ESTIMATING SMITH'S b FROM SAMPLE SURVEY DATA

Charles H. Proctor, North Carolina State University

1. Introduction

Although lists and directories are common frame materials for sample surveys it frequently happens that the space continuum, represented by a map, or the time continuum, reflected in clock readings, becomes the frame of a sampling investigation. One of the problems of design in such studies is choice of the best size of sampling unit. In order to solve this problem one needs to express survey costs as a function of sampling unit size and also to find the function relating sampling unit variance to size of unit. For surveys of agricultural crops such a variance function has been suggested by H. Fairfield Smith (1938).

The form of function is $S_C^2 = S_E^2 M^{-b}$ where S_C^2 is the variance of means of sampling units (clusters) containing M elements and S_E^2 is the variance among elements. If the M elements had been randomly selected to constitute the sampling units then the form of function would be $S_R^2 = S_E^2 M^{-1}$ where S_R^2 is the variance of means of such randomly created units. When adjacent elements are joined to form larger units, as is natural for a spatial frame, then b is found to be less than one. This reflects the fact that adjacent elements are more similar than those that are further separated.

2. Multi-stage Notation

The sampling units at the level of nesting of smallest size units will be called elementary units or elements. The levels will be indexed $\ell = 1,2,\ldots,L$ where the Lth level contains the elements. It will be supposed that units within each level are all of the same size, although the results should hold reasonably well for unequal sizes, with average sizes substituted for the single size. The number of first stage units (PSU's) will be denoted N_1 and the number of second stage units (SSU's) in each PSU will be written N_2, etc. The number of elements in each PSU is written M_1 and the number in each SSU is denoted M_2, etc. Notice that

$$M_\ell = \prod_{j=\ell+1}^{L} N_j = M_{\ell+1}N_{\ell+1} \quad (\ell = 0,1,2,\ldots,L-1), \quad (2.1)$$

and $M_L = 1$, while M_0 is the total number of element in the population.

The PSU's themselves will be indexed $i_1 = 1,2,\ldots,N_1$. The SSU's will be indexed by $(i_1, i_2) = (1,1), (1,2), \ldots, (1,N_2), \ldots, (N_1,N_2)$. An under squiggle will be used as shorthand for the ℓ-tuples such as $\underset{\sim}{i}_1 = i_1$, $\underset{\sim}{i}_2 = (i_1,i_2)$, $\underset{\sim}{i}_3, \ldots, \underset{\sim}{i}_L$. That is, there are M_0 L-tuples of the type $\underset{\sim}{i}_L$. The finite population of M_0 values of the Y variable may be written collectively as $\{Y_{\underset{\sim}{i}_L}\}$.

Population quantities of interest are:

Means: $\bar{Y}_0 = \bar{Y} = \sum_{\underset{\sim}{i}_L} Y_{\underset{\sim}{i}_L} / M_0$, $\bar{Y}_{\underset{\sim}{i}_\ell} = \sum_{\underset{\sim}{i}_{\ell+1} \in \underset{\sim}{i}_\ell} \bar{Y}_{\underset{\sim}{i}_{\ell+1}} / N_{\ell+1}$

Deviations: $D_{\underset{\sim}{i}_\ell} = \bar{Y}_{\underset{\sim}{i}_\ell} - \bar{Y}_{\underset{\sim}{i}_{\ell-1}} \quad (\underset{\sim}{i}_\ell \in \underset{\sim}{i}_{\ell-1})$,

where "$\underset{\sim}{i}_\ell \in \underset{\sim}{i}_{\ell-1}$" means summation over all $\underset{\sim}{i}_\ell$ having the same initial $\ell-1$ components as $\underset{\sim}{i}_{\ell-1}$.

The variances are:

Variance Components: $S_\ell^2 = \sum_{\underset{\sim}{i}_\ell} D_{\underset{\sim}{i}_\ell}^2 / N_1 N_2 \ldots N_{\ell-1}(N_\ell - 1)$

Cluster Variances: $S_{C,\ell}^2 = \sum_{\underset{\sim}{i}_\ell} (\bar{Y}_{\underset{\sim}{i}_\ell} - \bar{Y}_0)^2 / (U_\ell - 1)$,

where $U_\ell = \prod_{j=1}^{\ell} N_j = M_0 / M_\ell$.

