Aligarh Journal of Statistics
Vol. 27 (2007), 35-42

FORMAL ALGORITHM FOR BRYANT, HARTLEY AND JESSEN'S METHOD

Syed Shakir Ali Ghazali

ABSTRACT

A formal algorithm of a two-way stratification design described by Bryant *et al.* (1960) is presented. Also a simple modification is proposed which improves the design.

1. INTRODUCTION

Stratified random sampling is one of the most widely used sampling techniques. In practice in many surveys, it is possible to stratify the population with respect to a number of variables. But the number of strata that can be formed from the combination of several stratifying variables may be very large and possibly even greater than the permitted sample size. In the simplest case of two stratifying variables with J and K categories, the population of N units can be partitioned into $J \times K$ strata. In order to estimate the population mean, a sample of al least JK units is to be selected; if variance estimation is required, then the sample must be at least $2JK$.

Goodman and Kish (1950) described a method, called controlled selection, for drawing samples for two or more-way stratified population. Hess *et al.* (1961); Jessen (1969, 1970, 1975); Patterson (1954) and Waterton (1983) also presented multiple stratified designs. Ghazali (1996a) has shown that the procedure developed by Waterton (1983) could fail in simple cases. Some modifications were proposed to remedy the problems. Ghazali (2001) and Ghazali (1996b) has also modified Jessen's method -2 and method -3. Bryant *et al.* (1960) presented a design for two-way stratification which, for samples of size n, where n is the largest of J or K, allows unbiased estimate of the population mean; for variance estimation the required sample size is at least $2n$. They gave a verbal description but here a formal algorithm is presented. Also, a simple method for reducing discrepancies between the expected cell frequencies under proportional allocation and the expected cell frequencies under the design is proposed.

2. NOTATION

Suppose a population of N units is stratified by two stratifying variables. One variable having J categories and the other has K categories. Thus there are $J \times K$ strata cells.

Let N_{jk} be the number of population units in the $jk-$th cell and $W_{jk} = \frac{N_{jk}}{N}$ be the proportion of population units in the $jk-$th cell. Further, amongst a sample of n units let n_{jk}, $n_{j.}$ and $n_{.k}$ be the number of sample units allocated to the $jk-$th cell, $j-$th row and $k-$th column respectively. Let $E(n_{jk})$ be the average number of sample units, over all possible samples, to be drawn from the $jk-$th cell by a given sampling design.

Given the sample size n, let $E_{jk} = n\, W_{jk}$ be the expected number of units to be drawn from the $jk-$th cell under proportional stratification.

E_{jk} may be written as

$$E_{jk} = n^*_{jk} + P_{jk}, \tag{2.1}$$

where n^*_{jk} is an integer and $0 \le P_{jk} < 1$.

Similarly,

$$E_{j.} = \sum_{k=1}^{K} E_{jk} = n^*_{j.} + P_{j.} \tag{2.2}$$

and $E_{.k} = \sum_{j=1}^{J} E_{jk} = n^*_{.k} + P_{.k}$, (2.3)

where $n^*_{j.}$ and $n^*_{.k}$ are integers and $0 \le p_{j.} < 1$ and $0 \le p_{.k} < 1$.

Note that, in general, $\sum_{k=1}^{K} n^*_{jk} \ne n^*_{j.}$ and $\sum_{j=1}^{J} n^*_{jk} \ne n^*_{.k}$.

3. BRYANT, HARTLEY AND JESSEN METHOD

For a sample of size n, assume $P_{j.}$ and $P_{.k}$, as given in (2.2) and (2.3), respectively, are equal to zero for all j and k and set $n_{j.} = n^*_{j.}$ and $n_{.k} = n^*_{.k}$; it is also assumed that $n_{j.} \ge 1$ and $n_{.k} \ge 1$. In order to draw a sample by the Bryant *et al.* (1960) (BR method in what follows), given below, a square of size $n \times n$ is constructed and for each row (column) a column (row) is selected at random without replacement and "1" is placed in the corresponding cell. This process is continued until all rows and columns have exactly one cell with one "1". By adding $n_{j.}$ adjacent rows and $n_{.k}$ adjacent columns a matrix $[n_{jk}]$ of size $J \times K$ results, n_{jk} being the number of 1s in the $jk-$th cell.

4. THE ALGORITHM

The formal algorithm is presented as below:

Let U_{rs} and I_s be $0-1$ indicator variables. Then, the procedure to determine the n_{jk} is given as below:

Step 1: Set $U_{rs}=0$, $r=1,\cdots,n$; $s=1,\cdots,n$ and $I_s=0$.

