

COMPUTER SIMULATION OF VARIANCE COMPONENTS ESTIMATES\*

S. R. Searle

Biometrics Unit, Cornell University, Ithaca, N.Y. U.S.A.

BU-233-M

January, 1967

Abstract

Preliminary results are discussed of investigations into the effects of unbalancedness of data on the customary estimate of the between-groups variance component in a 1-way classification. Variances of the estimate are given and, based on computer simulation results, tentative suggestions are made about the character of the distribution of the estimate (similar to a chi-square, although tending to exponential in certain situations). Monte Carlo techniques, distinct from straight simulation, are also discussed, and approximations to the distribution are considered.

---

\* Paper prepared for conference pre-print, 6<sup>th</sup> International Biometric Conference, Sydney, Australia, August 1967.

COMPUTER SIMULATION OF VARIANCE COMPONENTS ESTIMATES\*

S. R. Searle

Biometrics Unit, Cornell University, Ithaca, N.Y. U.S.A.

BU-233-M

January, 1967

Introduction

The customary random effects model for a between- and within-groups analysis is well known, viz. Model II of Eisenhart (1947). If data consist of  $n_i$  observations in the  $i$ 'th group, for  $i = 1, 2, \dots, c$ , the equation of the model for  $y_{ij}$ , the  $j$ 'th observation in the  $i$ 'th group is

$$y_{ij} = \mu + a_i + e_{ij}$$

where  $\mu$  is a general mean,  $a_i$  is the effect due to the  $i$ 'th group and  $e_{ij}$  is a random error term. In the random model the  $a_i$  are assumed to be a random sample of  $a$ 's from a population having zero mean and variance  $\sigma_a^2$ , being uncorrelated with each other and with the  $e_{ij}$ -terms which themselves are assumed to have zero mean and variance  $\sigma_e^2$ , they too being uncorrelated with each other. In this context the matter of interest is to estimate the variance components  $\sigma_a^2$  and  $\sigma_e^2$  from the  $N (= \sum_{i=1}^c n_i)$  observations  $y_{ij}$  for  $j = 1, 2, \dots, n_i$ , and  $i = 1, 2, \dots, c$ . The usual procedure, vide Henderson (1953), is to calculate the between- and within-group mean squares,

$$MSB = \frac{1}{c-1} \left[ \sum_{i=1}^c \left( \sum_{j=1}^{n_i} y_{ij} \right)^2 / n_i - (\sum_{ij} y_{ij})^2 / N \right] \quad \dots (1)$$

and

$$MSW = \frac{1}{N-c} \left[ \sum_{i=1}^c \sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^c \left( \sum_{j=1}^{n_i} y_{ij} \right)^2 / n_i \right]$$

---

\* Paper prepared for conference pre-print, 6<sup>th</sup> International Biometric Conference, Sydney, Australia, August 1967.

and equate these values to their expectations. Solving the resulting equations for  $\sigma_a^2$  and  $\sigma_e^2$  leads to estimators

$$\hat{\sigma}_e^2 = \text{MSW} , \quad \text{--- (2)}$$

$$\text{and } \hat{\sigma}_a^2 = \frac{\text{MSB} - \text{MSW}}{\frac{c}{N^2 - \sum_{i=1}^c n_i^2}} \cdot \frac{N(c-1)}{N(c-1)} .$$

These estimators are not maximum likelihood, as are those of Herbach (1959) and Thompson (1962) nor are they admissible, Robson (1965). They are, however, unbiased no matter what underlying form of distribution is attributed to the  $a_i$ 's and  $e_{ij}$ 's - so long as they have zero means, variances  $\sigma_a^2$  and  $\sigma_e^2$  and are uncorrelated as previously described. Should their distributions be normal the variances of these estimators are, as given in Searle (1956),

$$\text{var}(\hat{\sigma}_a^2) = \frac{2N^2(c-1)}{(N^2 - S_2)^2} \left[ \frac{(N-1)\sigma_e^4}{N-c} + \frac{2(N^2 - S_2)\sigma_e^2\sigma_a^2}{N(c-1)} + \frac{(N^2S_2 + S_2^2 - 2NS_3)\sigma_a^4}{N^2(c-1)} \right]$$

and --- (3)

$$\text{var}(\hat{\sigma}_e^2) = \frac{2\sigma_e^4}{N-c} ,$$

where  $S_2 = \sum_{i=1}^c n_i^2$  and  $S_3 = \sum_{i=1}^c n_i^3$ ; and the covariance between the estimators is

$$\text{cov}(\hat{\sigma}_a^2, \hat{\sigma}_e^2) = \frac{-N(c-1)\text{var}(\hat{\sigma}_e^2)}{N^2 - S_2} .$$