Conversion back and forth between the S_ℓ^2 and the $S_{C,\ell}^2$ is relatively easy. Notice that

$$(\bar{Y}_{\underset{\sim}{i}_\ell} - \bar{Y}) = \bar{Y}_{\underset{\sim}{i}_\ell} - \bar{Y}_{\underset{\sim}{i}_{\ell-1}} + \bar{Y}_{\underset{\sim}{i}_{\ell-1}} - \bar{Y}$$
$$= D_{\underset{\sim}{i}_\ell} + (\bar{Y}_{\underset{\sim}{i}_{\ell-1}} - \bar{Y}),$$

so that

$$S_{C,\ell}^2 = [\sum D_{\underset{\sim}{i}_\ell}^2 + \sum (Y_{\underset{\sim}{i}_{\ell-1}} - \bar{Y})^2] / (U_\ell - 1)$$
$$= [(U_\ell - U_{\ell-1}) S_\ell^2 + (U_{\ell-1} - 1) N_\ell S_{C,\ell-1}^2] / (U_\ell - 1), \quad (2.2)$$

or equivalently:

$$S_\ell^2 = [U_\ell (S_{C,\ell}^2 - S_{C,\ell-1}^2) + N_\ell S_{C,\ell-1}^2 - S_{C,\ell}^2] / (U_\ell - U_{\ell-1}). \quad (2.3)$$

Equation (2.2) applies to $\ell = 2,3,\ldots,L$ while $S_{C,1}^2 = S_1^2$. This shows that the $S_{C,\ell}^2$ may be obtained by a linear transformation on the S_ℓ^2. If the quantities $S_{C,\ell}^2$ be placed in an L-component column vector, say $\underset{\sim}{CV}$, and the S_ℓ^2 in another, say $\underset{\sim}{VC}$ then equation (2.2) defines the matrix B in

$$\underset{\sim}{CV} = B \, \underset{\sim}{VC} \quad . \quad (2.4)$$

3. Sample Data and ANOVA Estimates of $\underset{\sim}{VC}$

In parallel to the population quantities $\{Y_{\underset{\sim}{i}_L}\}$ we denote sample observations as $\{y_{\underset{\sim}{u}_L}\}$. The draw-order subscripts $\underset{\sim}{u}_\ell$ stand for the vectors $(u_1, u_2, \ldots, u_\ell)$ with $\ell = 1,2,\ldots,L$ where u_ℓ denotes order of draw of a cluster into a simple random sample (SRS) of size n_ℓ from a finite population of N_ℓ clusters. The simple balanced multi-stage sample design is completely specified by the sample sizes $n_1, n_2, \ldots, n_L$. The whole collection may be denoted S and the number of such samples is

$$(S) = \binom{N_1}{n_1} \binom{N_2}{n_2}^{n_1} \cdots \binom{N_L}{n_\ell}^{n_1 n_2 \cdots n_{L-1}} \quad . \quad (3.1)$$

A randomly selected such sample will be written s. In fact, s is a correspondence, $s(u_\ell) = i_\ell$, showing which clusters were drawn and in what order into the sample.

From a thoroughgoing, sampling-from-a-finite-population viewpoint the $\{Y_{i_L}\}$ are a fixed collection of quantities from which the $\{D_{i_\ell}\}$ may be calculated, as well as the S_ℓ^2, and so forth. The randomness of the sample observations is due entirely to the simple random subsampling of clusters. The sample unit means may be represented as:

$$\bar{y}_{u_\ell} = \bar{Y} + D_{u_1 \in u_\ell} + D_{u_2 \in u_\ell} + \ldots + D_{u_\ell} + \bar{D}_{u_\ell \ni u_{\ell+1}} + \ldots + \bar{D}_{u_\ell \ni u_L} , \quad (3.2)$$

where the quantities, $D_{u_2 \in u_\ell}$, for example, are single values of deviations while the, for example, $\bar{D}_{u_\ell \ni u_{\ell+1}}$ are means of finite population deviations. The notation "$\in$" refers to agreement of components of the indexes and the symbol "$\ni$" refers to averaging over the extra components. For example $\bar{D}_{u_\ell \ni u_{\ell+2}}$ is an average of $n_{\ell+1}\, n_{\ell+2}$ quantities.

