Journal of the
Indian Society of Agricultural Statistics

Vol. XVI] 1964 [No. 2

UNBIASED RATIO AND REGRESSION ESTIMATORS IN MULTI-STAGE SAMPLING

By J. N. K. Rao

Iowa State University

I. Introduction

Ratio and regression estimators are widely used in sample surveys for estimating the population mean or total of a characteristic 'y' by utilising the supplementary characteristic 'x' for which the population mean is known. In recent years, considerable attention has been given in the literature to the construction of unbiased ratio and regression estimators, since the classical ratio and regression estimators are biased. Mickey (1959) has given a method of constructing a broad class of unbiased ratio and regression estimators in simple random sampling. This class includes the well-known Hartley-Ross unbiased ratio estimator. However, since in practice multi-stage sampling is often employed, it seems necessary to extend Mickey's method to multi-stage designs. Ross (1960), in the unpublished part of his thesis, has extended Hartley-Ross unbiased ratio estimator to stratified and two-stage sampling. He has also explicitly derived the formulae for variance and estimated variance. It might be mentioned that Nieto's (1961) 'combined' unbiased ratio estimator in stratified sampling is applicable only for equal sample size per stratum and sampling with replacement, whereas the 'combined' estimator of Ross is applicable for unequal sample sizes and sampling without replacement in the strata. Williams (1961) has considered certain 'combined' unbiased ratio and regression estimators for stratified and two-stage sampling. He has also shown (Williams, 1962) how Mickey's method can be modified to construct his estimators in simple random sampling.

The purpose of the present paper is to extend Mickey's method to two-stage sampling and construct two different classes of 'combined' unbiased ratio and regression estimators. (The two classes are identical in the special case of stratified sampling.) In Section II it is shown that Williams' estimator for two-stage sampling can be obtained from our class 1 estimators. It is also shown that our class 2 includes the unbiased ratio estimator of Ross for two-stage sampling. Moreover, from class 2 a more general estimator (of Williams' type) is obtained. The estimators in class 1 depend on the sum of the population totals X_i of the n primaries selected in the sample* (which may not be known in practice), whereas the estimators in class 2 depend only on the overall population mean $\bar{X} = \sum_{1}^{N} X_i/M_0$, where $M_0 = \sum_{1}^{N} M_i$ and M_i is the number of secondaries in the i-th primary and N is the number of primaries in the population. Therefore, the estimators in class 2 may become more useful in practice. As pointed out by Mickey (1959), the question of relative efficiency of the estimators seems to be difficult. In specific cases it is possible to derive the variance explicitly (*e.g.*, Ross, 1960). However, the resulting expressions become very unwieldy so that it is difficult to make any meaningful efficiency comparisons. Also, the formulae for estimator of variance obtained by estimating each term in the variance formulae become very complex and will be useless in practice. Therefore, in Section III, a simpler method of variance estimation, similar to that proposed by Goodman and Hartley (1958) for the unbiased ratio estimator in simple random sampling, is considered.

The extension to three and higher stage sampling is relatively straightforward and therefore will not be considered here.

II. Two-Stage Sampling

Let y_{ij} and x_{ij} denote the values of 'y' and 'x' respectively for the j-th secondary in the i-th primary ($j = 1, \cdots, M_i$; $i = 1, \cdots, N$). The sample consists of n primaries selected with equal probabilities and without replacement and if the i-th primary is selected in the sample m_i secondaries are drawn with equal probabilities and without replacement.

---

* There seems to be a mistake in equation (11) of Williams (1961). $\sum_{1}^{r} X_i/(n\bar{M})$ should be inserted in the place of μ_x.

A class of 'separate' estimators is easily obtained by using Mickey's class of unbiased estimators, $T_{\beta_i}^{(i)}$, in each selected primary, where

$$T_{\beta_i}^{(i)} = a_i(Z_{\beta_i})\bar{X}_i + \frac{(M_i - \beta_i)}{M_i(m_i - \beta_i)}\{\bar{y}_i - a_i(Z_{\beta_i})\bar{x}_i\}$$

$$- \frac{(M_i - m_i)\beta_i}{M_i(m_i - \beta_i)}\{\bar{y}_i(\beta_i) - a_i(Z_{\beta_i})\bar{x}_i(\beta_i)\} \quad (1)$$