Under the assumption of normality MSW, apart from a constant, is a  $\chi^2$  variable and hence so is  $\hat{\sigma}_e^2$ . And, when the numbers of observations,  $n_i$ , are all the same (balanced data) MSB is a  $\chi^2$  variable multiplied by a constant and the distribution and moments of  $\hat{\sigma}_a^2$  are of the form indicated in Wang (1966). But when there are unequal numbers of observations in the groups (unbalanced data) MSB is not  $\chi^2$  and the distribution of  $\hat{\sigma}_a^2$  is unknown. Its variance and covariance with  $\hat{\sigma}_e^2$  are known, as above, but its distribution is not. It is of interest, therefore, to study the manner in which unbalancedness of data

(i.e. the divergence from having equal numbers of observations in the groups) affects the distribution of the estimator  $\hat{\sigma}_a^2$ . Because analytical methods appear intractable, and because there is virtually no limit to the different kinds of unbalancedness of data, computer simulations are being used as a means of investigation. Different sets of values of the  $n_i$  are being used in combination with different  $\sigma_a^2$  values, and for each combination a frequency polygon is being made of estimated components  $\hat{\sigma}_a^2$  derived from 2000 simulations. The work is currently in progress, and the results which follow are those of preliminary investigations only.

#### Variance of estimated between-groups component

The sampling variance of  $\hat{\sigma}_a^2$ , the estimate of the between-groups variance component, is given by (3). This expression is not readily amenable to analytic study to ascertain the effect on  $\text{var}(\hat{\sigma}_a^2)$  of unbalancedness. Computed values of  $\text{var}(\hat{\sigma}_a^2)$  were therefore obtained for a series of  $\sigma_a^2$  values for each of several sets of  $n_i$ -values, in the hope that they might indicate at least gross effects of unbalancedness on the estimation procedure. This approach would, of course, be greatly aided by an ability to characterize unbalancedness in some fashion. Ideally one would like to have some statistic of unbalancedness so as to consider the effects of different values of this measure on the variance component estimation procedure. Alternately the problem could be looked at the other way: if the effects of different kinds or degrees of unbalancedness on  $\hat{\sigma}_a^2$  were known then maybe it would be possible to use these to establish a measure of unbalancedness, one that was particularly suited to the problem of characterizing its effect on the estimation of  $\sigma_a^2$ . From either point of view the need for a measure of unbalancedness is apparent. The variance of the  $n_i$ 's is an obvious suggestion, relative to its maximum value  $(N - c)^2/c$  perhaps. This, and other possible statistics are currently being considered but in the meantime, in the absence of anything wholly satisfactory, the effects of unbalancedness are being studied by considering sets of  $n_i$ -values that range all the way from balanced data to situations that would customarily be considered very unbalanced. Initially, 5 sets of  $n_i$ -values have been used, each of them being for 5 groups with a total of 25 observations:

$P_1$	:	5	5	5	5	5	(balanced data)
$P_2$	:	1	1	3	10	10	} (moderately unbalanced)
$P_3$	:	1	1	7	8	8	
$P_4$	:	1	1	1	11	11	} (seriously unbalanced) .
$P_5$	:	1	1	1	1	21	

We refer to these as n-patterns; thus an n-pattern is simply a set of  $n_i$ -values for some given  $c$  and  $N$ , the number of groups and number of observations, respectively ( $N = \sum_{i=1}^c n_i$ ). The five n-patterns considered above range in style from balanced to very unbalanced. For additional comparison four further n-patterns were also used:

$P_6$	:	1	1	1	21	21	} (for comparisons) .
$P_7$	:	5	5	5	55	55	
$P_8$	:	5	5	5	5	105	
$P_9$	:	5	5	5	105	105	

The first of these,  $P_6$  is merely  $P_5$  with 20 observations added to one of the single-observation groups; and  $P_7$ ,  $P_8$  and  $P_9$  are just  $P_4$ ,  $P_5$  and  $P_6$  with five times as many observations in each group.

