ALLOCATION IN STRATIFIED SAMPLING BASED ON PRELIMINARY TEST OF SIGNIFICANCE II

Victor K. T. Tang and B. V. Sukhatme
California State University, Humboldt and Iowa State University, Ames

I. Introduction

Let a population consisting of N units be classified into k strata, the i-th stratum having N_i units $i = 1,2,\ldots,k$ so that $\sum_{i=1}^{k} N_i = N$. Let Y be the characteristic under study and consider the problem of estimating the population mean $\bar{y}_N = \frac{1}{N}\sum_{i=1}^{N} y_i$ from a stratified random sample of size $n = \sum_{i=1}^{k} n_i$ where n_i units are drawn by simple random sampling without replacement from the i-th stratum $i = 1,2,\ldots,k$. An unbiased estimate of the mean $\bar{y}_N$ is given by

$$\bar{y}_{st} = \sum_{i=1}^{k} W_i \, \bar{y}_{n_i} \tag{1.1}$$

where W_i is the proportion of units in the i-th stratum and $\bar{y}_{n_i}$ is the simple mean unbiased estimate of $\bar{y}_{N_i}$, the mean for the i-th stratum. If N_i is so large that $\frac{N_i}{N_i - 1} \simeq 1$, $V(\bar{y}_{st})$ can be written as

$$V(\bar{y}_{st}) = \sum_{i=1}^{k} \frac{W_i^2 \sigma_i^2}{n_i} - \frac{1}{N}\sum_{i=1}^{k} W_i \sigma_i^2 \, . \tag{1.2}$$

If the total sample size n is fixed in advance, the classical problem of allocation of sample sizes in stratified sampling is to determine a vector $(n_1, n_2, \ldots, n_k)$ of k non-negative integers such that $\sum_{i=1}^{k} n_i = n$ and for which $V(\bar{y}_{st})$ is minimum. The allocation so determined, commonly known as Neyman allocation (Neyman, 1934) is given by

$$n_i = nW_i\sigma_i / \sum_{i=1}^{k} W_i\sigma_i \, . \tag{1.3}$$

Neyman allocation however depends on strata variances σ_i^2 which are generally not known. One way out of this difficulty (Sukhatme, P. V., 1935) is to draw an initial sample of fixed size m from each stratum to estimate σ_i^2 which in turn are used to estimate n_i from (1.3). In this case, n_i is estimated by

$$n_i = nW_i s_i / \sum_{i=1}^{k} W_i s_i \tag{1.4}$$

where s_i^2 is an unbiased estimate of σ_i^2. The allocation (1.4) will be called Modified Neyman allocation.

Another allocation which is frequently used in practice and does not require knowledge of strata variances σ_i^2 is proportional allocation in which n_i is given by

$$n_i = nW_i \, . \tag{1.5}$$

If the strata variances σ_i^2 do not differ significantly among themselves, modified Neyman allocation may turn out to be less efficient than proportional allocation (Evans, 1951).

Before deciding on the method of allocation, it is therefore proposed to carry out a preliminary test of significance concerning the homogeneity of strata variances. If on the basis of the test of significance the strata variances are found to be homogeneous, the sample sizes to be drawn from the different strata will be determined according to proportional allocation. This allocation based on preliminary test of significance will be called 'sometimes proportional allocation'. In an earlier paper (Sukhatme, B.V. and Tang, 1970), some results concerning the efficiency of sometimes proportional allocation with respect to proportional allocation and modified Neyman allocation were presented for the relatively simple case of two strata when $\sigma_1^2 \leq \sigma_2^2$. In this paper, some further results are presented concerning the efficiency of sometimes proportional allocation and optimum choice of level of significance for the case of two strata when $\sigma_1^2 \neq \sigma_2^2$. The results for three or more strata will be presented at a later date.

