RESEARCH REPORT SERIES
(Statistics #2009-08)

Sample Allocation and Stratification

William E. Winkler

Statistical Research Division
U.S. Census Bureau
Washington, D.C. 20233

Report Issued: November 16, 2009

Sample Allocation and Stratification

William E. Winkler   2004.03.04

This chapter covers stratification and sample allocation for one-variable and multi-variable selection schemes. The purpose of sampling is to reduce cost by taking a subset of a population while assuring that the accuracy of one or more estimates are preserved. Stratification and clustering are methods for subdividing a population into subsets in which efficient sampling can be performed. Two implicit assumptions are typically made. The first is that the population file contains all members of the desired population and is free of duplicates. The second is that the quantitative and other variables used in the scientific design are accurately recorded in computer files and represent information that correspond to the desired population estimates. These assumptions can be relaxed in some situations.

The outline of this chapter is as follows. In the first section, we describe univariate methods of stratifying and sampling for one variable. These methods are due to Dalenius and Hodges (1959) and Ekman (1959). Lavallée and Hidiroglou (1988) and Winkler (1998) have given extensions for situations in which the underlying population distribution for continuous variables has significant gaps. The gaps can affect the stratification procedures of Dalenius and Hodges in particular. In the second section, we cover methods of stratification for more than one variable. The stratification ideas begin with independent stratification of two or more variables that are extended to a two or more way stratification of a population. The original ideas are due to Tepping, Hurwitz, and Demming (1943) with extensions by Goodman and Kish (1950) and Bryant, Hartley, and Jessen (1960). Modern extensions using non-trivial computational methods are due to Rao and Nigam (1990, 1992), Sitter and Skinner (1994), and Winkler (2001). In the final section, we give concluding remarks.

1. Univariate Stratification and Sampling

In this section, we describe methods for one-way stratification and sampling. The first three subsections consist of background notation, a description of several methods for one-way stratification, and an empirical comparison of the methods on highly skewed populations. The fourth subsection consists of remarks giving relationships to other closely related concepts in the literature. The fifth subsection describes additional concepts that are often used for grouping a population into subsets for sampling and are not closely related to methods for stratification.

1.1 Notation and Definitions

In this section, we summarize notation that is standard in sampling theory (e.g., Cochran 1977). A population of size N is partitioned into subsets called strata. The subscript h denotes the stratum and the subscript i denotes the unit within the stratum. The following all refer to stratum h:

N_h, total number of units,
n_h, number of units in sample from stratum h,
y_{hi}, value obtained for i^{th} unit of stratum h,

y_h, breakpoint between stratum h and stratum $h+1$,

$$Y_h = \sum_{i=1}^{N_h} y_{hi}, \quad \text{true total}$$

$$\hat{Y}_h = (N_h / n_h) \sum_{i=1}^{n_h} y_{hi}, \text{ estimated total, and}$$

$$S_h^2 = (1/(N_h - 1)) \sum_{i=1}^{N_h} (y_{hi} - (Y_h / N_h))^2, \text{ variance } y_{hi}.$$

In most situations, we will only consider simple random sampling within each stratum. We always assume $n_h > 1$. If L is the number of strata, then the population total is given by

$Y = \sum_{h=1}^{L} Y_h$ and the estimate of the population total is given by

$$\hat{Y} = \sum_{h=1}^{L} \hat{Y}_h.$$

In many situations, we wish to allocate the sample to strata to minimize variance under a fixed sample size or minimize sample size under a fixed variance. We define the terms

$$C = n = \sum_{h=1}^{L} n_h, \text{ fixed overall cost or sample size, and}$$

$$V = V(\hat{Y}) = \sum_{h=1}^{L} (N_h (N_h - n_h) S_h^2 / n_h), \text{ variance of the estimate of the population total.}$$

The unvariate stratification problem is to minimize the variance V under the fixed cost C. Alternatively, we may keep the variance V below a fixed bound while minimizing the total sample size or cost C. If the number of strata L and stratum boundaries given by the breakpoints y_h are fixed, then Neyman allocation (e.g., Cochran 1977, p. 97) can minimize C for fixed V or minimize V for fixed C. The sample size n_h is determined by the formula

$$n_h / n = (N_h S_h / \sum_{h=1}^{L} N_h S_h).$$

Because the underlying population distribution is discrete rather than continuous, the terms n_h/n on the left hand of the equation can only be approximately equal to the terms on the right hand side. If the population distribution is quite skewed, then a number of sample units can be allocated to a certainty stratum. The *finite population correction* (*fpc*) is given by $(N_h - n_h)/N_h$. In the certainty stratum, $n_h = N_h$. For the variance component of V associated with the certainty stratum, the *fpc* causes the variance component to be zero. How the sample is allocated to the certainty stratum and noncertainty strata will cause further complications in the use of the Neyman and other allocation formulas.

We always use Neyman allocation. Although *proportional* allocation (allocating in proportion to the number of units N_h in stratum h) has often been used, we do not use it. Neyman allocation is known to be theoretically optimal in comparison with proportional allocation. In some empirical work, proportional allocation often yields results that are similar to those from Neyman allocation (Kish, 1965, pp. 82-92). Proportional allocation is easier for hand calculation. Because the stratum variances S_h^2 are easily computed in modern computing environments, we prefer Neyman allocation that needs the stratum variances S_h^2. We note that the general problem of stratification necessitates that we simultaneously determine stratum boundaries y_h and the allocations n_h. A problem that is equivalent to the problem to minimizing variance for a given sample size is minimizing the coefficient of variation (*cv*) for the fixed sample size. For our purposes of bounding variances, we also use the coefficient of variation $cv = \sqrt{V}/Y$. The *cv* is practical because it removes dependence on the scale (or range) of the values associated with the *y*-variables.

1.2 Univariate (one-way) Stratification Methods

To assure proper understanding of the empirical comparisons, we summarize the methods of Dalenius and Hodges and of Ekman (see e.g., Cochran 1977, pp. 127-131). We assume that the population of values is sorted (here, decreasing). To simplify notation slightly, we denote population values by z_i. Then the decreasing sorted values are $z_1 \geq z_2 \geq \cdots \geq z_N$ where N is the size of the population. We divide the population into L strata by finding stratum breakpoints y_h. If we let $y_0 = z_1$ and $y_L = z_N$, then

$$y_{h-1} \geq y_{h1} \geq \cdots \geq y_{hn_h} \geq y_h \quad for\ h = 1, \cdots, L$$

where $y_{hi}\ i = 1, \cdots, n_h$ is an enumeration of the z_is that are between y_{h-1} and y_h under the restriction that each z_i and y_{hi} is only counted in one stratum h.

Dalenius and Hodges (1959) stratify to make the quantities

$$N_h^{1/2}(y_h - y_{h-1}),\ h = 1, \cdots, L \qquad (1)$$

approximately equal where the number of strata L are decided a priori.

