**Example 6.4 (for $N \neq nk$)**

On a particular day, 162 boats had gone to sea from the coast for fishing. It was desired to estimate the total catch of fish at the end of the day. As it was not possible to weigh the catch for all the 162 boats, it was decided to weigh fish for only 15 boats selected using circular systematic sampling. Discuss the selection procedure, and obtain the estimate of total catch of fish using data on the 15 sample boats given in table 6.4.

**Table 6.4** Catch of fish (in quintals) for 15 selected boats

| Serial No. of boat | Catch of fish | Serial No. of boat | Catch of fish | Serial No. of boat | Catch of fish |
|---|---|---|---|---|---|
| 73 | 5.614 | 128 | 9.225 | 21 | 8.460 |
| 84 | 8.202 | 139 | 6.640 | 32 | 10.850 |
| 95 | 6.115 | 150 | 7.350 | 43 | 6.970 |
| 106 | 9.765 | 161 | 5.843 | 54 | 5.524 |
| 117 | 8.550 | 10 | 6.875 | 65 | 7.847 |

**Solution**

In this case, we have N=162 and n=15. Since N/n=162/15=10.8 is not a whole number, the value of sampling interval k is taken as 11, an integer nearest to 10.8, and circular systematic sampling is used for selection of boats. If the selected random number r, $1 \leq r \leq 162$, is 73, then the boats bearing serial numbers 73, 84,..., 65 will be included in the sample. The serial numbers of selected boats, along with the corresponding catch of fish, are presented in table 6.4. We now proceed to estimate the total catch of fish using (6.1). This estimate is

$$\hat{Y}_{sy} = N\bar{y}_{sy} = \frac{N}{n} \sum_{i=1}^{n} y_i$$

$$= \frac{162}{15} (5.614 + 8.202 + \ldots + 7.847)$$

$$= \frac{(162)(113.83)}{15}$$

$$= 1229.364$$

The estimate of variance $V(\hat{Y}_{sy})$ is then computed by using the expression (6.4). Thus,

$$v(\hat{Y}_{sy}) = N^2 v(\bar{y}_{sy}) = \frac{N(N-n)}{2n(n-1)} \sum_{i=1}^{n-1} (y_{i+1} - y_i)^2$$

$$= \frac{162\,(162-15)}{2(15)\,(14)} [(8.202-5.614)^2 + (6.115-8.202)^2 + \ldots + (7.847-5.524)^2]$$

$$= \frac{(162)\,(162-15)\,(67.596)}{2(15)\,(14)}$$

$$= 3832.693$$

The confidence interval, for the total catch of fish for 162 boats, can then be calculated from

$$\hat{Y}_{sy} \pm 2\sqrt{v(\hat{Y}_{sy})}$$

$$= 1229.364 \pm 2\sqrt{3832.693}$$

$$= 1105.547,\ 1353.181$$

Thus, the estimate of total catch of fish obtained from a single sample is 1229.364 quintals. The confidence limits, obtained above, indicate that the total catch from all the 162 boats is likely to fall in the interval [1105.547, 1353.181] quintals. ■

# Sample Size Determination:

**Sample size for estimating mean/total with a permissible error B :**

$$n = \frac{Ns_1^2}{ND + s_1^2} \tag{6.11}$$

where

$$D = \frac{B^2}{4} \qquad \text{(when estimating mean)}$$

$$D = \frac{B^2}{4N^2} \qquad \text{(when estimating total)}$$

with $s_1^2$ defined in (6.10). If $n_1 \geq n$, the sample size $n_1$ is sufficient, otherwise, $(n-n_1)$ additional units need to be selected.

**Example 6.6**

Assuming the population in example 6.3 to be in random order, and treating the sample of 36 trees selected there as the preliminary sample, determine the sample size required to estimate total timber volume with a tolerable error of 400 cubic meters.

**Table 6.3** Timber volume (in cubic meters) for 36 selected trees

| Serial No. of tree | Timber volume | Serial No. of tree | Timber volume | Serial No. of tree | Timber volume |
|---|---|---|---|---|---|
| 28 | 1.72 | 1228 | 2.17 | 2428 | 1.89 |
| 128 | 1.29 | 1328 | 1.63 | 2528 | 1.63 |
| 228 | 1.08 | 1428 | 1.91 | 2628 | 2.23 |
| 328 | 2.29 | 1528 | 1.66 | 2728 | 2.40 |
| 428 | 2.01 | 1628 | 1.56 | 2828 | 2.51 |
| 528 | 1.77 | 1728 | 2.26 | 2928 | 2.57 |

**Table 6.3** continued ...

| Serial No. of tree | Timber volume | Serial No. of tree | Timber volume | Serial No. of tree | Timber volume |
|---|---|---|---|---|---|
| 628 | 1.63 | 1828 | 2.49 | 3028 | 1.26 |
| 728 | 1.20 | 1928 | 2.26 | 3128 | 1.46 |
| 828 | 2.03 | 2028 | 2.31 | 3228 | 1.00 |
| 928 | 1.17 | 2128 | 1.60 | 3328 | 1.94 |
| 1028 | 2.47 | 2228 | 1.64 | 3428 | 1.80 |
| 1128 | 1.86 | 2328 | 1.43 | 3528 | 1.60 |