Each of these nine n-patterns has been used in combination with eleven  $\sigma_a^2$  values. Since the expression for  $\text{var}(\hat{\sigma}_a^2)$  in (3) is homogeneous in  $\sigma_a^2$  and  $\sigma_e^2$ , the value of unity has been used for  $\sigma_e^2$  at all times. (In this way  $\sigma_a^2$  hereafter represents the ratio  $\sigma_a^2/\sigma_e^2$ .) And in this context the eleven values used for  $\sigma_a^2$  were

$$\sigma_a^2 : 0, \frac{1}{4}, \frac{1}{2}, \frac{3}{4}, 1, 2, 3, 4, 5, 10 \text{ and } 20 .$$

For each of these values of  $\sigma_a^2$  (with  $\sigma_e^2 = 1$ ) used in combination with each of the nine n-patterns, the value of  $\text{var}(\hat{\sigma}_a^2)$  calculated from (3) is shown in Table 1. As would be expected, for each n-pattern this variance increases as  $\sigma_a^2$  increases; and, as indicated in footnotes to the table, several other points are evident, as follows.

(1) In n-patterns  $P_1$  through  $P_5$ , the largest value of  $\text{var}(\hat{\sigma}_a^2)$  for any given  $\sigma_a^2$  is either in  $P_4$ , (1, 1, 1, 11, 11), or in  $P_5$ , (1, 1, 1, 1, 21). The indication is that for given N and c the largest value of  $\text{var}(\hat{\sigma}_a^2)$  for any  $\sigma_a^2$  may be when the n-pattern is of the form (1, 1, ..., 1, k) or (1, 1, 1, ..., q, q) where  $k = N-c+1$  and  $q = \frac{1}{2}(N-c+2)$ .

(2) If, for given c and N, the n-pattern giving largest  $\text{var}(\hat{\sigma}_a^2)$  is of the form (1, 1, ..., 1, q, q) then the pattern (1, 1, ..., 1, k) does not necessarily give the next largest. For example, with  $\sigma_a^2 = 10$  pattern  $P_4$  gives  $\text{var}(\hat{\sigma}_a^2) = 113$  and  $P_2$ , not  $P_5$ , gives the next largest value of  $\text{var}(\hat{\sigma}_a^2)$ , namely 90.

(3) Increasing the total number of observations can, in some situations, actually increase  $\text{var}(\hat{\sigma}_a^2)$ . Thus for  $\sigma_a^2 \geq 1$ , all values of  $\text{var}(\hat{\sigma}_a^2)$  are greater with  $P_6$  than they are with  $P_5$ ; and for  $\sigma_a^2 \geq \frac{1}{2}$  the values are greater with  $P_9$  than with  $P_8$ . In both instances N is increased solely by the addition of more observations to one group. In similar fashion, comparisons of  $P_4$  and  $P_5$  with  $P_8$  and  $P_9$  appear to indicate that increasing the numbers of observations in all groups will reduce  $\text{var}(\hat{\sigma}_a^2)$  appreciably only when  $\sigma_a^2$  (in reality  $\sigma_a^2/\sigma_e^2$ ) is small. This is as one might expect, and it would appear to indicate the reasonable suggestion that any definition of a statistic for unbalancedness will have to depend on the value of  $\sigma_a^2$ .

Indications of the above nature suggested by Table 1 must be considered tentative, due to the limited nature of the table with regard to n-patterns. Considerable extension is needed and is being undertaken, as are further attempts to manipulate the expression for  $\text{var}(\hat{\sigma}_a^2)$  analytically.

#### Frequency distributions of simulated components

For each of the 9 n-patterns,  $P_1$  through  $P_9$ , with each of the 11 values of  $\sigma_a^2$ , 2000 simulations of  $\hat{\sigma}_a^2$  were made. On each occasion  $\hat{\sigma}_e^2$  was simulated as a  $\chi_{N-c}^2$ -variate, using procedures given (a) in U.S. Steel (1962) for  $N-c \leq 30$  and (b) in Zelen and Severo (1964) for  $N-c > 30$ . Group means were simulated from a random sampling of 1000 values representing medians of 1000 equiprobable areas of the standardized normal distribution, Searle (1966). From

the simulated  $\hat{\sigma}_e^2$  and group means,  $\hat{\sigma}_a^2$  was calculated from (1) and (2) and a frequency distribution made of the resulting values, using intervals defined by  $\sigma_a^2 + k[s.e.(\hat{\sigma}_a^2)]$  for  $k$  ranging from -2.0 to +5.0 in steps of 1/14. This provided just under 100 intervals for the frequency distribution. In this manner frequency distributions, cumulative distributions and frequency polygons were produced for the 2000 simulations of each  $n$ -pattern  $\times \sigma_a^2$  combination studied.