2. Variance of $\bar{y}_{st}$ under sometimes proportional allocation

The sometimes proportional allocation is defined as

$$\begin{aligned} n_i &= nW_i \qquad \text{if } \frac{s_j^2}{s_i^2} < \lambda \quad \text{for } i,j = 1,2 \text{ and } i \neq j \\ &= n\frac{W_i s_i}{\sum_{i=1}^{2} W_i s_i} \qquad \text{otherwise,} \end{aligned} \tag{2.1}$$

where λ is a known constant. Let the event A_0' be defined by $\{\frac{s_j^2}{s_i^2} < \lambda$ for $i,j = 1,2$ and $i \neq j\}$ and A_1' be the complementary event of A_0'. The variance of $\bar{y}_{st}$ is given by

$$V(\bar{y}_{st})_S = \sum_{i=0}^{1} E_i\{V(\bar{y}_{st}|A_i')\}P(A_i'), \tag{2.2}$$

where E_i denotes that the expectations are taken with reference to the set A_i' and S stands for sometimes proportional allocation. To evaluate the variance, it will be assumed that $\frac{(m-1)s_i^2}{\sigma_i^2}$ is approximately distributed as chi-square with $t = m-1$ degrees of freedom. It can be seen that

$$V(\bar{y}_{st})_S = \frac{\sigma_1^2}{n}(W_1^2 + W_2^2\theta_{21}) - \frac{\sigma_1^2}{N}(W_1 + W_2\theta_{21})$$
$$+ \frac{W_1W_2}{n}\sigma_1^2[(1+\theta_{21})\{I_{p_{21}^{(1)}}(\tfrac{t}{2}, \tfrac{t}{2})$$
$$- I_{p_{21}}(\tfrac{t}{2}, \tfrac{t}{2})\} + G\theta_{21}^{1/2}\{I_{q_{21}^{(1)}}(\tfrac{t-1}{2}, \tfrac{t-1}{2})$$
$$+ I_{p_{21}}(\tfrac{t-1}{2}, \tfrac{t-1}{2})\}] , \qquad (2.3)$$

where $\theta_{21} = \frac{\sigma_2^2}{\sigma_1^2}$, $p_{21} = \frac{\theta_{21}}{\lambda + \theta_{21}}$,

$$q_{21}^{(1)} = 1 - p_{21}^{(1)} = \frac{1}{1 + \lambda\theta_{21}} ,$$

$G = \frac{2\Gamma(\frac{t-1}{2})}{\{\Gamma(\frac{t}{2})\}^2}$ and $I_{\cdot}(.,.)$ is the incomplete beta distribution.

If we let $\lambda \to \infty$, we obtain the variance of $\bar{y}_{st}$ under proportional allocation, namely,

$$V(\bar{y}_{st})_P = (\tfrac{1}{n} - \tfrac{1}{N})\sigma_1^2(W_1 + W_2\theta_{21}) , \qquad (2.4)$$

where P stands for proportional allocation.

If we put $\lambda = 1$, we get the variance of $\bar{y}_{st}$ under modified Neyman allocation, namely,

$$V(\bar{y}_{st})_{\underline{N}} = \frac{\sigma_1^2}{n}(W_1^2 + W_2^2\theta_{21}) - \frac{\sigma_1^2}{N}(W_1 + W_2\theta_{21})$$
$$+ \frac{W_1W_2}{n}\sigma_1^2 G\theta_{21}^{1/2} , \qquad (2.5)$$

where $\underline{N}$ stands for modified Neyman allocation.

