

ON THE PRECISION OF AN ESTIMATOR OF MEAN FOR DOMAINS IN DOUBLE SAMPLING FOR INCLUSION PROBABILITIES

GODWIN A. UDOFIA

(Received 5 February 2002; Revision accepted 6 June 2002)

ABSTRACT

In sample surveys, estimates are often required for small subclasses of the population under study. In many sample situations, the size of the domain and hence the number of sample units that fall in the domain are unknown before the start of the survey. In this paper, we examine the effect of the randomness of the domain size on the precision of an estimator for the domain under double sampling for inclusion probabilities. It is assumed that the population can be divided into different strata. The results show that there is a positive contribution to the variance of the estimator which varies from one stratum to another. This addition vanishes where the domain coincides with a stratum. The total sampling variance depends only on components of variance for the domain and is inversely related to the total sample size in each phase of the survey.

Key Words: Domain, Double Sampling, Unequal Probabilities.

INTRODUCTION

In the analysis of sample survey data, estimates are usually required for small subclasses of the population under study. Such subclasses have been tagged domains of study by the United Nation. Statistical Office (1950). For example, in a fertility survey, estimates of parity may be needed for a certain class of women who are gainfully employed. The problem that the survey statistician ordinarily faces is that the domain often cuts across the various strata of the population with unknown weights and hence the domain sample size is random.

Yates (1953) first considered in detail some of the problems associated with the estimation of domain totals, means and proportions. Derivation of Yates' results was given by Durbin (1958) and by Hartley (1959). Kish (1969) considers allocation of resources when domains of study are of primary interest. Scott and Smith (1971) discuss the application of Bayesian approach to estimation for domains. Tin and Toe (1972) extend the available results to Multistage sampling while Tripath (1988) extends the results to sampling on two occasions. The results obtained in the above studies indicate that the sampling variance of an estimator increases by different factors under the different sample designs adopted.

It is well known that under certain conditions, sampling with probability proportional to size gives a more precise result than equal probability sampling: see for example Raj (1954, 1958, 1968),

Cochran (1977) and Foreman and Brewer (1977). Furthermore, keeping the selection probabilities proportional to the chosen size is difficult in sampling without replacement and soon becomes impossible as n increases whereas selection with replacement leads to simple formulae for the true and the estimated variances of the estimators: see Cochran (1977) section 9A.2. For this reason we consider in this paper, selection with probabilities proportional to size and with replacement. Where information on the size of each sampling unit is lacking, double sampling for unequal probabilities is the sample design often used. It is therefore also of interest to determine the manner and direction of effect of randomness of the domain sample size on the sampling variance of a proposed estimator. This paper is an attempt in this direction.

SAMPLE DESIGN

We assume that the population under study consists of H strata with N_h elements in the h^{th} stratum $h=1,2,\dots,H$. We also assume that there is no reliable information on the size, X , of each element in any given stratum. Let D_{hj} denote the part of domain j ; $j=1,2,\dots,M$, in stratum h . The number of elements,

N_{hj} , in D_{hj} (and hence $N_j = \sum_h N_{hj}$) is unknown.

An initial sample, S_{1h} , of size n_{1h} , is drawn by simple random sampling without replacement independently from each stratum and the auxiliary variable X is measured on it. Suppose that out of n_{1h} units of the initial sample n_{1hj} units fall in D_{hj} . For any fixed n_{1h} , n_{1hj} is a random sample and $0 < n_{1hj} < n_{1h}$. A subsample, S_{2h} , of n_{2h} , $n_{2h} < n_{1h}$, $h=1,2,\dots,H$, units is drawn from S_{1h} with probability proportional to X and with replacement. The study variable, Y , is measured on S_{2h} . Let n_{2hj} denote the number of units in S_{2h} that belong to D_{hj} . Also n_{2hj} is a random variable.

The Sampling Variance of an Estimator of Domain Total

Let

$$\begin{aligned} y_{hi}^j &= Y_{hij} \text{ if the } i^{\text{th}} \text{ element is in } D_{hj} \\ &= 0 \text{ otherwise.} \end{aligned} \quad \dots (1)$$

Then an unbiased estimator of the total of Y for domain j defined as

$$Y_j = \sum_{h=1}^H \sum_{i=1}^{N_{hj}} Y_{hij}$$

is given by

$$\hat{Y}_j = \sum_{h=1}^H \frac{N_h}{n_{1h}} \frac{1}{n_{2h}} \sum_{i=1}^{n_{2h}} \frac{y_{hi}^j}{p_{hi}}$$

where

$$p'_{hi} = x'_{hi} / x'_{1h}, \quad x'_{1h} = \sum_{i=1}^{n_h} x'_{hi}$$

is the probability that the i th ($i=1, 2, \dots, n_{1h}$) of the first sample will be included in the second sample.