so that the overall population mean $\bar{Y}$ is estimated by

$$T_s = \frac{N}{n}\sum_1^n \frac{M_i}{M_0} T_{\beta_i}^{(i)}. \quad (2)$$

Here Z_{β_i} denotes the ordered set of observations on the first β_i, $(0 < \beta_i < m_i)$, sampled secondaries from the i-th selected primary; $a_i(Z_{\beta_i})$ denotes a function of these observations, and $(\bar{y}_i, \bar{x}_i)$ and $(\bar{y}_i(\beta_i), \bar{x}_i(\beta_i))$ are the i-th primary sample means based respectively on m_i and β_i secondaries. From Mickey's results, $T_{\beta_i}^{(i)}$ is an unbiased estimator of the i-th primary population mean $\bar{Y}_i$. This follows by observing that, given the ordered set of β_i secondaries, the remaining $(m_i - \beta_i)$ secondaries constitute a random sample from the finite population of $(M_i - \beta_i)$ secondaries obtained by excluding the β_i secondaries in the ordered set from the M_i secondaries in the i-th primary. Therefore, T_s is an unbiased estimator of $\bar{Y}$. One can now obtain the class of estimators T_s^* from T_s by averaging each $T_{\beta_i}^{(i)}$ over all permutations in the ordering of β_i sampled secondaries from the i-th primary. It may be noted that it is necessary to know the population means, $\bar{X}_i$, of the n primaries in the sample in order to construct T_s^* or T_s. In practice, we are interested in T_s^* since it does not depend on the order in which the sample is drawn. Also, as shown by Mickey (1959), the variance of T_s^* is not greater than the variance of T_s.

Special case.—Let $\beta_i = 1$ and $a_i(Z_{\beta_i}) = y_{ij}/x_{ij} = r_{ij}$. Then

$$T_s^* = \frac{N}{n}\sum_1^n \frac{M_i}{M_0}\left\{\bar{r}_i\bar{X}_i + \frac{(M_i - 1)m_i}{M_i(m_i - 1)}(\bar{y}_i - \bar{r}_i\bar{x}_i)\right\} \quad (3)$$

where $\bar{r}_i = \sum_1^{m_i} r_{ij}/m_i$. The term in the curly bracket is the Hartley-

Ross unbiased ratio estimator for the i-th primary. For the case $n = N$, (3) reduces to a 'separate' unbiased ratio estimator in stratified sampling.

Let us now consider the construction of 'combined' unbiased ratio and regression estimators in two-stage sampling. We obtain two different classes of estimators.

Class 1

From the n primaries in the sample choose any α primaries $(1 \leqslant \alpha \leqslant n)$. Let $Z_{\alpha \cdot \beta_i}$ denote the ordered set of observations on the first $\sum\limits_{1}^{\alpha} \beta_i$ secondaries sampled from these α primaries $(0 < \beta_i < m_i)$, and $a\,(Z_{\alpha,\beta_i})$ denotes a function of these observations. Then, for a fixed set of n primaries and a given Z_{α,β_i},

$$T_{c(1)} = \frac{N}{nM_0}\Bigg[\left(\sum_{1}^{\alpha} \frac{M_i - \beta_i}{m_i - \beta_i} \sum_{1}^{m_i-\beta_i} y_{ij} + \sum_{1}^{n-\alpha} M_i \bar{y}_i\right)$$

$$- a\,Z(\alpha,\beta_i)\left\{\sum_{1}^{\alpha} \frac{M_i - \beta_i}{m_i - \beta_i} \sum_{1}^{m_i-\beta_i} x_{ij} + \sum_{1}^{n-\alpha} M_i \bar{x}_i\right.$$

$$\left. - \left(\sum_{1}^{n} X_i - \sum_{1}^{\alpha} \sum_{1}^{\beta_i} x_{ij}\right)\right\} + \sum_{1}^{\alpha} \sum_{1}^{\beta_i} y_{ij}\Bigg] \qquad (4)$$

is an unbiased estimator of $(N/nM_0 \sum\limits_{1}^{n} Y_i)$, so that $T_{c\,(1)}$ is unbiased independent of Z_{α,β_i}. Therefore, the estimator $T^*_{c(1)}$ obtained from $T_{c(1)}$ by averaging over all permutations in the ordering of the β_i secondaries in the i-th primary and then by averaging over all possible choices of α primaries is an unbiased estimator of $N/nM_0 \sum\limits_{1}^{n} Y_i$. Since the average of $N/nM_0 \sum\limits_{1}^{n} Y_i$ over all possible samples of n primaries is $\bar{Y}$, it follows that $T^*_{c(1)}$ is an unbiased estimator of $\bar{Y}$. It may be noted that to construct $T^*_{c(1)}$ the sum of the X_i of the n primaries in the sample must be known.