The unweighted sample mean can be written

$$\bar{y} = \bar{Y} + \bar{D}_1 + \bar{D}_2 + \ldots + \bar{D}_L \quad (3.3)$$

in which $\mathcal{E}(\bar{D}_\ell) = 0$ for $\ell = 1, 2, \ldots, L$ and thus $\mathcal{E}(\bar{y}) = \bar{Y}$. Now consider the differences $\bar{y}_{u_1} - \bar{y}$ for $u_1 = 1, 2, \ldots, n_1$. These appear as:

$$\bar{y}_{u_1} - \bar{y} = (D_{u_1} - \bar{D}_1) + (\bar{D}_{u_1 \ni u_2} - \bar{D}_2) + \ldots + (D_{u_1 \ni u_L} - \bar{D}_L) . \quad (3.4)$$

Notice that $\bar{D}_2$ is the mean of $n_1 n_2$ deviation quantities or of n_1 submeans each based on n_2 quantities, while each $\bar{D}_{u_1 \ni u_2}$ is a submean of n_2 deviations. Because the sample of n_1 first stage units from the N_1 of the population was SRS, it is possible to construct unbiased estimates of population variances by squaring the n_1 differences, adding them and dividing by $n_1 - 1$ (Cochran's Theorem 2.2). In routine computation of the sample analysis of variance (ANOVA) this is done to find mean squares as, for example:

$$ms_1 = \sum_{u_1} \ldots \sum_{u_L} (y_{u_1} - \bar{y})^2/(n_1 - 1) = n_2 n_3 \ldots n_L \sum_{u_1} (y_{u_1} - \bar{y})^2/(n_1 - 1) . \quad (3.5)$$

Upon squaring the RHS of (3.4) and then taking expectation over random sampling the cross product terms will equal zero because the average of deviations at any one level equals zero within any more inclusive unit. Consequently,

$$\mathcal{E}(ms_1) = n_2 n_3 \ldots n_L [S_1^2 + (1 - f_2)S_2^2/n_2 + (1 - f_3)S_3^2/n_2 n_3 + \ldots + (1 - f_L)S_L^2/n_2 n_3 \ldots n_L] = (1 - f_L)S_L^2 + n_L(1 - f_{L-1})S_{L-1}^2 + \ldots + n_L n_{L-1} \ldots n_2 S_1^2 \quad (f_\ell = n_\ell/N_\ell) . \quad (3.6)$$

By a similar argument one finds:

$$\mathcal{E}(ms_\ell) = (1 - f_L)S_L^2 + \ldots + n_L n_{L-1} \ldots n_{\ell-1} S_\ell^2 , \quad (3.7)$$

as given in, for example, Bennett and Franklin (1954).

It is frequently more realistic to suppose that there have been added to each sample observation some random, so called, nonsampling error quantities. We will suppose, rather unrealistically I'm afraid, that they are independent of the underlying Y_{i_L} values and that their expectations are zero for each level with some characteristic variance at that level.

To see their effect on expectations of sample mean squares we go back to the discussion of $\bar{D}_{u_1 \ni u_2}$ as a submean of n_2 deviations. We still have these deviations, but there are now also n_2 nonsampling error quantities in each submean. Whereas we had $(1 - f_2)S_2^2/n_2$ for the variance being estimated we now find it to be $[(1 - f_2)S_2^2/n_2 + \sigma_2^2/n_2]$ say.

