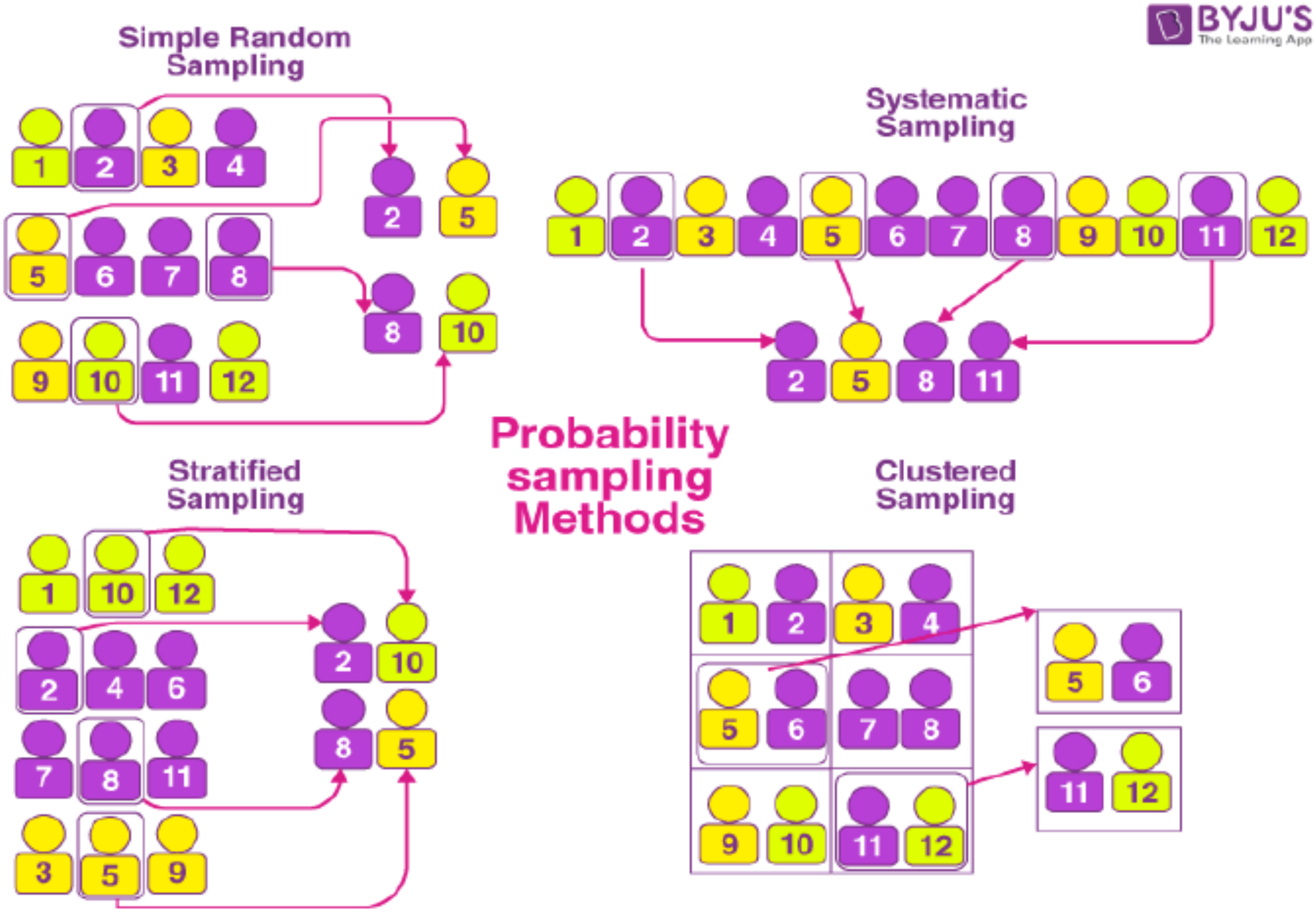
Definition 10.1 The *cluster sampling* consists of forming suitable clusters of contiguous population units, and surveying all the units in a sample of clusters selected according to an appropriate sampling scheme.



