1964

Sample Selection and the Choice of Estimator in Two-Way Stratified Populations

William (Bill) H. Williams
CUNY Hunter College

SAMPLE SELECTION AND THE CHOICE OF ESTIMATOR IN TWO-WAY STRATIFIED POPULATIONS

BY

W. H. WILLIAMS

Reprinted from the JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
December 1964, Vol. 59
pp. 1054–1062

SAMPLE SELECTION AND THE CHOICE OF ESTIMATOR IN TWO-WAY STRATIFIED POPULATIONS

W. H. WILLIAMS

Bell Telephone Laboratories, Inc.

The paper studies the effect on variance of the method of sample selection and the choice of estimator in a two-way stratified population. Three methods of sampling, simple random and one and two-way stratified, are studied in conjunction with estimators which use no population information, one and two sets of marginal frequencies and cell frequencies. It is shown that whether the information in the population frequencies is used in the sample draw or in the estimator, the sampling variances are approximately the same. Hence efficiency can be associated with the method of sample selection and the estimator by the amount of population information used by each.

Some of the estimators involve post-stratified weighting and hence are subject to biases and inflated variances by the occurrence of small (or zero) sample sizes in the strata; in addition, the comparisons made are unequally affected by small sample sizes, but they are valid for sample sizes large enough to make the ratio estimators become stable.

Data from a real two-way stratified population are used for illustration.

1. INTRODUCTION

Good precision may be derived from either, or both, of the sampling procedure and the form of estimator. Thus it is of interest to ask whether information available in advance of sampling is best used in the sample draw, or at the later stage of the formation of the estimator. What procedures are likely to yield high precision and what other factors, if any, are involved? This paper studies these questions in connection with the estimation of the mean of a two-way stratified population.

2. DESCRIPTION OF THE ESTIMATORS

The population is stratified in two directions with ${}_iM_t$ units in the (i, t)th cell; $i=1, 2, \cdots L$; $t=1, 2, \cdots N$; and by definition,

$$\sum_{i=1}^{L} {}_iM_t = M_t; \qquad \sum_{t=1}^{N} {}_iM_t = {}_iM; \qquad \sum_{i=1}^{L}\sum_{t=1}^{N} {}_iM_t = M; \tag{2.1}$$

also ${}_iY_t$ is the total of the ${}_iM_t$ values and analogously,

$$\sum_{i=1}^{L} {}_iY_t = Y_t; \qquad \sum_{t=1}^{N} {}_iY_t = {}_iY; \qquad \sum_{i=1}^{L}\sum_{t=1}^{N} {}_iY_t = Y. \tag{2.2}$$

Lower case letters denote the corresponding sample values, and bars over letters are used to denote means. When a single direction of stratification is discussed the marginal notation given above will be used to describe it.

A distinction among certain stratified estimators may be made according to the amount of information which each estimator utilizes. More specifically, estimators which use, one set of marginal frequencies, two sets of marginal frequencies, and cell frequencies are referred to as marginally weighted, bi-marginally weighted, and cell weighted estimators, respectively. Three methods of sample selection are discussed, simple random, one-way stratified and cell stratified sampling. Table 1 gives notation for reference to the methods of drawing the sample in combination with the choices of estimator. Two combinations in Table 1 are not considered because they suggest using less information than was used in the sample draw and this is not consistent with the objectives described in Section 1.

TABLE 1. NOTATION FOR REFERENCE TO SAMPLING SCHEMES

		Sample Selection		
		Completely Random (A)	One-Way Stratified (B)	Cell Stratified (C)
Type of Estimator	Marginally Weighted (I)	C(A, I)	C(B, I)	Not Considered
	Bi-Marginally Weighted (II)	C(A, II)	C(B, II)	Not Considered
	Cell Weighted (III)	C(A, III)	C(B, III)	C(C, III)

To discuss some of the classifications of Table 1 from the point of view of someone considering their relative merits in advance of sampling, it is necessary to introduce post-stratified weighting. For example, in simple random sampling none of the two-way frequency information is used in the selection process and so, in a sense, the simple mean, $\bar{y}$, is the natural estimator. However introducing post-stratification and the marginal frequencies, combination C(A, I)[1] can be obtained. Similarly, post-stratification with the bi-marginal frequencies yields C(A, II), and with the cell frequencies yields C(A, III). The sampling variance of the estimators with post-stratified weighting can be found using a method described by Williams [3].

