ALLOCATION IN STRATIFIED SAMPLING BASED ON PRELIMINARY TEST OF SIGNIFICANCE

B. V. Sukhatme and Victor K. T. Tang
Iowa State University and Humboldt State College

1. Introduction

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

$$\bar{y}_W = \sum_{i=1}^{k} W_i \bar{y}_{n_i} \qquad (1.1)$$

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

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

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 $V(\bar{y}_W)$ is minimum. The allocation so determined, commonly known as Neyman allocation (Neyman, 1934) is given by

$$n_i = nW_i\sigma_i / \sum_{i=1}^{k} W_i \sigma_i \quad . \qquad (1.3)$$

Neyman allocation however depends on strata variances σ_i^2 which are generally not known. One way out of this difficulty (Sukhatme, 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

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

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. 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'. This paper will consider in detail the sometimes proportional allocation and discuss its efficiency with respect to proportional allocation and modified Neyman allocation for the relatively simple case of two strata when $\sigma_1^2 \leq \sigma_2^2$. The results for three or more strata will be presented in a separate communication.

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

The sometimes proportional allocation may be defined as

$$\left.\begin{aligned} n_i &= nW_i & &\text{if } \frac{s_2^2}{s_1^2} < \lambda \\ &= nW_i s_i / \sum_{i=1}^{2} W_i s_i & &\text{otherwise} \end{aligned}\right\} \qquad (2.1)$$

where λ is a known constant. Clearly, the variance of $\bar{y}_W$ is given by

$$V(\bar{y}_W)_S = E\{V(\bar{y}_W \mid \frac{s_2^2}{s_1^2} < \lambda)\}\, P(\frac{s_2^2}{s_1^2} < \lambda)$$

$$+ E\{V(\bar{y}_W \mid \frac{s_2^2}{s_1^2} \geq \lambda)\}\, P(\frac{s_2^2}{s_1^2} \geq \lambda) \qquad (2.2)$$

where the expectation in each term is taken with reference to the corresponding set and S stands for sometimes proportional allocation. To evaluate the various terms, it will be assumed that $\frac{(m-1)s_i^2}{\sigma_i^2}$ is distributed as chi-square with f = m-1 degrees of freedom. It can then be seen that

$$V(\bar{y}_W)_S = \frac{\sigma_1^2}{n}(W_1^2 + W_2^2\,\theta_{21}) - \frac{\sigma_1^2}{n}(W_1 + W_2\,\theta_{21})$$

$$+ \frac{W_1W_2\sigma_1^2}{n}\,[(1+\theta_{21})\,I_{q_{21}}(\tfrac{f}{2}, \tfrac{f}{2})$$

$$+ G\,\theta_{21}^{1/2}\,I_{p_{21}}(\tfrac{f-1}{2}, \tfrac{f-1}{2})] \qquad (2.3)$$

where $\theta_{21} = \sigma_2^2/\sigma_1^2$, $p_{21} = \theta_{21}/(\lambda + \theta_{21})$, $q_{21} = 1 - p_{21}$, $G = 2\Gamma(\frac{f-1}{2})/[\Gamma(\frac{f}{2})]^2$ and $I_{\cdot}(\cdot,\cdot)$ is the incomplete beta distribution.

If we let $\lambda \longrightarrow \infty$, we obtain the variance under proportional allocation, namely,

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

where P stands for proportional allocation.