Special cases.—(*a*) Let $\alpha = 1$, $\beta_i = 1$ and $a(Z_{,\beta_i}) = y_{ij}/x_{ij} = r_{ij}$. Then, averaging $T_{c(1)}$ over all permutations in the ordering of $\beta_i = 1$ sampled secondary, we get

$$T'_{c(1)} = \frac{N}{nM_0}\left\{\sum_1^n M_t\bar{y}_t + \bar{r}_i\left(\sum_1^n X_t - \sum_1^n M_t\bar{x}_t\right)\right.$$

$$\left. + \frac{M_i - m_i}{M_i - 1}(\bar{y}_i - \bar{r}_i\bar{x}_i)\right\}. \quad (5)$$

Now taking a weighted average of $T'_{c(1)}$ with weights $M_i / \sum_1^n M_i$ over all possible choices of $\alpha = 1$ sampled primary, we obtain

$$T^*_{c(1)} = \frac{N}{nM_0}\left\{\sum_1^n M_t\bar{y}_t + \frac{\sum_1^n M_t\bar{r}_t}{\sum_1^n M_t}\left(\sum_1^n X_t - \sum_1^n M_t\bar{x}_t\right)\right.$$

$$\left. + \sum_1^n \frac{(M_t - m_t)\, M_t}{(M_t - 1)\sum_1^n M_t}(\bar{y}_t - \bar{r}_t\bar{x}_t)\right\}. \quad (6)$$

For the special case of $x_{ij} = 1$ (for all i and j), $X_i = M_i$ and (6) reduces to

$$T^*_{c(1)} = \frac{N}{nM_0}\sum_1^n M_t\,\bar{y}_t$$

which is the usual unbiased estimator in two-stage sampling.

For the special case of stratified sampling, $n = N$ and (6) reduces to

$$T^*_{c(1)} = \bar{X}\bar{r}_{st} + (\bar{y}_{st} - \bar{x}_{st}\bar{r}_{st}) + \sum_1^N \frac{M_t^2}{M_0^2}\,\frac{M_t - m_t}{M_t(m_t - 1)}(\bar{y}_t - \bar{r}_t\bar{x}_t), \quad (7)$$

where $\bar{r}_{st} = \sum_1^N M_t\bar{r}_t/M_0$ and similar expressions for $\bar{y}_{st}$ and $\bar{x}_{st}$. Equation (7) is identical with the one given in Ross (1960).

(*b*) To find Williams' (1961) estimator in two-stage sampling, consider the special case $M_i = \bar{M}$ and $m_i = \bar{m}$ and let $\bar{m}/k$ be a positive integer. Then split at random the $\bar{m}$ sampled secondaries in each of the n selected primaries into k groups each of size $\bar{m}/k$. Now choose $\alpha = n$ and $\beta_i = \bar{m}/k$ and consider the estimator (4) with each of the k groups is turn as the conditional group. Then the mean of these k estimators reduces to

$$T''_{c(1)} = \bar{\bar{y}} + \bar{b}\left(\frac{1}{n\bar{M}}\sum_{1}^{n} X_i - \bar{\bar{x}}\right)$$

$$+ \left(1 - \frac{\bar{m}}{\bar{M}}\right)\frac{1}{k(k-1)}\sum_{1}^{k}(b^i - \bar{b})(\bar{\bar{x}}^i - \bar{\bar{x}}) \qquad (8)$$

where $\bar{\bar{y}}^i = (k/n\bar{m}) \sum_{1}^{n} \sum_{1}^{\bar{m}/k} y_{tj}$ is the mean for the i-th group, $\bar{\bar{y}} = \sum_{1}^{k} \bar{\bar{y}}^i/k$ is the overall semple mean, $b^i = a\left(Z_{n,\bar{m}/k}\right)$ and $\bar{b} = \sum_{1}^{k} b^i/k$ and similar expressions for 'x'. Equation (8) is identical with equation (11) of Williams (1961). The choice of b^i is arbitrary. For example, a regression-type estimator is obtained from (8) by choosing

$$b^i = \frac{\sum_{1}^{n} \sum_{1}^{\bar{m}/k} (y_{tj} - \bar{y}_t^{\,i})(x_{tj} - \bar{x}_t^{\,i})}{\sum_{1}^{n} \sum_{1}^{\bar{m}/k} (x_{tj} - \bar{x}_t^{\,i})^2}$$

where $\bar{y}_t^{\,i} = k \sum_{1}^{\bar{m}/k} y_{tj}/\bar{m}$ and a similar expression for $\bar{x}_t^{\,i}$.