Ekman (1959) stratifies to make the quantities

$$N_h(y_h - y_{h-1}),\ h = 1, \cdots, L \qquad (2)$$

approximately equal.

To minimize the variance when sample size is fixed, Winkler (1998) tries to assure that the quantities

$$N_h^{1/2}(N_h - n_h)^{1/2}(S_h / n_h),\ h = 1, \cdots, L \qquad (3)$$

are approximately equal. Equation (3) implicitly accounts for the range of values in the h^{th} stratum by using the square root of the population variance given by S_h. It also explicitly makes use of the finite population correction (*fpc*).

As the underlying population distributions are discrete, we choose the stratification among a finite number of stratifications for which approximate equality in (3) holds. The chief differences between method of Winkler and the methods of Dalenius and Hodges and of Ekman are:

(a) Winkler accounts for the finite population correction
(b) Winkler uses the standard deviation S_h instead of the surrogate $(y_h - y_{h-1})$, and
(c) Winkler allows a choice among a finite number of stratifications for which approximate equality in (3) holds.

In actual practice, statisticians using the methods of Dalenius and Hodges and of Ekman also allow choice among stratifications approximately satisfying formulas (1) and (2), respectively.

Winkler provides a second stratification:

$$N_h^{1/2}(N_h - n_h)^{1/2}(S_h / n_h^{1/2}),\ h = 1, \cdots, L \qquad (4)$$

for which each term must be approximately equal. Equality in (3) assures that each of the terms in the variance *V* are equal. If each n_h is equal, then the stratification yielding equalities in (3) agrees with the stratification yielding equalities in (4). We denote the first Winkler stratification method by S1 and the second by S2. We denote the Dalenius-Hodges method by D-H and the Ekman method by Ek.

If we allow the number of strata *L* to vary, then the minimum *cv* for a fixed sample size *n* is obtained when each noncertainty stratum has a sample allocation of two units. In the following, we will assume that the sample allocation scheme has one certainty stratum and *L-1* noncertainty strata. With very skewed population distributions, optimization can primarily depend on determining the number of units sampled with certainty. The procedure for finding the minimum *cv* consists of a straightforward grid search for the *L* strata with corresponding break points obtained using the method of the theorem. If we increase *L* (i.e., decrease the size of the certainty stratum) and the *cv* increases, then we reverse the direction of our search (i.e., decrease *L*). There are necessarily *n-2(L-1)* units in the certainty stratum. With skewed populations, the *L* yielding a minimum *cv* is generally obtained when the first noncertainty stratum is sampled at a rate less than 50 percent.

1.3 Empirical Results

The database used in the empirical analyses consists of 1106 records, each having three quantitative data elements. The empirical data are an anonymized version of actual economic survey data. The distribution of each variable is quite skewed. Nonzero quantitative values vary from 61,000 to 1 for the first variable, from 76,000 to 1 for the second, and from 241,000 to 1 for the third. For each variable, a small number of values cover a moderate proportion of the population total. In each case, the number of strata and the sample size within strata are fixed. Strata boundaries are then chosen that approximately minimize the variance.

The empirical results show that the two methods of Winkler are roughly equivalent and slightly better than the method of Ekman. All three methods are better than the method of Dalenius and Hodges (Table 1) because they yield lower *cv*s. With three exceptions, method S1 of this paper performs best. The first two exceptions are associated with stratification *C* for variable 1 and stratification A with variable 3. Method S2 of this paper performs slightly better (.1518 versus .1541 and .1985 versus .2065, respectively). In the footnote to Table 1, the numbers following the word 'certainty' indicate the number of noncertainty strata and the sample allocations within each. For instance, '2,2,2,2,2' indicates that there are five noncertainty strata and that the sample allocations to each are 2 units.

Table 1. Comparison of *cv*s
Different Stratifying Methods

Method	Variable and Stratification							
	1A	1B	1C	2A	2B	2C	3A	3B
S1	.0810	.0965	.1541	.0793	.0832	.1468	.2065	.1041
S2	.0810	.1007	.1518	.0796	.0833	.1478	.1985	.1052
EK	.0815	.1091	.1556	.0801	.0818	.1575	.2081	.1144
D-H	.1147	.1023	.1720	.1048	.1462	.1620	.2207	.1285

1A- Var 1 Sample: 8 certainty 2,2,2,2,2 noncertainty, total-18.
1B- Var 1 Sample: 3 certainty, 2,2,4,8 noncertainty, total-19.
1C- Var 1 Sample: 4 certainty, 2,3,5 noncertainty, total-14.
2A- Var 2 Sample: 8 certainty, 2,2,2,2,2 noncertainty, total-18.
2B- Var 2 Sample: 6 certainty, 4,4,8 noncertainty, total-16.
2C- Var 2 Sample: 3 certainty, 4,4,8 noncertainty, total-19.
3A- Var 3 Sample: 4 certainty, 4,8 noncertainty, total-16.
3B- Var 3 Sample: 4 certainty, 3,4,5 noncertainty, total 16.

For general skewed populations, we would use either the stratification methods of Winkler or of Ekman. With the exception of 1B in which D-H does better than EK but worse than S1 and S2, D-H does worse than the other three methods. As each of the methods is easy to program, we prefer having printouts that give a side-by-side comparison. The method of Dalenius and Hodges performs poorly primarily because the underlying probability density function is not constant in stratification intervals and the finite population correction cannot be ignored.

1.4 Remarks on One-way Stratification

Cochran (1977, Chapter 5A) observed that, for a fixed and sufficiently large sample size, increasing the number of strata from a small number to a somewhat larger number will generally cause the overall *cv* (coefficient of variation) to decrease. He assumed that a minimum of two units would be allocated (sampled) in each of the individual strata. In the single-variable situation, he observed that, if the number of individual strata were increased, then there was a point at which further increase in the number of strata yielded a negligible decrease in the estimated *cv*. Cochran's (1977, p. 133) heuristic rule for the single-variable situation was that increasing the number of strata beyond six would yield little benefit.

Cochran (1977) also observed that when the distribution of a single variable was quite skewed, then it was always advantageous to put the largest units in a certainty stratum. Because the largest units are sampled with certainty, the finite population correction (*fpc*) yields a variance term for the certainty stratum that contributes zero to the *cv*. Because the largest units can contribute significantly to the population total, the decrease in the overall *cv* for a given sample size can be substantial. In actual skewed populations, there can be substantial gaps between the values of the largest units. For a given fixed sample size, Lavallée and Hidiroglou (1988) gave an algorithm for automatically finding the number of units to put in the certainty stratum. Later work by Chen (1989), Slanta and Krenzke (1994), and Rivest (2002) indicated that the algorithm does not always work well in practice. The basic reasons are due to numeric instability of the programmed versions of the algorithm and convergence to local minima in distributions that were often multimodal. We can also observe that Lavallée and Hidiroglou (1988) made distributional assumptions similar to those of Dalenius and Hodges (1959). It is possible that a straightforward grid search based on the Ekman ideas (equation (2)) or the Winkler ideas (equation (3)) may yield more suitable results.

