2.Estimation of population total Y':

WR:

- Unbiased estimator of population total Y': $y' = N\,\bar{y}$ (3.11)
- variance of y': $V(y') = N^2 V(\bar{y}) = N^2\,\frac{\sigma^2}{n}$ (3.12)
- Unbiased estimator of $V(y') = v(y') = N^2 v(\bar{y}) = N^2\,\frac{s^2}{n}$ (3.13)

WOR:

- Unbiased estimator of population total Y': $y' = N\,\bar{y}$
- Sample variance of y': $V(y') = N^2 V(\bar{y}) = N(\frac{N-n}{n})S^2$ (3.14)
- Unbiased estimator of $V(y') = v(y') = N^2\,v(\bar{y}) = N(\frac{N-n}{n})s^2$ (3.15)

Example 3.5

From the data of WOR sample comprising of 10 villages in example 3.1, estimate the total number of tractors in the development block of 69 villages. Also, set up the confidence interval for it.

Solution

We have N=69 and n=10. For the sake of convenience, the data has been presented in table 3.7 along with some other computations. From (3.11), the estimate of total number of tractors in the block is

$$y' = N\,\bar{y}$$

$$= (69)\,(16.7) \qquad \text{(from example 3.4)}$$

$$= 1152.3$$

$$\approx 1152$$

The estimate of variance of y' is provided by (3.13). Thus,

$$v(y') = N^2 \, v(\bar{y})$$

Substituting the value of $v(\bar{y})$ from example 3.4, one gets

$$v(y') = (69)^2 (6.956)$$
$$= 33117.5$$

We now work out confidence interval for the population total. Following (2.8), the required confidence limits will be

$$y' \pm 2\sqrt{v(y')}$$

$$= 1152.3 \pm 2\sqrt{33117.5}$$

$$= 788.3,\ 1516.3$$

$$\approx 788,\ 1516$$

Often it happen that the characteristic under study is of qualitative nature, for example we might be interested to study the academic status of individuals (literate or illiterate) for a population. In such situations the individuals

in the population or in the sample will either possess the characteristic or will not possess the characteristic. The variable under study in this case will be binary variable (1 if individual has the characteristic and 0 otherwise) and the measure under study is the *proportion of individuals who possess the characteristic*. If Y is the variable and study and A is the total number of individuals in population who possess the characteristic under study then the population proportion is defined as

$$P = \frac{No.\ of\ Individuals\ with\ Characteristic}{Total\ Number\ in\ the\ Population} = \frac{A}{N}.$$

The sample proportion is similarly defined as

$$p = \frac{No.\ of\ Individuals\ in\ Sample\ with\ Characteristic}{Total\ Number\ in\ the\ Sample} = \frac{a}{n}.$$

3.Estimation of population proportion P:

- Unbiased estimator of population proportion P:

$$p = \frac{a}{n} \qquad (3.22)$$

<u>WR:</u>

- variance of p: $V(p) = \frac{PQ}{n}$; $P = \frac{A}{N}$,$Q = 1 - P$ (3.23)
- Unbiased estimator of $V(p) = v(p) = \frac{pq}{n-1}$ (3.24)

<u>WOR:</u>

- Sample variance of P: $V(p) = \frac{(N-n)}{N-1}\frac{PQ}{n}$ (3.25)
- Unbiased estimator of $V(p) = v(p) = \frac{(N-n)}{N}\frac{pq}{n-1}$ (3.26)

Example 3.10

Punjab Agricultural University, Ludhiana, is interested in estimating the proportion P of teachers who consider semester system to be more suitable as compared to the trimester system of education. A with replacement simple random sample of n=120 teachers is taken from a total of N=1200 teachers. The response is denoted by 0 if the teacher does not think the semester system suitable, and 1 if he/she does. From the sample observations given below, estimate the proportion P along with the standard error of your estimate. Also, work out the confidence interval for P.

Teacher	:	1	2	3	4	5	6	. . .	119	120	Total
Response	:	1	0	1	1	0	1	. . .	0	1	72

Solution

Estimate of P is given by (3.22). Thus,

$$p = \frac{72}{120} = .6$$

Estimate of the standard error of p will then be obtained as

$$\begin{aligned} se(p) &= \sqrt{v(p)} \\ &= \sqrt{\frac{pq}{n-1}} \\ &= \sqrt{\frac{(.6)\,(.4)}{119}} \\ &= .04491 \end{aligned}$$

The confidence limits for P would be obtained following (2.8). Thus,

$$
\begin{aligned}
& p \pm 2\sqrt{v(p)} \\
= \; & .6 \pm 2(.04491) \\
= \; & .5102, .6898
\end{aligned}
$$

The proportion of teachers in the university favoring semester system is, therefore, likely to be in the closed interval [.5102, .6898]. ■

In certain situations, the objective could be to estimate the number A of units possessing the attribute of interest. The unbiased estimator A' of A would be N times the sample proportion p defined in (3.22). The variance and its estimator respectively for A' would be N^2 times the variance and the estimator of variance for p.