It is known that this estimator is unbiased: see Udofia (2001). We now examine the effect of the randomness of the domain sample size of the sampling variance of the estimator.

By conditional variance formula,

$$V(\hat{Y}_j) = V_1 E_2(\hat{Y}_j) + E_1 V_2(\hat{Y}_j). \quad \dots (2)$$

See Raj (1956).

Now

$$V_1 E_2(\hat{Y}_j) = V_1 \left(\sum_{h=1}^H N_h \bar{y}'_{1h} \right) = \sum_{h=1}^H N_h^2 V_1(\bar{y}'_{1h})$$

The crossproduct terms vanish because of independence of sampling within each stratum.

Since \bar{y}'_{1h} is the mean of a simple random sample of size n_{1h} from N_h elements,

substitute for $V_1(\bar{y}'_{1h})$ from simple random sample theorem and obtain the

result.

$$V_1 E_2(\hat{Y}_j) = \sum_{h=1}^H N_h^2 \left(\frac{1}{n_{1h}} - \frac{1}{N_h} \right) S_{y'h}^2 \quad \dots (3)$$

Let

$$\begin{aligned} d'_{hi} &= 1 \text{ if the } i^{\text{th}} \text{ element of stratum } h \text{ is in } D_{in} \\ &= 0 \text{ otherwise} \end{aligned} \quad \dots (4)$$

Then

$$\bar{d}'_{2i} = \frac{1}{n_{2h}} \sum_{i=1}^{n_{2h}} d'_{hi} = \frac{n_{2hj}}{n_{2h}}$$

$$D'_h = \frac{1}{N_h} \sum_{i=1}^{N_h} y'_{hi}$$

$$\bar{Y}'_{2h} = \frac{1}{n_{2h}} \sum_{i=1}^{n_{2h}} y'_{hi} = \frac{n_{2hj}}{n_{2h}} \bar{Y}'_{2hj}$$

$$\bar{Y}'_{1h} = \frac{1}{n_{1h}} \sum_{i=1}^{n_{1h}} y'_{hi} = \frac{n_{1hj}}{n_{1h}} \bar{Y}'_{1hj}$$

$$\bar{Y}'_h = \frac{1}{N_h} \sum_{i=1}^{N_h} y'_{hi} = \frac{N_{hj}}{N_h} \bar{Y}'_{hj} \dots (5)$$

$$(N_h - 1) S_{y'h}^2 = \sum_{i=1}^{N_h} (y'_{hi} - \bar{Y}'_h)^2 \dots \dots \dots (6)$$

Let A be any constant and

$$U_{hi} = y'_{hi} - A \dots (7)$$

so that

$$y'_{hi} = U_{hi} + A$$

Then

$$\bar{Y}'_h = \frac{1}{N_h} \sum_{i=1}^{N_h} y'_{hi} = \frac{1}{N_h} \sum_{i=1}^{N_h} (u_{hi} + A) = \frac{1}{N_h} \sum_{i=1}^{N_h} u_{hi} + A$$

and hence (6) becomes

$$(N_h - 1) S_{y'h}^2 = \sum_{i=1}^{N_h} (y'_{hi} - \bar{Y}'_h)^2 = \sum_{i=1}^{N_h} \left(u_{hi} - \frac{1}{N_h} \sum_{i=1}^{N_h} u_{hi} \right)^2$$

Substitution for u_{hi} from (7) in the last expression gives

$$(N_h - 1) S_{y'h}^2 = \sum_{i=1}^{N_h} \left[y'_{hi} - A - \frac{1}{N_h} \sum_{i=1}^{N_h} (y'_{hi} - A) \right]^2$$