Another issue that typically occurs is that several variables in a survey must be estimated. If that is the case, then a method of Jarque (1981) for creating a new variable that is a linear combination of the existing variables and applying one of the one-way stratification methods may be effective. If we only consider two variables that each have skewed distributions and are uncorrelated, then a linear combination of the two variables may not be highly correlated with either variable. In those situations, it may be useful to consider alternative methods. We note that stratifying on one skewed variable may do little to reduce the *cv* of a second variable that is uncorrelated with the first. Cochran (1977, p. 132) provides a concise mathematical justification.

1.5. Additional concepts

This subsection considers three ideas (1) post-stratification, (2) cluster sampling, and (3) *two stage sampling* with *psu*s (*primary sampling units*) versus stratification. First, *post-stratification* is a method for stratifying a sample (possibly simple random) after the sample has been selected. For instance, for some variables such as age, the stratum to which a unit belongs may not be known. As the age is obtained, units are post-stratified into strata where the strata sizes N_h may be obtained from another survey such as a census. If the sample size in stratum h is given by $n_h > 0$, then, we can get a population estimate for the mean

$$\bar{y} = \sum_{h=1}^{L} (N_h / N) \bar{y}_h,$$

where $\bar{y}_h = \hat{Y}_h / n_h$ is the arithmetic mean of the sample units in stratum *h*. If we ignore the *fpc*, then we get that the variance is

$$V(\bar{y}) = \sum_{h=1}^{L} (N_h^2 / N^2) S_h^2 (\bar{y}_h),$$

where $S_h^2(\bar{y}_h)$ is the variance of $\bar{y}_h$. If each of the n_h is large, then Cochran (1977, p. 134) indicates that the increase in the variance due to post-stratification can be small in contrast to a (pre-)stratification where the strata are known prior to sampling and proportional allocation is the standard for comparison. For further reading on post-stratification, see Holt and Smith (1979) or Little (1993).

Second, *cluster* sampling divides a population of *N* units into *M* subpopulations consisting of groups of units that are then sampled. In the situation of housing surveys or surveys of areas, cluster sampling is a means of sampling that is intended to reduce costs where traveling may be expensive. If we wish to select a sample of 1000 households in a city, then rather than draw a random sample, we may cluster the city by blocks having approximately 50 units per block. Then, our sample would consist of 20 blocks. The idea is to reduce travel costs significantly. Rather than make clusters as homogeneous as possible as in one-way stratification, it is often helpful to have clusters be as heterogeneous as possible (Jessen 1978, p. 192). If the intraclass correlation coefficient ρ is negative, then it is possible that the overall variance can be reduced when compared to simple random sampling. Cochran (1977), Hansen et al. (1953), and Jessen (1978) are good references on clustering and the intraclass correlation coefficient ρ.

Third, *two-stage* sampling divides a population of *N* units into *M* subpopulations consisting of groups of units that that are sampled in the first stage. We refer to the *M* subpopulation groups as *primary sampling units* (*psus*). In the second stage, we sample units within each of the *psus*. If we wish to sample 200 households in a city, then we might first cluster the households in the city by blocks consisting of approximately 50 units each. At the first stage, we could sample 20 blocks (*psus*) and then subsample 10 units (households) per block. This type of sampling is suitable if the characteristics of the units within blocks that we wish to estimate are reasonably homogeneous. Sampling within *psus* can be done using a variety of methods (e.g., Cochran 1977, Chapters 10 and 11).

2. Multivariate Stratification and Sampling

The method of this section involves multi-way stratification where single-way designs are combined into a multi-way design. Overall sample size is kept relatively small by not allocating the sample to every multi-way cell. The sample allocation has similarity to Latin Square designs as was observed by Tepping et al. (1943).

In most survey settings, we wish to control the *cv*s of several variables. With single variables, the single-way methods of stratification work well. If we stratify on $p \geq 2$

variables, then we can stratify the population with the *p*-way stratification induced by the *p* single-way designs. If this type of design is naively used, then sample must be allocated to every *p*-way cell and sample size can become very large (e.g., Cochran 1977, pp. 124-126). One alternative is to define a new variable that is a linear combination of the *p* variables and apply a single-way design to all of the *p*-way cells. This alternative strategy will only partially control the *cv*s of the *p*-variables.

Another alternative *p*-way stratification strategy is to develop a Latin-Squares type of sampling strategy that does not require the sample to be allocated to every *p*-way cell. The earliest such strategy for two-way designs was introduced by Tepping et al. (1943) for use in population sampling in which the underlying distributions were not too skewed. Bryant et al. (1960) extended the two-way designs to populations that were slightly more skewed. Since the two-way method of Bryant et al. (1960, hereafter BHJ) gives crucial insights and is a special case of more general methods, we cover it detail. BHJ considered the 2-dimensional situation when two single-variable stratifications $F = (n_{i\cdot})$, $1 \le i \le n_f$, and $G = (n_{\cdot j})$, $1 \le j \le n_g$, have been created. The $(n_{i\cdot})$, $1 \le i \le n_f$, and $(n_{\cdot j})$, $1 \le j \le n_g$, are the counts associated with the one-way samples. We assume that the two one-way sample sizes $n = \sum_{i=1}^{n_f} n_{i\cdot} = \sum_{j=1}^{n_g} n_{\cdot j}$ are equal. The two stratifications induce a two-way stratification with population cells having proportions of the total population size P_{ij}, $1 \le i \le n_f$, $1 \le j \le n_g$. We wish to devise a sampling strategy that assigns two-way sample size (n_{ij}), $1 \le i \le n_f$, $1 \le j \le n_g$, so that its marginal counts $(n_{i\cdot})$, $1 \le i \le n_f$, and $(n_{\cdot j})$, $1 \le j \le n_g$, agree with the original two one-way sample sizes. The two-way stratification has $n_{f\cdot} \cdot n_g$ two-way cells. If we were to use naïve sampling, then we might assign a minimum sample size of 2 in each two-way cell for a total sample size of at least 2 $n_f \cdot n_g$. We wish to develop a method of allocating a sample such as in Latin-squares in which we do not need to sample in each two-way cell. The overall sample sizes are controlled by the marginal values *F* and *G*.