If we put $\lambda = 0$, we get the variance under modified Neyman allocation, namely

$$V(\bar{y}_W)_{\underline{N}} = \frac{\sigma_1^2}{n}(W_1^2 + W_2^2\,\theta_{21}) - \frac{\sigma_1^2}{N}(W_1 + W_2\theta_{21})$$

$$+ \frac{W_1W_2}{n}\,\sigma_1^2\,G\,\theta_{21}^{1/2} \qquad (2.5)$$

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

$$e_1(\lambda, \theta_{21}) = \frac{V(\bar{y}_W)_P}{V(\bar{y}_W)_S}$$

$$= 1/\,[1 - \frac{W_1W_2}{W_1+W_2\theta_{21}}\,\{(1+\theta_{21})I_{p_{21}}(\tfrac{f}{2}, \tfrac{f}{2})$$

$$- G\,\theta_{21}^{1/2}\,I_{p_{21}}(\tfrac{f-1}{2}, \tfrac{f-1}{2})\}\,] \qquad (3.1)$$

Clearly, if $e_1(\lambda, \theta_{21}) \geq 1$, sometimes proportional allocation is at least as efficient as proportional allocation. We shall now obtain some results concerning the behavior of the efficiency function. We shall first consider the case when λ is an arbitrary but fixed number such that

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

Then it can be seen that

i) $\lim_{\theta_{21} \to 1} e_1(\lambda, \theta_{21}) \leq 1$

ii) $\exists\, \theta' \ni \frac{\partial}{\partial\theta_{21}} e_1(\lambda, \theta_{21}) > 0$ for every $\theta_{21} > \theta'$

and

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

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

<u>Theorem 3.1.</u> Let λ be an arbitrary but fixed number in the set $[0, \infty)$ such that

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

Then $\exists\, \theta_0 \ni e_1(\lambda, \theta_0) = 1$ and $e_1(\lambda, \theta_{21}) > 1$ for every $\theta_{21} > \theta_0$.

Theorem 3.1 assures us that there exists a θ_0 such that for each $\theta_{21} > \theta_0$, 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 greater than or equal to $\frac{1}{2}(G^2 - 2 + G\sqrt{G^2 - 4})$. Then it is easy to see that $e_1(0, \theta_{21}) > 1$. Further, it can be shown that $e_1(\lambda, \theta_{21})$ is increasing if $\lambda < 1$ or $\lambda > \theta_{21}^2$ and decreasing for $1 < \lambda < \theta_{21}^2$. It follows that $e_1(\lambda, \theta_{21})$ reaches its maximum at $\lambda = 1$ and its minimum at $\lambda = \theta_{21}^2$. Also, it is not difficult to see that $\lim_{\lambda \to \infty} e_1(\lambda, \theta_{21}) = 1^-$. It is now 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 greater than or equal to $\frac{1}{2}(G^2 - 2 + G\sqrt{G^2 - 4})$. Then $\exists\, \lambda_0 \ni e_1(\lambda_0, \theta_{21}) = 1$ and $e_1(\lambda, \theta_{21}) > 1$ for every $\lambda < \lambda_0$.

Armed with the above results, it is now possible to prove the existence of a pair $(\lambda_1^*, \lambda_2^*)$ with $\lambda_1^* \leq \lambda_2^*$ such that for every 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 $e_0 < 1$. This result is stated in Theorem 3.3.

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

Proof: To prove the theorem, let θ_{21} be a fixed number greater than or equal to 1.

First consider the case when $\inf_{\lambda} e_1(\lambda, \theta_{21}) \geq e_0$. Then $e_1(\lambda, \theta_{21}) \geq e_0$ for every λ. If we take $\lambda_1^* = \lambda_2^*$ to be any real number greater than 1, then the theorem is true.

Now consider the case when $\inf_{\lambda} e_1(\lambda, \theta_{21}) < e_0$. Then for some values of λ, $e_1(\lambda, \theta_{21}) < e_0$. But $e_1(\lambda, \theta_{21})$ is decreasing when $1 < \lambda < \theta_{21}^2$. Also $\lim_{\lambda \to 1} e_1(\lambda, \theta_{21}) > 1$. It follows that $\exists \underline{\lambda} \ni e_1(\lambda, \theta_{21}) \geq e_0$ for every $\lambda < \underline{\lambda}$.