It may be noted that the unbiased ratio-type estimator (6) is based on the ratios y_{ij}/x_j at the secondary level, whereas to reduce the variation among the primaries it seems desirable to construct estimators based on the ratio $\bar{y}_i/\bar{x}_i$ at the primary level. Since β_i is less than m_i for constructing $T^*_{c(1)}$, it is not possible to obtain estimators based on the primary ratios $\bar{y}_i/\bar{x}_i$ from $T^*_{c(1)}$. However, if m_i is moderately large, one can obtain estimators from $T^*_{c(1)}$ based on the ratios $\bar{y}_i/\bar{x}_i$ (approximately) by taking $\beta_i = m_i - 1$. But this is not quite adequate since in many sample surveys the tendency is to keep m_i small. Therefore, by using a different argument, we shall now develop estimators based on the ratios $\bar{y}_i/\bar{x}_i$.

From the n sampled primaries choose any α primaries and let Z_{α,m_i} denote the set of observations on the $\sum_1^{\alpha} m_i$ secondaries sampled from these α primaries, and $a(Z_{\alpha,m_i})$ denotes a function of these observations. Then, for a fixed set of n primaries and a given Z_{α,m_i},

$$\tilde{T}_{c(1)} = \frac{N}{nM_0}\left\{\sum_1^{n-\alpha} M_i\bar{y}_i - a(Z_{\alpha,m_i})\left(\sum_1^{n-\alpha} M_i\bar{x}_i - \sum_1^{n-\alpha} X_i\right)\right\} \tag{9}$$

is an unbiased estimator of $N\left(\sum_1^n Y_i - \sum_1^{\alpha} Y_i\right)\Big/(nM_0)$. Now, averaging (9) over all possible choices of α sampled primaries, we obtain. $\tilde{T}^*_{c(1)}$ which is an unbiased estimator of $N/nM_0\,(n-\alpha)/n \sum_1^n Y_i$. Therefore, $n/(n-\alpha)\,\tilde{T}^*_{c(1)}$ is an unbiased estimator of $\bar{Y}$ when averaged over all possible samples of n primaries. As a particular case, let $\alpha = 1$ and $a(Z_{\alpha,m_i}) = \bar{y}_i/\bar{x}_i$. Then

$$\frac{n}{n-1}\,\tilde{T}^*_{c(1)} = \frac{n}{(n-1)}\,\frac{N}{nM_0}\left\{\sum_1^n M_i\bar{y}_i + \frac{1}{n}\sum_1^n \frac{\bar{y}_i}{\bar{x}_i} \times\left(\sum_1^n X_i - \sum_1^n M_i\bar{x}_i\right) - \frac{1}{n}\sum_1^n \frac{\bar{y}_i}{\bar{x}_i}X_i\right\} \tag{10}$$

is an unbiased estimator of $\bar{Y}$ and is based on the primary ratios $\bar{y}_i/\bar{x}_i$. It seems that the estimator (10) should be quite useful in practice, provided the population totals X_i of the n primaries in the sample are known.

Class 2

We shall now construct a class of estimators, using a different argument, which depends only on the overall population mean $\bar{X}$. Let Z'_{α,β_i}, $(1 \leqslant \alpha < n,\ 1 \leqslant \beta_i < m_i)$, denote the ordered set of observations on the first $\sum_1^{\alpha} \beta_i$ sampled secondaries obtained from the first α sampled primaries and $a(Z'_{\alpha,\beta_i})$ denotes a function of these observations. Then, for a given Z'_{α,β_i},

$$\frac{N-\alpha}{n-\alpha}\sum_1^{n-\alpha} M_i\bar{y}_i + \sum_1^{\alpha}\frac{M_i-\beta_i}{m_i-\beta_i}\sum_1^{m_i-\beta_i} y_{ij} \tag{11}$$