3. DERIVATION OF THE ESTIMATORS

3a. Simple Random Sampling

Simple random sampling uses no marginal population information and the sampling variance of the mean, $\bar{y}$, is well known. If the marginal frequencies, M_t, are introduced, the post-stratified estimator in C(A, I) is

$$\hat{y}_{A,I} = \sum_{t=1}^{N} \frac{M_t}{M} \bar{y}_t \tag{3.1}$$

[1] Hereafter the combinations will be referred to simply as C(A, I). This should be read as "the combination of sample selection procedure A and estimator weighting I."

with

$$V(\hat{y}_{\mathrm{A,I}}) = \left(1 - \frac{m}{M}\right)\frac{1}{m}\,\frac{1}{M-1}\sum_{t=1}^{N}\sum_{j=1}^{M_t}(y_{tj} - \overline{Y}_t)^2. \tag{3.2}$$

This sampling variance is approximately equal to the conditional variance given a fixed number of proportionally allocated observations in each stratum.

C(A, II) requires the use of both sets of marginal totals, ${}_iM$ and M_t for all i and t. First, post-stratification is applied with t as the post strata index giving, in Equation (3.3), an estimator of the total in the ith stratum,

$$_i\hat{Y} = \sum_{t=1}^{N} M_t\,\frac{_iy_t}{m_t}\,. \tag{3.3}$$

Similarly, Equation (3.4) gives an estimator of the total number in the ith stratum

$$_i\hat{M} = \sum_{t=1}^{N} M_t\,\frac{_im_t}{m_t}\,. \tag{3.4}$$

Then by weighting the ratio, ${}_i\hat{Y}/{}_i\hat{M}$, by the known weights ${}_iM/M$, $\hat{y}_{\mathrm{A,II}}$ is obtained as an estimator of the over-all population mean;

$$\hat{y}_{\mathrm{A,II}} = \sum_{i=1}^{L}\frac{_iM}{M}\cdot\frac{_i\hat{Y}}{_i\hat{M}}\,. \tag{3.5}$$

Furthermore,

$$V(\hat{y}_{\mathrm{A,II}}) = \left(1 - \frac{m}{M}\right)\frac{1}{m}\,\frac{1}{M-1}$$

$$\cdot\left\{\sum_{i=1}^{L}\sum_{t=1}^{N}\sum_{j=1}^{_iM_t}({}_iy_{tj} - {}_i\overline{Y}_t)^2 + \sum_{i=1}^{L}\sum_{t=1}^{N}{}_iM_t({}_i\overline{Y}_t - \overline{Y}_i - \overline{Y}_t + \overline{\overline{Y}}_t)^2\right\},$$

where

$$\overline{\overline{Y}}t = \sum_{i=1}^{L}\frac{_iM_t}{M_t}\,{}_i\overline{Y}. \tag{3.6}$$

C(A, III) uses a simple random sample and the cells frequencies, ${}_iM_t$, as post strata. This is analytically equivalent to one-way stratification and the formulas are easily spelled out.

3b. Stratified Sampling

Stratified sampling in one direction, involves one set of marginal totals, M_t, and an estimator which utilizes this information is the usual stratified estimator. It is the "natural" estimator to use in the same sense that the simple mean, $\bar{y}$, is the natural estimator for simple random sampling.