In order to minimize disruption to our notational scheme we propose to denote total variance, which may be defined as $\sigma_2^2 + S_2^2$, by S_2^2 itself and to denote the sampling error portion by $\gamma_2 S_2^2$. Thus each submean of n_2 FP deviations plus n_2 nonsampling error quantities has variance:

$$[(1 - f_2)\gamma_2 S_2^2 + (1 - \gamma_2)S_2^2]/n_2 = (1 - \gamma_2 f_2)S_2^2/n_2 . \quad (3.8)$$

This leads finally to a generalization of (3.7) as:

$$E(ms_\ell) = (1 - \gamma_L f_L)S_L^2 + n_L(1 - \gamma_{L-1} f_{L-1})S_{L-1}^2 + \ldots + n_L n_{L-1} \ldots n_{\ell-1} S_\ell^2 \quad (3.9)$$

This linear transformation can be represented by the matrix A in

$$\hat{\underset{\sim}{VC}} = A\, \underset{\sim}{ms} , \quad (3.10)$$

where $\hat{\underset{\sim}{VC}}$ is the vector of $\hat{S}_\ell^2$ quantities and $\underset{\sim}{ms}$ contains the ms_ℓ. Notice that by using (2.4) we obtain estimates of the cluster variances in:

$$\hat{\underset{\sim}{CV}} = BA\, \underset{\sim}{ms} . \quad (3.11)$$

4. Estimating b From Sample Mean Squares

Smith's relation is written

$$S_{C,\ell}^2 = S_{C,L} M_\ell^{-b} , \quad (4.1)$$

while a linearized form to be fit to data can be obtained as:

$$y = x\beta + d \quad , \qquad (4.2)$$

where $y = \ell og(\hat{\underset{\sim}{C}}\underset{\sim}{y})$, x contains logs of the M_ℓ , β has two components , $\beta_1 = \ell og(s^2_{C,L})$ and $\beta_2 = b$, while d is an error vector.

If the observations $\{y_{u_{\sim L}}\}$ can be viewed as randomly drawn from a normally distributed additive model then the sample means squares would be independently chi-square distributed. Each mean square would have variance equal to twice its squared expectation divided by its degrees of freedom. Although such an assumption is not exactly satisfied, there may be sufficiently close correspondence to warrant its provisional acceptance.

If we denote by D_{msv} a diagonal matrix with entries $2\, ms^2_\ell/d_\ell$ where $d_\ell = n_1 n_2 \cdots n_{\ell-1}(n_\ell - 1)$ and further denote by D_{VC} the diagonal matrix with entries from $\hat{\underset{\sim}{VC}}$ then the following computation for V would provide a covariance matrix for d under the just stated assumptions:

$$V = D^{-1}_{VC}\, B\, A\, D_{msv} A^T B^T D^{-1}_{VC} \quad . \qquad (4.3)$$

Applying generalized least squares to the estimation of β leads to:

$$\hat{\beta} = (x^T V^{-1} x)^{-1}\, x^T V^{-1} y \quad . \qquad (4.4)$$

The estimated covariance matrix for $\hat{\beta}$ is $(x^T V^{-1} x)^{-1}$, while the error sum of squares is found as:

$$ESS = y^T V^{-1} y - y^T V^{-1} x \hat{\beta} \quad . \qquad (4.5)$$

This quantity ESS may be compared to a chi-squared distribution on L - 2 degrees of freedom and large values taken as evidence of failure of fit of the model. With reasonable fit the square roots of diagonal elements of $(x^T V^{-1} x)^{-1}$ are useful standard errors for judging the precision of a and b . These results are essentially those given by Hathaway and Williams (1958).

After the model had been fit to several sets of data we found that evidence of lack of fit was rather common. Such a result could arise from non-normality or nonindependence of the deviation at the various levels as well as from nonlinearity at the overall level of logs of cluster variances. If one were to relax the distributional assumptions the logical or "default" direction would seem to be more toward unweighted least squares. However, it is unfortunate to have to give up knowledge of the degrees of freedom involved when estimating standard errors. That is, when the data are for only, say, four levels of nesting then the standard errors are based on just these four observations when using unweighted least squares, although they are based on the multi-stage sample sizes behind the ANOVA mean squares when the more specific model is used.