is an unbiased estimator of $M_0\bar{Y} - \sum_1^{a}\sum_1^{\beta_i} y_{ij}$. Similarly, the estimator obtained from (11) by replacing y_{ij} by x_{ij} is an unbiased estimator of

$$M_0\bar{X} - \sum_1^{a}\sum_1^{\beta_i} x_{ij}.$$

Therefore, for a given Z'_{a,β_i},

$$T_{c(2)} = \frac{1}{M_0}\left\{\frac{N-a}{n-a}\sum_1^{n-a} M_i\bar{y}_i + \sum_1^{a}\frac{M_i-\beta_i}{m_i-\beta_i}\sum_1^{m_i-\beta_i} y_{ij} - a\,(Z'_{a,\beta_i})\left(\frac{N-a}{n-a}\sum_1^{n-a} M_i\bar{x}_i + \sum_1^{a}\frac{M_i-\beta}{m_i-\beta_i}\sum_1^{m_i-\beta_i} x_{ij}\right) + \sum_1^{a}\sum_1^{\beta_i} y_{ij}\right\} \tag{12}$$

is an unbiased estimator of $\bar{Y}$, and hence $T_{c(2)}$ is unconditionally unbiased. Equation (12) can be written as

$$T_{c(2)} = \left\{\frac{N-a}{N(n-a)}\bar{y}_c - \frac{N-a}{r-a}\sum_1^{a}\frac{M_i\bar{y}_i}{M_0} + \sum_1^{a}\frac{(M_i-\beta_i)\,m_i}{M_0(m_i-\beta_i)}\bar{y}_i - \sum_1^{a}\frac{(M_i-\beta)\,\beta_i}{M_0(m_i-\beta_i)}\bar{y}_i(\beta_i)\right\} - a\,(Z'_{a,\beta_i})$$
$$\times\left\{\frac{N-a}{N(n-a)}\bar{x}_c - \frac{N-a}{n-a}\sum_1^{a}\frac{M_i\bar{x}_i}{M_0} + \sum_1^{a}\frac{(M_i-\beta_i)\,m_i}{M_0(m_i-\beta_i)}\bar{x}_i - \sum_1^{a}\frac{(M_i-\beta_i)\,\beta_i}{M_0(m_i-\beta_i)}\bar{x}_i(\beta_i) - \bar{X} + \frac{1}{M_0}\sum_1^{a}\sum_1^{\beta_i} x_{ij}\right\} \tag{13}$$

where $\bar{y}_c = N\sum_1^{n} M_i\bar{y}_i/(nM_0)$ and a similar expression for $\bar{x}_c$. One can now obtain $T^*_{c(2)}$ from $T_{c(2)}$ by averaging over all permutations in the ordering of the first β_i sampled secondaries from the i-th primary ($i = 1, \cdots, \alpha$) and then by averaging over all permutations in the ordering of the first α sampled primaries.

Special cases.—(*a*) Let $\alpha = 1$, $\beta_i = 1$ and $a(Z'_{\alpha,\beta_i}) = y_{ij}/x_{ij}$. After the averaging process we obtain, from (13),

$$T^*_{c(2)} = \bar{X}\bar{r} + \frac{(N-1)\,n}{N(n-1)}(\bar{y}_c - \bar{r}\bar{x}_c)$$

$$+ \frac{1}{n}\sum_1^n \frac{M_i}{M_0}\left\{\frac{(M_i - 1)\,m_i}{M_i(m_i - 1)} - \frac{N-1}{n-1}\right\}(\bar{y}_i - \bar{r}_i\bar{x}_i). \quad (14)$$

For the case of $x_{ij} = 1$ (for all i and j), (14) reduces to

$$T^*_{c(2)} = \frac{1}{n}\sum_1^n \bar{y}_i + \frac{(N-1)\,n}{N(n-1)}$$

$$\times\left\{\bar{y}_c - \frac{N}{nM_0}\left(\frac{1}{n}\sum_1^n \bar{y}_i\right)\left(\sum_1^n M_i\right)\right\}. \quad (15)$$

