Survey Methodology, June 1988
Vol. 14, No. 1, pp. 33–43
Statistics Canada

On the Stratification of Skewed Populations

PIERRE LAVALLÉE and MICHEL A. HIDIROGLOU[1]

ABSTRACT

For a given level of precision, Hidiroglou (1986) provided an algorithm for dividing the population into a take-all stratum and a take-some stratum so as to minimize the overall sample size assuming simple random sampling without replacement in the take-some stratum. Sethi (1963) provided an algorithm for optimum stratification of the population into a number of take-some strata. For the stratification of a highly skewed population, this article presents an iterative algorithm which has as objective the determination of stratification boundaries which split the population into a take-all stratum and a number of take-some strata. These boundaries are computed so as to minimize the resulting sample size given a level of relative precision, simple random sampling without replacement from the take-some strata and use of a power allocation among the take-some strata. The resulting algorithm is a combination of the procedures of Hidiroglou (1986) and Sethi (1963).

KEY WORDS: Iterative algorithm; Optimum boundaries; Take-all; Take-some.

1. INTRODUCTION

Efficient sampling of highly skewed populations such as those displayed by business surveys require that they be stratified into a take-all stratum and a number of take-some strata. The whole of units the take-all stratum is selected with certainty whereas units in the take-some strata are selected by a probability mechanism. Approximate cut-off rules for stratifying a population into a take-all and a single take-some stratum have been given by Glasser (1962) and Hidiroglou (1986). Glasser (1962) provided the cut-off value under the assumption that a fixed total sample size was to be drawn from the take-all and take-some stratum, and that the take-some sampled units were to be selected without replacement using simple random sampling. Hidiroglou (1986) provided the cut-off value under the assumption that a required level of precision had to be satisfied. These two approaches are dual in the sense that Glasser's objective was to minimize sampling variance for fixed sample size, whereas Hidiroglou's objective was to minimize sample size for fixed sampling variance.

In this article, an algorithm for stratifying a highly skewed population into a take-all stratum and a number of take-some strata will be presented. The objective will be to minimize the overall sample size given the coefficient of variation of the estimator and the allocation scheme of the sample to the take-some strata. The strata boundaries will be derived in term of an auxiliary variable which is closely related to the information being collected by the survey. For example, for a census of retailers, if yearly sales is one of the variables measured, this auxiliary variable can be used to determine the strata boundaries for a single-purpose survey which collect sales on a monthly basis. For a multi-purpose survey, given that the strata boundaries have been determined using

[1] Pierre Lavallée is Methodologist and Michel A. Hidiroglou is Chief, Business Survey Methods Division, Statistics Canada, Ottawa, Ontario K1A 0T6, Canada. The authors would like to acknowledge France Bilocq, Business Survey Methods Division, Statistics Canada, for programming the examples.

an auxiliary variable closely related to the main variable, the optimality of these boundaries will diminish for other variables which are not well correlated with it. The algorithm is a modification of Sethi's (1963) method for stratifying a population. The resulting boundaries, which are optimal, will provide the required minimum sample size.

The allocation scheme which has been chosen to illustrate the method is the power allocation. The use of this type of allocation enables the publication of strata estimates which do not have markedly different coefficients of variation. Power allocation has been proposed by Carrol (1970), Fellegi (1981) and Bankier (1988). It is found to offer in practice a compromise between Neyman allocation and the requirement to have equal coefficients of variation for each stratum. A disadvantage of Neyman allocation is that if estimates are required for each stratum, the associated coefficients of variation may be quite different between the strata. Alternatively, an allocation which achieves equal coefficients of variation amongst the strata may require sample size which is much larger than the one required under Neyman allocation. In our context, power allocation would enable the publication of estimates for strata of varying sizes (small, medium and large) companies with similar coefficients of variation.

The method developed in the paper will be numerically compared, in terms of boundary values and sample size, to the Dalenius — Hodges (1959) cumulative square root f rule, as well as to a mixture of the Hidiroglou (1986) and the Dalenius — Hodges (1959) stratification methods. The algorithm, which is recursive in nature, is simple to program and converges rapidly to the optimum boundary points. It also offers substantial savings in terms of sample size for given reliability criteria.

