Variance Components in the Two-Way Nested Model with Incomplete Nesting Information
Author(s): Ruiguang Song and Stanley A. Shulman

Source: *Technometrics*, Vol. 39, No. 1 (Feb., 1997), pp. 71-80
Published by: Taylor & Francis, Ltd. on behalf of American Statistical Association and American Society for Quality
Stable URL: https://www.jstor.org/stable/1270774

Variance Components in the Two-Way Nested Model With Incomplete Nesting Information

Ruiguang SONG

Computer Sciences Corporation/NIOSH
Cincinnati, OH 45226

Stanley A. SHULMAN

U.S. Department of Health and Human Services,
Public Health Service
Centers for Disease Control and Prevention
National Institute for Occupational Safety and Health
Cincinnati, OH 45226

When part of the nesting information in a two-way nested model is missing, there has been no way to use all of the data in the analysis of variance components. Discarding the data with missing nesting information loses useful information, and in some circumstances, this approach cannot separate variance components from one another. To make use of the data with missing nesting information, computable sums of squares for the data with missing nesting information can be linearly combined with sums of squares for the data with complete nesting information. Prespecified weights are needed for the combination. Different estimates are obtained by using different weights. Because all of these estimators are unbiased, variances and covariances of these estimators are derived and used to compare these estimators. In addition, a simulation study is conducted to provide evidence for the reliability of the variances and covariances formulas. Finally, as an application, the proposed methods are applied to the analysis of a proficiency testing program.

KEY WORDS: Missing nesting information, Prespecified weights, Unbiased estimators.

1. THE MISSING–NESTING–INFORMATION PROBLEM

In regression analysis, independent variables may have missing values in practice. It is also likely that nesting information (which group or subgroup an observation belongs to) in the analysis of variance is missing. The nesting information in a nested model has the same importance as the independent variables in regression analyses. Without the nesting information, the variance components in the nested model cannot be separated from one another. Although the missing-data problem has been studied extensively (see Little and Rubin 1987; Little 1992), the missing-nesting-information problem has not been discussed in literature. In this article, we introduce the missing-nesting-information problem with a two-way nested model and investigate estimation methods that can use all of the data including the data with missing nesting information. The two-way nested model considered in this article is

$$y_{ijk} = \mu + \alpha_i + \beta_{ij} + e_{ijk},$$
$$i = 1, \ldots, a;\ \ j = 1, \ldots, b_i;\ \ k = 1, \ldots, n_{ij}, \quad (1)$$

where μ is the general mean and the α_i's, β_{ij}'s, and e_{ijk}'s are mutually independent random variables with zero means and variances $\sigma^2_\alpha, \sigma^2_\beta$, and σ^2_e, respectively.

In the two-way nested model, each observation is associated with two nesting (group) variables i and j. The subscripts i and j represent the main group and the subgroup that an observation belongs to, respectively. Like the missing-data problem in regression analysis, the nesting information could be lost, or perhaps it was not recorded. The information could be incomplete in different ways. To be specific, we consider the case in which the missing information occurs on subgroup j only. That is, for each observation, we know which main group i, but may not know which subgroup j, that the observation comes from. In each main group i, the subgroup information may be missing (a) completely (no observation in the main group has subgroup information), (b) partly (some of the observations in the main group have missing subgroup information), or (c) not at all on any observation in the main group. The three types of missing information can be described by the example shown in Table 1, with $b_i = 3$ and $n_{ij} = 3, j = 1, 2, 3$, in one of the main groups of Model (1).

It is very likely in practice that in a main group, either no observation or all observations in the main group have missing subgroup information. To avoid overly complicated formulas, we consider data completely missing the nesting information (a) or with no data missing the nesting information (c) in each main group. By following the method presented in this article and tedious algebra, one can derive similar formulas for data with partially missing nesting information (b) in some main groups.

One example showing the missing-nesting-information problem is found in the asbestos-fiber-count data from the American Industrial Hygiene Association (AIHA) Asbestos Analyst Registry (AAR) Program. This program was developed by AIHA in cooperation with researchers at the National Institute for Occupational Safety and Health (NIOSH) to ensure the quality of asbestos measurements needed to avoid exposing workers and the public to hazardous airborne concentrations of asbestos. In this personnel certification program, samples are produced in a generation chamber. Filters are used to collect asbestos fibers.

Table 1. Examples of Data With Different Types of Missing Nesting Information

Type of missing information	Data with complete nesting information	Data with incomplete nesting information
(a)	$\underline{y_{i11}}, \underline{y_{i12}}, \underline{y_{i13}}$	$y_{i??}, y_{i??}, y_{i??}, y_{i??}, y_{i??}$
	$\underline{y_{i21}}, \underline{y_{i22}}, \underline{y_{i23}}$	$y_{i??}, y_{i??}, y_{i??}, y_{i??}$
	$\underline{y_{i31}}, \underline{y_{i32}}, \underline{y_{i33}}$	
(b)	$y_{i11}, y_{i12}, y_{i13}$	$y_{i11}, y_{i12}, y_{i13}$
	$y_{i21}, y_{i22}, \underline{y_{i23}}$	y_{i21}, y_{i22}
	$\underline{y_{i31}}, \underline{y_{i32}}, \underline{y_{i33}}$	$y_{i??}, y_{i??}, y_{i??}, y_{i??}$
(c)	$y_{i11}, y_{i12}, y_{i13}$	$y_{i11}, y_{i12}, y_{i13}$
	$y_{i21}, y_{i22}, y_{i23}$	$y_{i21}, y_{i22}, y_{i23}$
	$y_{i31}, y_{i32}, y_{i33}$	$y_{i31}, y_{i32}, y_{i33}$

NOTE: The third column is the outcome of the second column when the nesting information is missing for the data with underscores.

Four runs of the generation system are used to produce four batches of samples (labeled as samples A, B, C, and D), each at a different concentration. Taking one sample from each of these batches produces a set of four samples (filters). Participating organizations receive one or more sets of samples each quarter (termed a round). Each counter analyzes a wedge from each filter in one of these sets. According to the AAR guidelines, up to five counters in an organization can share a set of samples. Because counters within an organization may be located at various sites, the participating organization takes the filters containing asbestos fibers, cuts the filters into wedges (up to five wedges per filter), mounts the wedges onto microscope slides, and distributes the slides to the organization's counters. For example, a participating organization with 10 counters will receive two sets of samples. Each filter in a set will be cut into five wedges. Five of the 10 counters will share a sample set, and each counter analyzes a wedge from each of the four filters in the set (see Fig. 1). Each counting result is reported with the counter's identification number and his or her organization's identification number. Unfortunately, the filter from which the counter counts a wedge was not identified and reported. This important information was not recorded because the program was not designed to estimate each individual variance. Instead, its aim is counter certification. The need for estimates of each variance component was not recognized until recently, when a study about asbestos counter performance was undertaken at NIOSH. To evaluate the sample quality, counters' performances, and the differences among organizations, all of these variance components are important and must be estimated separately. Estimation of these variance components without knowing the beforementioned nesting information was an important problem in making full use of the AAR Program data. The analysis of the AAR data is discussed further in Section 4.

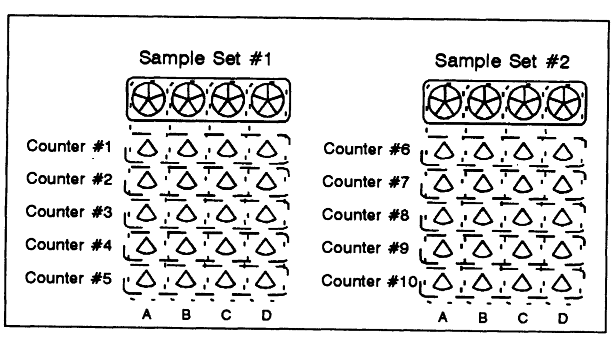

Figure 1. Sample Distribution Within an Organization With 10 Counters.

Some procedures developed for data with complete nesting information cannot be applied when this information is incomplete. Particularly, likelihood-based procedures cannot be applied because the variance–covariance matrix of the joint distribution of the y_{ijk}'s cannot be specified when nesting information is missing. For example, if u_1 and u_2 are two of the y_{ijk}'s, and if $u_1 = y_{111}$ and $u_2 = y_{112}$, then $\text{cov}(u_1, u_2) = \sigma_\alpha^2 + \sigma_\beta^2$. On the other hand, if $u_1 = y_{111}$ and $u_2 = y_{121}$, then $\text{cov}(u_1, u_2) = \sigma_\alpha^2$. In other words, the covariance between two observations cannot be specified unless we know the i and j values associated with the observations.