The covariance matrix of the data points for unweighted least squares is simply a constant diagonal matrix and thus a compromise would be attained by some linear combination of the identity matrix I and V . By way of preserving information on standard errors contained in V it was decided to maintain the trace of the compromise covariance matrix equal to the trace of V . The new covariance matrix will be denoted σ and is defined as:

$$\sigma = \alpha I + (1 - \alpha/\bar{\gamma})\, V \quad , \qquad (4.6)$$

where $\bar{\gamma}$ is the arithmetic average of the eigenvalues of V . As the chi-square distribution approaches normality with large degrees of freedom the ℓog likelihood approaches the following multivariate normal form:

$$LLF = -.5\, [\ell og\, 2 + \ell og(\sigma) + (y - x\beta)^T \sigma^{-1}(y - x\beta)] \quad . \qquad (4.7)$$

Maximizing this likelihood by choice of β and α furnishes some information about the advisability of using $\alpha \neq 0$.

5. Applications

After a number of experiences with datasets ranging from small scale classroom examples, through Smith's original wheat yield data and the Hathaway and Williams' examples for soybean yields and timber volumes plus tobacco diseases, satellite imagery of wheat fields, Ghana agricultural census, Detroit Area Study educational levels, and many more, we arrived at a computational routine with the following features:

1) Input consists of mean squares (ms_ℓ), their degrees of freedom (d_ℓ) , population sizes (N_ℓ), and proportions of sampling variances (γ_ℓ) . If the mean square at any level is in fact a residual mean square after removal of a systematic effect, then its actual degrees of freedom must also be input. Sample sizes (n_ℓ) are computed from the degrees of freedom and may be noninteger when the sample data are unbalanced.

2) Estimates of b are furnished for unweighted least squares, for $\alpha = 0$, for $\hat{\alpha}$ and for an imposed value of α , usually $\alpha = .01$. The likelihood function is displayed and chi-square goodness of fit statistics are provided in order to judge the reasonableness of the models.

Whencontemplating use of a nearly balanced multi-stage sampling design one faces a sample variance of the form:

$$V(\bar{y}) = s^2_1(n_1^{-1} - \gamma_1 N_1^{-1}) + s^2_2(n_2^{-1} - \gamma_2 N_2^{-1})n_1^{-1} + \cdots + s^2_L(n_L^{-1} - \gamma_L N_L^{-1})n_1^{-1} n_2^{-1} \cdots n_{L-1}^{-1}, \qquad (5.1)$$

with a cost function possibly of the form:

$$C = C_0 + C_1 n_1 + C_2 n_1 n_2 + \cdots + C_L n_1 n_2 \cdots n_L \quad . \qquad (5.2)$$

Minimizing $V(\bar{y})$ for fixed C as in Cochran (1977, p. 288) results in:

$$n^2_{\ell,opt} = \frac{s^2_\ell - \gamma_{\ell+1} s^2_{\ell+1}/N_{\ell+1}}{s^2_{\ell-1} - \gamma_\ell s^2_\ell/N_\ell} \times \frac{C_{\ell-1}}{C_\ell} \quad , \qquad (5.3)$$

for $\ell = 2, 3, \ldots, L$ and $\gamma_{L+1}\, s^2_{L+1}/N_{L+1}$ may be set to zero. The quantities $s^2_\ell - \gamma_{\ell+1} s^2_{\ell+1}/N_{\ell+1}$ may be estimated in the program by setting the $\gamma_\ell = 0$ and obtaining the ANOVA-estimated components of variance. Using an approximate form of (2.3) and denoting the value of Smith's b just obtained as b^* we find:

$$n_{\ell,opt} \doteq N_\ell^{b^*/2} \sqrt{C_{\ell-1}/C_\ell} \quad , \qquad (5.4)$$

which may be of use in deciding on provisional subsample sizes when the cluster sizes are fixed.

A somewhat more familiar problem is that of choosing an optimum size of cluster, or M_ℓ, when cluster size may be varied in the neighborhood of some M_ℓ. The solution as derived by Smith (1938) is

$$M_{\ell,opt} \doteq b\, C_\ell/(1-b)C_{\ell+1} \qquad (5.5)$$

If there is seen to be lack of fit in the model then one may use a b-value closer to the inter-point b-values at level ℓ-1 to ℓ and ℓ to ℓ+1. The optimizing equation then becomes:

$$M_{\ell,opt} \doteq b_\ell C_\ell/(1-b_\ell)\, C_{\ell+1} \quad . \qquad (5.6)$$

Since Smith's b is to some extent an alter-native to use of the intracluster correlation coefficient and design effect it may be of interest to show how these are related, namely as:

$$deff_\ell = M_\ell^{1-b} \text{ and } \rho_\ell = [M_\ell^{(1-b)} - 1]/(M_\ell - 1). \qquad (5.7)$$

References

Bennett, Carl A. and Franklin, Normal L. (1954) Statistical Analysis in Chemistry and the Chemical Industry, Wiley, NY.