Step 2: Define $A=\{s: I_s=0, 1\leq s\leq n\}$.

For $r=1$, select a value of s at random such that $s\in A$. For this chosen value of s, set $I_s=1$ and $U_{rs}=1$.

Repeat step 2 for $r=2,3,\cdots,n$.

Step 3: Let $n_{R0}=n_{C0}=0$

$$n_{Rj}=\sum_{i=1}^{j} n_{i.}\ ;\quad j=1,2,\cdots,J$$

and $n_{Ck}=\sum_{i=1}^{k} n_{.i}$; $k=1,2,\cdots,K$.

Note that, $n_{Rj}=n_{Ck}=n$.

Define,

$$Z_{js}=\sum_{r=n_{R(j-1)}+1}^{n_{Rj}} U_{rs}\ ;\quad j=1,\cdots,J$$

and $$n_{jk}=\sum_{r=n_{C(k-1)}+1}^{n_{Ck}} z_{js}\ ;\quad k=1,\cdots,K.$$

Then n_{jk} is the number of sample units to be drawn from the $jk-$th cell.

5. BIAS OF THE BIASED ESTIMATOR

Bryant *et al.* (1960) have given two estimators, biased and unbiased, of the population mean. They show that the procedure is particularly effective, compared to $1-$way stratification with respect to either of the two stratifying variables, if the population cell frequencies are proportional to both marginal frequencies i.e. if the stratifying variables are independent. However if the cell frequencies are not proportional to both marginal frequencies, the bias of the biased estimator may be of some concern. The amount of bias as given by Bryant *et al.* (1960) is

$$\text{Bias} = \frac{1}{n}\sum_{ik}\{E(n_{jk}) - E_{jk}\}\overline{Y}_{jk} ,$$

where $\overline{Y}_{jk}$ is the population mean of the $jk-$th cell. Thus, the amount of bias is due to the differences between $E(n_{jk})$ and E_{jk}. The discrepancies between the cell frequencies also inflate the variance of the unbiased estimator. Bryant *et al.* (1960) proposed a method for reducing the differences between $E(n_{jk})$ and E_{jk} (PR method in what follows). In PR method some sample units are allocated to the cells arbitrarily by an iterative procedure. The remaining sample units are allocated by the BR method. The final sample constitutes the fixed allocations plus the random allocations. The PR method will not be given in detail.

6. ALTERNATIVE METHOD FOR REDUCING DIFFERENCES BETWEEN $E(n_{jk})$ AND E_{jk}

Now a simple alternative method of correcting disproportions of cell frequencies is presented.

Let E_{jk} be as defined in (2.1) then n_{jk}^{*} is taken as the fixed allocation for the $jk-$th cell and u sample units are assigned to the cells using BR method with $u_{j.}$ and $u_{.k}$ replacing $n_{j.}$ and $n_{.k}$, respectively, where,

$$u_{j.} = \sum_{k} P_{jk} ,$$

$$u_{k.} = \sum_{j} P_{jk}$$

and $u = \sum_{k} u_{j.} = \sum_{k} u_{.k} = n - \sum_{j}\sum_{k} n_{jk}^{*}$;

u is an integer and $u_{j.}$ and $u_{.k}$ are assumed to integers.

The final sample constitutes the random allocation plus the fixed allocations.

7. COMPARISON

Now the three designs, the BR method, the PR method and the alternate method are compared. The data given by Bryant *et al.* (1960) in Table 1.2.3 are used and are presented in Table 7.1.

The expected number of sample units i.e. $E(n_{jk})$'s for the BR method are given in Table 7.2.

Table 7.1: Data from Bryant *et al.* (1960), Table 1.2.3

	E_{jk}			$E_{j.}$
	2.00	1.00	1.00	4
	0.40	0.60	1.00	2
	0.40	1.20	2.40	4
	1.20	3.60	1.20	6
	2.00	1.60	0.40	4
$E_{.k}$	6	8	6	20

Table 7.2: $E(n_{jk})$ for BR method

$E(n_{jk})$		
1.20	1.60	1.20
0.60	0.80	0.60
1.20	1.60	1.20
1.80	2.40	1.80
1.20	1.60	1.20

Table 7.3a: Fixed allocations by PR method

	m_{jk}			$m_{j.}$
	1	0	0	1
	0	0	0	0
	0	0	1	1
	0	1	0	1
	1	0	0	1
$m_{.k}$	2	1	1	4

Table 7.3b: Number of units to be allocated by BR method

	$u_{jk} = E_{jk} - m_{jk}$			$u_{j.}$
	1.00	1.00	1.00	3
	0.40	0.60	1.00	3
	0.40	1.20	1.40	3
	1.20	2.60	1.20	5
	1.00	1.60	0.40	3
$u_{.k}$	4	7	5	16

Table 7.3a and Table 7.3b show the fixed allocations, denoted by m_{jk}, obtained by PR method and the remaining sample units to be allocated randomly by the BR method respectively. For this case, the expected numbers of sample units for the whole sample are given in Table 7.3c.