In Section 2, we start with a simple method that discards the data with missing nesting information and uses the data with complete nesting information only. Then we consider how to make use of the data with missing nesting information. Based on an analysis of variance (ANOVA) procedure, the computable sums of squares from the data with missing nesting information are combined with the sums of squares from the data with complete nesting information. These sums of squares are combined in three different ways. One uses equal weights, one uses the weights that minimize the variances of the combined sums of squares, and one uses the weights that minimize the variances of resulting estimators. The sampling variances and covariances of the derived estimators are also obtained. Estimators are compared in Section 3. A simulation study is conducted to show the accuracy of the variance and covariance formulas. Comparison and simulation results are shown in Tables 2, 3, and 4. Finally, an application to the AAR data analysis is discussed in Section 4. One of the datasets from sample B of round 2 is listed in Table 5, and results from the first 10 rounds are shown in Table 6.

2. ESTIMATES USING DATA WITH MISSING NESTING INFORMATION

Let a' and $a'' = a - a'$ be the numbers of main groups with no missing (c) and complete missing (a) subgroup nesting information, as discussed in Section 1, respectively. Let the first a' main groups be the groups with no missing nesting information. In this article, we assume that all b_i's and n_{ij}'s in Model (1) are known. An example of this case is when all subgroups have equal numbers of observations (all $n_{ij} = n$, a known constant) and the total number of observations in each main group is known ($n_{i\cdot} = \sum_{j=1}^{b_i} n_{ij}$). Then the number of subgroups can be determined as $b_i = n_{i\cdot}/n$.

Let $N = \sum_i n_{i\cdot}$, and let $\text{SS}_e = T_0 - T_{ab}$, $\text{SS}_b = T_{ab} - T_a$, $\text{SS}_a = T_a - T_\mu$ denote the within-subgroups, between-subgroups, within-main-groups, and between-main-groups sums of squares, respectively, where

Table 2. Simulation Results With 1,000 Replicates for Balanced Design with a = 20, b = 2, and n = 5: a′ = 5 —75% Data With Missing Nesting Information

σ_e^2	σ_β^2	σ_α^2	(M)	r_b	r_a	$m[\tilde\sigma_e^2]$	$s[\tilde\sigma_e^2](\sigma[\tilde\sigma_e^2])$	$m[\tilde\sigma_\beta^2]$	$s[\tilde\sigma_\beta^2](\sigma[\tilde\sigma_\beta^2])$	$m[\tilde\sigma_\alpha^2]$	$s[\tilde\sigma_\alpha^2](\sigma[\tilde\sigma_\alpha^2])$	$\hat\rho_{e\beta}(\rho_{e\beta})$	$\hat\rho_{e\alpha}(\rho_{e\alpha})$	$\hat\rho_{\beta\alpha}(\rho_{\beta\alpha})$
.04	.01	.01	(1)	1.00	1.00	.0399	.0088(.0089)	.0105	.0117(.0115)	.0093	.0149(.0146)	−.22(−.16)	.02(.00)	−.40(−.39)
			(2)	.50	.50			.0102	.0150(.0151)	.0096	.0093(.0093)	−.81(−.83)	.53(.58)	−.71(−.73)
			(3)	.89	.78			.0104	.0108(.0105)	.0095	.0093(.0090)	−.57(−.55)	.22(.22)	−.52(−.53)
			(4)	.91	.50			.0104	.0107(.0104)	.0095	.0082(.0080)	−.52(−.49)	.21(.21)	−.59(−.60)
		.04	(1)	1.00	1.00	.0401	.0088(.0089)	.0103	.0115(.0115)	.0402	.0340(.0351)	−.18(−.16)	−.04(.00)	−.12(−.16)
			(2)	.50	.50			.0096	.0146(.0151)	.0412	.0178(.0177)	−.83(−.83)	.27(.30)	−.34(−.38)
			(3)	.89	.78			.0100	.0103(.0105)	.0407	.0198(.0202)	−.56(−.55)	.06(.10)	−.19(−.23)
			(4)	.91	.50			.0101	.0103(.0104)	.0410	.0172(.0170)	−.51(−.49)	. 07(.10)	−.25(−.28)
		.09	(1)	1.00	1.00	.0401	.0092(.0089)	.0105	.0120(.0115)	.0884	.0696(.0702)	−.17(−.16)	.07(.00)	−.13(−.08)
			(2)	.50	.50			.0100	.0154(.0151)	.0901	.0338(.0337)	−.84(−.83)	.17(.16)	−.22(−.20)
			(3)	.89	.78			.0103	.0108(.0105)	.0894	.0400(.0400)	−.56(−.55)	.10(.05)	−.19(−.12)
			(4)	.91	.50			.0104	.0108(.0104)	.0899	.0337(.0334)	−.50(−.49)	.06(.05)	−.20(−.14)
	.04	.01	(1)	1.00	1.00	.0401	.0088(.0089)	.0399	.0302(.0304)	.0105	.0286(.0284)	−.07(−.06)	.02(.00)	−.54(−.53)
			(2)	.50	.50			.0397	.0207(.0206)	.0104	.0148(.0150)	−.63(−.61)	.40(.36)	−.66(−.65)
			(3)	.79	.78			.0398	.0206(.0206)	.0104	.0167(.0169)	−.42(−.40)	.22(.19)	−.59(−.59)
			(4)	.67	.50			.0397	.0198(.0198)	.0104	.0146(.0148)	−.55(−.52)	.33(.29)	−.64(−.64)
		.04	(1)	1.00	1.00	.0400	.0090(.0089)	.0393	.0308(.0304)	.0382	.0489(.0477)	−.09(−.06)	.05(.00)	−.28(−.32)
			(2)	.50	.50			.0399	.0208(.0206)	.0391	.0239(.0235)	−.60(−.61)	.25(.23)	−.41(−.42)
			(3)	.79	.78			.0397	.0207(.0206)	.0388	.0277(.0275)	−.42(−.40)	.16(.12)	−.31(−.36)
			(4)	.67	.50			.0398	.0199(.0198)	.0392	.0236(.0233)	−.53(−.52)	.21(.18)	−.38(−.40)
		.09	(1)	1.00	1.00	.0402	.0087(.0089)	.0404	.0313(.0304)	.0903	.0812(.0820)	−.07(−.06)	−.03(.00)	−.19(−.18)
			(2)	.50	.50			.0396	.0203(.0206)	.0889	.0373(.0392)	−.57(−.61)	.15(.14)	−.26(−.25)
			(3)	.79	.78			.0399	.0210(.0206)	.0894	.0452(.0468)	−.37(−.40)	.06(.07)	−.21(−.21)
			(4)	.67	.50			.0398	.0198(.0198)	.0888	.0372(.0392)	−.49(−.52)	.12(.11)	−.25(−.24)
	.09	.01	(1)	1.00	1.00	.0398	.0088(.0089)	.0898	.0607(.0620)	.0097	.0500(.0520)	−.06(−.03)	−.02(.00)	−.55(−.60)
			(2)	.50	.50			.0902	.0341(.0340)	.0100	.0254(.0258)	−.40(−.37)	.21(.21)	−.63(−.65)
			(3)	.76	.78			.0900	.0371(.0375)	.0099	.0291(.0301)	−.27(−.23)	.11(.12)	−.58(−.62)
			(4)	.55	.50			.0902	.0339(.0338)	.0101	.0253(.0258)	−.39(−.35)	.20(.20)	−.63(−.64)
		.04	(1)	1.00	1.00	.0400	.0089(.0089)	.0867	.0594(.0620)	.0466	.0727(.0701)	−.01(−.03)	.01(.00)	−.34(−.44)
			(2)	.50	.50			.0882	.0325(.0340)	.0418	.0350(.0340)	−.36(−.37)	.12(.16)	−.42(−.49)
			(3)	.76	.78			.0877	.0356(.0375)	.0435	.0417(.0402)	−.22(−.23)	.07(.09)	−.36(−.46)
			(4)	.55	.50			.0881	.0323(.0338)	.0418	.0350(.0340)	−.34(−.35)	.12(.15)	−.42(−.49)
		.09	(1)	1.00	1.00	.0402	.0089(.0089)	.0938	.0656(.0620)	.0841	.0994(.1031)	−.05(−.03)	.00(.00)	−.29(−.30)
			(2)	.50	.50			.0921	.0352(.0340)	.0876	.0488(.0493)	−.42(−.37)	.09(.11)	−.32(−.34)
			(3)	.76	.78			.0927	.0391(.0375)	.0864	.0579(.0587)	−.27(−.23)	.05(.06)	−.30(−.32)
			(4)	.55	.50			.0922	.0351(.0338)	.0876	.0488(.0492)	−.40(−.35)	.08(.10)	−.32(−.34)