The second type B scheme uses both the marginal totals in the estimator,

$$\hat{y}_{B,\mathrm{II}} = \sum_{i=1}^{L}\frac{_iM}{M}\,{}_i\hat{y}, \tag{3.7}$$

where ${}_i\hat{y}$ is a ratio estimator of the mean of the ith post-stratum obtained from the original stratified sample. The sampling variance of $\hat{y}_{B,II}$ can be written as

$$V(\hat{y}_{B,II}) = \sum_{t=1}^{N} \frac{M_t^2}{M^2}\left(1 - \frac{m_t}{M_t}\right)\frac{1}{m_t}\frac{1}{M_t - 1}$$
$$\cdot\left\{\sum_{i=1}^{L}\sum_{j=1}^{{}_iM_t} ({}_iy_{tj} - {}_i\overline{Y}_t)^2 + \sum_{i=1}^{L} {}_iM_t({}_i\overline{Y}_t - {}_i\overline{Y} - \overline{Y}_t + \overline{\overline{Y}}_t)^2\right\}. \quad (3.8)$$

C(B, III) uses a cell weighted estimator obtained by post-stratifying the tth stratum separately to give

$$\hat{y}_t = \sum_{i=1}^{L} \frac{{}_iM_t}{M_t}\, {}_i\hat{y}_t \quad (3.9)$$

as an estimator of the mean of the tth original stratum. Then

$$\hat{y}_{B,III} = \sum_{t=1}^{N} \frac{M_t}{M}\hat{y}_t \quad (3.10)$$

is an estimator of the population mean and its variance is given by

$$V(\hat{y}_{B,III}) = \sum_{t=1}^{N} \frac{M_t^2}{M^2}\sum_{i=1}^{L}\left(1 - \frac{m_t}{M_t}\right)\frac{{}_iM_t - 1}{M_t}\,\frac{1}{m_t}\,\frac{1}{{}_iM_t - 1}\sum_{j=1}^{{}_iM_t} ({}_iy_{tj} - {}_i\overline{Y}t)^2. \quad (3.11)$$

C(C, III) is a two-way stratification with the cell weighted estimator,

$$\hat{y}_{C,III} = \sum_{i=1}^{L}\sum_{t=1}^{N} \frac{{}_iM_t}{M}\, {}_i\bar{y}_t, \quad (3.12)$$

which has a sampling variance given by,

$$V(\hat{y}_{C,III}) = \sum_{i=1}^{L}\sum_{t=1}^{N} \frac{{}_iM_t^2}{M^2}\left(1 - \frac{{}_im_t}{{}_iM_t}\right)\frac{1}{{}_im_t}\,\frac{1}{{}_iM_t - 1}\sum_{j=1}^{{}_iM_t} ({}_iy_{tj} - {}_i\overline{Y}_t)^2. \quad (3.13)$$

Variance estimation is not discussed but follows in a straightforward manner from Williams [3].

For convenient reference in Section 4, the results of this section have been summarized in Table 2. For ease in comparison, some of the formulas have been written in slightly different forms.

4. DISCUSSION OF THE ESTIMATORS

In Table 2, $V(\hat{y}_{A,I})$ has been rewritten in a form which shows clearly that $V(\hat{y}_{A,I})$ is smallest when the variance among cells within strata is a minimum. Also, the relative precision of $V(\hat{y}_{A,I})$ and $V(\hat{y}_{A,II})$ depends upon the relative magnitudes of the last term on the right side of each of $V(\hat{y}_{A,II})$ and $V(\hat{y}_{A,I})$, so that, if the one-way stratification has been reasonably successful in making