Let $L_1 = \{\underline{\lambda} \mid e_1(\lambda, \theta_{21}) \geq e_0$ for every $\lambda < \underline{\lambda}$ and $\theta_{21} \geq 1$ and fixed$\}$.

Clearly $\inf_{\theta_{21} \geq 1} L_1$ is the required λ_1^*. Q.E.D.

On the other hand, since $e_1(\lambda, \theta_{21})$ is increasing when $\lambda > \theta_{21}^2$ and $\lim_{\lambda \to \infty} e_1(\lambda, \theta_{21}) = 1^-$, $\exists \bar{\lambda} \ni e_1(\lambda, \theta_{21}) \geq e_0$ for every $\lambda > \bar{\lambda}$. Let

$L_2 = \{\bar{\lambda} \mid e_1(\lambda, \theta_{21}) \geq e_0$ for every $\lambda > \bar{\lambda}$ and $\theta_{21} \geq 1$ and fixed$\}$

clearly, $\sup_{\theta_{21} \geq 1} L_2$ is the required λ_2^*.

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

$$e_2(\lambda, \theta_{21}) = V(\bar{y}_W)_{\underline{N}} / V(\bar{y}_W)_S = \frac{1}{D} [W_1^2 + W_2^2 \theta_{21} + W_1 W_2 G \theta_{21}^{1/2}] \tag{3.2}$$

where

$$D = (W_1^2 + W_2^2 \theta_{21} + W_1 W_2 G \theta_{21}^{1/2}) - W_1 W_2 G \theta_{21}^{1/2} I_{q_{21}}\left(\frac{f-1}{2}, \frac{f-1}{2}\right) - W_1 W_2 (1 + \theta_{21}) I_{q_{21}}\left(\frac{f}{2}, \frac{f}{2}\right).$$

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

Theorem 3.4. Let λ be an arbitrary but fixed number in the set $[0, \infty)$. Then $\exists \theta_0 \geq 1 \ni e_2(\lambda, \theta_0) = 1$ and $e_2(\lambda, \theta_{21}) \geq 1$ for every $\theta_{21} \leq \theta_0$.

Theorem 3.5. Let θ_{21} be an arbitrary but fixed number greater than or equal to $\frac{1}{2}[G^2 - 2 + G\sqrt{G^2 - 4}]$. Then $\exists \lambda_0 \ni e_2(\lambda_0, \theta_{21}) = 1$ and $e_2(\lambda, \theta_{21}) \geq 1$ for every $\lambda \leq \lambda_0$.

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

4. Numerical illustration

For the purpose of illustration, consider the problem of sampling households in a town in order to estimate the average amount of assets per household that are readily convertible into cash. The households are stratified into a high-rent and a low-rent stratum. The variance σ_2^2 in the high-rent stratum is considerably larger than the variance σ_1^2 in the low-rent stratum. On the basis of preliminary evidence, it is guessed that $\theta_{21} \leq 9$. It is known that

$N = 24{,}000$, $W_1 = 5/6$ and $W_2 = 1/6$

N_1 and N_2 are sufficiently large, so that finite correction factors can be ignored. Further, let $f = 7$ and $\lambda = 2$. The table below gives the relative efficiency of sometimes proportional allocation with respect to proportional allocation as also with respect to modified Neyman allocation for different values of θ_{21}.

Relative efficiency of sometimes proportional allocation

With respect to	$\theta_{21}=1$	$\theta_{21}=3$	$\theta_{21}=5$	$\theta_{21}=7$	$\theta_{21}=9$
Proportional Allocation	0.99	1.02	1.10	1.18	1.26
Modified Neyman Allocation	1.014	0.998	0.995	0.997	0.998

It is seen that for appropriate choice of the level of significance as determined by λ (in this case λ=2), 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} closer to 9.

Acknowledgement

This work was supported 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. (1951). "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, P. V. (1935). "Contribution to the theory of the representative method," Journal of the Royal Statistical Society, Supplement 2, 253-268.