Since A can be any constant, we replace it by $d'_{hi} \bar{Y}_j$ which is the mean

of y'_{hi} , a constant value, for D_{hi} and obtain

$$\begin{aligned} (N_h - 1) S_{y'h}^2 &= \sum_{i=1}^{N_h} \left[y'_{hi} - d'_{hi} \bar{Y}_j - \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (y'_{hi} - d'_{hi} \bar{Y}_j) \right]^2 \\ &= \sum_{i=1}^{N_h} (y'_{hi} - d'_{hi} \bar{Y}_j)^2 - N_h (\bar{Y}_h - \bar{D}'_h \bar{Y}_j)^2 \end{aligned}$$

Substitution from (1), (4) and (5) for y'_{hi} , d'_{hi} and \bar{Y}'_h respectively in the expression on the right-hand-side gives the result

$$(N_h - 1) S_{y'h}^2 = \sum_{i=1}^{N_{hj}} (Y_{hij} - \bar{Y}_j)^2 - \frac{N_{hj}^2}{N_h} (\bar{Y}_{hj} - \bar{Y}_j)^2 \dots (8)$$

Now

$$\begin{aligned} \sum_{i=1}^{N_{hj}} (Y_{hij} - \bar{Y}_j)^2 &= \sum_{i=1}^{N_{hj}} \left[Y_{hij} - \bar{Y}_{hj} + (\bar{Y}_{hj} - \bar{Y}_j) \right]^2 \\ &= \sum_{i=1}^{N_{hj}} (Y_{hij} - \bar{Y}_{hj})^2 + N_{hj} (\bar{Y}_{hj} - \bar{Y}_j)^2 \end{aligned}$$

We substitute this result in (8) and obtain

$$S_{y'h}^2 = \frac{1}{N_h - 1} \left\{ (N_{hj} - 1) S_{y(hj)}^2 + N_{hj} \left(1 - \frac{N_{hj}}{N_h} \right) (\bar{Y}_{hj} - \bar{Y}_j)^2 \right\} \dots (9)$$

We substitute (9) in (3) and obtain

$$V_1 E_2 (\hat{Y}_j) = \sum_{h=1}^H N_h^2 \left(\frac{1}{n_{1h}} - \frac{1}{N_h} \right) \frac{1}{N_h - 1} \left\{ (N_{hj} - 1) S_{y(hj)}^2 + N_{hj} \left(1 - \frac{N_{hj}}{N_h} \right) (\bar{Y}_{hj} - \bar{Y}_j)^2 \right\} \dots (10)$$

where

$$S_{Y^{(h)h}}^2 = \frac{1}{N_{1h} - 1} \sum_{i=1}^{N_{1h}} (Y_{hi} - \bar{Y}_h)^2$$

The second term on the right-hand-side of (2) is calculated as follows:

$$V_2(\hat{Y}_j) = V_2 \left(\sum_{h=1}^L \frac{N_h}{n_{1h}} \frac{1}{n_{2h}} \sum_{j=1}^{n_{2h}} \frac{Y'_{hi}}{P'_{hi}} \right) = \sum_{h=1}^L \frac{N_h^2}{n_{1h}^2} \frac{1}{n_{2h}} \sum_{j=1}^{n_{2h}} P'_{hi} \left(\frac{Y'_{hi}}{P'_{hi}} - \sum_{i=1}^{n_{2h}} Y'_{hi} \right)^2$$

since the covariance term vanishes as before.

From Raj, (1968), the above can be written as

$$V_2(\hat{Y}_j) = \sum_{h=1}^L \frac{N_h^2}{n_{1h}^2} \sum_{i=1}^{n_{1h}} \sum_{k>i}^{n_{1h}} x'_{hi} x'_{hk} \left(\frac{Y'_{hi}}{x'_{hi}} - \frac{Y'_{hk}}{x'_{hk}} \right)^2$$

and then the conditional expectation (for fixed n_{1h}) is calculated as follows:

$$E_1 V_2(\hat{Y}_j) = \sum_{h=1}^L \frac{N_h^2}{n_{1h}^2} \frac{1}{n_{2h}} \frac{n_{1h}(n_{1h}-1)}{N_h(N_h-1)} \sum_{i=1}^{N_h} \sum_{k>i}^{N_h} x'_{hi} x'_{hk} \left(\frac{Y'_{hi}}{x'_{hi}} - \frac{Y'_{hk}}{x'_{hk}} \right)^2$$