$$
\begin{aligned}
T_0 &= \sum_i \sum_{j=1}^{b_i} \sum_{k=1}^{n_{ij}} y_{ijk}^2 \\
T_{ab} &= \sum_i \sum_{j=1}^{b_i} \left(\sum_{k=1}^{n_{ij}} y_{ijk} \right)^2 \Big/ n_{ij} \\
T_a &= \sum_i \left(\sum_{j=1}^{b_i} \sum_{k=1}^{n_{ij}} y_{ijk} \right)^2 \Big/ n_{i\cdot} \\
T_\mu &= \left(\sum_i \sum_{j=1}^{b_i} \sum_{k=1}^{n_{ij}} y_{ijk} \right)^2 \Big/ N
\end{aligned}
\tag{2}
$$

are the uncorrected sums of squares. These notations and the following k notations are adopted from Searle, Casella, and McCulloch (1992):

$$
\begin{aligned}
k_1 &= \sum_i n_{i\cdot}^2 / N; \qquad k_{12} = \sum_i \left(\sum_{j=1}^{b_i} n_{ij}^2 \right) \Big/ n_{i\cdot} \\
k_3 &= \sum_i \sum_{j=1}^{b_i} n_{ij}^2 / N; \qquad k_4 = \sum_i \sum_{j=1}^{b_i} n_{ij}^3 \\
k_5 &= \sum_i \left(\sum_{j=1}^{b_i} n_{ij}^3 \right) \Big/ n_{i\cdot}; \qquad k_6 = \sum_i \left(\sum_{j=1}^{b_i} n_{ij}^2 \right)^2 \Big/ n_{i\cdot} \\
k_7 &= \sum_i \left(\sum_{j=1}^{b_i} n_{ij}^2 \right)^2 \Big/ n_{i\cdot}^2; \qquad k_8 = \sum_i n_{i\cdot} \left(\sum_{j=1}^{b_i} n_{ij}^2 \right) \\
k_9 &= \sum_i n_{i\cdot}^3.
\end{aligned}
\tag{3}
$$

In this article we define variables and constants without prime notations or with single or double prime notations as follows. If there is a notation without a prime,

Table 3. Simulation Results With 1,000 Replicates for Balanced Design With a = 20, b = 2, and n = 5: a′ = 10 —50% Data With Missing Nesting Information

σ_e^2	σ_β^2	σ_α^2	(M)	r_b	r_a	$m[\tilde\sigma_e^2]$	$s[\tilde\sigma_e^2](\sigma[\tilde\sigma_e^2])$	$m[\tilde\sigma_\beta^2]$	$s[\tilde\sigma_\beta^2](\sigma[\tilde\sigma_\beta^2])$	$m[\tilde\sigma_\alpha^2]$	$s[\tilde\sigma_\alpha^2](\sigma[\tilde\sigma_\alpha^2])$	$\hat\rho_{e\beta}(\rho_{e\beta})$	$\hat\rho_{e\alpha}(\rho_{e\alpha})$	$\hat\rho_{\beta\alpha}(\rho_{\beta\alpha})$
.04	.01	.01	(1)	1.00	1.00	.0399	.0063(.0063)	.0100	.0081(.0081)	.0098	.0100(.0098)	−.14(−.16)	−.02(.00)	−.45(−.40)
			(2)	.50	.50			.0104	.0098(.0099)	.0096	.0078(.0078)	−.63(−.64)	.35(.32)	−.61(−.58)
			(3)	.72	.50			.0102	.0079(.0080)	.0097	.0073(.0073)	−.50(−.51)	.22(.19)	−.54(−.50)
			(4)	.84	.50			.0101	.0076(.0076)	.0097	.0072(.0073)	−.36(−.37)	.14(.11)	−.53(−.49)
		.04	(1)	1.00	1.00	.0399	.0063(.0063)	.0103	.0082(.0081)	.0408	.0237(.0234)	−.13(−.16)	−.02(.00)	−.13(−.17)
			(2)	.50	.50			.0104	.0097(.0099)	.0399	.0168(.0170)	−.62(−.64)	.14(.15)	−.21(−.27)
			(3)	.72	.50			.0103	.0079(.0080)	.0399	.0167(.0167)	−.48(−.51)	.07(.08)	−.17(−.22)
			(4)	.84	.50			.0103	.0077(.0076)	.0399	.0167(.0167)	−.34(−.37)	−.03(.05)	−.18(−.21)
		.09	(1)	1.00	1.00	.0401	.0065(.0063)	.0100	.0080(.0081)	.0891	.0457(.0468)	−.13(−.16)	−.02(.00)	−.09(−.08)
			(2)	.50	.50			.0098	.0102(.0099)	.0894	.0331(.0333)	−.65(−.64)	.06(.08)	−.15(−.14)
			(3)	.72	.50			.0099	.0081(.0080)	.0893	.0330(.0332)	−.51(−.51)	.02(.04)	−.13(−.11)
			(4)	.84	.50			.0099	.0076(.0076)	.0893	.0330(.0332)	−.36(−.37)	.00(.02)	−.13(−.11)
	.04	.01	(1)	1.00	1.00	.0404	.0064(.0063)	.0403	.0219(.0215)	.0093	.0185(.0193)	−.08(−.06)	.00(.00)	−.57(−.55)
			(2)	.50	.50			.0401	.0173(.0172)	.0096	.0142(.0141)	−.38(−.37)	.18(.18)	−.61(−.59)
			(3)	.55	.50			.0401	.0171(.0169)	.0096	.0141(.0140)	−.36(−.34)	.17(.16)	−.61(−.59)
			(4)	.60	.50			.0401	.0171(.0169)	.0096	.0141(.0140)	−.33(−.31)	.15(.14)	−.61(−.59)
		.04	(1)	1.00	1.00	.0400	.0066(.0063)	.0398	.0226(.0215)	.0384	.0305(.0320)	−.09(−.06)	.02(.00)	−.38(−.33)
			(2)	.50	.50			.0403	.0176(.0172)	.0395	.0234(.0229)	−.42(−.37)	.11(.11)	−.37(−.37)
			(3)	.55	.50			.0402	.0174(.0169)	.0396	.0233(.0229)	−.39(−.34)	.10(.10)	−.36(−.36)
			(4)	.60	.50			.0402	.0174(.0169)	.0396	.0234(.0229)	−.36(−.31)	.09(.09)	−.37(−.36)
		.09	(1)	1.00	1.00	.0401	.0062(.0063)	.0390	.0211(.0215)	.0921	.0530(.0548)	−.07(−.06)	.03(.00)	−.22(−.20)
			(2)	.50	.50			.0394	.0167(.0172)	.0902	.0381(.0389)	−.33(−.37)	.07(.07)	−.21(−.22)
			(3)	.55	.50			.0393	.0165(.0169)	.0902	.0381(.0389)	−.31(−.34)	.07(.06)	−.21(−.21)
			(4)	.60	.50			.0393	.0164(.0169)	.0903	.0381(.0389)	−.29(−.31)	.06(.05)	−.21(−.21)
	.09	.01	(1)	1.00	1.00	.0397	.0062(.0063)	.0888	.0422(.0438)	.0096	.0344(.0354)	−.05(−.03)	.06(.00)	−.60(−.62)
			(2)	.50	.50			.0892	.0303(.0320)	.0101	.0247(.0253)	−.22(−.20)	.11(.10)	−.60(−.63)
			(3)	.51	.50			.0892	.0303(.0320)	.0101	.0247(.0253)	−.22(−.19)	.11(.10)	−.60(−.63)
			(4)	.53	.50			.0892	.0303(.0320)	.0101	.0247(.0253)	−.21(−.19)	.11(.09)	−.60(−.63)
		.04	(1)	1.00	1.00	.0401	.0064(.0063)	.0896	.0421(.0438)	.0396	.0460(.0473)	−.04(−.03)	−.03(.00)	−.44(−.46)
			(2)	.50	.50			.0889	.0320(.0320)	.0402	.0333(.0337)	−.24(−.20)	.09(.08)	−.49(−.47)
			(3)	.51	.50			.0889	.0319(.0320)	.0402	.0333(.0337)	−.24(−.19)	.09(.07)	−.49(−.47)
			(4)	.53	.50			.0889	.0318(.0320)	.0402	.0333(.0337)	−.23(−.19)	.09(.07)	−.49(−.47)
		.09	(1)	1.00	1.00	.0403	.0065(.0063)	.0900	.0422(.0438)	.0877	.0667(.0691)	−.02(−.03)	−.01(.00)	−.31(−.32)
			(2)	.50	.50			.0901	.0304(.0320)	.0872	.0474(.0490)	−.17(−.20)	.04(.05)	−.32(−.32)
			(3)	.51	.50			.0901	.0303(.0320)	.0872	.0473(.0490)	−.17(−.19)	.04(.05)	−.32(−.32)
			(4)	.53	.50			.0901	.0303(.0320)	.0872	.0473(.0490)	−.17(−.19)	.03(.05)	−.32(−.32)

we do not specify the range for i if the variable or constant is summed over i. The same notation with a single prime (double primes) is then defined as the same quantity summed over i from 1 to a' (from $a'+1$ to a, respectively). For example, we define $b_\cdot = \sum_i b_i$, then $b'_\cdot = \sum_{i=1}^{a'} b_i$ and $b''_\cdot = \sum_{i=a'+1}^{a} b_i$. For $k_1 = \sum_i n_{i\cdot}^2/N$, then $k'_1 = \sum_{i=1}^{a'} n_{i\cdot}^2/N'$ and $k''_1 = \sum_{i=a'+1}^{a} n_{i\cdot}^2/N''$.