BHJ developed a method in which the (n_{ij}), $1 \le i \le n_f$, $1 \le j \le n_g$, are randomly allocated in an *allocation* step for which

$$E^1(\hat{n}_{ij}) = P_{i\cdot} P_{\cdot j} n, \qquad (5)$$

where the first-stage expectation E^1 is taken with respect to the randomization that is associated with $\hat{n}_{ij}$, $P_{i\cdot} = \sum_j P_{ij}$, and $P_{\cdot j} = \sum_i P_{ij}$. Here $P_{i\cdot}$ is the proportion of the population associated with the sample stratification of the first variable and $P_{\cdot j}$ is the proportion of the population associated with the sample stratification of the second variable. For each realization of $\hat{n}_{ij}(s)$, we obtain a fixed integer array n_{ij} (s). Here *s* refers to a particular sample. After the fixed integer n_{ij}(s) array is determined, we can use simple random sampling within two-way cells. The second stage of sampling is denoted with expectation E^2. BHJ gave formulas for estimates of the population means and divided the variance (using E^1 and E^2 with some algebra) into two components. BHJ observed that their estimates could be biased due to the fact that their expected sample proportions for which $E^1(\hat{n}_{ij}) = P_{i\cdot} P_{\cdot j}\, n$ could deviate significantly from observed population proportions associated with the sample that are given by $P_{ij}\, n$. Although BHJ gave some adjustment procedures for the biases, we need to consider allocation procedures that are somewhat similar to the BHJ procedure related to equation (5) but better correspond to observed population proportions P_{ij}. The allocation procedures and associated probability models for creating $\hat{n}_{ij}$ must also generalize to dimensions higher than 2.

The natural way to extend BHJ is to use standard iterative proportional fitting (IPF) procedures (Bishop et al. 1975) to allocate two single-way designs to two-way population cells and that can generalize to higher dimensions. Winkler (1990, also 2001) showed that standard IPF procedures that allocated two single-way designs to two-way population cells could yield two-way sample sizes that exceed observed population sizes in certain cells. We also use Dykstra's Generalized Iterative Fitting Procedure (GIFP, Dykstra 1985) that holds for any dimension and allows convex constraints in addition to the linear constraints of classical IPF. With both IPF and GIFP, the starting point of the fitting procedure can be an array of population counts given in Table 2. We observe that the structural zero in population cell (3,1) of Table 2 is preserved in the fitted solutions given in Tables 3 and 4. The classical IPF yields the fitted solution given in Table 3 and Dykstra's GIFP yields the fitted solution given in Table 4. The classic IPF provides a fitted value in cell (1,1) of Table 3 that exceeds the population value of 2 that is in Table 2. The GIFP that also allows convex constraints gives a solution for which none of the fitted sample values exceed the corresponding population values. The convex constraint in this example is that the value in cell (1,1) restricted to be less than or equal 2. If there are structural zeros, then, by Lemma 1 of Winkler (1990), the GIFP procedure is guaranteed to converge to the correct unique limiting solution provided that the marginal constraints, the interactions that are fit, and the structural zeros are consistent (i.e., have at least one solution). In two dimensions, there are no interactions terms. The advantage of interaction terms in higher dimensions is that the fitted array may correspond much more closely to the population proportions P_{ij}. The output of the GIFP is the non-integer fitted array $M_{ij} = (n_{ij})$.

The second component of the *p*-way stratification procedure is an *allocation step* in which we define a random variable $\hat{n}_{ij}$ such that $E^1(\hat{n}_{ij}) = (n_{ij}) = M_{ij}$ where the right-hand-side is the output of the GIFP. In higher dimension (3 or greater), we use **i** to denote a general index. With equation (5), BHJ essentially found a convex sum

$$M_{ij} = \sum_t p_t M_{ijt} \quad where\ p_t > 0 \ \ and \sum_t p_t = 1. \qquad (6)$$

Here M_{ij} is the non-integer array (n_{ij}) by GIFP fitting. Each M_{ijt} is an integer array having margins that agree with the original one-way stratifications and samples. In the situation of BHJ, the fact that each $p_t = (1/n!)$ is due to their method of determining $\hat{n}_{ij}$ in the sample design. For 2-dimensions, Causey, Cox, Ernst (1985, hereafter CCE) gave an algorithm for finding a set of integer arrays M_{ijt} that was extended to higher dimensions by Winkler (2001). The inductive algorithm iterates through steps where at each step *t* an integer array M_{ijt} is found. The algorithms of CCE and Winkler are similar. The bound on the

number of terms in the equation (6) is the number of *p*-way cells. At each stage of the

Table 2. Population Counts **N** Induced by Two Single-Criteria Stratifications 1/

Var 1 Strata	Variable 2 Strata 1	2	3	4	5	Total
1	2	7	4	1	11	25
2	3	5	7	17	31	63
3	0	10	16	47	85	158
4	2	3	10	78	257	350
5	3	5	29	67	551	655
	10	30	66	210	935	1251

1/ The quantitative measures of size associated with the variables, by strata, are:
Variable 1- 1 : 20589-8116, 2 : 8115-3501, 3 : 3500-1501, 4 : 1500-501, and 5 : 500-0; and
Variable 2- 1 : 16423-5195, 2 : 5194-2260, 3 : 2259-648, 4 : 647-152, and 5 : 150-0.

Table 3. Fitted Sample Matrix **A** Obtained by Classical IPF Marginal Totals are Fixed

Variable 1 Strata	Variable 2 Strata 1	2	3	4	5	Total
1	2.172	2.393	0.997	0.097	0.341	6.
2	2.098	1.101	1.124	1.059	0.618	6.
3	0.0	1.641	1.914	2.182	1.263	7.
4	0.820	0.387	0.941	2.848	3.004	8.
5	0.910	0.478	2.024	1.814	4.774	10.
	6.	6.	7.	8.	10.	37.

Table 4. Fitted Sample Matrix **B** Obtained by Dykstra's GIFP Marginal Totals are Fixed

Variable 1 Strata	Variable 2 Strata 1	2	3	4	5	Total
1	2.000	2.483	1.052	0.103	0.362	6.
2	2.182	1.061	1.101	1.046	0.610	6.
3	0.0	1.614	1.914	2.200	1.272	7.
4	0.860	0.377	0.930	2.840	2.993	8.
5	0.958	0.465	2.003	1.811	4.763	10.
	6.	6.	7.	8.	10.	37.

algorithm, we need to be able to obtain the integer array M_{ijt} with the properties that it preserves the one-way sample sizes and the row and column entries add to the corresponding integer margins. In 2-dimensions, Cox and Ernst (1982) demonstrated that zero-controlled roundings (to base 1) always exist. If we *round* a number x to integer base b, then we either round to the largest multiple of b that is less than x or to the smallest multiple of b that is larger than x. A *zero-controlled rounding* of x to base b keeps x at any multiple of b. If the base is 1, then controlled integer rounding to base 1 assures that the marginal values (the one-way sample sizes) are preserved.