TABLE 2. SUMMARY OF ESTIMATORS AND SAMPLING VARIANCES

Type of Estimator	Sample Selection: Completely Random	Sample Selection: One-Way Stratified	Sample Selection: Cell Stratified
Marginally Weighted (1)	$\hat{y}_{A.I} = \sum_{t=1}^{N} \frac{M_t}{M} \bar{y}_t$ $V(\hat{y}_{A.I}) = \left(1 - \frac{m}{M}\right) \frac{1}{m} \frac{1}{M-1}$ $\left\{ \sum_{i=1}^{L} \sum_{t=1}^{N} \sum_{j=1}^{{}_iM_t} ({}_iy_{tj} - {}_i\bar{Y}_t)^2 \right.$ $\left. + \sum_{i=1}^{L} \sum_{t=1}^{N} {}_iM_t({}_i\bar{Y}_t - \bar{Y}_t)^2 \right\}$	$\hat{y}_{B.I} = \sum_{t=1}^{N} \frac{M_t}{M} \hat{y}_t$ $V(\hat{y}_{B.I}) = \sum_{t=1}^{N} \frac{M_t^2}{M^2} \left(1 - \frac{m_t}{M_t}\right)$ $\frac{1}{m_t} \frac{1}{M_t - 1} \left\{ \sum_{i=1}^{L} \sum_{j=1}^{{}_iM_t} ({}_iy_{tj} - {}_i\bar{Y}_t)^2 \right.$ $\left. + \sum_{i=1}^{L} {}_iM_t({}_i\bar{Y}_t - \bar{Y}_t)^2 \right\}$	Not Considered
Bi-Marginally Weighted (2)	$\hat{y}_{A.II} = \sum_{i=1}^{L} \frac{{}_iM}{M} \cdot \frac{{}_i\hat{Y}}{{}_i\hat{M}}$ $V(\hat{y}_{A.II}) = \left(1 - \frac{m}{M}\right) \frac{1}{m} \frac{1}{M-1}$ $\left\{ \sum_{i=1}^{L} \sum_{t=1}^{N} \sum_{j=1}^{{}_iM_t} ({}_iy_{tj} - {}_i\bar{Y}_t)^2 \right.$ $\left. + \sum_{i=1}^{L} \sum_{t=1}^{N} {}_iM_t({}_i\bar{Y}_t - {}_i\bar{Y} - \bar{Y}_t + \bar{\bar{Y}}_t)^2 \right\}$	$\hat{y}_{B.II} = \sum_{i=1}^{L} \frac{{}_iM}{M} {}_i\hat{y}$ $V(\hat{y}_{B.II}) = \sum_{t=1}^{N} \frac{M_t^2}{M^2} \left(1 - \frac{m_t}{M_t}\right)$ $\frac{1}{m_t} \frac{1}{M_t - 1} \left\{ \sum_{i=1}^{L} \sum_{j=1}^{{}_iM_t} ({}_iy_{tj} - {}_i\bar{Y}_t)^2 \right.$ $\left. + \sum_{i=1}^{L} {}_iM_t({}_i\bar{Y}_t - {}_i\bar{Y} - \bar{Y}_t + \bar{\bar{Y}}_t)^2 \right\}$	Not Considered
Cell Weighted (3)	$\hat{y}_{A.III} = \sum_{i=1}^{L} \sum_{t=1}^{N} \frac{{}_iM_t}{M} {}_i\bar{y}_t$ $V(\hat{y}_{A.III}) = \left(1 - \frac{m}{M}\right) \frac{1}{m} \frac{1}{M-1}$ $\sum_{i=1}^{L} \sum_{t=j}^{N} \sum_{j=1}^{{}_iM_t} ({}_iy_{tj} - {}_i\bar{Y}_t)^2$	$\hat{y}_{B.III} = \sum_{t=1}^{N} \frac{M_t}{M} \hat{y}_t$ $V(\hat{y}_{B.III}) = \sum_{t=1}^{N} \frac{M_t^2}{M^2} \sum_{i=1}^{L} \left(1 - \frac{m_t}{M_t}\right)$ $\frac{{}_iM_t - 1}{M_t - 1} \frac{1}{m_t} \frac{1}{{}_iM_t - 1} \sum_{j=1}^{{}_iM_t} ({}_iy_{tj} - {}_i\bar{Y}_t)^2$	$\hat{y}_{C.III} = \sum_{i=1}^{L} \sum_{t=1}^{N} \frac{{}_iM_t}{M} {}_i\bar{y}_t$ $V(\hat{y}_{C.III}) = \sum_{i=1}^{L} \sum_{t=1}^{N} \frac{{}_iM_t^2}{M^2} \left(1 - \frac{{}_im_t}{{}_iM_t}\right)$ $\frac{1}{{}_im_t} \frac{1}{{}_iM_t - 1} \sum_{j=1}^{{}_iM_t} ({}_iy_{tj} - {}_i\bar{Y}_t)^2$