BYJU'S
The Learning App
Simple Random Sampling
1 2 3 4
5 6 7 8
9 10 11 12
2 5
8 10
Systematic Sampling
1 2 3 4 5 6 7 8 9 10 11 12
2 5 8 11
Probability sampling Methods
Stratified Sampling
1 10 12
2 4 6
7 8 11
3 5 9
2 10
8 5
Clustered Sampling
1 2 3 4
5 6 7 8
9 10 11 12
5 6
11 12

10.2 NOTATIONS

In order to facilitate the understanding of the text, we first acquaint the reader with the notations to be used in the chapter. Let

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

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

$\overline{M} = M_o/N$ = average number of units per cluster in the population

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

$Y_{i.} = \sum_{j=1}^{M_i} Y_{ij}$ = i - th cluster total

$$\overline{Y}_i = \frac{1}{M_i} \sum_{j=1}^{M_i} Y_{ij} = \text{per unit i - th cluster mean}$$

$$y_{i.} = \sum_{j=1}^{M_i} y_{ij} = \text{i - th sample cluster total}$$

$$\overline{y}_i = \frac{1}{M_i} \sum_{j=1}^{M_i} y_{ij} = \text{per unit i - th sample cluster mean}$$

$$\overline{y}_c = \frac{1}{n} \sum_{i=1}^{n} y_{i.} = \text{mean per cluster in the sample}$$

$$\overline{Y}_N = \frac{1}{N} \sum_{i=1}^{N} \overline{Y}_i = \text{mean of cluster means in the population}$$

$$\overline{Y} = \frac{1}{M_o} \sum_{i=1}^{N} \sum_{j=1}^{M_i} Y_{ij} = \text{mean per unit of the population}$$

$$\overline{Y}_c = Y_{..}/N = \text{population mean per cluster}$$

Unbiased estimator of population mean when M_o is known :

$$\left.\begin{aligned}\bar{y}_{c1} &= \frac{N}{nM_o}\sum_{i=1}^{n} M_i\,\bar{y}_i \\ &= \frac{1}{\overline{M}n}\sum_{i=1}^{n} y_{i.}\end{aligned}\right] \qquad (10.1)$$

Variance of estimator $\bar{y}_{c1}$:

$$V(\bar{y}_{c1}) = \left(\frac{N-n}{Nn\overline{M}^2}\right)\frac{1}{N-1}\sum_{i=1}^{N}(Y_{i.} - \overline{Y}_c)^2 \qquad (10.2)$$

Estimator of variance $V(\bar{y}_{c1})$:

$$\left.\begin{aligned}v(\bar{y}_{c1}) &= \left(\frac{N-n}{Nn\overline{M}^2}\right)\frac{1}{n-1}\sum_{i=1}^{n}(y_{i.} - \overline{M}\,\bar{y}_{c1})^2 \\ &= \left(\frac{N-n}{Nn\overline{M}^2}\right)\frac{1}{n-1}\left[\sum_{i=1}^{n} y_{i.}^2 - n(\overline{M}\,\bar{y}_{c1})^2\right]\end{aligned}\right] \qquad (10.3)$$

Example 10.1

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 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									$y_{i.}$
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						

Table 10.1 continued...