2. THE PROBLEM

Consider a finite ordered population of N units:

$$y_{(1)}, y_{(2)}, \ldots, y_{(N)},$$

with $y_{(i)} \leq y_{(i+1)}$ for $i = 1, 2, \ldots, N-1$. This population is to be stratified into L strata. The number of units in each stratum is denoted by N_h, $h = 1, 2, \ldots, L$. The sampling scheme calls for n_h units to be drawn from each corresponding take-some stratum of size N_h $(h = 1, 2, \ldots, L-1)$ without replacement, using simple random sampling, with $n_L = N_L$. The mean to be estimated is

$$\bar{Y} = \sum_{h=1}^{L} \sum_{j=M_{h-1}+1}^{M_h} y_{(j)}/N, \tag{2.1}$$

where $M_h = \sum_{i=1}^{h} N_i$ for $h = 1, 2, \ldots, L$ and $M_o = 0$.

Given this set up, the estimator of population mean $\bar{Y}$ is

$$\hat{\bar{Y}} = \left[\sum_{h=1}^{L-1} \frac{N_h}{n_h} \sum_{j=m_{h-1}+1}^{m_n} z_j + \sum_{j=M_{L-1}+1}^{N} y_{(j)} \right] /N \tag{2.2}$$

where $y_{M_{h-1}+1} \leq z_j \leq y_{M_h}$ for $j = m_{h-1}+1, \ldots, m_h$ ($h = 1, 2, \ldots, L-1$), $m_h = \sum_{i=1}^{h} n_i$ for $h = 1, 2, \ldots, L$ and $m_o = 0$.

Assume that the desired level of precision for the estimated mean is specified by c (coefficient of variation) and that the proportion of sampled units to be allocated to each of the first $L-1$ strata is a_h ($h = 1, 2, \ldots, L-1$) where $\sum_{h=1}^{L-1} a_h = 1$. The term "a_h" is conveniently used to represent any type of allocation to the strata. For instance, in the case of N-proportional power allocation,

$$a_h = \frac{N_h^P}{\sum_{h=1}^{L-1} N_h^P} \quad (h = 1, 2, \ldots, L-1)$$

and in the case of Y-proportional power allocation,

$$a_h = \frac{Y_h^P}{\sum_{h=1}^{L-1} Y_h^P},$$

where $0 < p < \infty$. The power allocations have the property that under relatively simple assumptions and for a suitable choice of p, the coefficients of variation for the take-some strata tend to be equalized without a significant increase in the overall coefficient of variation. This equality of coefficients of variation is often desired by the users of the survey data.

In practice, the value of p is often chosen to be 1/2 or 1/3. A small value of p (i.e. p close to 0) usually yields similar stratum coefficients of variation while a larger value increases the discrepancy between the coefficients of variation but also increases the precision of the overall estimates.

It would be noted that these power allocations are equivalent to the allocation proposed by Bankier (1988) when the population coefficients of variation of the take-some strata are equal.

The variance of $\hat{\bar{Y}}$ is

$$V(\hat{\bar{Y}}) = \frac{1}{N^2} \sum_{h=1}^{L-1} \frac{N_h}{n_h} (N_h - n_h) S_h^2, \tag{2.3}$$

where S_h^2 denotes the population variance of each stratum h. In terms of the desired level of coefficient of variation c, $V(\hat{\bar{Y}})$ may be reexpressed as $V(\hat{\bar{Y}}) = c^2 \bar{Y}^2$. Substituting $n_h = (n - N_L)\, a_h$ and $V(\hat{\bar{Y}}) = c^2 \bar{Y}^2$ into (2.3) and solving for n obtains

$$n = N_L + \frac{\sum_{h=1}^{L-1} N_h^2 \, S_h^2 / a_h}{(N c \bar{Y})^2 + \sum_{h=1}^{L-1} N_h \, S_h^2}. \tag{2.4}$$

The problem is to find boundaries $b_{(1)}, b_{(2)}, \ldots, b_{(L-1)}$ (where $y_{(1)} < b_{(1)} < \ldots < b_{(L-1)} < y_{(N)}$) such that the overall sample size n is minimized, given the level of reliability c and the specific allocation scheme (represented by a_h).