Perusal of the frequency polygons appears to indicate the following results for the  $n$ -patterns studied.

(1) The distribution of  $\hat{\sigma}_a^2$  is akin to a  $\chi^2$  distribution.

(2) For balanced and near-balanced  $n$ -patterns there is relatively little change in the distribution of  $\hat{\sigma}_a^2$  for changes in  $\sigma_a^2$ . But the distribution tends toward the exponential for large values of  $\sigma_a^2$  in unbalanced  $n$ -patterns, especially in the (1, 1, 1, 11, 11) case.

(3) In almost all cases 95% of the estimates  $\hat{\sigma}_a^2$  lay between  $\sigma_a^2 - 1.5(s.e.)$  and  $\sigma_a^2 + 2(s.e.)$ .

(4) For small values of  $\sigma_a^2$  (i.e. of the ratio  $\sigma_a^2/\sigma_e^2$ ), there were many instances of negative estimates of  $\hat{\sigma}_a^2$ . Indication of the extent of this occurrence is shown in Table 2, wherein is given the percentage of the 2000 simulated values of  $\hat{\sigma}_a^2$  that were negative in various  $n$ -pattern  $\times \sigma_a^2$  combinations. It is clear that for situations in which  $\sigma_a^2$  is close to zero there may be many negative estimates  $\hat{\sigma}_a^2$ , but even when  $\sigma_a^2/\sigma_e^2$  is in the neighborhood of 0.25 to 0.50 there still seems to be an appreciable likelihood of getting a negative estimate. If this is indeed the case it gives credence to results often obtained by geneticists, animal breeders and others for whom the variance ratio is customarily in this range.

#### Monte Carlo methods

The procedure described above is purely one of simulation. It makes no use of a (conditional) distribution property already known about  $\hat{\sigma}_a^2$ . This can be utilized in a method which we call Monte Carlo, distinct from the method already discussed, which we henceforth call the simulation method.

With  $SSB = (c-1)MSB$  and  $SSW = (N-c)MSW$ , the between- and within-group sums of squares respectively, equation (2) can be written as

$$\hat{\sigma}_a^2 = \lambda_1 SSB + \lambda_2 (SSW/\sigma_e^2) \quad - - - (4)$$

for

$$\lambda_2 = \frac{N(c-1)\sigma_e^2}{(N^2 - S_2)(N-c)} \quad - - - (5)$$

and  $\lambda_1 = N/(N^2 - S_2)$ . From (4) it is seen at once that the conditional variable  $(\hat{\sigma}_a^2 | SSB)$  has a  $\chi^2$  distribution (multiplied by a constant). Therefore, for any interval  $I_k$  on the real line, one can simulate SSB and from tables, or directly, calculate the probability  $p_k = \Pr\{\hat{\sigma}_a^2 | SSB \in I_k\}$ . On dividing the real line into  $n$  intervals  $I_k$ ,  $k = 1, 2, \dots, n$ ,  $p_k$  can then be found for every interval for each simulated SSB, and averaging each  $p_k$  over a series of simulations would give an estimated probability density function of  $\hat{\sigma}_a^2$ . This procedure is described in detail in Searle et al (1966). Its apparent advantage over the simulation method is that each simulated SSB contributes information to every one of the  $I_k$  intervals, whereas in the simulation method each simulated  $\hat{\sigma}_a^2$  contributes information to only one interval. Hopefully, for equivalent information about the whole curve, this should mean that the Monte Carlo procedure would require less simulations (of SSB) than would the simulation method (of  $\hat{\sigma}_a^2$ ). Unfortunately this advantage does not always occur in practice. The difficulty is that even if the interval  $I_k$  has length  $[s.e.(\hat{\sigma}_a^2)]/14$  say, as used in the simulation method, then  $p_k$  is the probability that a  $\chi^2$ -variable lies in an interval of length  $[s.e.(\hat{\sigma}_a^2)]/14\lambda_2$ , with  $\lambda_2$  being as given in (5). And this interval can turn out to be so large that the probability content of five, or even fewer, adjacent intervals can be close to 1.00, leaving other intervals with zero probability. For example, Table 1 with  $\sigma_a^2 = 1$  and  $n$ -pattern (1, 1, 1, 11, 11)  $\text{var}(\hat{\sigma}_a^2) = 1.41$ , and so  $[s.e.(\hat{\sigma}_a^2)]/14 = 0.0842$ . But  $\lambda_2 = 25(4)/380(20) = 1/76$  so that the interval has length  $76(0.0842) = 6.4$ ; and the 1% and 99% points respectively of the  $\chi_{20}^2$  distribution are 8.26 and 37.57. It is clear that five adjacent intervals include nearly all of the probability. Hence in this case a simulated SSB would be contributing non-zero information not to all the intervals but only to about five of them. Furthermore, computer time for calculating the probabilities  $p_k$  exceeds that of calculating additional  $\hat{\sigma}_a^2$  values in the simulation method. Thus the apparent advantage of the Monte Carlo method does not materialize.