Due to missing nesting information in main groups from $a'+1$ to a, the subgroup sums $\sum_{k=1}^{n_{ij}} y_{ijk}$ in T''_{ab} are not computable. So the ANOVA procedure cannot be applied directly to the data for main groups $a'+1, \ldots, a$. To handle this situation simply, we may discard these data and then apply the ANOVA procedure to the data with complete nesting information. To do so, we set the sums of squares equal to their expectations:

$$\begin{aligned}
\mathrm{SS}'_e &= E(\mathrm{SS}'_e) = (N' - b'_\cdot)\sigma_e^2 \\
\mathrm{SS}'_b &= E(\mathrm{SS}'_b) = (N' - k'_{12})\sigma_\beta^2 + (b'_\cdot - a')\sigma_e^2 \\
\mathrm{SS}'_a &= E(\mathrm{SS}'_a) = (N' - k'_1)\sigma_\alpha^2 \\
&\quad + (k'_{12} - k'_3)\sigma_\beta^2 + (a'-1)\sigma_e^2. \qquad (4)
\end{aligned}$$

Solving these equations, we obtain the following unbiased estimators:

$$\begin{aligned}
\bar\sigma_e^2 &= \mathrm{SS}'_e/(N' - b'_\cdot) \\
\bar\sigma_\beta^2 &= [\mathrm{SS}'_b - (b'_\cdot - a')\bar\sigma_e^2]/(N' - k'_{12}) \\
\bar\sigma_\alpha^2 &= [\mathrm{SS}'_a - (k'_{12} - k'_3)\bar\sigma_\beta^2 - (a'-1)\bar\sigma_e^2]/(N' - k'_1). \qquad (5)
\end{aligned}$$

Note that discarding data is not simply a waste of information. It may make some variance components inestimable. For example, when $b_i = 1$ for $i \le a'$, then σ_α^2 and σ_β^2 cannot be separated by this method because no observations within the same main group are from different subgroups. The previously mentioned AAR dataset is such an example. In this case, $N' = k'_{12}, b'_\cdot = a'$, and hence $\mathrm{SS}'_b = 0$. The two-way nested model is actually reduced to a one-way nested model with the two variance components σ_α^2 and σ_β^2 confounded. Only the sum of these two components can be estimated:

$$\bar\sigma_\beta^2 + \bar\sigma_\alpha^2 = [\mathrm{SS}'_a - (a'-1)\bar\sigma_e^2]/(N' - k'_1). \qquad (6)$$

The two variance components can be separated by including the data with missing nesting information if not

Table 4. Simulation Results With 1,000 Replicates for Balanced Design With a = 20, b = 2, and n = 5: a′ = 15 —25% Data With Missing Nesting Information

σ_e^2	σ_β^2	σ_α^2	(M)	r_b	r_a	$m[\tilde\sigma_e^2]$	$s[\tilde\sigma_e^2](\sigma[\tilde\sigma_e^2])$	$m[\tilde\sigma_\beta^2]$	$s[\tilde\sigma_\beta^2](\sigma[\tilde\sigma_\beta^2])$	$m[\tilde\sigma_\alpha^2]$	$s[\tilde\sigma_\alpha^2](\sigma[\tilde\sigma_\alpha^2])$	$\hat\rho_{e\beta}(\rho_{e\beta})$	$\hat\rho_{e\alpha}(\rho_{e\alpha})$	$\hat\rho_{\beta\alpha}(\rho_{\beta\alpha})$
.04	.01	.01	(1)	1.00	1.00	.0398	.0052(.0052)	.0096	.0065(.0067)	.0102	.0081(.0079)	−.15(−.16)	−.01(.00)	−.40(−.41)
			(2)	.50	.50			.0097	.0073(.0074)	.0103	.0074(.0072)	−.42(−.42)	.14(.14)	−.48(−.48)
			(3)	.46	.22			.0097	.0076(.0077)	.0106	.0087(.0085)	−.43(−.43)	.14(.14)	−.42(−.43)
			(4)	.80	.50			.0096	.0062(.0064)	.0104	.0072(.0071)	−.26(−.26)	.05(.05)	−.44(−.44)
		.04	(1)	1.00	1.00	.0400	.0052(.0052)	.0099	.0068(.0067)	.0400	.0186(.0188)	−.13(−.16)	.02(.00)	−.18(−.17)
			(2)	.50	.50			.0101	.0074(.0074)	.0398	.0163(.0167)	−.37(−.42)	.07(.06)	−.20(−.21)
			(3)	.46	.22			.0101	.0077(.0077)	.0397	.0192(.0200)	−.39(−.43)	.06(.06)	−.14(−.18)
			(4)	.80	.50			.0100	.0065(.0064)	.0399	.0162(.0166)	−.22(−.26)	.03(.02)	−.18(−.19)
		.09	(1)	1.00	1.00	.0399	.0050(.0052)	.0103	.0064(.0067)	.0891	.0366(.0376)	−.13(−.16)	−.01(.00)	−.06(−.09)
			(2)	.50	.50			.0102	.0072(.0074)	.0905	.0332(.0332)	−.37(−.42)	.03(.03)	−.09(−.11)
			(3)	.46	.22			.0102	.0075(.0077)	.0922	.0421(.0400)	−.38(−.43)	.04(.03)	−.10(−.09)
			(4)	.80	.50			.0103	.0062(.0064)	.0905	.0331(.0331)	−.23(−.26)	.01(.01)	−.07(−.09)
	.04	.01	(1)	1.00	1.00	.0401	.0052(.0052)	.0398	.0175(.0176)	.0102	.0160(.0156)	−.05(−.06)	−.01(.00)	−.57(−.56)
			(2)	.50	.50			.0399	.0159(.0159)	.0104	.0142(.0138)	−.19(−.19)	.05(.07)	−.57(−.57)
			(3)	.29	.22			.0400	.0187(.0185)	.0107	.0168(.0164)	−.25(−.26)	.09(.11)	−.54(−.56)
			(4)	.57	.50			.0399	.0157(.0158)	.0104	.0142(.0138)	−.16(−.17)	.04(.06)	−.57(−.57)
		.04	(1)	1.00	1.00	.0401	.0050(.0052)	.0402	.0176(.0176)	.0406	.0260(.0257)	−.06(−.06)	.01(.00)	−.32(−.34)
			(2)	.50	.50			.0398	.0160(.0159)	.0408	.0231(.0227)	−.16(−.19)	.04(.05)	−.35(−.35)
			(3)	.29	.22			.0396	.0184(.0185)	.0409	.0279(.0272)	−.20(−.26)	.06(.07)	−.35(−.34)
			(4)	.57	.50			.0399	.0159(.0158)	.0407	.0230(.0227)	−.14(−.17)	.04(.04)	−.35(−.34)
		.09	(1)	1.00	1.00	.0403	.0051(.0052)	.0393	.0176(.0176)	.0890	.0440(.0440)	−.04(−.06)	−.00(.00)	−.20(−.20)
			(2)	.50	.50			.0391	.0159(.0159)	.0887	.0390(.0388)	−.18(−.19)	.01(.03)	−.19(−.20)
			(3)	.29	.22			.0390	.0184(.0185)	.0883	.0458(.0466)	−.25(−.26)	.00(.04)	−.17(−.20)
			(4)	.57	.50			.0391	.0158(.0158)	.0887	.0390(.0388)	−.16(−.17)	.00(.02)	−.19(−.20)
	.09	.01	(1)	1.00	1.00	.0401	.0052(.0052)	.0900	.0352(.0358)	.0096	.0287(.0286)	−.04(−.03)	.02(.00)	−.62(−.63)
			(2)	.50	.50			.0909	.0309(.0313)	.0094	.0253(.0251)	−.11(−.10)	.06(.04)	−.62(−.62)
			(3)	.26	.22			.0917	.0367(.0364)	.0092	.0302(.0298)	−.14(−.14)	.08(.07)	−.61(−.61)
			(4)	.52	.50			.0908	.0309(.0313)	.0094	.0253(.0251)	−.11(−.10)	.06(.04)	−.62(−.62)
		.04	(1)	1.00	1.00	.0399	.0052(.0052)	.0898	.0366(.0358)	.0425	.0390(.0381)	.05(−.03)	−.01(.00)	−.50(−.47)
			(2)	.50	.50			.0905	.0312(.0313)	.0408	.0340(.0335)	−.01(−.10)	.01(.03)	−.50(−.47)
			(3)	.26	.22			.0912	.0363(.0364)	.0386	.0388(.0400)	−.07(−.14)	.03(.05)	−.50(−.45)
			(4)	.52	.50			.0905	.0312(.0313)	.0408	.0340(.0335)	−.01(−.10)	.01(.03)	−.50(−.47)
		.09	(1)	1.00	1.00	.0401	.0053(.0052)	.0904	.0346(.0358)	.0889	.0556(.0555)	.02(−.03)	−.02(.00)	−.32(−.32)
			(2)	.50	.50			.0896	.0298(.0313)	.0895	.0494(.0489)	−.05(−.10)	.00(.02)	−.34(−.32)
			(3)	.26	.22			.0887	.0345(.0364)	.0901	.0611(.0586)	−.10(−.14)	.01(.03)	−.35(−.31)
			(4)	.52	.50			.0896	.0298(.0313)	.0895	.0494(.0489)	−.04(−.10)	−.00(.02)	−.34(−.32)