3. THE ALGORITHM

The approach used in this paper, for obtaining stratification boundaries for a desired level of precision, has first been used by Dalenius (1950) in the case of stratification boundaries for a given sample size. It is first assumed that the sampling is done from a population whose frequency distribution may with sufficient accuracy be represented by a continuous density $f(y)$. Then, for a given set of boundaries $b_{(1)}, \ldots, b_{(L-1)}$ the following quantities are defined:

$$W_h = \int_{b_{(h-1)}}^{b_{(h)}} f(y)\, dy, \tag{3.1}$$

$$\mu_h = \int_{b_{(h-1)}}^{b_{(h)}} y\, f(y)\, dy / W_h, \tag{3.2}$$

$$\sigma_h^2 = \int_{b_{(h-1)}}^{b_{(h)}} y^2 f(y)\, dy / W_h - \mu_h^2, \tag{3.3}$$

for $h = 1, \ldots, L$, with $b_{(o)} = -\infty$, $b_{(L)} = +\infty$.

Equation (2.4) can then be rewritten as

$$n = NW_L + \frac{N\left(\sum_{h=1}^{L-1} W_h^2\, \sigma_h^2 / a_h\right)}{N c^2 \mu^2 + \sum_{h=1}^{L-1} W_h\, \sigma_h^2}, \tag{3.4}$$

where

$$\mu = \int_{b_{(o)}}^{b_{(L)}} y\, f(y)\, dy.$$

It should be noted that even if the population is considered to be large, the finite population correction (f.p.c.) factor is still present in equation (3.4) - see Dalenius-Gurney (1951). By definition, the take-all stratum needs to have a finite population in order to get a finite sample size. Also, ignoring the f.p.c. would not lead to a zero variance for the take-all stratum.

The a_h in equation (2.3) can also be represented using the quantities (3.1), (3.2) and (3.3). In the case of the N-proportional power allocation, we get:

$$a_h = \frac{W_h^p}{\sum_{h=1}^{L-1} W_h^p}, \tag{3.5}$$

for $h = 1, \ldots, L-1$.

For the Y-proportional power allocation, the following is obtained:

$$a_h = \frac{(W_h \mu_h)^p}{\sum_{h=1}^{L-1} (W_h \mu_h)^p}, \tag{3.6}$$

where $0 < p < \infty$.

In this paper, the Y-proportional power allocation will mainly be considered but the calculations can also be performed for the N-proportional power allocation and, in fact, for any kind of allocation represented by some a_h where $\sum_{h=1}^{L-1} a_h = 1$. Putting equation (3.6) into (3.4), we get

$$n = N W_L + \frac{N \left[\sum_{h=1}^{L-1} (W_h \sigma_h)^2 (W_h \mu_h)^{-p} \right] \left[\sum_{h=1}^{L-1} (W_h \mu_h)^p \right]}{N c^2 \mu^2 + \sum_{h=1}^{L-1} W_h \sigma_h^2}. \tag{3.7}$$

In order to find the optimal boundaries $b_{(1)}, \ldots, b_{(L-1)}$ such that the sample size n will be minimum, the derivatives of equation (3.7) are taken with respect to $b_{(1)}, \ldots, b_{(L-1)}$, respectively, and equated to zero. The resulting equations are:

For $h = 1, \ldots, L-2$,

$$[F T_h - F T_{h+1}] \; b_{(h)}^2 +$$

$$[F K_h - 2\mu_h F T_h - F K_{h+1} + 2\mu_{h+1} F T_{h+1} + 2\mu_h AB - 2\mu_{h+1} AB] \; b_{(h)} +$$

$$[F T_h \mu_h^2 + F T_h \sigma_h^2 - F T_{h+1} \mu_{h+1}^2 - F T_{h+1} \sigma_{h+1}^2 - AB\mu_h^2 + AB\mu_{h+1}^2] = 0, \tag{3.8}$$