3. Efficiency of sometimes proportional allocation

We shall first discuss the relative efficiency of sometimes proportional allocation with respect to proportional allocation. If $e_1^*(\lambda, \theta_{21})$ denotes the relative efficiency of sometimes proportional allocation with respect to proportional allocation, it is easy to see that

$$e^*(\lambda, \theta_{21}) = \frac{V(\bar{y}_W)_P}{V(\bar{y}_{st})_S}$$

$$= 1/\Big\{1 - \frac{W_1W_2}{W_1 + W_2\theta_{21}}[(1+\theta_{21})\{I_0(p_{21}) \qquad (3.1)$$
$$+ I_0(q_{21}^{(1)})\} - G\theta_{21}^{1/2}\{I_{-\frac{1}{2}}(p_{21}) + I_{-\frac{1}{2}}(q_{21}^{(1)})\}]\Big\} ,$$

where $I_i(\alpha) = I_\alpha(\frac{t}{2} + i, \frac{t}{2} + i)$. Clearly, if $e_1^*(\lambda, \theta_{21}) \geq 1$, sometimes proportional allocation is at least as efficient as proportional allocation. We shall now present some results concerning the behavior of the efficiency function $e_1^*(\lambda, \theta_{21})$.

We shall first consider the case when λ is an arbitrary but fixed number. Then it can be seen that for any given $\lambda \geq 1$,

i) $\lim_{\theta_{21} \to 0} e_1^*(\lambda, \theta_{21}) > 1$,

ii) $\lim_{\theta_{21} \to 1} e_1^*(\lambda, \theta_{21}) \leq 1$,

iii) $\frac{\delta}{\delta\theta_{21}} e_1^*(\lambda, \theta_{21}) < 0$ for $0 < \theta_{21} < 1$

iv) $\exists\ \theta' > 1$ such that $\frac{\delta}{\delta\theta_{21}} e_1^*(\lambda, \theta_{21}) > 0$ for every $\theta_{21} > \theta'$ provided

$$1 - \frac{G}{2} + \frac{(\lambda^{t-\frac{1}{2}} + 1)(\lambda^{\frac{1}{2}} - 1)}{\lambda^{\frac{t}{2}} B(\frac{t}{2}, \frac{t}{2})} > 0 ,$$

and

v) $\lim_{\theta_{21} \to \infty} e_1^*(\lambda, \theta_{21}) > 1$.

As a consequence of the above, we obtain the following result.

Theorem 3.1 Let $\lambda \geq 1$ be an arbitrary but fixed number such that

$$1 - \frac{G}{2} + \frac{(\lambda^{t-\frac{1}{2}} + 1)(\lambda^{\frac{1}{2}} - 1)}{\lambda^{\frac{t}{2}} B(\frac{t}{2}, \frac{t}{2})} > 0 .$$

Then $\exists\ \theta_0^{(1)}$ in (0, 1) and $\theta_0^{(2)} > 1$ such that

$$e_1^*(\lambda, \theta_0^{(1)}) = e_1^*(\lambda, \theta_0^{(2)}) = 1$$

and

$$e_1^*(\lambda, \theta_{21}) > 1 \quad \forall\ \theta_{21} < \theta_0^{(1)} \text{ or } \theta_{21} > \theta_0^{(2)} .$$

Theorem 3.1 assures us that there exist $\theta_0^{(1)}$ between 0 and 1 and $\theta_0^{(2)}$ larger than 1 such that for each $\theta_{21} < \theta_0^{(1)}$ or $\theta_{21} > \theta_0^{(2)}$, sometimes proportional allocation is always more efficient than proportional allocation.

We shall now consider the case when Θ_{21} is an arbitrary but fixed number less than $\frac{1}{2}(G^2 - 2 - G\sqrt{G^2 - 4})$ or larger than $\frac{1}{2}(G^2 - 2 + G\sqrt{G^2 - 4})$. Then it is easy to see that $e_1^*(0, \Theta_{21}) > 1$. Further, $e_1^*(\lambda, \Theta_{21})$ tends to 1 as its horizontal asymptote from below. It is clear that there exists λ_0 such that $e_1^*(\lambda, \Theta_{21}) > 1$ for every $\lambda < \lambda_0$. We have thus proved the following result.