The second term on the right-hand side of (15) is the correction for bias in the usual biased estimator $\sum_1^n \bar{y}_i/n$. For the particular case of $M_i = \bar{M}$ and $m_i = \bar{m}$, (14) reduces to

$$T^*_{c(2)} = \bar{X}\bar{r} + \frac{(N-1)\,n}{N(n-1)}(\bar{\bar{y}} - \bar{r}\bar{\bar{x}}) + \frac{1}{nN}\left\{\frac{(M-1)\,m}{M(m-1)} - \frac{N-1}{n-1}\right\}$$

$$\times\left(n\bar{\bar{y}} - \sum_1^n \bar{r}_i\bar{x}_i\right). \quad (16)$$

Equation (16) is identical with the one given in Ross (1960). For the case of stratified sampling $n = N$ and taking a weighted average over the primaries with weights M_i/M_0 instead of a simple average, we obtain an estimator identical with the class 1 estimator derived earlier [equation (7)].

The estimator (14) is based on the secondary ratios y_{ij}/x_{ij}. As mentioned earlier, it is desirable to construct estimators based on the primary ratios $\bar{y}_i/\bar{x}_i$ instead of the secondary ratios. In class 1, we had constructed the estimator (10) based on the ratios $\bar{y}_i/\bar{x}_i$ but depending on the population totals X_i of the n primaries in the sample. It does not seem to be possible to construct estimators, based on the primary ratios, depending only on the overall population mean $\bar{X}$. This is seen from the following considerations.

Let Z'_{a,m_i} denote the set of observations on the $\sum_{1}^{a} m_i$ sampled secondaries obtained from the first a sampled primaries and $a(Z'_{a,m_i})$ denotes a function of these observations. Then, for a given Z'_{a,m_i},

$$\tilde{T}_{c(2)} = \frac{1}{M_0}\left[\frac{N-a}{n-a}\sum_{1}^{n-a} M_i \bar{y}_i - a(Z'_{a,m_i})\right.$$

$$\left.\times\left\{\frac{N-a}{n-a}\sum_{1}^{n-a} M_i \ \bar{x}_i - \left(M_0 X - \sum_{1}^{a} X_i\right)\right\}\right] \quad (17)$$

is an unbiased estimator of $\bar{Y} - \sum_{1}^{a} Y_i/M_0$. Now, taking the expectation over Z'_{a,m_i}, we find that $\tilde{T}_{c(2)}$ is an unbiased estimator of $(N-a)\bar{Y}/N$. Therefore, the estimator $\tilde{T}^*_{c(2)}$ obtained by averaging $\tilde{T}_{c(2)}$ over all permutations in the ordering of the first a sampled primaries is an unbiased estimator of $(N-a)\bar{Y}/N$, so that $N\tilde{T}^*_{c(2)}/(N-a)$ is an unbiased estimator of $\bar{Y}$. For the special case of $a=1$ and $a(Z'_{a,mi}) = \bar{y}_i/\bar{x}_i$,

$$\frac{N}{N-1}\tilde{T}^*_{c(2)} = \frac{N}{n(N-1)}\bar{X}\sum_{1}^{n}\frac{\bar{y}_t}{\bar{x}_t} + \frac{N}{(n-1)M_0}$$

$$\times\left\{\sum_{1}^{n} M_t\bar{y}_t - \frac{1}{n}\left(\sum_{1}^{n} M_t\bar{x}_t\right)\left(\sum_{1}^{n}\frac{\bar{y}_t}{\bar{x}_t}\right)\right\}$$

$$-\frac{N}{n(N-1)}\frac{1}{M_0}\sum_{1}^{n} X_t\frac{\bar{y}_t}{\bar{x}_t} \quad (18)$$

The estimator (18) depends on the population totals X_i of the n primaries in the sample. However, it may be noted that (18) is not the same as the class 1 estimator $n\tilde{T}^*_{c(1)}/(n-1)$ except when $n = N$, *i.e.*, the case of stratified sampling. In this connection it may be interesting to note that it is possible to construct unbiased ratio estimators based on the ratios $\bar{y}_i/\bar{x}_i$ and depending only on $\bar{X}$ provided the primaries are sampled *with* replacement. Ross (1960) has shown that, if the primaries are sampled *with* replacement,

$$\frac{\bar{X}}{n}\sum_1^n \frac{\bar{y}_i}{\bar{x}_i} + \frac{N}{(n-1)M_0} \times \left\{\sum_1^n M_i\bar{y}_i - \frac{1}{n}\left(\sum_1^n M_i\bar{x}_i\right)\left(\sum_1^n \frac{\bar{y}_i}{\bar{x}_i}\right)\right\} \tag{19}$$

is an unbiased estimator of $\bar{Y}$. Note that (19) is based on the primary ratios and depends only on $\bar{X}$.