all $b_i = 1$ for $i > a'$. Although missing subgroup information (the j value) makes T''_{ab} incomputable, it does not cause problems in computing T''_μ, T''_a, and T''_0. So, $\mathrm{SS}''_{be} = \mathrm{SS}''_e + \mathrm{SS}''_b = (T''_0 - T''_{ab}) + (T''_{ab} - T''_a) = T''_0 - T''_a$ and $\mathrm{SS}''_a = T''_a - T''_\mu$ are computable. Setting the two sums of squares equal to their expectations, we have

$$\begin{aligned} \mathrm{SS}''_{be} &= E(\mathrm{SS}''_{be}) = (N'' - k''_{12})\sigma_\beta^2 + (N'' - a'')\sigma_e^2 \\ \mathrm{SS}''_a &= E(\mathrm{SS}''_a) \\ &= (N'' - k''_1)\sigma_\alpha^2 + (k''_{12} - k''_3)\sigma_\beta^2 + (a'' - 1)\sigma_e^2. \end{aligned} \tag{7}$$

To include the data with missing nesting information in estimation, we have five (or four if $\mathrm{SS}'_b = 0$) equations and only three unknown parameters—$\sigma_e^2, \sigma_\beta^2$, and σ_α^2. Note that all coefficients in these equations are known because we assume that all b_i's and n_{ij}'s in Model (1) are known. To obtain unbiased estimates of the three variance components, we may linearly combine the five (or four) equations in (4) and (7) into three and then solve the combined equations. A natural way to combine these equations is to combine SS''_{be} and SS'_b with a prespecified weight r_b and combine SS''_a and SS'_a with a prespecified weight r_a. Let $\tilde\sigma_e^2, \tilde\sigma_\beta^2$, and $\tilde\sigma_\alpha^2$ be the resulting estimates. Then,

$$\begin{aligned} \mathrm{SS}'_e &= c'_{11}\tilde\sigma_e^2 \\ r_b\mathrm{SS}'_b + (1 - r_b)\mathrm{SS}''_{be} &= d_{21}\tilde\sigma_e^2 + d_{22}\tilde\sigma_\beta^2 \\ r_a\mathrm{SS}'_a + (1 - r_a)\mathrm{SS}''_a &= d_{31}\tilde\sigma_e^2 + d_{32}\tilde\sigma_\beta^2 + d_{33}\tilde\sigma_\alpha^2, \end{aligned} \tag{8}$$

where

$$\begin{aligned} d_{21} &= r_b c'_{21} + (1 - r_b)(c''_{11} + c''_{21}), \\ d_{22} &= r_b c'_{22} + (1 - r_b)c''_{22} \\ d_{3j} &= r_a c'_{3j} + (1 - r_a)c''_{3j}, \qquad j = 1, 2, 3 \end{aligned} \tag{9}$$

and

$$\begin{aligned} c_{11} &= N - b., \qquad c_{21} = b. - a, \qquad c_{22} = N - k_{12} \\ c_{31} &= a - 1, \qquad c_{32} = k_{12} - k_3, \qquad c_{33} = N - k_1. \end{aligned} \tag{10}$$

Solving the equations in (8) yields

$$\begin{aligned} \tilde\sigma_e^2 &= \mathrm{SS}'_e / c'_{11} \\ \tilde\sigma_\beta^2 &= [r_b\mathrm{SS}'_b + (1 - r_b)\mathrm{SS}''_{be} - d_{21}\tilde\sigma_e^2]/d_{22} \\ \tilde\sigma_\alpha^2 &= [r_a\mathrm{SS}'_a + (1 - r_a)\mathrm{SS}''_a - d_{31}\tilde\sigma_e^2 - d_{32}\tilde\sigma_\beta^2]/d_{33}. \end{aligned} \tag{11}$$

Table 5. AAR Data From Sample B of Round 2 (sorted by $n_{i.}$) and Basic Statistics

i	b_i	$n_{i.}$	y_{ijk}									
1	1	1	113									
2	1	1	400									
3	1	1	408									
4	1	1	417									
5	1	1	454									
6	1	1	660									
7	1	2	201	559								
8	1	2	271	436								
9	1	2	318	367								
10	1	2	345	355								
11	1	2	462	563								
12	1	2	471	501								
13	1	2	493	649								
14	1	2	531	550								
15	1	2	558	594								
16	1	3	245	360	414							
17	1	3	275	310	335							
18	1	3	379	410	413							
19	1	3	424	477	596							
20	1	3	467	545	678							
21	1	3	535	628	666							
22	1	4	209	310	508	510						
23	1	4	363	376	401	439						
24	1	4	447	563	612	724						
25	1	4	452	523	533	640						
26	1	5	354	389	402	443	460					
27	1	5	363	426	458	554	789					
28	2	6	319	343	370	398	490	607				
29	2	7	260	314	325	335	384	403	512			
30	2	8	276	295	318	332	336	347	373	390		
31	2	8	327	390	405	428	466	539	576	599		
32	2	9	365	499	506	545	545	553	589	662	759	
33	2	10	284	330	386	458	459	466	475	479	518	661
34	4	20	261	311	312	344	345	345	348	356	384	387
			388	416	417	422	426	438	450	469	501	551

$a' = 27$	$a'' = 7$	
$b'_. = 27$	$b''_. = 16$	
$N' = 68$	$N'' = 68$	
$k'_1 = 3.09$	$k''_1 = 11.6765$	
$k'_{12} = 68$	$k''_{12}(\mathrm{a}) = 29.127$	$k''_{12}(\mathrm{b}) = 31.5317$
$k'_3 = 3.09$	$k''_3(\mathrm{a}) = 4.3823$	$k''_3(\mathrm{b}) = 4.6176$
$k'_4 = 746$	$k''_4(\mathrm{a}) = 1340$	$k''_4(\mathrm{b}) = 1502$
$k'_5 = 210$	$k''_5(\mathrm{a}) = 125$	$k''_5(\mathrm{b}) = 149$
$k'_6 = 746$	$k''_6(\mathrm{a}) = 1336.06$	$k''_6(\mathrm{b}) = 1458.59$
$k'_7 = 210$	$k''_7(\mathrm{a}) = 124.508$	$k''_7(\mathrm{b}) = 142.819$
$k'_8 = 746$	$k''_8(\mathrm{a}) = 3,664$	$k''_8(\mathrm{b}) = 3772$
$k'_9 = 746$	$k''_9 = 11,312$	
$T'_0 = 31,081$	$T''_0 = 28,867$	
$T'_{\mathrm{ab}} = 30,853.69$	$T''_{\mathrm{ab}} = ?$	
$T'_{\mathrm{a}} = 30,853.69$	$T''_{\mathrm{a}} = 28,609.46$	
$T'_{\mu} = 30,412.52$	$T''_{\mu} = 28,446.34$	

NOTE: 1. For large organizations ($n_{.} > 5$), the subscript j for each result y_{ijk} is unknown.
2. The letters a and b in parentheses indicate the grouping methods.
3. The sums of squares are computed using (square root) transformed data.