Approximations to Wang's distribution

In the case of balanced data, with all  $n_1$  equal to  $n$  and  $N = cn$ ,  $SSB/(n\sigma_a^2 + \sigma_e^2)$  and  $SSW/\sigma_e^2$  are independently distributed as  $\chi_{c-1}^2$  and  $\chi_{N-c}^2$ . In this context  $\hat{\sigma}_a^2$  of (2) can be written as

$$\hat{\sigma}_a^2 = \mu_1 \chi_{c-1}^2 + \mu_2 \chi_{N-c}^2 \quad \dots (6)$$

where

$$\mu_1 = \frac{n\sigma_a^2 + \sigma_e^2}{n(c-1)} \quad \text{and} \quad \mu_2 = \frac{\sigma_e^2}{nc(n-1)} \quad \dots (7)$$

Wang (1966) has derived the distribution function of the general expression (6) and has suggested that with unbalanced data the distribution of  $\hat{\sigma}_a^2$  might be approximated by

$$\hat{\sigma}_a^2 = \alpha \chi_q^2 + \lambda_2 \chi_{N-c}^2, \quad \dots (8)$$

deriving values for  $\alpha$  and  $q$  by fitting the first two moments of  $\hat{\sigma}_a^2$ , with  $\lambda_2$  being the constant given in (5). Thus  $\alpha$  and  $q$  are determined from

$$E(\hat{\sigma}_a^2) = \sigma_a^2 = \alpha q + \lambda_2(N - c)$$

and

$$\text{var}(\hat{\sigma}_a^2) = 2\alpha^2 q + 2\lambda_2^2(N - c)$$

Using expression (3) for the left-hand side of the second of these equations they can be solved, writing  $\tau \equiv \sigma_e^2/\sigma_a^2$ , as

$$\alpha = \frac{\sigma_e^2 [N^2(c-1) + 2N(N^2 - S_2)\tau + (N^2 S_2 + S_2^2 - 2NS_3)\tau^2]}{(N^2 - S_2)[(N^2 - S_2)\tau + N(c-1)]} \quad \dots (9)$$

and

$$q = \frac{(c-1)[N^2(c-1) + 2N(N^2 - S_2)\tau + \frac{(N^2 - S_2)^2}{c-1} \tau^2]}{[N^2(c-1) + 2N(N^2 - S_2)\tau + (N^2 S_2 + S_2^2 - 2NS_3)\tau^2]} \quad \dots (10)$$

For balanced data these expressions reduce to  $\alpha = \mu_1$  and  $q = c-1$ , as one would expect from equation (6). Notice from (10) that  $q$ , the approximated degrees of freedom, is a multiple of  $(c-1)$ , the degrees of freedom in the balanced case, and would exceed this value in any situation in which  $(N^2 - S_2)^2 > (c-1)(N^2 S_2 + S_2^2 - 2NS_3)$ . Furthermore, for  $\sigma_e^2 = 0$ ,  $q = c-1$  on all occasions with  $\alpha$  then being  $\alpha = N\sigma_a^2/(N^2 - S_2)$ .

Computed values of  $\alpha$  and  $q$  (with  $\sigma_e^2 = 1$ ) for the  $n$ -patterns and  $\sigma_a^2$  values of Table 1 are shown in Table 3. Obvious trends in those values are that for each  $n$ -pattern  $\alpha$  increases and  $q$  decreases as  $\sigma_a^2$  increases; and for the  $n$ -patterns used there no value of  $q$  exceeds  $c-1$ .