and for $h = L-1$,

$$[F T_{L-1} - AB] \; b_{(L-1)}^2 +$$

$$[F K_{L-1} - 2\mu_{L-1} F T_{L-1} + 2\mu_{L-1} AB] \; b_{(L-1)} +$$

$$[F T_{L-1} \mu_{L-1}^2 + F T_{L-1} \sigma_{L-1}^2 - AB\mu_{L-1}^2 - F^2] = 0, \tag{3.9}$$

where

$$A = \sum_{h=1}^{L-1} (W_h \mu_h)^p,$$

$$B = \sum_{h=1}^{L-1} (W_h \sigma_h)^2 (W_h \mu_h)^{-p},$$

$$F = N c^2 \mu^2 + \sum_{h=1}^{L-1} W_h \sigma_h^2,$$

$$K_h = B p (W_h \mu_h)^{p-1} - A p (W_h \sigma_h)^2 (W_h \mu_h)^{-p-1},$$

$$T_h = A W_h (W_h \mu_h)^{-p}.$$

Labeling the coefficient of $b_{(h)}^2$ as α_h, the coefficient of $b_{(h)}$ as β_h and the remaining terms as γ_h, equations (3.8) and (3.9) can be represented as quadratic equations of the form $\alpha_h b_{(h)}^2 + \beta_h b_{(h)} + \gamma_h = 0$. However, as pointed out by Sethi (1963), the terms α_h, β_h and γ_h are themselves functions of $b_{(1)}, \ldots, b_{(L-1)}$ through the integrals (3.1), (3.2) and (3.3). Using Sethi's (1963) approach, equations (3.8) and (3.9) can easily be solved using the following iterative method:

STEP 1 : Start with some arbitrary boundaries $b_{(1)}' < \ldots < b_{(L-1)}'$.

STEP 2 : Calculate the proportions W_h', the means μ_h' and the variances $\sigma_h^{2'}$ (from equations (3.1), (3.2) and (3.3), respectively) based on these boundaries, $h = 1, \ldots, L-1$.

STEP 3 : Replace the initial set of boundaries by $b_{(1)}'', \ldots, b_{(L-1)}''$ where

$$b_{(h)}'' = \frac{-\alpha_h' + \sqrt{\beta_h'^2 - 4\alpha_h' \gamma_h'}}{2\alpha_h'}, \quad h = 1, \ldots, L-1. \tag{3.10}$$

STEP 4 : Repeat steps 2 and 3 till two consecutive sets are either identical or differ by negligible quantities, i.e.

$$\max_{h=1}^{L-1} | b_{(h)}'' - b_{(h)}' | < \epsilon \text{ for some } \epsilon > 0. \tag{3.11}$$

It should be noted that it can be proved that the sign before the square root ($\sqrt{\ }$) is positive because $b_{(h)}'$ lies between μ_h' and μ_{h+1}'.

The difficulty of using the above algorithm is that some knowledge of $f_{(y)}$, the approximate density, is required. Since the population considered is finite, it is possible to overcome this difficulty by replacing the quantities (3.1), (3.2) and (3.3) by corresponding expressions based on the finite population. Hence, proceeding as in Cochran (1977), the infinite population parameters given by expressions (3.1), (3.2) and (3.3) can be replaced by their finite population counterparts. That is:

$$W_h = \frac{N_h}{N}, \tag{3.12}$$

$$\bar{Y}_h = \frac{1}{N_h} \sum_{j=b(h-1)+1}^{b(h)} y_{(j)}, \tag{3.13}$$

$$S_h^2 = \frac{1}{N_{h-1}} \sum_{j=b(h-1)+1}^{b(h)} y_{(j)}^2 - N_h \bar{Y}_h^2, \tag{3.14}$$

for $h = 1, \ldots, L$.

Using these last quantities, the problem described in section 2 of finding boundaries $b_{(1)}, \ldots, b_{(L-1)}$ such that the overall sample size n is minimized for a given level of reliability c and a specific allocation scheme can easily be solved by the following iterative method:

STEP 0 : Sort the population $y_1, \ldots, y_N$ in ascending order and set $b_{(0)} = y_{(1)}$ and $b_{(L)} = y_{(N)}$.