The weights can be simply set to a constant or derived by some optimal procedures. Four kinds of weights are considered:

1. Set $r_b = r_a = 1$. This gives the estimates using the data with complete nesting information only.
2. Set $r_b = r_a = \frac{1}{2}$. This method gives equal weight to the sums of squares associated with both the complete and the incomplete nesting information.
3. Select r_b and r_a that minimize the variances of the combined sums of squares: $V(r_b \mathrm{SS}'_b + (1 - r_b)\mathrm{SS}''_{be})$ and $V(r_a \mathrm{SS}'_a + (1 - r_a)\mathrm{SS}''_a)$, respectively. This method requires prespecified values of the three variance components to evaluate the two weights (see formulas in App. B).
4. Select r_b and r_a that minimize the variances of the derived estimates: $V(\tilde{\sigma}^2_\beta)$ and $V(\tilde{\sigma}^2_\alpha)$, respectively. This method also requires prespecified values of the three variance components to evaluate the two weights (see formulas in App. B).

When $b_i = 1$ for $i \leq a'$, we have $\mathrm{SS}'_b = 0$, and hence r_b in methods 3 and 4 is undefined. In this case, only SS''_{be} can be used to estimate σ^2_β. The value of r_b can be set to any value except 1 and the estimator for σ^2_β will be the same using any one of the three methods (2–4).

For methods 3 and 4, one may use estimates of the variance components obtained by method 1 or method 2, but such derived estimates may no longer be unbiased because r_b and r_a will not be independent of the sums of squares.

Remark 1. As characteristic of ANOVA estimators, it is possible to get negative estimates for σ^2_β and σ^2_α. To avoid negative estimates, one may follow a commonly used procedure such as setting the variance component with a negative estimate to be 0. Be aware that in this case the estimators are no longer unbiased.

3. COMPARISONS OF ESTIMATORS DERIVED WITH DIFFERENT WEIGHTS

In this section, we compare the estimators derived in Section 2. Note that all estimators are unbiased because they are actually ANOVA-type estimators. A natural way to compare these estimators is to compare their variances and covariances. There is no need to compare the estimators for σ^2_e because they are the same.

Under the assumption that all three components in the two-way nested model are normal, we have (see App. A for details):

$$
\begin{aligned}
\operatorname{var}(\tilde{\sigma}_e^2) &= \operatorname{var}(\mathrm{SS}_e')/(c_{11}')^2 \\
\operatorname{cov}(\tilde{\sigma}_e^2, \tilde{\sigma}_\beta^2) &= -d_{21}\operatorname{var}(\tilde{\sigma}_e^2)/d_{22} \\
\operatorname{var}(\tilde{\sigma}_\beta^2) &= [r_b^2\operatorname{var}(\mathrm{SS}_b') + (1-r_b)^2(\operatorname{var}(\mathrm{SS}_b'') + \operatorname{var}(\mathrm{SS}_e'')) + d_{21}^2\operatorname{var}(\tilde{\sigma}_e^2)]/(d_{22})^2 \\
\operatorname{cov}(\tilde{\sigma}_e^2, \tilde{\sigma}_\alpha^2) &= -[d_{31}\operatorname{var}(\tilde{\sigma}_e^2) + d_{32}\operatorname{cov}(\tilde{\sigma}_e^2, \tilde{\sigma}_\beta^2)]/d_{33} \\
\operatorname{cov}(\tilde{\sigma}_\beta^2, \tilde{\sigma}_\alpha^2) &= [r_b r_a \operatorname{cov}(\mathrm{SS}_b', \mathrm{SS}_a')/d_{22} + (1-r_b)(1-r_a)(\operatorname{cov}(\mathrm{SS}_b'', \mathrm{SS}_a'') \\
&\quad + \operatorname{cov}(\mathrm{SS}_e'', \mathrm{SS}_a''))/d_{22} - d_{31}\operatorname{cov}(\tilde{\sigma}_e^2, \tilde{\sigma}_\beta^2) - d_{32}\operatorname{var}(\tilde{\sigma}_\beta^2)]/d_{33} \\
\operatorname{var}(\tilde{\sigma}_\alpha^2) &= [r_a^2\operatorname{var}(\mathrm{SS}_a') + (1-r_a)^2\operatorname{var}(\mathrm{SS}_a'') - d_{32}(r_b r_a \operatorname{cov}(\mathrm{SS}_b', \mathrm{SS}_a') \\
&\quad + (1-r_b)(1-r_a)\operatorname{cov}(\mathrm{SS}_b'', \mathrm{SS}_a''))/d_{22}]/(d_{33})^2 \\
&\quad - [d_{31}\operatorname{cov}(\tilde{\sigma}_e^2, \tilde{\sigma}_\alpha^2) - d_{32}\operatorname{cov}(\tilde{\sigma}_\beta^2, \tilde{\sigma}_\alpha^2)]/d_{33}. \qquad (12)
\end{aligned}
$$

These variances and covariances can be expressed as functions of the three unknown parameters using the formulas provided in Appendix A. To estimate these variances and covariances, we simply replace the three unknown parameters by their estimators. A confidence interval for σ_e^2 can be constructed because we know that $(N' - a')\tilde{\sigma}_e^2/\sigma_e^2$ has a chi-squared distribution. Interval estimates for σ_α^2 and σ_β^2 are difficult to derive, however, even though there is no missing nesting information in data (see Burdick and Graybill 1988, 1992). Approximate interval estimates can be constructed if sums of squares used in point estimates are jointly independent and are scaled chi-squares. Unfortunately, this method cannot be applied to the data with missing nesting information. For data with missing nesting information, the two sums of squares SS_e'' and SS_b'' are not separable and the combined sum of squares SS_{be}'' $(=\mathrm{SS}_e''+ \mathrm{SS}_b'')$ may not be scaled chi-squared because the sum of two scaled chi-squared variables may not be scaled chi-squared.

Simply examining how the weights are selected for each estimator, we might expect that method 2 is better than method 1 because more data are used in method 2. Method 3 may be better than method 2 and method 4 may be the best among the four methods because methods 3 and 4 are intended or designed to minimize the variances of estimates. Because the formulas for the variances of estimates are not simple, it is difficult to compare them analytically. Moreover, unbalanced data are difficult to compare. So comparisons are made for the balanced design only. In particular, $n = 5, b = 2$, and $a = 20$ are used in the comparison with $a' = 5, 10, 15$ corresponding to 75%, 50%, and 25% of data with missing nesting information.

The values of the variance components also need to be specified in the comparison. Because data scale is not important in the comparison, we may fix one of the three variance components. Because we are interested in comparing the estimates for σ_β^2 and σ_α^2, we set $\sigma_e^2 = .2^2$ and let the two variance components take the values $.1^2, .2^2, .3^2$ to form nine different combinations. For each design specified in Tables 2–4 (pp. 73–75) and each of the nine combinations, 1,000 datasets were generated. Results for each of the four methods discussed in Section 2 are shown in the rows with the method number specified in the fourth column of each table. The sample mean and sample standard deviation of the 1,000 estimates for each variance component are listed in columns headed with $m[\ \ ]$ and $s[\ \ ]$, respectively. The sample correlations are given in the last three columns. Corresponding values shown in parentheses are the expected values, computed according to the formulas in this section and those in Appendix A with the three unknown variance components replaced by their true values.

Examining the results, we see that method 4 does provide the most stable estimates in terms of smallest variances, but method 3 does not perform as well as we expected. In most cases, it is worse than the method 2 except when σ_β^2 is less than σ_e^2 and 50% or more of data has missing nesting information. Method 2 is better than method 1 in estimating σ_α^2 and in estimating σ_β^2, except for cases in which σ_β^2 is less than σ_e^2. So, properly including data with missing nesting information in estimation does improve the estimates for both σ_β^2 and σ_α^2, and the improvement is greater when more data have missing nesting information.

To validate our comparisons using the formulas for the variance, we evaluated the accuracies of the variance and covariance formulas using a simulation study. For each of the nine combinations, 1,000 datasets were generated and 1,000 estimates of each of the four kinds of estimates were obtained. Then the sample mean, sample standard deviation, and sample covariances of these estimates were computed. These values are quite close to the values computed by the variance and covariance formulas. The differences of the variances increase somewhat as the true values of the variances increase. Moreover, in most instances, the mean of the method 4 estimator is closer to the true value than the mean of the method 1 estimator.

A fact of interest is that the estimators are highly correlated. With incomplete nesting information, the estimators for σ_e^2 and σ_α^2 are no longer mutually independent. They are positively correlated, and each of them is negatively correlated with the estimator for σ_β^2. The correlations increase with the proportion of data with missing nesting information. The correlations between the estimators for σ_e^2 and σ_β^2 and those for σ_β^2 and σ_α^2 are usually larger than the correlations between the estimators for σ_e^2 and σ_α^2. For fixed percent missing, correlations of highest absolute value occur at low σ_α^2 or low σ_β^2 or combinations of these.