If the primaries are sampled without replacement and if the m_i are moderately large, we can, however, construct an estimator from (13), based on the primary ratios $\bar{y}_i/\bar{x}_i$ (approximately) and depending only on $\bar{X}$, by taking $\alpha = 1$, $\beta_i = m_i - 1$. This is considered below.

(*b*). Let $\alpha = 1$, $\beta_i = m_i - 1$ and $a\left(Z'_{\alpha,\beta_i}\right) = (m_i\bar{y}_i - y_{ij})/(m_i\bar{x}_i - x_{ij})$. If m_i is not small, we can write $a\left(Z'_{\alpha,\beta_i}\right) \doteqdot \bar{y}_i/\bar{x}_i$. Now, from (13), after the averaging process we obtain

$$T^*_{c(2)} = \frac{\bar{X}}{n}\sum_1^n \bar{R}^i_{(m_i-1)} + \frac{(N-1)n}{N(n-1)} \times \left\{\bar{y}_c - \frac{1}{n}\bar{x}_c\left(\sum_1^n \bar{R}^i_{(m_i-1)}\right)\right\} + \frac{1}{n}\sum_1^n \frac{M_i}{M_0} \times \left\{\frac{(M_i - m_i + 1)m_i}{M_i} - \frac{N-1}{n-1}\right\}(\bar{y}_i - \bar{R}^i_{(m_i-1)}\bar{x}_i) \tag{20}$$

where

$$\bar{R}^i_{(m_i-1)} = \frac{1}{m_i}\sum_1^{m_i} \frac{m_i\bar{y}_i - y_{ij}}{m_i\bar{x}_i - x_{ij}}.$$

For the case $x_{ij} = 1$, (20) reduces to (15). In the special case of stratified sampling, taking a weighted average of the primaries with weights M_i/M_0 instead of a simple average, we obtain

$$T^*_{c(2)} = \bar{X} \sum_{1}^{N} \frac{M_i}{M_0} \bar{R}^i_{(m_i-1)} + \left\{ \bar{y}_{st} - \left(\sum_{1}^{N} \frac{M_i}{M_0} \bar{R}^i_{(m_i-1)} \right) \bar{x}_{st} \right\}$$

$$+ \sum_{1}^{N} \frac{M_i^2}{M_0^2} \frac{(M_i - m_i)(m_i - 1)}{M_i} (\bar{y}_i - \bar{R}^i_{(m_i-1)} \bar{x}_i). \quad (21)$$

(c). We shall now develop a Williams' type estimator from (13) which depends only on $\bar{X}$. Consider again the case $M_i = \bar{M}$ and $m_i = \bar{m}$. The $\bar{m}$ sampled secondaries in each of the n sampled primaries are split at random into k groups each of size $\bar{m}/k$. Also, the n sampled primaries are split at random into r groups each of size n/r (assuming $\bar{m}/k$ and n/r to be positive integers). Choose now $\alpha = n/r$ and $\beta_i = \bar{m}/k$ in (13), and let $a\left(Z_{\alpha,\beta_i}\right) = b^{tp}$, $(t = 1, \cdots, r;\ p = 1, \cdots, k)$, denote a function of the set of observations on the p-th group of secondaries in the t-th group of primaries. For example

$$b^{tp} = \frac{\sum_{1}^{n/r} \sum_{1}^{\bar{m}/k} (y_{ij} - \bar{y}_i^p)(x_{ij} - \bar{x}_i^p)}{\sum_{1}^{n/r} \sum_{1}^{\bar{m}/k} (x_{ij} - \bar{x}_i^p)^2},$$

where $\bar{y}_i^p = k \sum_{1}^{\bar{m}/k} y_{ij}/\bar{m}$ and a similar expression for $\bar{x}_i^p$. Consider now the estimator (13) with the t-th group of primaries and with each of the k groups of secondaries in these primaries as the conditional group, and take the mean of these k estimators. Then consider each of the r groups of primaries as the conditional group and take the mean of these r estimators. After some simplification, we obtain

