Minnesota Forestry Research Notes

No. 280
August 1, 1982



# A PROCEDURE FOR COMBINING INDEPENDENT SAMPLE ESTIMATES

T. E. Burk and A. R. Ek[1]

## ABSTRACT

A statistical procedure is presented for combining estimates from independent forest surveys. An example is given and assumptions behind the procedure are also discussed.

## INTRODUCTION

Resource analysts frequently find two or more independent sample based estimates of the same quantity. These may arise from independent surveys conducted for different purposes and/or by different organizations. In some cases these surveys may also involve different sampling designs, plot sizes, and plot types. This note describes a procedure for combining such independent estimates to obtain an improved unbiased estimator of the quantity of interest. The precision of the resulting estimator in comparison to alternatives is also discussed.

## PROCEDURE

Assume that an unbiased estimate of mean timber volume per acre is to be developed for a subject area by probability sampling methods. Let the estimator and its sampling error be $z_1$ and $s_1$, respectively. Now let $z_i$, i=2, ...., k be k-1 other, independent, unbiased estimators of this same mean timber volume. Further, let $s_i$, i=2, ...., k be the estimators of the sampling errors of these $z_i$. It seems logical to utilize the k-1 independent estimates to improve upon the estimate from the current sampling effort. One approach to "pooling" all the information available for estimating mean timber volume per acre is to take the (unweighted) average of the k $z_i$. However, such an estimator ignores the important information provided by the sampling errors. For example if independent simple random samples of size 20 and 200 were taken in a subject area using identical 0.10 acre plots, the estimate provided by the sample of size 200 should logically be given more weight than the sample of size 20. One general method of combining independent estimates that does utilize the $s_i$ leads to the revised estimator

$$z^* = \frac{\sum_{i=1}^{k} w_i z_i}{w} \qquad [1]$$

whose sampling error may be estimated by[2]

$$s_{z^*} = [\frac{1}{w} + \frac{4}{w} \sum_{i=1}^{k} \frac{1}{n_i} \frac{w_i}{w} (1 - \frac{w_i}{w})]^{\frac{1}{2}} \qquad [2]$$

where

$w_i = 1/s_i^2$

$w = \sum_{i=1}^{k} w_i$

$n_i$ = degrees of freedom for the ith estimate

[1]Assistant Professor, Department of Forestry, Virginia Polytechnic Institute and State University, Blacksburg, VA and Professor, College of Forestry, University of Minnesota, St. Paul, MN. Research supported by the College of Forestry, University of Minnesota; Department of Forestry, Virginia Polytechnic Institute and State University; and the McIntire-Stennis Cooperative Forestry Research Program (Projects MIN-42-044 and 6324100).

[2]Although this is an approximation of sampling error, its performance has been shown to be good for moderate sized samples, say all $n_i > 10$ (Cochran and Carroll, 1953).

Published by the College of Forestry, 1530 North Cleveland Avenue, St. Paul, Minnesota 55108. E. Sucoff, editor.

The sampling error estimator [2] is from Meier (1953). The use of the combined estimator [1] can lead to substantial increases in precision in comparison to the use of any single estimator $z_i$ or the unweighted average of the $z_i$.

If desired, confidence intervals may be developed using [1], [2], and Student's-t distribution with degrees of freedom =

$$w^2/(\sum_{i=1}^{k} w_i^2/n_i)$$

(Meier, 1953).

## AN EXAMPLE

The following data represent three independent estimates of the mean cordwood volume per acre for a hypothetical jack pine pole-sized timber tract. Assume further that the three surveys involved different types of sampling units (plots) and sampling intensities.

| (i) | Survey | Mean Volume Per Acre ($z_i$) | Sampling Error[3] ($s_i$) | Degrees of Freedom for Estimate ($n_i$) |
|---|---|---|---|---|
| 1) | Reconnaissance Cruise | 19.1 | 3.2 | 30 |
| 2) | Permanent Plot Remeasurement | 16.8 | 5.2 | 12 |
| 3) | Timberland Acquisition Appraisal | 18.4 | 1.5 | 91 |

In this case the application of equations [1] and [2] led to

$$z^* = 18.4$$
$$s_z{}^* = 1.3$$

The ratio of sampling variances $s_1{}^2/s_z{}^{*2} = (3.2)^2/(1.3)^2 = 6.06$ suggests a 506 percent gain in efficiency for the combined estimator over the reconnaissance cruise. Alternatively, the combined estimator appears to provide a 33 percent gain in efficiency over the acquisition appraisal (i.e., $(1.5)^2/(1.3)^2 = 1.33$).

## DISCUSSION

The statistical model (or set of assumptions) for the above procedure is that the $z_i$ are independent, normally distributed random variables with the same expected value (i.e., a common mean). From a practical standpoint an important consideration before applying [1] and [2] is that, considering all facets of the measurement process, the $z_i$ are truly estimates of the same quantity.

For timber volume surveys, for example, the user must be sure that the subject surveys considered the same area and used the same merchantability specifications and volume tables. If that is not the case, appropriate adjustments should be made to the $z_i$ before attempting to combine them. Comparable considerations for other types of surveys are discussed by Lund and Schreuder (1980).

Several researchers have studied the potential gains from using z* or estimators similar to z* (see e.g., Burk et al., 1981, or Cochran and Carroll, 1953). At worst, z* will always be more precise than the least precise $z_i$. At best the true sampling error of z* can be much smaller than the true sampling error of the most precise $z_i$.[4] If the $s_i$ were equal to their corresponding true values rather than just being estimates of sampling error, z* would always be more precise than the most precise $z_i$. The magnitude of the gain in precision (over the most precise $z_i$) also is greater when combining estimators with similar precision. Realization of a potential gain in precision thus requires that the $s_i$ be good estimators of the corresponding true sampling errors. Further, the more similar the sampling errors of the $z_i$, the greater the chance that a gain in precision will be achieved. The authors speculate that actual gains (over the best of the $z_i$) of 20 to 50 percent will not be uncommon for timber volume surveys.

The estimator [1] is a more general form (for $k \geq 2$) of that presented by Wensel and John (1969) for combining data from different types of sampling units. The sampling error estimator [2] is likewise general and depends only on the respective sampling errors of the surveys being combined. Neither [1] or [2] require knowledge of the type of sampling units used.

An estimator similar to [1] is available for combining biased, dependent estimates, but such problems are beyond the scope of this discussion. See Burk et al. (1981) for more details on that topic.

---

[3]Estimate of the standard error of the mean for the survey.

[4]It can be shown that $s_z{}^*$ will always be smaller than the minimum $s_i$, i=1, ...., k when all $n_i \geq 4$.

## LITERATURE CITED


Burk, T.E., M.H. Hansen, and A.R. Ek. 1981 Combining sources of information for improved in-place inventory statistics. Proceedings of In-Place Inventories National Workshop, Orono, Maine, August 9-14. (in press).

Cochran, W.G. and S.P. Carroll. 1953. A sampling investigation of the efficiency of weighting inversely as the estimated variance. Biometrics 9:447-459.

Lund, H.G. and H.T. Schreuder. 1980. Aggregating inventories. Resources Evaluation Newsletter 4. USDI, Bureau of Land Management. 3 pp.

Meier, P. 1953. Variance of a weighted mean. Biometrics 9:59-73.

Wensel, L.C. and H.H. John. 1969. A statistical procedure for combining different types of sampling units in a forest inventory. Forest Science 15:307-317.



Published as Misc. Jour. Series No. 1827 of the University of Minnesota Agricultural Experiment Station.