4. APPLICATION TO THE AAR PROGRAM

The AAR program mentioned in Section 1 is operated on a quarterly schedule. Four batches of samples, at different concentrations, are generated each quarter. Reported results from the counters for each sample within a round can be described by the two-way nested model given in Section 1. The first component, α, is the random effect of organization, the second component, β, is the random effect of sample or filter, and the error is due to the counter-to-counter or wedge-to-wedge variation. As an example, we present the data from sample B of round 2 in Table 5 (p. 76). Because the filters were not identified in the process, the subgroup information (j) is missing in all organizations except for small (with five or fewer counters) ones because we know that one set of samples is enough for a small organization. So, the number of small organizations is the number with no missing nesting information. In other words, the number a' is the number of i's with $b_i = 1$. Note that this is the case mentioned in Section 2, where method 1 (discarding the data) fails to separate the variance components σ_β^2 and σ_α^2. Fortunately, not all $b_i = 1$, so we are able to separate the two variance components by the proposed methods (2–4).

To use the data with missing nesting information, we need to know all b_i's, the numbers of sample sets distributed to each organization, and all n_{ij}'s, the numbers of counters counting wedges from the same sample set. What we actually know with the AAR data, however, is $n_{i.}$, the number of counters in each organization. According to the rule that up to five counters can share a sample set, the number of sample sets distributed to each organization can be determined as the smallest integer that is greater than $n_{i.}/5$. As for the number of counters counting the same sample set, there is no unique solution if we do not know exactly how an organization allocates samples. For $n_{i.} = 11$, there are four possible ways to divide the 11 counters into three groups—(1) (4, 4, 3), (2) (5, 3, 3), (3) (5, 4, 2), (4) (5, 5, 1). The first and the last are the two extreme cases: (a) Counters are divided into subgroups with counters in each subgroup as equally as possible, and (b) all samples are fully used except the last one. When r_b and r_a are about the same for all cases, then the estimate for $\sigma_b^2(\sigma_a^2)$ by grouping (a) will be greater (less) than the estimate by grouping (b), and estimates by any other grouping will fall between the those two. This is because grouping only affects the values of k_{12}'' and k_3'' and does not affect other coefficients $(a'', b_.'', N'', k_1'',$ and in turn, $c_{21}, c_{31}, c_{33}, d_{21}, d_{31}, c_{33})$, which are used in the estimates. The k_{12}'' and k_3'' obtained by grouping (a) are less than those obtained by grouping (b), and in turn, c_{22} and $d_{22}(c_{32}$ and $d_{32})$ by (a) are greater (less) than those by (b).

In Table 6, we show our results for samples having fiber density closest to 500 fibers/mm^2 in the first 10 rounds. Round number (Rd), sample run (Sp), and sample mean $(\hat{\mu})$ are listed in the first three columns. Each variance component is expressed as a relative standard deviation—the standard deviation divided by the mean. Because the estimation method 4 is used, prespecified values for the three variance components are needed. The estimates from the previous round were used for each round except for the first round, in which the estimates obtained by method 2 were used. To avoid unreasonably low estimates, results below a prespecified value are replaced by the specified value. The prespecified value for $\tilde{\sigma}_e$ and $\tilde{\sigma}_\beta$ is $.1\hat{\mu}$ and the value for $\tilde{\sigma}_\alpha$

Table 6. Estimates of Variance Components in the AAR Data Analysis

Rd	Sp	$\hat{\mu}$	a'	a''	b'	b''	N'	N''		r_b	r_a	$\tilde{\sigma}_e/\hat{\mu}$ $(s[\tilde{\sigma}_e]/\hat{\mu})$	$\tilde{\sigma}_\beta/\hat{\mu}$ $(s[\tilde{\sigma}_\beta]/\hat{\mu})$	$\tilde{\sigma}_\alpha/\hat{\mu}$ $(s[\tilde{\sigma}_\alpha]/\hat{\mu})$	$\hat{\rho}(\tilde{\sigma}_e^2, \tilde{\sigma}_\beta^2)$	$\hat{\rho}(\tilde{\sigma}_e^2, \tilde{\sigma}_\alpha^2)$	$\hat{\rho}(\tilde{\sigma}_\beta^2, \tilde{\sigma}_\alpha^2)$
1	B	510.3	12	3	12	6	32	25	(a)	0	.35	.3114(.0153)	.1000(.0387)	.2335(.0345)	−.700	.472	−.756
									(b)	0	.33		.1000(.0432)	.2319(.0382)	−.703	.509	−.795
2	B	432.7	27	7	27	16	68	68	(a)	0	.40	.2264(.0106)	.1000(.0222)	.1537(.0188)	−.749	.474	−.711
									(b)	0	.39		.1000(.0236)	.1522(.0199)	−.751	.500	−.738
3	D	535.0	37	9	37	20	89	89	(a)	0	.29	.1933(.0052)	.1124(.0116)	.1887(.0089)	−.732	.477	−.724
									(b)	0	.28		.1137(.0118)	.1877(.0091)	−.733	.486	−.733
4	D	409.5	51	20	51	54	137	223	(a)	0	.59	.1768(.0028)	.3031(.0053)	.0500(.0059)	−.753	.346	−.520
									(b)	0	.59		.3147(.0056)	.0500(.0061)	−.758	.383	−.562
5	B	477.8	61	22	61	66	162	284	(a)	0	.03	.2423(.0023)	.1733(.0117)	.1820(.0069)	−.269	.130	−.575
									(b)	0	.04		.1771(.0122)	.1780(.0075)	−.271	.141	−.592
6	C	410.6	68	26	68	73	191	308	(a)	0	.42	.2036(.0037)	.2030(.0075)	.1398(.0065)	−.692	.344	−.568
									(b)	0	.41		.2084(.0078)	.1315(.0068)	−.697	.371	−.597
7	B	558.6	69	30	69	85	192	356	(a)	0	.32	.1950(.0026)	.2341(.0063)	.1245(.0051)	−.571	.268	−.552
									(b)	0	.32		.2407(.0066)	.1111(.0054)	−.576	.294	−.582
8	B	399.5	66	28	66	78	167	333	(a)	0	.23	.1883(.0025)	.2906(.0078)	.1177(.0054)	−.459	.221	−.565
									(b)	0	.24		.2981(.0080)	.0972(.0057)	−.467	.244	−.590
9	C	591.5	81	39	81	118	223	496	(a)	0	.16	.1911(.0021)	.1574(.0081)	.1576(.0055)	−.356	.156	−.535
									(b)	0	.17		.1623(.0085)	.1518(.0059)	−.363	.178	−.563
10	C	533.4	91	42	91	124	243	523	(a)	0	.46	.1822(.0021)	.1725(.0042)	.0979(.0038)	−.681	.301	−.513
									(b)	0	.46		.1772(.0043)	.0890(.0039)	−.687	.331	−.545

is $.05\hat{\mu}$. Because σ_β is estimated after σ_e and σ_α is estimated after σ_β, the replacement of estimates will affect successive estimates. If the replacement occurs to the estimate for σ_e, then it will affect the estimates for σ_β and σ_α. If it occurs to the estimate for σ_β, then it will affect the estimates for σ_α only. As we discussed in Section 3, these estimates are highly correlated. Underestimating $\sigma_e(\sigma_\beta)$ will result in overestimating σ_β (σ_α, respectively). So, the replacement should provide us with better estimates. Sampling errors of the estimates of the three components are devided by sample mean and shown in parentheses.

From the results shown in Table 6, the numbers of big (with six or more counters) and small organizations (a'' and a') and the total numbers of counters from both organizations (N'' and N') increased round by round. The number of counters from big organizations increased faster, however, and was more than two times the number from small organizations after round 8. Thus, there has been progressively more and more missing nesting information.

Because $\mathrm{SS}'_b = 0, r_b$ is set to 0, the two r_a's corresponding to grouping (a) and grouping (b) are almost equal. So, the estimates for $\sigma_\beta/\mu(\sigma_\alpha/\mu)$ by grouping (a) are greater (less) than the estimates by grouping (b). The differences are not great, however. Correlations among the three estimates are high. Sampling errors for $\tilde{\sigma}_\beta/\hat{\mu}$ and $\tilde{\sigma}_\alpha/\hat{\mu}$ are about the same, and they are two to three times the sampling error for $\tilde{\sigma}_e/\hat{\mu}$.

Note that the results are based on the assumption that counter variation is homogeneous; in particular, variability among counters in small organizations is assumed to be the same as that in large organizations. Because all the nesting information (in large organizations) is missing, there is no way to check this assumption. If the counter variance in small organizations is less than that in large organizations, then the sample variance, σ^2_β, would be overestimated (positively biased) because in this case the sample variance will include part of the extra variation resulting from counters in large organizations.

Plots of the AAR data indicate normality on the square-root scale. Because normality is needed to obtain sampling variance and covariance of estimators, a square-root transformation was used. Estimates of the variance components were transformed back to the original scale, but the correlations were not transformed back to the original scale because the main interest is in showing the effect of the missing nesting information. Actually, changing the scale does not greatly affect the correlation.