Substitution for x'_{hi} and y'_{hi} leads to the result

$$E_1 V_2(\hat{Y}_j) = \sum_{h=1}^L \frac{N_h^2}{n_{1h}^2} \frac{1}{n_{2h}} \frac{n_{1h}(n_{1h}-1)}{N_h(N_h-1)} \sum_{i=1}^{N_h} \sum_{k>i}^{N_h} x_{hij} x_{hkj} \left(\frac{y_{hij}}{x_{hij}} - \frac{y_{hkj}}{x_{hkj}} \right)^2$$

$$= \sum_{h=1}^L \frac{N_h}{N_h-1} \frac{n_{1h}-1}{n_{1h}n_{2h}} V_P(Y)_{hj} \quad \dots (11)$$

where

$$V_P(Y)_{hj} = \sum_{i=1}^{N_{hi}} \sum_{k>i}^{N_{hi}} x_{hij} x_{hkj} \left(\frac{y_{hij}}{x_{hij}} - \frac{y_{hkj}}{x_{hkj}} \right)^2$$

Finally (10) and (11) are substituted into (2) to obtain

$$V(\hat{Y}_j) = \sum_{h=1}^L \frac{N_h}{N_h-1} \frac{n_{1h}-1}{n_{1h}n_{2h}} V_p(y)_{hj} + \sum_{h=1}^L \frac{N_h(N_h-n_{1h})}{(N_h-1)n_{1h}} \left\{ (N_{hj}-1) S_{y^{(hj)}}^2 + N_{hj} \left(1 - \frac{N_{hj}}{N_h} \right) (\bar{Y}_{hj} - \bar{Y}_j)^2 \right\} \quad (12)$$

An unbiased estimator of $S_{y^{(hj)}}^2$ based on the initial sample is

$$S_{y^{(hj)}}^2 = \frac{1}{n_{1hj}-1} \sum_{i=1}^{n_{1hj}} (y_{hij} - \bar{y}_{1hj})^2; \bar{y}_{1hj} = \frac{1}{n_{1hj}} \sum_{i=1}^{n_{1hj}} y_{hij}.$$

Since values of Y are not available for the initial sample, we use the second phase sample values for the computation as follows

$$\begin{aligned} S_{y'h}^2 &= \frac{1}{n_{1h}-1} \left\{ \frac{1}{n_{2h}} \sum_{i=1}^{n_{2h}} \frac{Y_i}{P_{hi}} - \frac{1}{n_{1h}n_{2h}(n_{2h}-1)} \left[\left(\sum_{i=1}^{n_{2h}} \frac{Y_{hi}'}{P_{hi}'} \right)^2 - \sum_{i=1}^{n_{2h}} \frac{Y_{hi}^2}{P_{hi}^2} \right] \right\} \\ &= \frac{1}{n_{2h}(n_{1h}-1)} \left\{ X'_{1h} \sum_{i=1}^{n_{2h}} \frac{Y_{hi}^2}{x_{hi}'} - \frac{X_{1h}^2}{n_{1h}(n_{2h}-1)} \left[\left(\sum_{i=1}^{n_{2h}} \frac{Y_{hi}'}{x_{hi}'} \right)^2 - \sum_{i=1}^{n_{2h}} \frac{Y_{hi}^2}{x_{hi}^2} \right] \right\} \end{aligned}$$

See Raj (1968), section 7.3.

Substitution for X'_{hi} and y'_{hi} leads to the result

$$S_{y^{(hj)}}^2 = \frac{1}{n_{2h}} \left\{ X_{1hj} \sum_{i=1}^{n_{2hj}} \frac{Y_{hij}}{x_{hij}} - \frac{X_{1hj}^2}{n_{1h}(n_{2h}-1)} \left[\left(\sum_{i=1}^{n_{2hj}} \frac{Y_{hij}}{x_{hij}} \right)^2 - \sum_{i=1}^{n_{2hj}} \frac{Y_{hij}^2}{x_{hij}^2} \right] \right\}$$

where

$$X_{1hj} = \sum_{i=1}^{n_{1hj}} x_{hij}.$$

This is an unbiased estimator of $S_{y^{(hj)}}^2$ in (3). Thus an unbiased estimator of $V_1 E_2(\hat{Y}_j)$ in (10) is given by