Hatheway, W. H. and William, E. J. (1958) Efficient estimation of the relationship between plot size and the variability of crop yields, Biometrics, 14: 207-222.

Smith, H. Fairfield (1938) A empirical law describing heterogeneity in the yield of agricultural crops, Journal of Agricultural Science 28: 1-23.

Appendix on Two Numerical Examples

A population of zero-one data on a lattice was created by laying a 20 by 20 square grid of points on an aerial photo and scoring one if the point was on agricultural land - zero otherwise. First stage units were defined as 4 by 4 square sets of points, second stage units as 2 by 2 squares and individual points were elements. Thus $N_1 = 25$, $N_2 = 4$ and $N_3 = 4$. A simple random multi-stage sample was drawn with $n_1 = 5$, $n_2 = 4$ and $n_3 = 4$; that is, five of the 25 PSU's were selected by SRS and all points were observed in each. The mean squares in the nested ANOVA were found to be $ms_1 = 1.356$, $ms_2 = .308$ and $ms_3 = .159$ with $d_1 = 4$, $d_2 = 15$ and $d_3 = 60$.

Since $f_3 = f_2 = 1$ the expectations of the mean squares are: $E(ms_3) = S_3^2$, $E(ms_2) = 4\, S_2^2$ and $E(ms_1) = 16\, S_1^2$ so that $\hat{S}_3^2 = .159$ and $\hat{S}_2^2 = .308/4 = .077$ while $\hat{S}_1^2 = 1.356/16 = .08475$ in accord with (3.9). Using (2.3) we find $\hat{S}_{C,2}^2 = [25(3)\ .077 + 24(3)\ .08475]/99 = .14051515$ and $\hat{S}_{C,3}^2 = .25900752$. The matrices A and B are:

$$A = \begin{pmatrix} .0625 & 0 & 0 \\ 0 & .25 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ and } B = \begin{pmatrix} 1 & 0 & 0 \\ .9697 & .7576 & \\ .9624 & .7519 & .7519 \end{pmatrix}.$$

Under the approach based on unweighted regression of $\log \hat{S}_{C,\ell}^2$ on $\log M_\ell$ the estimate of Smith's b was found as $b_u = .403$ with a standard error, derived from the one degree of freedom in the regression residuals, as .022. Consistent with this unweighted approach to estimating b are the, so called, local b-values. These are obtained as ratios of the negative first differences of $\log \hat{S}_{C,\ell}^2$ divided by those of $\log M_\ell$. For example,

$$(\log \hat{S}_{C,1}^2 - \log \hat{S}_{C,2}^2)/(\log M_2 - \log M_1)$$

$$= (-2.468 + 1.962)/(1.386 - 2.773) = .365 \ .$$

Their standard errors may be obtained from V (see below) which contains estimated covariances of the $\log \hat{S}_{C,\ell}^2$ quantities. For example, the standard error to attach to .365 is found as:

$$[.5 + .19401 - 2(.29243)]^{\frac{1}{2}}/(1.386 - 2.773) = .238 \ .$$

In accord with the Hatheway-Williams (1958) procedure based on $\alpha = 0$, the significance probability of the fit of the data turned out to be .68 and so the value $\alpha = 0$ seems reasonable. The accompanying estimate of Smith's b is $b_0 = .438$ with standard error based on the supposed chi-square distributed mean squares and good fit to the model as .159. Smoothed cluster variances, to be written $\tilde{S}_{C,\ell}^2$, were calculated by exponentiating fitted values from the prediction equation, while smoothed variance components, written $\tilde{S}_\ell^2$, were obtained from these by using (2.3). We found $\tilde{S}_{C,1}^2 = .076$, $\tilde{S}_{C,2}^2 = .139$ and $\tilde{S}_{C,3}^2 = .254$, while $\tilde{S}_1^2 = .076$, $\tilde{S}_2^2 = .086$ and $\tilde{S}_3^2 = .155$.