<u>Theorem 3.2</u> Let Θ_{21} be an arbitrary but fixed number less than $\frac{1}{2}(G^2 - 2 - G\sqrt{G^2 - 4})$ or larger than $\frac{1}{2}(G^2 - 2 + G\sqrt{G^2 - 4})$. Then $\exists\ \lambda_0$ such that

$$e_1^*(\lambda_0, \Theta_{21}) = 1$$

and

$$e_1^*(\lambda, \Theta_{21}) > 1 \quad \text{for every} \quad \lambda < \lambda_0 .$$

After having obtained the above results, it is now possible to prove the existence of a pair of numbers $(\lambda_1^*, \lambda_2^*)$ with $\lambda_1^* \leq \lambda_2^*$ such that for each λ outside the interval $(\lambda_1^*, \lambda_2^*)$ the relative efficiency of sometimes proportional allocation with respect to proportional allocation is never less than a preassigned value $e_0 < 1$. The result is stated in Theorem 3.3 without proof.

<u>Theorem 3.3</u> Let e_0 be a real number such that $0 < e_0 < 1$. Then $\exists\ \lambda_1^* \leq \lambda_2^*$ such that $e_1^*(\lambda, \Theta_{21}) \geq e_0$ for every λ outside the interval $(\lambda_1^*, \lambda_2^*)$.

We shall now discuss the relative efficiency of sometimes proportional allocation with respect to modified Neyman allocation which is given by

$$e_2^*(\lambda, \Theta_{21}) = \frac{V(\bar{y}_{st})_{\underline{N}}}{V(\bar{y}_{st})_S}$$

$$\doteq 1/[1 - \frac{W_1W_2D^*}{W_1^2 + W_2^2\Theta_{21} + W_1W_2G\Theta_{21}^{1/2}}] ,$$

where

$$D^* = G\Theta_{21}^{1/2}\{I_{-\frac{1}{2}}(p_{21}^{(1)}) - I_{-\frac{1}{2}}(p_{21})\}$$

$$- (1 + \Theta_{21})\{I_0(p_{21}^{(1)}) - I_{-\frac{1}{2}}(p_{21})\} .$$

The results concerning the behavior of $e_2^*(\lambda, \Theta_{21})$ can be obtained in a similar manner and are stated below.

<u>Theorem 3.4</u> Let $\lambda > 1$ be an arbitrary but fixed number. $\exists\ \Theta_0^{(1)}$ in $(0, 1)$ and $\Theta_0^{(2)} > 1$ such that

$$e_2^*(\lambda, \Theta_0^{(1)}) = e_2^*(\lambda, \Theta_0^{(2)}) = 1$$

and

$$e_2^*(\lambda, \Theta_{21}) \geq 1 \quad \text{for every} \quad \Theta_0^{(1)} \leq \Theta_{21} \leq \Theta_0^{(2)} .$$

<u>Theorem 3.5</u> Let Θ_{21} be an arbitrary but fixed number such that $\frac{1}{2}(G^2 - 2 - G\sqrt{G^2 - 4}) \leq \Theta_{21} \leq \frac{1}{2}(G^2 - 2 + G\sqrt{G^2 - 4})$. Then $\exists\ \lambda_0$ such that

$$e_2^*(\lambda_0, \Theta_{21}) = 1$$

and

$$e_2^*(\lambda, \Theta_{21}) \geq 1 \quad \text{for every} \quad \lambda \geq \lambda_0 .$$

<u>Theorem 3.6</u> Let e_0 be a real number such that $0 < e_0 < 1$. Then $\exists\ \lambda_1^* \leq \lambda_2^*$ such that $e_2^*(\lambda, \Theta_{21}) \geq e_0$ for every λ outside the interval $(\lambda_1^*, \lambda_2^*)$.