Notation: $({}_iM_t - 1)\,{}_iS_t^2 = \sum_{j=1}^{{}_iM_t} ({}_iy_{tj} - {}_i\bar{Y}_t)^2$; $(M_t - 1) S_t^2 = \sum_{i=1}^{L} \sum_{j=1}^{{}_iM_t} ({}_iy_{tj} - \bar{Y}_t)^2$; $({}_iM - 1)\,{}_iS^2 = \sum_{t=1}^{N} \sum_{j=1}^{{}_iM_t} ({}_iy_{tj} - {}_i\bar{Y})^2$; $\bar{\bar{Y}}_t = \sum_{i=1}^{L} \frac{{}_iM_t}{M_t} {}_i\bar{Y}$.

homogeneous strata and the "interaction" in $V(\bar{y}_{A,II})$ is large than $C(A, II)$ may have larger variance than $C(A, I)$ in spite of the fact that both the ${}_iM$'s and M_t's have been used. When all ${}_i\bar{Y}$ are equal, $\bar{\bar{Y}}_t = {}_i\bar{Y}$, the two variance formulas are equal and there is no gain from the second direction of stratification. In general however one would expect that the second direction of stratification would contribute to a reduction in variance and that $V(\bar{y}_{A,II})$ would be less than $V(\bar{y}_{A,I})$. Next, from Table 2, $V(\bar{y}_{A,III})$ is seen to be equal to the first component of $V(\bar{y}_{A,I})$, or the first component of $V(\bar{y}_{A,II})$, so that $V(\bar{y}_{A,III})$ is always smaller than $V(\bar{y}_{A,II})$ which in turn would usually be smaller than $V(\bar{y}_{A,I})$.

Next, in Table 2, consider the column for stratified sampling. It would be expected that if the units were allocated to strata in $C(B, I)$ by proportional allocation, that $V(\bar{y}_{B,I})$ would be approximately the same as $V(\bar{y}_{A,I})$ and this is indeed the case, for if $m_t/m = M_t/M$ and $M_t - 1 \simeq M_t$ and $M \simeq M-1$ then the weighting of the within strata variances in both $V(\bar{y}_{A,I})$ and $V(\bar{y}_{B,I})$ is identical.

In $C(B, II)$, if the two classifications are statistically independent in the sense that

$$\frac{{}_iM_t}{M_t} = \frac{{}_iM}{M},$$

then $\bar{\bar{Y}}_t = \bar{Y}$ and the last component of $V(\bar{y}_{B,II})$ is a usual interaction term; hence the combination will give a higher precision if the effects of the strata are additive.

As for comparisons, if the allocation of units to strata in $C(B, II)$ is proportional to size, then $C(B, II)$ is equivalent to $C(A, II)$; however, if the allocation in $C(B, II)$ is Neyman optimum rather than proportional, $C(B, II)$ may be more precise than $C(A, II)$. Also, it is easily seen that $C(B, I)$ and $C(B, II)$ compare in exactly the same manner as $C(A, I)$ and $C(A, II)$; that is, $V(\bar{y}_{B,II})$ is not necessarily smaller than $V(\bar{y}_{B,I})$ but in most cases the second direction of stratification will cause $V(\bar{y}_{B,II})$ to be less than $V(\bar{y}_{B,I})$.

In the comparison of $V(\bar{y}_{B,III})$ with the variances of $C(B, I)$ and $C(B, II)$, the same remarks apply as in the comparison of $C(A, III)$ to $C(A, I)$ and $C(A, II)$; specifically, $C(B, I)$ and $C(B, II)$ each have a component in their variance formulas which does not appear in the formula $V(\bar{y}_{B,III})$. These components are the price which must be paid for not using, or not knowing, the population substrata sizes. As would be expected both of these extra components are zero if ${}_i\bar{Y}_t = {}_i\bar{Y} = \bar{Y}_t$. Finally, if proportional allocation is used in the strata of $C(B, III)$, then $C(A, III)$ and $C(B, III)$ are identical, and if optimum allocation is used, $C(B, III)$ is likely to be better than $C(A, III)$.