**Solution**

Here, we are given B=400 cubic meters. From example 6.3, we have N=3600, $n_1$=36, and

$$\bar{y}_{sy1} = \frac{1}{36}(1.72 + 1.29 + ... + 1.60)$$
$$= 1.826$$

so that,

$$s_1^2 = \frac{1}{n_1 - 1}\sum_{i=1}^{n_1}(y_i - \bar{y}_{sy1})^2$$

$$= \frac{1}{n_1 - 1}(\sum_{i=1}^{n_1} y_i^2 - n_1\bar{y}_{sy1}^2)$$

$$= \frac{1}{36-1}[(1.72)^2 + (1.29)^2 + ... + (1.60)^2 - 36\,(1.826)^2]$$

$$= .1919$$

Also,

$$ND = \frac{B^2}{4N}$$

$$= \frac{(400)^2}{4(3600)} = 11.1111$$

Then, from (6.11), we can determine the required sample size as

$$n = \frac{Ns_1^2}{ND + s_1^2}$$

$$= \frac{(3600)\,(.1919)}{11.1111 + (.1919)}$$

$$= 61.12$$

$$\approx 61$$

Since the sample size required to estimate the total timber volume with a permissible error of 400 cubic meters is 61, the investigator will, therefore, need to select 61-36=25 more trees to get the estimate with specified magnitude of tolerable error. ■

# Estimating the Proportion:

**Estimator of proportion P :**

$$p_{sy} = \frac{n_1}{n} \tag{6.12}$$

**Variance of estimator $p_{sy}$ :**

$$V(p_{sy}) = \frac{1}{k} \sum_{r=1}^{k} (p_{sy} - P)_r^2 \quad \text{(for LS sampling)} \tag{6.13}$$

$$= \frac{1}{N} \sum_{r=1}^{N} (p_{sy} - P)_r^2 \quad \text{(for CS sampling)} \tag{6.14}$$

# Estimating the Proportion:

**Estimator of variance $V(p_{sy})$ :**

$$v(p_{sy}) = \frac{(N-n)\,R}{2Nn\,(n-1)} \tag{6.15}$$

where R is the total number of times that 0 follows 1 or 1 follows 0 in the ordered sequence of observations for the n sample units.

**Estimator of $V(p_{sy})$ for population in random order :**

$$v(p_{sy}) = \frac{(N-n)}{N}\left[\frac{p_{sy}(1-p_{sy})}{(n-1)}\right] \tag{6.16}$$

**Example 6.7**

On public complaint that some gas cylinders supplied for domestic use were underweight, an inquiry committee was set up. The committee decided to examine 1-in-50 cylinders from the 8000 cylinders stored in a warehouse, arranged in rows by the gas company. The committee found 18 cylinders to be underweight from the 160 sampled cylinders. Estimate the total number $N_1$, and also the proportion, of underweight cylinders in the warehouse. Also, build up confidence interval for these parameters.

**Solution**

We have $N = 8000$, $n=160$, and $n_1=18$. Then the estimate of proportion of underweight cylinders in the warehouse is

$$p_{sy} = \frac{n_1}{n}$$

$$= \frac{18}{160}$$

$$= .1125$$

Estimated total number of underweight cylinders would then be

$$\hat{N}_1 = Np_{sy}$$

$$= (8000)(.1125)$$

$$= 900$$

Assuming that the population units (gas cylinders) under study were placed in random order before drawing the sample, the estimate of variance $V(\hat{N}_1)$ is computed using (6.16) as

$$v(\hat{N}_1) = N^2 v(p_{sy})$$

$$= \frac{N(N-n)\,p_{sy}(1-p_{sy})}{(n-1)}$$

$$= \frac{8000\,(8000-160)\,(.1125)\,(1-.1125)}{159}$$

$$= 39384.90$$

Also, one can work out the confidence interval for $N_1$ from

$$\hat{N}_1 \pm 2\sqrt{v(\hat{N}_1)}$$

$$= 900 \pm 2\sqrt{39384.90}$$

$$= 503.09,\ 1296.91$$

$$\approx 503,\ 1297$$

Thus, the inquiry committee estimated 900 underweight cylinders from this particular sample. The committee also feels that the total number of underweight cylinders in the warehouse is likely to be between 503 and 1297.

From the above calculations, we find that $p_{sy} = .1125$ is the estimate of the proportion of underweight cylinders in the warehouse. Also,

$$v(p_{sy}) = \frac{v(\hat{N}_1)}{N^2}$$

$$= \frac{39384.90}{(8000)^2}$$

$$= .0006154$$

The required confidence interval for the proportion of underweight cylinders is given by

$$p_{sy} \pm 2\sqrt{v(p_{sy})}$$

$$= .1125 \pm 2\sqrt{.0006154}$$

$$= .0629,\ .1621$$

Thus, the proportion of underweight cylinders in the warehouse is most probably in the range .0629 to .1621. ■

It can be noted here that the lower and upper limits for the confidence interval of P can alternatively be obtained by dividing corresponding limits for the confidence interval for $N_1$ by N.

# HW

- Do questions 6.11, 6.12, and 6.13 in page 164