STEP 1 : Start with some arbitrary boundaries such that $b_{(0)} < b'_{(1)} < \ldots < b'_{(L\text{-}1)} < b_{(L)}$.

STEP 2 : Calculate the proportions W'_h, the mean $\bar{Y}'_h$ and the variance $S_h^{2'}$ (from equations (3.12), (3.13) and (3.14) respectively) based on these boundaries, $h = 1, \ldots, L\text{-}1$.

STEP 3 : Replace the initial set of boundaries by $b''_{(1)}, \ldots, b''_{(L-1)}$ where

$$b''_{(h)} = \frac{-\alpha'_h + \sqrt{\beta_h'^2 - 4\,\alpha'_h\,\gamma'_h}}{2\,\alpha'_h}, \; h = 1, \ldots, L-1.$$

STEP 4: Repeat step 2 and 3 till two consecutive sets are either identical or differ by negligible quantities, i.e.

$$\max_{h=1}^{L-1} | b''_{(h)} - b'_{(h)} | < \epsilon \text{ for some } \epsilon < 0.$$

The use of this algorithm with real data will be compared to others in the next section.

4. SOME ILLUSTRATIONS

In order to display results given in Section 3, we will use data obtained from the Annual Retail Trade and Wholesale Trade Surveys conducted at Statistics Canada. These surveys measure the sales of companies whose principal business is retailing or wholesaling respectively. Three populations have been used to illustrate the algorithm. They are, respectively, other products in Wholesale in Quebec (Population 1), other foods in Wholesale in Manitoba (Population 2), and appliances, television, radio and stereo stores in Retail in Quebec (Population 3). Those populations have been chosen to reflect different combinations of population sizes: high, medium and low. The skewness for these populations is 24.2 (for Population 1), 6.5 (for Population 2) and 13.6 (for Population 3).

The numerical results provided by the algorithm will be compared to those obtained using two other methods. The first method is to simply stratify the population using the cumulative square root f rule given by Dalenius-Hodges (1959). The second method is to determine the cut-off boundary between take-all and take-some strata using the approximation given by Hidiroglou (1986) and then to apply the cumulative square root f rule to stratify the non take-all population into a number of take-some strata. The different methods will respectively be labelled as i) Cum $f^{½}$ rule, ii) mixture, and iii) optimum, for the currently proposed algorithm. The sole use of the Dalenius-Hodges (1959) method is not realistic because it would, in practice, only be used after the take-all stratum had been identified using some given arbitrary rule. However, we display the sole use of this method to caution against its blind use in the context of highly skewed populations.

The Hidiroglou (1986) cut-off point is obtained via the following iterative process:

$$b_{TA}'' = \mu_{[N-t']} = \left\{\frac{N-t'-1}{(N-t')^2} N^2 c^2 \bar{Y}^2 + S^2_{[N-t']}\right\}^{½}, \tag{4.1}$$

where

$$\mu_{[N-t']} = \frac{1}{N-t'} \sum_{i=1}^{N-t'} y_{(i)} \tag{4.2}$$

Table 1

Effect of Varying Coefficient of Variation and Power Allocation on Sample Sizes for Three Stratification Methods (Population 1 — Size = 1221)

			Stratification Method								
			Cum $f^{½}$ Rule			Mixture			Optimum		
c	p	Strata	N_h	n_h	$b_{(h)}$	N_h	n_h	$b_{(h)}$	N_h	n_h	$b_{(h)}$
0.05	0.25	1	1196	177*		1017	16		891	11	
		2	20	20	3,715,320	152	14	465,180	290	13	302,912
		3	5	5	14,786,280	52	52	1,131,961	40	40	1,835,930
		Total		202			82			64	
0.05	0.50	1	1196	178*		1017	16		863	10	
		2	20	20	3,715,320	152	13	465,180	318	14	289,422
		3	5	5	17,786,280	52	52	1,131,961	40	40	1,832,038
		Total		203			81			64	
0.01	1.00	1	1196	616*		751	37		687	36	
		2	20	20	3,715,320	215	34	196,840	374	78	162,068
		3	5	5	14,786,280	255	255	383,033	160	160	564,076
		Total		641			326			274	
0.05	1.00	1	1196	180*	3,715,320	1017	16		858	8	
		2	20	20	14,786,280	152	11	465,180	323	16	271,920
		3	5	5		52	52	1,131,961	40	40	1,867,254
		Total		205			79			64	
0.10	1.00	1	1196	56*		1073	7		1007	7	
		2	20	20	3,715,320	109	4	592,900	191	9	442,357
		3	5	5	14,786,280	39	39	1,953,113	23	23	4,032,950
		Total		81			50			39	