In the column of cell stratified sampling $\bar{y}_{C,III}$ is the only estimator considered; the other combinations would use less information than was used in the sample selection. If the allocation of the units to cells is proportional, $V(\bar{y}_{C,III})$ is identical to $V(\bar{y}_{A,III})$ and to $V(\bar{y}_{B,III})$, provided that the allocation in $C(B, III)$ is also proportional.

To summarize the comparisons, it has been shown that $C(A, III)$, $C(B, III)$, and $C(C, III)$ all have approximately the same sampling variances attached to their estimators. Similarly, $C(A, II)$ and $C(B, II)$ have approximately the same variances, as do $C(\mathrm{A}, \mathrm{I})$ and $C(\mathrm{B}, \mathrm{I})$. Consequently, combinations which utilize the same amount of population information have approximately the same precision. On the other hand, combinations which utilize different amounts of population information have different sampling variances, and the differences in these variances depend only upon the amount of information used and not upon whether it was used in the sample selection or the estimator. The same precision can be achieved whether the information is used in the estimator alone, the method of sampling alone, or in some combination; and as the amount of information increases the sampling variance gets smaller.

5. POST-STRATIFICATION

In many applied situations post-stratification may be necessary in that the stratum to which a unit belongs may not be identifiable in advance, or the observations may be multivariate and a given stratification may be appropriate for some but not for others. In this paper post-stratification has been used to obtain the point of view of one considering the population information available in advance of sampling; it is not the intent to advocate post-stratified weighting when prestratification is possible, for certain difficulties can be encountered. In particular, there is a non-zero probability that not all the strata will appear in the sample, and so, the post-stratified estimators are biased estimators of the over-all population mean; in addition, they are ratio estimators and have sampling variances which are inflated by the occurrence of small sample sizes in the strata. Consequently, since different groups were used as post-strata, the comparisons made in Section 4 are unequally affected by small sample sizes, but they are appropriate for samples large enough to produce stable ratio estimators.

After the sample has been selected, the conditional estimator, given the realized distribution of the sample, is frequently the most relevant. This estimator is the usual stratified estimator and is unbiased and has an exact variance. It should be realized however that these properties refer to a possibly restricted reference set, specifically, the one consisting of the over-all population exclusive of any strata which do not appear. Furthermore, allocation of the units after sampling does not allow optimum allocation to the strata and this also affects the conditional estimator.

Similar comments can be made about the variance estimator when fewer than two units fall into a stratum.

The conditional variance can be combined with a result of Stephan [2] which says that

$$E\left(\frac{1}{{}_im_t}\right) \cong \frac{1}{m}\,\frac{M}{{}_iM_t} - \frac{1}{m^2}\,\frac{M}{{}_iM_t} + \frac{1}{m^2}\,\frac{M^2}{{}_iM_t} \tag{5.1}$$

to terms in m^{-2}. For example, the conditional variance of $\hat{y}_{\mathrm{A,III}}$, ignoring the finite population correction, combines with Equation (5.1) to give an over-all variance of

$$V(\hat{y}_{\mathrm{A,III}}) = \sum_{i=1}^{L}\sum_{t=1}^{N} \frac{{}_iM_t}{M}\,\frac{{}_iS_t^2}{m} + \frac{1}{m^2}\sum_{i=1}^{L}\sum_{t=1}^{N} {}_iS_t^2\left(1 - \frac{{}_iM_t}{M}\right). \tag{5.2}$$

TABLE 3. GROSS RETAIL SALES OF FOOD AND DRINK ESTABLISHMENTS IN SOUTHWESTERN WYOMING, 1953
(Means in Thousands of Dollars)