Village	M_i	Nitrogen applied (in kg) by a farmer									$y_{i.}$
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

Here we have $N = 170$, $M_o = 2890$, and $n = 12$. It gives

$$\overline{M} = \frac{M_o}{N} = \frac{2890}{170} = 17$$

As the sample cluster totals $y_{i.}$ will be required for computing the estimates, these are worked out below and are presented in the last column of the table 10.1 above.

$$\begin{array}{lcl}
\text{Cluster 1} & : & y_{1.} = 105 + 128 + \ldots + 122 = 1843 \\
\text{Cluster 2} & : & y_{2.} = 135 + 128 + \ldots + 131 = 2206 \\
\cdot & & \cdot \\
\cdot & & \cdot \\
\cdot & & \cdot \\
\text{Cluster 12} & : & y_{12.} = 124 + 121 + \ldots + 132 = 1935
\end{array}$$

Estimate of the average amount of nitrogen used per hectare, by a farmer, follows from (10.1) as

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

$$= \frac{1}{(17)(12)} (1843 + 2206 + ... + 1935)$$

$$= \frac{23652}{(17)(12)}$$

$$= 115.941$$

We then work out the estimate of variance using (10.3). Thus,

$$v(\bar{y}_{c1}) = \left(\frac{N-n}{Nn\bar{M}^2}\right) \frac{1}{n-1} \sum_{i=1}^{n} (y_{i.} - \bar{M}\,\bar{y}_{c1})^2$$

where

$$\bar{M}\,\bar{y}_{c1} = (17)(115.941) = 1970.997$$

Hence,

$$v(\bar{y}_{c1}) = \left(\frac{170-12}{(170)(12)(17)^2}\right)\frac{1}{11}[(1843-1970.997)^2 + (2206-1970.997)^2$$

$$+ \ldots + (1935-1970.997)^2]$$

$$= \left(\frac{170-12}{(170)(12)(17)^2}\right)\frac{1}{11}[(1843)^2 + (2206)^2 + \ldots + (1935)^2$$

$$- 12(1970.997)^2]$$

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

$$= 105.519$$

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

$$se(\bar{y}_{c1}) = \sqrt{105.519}$$

$$= 10.272$$

Following (2.8), the required confidence limits for population mean are obtained as

$$\bar{y}_{c1} \pm 2\sqrt{v(\bar{y}_{c1})}$$

$$= 115.941 \pm 20.544$$

$$= 95.397,\ 136.485$$

The above confidence limits reasonably ensure that per hectare average dose of nitrogen used by a farmer in the target population is likely to be within the range 95.397 to 136.485 kg. ■

10.4 ESTIMATION OF TOTAL USING SIMPLE RANDOM SAMPLING

An estimator of population total can be easily obtained by multiplying any one of the corresponding estimators of mean given in (10.1), (10.4), and (10.8) by M_o.

Estimators of population total Y:

$$\hat{Y}_{c1} = \frac{N}{n} \sum_{i=1}^{n} y_{i.} \quad (10.12)$$

Expressions for variances and their estimators for the above estimators of population total, can be easily obtained by multiplying their counterparts for mean by M_o^2.

```
#example 10.1 by R

> # N = number of clusters in Population
>   # M = number of elements in the Population
>   # m = vector of the cluster sizes in the Sample
>   # y = either a vector of totals per cluster, or a list of the observations per cluster (this is set by total)
> # n= number of sampled clusters
>
>  # If M is unknown, else M is known
> y <- c(1843,2206,3085,2582,1292,1627,2686,1471,1234,2449,1242,1935)
> M.vec <-c(15,18,25,21,11,13,22,12,10,20,10,16)
> Mo=2890
> N= 170
> n=length(y)
>
>     ybar.c = (1/(n*(Mo/N)))*sum(y)
>     ybar.c
[1] 115.9412
```

```
>    s2.c= sum((y-mean(y))^2 )/(n-1)
>    var.c = ((N-n)/(N*n*(Mo/N)^2))*s2.c
>    var.c
[1] 105.5155
>
ME = 2*sqrt(var.c) #margin of error
>   lower.CI=  ybar.c-ME
>   upper.CI=  ybar.c+ME
>   lower.CI;  upper.CI
[1] 95.39703
[1] 136.4853

>   ytotal.c= Mo*ybar.c
>   ytotal.c
[1] 335070
>
>   var_ytotal.c=  Mo^2 *var.c
>   var_ytotal.c
[1] 881276193
```