*Requires over allocation to satisfy coefficient of variation.

Table 2

Effect of Varying Coefficient of Variation and Power Allocation on Sample Sizes for Three Stratification Methods (Population 2 — Size = 44)

			Stratification Method								
			Cum $f^{1/2}$ Rule			Mixture			Optimum		
c	p	Strata	N_h	n_h	$b_{(h)}$	N_h	n_h	$b_{(h)}$	N_h	n_h	$b_{(h)}$
0.05	0.25	1	42	38		32	1		29	1	
		2	1	1*	137,939,900	6	1	4,708,409	11	1	3,029,455
		3	1	1	459,739,000	6	6	10,622,301	4	4	17,461,464
		Total		40			8			6	
0.05	0.50	1	42	38		32	1		28	1	
		2	1	1*	137,939,900	6	1	4,708,409	12	1	2,582,819
		3	1	1	459,739,000	6	6	10,622,301	4	4	17,640,325
		Total		40			8			6	
0.01	1.00	1	42	42		25	1		25	1	
		2	1	1	137,939,900	5	1	1,059,550	10	4	1,153,322
		3	1	1	459,739,000	14	14	3,742,377	9	9	5,969,271
		Total		44			16			14	
0.05	1.00	1	42	38		32	1		26	1	
		2	1	1*	137,939,900	6	1	4,708,409	14	2	1,779,500
		3	1	1	459,739,000	6	6	10,622,301	4	4	17,349,902
		Total		40			8			7	
0.10	1.00	1	42	30		34	1		28	1	
		2	1	1*	137,939,900	6	1	4,848,218	13	1	2,413,800
		3	1	1	459,739,000	4	4	16,749,625	3	3	30,091,449
		Total		32			6			5	

*Requires over allocation to satisfy coefficient of variation.

and

$$S^2_{[N-t']} = \frac{1}{N-t'-1} \sum_{i=1}^{N-t'} (y_{(i)} - \mu_{[N-t']})^2.$$

The number of take-all units obtained for each step of this iterative process is t'. The starting point for this approximation is

$$b'_{TA} = \mu_{[N]} + \{N c^2 \bar{Y}^2 + S^2_{[N]}\}^{1/2} \tag{4.3}$$

The stopping point for (4.1) is reached when the following inequality is satisfied:

$$0 \leq 1 - n(t'')/n(t') < 0.10 \tag{4.4}$$

Table 3

Effect of Increasing the Number of Strata on Sample Sizes for Two Stratification Methods
$p = 1, c = 0.05$

Population 1 (N = 1221) Stratification Method	Strata	Number of Strata 3 N_h	3 n_h	3 $b_{(h)}$	4 N_h	4 n_h	4 $b_{(h)}$	5 N_h	5 n_h	5 $b_{(h)}$
Mixture	1	1017	16		897	6		823	3	
	2	152	11	465,180	194	5	311,117	194	2	245,090
	3	52	52	1,131,961	78	4	641,252	101	2	465,180
	4				52	52	1,131,961	51	2	751,297
	5							52	52	1,131,961
	Total		79			67			61	
Optimum	1	858	8		704	3		655	2	
	2	323	16	271,920	373	7	173,981	358	4	149,327
	3	40	40	1,867,254	112	6	604,869	163	5	453,114
	4				32	32	2,676,449	29	4	1,522,329
	5							16	16	5,810,487
	Total		64			48			31	
Population 3 (N = 161)										
Mixture	1	106	6		84	2		71	1	
	2	39	6	265,480	38	2	185,320	35	1	155,260
	3	16	16	553,255	23	2	335,620	22	1	265,480
	4				16	16	553,255	17	1	385,720
	5							16	16	553,255
	Total		28			22			20	
Optimum	1	86	4		55	1		34	1	
	2	65	9	199,415	61	3	125,572	51	1	83,594
	3	10	10	680,942	39	5	312,769	42	2	192,215
	4				6	6	826,942	29	3	382,236
	5							5	5	906,894
	Total		23			15			12	