4. <u>Optimum choice of the level of significance of the preliminary test</u>

As we have seen in Section 3, the relative efficiency of sometimes proportional allocation with respect to proportional allocation as also modified Neyman allocation depends on W_1, Θ_{21} and λ. Generally W_1 is known while Θ_{21} is not known. The question naturally arises concerning the choice of the level of significance as determined by λ. We would like to choose that value of λ for which the relative efficiency of sometimes proportional allocation with respect to either of the other two allocations is as high as possible. For example, if Θ_{21} is likely to be very much different from 1, it would seem desirable to choose λ as small as possible. If on the other hand, Θ_{21} is likely to be closer to 1, it would seem desirable to choose λ as large as possible. If however, nothing is known concerning the likely range of values of Θ_{21}, difficulty arises concerning the choice of λ. Theorems 3.3 and 3.6 provide useful results from this point of view. Let e_0 be a real number such that $0 < e_0 < 1$. Then we shall restrict our choice to those values of λ for which the relative efficiency of sometimes proportional allocation with respect to proportional allocation or modified Neyman allocation is at least e_0. Using this criterion, the following sets of values of λ are obtained for different values of m.

Within a particular set of values of λ, we shall choose that value of λ for which gain in efficiency of sometimes proportional allocation with respect to either of the other two allocations is maximum. Using this criterion, certain values for λ have been suggested in the last col-

Table 1

Sets of values of λ for which the relative efficiency of sometimes proportional allocation with respect to either of the other two allocations is at least e_0

m	e_0	Set of values of λ	Suggested value of λ
6	0.96	$3.3 \leq \lambda \leq 4.0$	3.3
7	0.97	$3.2 \leq \lambda \leq 3.7$	3.2
8	0.98	3.1	3.1
9	0.98	$2.6 \leq \lambda \leq 3.0$	2.6

Table 2

Relative efficiency of sometimes proportional allocation

		Θ_{21}								
λ	with respect to	0.4	0.7	1.0	3.0	5.0	7.0	9.0	11.0	13.0
2.6	Proportional allocation	1.003	0.984	0.982	1.017	1.076	1.131	1.178	1.218	1.254
	Modified Neyman allocation	0.992	1.004	1.009	0.990	0.990	0.993	0.995	0.997	0.998
3.1	Proportional allocation	0.998	0.983	0.981	1.009	1.065	1.119	1.166	1.208	1.245
	Modified Neyman allocation	0.991	1.008	1.012	0.986	0.983	0.986	0.990	0.992	0.994
3.2	Proportional allocation	0.993	0.978	0.976	1.004	1.058	1.111	1.158	1.198	1.235
	Modified Neyman allocation	0.992	1.008	1.012	0.989	0.983	0.985	0.980	0.990	0.992
3.3	Proportional allocation	0.986	0.969	0.967	0.998	1.049	1.100	1.145	1.186	1.222
	Modified Neyman allocation	0.995	1.009	1.013	0.990	0.983	0.984	0.986	0.988	0.990

umn of Table 1. Table 2 gives the relative efficiency averaged over W_1 of sometimes proportional allocation with respect to proportional allocation as also with respect to modified Neyman allocation for suggested values of λ over a wide range of values of Θ_{21}.

It is seen that sometimes proportional allocation is almost as efficient as modified Neyman allocation. It is also seen that sometimes proportional allocation is almost as efficient as proportional allocation for values of Θ_{21} close to 1 while it is considerably more efficient than proportional allocation for values of Θ_{21} away from 1.

Acknowledgement

This work was supported in part under OEC Contract No. 0-70-3655 by the U. S. Office of Education, Department of Health, Education and Welfare.

References

[1] Evans, W. D. (1934). "On stratification and optimum allocation," Journal of the American Statistical Association, 46, 95-104.

[2] 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.

[3] Sukhatme, B. V. and Tang, Victor K. T. (1970). "Allocation in stratified sampling based on preliminary test of significance," American Statistical Association Proceedings of the Social Statistics Section, 368-371.

[4] Sukhatme, P. V. (1935). "Contribution to the theory of the representative method," Journal of the Royal Statistical Society, Supplement 2, 253-268.