	1935
		93	115	120	124	121	130	132			

Solution:

$$N = 170, M = 2890 \text{ , and } n = 12$$

$$\bar{M} = \frac{M}{N} = \frac{2890}{170} = 17$$

Estimate of the average amount of nitrogen used per hectare, by a farmer:

$$\bar{y}_{clu} = \frac{N}{n}\frac{\sum_{i=1}^{n} t_i}{M} = \frac{1}{\bar{M}n}\sum_{i=1}^{n} t_i$$

$$= \frac{1}{17 * 12}(1843 + \cdots + 1935) = \frac{23652}{17 * 12} = 115.941$$

Estimate of variance:

$$var(\bar{y}_{clu}) = \frac{(N-n)}{Nn\,\bar{M}^2}\frac{1}{n-1}\sum_{i=1}^{n}(t_i - \bar{M}\,\bar{y}_{clu})^2$$

$$\bar{M}\,\bar{y}_{clu} = (17)(115.941) = 1970.997$$

$$= \left(\frac{170-12}{170 * 12 * (17)^2}\right)\frac{1}{11}\,[(\mathbf{1843} - 1970.997)^2 + (\mathbf{2206} - 1970.997)^2$$

$$+ \cdots + (\mathbf{1935} - 1970.997)^2]$$

$$= \frac{(170-12)(4331060)}{(170)(12)(17)^2(11)} = 105.519$$

Using above calculated estimate of variance, the standard error of mean will be

$$sd(\bar{y}_{clu}) = \sqrt{var(\bar{y}_{clu})} = \sqrt{105.519} = 10.272$$

The required confidence limits for population mean are obtained as

$$\bar{y}_{clu} \pm Z_{1-\frac{\alpha}{2}}\sqrt{var(\bar{y}_{clu})}$$

$$115.941 \pm 1.96(10.272)$$

[95.8079, 136.0741]

H.W

Example 3:

The Department of Education of a state has been providing fixed medical allowance at the rate of $ 60 per head, for a quarter, to its teachers and their dependents for the last five years. With a view to examine the rationality of this policy today, when the price index has gone up about 1.5 times during the preceding five years, a simple random WOR sample of 10 schools was drawn from a total of 104 schools in a development block by the investigator. Since some of the teachers might be on long leave, the total number of teachers available could not be known in advance. All the teachers (M), except those on long leave, in the sample schools were interviewed. They were requested to give per head medical expenses (in rupees), for themselves and their dependents, during the past 3 months. The results are as follows :

School	M_i	Per head medical expenses for 3 months						
1	4	50	100	120	110			
2	4	90	70	40	140			
3	5	85	33	122	60	105		
4	6	55	80	130	70	240	80	
5	4	130	70	40	120			
6	3	85	65	45				
7	4	30	75	65	115			
8	6	150	105	0	25	185	100	
9	5	100	60	130	40	125		
10	7	50	45	110	120	60	140	95

1. Estimate the average per head money spent as medical expenses during the past 3 months.
2. Build up the 95% confidence interval for the population average.
3. Use R Program to solve previous part .