Ernst (1989) provided a counter example showing that neither roundings nor zero controlled roundings to base 1 generally exist in dimensions higher than 2. If the desired rounding base is b, then he further provided a proof that showed that in 3-dimensions, controlled roundings to base 2b always exist. His proof can be extended to show that controlled roundings to base $2^{p-1}b$ always exist in p dimensions. Further, it is straightforward to show in Ernst's 3-dimensional situation and in general p-way situations that the margins are preserved to base b. This result means that if the rounding is to base 1, then the original one-way sample sizes (i.e., margins) can be preserved in any p-way stratification. Ernst's result seems to suggest that the p-way stratification problem may be unsolvable because in 3-dimensions, the best that his lemma can yield is that 0.5 (in interior cells) can be rounded to –1, 0, 1, or 2. Clearly, we cannot have a negative sample size. If a rounding to 2 exceeds the margin, then the sample allocation may exceed the available population size in a cell.

In practice, we do not need the existence of a controlled rounding to base b for every cell. Winkler's implementation of the algorithm for p-way tables uses the SAS LP procedure with capacities that allow specified lower and upper bounds on the rounding in individual p-way cells. In a three-way stratification situation with 3 variables from skewed and pair-wise uncorrelated populations distributions, he observed that if a few three-way cells having $n_{ij} > 3$ were allowed to vary by 2 and the remaining cells to vary by one, then the controlled roundings typically existed. Integer values remained as integers. Most cells were rounded to base 1. The overall algorithm converged by successively completing 20 or more roundings (matrices M_{ijt}). Further, Winkler observed that, if the individual one-way strata boundaries were varied somewhat and the one-way sample sizes changed slightly, then the algorithm completed even when the controlled roundings were to base 1 in 3-dimensions. Although Ernst's work and later work by others suggest that there are many situations when controlled roundings cannot be found, the empirical work suggests that there are many situations for which controlled roundings can be successively found.

The final facet of the p-way stratification is a method of estimating means and variances analogous to the methods of BHJ. The theorem of J. N. K. Rao (1975) provides a general setting in which variances can be estimated by first taking expectation E^1 with respect to the allocation mechanism and then taking expectation E^2 with respect to the selection mechanism given the specific allocation. For arbitrary dimensions, we use an index $\mathbf{i} \in \mathbf{I}$ for the cells in the array and obtain a representation analogous to equation (5).

$$M_{\mathbf{i}} = \sum_t p_t M_{\mathbf{i}t} \qquad (7)$$

At the first stage of the sampling, we choose an array $\boldsymbol{M_{is}}$ with probability proportion to size p_s. This gives the allocation variable $\hat{n}_{\mathbf{i}}(s)$ associated with the first stage of the sampling. At the second stage, we choose random samples of size $n_{\mathbf{i}s}$ in cell $\mathbf{i} \in \mathbf{I}$. In practice, it is possible for the allocation in a cell $n_{\mathbf{i}}$ to be 1. In those situations, we can still approximate within-cell variances by collapsing the cell with all other cells across the same dimension. For instance, in two dimensions, we collapse within the same row or column. We still bound respective single-variable variances because they are controlled by the one-way stratifications. BHJ had an analogous strategy for approximating the within-cell variances. We let $\boldsymbol{Y_i}$ be the population total in cell $\mathbf{i}$, $y_{\mathbf{i}j}$ be the quantitative value associated with the jth unit in the $\mathbf{i}^{th}$ cell, and $\mathbf{1}_{\mathbf{i}j}$ be an indicator the jth unit in cell $\mathbf{i}$ is sampled. We also let $N_{\mathbf{i}}$ be the population size in cell $\boldsymbol{i}$ and $\hat{\sigma}_{\mathbf{i}}^{\mathbf{2}}$ be an unbiased estimator of the population variance $\sigma_{\mathbf{i}}^{\mathbf{2}}$ in cell $\mathbf{i}$. Our estimator of the population total in cell $\mathbf{i}$ is given by

$$\hat{Y}_{\mathbf{i}} = (N_{\mathbf{i}} / \hat{n}_{\mathbf{i}}) \sum_j y_{\mathbf{i}j} \cdot 1_{\mathbf{i}j} \tag{8}$$

An unbiased estimator of the population total T is

$$\hat{T} = \sum_{\mathbf{i} \in \mathbf{I}} (\hat{n}_{\mathbf{i}} / n_{\mathbf{i}}) \hat{Y}_{\mathbf{i}} \tag{9}$$

Applying the main theorem of Rao (1975), we obtain the unbiased estimator of the variance

$$\hat{v} = \sum_{\mathbf{i} \in \mathbf{I}} (\hat{n}_{\mathbf{i}} / n_{\mathbf{i}} - 1)^2 \cdot \hat{Y}_{\mathbf{i}}^2 + \sum_{\mathbf{i} \neq \mathbf{j} \in \mathbf{I}} (\hat{n}_{\mathbf{i}} / n_{\mathbf{i}} - 1)\, (\hat{n}_{\mathbf{j}} / n_{\mathbf{j}} - 1)\, \hat{Y}_{\mathbf{i}}\, \hat{Y}_{\mathbf{j}} +$$

$$\sum_{\mathbf{i} \in \mathbf{I}} (2 \cdot \hat{n}_{\mathbf{i}} / n_{\mathbf{i}} - 1) N_{\mathbf{i}} (N_{\mathbf{i}} - \hat{n}_{\mathbf{i}}) \hat{\sigma}_{\mathbf{i}}^2 / \hat{n}_{\mathbf{i}} \tag{10}$$

The summations in (10) are over $\hat{n}_{\mathbf{i}} > 0$. The cells for which $\hat{n}_{\mathbf{i}} = 0$ have no sample allocated.

The algorithm giving representation (7) is applied to the fitted 5×5 matrix of Table 4 to yield Table 5. The first matrix $\mathbf{M}_0$ is the fitted matrix derived using Dykstra's GIFP. The next fifteen matrices $\mathbf{M}_k$, k = 1, ⋯, 15, are the nonnegative integer matrices obtained by the iterative integer LP procedure. The final matrix $\mathbf{M}_{16}$ is the convex sum (with the entries in the column headed by p_k used as the coefficients) of the integer matrices $\mathbf{M}_k$, k = 1, ⋯, 15. Because of space limitations, we do not present a three-way example of the iterated allocation procedure.