where

$$n(t') = t' + \frac{(N-t')^2 S^2_{[N-t']}}{(N c \bar{Y})^2 + (N-t') S^2_{[N-t']}} . \qquad (4.5)$$

Tables 1 and 2 display the results for a large population (Population 1) and a small population (Population 2) for a number of different coefficients of variation and power allocations. Table 3 displays the results for the large population (Population 1) and a medium population (Population 3) by varying the number of strata. For all three tables, the allocation of the sample to the take-some strata is the power Y-proportional scheme.

The following conclusions can be drawn from Tables 1 and 2. The use of the cumulative square root f rule to determine boundary points is very inefficient in the present context. Substantial gains, in terms of sample size reduction, are made by using the mixture rule. For the three strata used in those two tables, further reductions in sample size of the order of 20% can be achieved by using the optimum rule. For a given fixed coefficient of variation, the variation of the power "p" has a minor impact on the resulting sample size. As expected, sample sizes increase when the required coefficient of variation, c, is decreased (for a fixed power allocation). The optimum method declares less take-all units (stratum 3) than the mixture method, or stated another way, the take-all stratum boundary is higher for the optimum than for the mixture. The cumulative square root rule loses its efficiency in the take-all stratum boundary determination. It is readily observed that the boundary for this method is significantly higher than those obtained with the other methods.

In Table 3, we only compare the mixture and optimum methods for two populations, varying the number of strata, for a fixed coefficient of variation and Y-proportional power allocation. Similar conclusions to those drawn from Tables 1 and 2 hold. The effect of increasing the number of strata is to reduce the number of sampled units for both methods. However, the reduction becomes more pronounced for the optimum method as the number of strata increases.

5. CONCLUSION

The optimal stratification, of a skewed population into a take-all stratum and a number of take-some strata, has provided a substantial reduction in overall sample size for given relative precision. The method can be adapted to any type of allocation and to any number of strata. The take-all condition can also be excluded.

The algorithm, which is recursive in nature, converges quickly. It is simple to implement on the computer using SAS, FORTRAN, or any other high level language.

REFERENCES

BANKIER, M.D. (1988). Power allocations, determining sample sizes for sub-national areas. To appear in *The American Statistician*.

CARROL, J. (1970). Allocation of a sample between States. Unpublished memorandum of Australian Bureau of Census and Statistics.

COCHRAN, W.G. (1977). *Sampling Techniques*, (3rd. ed.). New York: John Wiley & Sons.

DALENIUS, T. (1950). The problem of optimum stratification. *Skandinavisk Aktuarietidskrift*, 33, 203-213.

DALENIUS, T., and GURNEY, M. (1951). The problem of optimum stratification. *Skandinavisk Aktuarietidskrift*, 34, 133-148.

DALENIUS, T., and HODGES, J.L.Jr. (1959). Minimum variance stratification, *Skandinavisk Aktuarietidskrift*, 54, 88-101.

FELLEGI, I.P. (1981). Should the census counts be adjusted for allocation purposes? - Equity considerations. In *Current Topics in Survey Sampling* (Eds. D. Krewski, R. Platek and J.N.K. Rao). New York: Academic Press, 47-76.

GLASSER, G.J. (1962). On the complete coverage of large units in a statistical study. *International Statistical Review*, 30, 28-32.

HIDIROGLOU, M.A. (1986). The construction of a self-representing stratum of large units in survey design. *The American Statistican*, 40, 27-31.

SETHI, V.K. (1963). A note on optimum stratification of populations for estimating the population means. *Australian Journal of Statistics*, 5, 20-33.