$$\sum_{h=1}^L N_h^2 \left(\frac{1}{n_{1h}} - \frac{1}{N_h} \right) \frac{1}{N_h-1} \frac{1}{n_{2h}} \left\{ X_{1hj} \sum_{i=1}^{n_{2hj}} \frac{Y_{hij}^2}{x_{hij}} - \frac{X_{1hj}^2}{n_{1h}(n_{2h}-1)} \left[\left(\sum_{i=1}^{n_{2hj}} \frac{Y_{hij}}{x_{hij}} \right)^2 - \sum_{i=1}^{n_{2hj}} \frac{Y_{hij}^2}{x_{hij}^2} \right] \right\} +$$

$$+ \frac{n_{1hj}}{n_{1h}} N_h \left(1 - \frac{n_{1hj}}{n_{1h}} \right) X_{1hj}^2 \left(\frac{1}{n_{2hj}} \sum_{i=1}^{n_{2hj}} \frac{Y_{hij}}{x_{hij}} - \frac{1}{n_{2j}} \sum_{h=1}^H \sum_{i=1}^{n_{2hj}} \frac{Y_{hij}}{x_{hij}} \right)^2 \dots (13)$$

where we have used $N_{hj} = (n_{1hj}/n_{1h})N_h$ following Durbin (1958).

Also from Raj (1968), an unbiased estimator of $E_1 V_2(\hat{Y}_j)$ is obtained as

$$\sum_{h=1}^L \frac{N_h^2}{n_{1h}^2} \frac{X_{1hj}^2}{n_{2h}(n_{2h}-1)} \sum_{i=1}^{n_{2hj}} \left(\frac{Y_{hij}}{x_{hij}} - \frac{1}{n_{2h}} \sum_{i=1}^{n_{2hj}} \frac{Y_{hij}}{x_{hij}} \right)^2$$

Substitution for x'_{hi} and y'_{hi} gives the result

$$\sum_{h=1}^L \frac{N_h^2}{n_{1h}^2} \frac{X_{1hj}^2}{n_{2h}(n_{2h}-1)} \sum_{i=1}^{n_{2hj}} \left(\frac{Y_{hij}}{x_{hij}} - \frac{1}{n_{2h}} \sum_{i=1}^{n_{2hj}} \frac{Y_{hij}}{x_{hij}} \right)^2 \dots (14)$$

By using (12) and (13) in (3), an unbiased estimator of $V(\hat{Y}_j)$ is obtained as

$$\begin{aligned} \hat{V}(\hat{Y}_j) &= \sum_{h=1}^L \frac{N_h^2}{n_{1h}^2} \frac{X_{1hj}^2}{n_{2h}(n_{2h}-1)} \sum_{i=1}^{n_{2hj}} \left(\frac{Y_{hij}}{x_{hij}} - \frac{1}{n_{2h}} \sum_{i=1}^{n_{2hj}} \frac{Y_{hij}}{x_{hij}} \right)^2 + \\ &+ \sum_{h=1}^L \frac{N_h(N-n_{1h})}{(N_h-1)n_{1h}n_{2hj}} \left[X_{1hj} \sum_{i=1}^{n_{2hj}} \frac{Y_{hij}^2}{x_{hij}} - \frac{X_{1hj}}{n_{1h}(n_{2h}-1)} \left(\left(\sum_{i=1}^{n_{2hj}} \frac{Y_{hij}}{x_{hij}} \right)^2 - \sum_{i=1}^{n_{2hj}} \frac{Y_{hij}^2}{x_{hij}^2} \right) \right] + \\ &+ \frac{n_{1hj}}{n_{1h}} N_h \left(1 - \frac{n_{1hj}}{n_{1h}} \right) X_{1hj}^2 \left(\frac{1}{n_{2hj}} \sum_{i=1}^{n_{2hj}} \frac{Y_{hij}}{x_{hij}} - \frac{1}{n_{2j}} \sum_{h=1}^H \sum_{i=1}^{n_{2hj}} \frac{Y_{hij}}{x_{hij}} \right)^2 \end{aligned}$$