A key feature observed by Tepping et al. (1943) is that the allocation procedure could yield substantial reductions in the across-cell component of the variance in the types of distributions (not highly skewed) that they used. BHJ observed that their allocation procedure could also yield (possibly modest) reductions in the across-cell component of

Table 5. Example of Iterative Integer LP Procedure

k	p_k	Cell of Matrix M_k (1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(2,1)	(2,2)	(2,3)
0	0.000	2.000	2.483	1.052	0.103	0.362	2.182	1.061	1.101
1	0.140	2.000	3.000	1.000	0.000	0.000	3.000	1.000	1.000
2	0.042	2.000	3.000	1.000	0.000	0.000	3.000	1.000	1.000
3	0.007	2.000	2.000	1.000	0.000	1.000	2.000	2.000	1.000
4	0.054	2.000	3.000	1.000	0.000	0.000	2.000	2.000	1.000
5	0.029	2.000	3.000	1.000	0.000	0.000	2.000	1.000	2.000
6	0.114	2.000	3.000	1.000	0.000	0.000	2.000	1.000	1.000
7	0.104	2.000	3.000	1.000	0.000	0.000	2.000	1.000	1.000
8	0.017	2.000	2.000	2.000	0.000	0.000	2.000	1.000	1.000
9	0.035	2.000	2.000	2.000	0.000	0.000	2.000	1.000	1.000
10	0.003	2.000	2.000	1.000	0.000	1.000	2.000	1.000	1.000
11	0.032	2.000	2.000	1.000	0.000	1.000	2.000	1.000	2.000
12	0.043	2.000	2.000	1.000	0.000	1.000	2.000	1.000	1.000
13	0.237	2.000	2.000	1.000	0.000	1.000	2.000	1.000	1.000
14	0.040	2.000	2.000	1.000	0.000	1.000	2.000	1.000	2.000
15	0.103	2.000	2.000	1.000	1.000	0.000	2.000	1.000	1.000
16	1.000	2.000	2.483	1.052	0.103	0.362	2.182	1.061	1.101

k	Cell of Matrix M_k (2,4)	(2,5)	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(4,1)	(4,2)
0	1.046	0.610	0.000	1.614	1.914	2.200	1.272	0.860	0.377
1	1.000	0.000	0.000	1.000	2.000	2.000	2.000	0.000	1.000
2	1.000	0.000	0.000	1.000	2.000	2.000	2.000	1.000	0.000
3	1.000	0.000	0.000	1.000	2.000	2.000	2.000	1.000	1.000
4	1.000	0.000	0.000	1.000	2.000	2.000	2.000	1.000	0.000
5	1.000	0.000	0.000	1.000	1.000	3.000	2.000	1.000	1.000
6	1.000	1.000	0.000	1.000	2.000	3.000	1.000	1.000	1.000
7	1.000	1.000	0.000	2.000	2.000	2.000	1.000	1.000	0.000
8	1.000	1.000	0.000	2.000	1.000	3.000	1.000	1.000	1.000
9	1.000	1.000	0.000	2.000	2.000	2.000	1.000	1.000	1.000
10	2.000	0.000	0.000	2.000	2.000	2.000	1.000	1.000	1.000
11	1.000	0.000	0.000	2.000	2.000	2.000	1.000	1.000	1.000
12	2.000	0.000	0.000	2.000	2.000	2.000	1.000	1.000	0.000
13	1.000	1.000	0.000	2.000	2.000	2.000	1.000	1.000	0.000
14	1.000	0.000	0.000	2.000	1.000	3.000	1.000	1.000	0.000
15	1.000	1.000	0.000	2.000	2.000	2.000	1.000	1.000	0.000
16	1.046	0.610	0.000	1.614	1.914	2.200	1.272	0.860	0.377

Table 5. Example of Iterative Integer LP Procedure (continued)

	Cell of Matrix M_k							
k	(4,3)	(4,4)	(4,5)	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)
0	0.930	2.840	2.993	0.958	0.465	2.003	1.811	4.763
1	1.000	3.000	3.000	1.000	0.000	2.000	2.000	5.000
2	1.000	3.000	3.000	0.000	1.000	2.000	2.000	5.000
3	1.000	3.000	2.000	1.000	0.000	2.000	2.000	5.000
4	1.000	3.000	3.000	1.000	0.000	2.000	2.000	5.000
5	1.000	2.000	3.000	1.000	0.000	2.000	2.000	5.000
6	1.000	2.000	3.000	1.000	0.000	2.000	2.000	5.000
7	1.000	3.000	3.000	1.000	0.000	2.000	2.000	5.000
8	1.000	2.000	3.000	1.000	0.000	2.000	2.000	5.000
9	0.000	3.000	3.000	1.000	0.000	2.000	2.000	5.000
10	0.000	3.000	3.000	1.000	0.000	3.000	1.000	5.000
11	0.000	3.000	3.000	1.000	0.000	2.000	2.000	5.000
12	1.000	3.000	3.000	1.000	1.000	2.000	1.000	5.000
13	1.000	3.000	3.000	1.000	1.000	2.000	2.000	4.000
14	1.000	3.000	3.000	1.000	1.000	2.000	1.000	5.000
15	1.000	3.000	3.000	1.000	1.000	2.000	1.000	5.000
16	0.930	2.840	2.993	0.958	0.465	2.003	1.811	4.763

variance provided that the fitted array (n_i) had proportions that corresponded to the population proportions P_i. Because the procedure of Winkler (1990) can yield (n_i) that correspond quite closely to the population proportions P_i, it also has the potential to reduce the across-cell variances.

We are concerned about how the *p*-way stratification may improve the one-way stratifications in terms of reduced *cv*s for a given fixed sample size. We use the database of Tables 2-4. Each record contains synthetic data corresponding to volumetric data representing the sales of distillate fuel oil in residential and nonresidential end-use sectors. We use Winkler method S1 for the one-way stratification. The one-way designs are highly non-proportional. The magnitudes of the nonzero values range from 72,000 to 1 for the first variable and from 85,000 to 1 for the second. The second variable only takes nonzero values for 60 percent of the records.

To reduce *cv*s, each one-way design allocates the same 13 records with certainty. The remaining 1251 records are stratified into two different sets of five strata having a total sample allocation of 37. Each design has strata sample allocations of 6, 6, 7, 8, and 10. The *cv*s are 0.044 and 0.028, respectively. The first 13 records are also allocated with certainty under the two-way design. Single-criteria strata boundaries apportion the remaining 1251 records into a 5×5 array.

The results of the two single-criteria stratifications and the corresponding two-way design are compared with two optimal single-criteria designs (Table 6). Given the sample size of 50, the optimal designs yield minimal *cv*s for exactly one variable. To do this, the

Table 6. CVs for Single-Criteria Designs and two-way Design 1/

	CVs Var 1	CVs Var 2	Total Variation 2/
Optimal Single-Criteria			
Stratifying Var 1	.012	.334	.334
Stratifying Var 2	.407	.009	.407
Single-Criteria Designs for Multi-Way Design			
Stratifying Var 1	.044	.289	.292
Stratifying Var 2	.282	.028	.283
Multi-Way Design	.104	.041	.112

1/ Each of the second set of single-criteria designs and the two-way design allocate the same 13 population members with certainty and 37 randomly to the strata containing the remaining 1251 population members.
2/ Square root of the sum of squares of two *cv* columns.

optimal designs allocate fewer than 13 records with certainty. Sufficiently many noncertainty strata are created so that only two sample elements are allocated in each.