From the entries in Table 3 it would be dangerous to speculate on the significance of trends in  $\alpha$  and/or  $q$  in terms of unbalancedness, because the table is so limited in extent. Hopefully a more extensive tabulation might indicate possible lines of approach for developing an index of unbalancedness. Clearly it will depend on  $\sigma_a^2$ .

As final comment we can report that plotting some of the approximated distributions alongside the frequency distribution of simulated  $\hat{\sigma}_a^2$  values has indicated good agreement in the cases tried so far. An example, of  $P_2$ , is shown in Wang (1966).

#### Acknowledgements

Grateful thanks go to D. A. Evans and E. C. Townsend for their work in preparing the computer programs in this study.

References

- Eisenhart, C. (1947). The assumptions underlying the analysis of variance. *Biometrics* 3, 1.
- Henderson, C. R. (1953). Estimation of variance and covariance components. *Biometrics* 9, 226-252.
- Herbach, L. H. (1959). Properties of Model-II type analysis of variance tests. *Ann. Math. Stat.* 30, 939-959.
- Robson, D. S. (1965). On the admissibility of non-negative maximum likelihood estimators of the "between-groups" variance component. Paper No. BU-179-M in the mimeo series, Biometrics Unit, Cornell University, Ithaca, N.Y.
- Searle, S. R. (1956). Matrix methods in variance and covariance components analysis. *Ann. Math. Stat.* 27, 737.
- Searle, S. R. (1966). Properties of certain discrete distributions for generating approximately normal variables. Paper No. BU-228-M, mimeo series, Biometrics Unit, Cornell University, Ithaca, N.Y.
- Searle, S. R., Robson, D. S. and Wang, Y. Y. (1966). Monte Carlo procedures for variance components. Paper No. BU-214-M in the mimeo series, Biometrics Unit, Cornell University, Ithaca, N.Y.
- Thompson, W. A. (1962). The problem of negative estimates of variance components. *Ann. Math. Stat.* 33, 273-289.
- U. S. Steel (1962). The generation of random samples from common statistical distributions. Technical Report 25.17-016, U. S. Steel, Monroeville, Pa.
- Wang, Y. Y. (1966). The distributions and moments of some variance component estimators. **Accepted** for publication in *Biometrika*.
- Zelen, M., and Severo, N. C. (1964). Probability functions. Chapter 26, "Handbook of Mathematical Functions", Nat. Bur. of Standards.

Table 1. Values of  $\text{var}(\hat{\sigma}_a^2)$  for 9 n-patterns and 11 sets of  $\sigma_a^2$ , with  $\sigma_e^2 = 1$ .

n-pattern		Values of $\sigma_a^2$										
		0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1	2	3	4	5	10	20
P <sub>1</sub>	5 5 5 5 5	.024	.105	.25	.46	.72	2.4	5.1	9	14	52	204
P <sub>2</sub>	1 1 3 10 10	.035	.150	.37	.71	1.15	4.0	8.6	15	23	90 <sup>†</sup>	354
P <sub>3</sub>	1 1 7 8 8	.030	.134	.33	.62	1.01	3.5	7.5	13	20	78	308
P <sub>4</sub>	1 1 1 11 11	.042	.177	.45	.86	1.41	5.0*	10.8	19	29	113 <sup>†</sup>	448
P <sub>5</sub>	1 1 1 1 21	.185	.374	.66	1.05*	1.53*	4.5	9.0	15	23	85	330
P <sub>6</sub>	1 1 1 21 21	.014	.141	.44	.92	1.56 <sup>⊗</sup>	5.9	13.1	23	36	141	562
P <sub>7</sub>	5 5 5 55 55	.001	.084	.30	.66	1.16	4.5	10.1	18	28	112	444
P <sub>8</sub>	5 5 5 5 105	.006	.084	.26	.54	.91	3.4	7.5	13	20	81	320
P <sub>9</sub>	5 5 5 105 105	.001	.096	.37	.81	1.43	5.7	12.7	22	35	140	560

\* For given N and  $\sigma_a^2$ ,  $\text{var}(\hat{\sigma}_a^2)$  is usually maximum for n-pattern (1, 1, 1, 1, k) or (1, 1, 1, k, k).