When the value of $\alpha = .01$ is imposed then the estimate of Smith's b becomes $b_i = .421$ with standard error, based on the resulting error covariance matrix, of .173. The results for $\alpha = .01$ are somewhat of a compromise between the $\alpha = 0$ case and that of unweighted regression. The matrix V and its inverse, denoted W, are found to be:

$$V = \begin{pmatrix} .5 & .29243 & .157455 \\ .29243 & .19401 & .104462 \\ .157455 & .104462 & .0633472 \end{pmatrix}$$

$$\text{and } W = \begin{pmatrix} 16.886 & -25.4523 & 0 \\ -25.4523 & 84.343 & -75.8208 \\ 0 & -75.8208 & 140.817 \end{pmatrix}$$

Notice that the sum of entries in W is 39.5 and is equal to half the sum of degrees of freedom of the mean squares being fit as will always be the case.

All of the above calculations were obtained with the values of γ_ℓ equal one as befits the case of careful measurement with no nonsampling error variance. When the γ_ℓ are set to zero and the fitting routine is rerun then b^* is found, using the $\alpha = 0$ case, as .418. The ANOVA-estimated planning variance components were found as .0655, .03725 and .159 for $\ell = 1,2,3$ and when

smoothed they become: .0655 , .09173 and .2106 .

A second example uses data from Smith's (1938) original article. These refer to a wheat field in Australia on which a uniformity trial was conducted and yield observations recorded for every 6-inch by 12-inch plot - all 1,080 of them. This collection of elementary plots was used to create compact clusters of many shapes and sizes on which variances of average wheat yields could be calculated.

The present method was applied to the series of cluster variances for plots of sizes (in feet): 3 x 12, 3 x 6, 1 x 6, $\frac{1}{2}$ x 6, $\frac{1}{2}$ x 2, and $\frac{1}{2}$ x 1 . The cluster variances (with numbers of elementary units, M_ℓ , in brackets) as reported by Smith in his Table 1a (p. 7) are: 102(72), 155(36), 383(12), 524(6), 1220(2) and 2201(1) . The corresponding mean squares (with d_ℓ in brackets) were found to be: 7344(14), 3934(15), 4120(60), 1708(90), 2090 (360) and 1962(540) .

The following results were obtained upon applying the fitting routines. The unweighted regression estimate was $b_0 = .712$; with $\alpha = 0$ the estimate was $b_0 = .774 \pm .049$ but with $P_0 = .0002$ as level of significance for the test of H_0: $\alpha = 0$; with likelihood maximized at $\hat{\alpha} = .006$ (and goodness of fit $\chi^2 = 3.34$ on 4 degrees of freedom) we found $b_\alpha = .724 \pm .061$; finally with α set at .01 the b-value was $b_i = .721 \pm .064$. Smith reported $b' = .7486$ with standard error of .0132. The following series of ratios of the cluster variances to the smoothed values gives some evidence of the goodness of fit: 1.08 , .99 , 1.11 , .92 , .96 and 1.05 .

The agreement is certainly close between our methods and Smith's except for the standard error. Since the present method uses such a small portion of Smith's cluster variances, only six of the 45, it behooves us to inquire as to a possible loss of information on the part of the nested ANOVA approach or a possible manufacturing of data by Smith's calculations. It appears that both are happening. That is, examination of just one nested series looses some information on the effect of plot shape. However, whatever is the point-to-point pattern of spatial correlation governing the yield, it will be rather thoroughly reflected in one series of nested mean squares, so that, given knowledge of that one, alternative series must show a similar result.

Unfortunately, without a more specific model for generation of the data these arguments cannot be stated more precisely. Nonetheless, one should be rather cautious in accepting a standard error calculation, or an estimate of b , that is affected by the number of plot shapes chosen to be examined when that number is essentially unlimited. That is, the standard error obtained from our procedure, while it needs to be judged as only a provisional estimate, is likely a more reasonable one to use. Similarly, the estimate is somewhat more objectively arrived at and there are signals of possible lack of fit at stages of the estimation calculation.