The first set of *four cvs* is for optimal univariate designs in which the stratifying variable agrees with one of the variables being estimated. Diagonal elements are low (0.012 and 0.009). Off-diagonal elements (0.334 and 0.407) are dramatically higher because the stratifying variables are not highly correlated with the variables for which the *cv*s are computed. Regression using the two variables yields an R-square value less than 0.2. If, however, we apply standard contingency table techniques (Bishop, Fienberg, and Holland 1975) to the underlying population matrix **N** (Table 2), we reject independence at the 95 percent level of confidence.

The second set of numbers is for one-way designs used in creating the two-way stratification. Diagonal elements (0.044 and 0.028) are higher than the diagonal elements in the first matrix. Off-diagonal entries, 0.289 and 0.282, are lower. The final set consists of the *cv*s 0.104 and 0.041 for the two-way design. They are also higher than the diagonal entries in the first matrix and approximately 1.5 times as high the diagonal entries in the second matrix. They are substantially lower than the highest of the off-diagonal entries for the respective variables (0.334 and 0.407). Using the stratification given by the first row of the first matrix, we have total variation 0.334 while the *p*-way design yields total variation 0.112. Ignoring the finite population correction (*fpc*), sample size must increase by a factor greater than 6 to equal the total variation and *cv*s of the *p*-way case.

We provide an example of estimation from a three-way design. This example involves a three-way design using three single-criteria stratifications. The database and procedures are similar to those in the previous example (Table 6) except that a third variable is included.

The third variable is both uncorrelated with the first two variables and even more highly skewed. The one-way designs had the same 15 units allocated with certainty and a

Table 7. CVs for Single-Criteria Designs and three-way Design 1/

	CVs Var 1	CVs Var 2	CVs Var 3	Total Variation 2/
Optimal Single-Criteria				
Stratifying Var 1	.001	.490	.372	.615
Stratifying Var 2	.306	.001	.351	.466
Stratifying Var 3	.604	.884	.001	1.071
Single-Criteria Designs for Multi-Way Design				
Stratifying Var 1	.047	.291	.221	.368
Stratifying Var 2	.178	.043	.194	.267
Stratifying Var 3	.207	.272	.065	.349
Multi-Way Design	.071	.055	.071	.115

1/ Each of the second set of single-criteria designs and the three-way design allocate the same 15 population members with certainty and 50 randomly to the strata containing the remaining population members.
2/ Square root of the sum of squares of three *cv* columns.

remaining 50 units allocated in 4, 4, and 2 strata that were determined by three different measures of size, respectively. Thus, the three-way design array is 4×4×2. In this situation, the allocation procedure used in obtaining equation (7) works with base 1. As we noted earlier, the procedure with a variety of marginal constraints corresponding to three one-way designs always converged provided we allowed a few of the population cells associated with larger sample sizes to vary by base 2.

Total variation ranged from 0.466 (the best in the case of standard single-criteria stratification techniques) to 0.115 (multi-way) (Table 7). Ignoring the *fpc*, we would have to increase the sample size by a factor greater than 16 to equal the total variation and *cv*s given for the multi-way design. Based on 200 independent samples, the empirical biases of the multi-way estimators of the three variables were -0.004, 0.000, and -0.004, respectively. The empirical *cv*s were 0.070, 0.056, and 0.73, and the biases of the estimates of the *cv*s were 0.001, 0.001, and -0.002, respectively.

For the three-way design example, the within-cell variance component (last summation in equation (10) contributes an average of 67, 83, and 86 percent to the total variance for variables 1, 2, and 3, respectively (Table 8). The average is based on fifteen samples. The extremes are 41 and 100 percent, 27 and 100 percent, and 45 and 100 percent for variables 1, 2, and 3, respectively. We include Table 8 because it gives results that are consistent with Tepping et al. (1943) and Bryant et al. (1959). The *p*-way stratification method is effective in significantly reducing the across-cell variance component due to cancellations. This is true even though the three individual variables are skewed.

Because the *p*-way stratification does a good job of bounding the *cv*s of several variables, it will also work well for any variable that can be expressed as an approximate linear

Table 8. Components of Variance for three-way Design Example
Fifteen Samples

	Estimated Total 1/ Variable			CV Variable			Within-Cell Variance Proportion Variable		
	1	2	3	1	2	3	1	2	3
	1,423	1,195	1,684	.101	.056	.083	.48	.99	.75
	1,384	1,297	1,576	.059	.056	.042	1.00	.92	.89
	1,337	1,269	1,450	.083	.050	.021	.54	.78	.80
	1,278	1,246	1,469	.060	.045	.035	.82	.81	.97
	1,387	1,246	1,696	.048	.059	.090	.60	.96	1.00
	1,345	1,119	1,630	.056	.074	.038	.92	.83	.95
	1,211	1,194	1,554	.061	.045	.046	.91	.84	.92
	1,460	1,226	1,636	.104	.049	.061	.50	.79	.98
	1,355	1,191	1,540	.057	.039	.052	.42	1.00	.72
	1,428	1,250	1,843	.100	.069	.098	.41	.98	.99
	1,345	1,259	1,845	.071	.087	.083	.51	.71	.99
	1,271	1,229	1,808	.063	.075	.089	.76	.62	.86
	1,424	1,336	1,550	.047	.067	.043	.95	.27	.82
	1,436	1,115	1,637	.085	.029	.054	.49	.98	.45
	1,497	1,285	1,833	.051	.066	.095	.81	.97	.88
AVER	1,372	1,230	1,650	.073	.060	.070	.67	.83	.86

1/ True values are 1354, 1254, and 1654.

combination of the variables whose *cv*s are bounded.

We now discuss the relationship of the *p*-way stratification method of this chapter to the *p*-way stratification method of Sitter and Skinner (1994, hereafter SS) that generalizes the *p*-way method of Rao and Nigam (1992, 1990). We will indicate a reason why their method improved over that of BHJ for the type of simulated data that they considered. We will also demonstrate why the methods of this paper are more general and computationally more tractable in large situations.

SS explicitly wanted to restrict their two-way designs to the form:

$$\sum_{s \in S_n} n_{ij}(s)\, p(s) = nP_{ij} \quad for\ i = 1, \cdots, n_f \ \ and\ \ j = 1, \cdots, n_g . \qquad (11)$$

Here $P_{ij} = N_{ij}/N$, n is the sample size, and S_n are designs having sample size n.