Table 7.3c: $E(n_{jk})$ for PR method

$E(n_{jk})$		
1.75	1.31	0.94
0.50	0.88	0.63
0.75	1.31	1.94
1.25	3.19	1.56
1.75	1.31	0.94

Table 7.4a: Fixed allocations by alternative method

	m_{jk}			$m_{j.}$
	2	1	1	4
	0	0	1	1
	0	1	2	3
	1	3	1	5
	2	1	0	3
$m_{.k}$	5	6	5	16

Table 7.4b: Number of units to be allocated by BR method

	$u_{jk} = E_{jk} - m_{jk}$			$u_{j.}$
	0.00	0.00	0.00	0
	0.40	0.60	0.00	1
	0.40	0.20	0.40	1
	0.20	0.60	0.20	1
	0.00	0.60	0.40	1
$u_{.k}$	1	2	1	4

The fixed allocations obtained by using the alternative method are given in Table 7.4a and Table 7.4b contains the remaining sample units to be allocated by BR method. The average number of sample units in the $jk-$th cell for the whole sample by alternative method are shown in Table 7.4c.

Table 7.4c: $E(n_{jk})$ for alternative method

$E(n_{jk})$		
2.00	1.00	1.00
0.25	0.50	1.25
0.25	1.50	2.25
1.25	3.50	1.25
2.25	1.50	0.25

In order to compare the Table 7.2, 7.3c and 7.4c, we compute $D_r = \sum_{jk} \{E(n_{jk}) - E_{jk}\}^2$, $(r = 1, 2, 3)$ for each of these tables; the results are given below:

$D_1 = 6.96$

$D_2 = 1.48$

$D_3 = 0.34$.

Although the PR method reduces the discrepancies, the alternative method does even better and it is relatively simple. Further, note that alternative method also limits the deviations between the number of units allocated to a cell in a sample and the respective E_{jk}. If using Table 7.4b, 4 sample units are allocated to cells randomly; at most one unit would be allocated to any cell because $n_{j.} \leq 1$ for all j and, therefore, for any sample drawn by alternative method, $|n_{jk} - E_{jk}| < 1$.

Acknowledgement

The author would like to acknowledge with gratitude M.H. Tahir for careful reading of the manuscript, and also pointing out some corrections and missing references, which led to the improvement in this paper.

REFERENCES

Bryant, E.C., Hartley, H.O. and Jessen, J.R. (1960): Design and estimation in two-way stratification. *J. Amer. Statist. Assoc.*, **55**, 105-124.

Ghazali, S.S.A. (1996a): Modification of Waterton's controlled selection method. *Statistician*, **45,** 237-242.

Ghazali, S.S.A. (1996b): Modification of Jessen's method-3. *Pakistan J. Statist.*, **12,** 179-187.

Ghazali, S.S.A. (2001): Modification of Jessen's method-2. *J. Stat. Comput. Simul.*, **70,** 45-54.

Goodman, R. and Kish, L. (1950): Controlled selection: a technique in probability sampling. *J. Amer. Statist. Assoc.*, **45**, 350-372.

Hess, I., Riedel, D.C. and Fitzpatric, T.B. (1961): Probability sampling of hospitals and patients. *Ann. Arbor. Michigan, Bureau of Hospital Administration.*

Jessen, R.J. (1969): Some methods of probability non-replacement sampling. *J. Amer. Statist. Assoc.*, **64**, 175-193.

Jessen, R.J. (1970): Probability sampling with marginal constraints. *J. Amer. Statist. Assoc.*, **65**, 776-796.

Jessen, R.J. (1975): Square and cubic lattice sampling. *Biometrics*, **31**, 449-471.

Patterson, H.D. (1954): The errors of lattice sampling. *J. Roy. Statist. Soc. Ser. B*, **16,** 140-149.

Waterton, J.J. (1983): An exercise in controlled selection. *Appl. Statist.*, **32**, 150-164.

Received : 13-05-2006

Post-Graduate Department of Statistics,
Govt. S.E. College Bahawalpur,
Bahawalpur, Pakistan.
e-mail: shakir_ali_ghazali@yahoo.com