Total Payroll		Grocery ($t=1$)	Cafe ($t=2$)	Bar ($t=3$)		
		Type of Establishment				
Under $5,000 ($i=1$)	${}_1M_t$	280	170	130	${}_1M$	580
	${}_1\overline{Y}_t$	52.50	15.94	25.54	${}_1\overline{Y}$	35.74
	${}_1S_t^2$	1500	134	145	${}_1S^2$	1057
	${}_1m_t$	28	17	13	${}_1m$	58
$5,000 & over ($i=2$)	${}_2M_t$	100	110	160	${}_2M$	370
	${}_2\overline{Y}_t$	151.90	66.27	53.94	${}_2\overline{Y}$	84.08
	${}_2S_t^2$	4835	1301	575	${}_2S^2$	4144
	${}_2m_t$	10	11	16	${}_2m$	37
	M_t	380	280	290	M	950
	$\overline{Y}_t$	78.66	35.71	41.21	$\overline{Y}$	54.57
	S_t^2	4779	1188	577	S^2	2790
	m_t	38	28	29	m	95
	$\overline{\overline{Y}}_t$	48.46	54.73	62.41		

The additional component in $V(\hat{y}_{A,III})$ is always positive but decreases as m increases and as a proportion of the leading term of the formula is

$$P = \frac{1}{m}\left\{\frac{\sum_{t=1}^{N}\sum_{i=1}^{L} {}_iS_t^2}{\sum_{i=1}^{L}\sum_{t=1}^{N}\frac{{}_iM_t}{M}\,{}_iS_t^2} - 1\right\}. \tag{5.3}$$

If all the within cell variances are equal, then

$$P = \frac{NL - 1}{m}$$

and the influence of this term varies inversely with the sample size and directly with the number of strata. If all strata are equal in size, with proportion $\overline{W}$, then

$$P = \frac{1}{m}\,\frac{1 - \overline{W}}{\overline{W}},$$

which shows that P gets bigger as $\overline{W}$ gets smaller, i.e., the number of strata gets larger.

The conditional approach can be applied to other combinations to assess the effects of small sample sizes. For example the conditional variance of $\hat{y}_{B,III}$ can be combined with a formula analogous to Equation (5.1). In this latter case, since m_t is smaller than m, it might be desirable to include terms in m_t^{-3}.

TABLE 4. SAMPLING VARIANCES OF THE ESTIMATORS OF THE MEAN (SALES DATA)

		Sample Selection		
		Completely Random (A)	One-Way Stratified (B)	Cell Stratified (C)
Type of Estimator	Marginally Weighted (I)	21.23	21.26	
	Bimarginally Weighted (II)	13.90	13.93	
	Cell Weighted (III)	11.70	11.72	11.77

6. A NUMERICAL ILLUSTRATION

To illustrate the discussion, Table 3 contains a summary of a complete enumeration of ninety-five food and drink establishments in southwestern Wyoming. These ninety-five businesses are considered to be a ten per cent sample of a larger universe with known means and variances. The two-way stratification of these establishments is by type, grocery stores, bars and cafes; and payroll, total annual payroll under \$5,000 and a total annual payroll over \$5,000. The sampling variances for the various schemes are displayed in Table 4. The data were taken from Bryant [1] with Harmston of the Division of Economic Analysis, University of Wyoming as the original source.

The variance of the sample mean with simple random sampling is equal to 26.43 showing that a reduction in variance of about twenty per cent could be expected from one-way stratification.

Table 4 clearly illustrates the fact that reduction in variance comes with the amount of information used in *both* the sample selection and the estimator and that this reduction can be achieved by the sample selection alone or by the estimator alone. In this example, the strata effects seem to be additive; and the difference between the means of the two payroll classifications is large, so that the bi-marginally weighted estimators perform significantly better than the marginally weighted estimators.

Calculations using payroll as the prestratification in type B combinations will be found to yield results entirely similar to those presented in Table 4.

REFERENCES

[1] Bryant, Edward C., "An Analysis of Some Two Way Stratifications," Ph.D. thesis. Ames, Iowa: Iowa State University Library, 1955.

[2] Stephan, F. F., "The expected value and variance of the reciprocal and other negative powers of a positive Bernouillian variate," *Annals of Mathematical Statistics*, 16 (1945), 50–61.

[3] Williams, W. H., "The variance of an estimator with post-stratified weighting," *Journal of the American Statistical Association*, 57 (1962), 622–7.