† If n-pattern (1, 1, 1, k, k) gives maximum  $\text{var}(\hat{\sigma}_a^2)$ , pattern (1, 1, 1, 1, k) does not necessarily give next largest.

⊗ Increasing N by increasing n in just one group can increase  $\text{var}(\hat{\sigma}_a^2)$ .

Table 2. Approximate percentage of negative estimates of  $\sigma_a^2$  in 2000 simulations.

n-pattern						Value of $\sigma_a^2$						
						0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1	2	3
<u>Per Cent</u>												
P <sub>1</sub>	5	5	5	.5	5	56%	23	14	7	5	4	1
P <sub>2</sub>	1	1	3	10	10	56	31	19	13	13	6	2
P <sub>3</sub>	1	1	7	8	8	58	29	20	15	12	6	3
P <sub>4</sub>	1	1	1	11	11	59	32	23	18	14	7	6
P <sub>5</sub>	1	1	1	1	21	59	40	31	29	23	15	8
P <sub>6</sub>	1	1	1	21	21	60	29	19	17	15	8	8
P <sub>9</sub>	5	5	5	5	105	63	17	10	6	6	3	1

Table 3. Values of  $\alpha$  and  $q$  in the approximation  $\hat{\sigma}_a^2 = \alpha \chi_q^2 - \lambda_2 \chi_{N-c}^2$  obtained by fitting the first two moments of  $\hat{\sigma}_a^2$  [see equations (9) and (10)].

n-pattern		Values of $\sigma_a^2$										
		0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1	2	3	4	5	10	20
		<u>The multiplier <math>\alpha</math>:</u>										
P <sub>1</sub>	5 5 5 5 5	.05	.11	.18	.24	.30	.55	.80	1.05	1.30	2.55	4.05
P <sub>2</sub>	1 1 3 10 10	.06	.15	.25	.35	.46	.89	1.33	1.77	2.20	4.39	8.76
P <sub>3</sub>	1 1 7 8 8	.06	.14	.23	.31	.41	.79	1.17	1.55	1.93	3.82	7.62
P <sub>4</sub>	1 1 1 11 11	.07	.17	.29	.42	.56	1.10	1.65	2.20	2.76	5.52	11.05
P <sub>5</sub>	1 1 1 1 21	.14	.21	.30	.39	.48	.87	1.26	1.66	2.06	4.04	8.02
P <sub>6</sub>	1 1 1 21 21	.04	.17	.34	.51	.68	1.37	2.07	2.77	3.46	6.96	13.94
P <sub>7</sub>	5 5 5 55 55	.01	.14	.28	.41	.55	1.10	1.66	2.21	2.76	5.53	11.06
P <sub>8</sub>	5 5 5 5 105	.03	.12	.21	.31	.41	.81	1.20	1.60	2.00	3.99	7.97
P <sub>9</sub>	5 5 5 105 105	.01	.17	.34	.51	.69	1.39	2.09	2.78	3.49	6.98	13.97
		<u>Degrees of freedom <math>q</math>:</u>										
P <sub>1</sub>	5 5 5 5 5	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00
P <sub>2</sub>	1 1 3 10 10	4.00	3.35	2.98	2.80	2.69	2.50	2.43	2.40	2.38	2.33	2.31
P <sub>3</sub>	1 1 7 8 8	4.00	3.50	3.20	3.06	2.97	2.82	2.76	2.73	2.71	2.67	2.66
P <sub>4</sub>	1 1 1 11 11	4.00	3.11	2.63	2.40	2.27	2.05	1.97	1.93	1.91	1.85	1.83
P <sub>5</sub>	1 1 1 1 21	4.00	3.78	3.53	3.35	3.21	2.93	2.81	2.75	2.70	2.61	2.56
P <sub>6</sub>	1 1 1 21 21	4.00	2.38	1.96	1.80	1.71	1.57	1.53	1.50	1.49	1.46	1.45
P <sub>7</sub>	5 5 5 55 55	4.00	2.19	2.00	1.94	1.91	1.86	1.84	1.83	1.83	1.82	1.81
P <sub>8</sub>	5 5 5 5 105	4.00	3.12	2.86	2.76	2.70	2.61	2.57	2.56	2.55	2.53	2.52
P <sub>9</sub>	5 5 5 105 105	4.00	1.65	1.54	1.51	1.49	1.46	1.45	1.45	1.44	1.44	1.43