COMPARISON WITH GLOBAL ESTIMATOR

Under double sampling for probabilities proportional to size design, a corresponding unbiased estimator of the total of Y for the entire population described above, which we designate as a global estimator in this paper, is

$$\hat{Y} = \sum_{h=1}^H \frac{N_h}{n_{1h}} \frac{1}{n_{2h}} \sum_{i=1}^{n_{2hj}} \frac{Y_{hij}}{P_{hj}}$$

with variance

$$V(\hat{Y}) = \sum_{h=1}^L \frac{N_h}{N_h-1} \frac{n_{1h}-1}{n_{1h}n_{2h}} \nabla_p(Y)_h + \sum_h \frac{N_h(N_h-n_{1h})}{n_{1h}} S_{Y_{hj}}^2 \dots (15)$$

where

$$S_{yh}^2 = \frac{1}{N_h - 1} \sum_{i=1}^{N_h} (Y_{hi} - \bar{Y}_h)^2 \text{ and } V_p(Y) = \sum_{i=1}^N \sum_{h=1}^H x_{hi} x_{hk} \left(\frac{Y_{hi}}{x_{hi}} - \frac{Y_{hk}}{x_{hk}} \right)^2$$

A comparison of (12) and (15) shows that whereas $V(Y)$ depends on components of variance of Y for the entire population, $V(\bar{Y})$ depends on components of variance for domain j under study and on both the variability of the domain size and the variability of the domain mean, \bar{Y}_j . Thus the effect of lack of prior knowledge of the domain size is the addition of a positive quantity,

$$\sum_{h=1}^H N_h^2 \left(\frac{1}{n_{1h}} - \frac{1}{N_h} \right) \frac{N_{hj}}{N_h - 1} \left(1 - \frac{N_{hj}}{N_h} \right) (\bar{Y}_{hj} - \bar{Y}_j)^2 \dots \dots \dots (16)$$

to $V(\bar{Y})$ in (12) as a result of which the domain estimator is less precise than the Global estimator. Under simple random sampling theory,

$$\frac{N_{hj}}{N_h - 1} \left(1 - \frac{N_{hj}}{N_h} \right) / n_{1h} = P_{hj} (1 - P_{hj}) / n_{1h}$$

under the summation sign over h in (16) is the variance of the proportion of the domain elements in stratum h , $h = 1, 2, \dots, H$, that fall in the initial sample. The above positive contribution to the sampling variance of the domain estimator vanishes if the domain coincides with the h^{th} stratum in which case N_{hj} equals N_h . For such a case, no special theory is necessary.

REFERENCES

- Cochran, W. G., 1977. *Sampling Techniques*. John Wiley and Sons Inc., New York.
- Durbin, J., 1958. Sampling Theory for estimates based on fewer and individuals than the number selected. *Bull. Int. Stat. Inst.*, 36(3): 113-119.
- Foreman, E. K. and Brewer, K. W. R., 1971. The efficient use of supplementary information in standard sampling procedures. *Jour. Roy. Stat. Soc.*, B33: 391-400.
- Hartley, H.O., 1959. *Analytical Studies of Survey Data*. Instituto di Statistica. Rome.
- Kish, L., 1969. Design and estimation for subclass, comparative and analytical statistics. In: N.L. Smith and Jr. Harry (Editors), *New Development in Survey Sampling*, John Wiley and Sons Inc., New York.
- Raj, D., 1954. On Sampling with probabilities proportional to size. *Ganita*, 5: 175-182.
- Raj, D., 1956. Some estimators in Sampling with varying probabilities without replacement. *Jour. Ind. Doc. Agric. Stat.*, 6: 127 - 133.
- Raj, D., 1958. On the relative accuracy of some sampling techniques. *Jour. Amer. Stat. Assoc.*, 53 : 98-101.

- Raj, D., 1968. *Sampling Theory*. McGraw Hill Book Co., New York.
- Scott, A. and Smith, T.M.F., 1971. Bayes estimates for subclasses in stratified sampling. *Jour. Amer. Stat. Assoc.*, 66 : 834 - 836.
- Tin, M. and Toe, T., 1972. Estimation for domains in multistage sampling. *Jour. Amer. Stat. Assoc.*, 67 : 913 - 916.
- Udofia, G., 2001. Estimation for mains in double sampling for probabilities proportional to size. *Sankhyā, B*, in print.
- Tripathi, T. P., 1988. Estimation for domains in sampling on two occasions. *Sankhya*, 50: 103 - 110.
- U.N. Statistical Office, 1950. *The Preparation of Sample Survey Reports: Stat. Papers Series C*. No. 1.
- Yates, F., 1953. *Sampling Methods for Censuses and Surveys*. Charles Griffin and Co., London.