PROJECTING VARIANCES FOR TWO METHODS OF ESTIMATING WEED ABUNDANCES

C. H. Proctor, North Carolina State University

During the past five years a wide variety of data have been collected in 50 apple orchards in Henderson County, NC. Similar projects have been done in New York and Michigan orchards. Part of the intent is to utilize pesticides more effectively by collecting rather extensive data on pest incidences. The present project concerns the estimation of relative coverage of the orchard floor by various weeds.

Data on weeds under about 500 trees in the 50 orchards provided a basis for estimating Smith's b for 35 species. Estimating Smith's b was the topic of my paper at last year's meetings. Let me just reiterate that it seems to be a useful way to describe spatial correlation, especially in handling a variety of plot sizes.

The present problem arose from a special study done by two weed scientists, Walt Skroch and Jeff Conn. They suspected that scanning by walking through the orchard might be as good a way to estimate weed coverages as plunking down frames at randomly chosen locations under randomly selected trees and painstakingly looking at the plants within the frames. They did come by my office before designing the experiment but, as I recall, I simply asked them to keep track of time spent on each operation and use their best judgement as to plot sizes and definition of the scanning operation. I must say they did an excellent job.

There were three observers who worked on three orchards. They first scanned by walking odd-numbered rows under one type of scan and then did even-numbered rows under the other scan variant--recording percentages for each row scanned. They each then used three sizes of frames to do plot sampling and recording percentages at one plot of each size under each of 12 trees per orchard.

The details of the three analyses of variance (overall, scans, and plots) are discussed in the expanded version of the paper. The overall comparisons show the presence of effects where they should be, as for orchards and some observer difference, and their absence where they could have been a nuisance, as for plot size. The degrees of freedom are so few that the conclusions can only be tentative, except that when they agree with prevailing understanding they can be more firmly established.

Rather than attempt to explain all the steps involved in estimating variance components, in deriving the "projection" formulas for variance of the estimates, and comparing the scans variance with the plots variance let me just discuss the strategy involved. The goal of this study is to say which method is better - scanning or plots. In "general statistics" one might hope to find a relative efficiency and answer the question once and for all. In the survey business there are relative efficiencies. That is, some methods work better for small scale studies while others do better for large scale studies.