Sample size

- When we want to conduct a sample survey from a finite population then first thing to decide is size of the sample to be selected.
- The size of the sample is based upon the population variance and amount of error which we want to afford in our estimates.
- Suppose that the population variance is S^2 and the amount of error which we have is $B = 2\sqrt{Var(estimate)}$
- Then sample size for estimation of population mean and population proportion can be easily computed as in the following silds.

Determining sample size for estimating population mean($\bar{Y}$)/total (Y')

Population mean:

- $n = \frac{NS^2}{ND+S^2}$ WOR
- $n = \frac{S^2}{D}$ WR

where $D = \frac{B^2}{4}$

Population total:

- $n = \frac{NS^2}{ND^*+S^2}$ WOR
- $n = \frac{S^2}{D^*}$ WR

where$D^* = \frac{B^2}{4N^2}$

Example 11: A survey is to be conducted to estimate the average monthly income of a locality with 5000 households. It is known that the squared variability in the income is 250000 Rayal. How many households should we select so that the standard error of estimated income is no more than 100 Rayals?

Solution: We have $N = 5000$, $S^2 = 250000$ and $SE(\bar{y}) = 100$. Now

$$\begin{aligned}
D &= 2\sqrt{Var(\bar{y})} = 2 \times SE(\bar{y}) = 2 \times 100 = 200 \\
A &= \frac{B^2}{4} = \frac{200^2}{4} = 10000. \\
n &= \frac{NS^2}{N\,D + S^2} = \frac{(5000)(250000)}{(5000)(10000) + 250000} = 24.87 \simeq 25.
\end{aligned}$$

So a sample of 25 households is required to conduct the survey.

Determining sample size for estimating population proportion

Population proportion

- $n = \frac{NPQ}{(N-1)D+PQ}, \quad D = \frac{B^2}{4}$

- $n = \frac{N}{(N-1)B^2+1}$; P is unknown, P=Q=1/2

Example 12: It is known that the proportion of smokers in a society of 4000 individuals is 0.2. How many individuals we need to select if we want to estimate the proportion of smokers with maximum standard error in estimate equal to 2%.

Solution: Here we have $N = 4000$, $P = 0.2$ and $SE(p) = 0.02$. Now

$$
\begin{aligned}
B &= 2\sqrt{Var(p)} = 2 \times SE(p) = 2 \times 0.02 = 0.04 \\
D &= \frac{B^2}{4} = \frac{0.04^2}{4} = 0.0004. \\
n &= \frac{N\,P\,Q}{(N-1)\,D + P\,Q} = \frac{(4000)(0.2)(0.8)}{(3999)(0.0004) + (0.2)(0.8)} \\
&= 363.719 \Rightarrow n \simeq 364.
\end{aligned}
$$

Hence we need a sample of 364 individuals to estimate the proportion of smokers.

Example 13: A survey is to be conducted to estimate the proportion of families who have smart LCD's. It is known that total number of families in the locality is 6000. How many families we need to select so that the error in estimated proportion is not more that 5%.

Solution: Here we have $N = 5000$ and $B = 5\% = 0.05$. Since proportion of families with samart LCD's is unknown we assume that $P = Q = 1/2$ and hence the sample size is

$$n = \frac{N}{(N-1)\,B^2 + 1} = \frac{6000}{(5999 \times 0.05^2) + 1} = 375.06 \simeq 375.$$

Hence we need to selec 375 families.

HW

- If a simple random sample without replacement (WOR) of 10 universities are selected from a population of 50 universities in a particular country. The numbers of statistics professors in the sample university are: 23, 14, 38, 11, 7, 31, 9, 18, 12, 25. Solve by R:

1. Estimate the mean number of statistics professors in this population, $\bar{y}$.
2. Estimate the variance of your estimator, $\{v(\bar{y})\}$.
3. Estimate the total number of statistics professors in this population, y'.
4. Estimate the variance of your estimator, $\{v(y')\}$.
5. How many possible samples can be drawn.
6. Give an approximate 95% confidence interval for the population mean.
7. Give an approximate 95% confidence interval for the population total.

- Page 63 questions: 3.10, 3.11, 3.12, 3.13