ADD-ONS: OBTAINING DESIRED STRATUM SIZES WITH POST-STRATIFICATION

Martin Rosenzweig, University of Pennsylvania

INTRODUCTION

The use of post-stratification may be dictated by basically two situations in practice: First, a reasonably useful population list, or frame, exists, but it is not organized for stratification. Second, the frame may not list the characteristic to be used to stratify the sample.

In the first case, the standard procedure is to re-organize the frame into strata, edit the new frame, and sample this new frame according to some specified sampling plan. However, resource constraints - lack of time, lack of money - may make this unfeasible, particularly for a small study.

On the other hand, if the sample is to be stratified by age or income, for example, no frame exists. Of course, census data does exist for these, and other, characteristics which suggest post-stratification may lead to gain in precision.

The difficulty with post-stratification is that with large samples, it often leads only to modest gains over random sampling (say 25% or so), and with smaller samples, you face the possible embarrassment of empty strata. In view of this, an alternative approach is suggested--

"ADD-ONS"

If there are L strata, and n_h^* is the desired allocation for stratum h, then the procedure is (1) take a series of independent random samples, sampling until

$$n_h \geq n_h^* \qquad h = 1, 2, \text{---}, L$$

where n_h is the number identified as belonging to stratum h, and (2) if

$$n_h > n_h^*$$

sub-sample the n_h members of stratum h to achieve the desired stratum size n_h^*. Then (3), to estimate the mean, for example, if

$$n_h^* = n_{h1} + n_{h2} + .. + n_{hk},$$

where n_{hi} is the sample number in stratum h in the ith sample of the series, in each stratum we use

$$\bar{y}_h = (n_{h1}\bar{y}_{h1} + n_{h2}\bar{y}_{hr} + .. + n_{hk}\bar{y}_{hk}) \mid n_h^*$$

(Note the above is an identity, we need only compute the final mean). Then we see

$$V(\bar{y}_h) = \frac{S_h^2}{n_h^*} \text{ (Ignoring the finite correction)}$$

which is exactly the variance of the mean of a sample of size n_h^* where S_h^2 is the stratum variance. Also, we can now construct our stratified estimator

$$\bar{y}_{st} = \sum_{h=1}^{L} w_h \bar{y}_h$$

where $w_h = n_h^* \mid n$ is the allocation we seek. So that we have

$$V(\bar{y}_{st}) = \Sigma\, w_h^2\, V(\bar{y}_h)$$

the usual variance for a stratified estimator.

Where the characteristic with which we want to stratify is not directly available, the above procedure may be modified, for example, as follows: Draw a random sample of size n_1, then screen (perhaps, by phone) to find the n_{h1}, that is, the number in each stratum. If $n_h < n_h^*$ for any stratum, repeat this procedure. Continue until $n_h \geq n_h^*$ $h = 1, 2, .., L$, as before.

It is not difficult to show with 2 strata of sizes N_1 and N_2, $N_1 + N_2 = N$, and desired allocation n_1^* and n_2^*, respectively, $n_1^* + n_2^* = n$, that

$$P(n = n_0) = \frac{n_1^* \binom{N_1}{n_1^*}\binom{N_2}{n_0 - n_1^*}}{n_0 \binom{N}{n_0}} + \frac{n_2^* \binom{N_1}{n_0 - n_2^*}\binom{N_2}{n_2^*}}{n_0 \binom{N}{n_0}} .$$

However, with increasing number of strata this formula quickly becomes awkward. Even in the 2 strata case it is not convenient for calculating E(n). If we were able to find E(n), then

$$\text{R.E.} = \frac{n_{PS}}{E(n)} ,$$

where n_{PS} is the sample size required for the standard post-stratified estimator to yield the same variance as this estimator, would give a measure of efficiency.

Some preliminary numerical results indicate that in the favorable case where N_1 and N_2 do not differ radically and S_1 and S_2 are known, substantial gains in efficiency are possible.

CONCLUSION

We have now a technique for post-stratification situations which: (1) enables us to achieve <u>any</u> pre-selected allocation, hence achieve <u>any</u> precision required, and (2) is simple to apply. In the first type of case, the additional costs of this technique are trivial, and in the second case, will probably be relatively modest.