$$T''_{e2} = \bar{\bar{y}} + \bar{\bar{b}}(\bar{X} - \bar{\bar{x}}) + \left(1 - \frac{n}{N}\right)$$

$$\times \frac{1}{r(r-1)} \sum_{1}^{r} (\bar{b}^t - \bar{\bar{b}})(\bar{\bar{x}}^t - \bar{\bar{x}}) + \frac{1}{N}\left(1 - \frac{\bar{m}}{\bar{M}}\right)$$

$$\times \frac{1}{rk(k-1)} \sum_{1}^{r} \sum_{1}^{k} (\bar{x}^{tp} - \bar{\bar{x}}^t)(b^{tp} - \bar{b}^t) \quad (22)$$

where

$$\bar{x}^{tp} = rk \sum_{1}^{n/r} \sum_{1}^{m/k} \frac{x_{tj}}{(n\bar{m})}, \quad \bar{\bar{x}}^{t} = \sum_{1}^{k} \frac{\bar{x}^{tp}}{k}, \quad \bar{b}^{t} = \sum_{1}^{k} \frac{b^{tp}}{k}$$

and

$$\bar{\bar{b}} = \sum_{1}^{r} \frac{\bar{b}^{t}}{r}.$$

One can, of course, obtain the estimator (22) directly using Williams' procedure for constructing unbiased ratio and regression estimators.

III. Variance Estimation

As mentioned earlier in Section I, the direct evaluation of the estimator of variance by estimating each term in the variance formula is complex and, in any case, the resulting expression is useless in practice since it is very unwieldy. Mickey (1959) has suggested an alternative method of variance estimation for a sub-class of his estimators (which seems to be fairly manageable). It is possible to extend his method to multi-stage sampling. However, we shall not discuss this extension here and, instead, apply a simpler method of variance estimation suggested by Goodman and Hartley (1958). This method gives an unbiased estimator of variance, but this estimator of variance will usually be affected by a large sampling error.

With this method it is convenient to modify the sampling scheme as follows: Suppose that $n = 2n'$ is even and $n' \geq 2$. Draw a first random sample of n' primaries without replacement and, if the i-th primery is selected in the sample, take a sample of m_i secondaries at random from the i-th primary. Then construct an unbiased ratio or regression estimator (either of class 1 or 2) from this sample. Replace the whole sample and draw a second sample by the same procedure and construct an estimator belonging to class 1 or 2. If the second sample is identical with the first, it must be rejected and a new sample is drawn. The total number of possible samples is,

$$S = \sum^{\binom{N}{n'}} \prod_{1}^{n'} \binom{M_i}{m_i}.$$

When $m_i = \bar{m}_i$ and $M_i = \bar{M}$, S reduces to $\binom{N}{n'}\binom{\bar{M}}{\bar{m}}^{n'}$. If $T_1{}^*$ and

T_2^* denote the unbiased estimators of $\bar{Y}$ from the two samples, then $T^* = (T_1^* + T_2^*)/2$ is an unbiased estimator of $\bar{Y}$ and an unbiased estimator of variance is given by

$$v(T^*) = \frac{1}{4} \cdot \frac{(S-2)}{S} (T_1^* - T_2^*)^2. \tag{23}$$

In most cases S is large, so that $(S-2)/S$ can be replaced by 1. The estimator (22) is based only on a "single degree of freedom". If more "degrees of freedom" are desired, one can draw k samples each of size n/k, $(k > 2)$.

The author wishes to thank Dr. Alan Ross for helpful comments.

References

Goodman, L. A. and Hartley, H. O. .. "The precision of unbiased ratio-type estimators," *Journal of the American Statistical Association*, 1958, **53**, 491-508.

Mickey, M. R. .. "Some finite population unbiased ratio and regression estimators," *Ibid.*, 1959, **54**, 594–612.

Nieto, J. .. "Unbiased ratio estimators in stratified sampling," *Ibid.*, 1961, **56**, 70–87.

Ross, A. .. "On two problems in sampling theory: Unbiased ratio estimators and variance estimates in optimum sampling designs," *Unpublished Ph.D. Thesis*, Iowa State University Library, Ames, Iowa, 1960.

Williams, W. H. .. "Generating unbiased ratio and regression estimators," *Biometrics*, 1961, **17**, 267–74.

———— .. "On two methods of unbiased estimation with auxiliary variates," *Journal of the American Statistical Association*, 1962, **57**, 184–86.