CHAPTER 17

SAMPLING METHODS

17.1 Introduction

Sampling is a part and parcel of our daily life. The housewife uses the technique of sampling in taking a decision whether the rice is cooked properly or not by inspecting a sample of grains from a cooking vessel A businessman inspects a sample of goods for ordering a large consignment. In industry, a sample would be observed to assess the quality of a product (or products). A farmer would estimate his crop prospects by observing a sample of earheads (or the plants). In the above situations, sampling is being followed to save money and time to arrive at an idea of the characteristic in the population. If there would be a considerable variation in the population, sampling adopted in the usual way might not give correct picture about the population. For example, the consumer wants to purchase rice by inspecting a handful of it from the upper portion of bag. If the quality of rice is not uniform throughout the bag the decision he takes on the basis of inspecting an upper portion of the material may bring him a monetary loss. Similarly the decision taken on only few bags out of large consignment of bags which are not having uniform quality would be of serious consequence. Hence different sampling procedures were evolved for different situations to estimate the population characteristics with minimum risk These sampling methods were developed based on probability theory. There is also a sampling method called 'purposive sampling' which do not use probability theory. The main drawback of 'purposive sampling' is that it is not possible to provide the error involved in arriving at an estimate of the population, and also the confidence intervals for the population characteristic.

17.2 Simple random sampling

In this method every unit in the population will have equal probability of being selected in the sample. Alternatively, the simple random sampling is the method of selecting 'n' sampling units out of total N units such that all the possible $\binom{N}{n}$ samples would have equal chance of being selected.

17.2.1 Sample random sampling with replacement (SRSWR): A sample is drawn such that every sampling unit drawn would be replaced back in the population. In this way the sample may contain repeated elements and any number of samples could be drawn.

17.2.2 Simple random sampling without replacement (SRSWOR): A sample is drawn such that every sampling unit drawn would not be replaced back. The sample would contain all distinct elements. If there are N units in the population and n units in the sample, there would be $\binom{N}{n}$ distinct samples by this method.

17.2.3 Selection of a random sample: List of units would be prepared by serially numbering all the sampling units from 1 to N and n random numbers would be selected from the column (or row) of a table of random numbers either by SRSWR or SRSWOR. For example, if $N=40$ and $n=5$, two columns would be selected from the table of random numbers. The maximum figure in two column table would be 99. The numbers 81 to 99 would be rejected since they have more probability than the numbers from 1 to 80. Supposing that 75 would be selected in the first draw, the actual random numbers would be the remainder after dividing 75 by 40 i. e., 35. If 80 would be selected in a particular draw the random number selected would be 40 since the remainder would be zero. In this way all the five numbers would be selected either by with or without replacement.

The method of providing estimates of population mean, standard error of mean and the confidence intervals for population mean are given as follows. Let Y_i be the i-th observational value for the character under study.

Sample

n = size of the sample

$\overline{Y}=\frac{1}{n}\sum_{i=1}^{n} Y_i$ = mean of the sample and is an unbiased estimate of the population mean, $\overline{Y}_N$.

$\hat{Y}=N.\overline{Y}_n$ = estimate of the population total, $\overline{Y}$ and is an unbiased estimate.

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\overline{Y})^2$$

$$=\frac{1}{n-1}\left[\sum_{i=1}^{n}Y_i^2-\frac{(\sum_{i=1}^{n}Y_i)^2}{n}\right]$$ = mean square in the sample and is an unbiased estimate of S^2.

$\text{Est}.V(\overline{Y}_n)=\frac{N-n}{Nn}s^2$ = estimate of the variance of the sample mean and is an unbiased estimate of $V(\overline{Y}_n)$.

$\text{Est. S.E}(\overline{Y}_n)=\sqrt{\text{Est } V(\overline{Y}_n)}$ = estimate of the standard error of sample mean and is an unbiased estimate of $S.E(\overline{Y}_n)$.

Population

N = size of the population

$\overline{Y}_N=1/N\sum_{i=1}^{N}Y_i$ = mean of the population.

$Y=\sum_{i=1}^{N}Y_i$ = Population total

$$S^2=\frac{1}{N-1}\sum_{i=1}^{N}(Y_i-\overline{Y}_N)^2$$

$$=\frac{1}{N-1}\left[\sum_{i=1}^{N}Y_i^2-\frac{(\sum Y_i)^2}{N}\right]$$

= mean square in the population.

$V(\overline{Y}_n)=\frac{N-N}{Nn}S^2$ = variance of sample mean in the population, and $\frac{N-n}{N}$ is called the finite population correction. $S.E(\overline{Y}_n)=\sqrt{V(\overline{Y}_n)}$ = standard error of sample mean in the population.

$\text{Est } V(\hat{Y}) = \frac{N^2(N-n)}{Nn} s^2$ = estimate of variance of the estimate of total and is an unbiased estimate of $V(\hat{Y})$.

Confidence limits: If S^2 is not known and the size of sample is small, the confidence limits for population mean, $\overline{Y}_N$ are given as

$$\overline{Y}_n \pm t_{(n-1)} \times \text{Est.S.E.}(\overline{Y}_n)$$

as upper and lower limits. These limits can be written as

$$\overline{Y}_n - t_{(n-1)} \times \text{Est S.E.}(\overline{Y}_n)$$

and $$\overline{Y} + t_{(n-1)} \times \text{Est.S.E.}(\overline{Y}_n)$$

where $t_{(n-1}$ is tabulated value of student's t-distribution with $(n-1)$ d.f.

$V(\hat{Y}) = N^2 \frac{(N-n)}{Nn} S^2$ = variance of sample total in the population.

EXAMPLE: A sample of 50 progressive farmers were selected from a district containing 800 progressive farmers by simple random sampling method so as to estimate the total area under high yielding variety of paddy. The list of selected farmers along with corresponding areas under high yielding variety (HYV) is given in Table 17.1. Estimate the mean area under HYV, standard error and confidence limits for the mean area in the district.

TABLE 17.1

Holding	*Area (Hectares)*	*Holding*	*Area (Hectares)*	*Holding*	*Area (Hectares)*
1	3.5	18	4.2	35	2.1
2	3.2	19	6.1	36	2.4
3	2.5	20	1.1	37	1.5
4	4.0	21	1.0	38	1.1
5	3.2	22	1.7	39	0.7
6	2.0	23	2.3	40	3.1
7	2.2	24	5.2	41	3.3
8	1.5	25	4.6	42	2.8
9	2.6	26	0.8	43	2.2
10	2.8	27	1.9	44	4.3
11	3.5	28	2.5	45	3.8
12	3.0	29	2.6	46	6.2
13	1.4	30	3.1	47	5.0
14	1.2	31	6.2	48	0 7
15	1.3	32	5.4	49	0.9
16	3.6	33	3.6	50	1.2
17	3.2	34	4.5		

$\Sigma Y = 141.6, \ \Sigma Y^2 = 517.74$

$$\overline{Y}_n = \frac{141.6}{50} = 2.83, \quad s^2 = \frac{1}{50-1}\left[517.74 - \frac{(141.6)^2}{50}\right] = 2.38$$

$$\text{Est.}\, V(\overline{Y}_n) = \frac{800-50}{800 \times 50} \times 2.38 = 0.0446$$

$$\text{Est.}\, SE(\overline{Y}_n) = \sqrt{0.0446} = 0.2112$$

Confidence limits of $\overline{Y}_N$;

Lower limit: $2.83 - 1.96 \times 0.2112 = 2.42$

Upper limit: $2.83 + 1.96 \times 0.2112 = 3.24$