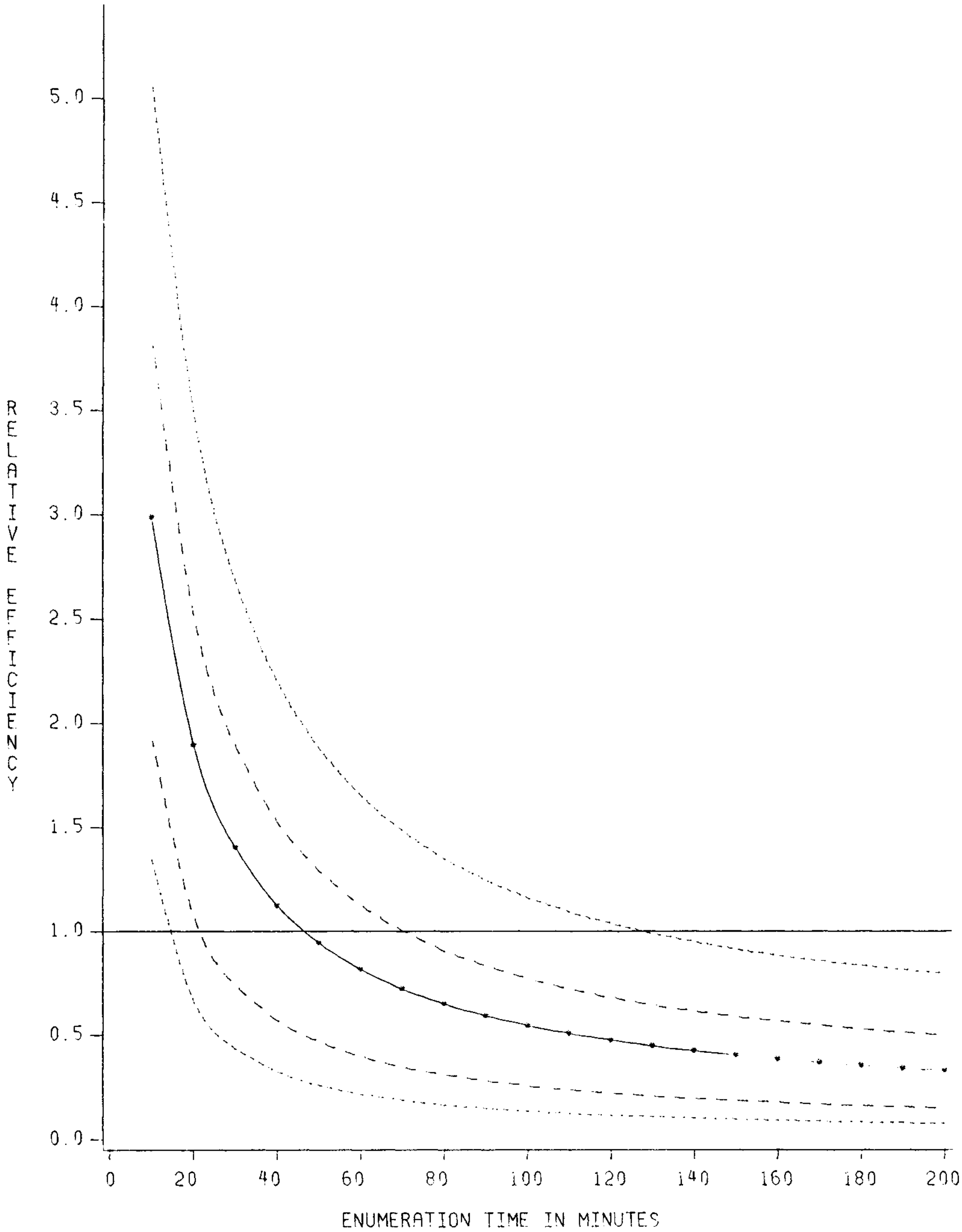
Roughly speaking, the mean square error of an estimate decreases with increasing survey cost as an equilateral hyperbola (as n^{-1}). Different methods differ: (1) in the sideways positioning of the hyperbola because of fixed costs, (2) in the vertical positioning on a floor of high or low biases, and (3) in the steepness or sharpness of curvature of the hyperbola because of cost coefficients. They may also differ (4) in height of the hyperbola because of differing plot size or bases of stratification or other conventional sample design features. In such cases relative efficiencies may stay constant for all survey cost and one design or design feature will be declared clearly superior.

In the present case one can show that a plot size between the two smaller plot sizes was better among the plots methods for all survey costs. This results from a nearly linear cost function and near conformity of plot variances to Smith's empirical law. Actually the optimum size differs slightly by species in accord with species differences in Smith's b.

In comparing scans to the best of the plots we are left with two equilateral hyperbolas. If we divide one by the other we have a relative mean square error curve and this is the end result of our calculations. It is shown as Figure 2, the central curve, the others are confidence limits. The general conclusion is that if relatively little time is to be spent for enumeration, as for a single orchard, one is better off using the scan. The five species were found to differ in this regard primarily because of the appearance of considerable observer biases in detecting some species by scanning. Thus the recommendation must be tempered by the fact that emergent phases of certain weed species can only be detected by carefully looking at plots, not just by scanning.

The statistical aspects of the study concern the formulation of model equations that serve to define variance components. Calculation of coefficients of expected mean squares are illustrated in their tedious detail. A program for hands-off variance component estimation and testing has been written in PROC MATRIX of SAS and tried out on text book analyses as well as on the present data and seems worthy of wider usage. Expressions for mean square error were derived which are seen to involve linear combination of the estimated variance components. Using the covariance matrix of the estimated variance components we could calculate effective degrees of freedom (a la Satterthwaite) for the projected mean square errors and then using the Wilson-Hilferty approximation to the F-ratio we found the confidence limits shown in Figure 2.

FIGURE 2. RELATIVE EFFICIENCY OF SCANNING TO PLOT SAMPLING FOR NIMBLEWILL



BANDS FOR 80% AND 50% CONFIDENCE COEFFICIENTS GIVEN BY DASHED LINES