The symbol $n_{ij}(s)$ represents the integer allocations associated with each sample $s \in S_n$ such that $\sum_{i,j} n_{ij}(s) = n$ and $\sum_{s \in S_n} p(s) = 1.$

Representation (11) corresponds to the representation (7) of this chapter. A key difference and difficulty is that SS needed to know all of the integer arrays $n_{ij}(s)$ prior to running their LP algorithms. The integer arrays $n_{ij}(s)$ are part of the restraints of the LP problem. As SS indicated, when there are more constraints, equation (11) is difficult to solve using the LP method of their paper but should be easier to solve using a method similar to the method of CCE. The method of this paper solves the equivalent situation for higher dimensional situations and for populations of variables that are reasonably highly skewed. It should be far faster than the LP method of SS. The method of CCE was not extended beyond two dimensions. It is not clear how individuals could find all integer arrays $n_{ij}(s)$ corresponding to a design $n_{\mathbf{i}}$ in three or more dimensions. If we do not have all of the integer arrays $n_{ij}(s)$ prior to solving (11) using the LP method of SS, then we cannot generally solve (11). Thus, the challenges go beyond the straightforward computational difficulties pointed out by SS.

SS explicitly minimized a loss function of the form:

$$w(s) = \sum_i (n_{i\cdot}(s) - nP_{i\cdot})^2 - \sum_j (n_{\cdot j}(s) - nP_{\cdot j})^2 \qquad (12)$$

As SS indicated, there are designs for which the minimal solution of (12) is zero. The designs of this paper (even in higher dimension) yield minimal solution of zero for equation (12). SS (section 2.3 of their paper) indicated how their LP method could be extended to higher dimensions using a squared-error loss function for which different variables could be weighted by various factors. Their method does not generally yield a minimal solution of zero for the general equations with three or more summations that correspond to equation (12). SS indicated that their method is better at controlling the covariances $Cov(n_{ij}, n_{i'j'})$ which directly affect the between-cell variances. Another explanation, as shown in this paper, is that their method was better at controlling the allocated proportions n_{ij} to be closer to nP_{ij} than the method of BHJ.

3. Concluding Remarks

This chapter provides an overview of one-way and *p*-way stratification methods under Neyman (optimal) allocations of samples. The emphasis is on reducing the *cv*s of one or more variables under fixed sample size constraints. Other related methods are also described.

Disclaimer: This report is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U. S. Census Bureau. The author wishes to thank the editor, Dr. Lynn Weidman, and Dr. Tommy Wright for a number of detailed comments that led to improvements in the exposition.

References

Bishop, Y., Fienberg, S., and Holland, P. (1975), *Discrete Multivariate Analysis*, MIT Press,

Cambridge, MA.
Bryant, E., Hartley, H., and Jessen, R. (1960), "Design and Estimation in Two-Way Stratification," *Journal of the American Statistical Association*, 55, 105-124.
Causey, B., Cox, L., and Ernst, L. (1985), "Applications of Transportation Theory to Statistical Problems,"*Journal of the American Statistical Association*, 80, 903-909.
Chen, W. (1989), ”Stratification of Population: Programming of Lavallée and Hidiroglou Algorithm,” American Statistical Association, *Proceedings of the Section on Survey Research Methods*, 620-624.
Chernick, M. R. and Wright, T. (1980), “Estimation of a Population Mean with Two-Way Stratification Using a Systematic Allocation Scheme, American Statistical Association, *Proceedings of the Section on Survey Research Methods*, 185-187.
Cochran, W. (1977), *Sampling Techniques*, Third Edition, J. Wiley, New York.
Cox, L. and Ernst, L. (1982), "Controlled Rounding," *Informatics*, 20, 423-432.
Dalenius, T. and Hodges, J. L. (1959), “Minimum Variance Stratification,” *Journal of the American Statistical Association*, 54, 88-101.
Dykstra, R. L. (1985), “An Iterative Algorithm for Finding *I*-Projections on the Intersection of Convex Sets,” *Annals of Probability*, 13, 975-984.
Ekman, G. (1959), “An Approximation Useful in Univariate Stratification,” *Annals of Mathematical Statistics*, 30, 219-229.
Ernst, L. R. (1989), "Further Applications of Linear Programming to Sampling Problems," American Statistical Association, *Proceedings of the Section on Survey Research Methods*, 625-630.
Goodman, R. and Kish, L. (1950), “Controlled Selection – A Technique in Probability Sampling,” *Journal of the American Statistical Association*, 45, 350-372.
Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953*), Sample Survey Methods and Theory, Volume I*, John Wiley, New York.
Holt, D. and Smith, T. M. F. (1979), “Post Stratification,” *Journal of the Royal Statistical Society, Ser. A*., 142, 33-46.
Jarque, C. (1981), “A Solution to the Problems of Optimum Stratification in Multivariate Sampling,” *Applied Statistics*, 30, 163-169.
Jessen, R. J. (1978), *Statistical Survey Techniques*, J. Wiley, New York.
Kish, L. (1965), *Survey Sampling*, J. Wiley, New York.
Lavallée, P. and Hidiroglou, M. A. (1988), “On the Stratification of a Skewed Population,” *Survey Methodology*, 14, 33-43.
Little, R. J. A. (1993), “Post-Stratification: A Modeler’s Perspective,” *Journal of the American Statistical Assocation*, 88, 1001-1012.
Neyman, J. (1934), "On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection," *Journal of the Royal Statistical Society*, 97, 558-606.
Rao, J. N. K. (1975), "Unbiased Variance Estimation for Multistage Designs," *Sankhya, Series C*, 37, 133-139.
Rao, J. N. K. and Nigam, A. K. (1990), “Optimal Controlled Sampling Designs,” *Biometrika*, 77, 907-814.
Rao, J. N. K. and Nigam, A. K. (1992), “Optimal Controlled Sampling: a Unified Approach,” *International Statistical Review*, 60, 89-98.
Rivest, L.-P. (2002), “A Generalization of Lavallée and Hidiroglou Algorithm for Stratification in Business Surveys,” *Survey Methodology*, 28, 191-198, also at http://www.mat.ulaval.ca/pages/lpr/pdf/algo1.pdf .
Sitter, R. R. and Skinner, C. J. (1994), “Multi-way Stratification by Linear Programming,” *Survey Methodology*, 20, 65-73.
Slanta, J. and Krenzke, T. (1994), “Applying the Lavallée and Hidiroglou Method to Obtain Stratification Boundaries for the Census Bureau’s Annual Capital Expenditures Survey,” American Statistical Association, *Proceedings of the Section on Survey Research Methods*, 693-698.
Tepping, B., Hurwitz, W. N., and Deming, W. E. (1943), “On the Efficiency of Deep Stratification In Block Sampling,” *Journal of the American Statistical Association*, 38, 93-100.
Winkler, W. E. (1990), “On Dykstra’s Iterative Fitting Procedure,” *Annals of Probability*, 18, 1410-1415.

Winkler, W. E. (1998), "Strata Boundary Determination," U.S. Bureau of the Census, Statistical Research Division Report rr98/03 (at http://www.census.gov/srd/www/byyear.html)
Winkler, W. E. (2001), "Multi-way Survey Stratification and Sampling," U.S. Bureau of the Census, Statistical Research Division Report RRS 2001/01 (at http://www.census.gov/srd/www/byyear.html).