5. SUMMARY

In this article, a missing nesting information problem in a two-way nested model is considered. Due to the missing information, currently available estimation methods are difficult to be applied to this case. Some methods require the variance–covariance matrix to be specified. Unfortunately, that matrix cannot be specified without the nesting information. The ANOVA method, however, can be used as long as the sums of squares can be computed. With incomplete nesting information, some sums of squares are still computable. The computable sums of squares do provide information to improve the estimates.

Although the methods are developed for a two-way nested random-effects model, they can be applied to higher-order nested models and to fixed or mixed-effects models. Given the difficulty and importance of the missing-information problem, we would like to see more research on this issue.

ACKNOWLEDGMENTS

We thank the editor, the associate editor, and referees for their helpful comments and suggestions which led to a significant improvement of the original manuscript. Mention of a company or product does not necessarily constitute endorsement by the National Institute for Occupational Safety and Health, Centers for Disease Control and Prevention.

APPENDIX A: SAMPLING VARIANCES AND COVARIANCES

Under the normality assumption, the variances and covariances for the uncorrected and corrected sums of squares can be obtained from Searle (1961). With the notations defined in Section 2, these variances and covariances are given by

$$\begin{aligned}
\mathrm{var}(T_a) &= 2(Nk_1\sigma^4_\alpha + k_7\sigma^4_\beta + a\sigma^4_e) \\
&\quad + 4(Nk_3\sigma^2_\alpha\sigma^2_\beta + N\sigma^2_\alpha\sigma^2_e + k_{12}\sigma^2_\beta\sigma^2_e) \\
\mathrm{var}(T_{ab}) &= \mathrm{var}(T_a) + 2[(Nk_3 - k_7)\sigma^4_\beta \\
&\quad + (b. - a)\sigma^4_e] + 4(N - k_{12})\sigma^2_\beta\sigma^2_e \\
\mathrm{var}(T_\mu) &= 2(k_1\sigma^2_\alpha + k_3\sigma^2_\beta + \sigma^2_e) \\
\mathrm{cov}(T_a, T_{ab}) &= \mathrm{var}(T_a) + 2(k_5 - k_7)\sigma^4_\beta \\
\mathrm{cov}(T_a, T_\mu) &= 2(k_9\sigma^4_\alpha + k_6\sigma^4_\beta + N\sigma^4_e)/N \\
&\quad + 4(k_8\sigma^2_\alpha\sigma^2_\beta/N + k_1\sigma^2_\alpha\sigma^2_e + k_3\sigma^2_\beta\sigma^2_e) \\
\mathrm{cov}(T_{ab}, T_\mu) &= \mathrm{cov}(T_a, T_\mu) + 2\sigma^4_\beta(k_4 - k_6)/N
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{var}(\mathrm{SS}_e) &= 2c_{11}\sigma^4_e \\
\mathrm{var}(\mathrm{SS}_b) &= \mathrm{var}(T_a) - 2\ \mathrm{cov}(T_a, T_{ab}) + \mathrm{var}(T_{ab}) \\
\mathrm{var}(\mathrm{SS}_a) &= \mathrm{var}(T_a) - 2\ \mathrm{cov}(T_a, T_\mu) + \mathrm{var}(T_\mu) \\
\mathrm{cov}(\mathrm{SS}_e, \mathrm{SS}_b) &= 0 \\
\mathrm{cov}(\mathrm{SS}_e, \mathrm{SS}_a) &= 0 \\
\mathrm{cov}(\mathrm{SS}_b, \mathrm{SS}_a) &= \mathrm{cov}(T_a, T_{ab}) - \mathrm{cov}(T_{ab}, T_\mu) \\
&\quad - \mathrm{var}(T_a) + \mathrm{cov}(T_a, T_\mu).
\end{aligned}$$

Replacing the no-prime notations with the single-prime and double-prime notations produces the formulas for both the data with complete nesting information and the data with missing nesting information cases. Although the sum of squares T''_{ab} cannot be computed, its variance and covariances are computable because we assume that all the numbers b_i and n_{ij} in Model (1) are known.

Note that the data with complete nesting information and the data with missing nesting information are from different main groups. So SS′'s are independent of SS″'s. With the formulas provided here, it is not too difficult to derive the variances and covariances of the derived estimators.

APPENDIX B: DETERMINATION OF WEIGHTS r_b AND r_a

Let X and Y be two random variables with nonzero variances and let p and q be two positive numbers. Define a variance function

$$f(r) = \mathrm{var}\left(\frac{rX + (1-r)Y}{rp + (1-r)q}\right), \quad \text{for} \quad 0 \le r \le 1.$$

Then, there exists r_0 such that $f(r_0) = \min_{0\le r\le 1} f(r)$ and

$$r_0 = \frac{p\ \mathrm{var}(Y) - q\ \mathrm{cov}(X,Y)}{p\ \mathrm{var}(Y) - (p+q)\ \mathrm{cov}(X,Y) + q\ \mathrm{var}(X)}.$$

The proof is simple: Set $f'(r) = 0$ and solve this equation for r.

In particular, if $p = q = 1$, and X, Y are independent, then $r_0 = \mathrm{var}(Y)/[\mathrm{var}(Y) + \mathrm{var(X)}]$ will minimize the variance function $f(r) = \mathrm{var}(rX + (1-r)Y)$.

In method 3, $p = q = 1, X = \mathrm{SS}'_b, Y = \mathrm{SS}''_{be}$ or $X = \mathrm{SS}'_a, Y = \mathrm{SS}''_a$, and X, Y are independent. So,

$$\begin{aligned} r_b &= \mathrm{var}(\mathrm{SS}''_{be})/(\mathrm{var}(\mathrm{SS}'_b) + \mathrm{var}(\mathrm{SS}''_{be})) \\ r_a &= \mathrm{var}(\mathrm{SS}''_a)/(\mathrm{var}(\mathrm{SS}'_a) + \mathrm{var}(\mathrm{SS}''_a)). \end{aligned}$$

In method 4, let

$$\begin{aligned} X_b &= \mathrm{SS}'_b - c'_{21}\tilde\sigma^2_e, \quad Y_b = \mathrm{SS}''_{be} - (c''_{11} + c''_{21})\tilde\sigma^2_e \\ X_a &= \mathrm{SS}'_a - c'_{31}\tilde\sigma^2_e - c'_{32}\tilde\sigma^2_\beta, \quad Y_a = \mathrm{SS}''_a - c''_{31}\tilde\sigma^2_e - c''_{32}\tilde\sigma^2_\beta. \end{aligned}$$

Then,

$$\begin{aligned} \tilde\sigma^2_\beta &= (r_bX_b + (1-r_b)Y_b)/(r_bc'_{22} + (1-r_b)c''_{22}) \\ \tilde\sigma^2_\alpha &= (r_aX_a + (1-r_a)Y_a)/(r_ac'_{33} + (1-r_a)c''_{33}) \end{aligned}$$

and hence

$$\begin{aligned} r_b = {} & [-c''_{22}\mathrm{cov}(X_b, Y_b) + c'_{22}\mathrm{var}(Y_b)]/ \\ & [c''_{22}\mathrm{var}(X_b) - (c''_{22} + c'_{22})\mathrm{cov}(X_b, Y_b) + c'_{22}\mathrm{var}(Y_b)] \end{aligned}$$

$$\begin{aligned} r_a = {} & [-c''_{33}\mathrm{cov}(X_a, Y_a) + c'_{33}\mathrm{var}(Y_a)]/ \\ & [c''_{33}\mathrm{var}(X_a) - (c''_{33} + c'_{33})\mathrm{cov}(X_a, Y_a) + c'_{33}\mathrm{var}(Y_a)], \end{aligned}$$

where

$$\begin{aligned} \mathrm{var}(X_b) &= \mathrm{var}(\mathrm{SS}'_b) + (c'_{21})^2\mathrm{var}(\tilde\sigma^2_e) \\ \mathrm{var}(Y_b) &= \mathrm{var}(\mathrm{SS}''_b) + \mathrm{var}(\mathrm{SS}''_e) \\ &\quad + (c''_{11} + c''_{21})^2\mathrm{var}(\tilde\sigma^2_e) \\ \mathrm{cov}(X_b, Y_b) &= c'_{21}(c''_{11} + c''_{21})\mathrm{var}(\tilde\sigma^2_e) \end{aligned}$$

and

$$\begin{aligned} \mathrm{var}(X_a) &= \mathrm{var}(\mathrm{SS}'_a) + (c'_{31})^2\mathrm{var}(\tilde\sigma^2_e) + (c'_{32})^2\mathrm{var}(\tilde\sigma^2_\beta) \\ &\quad - 2c'_{32}\mathrm{cov}(\mathrm{SS}'_a, \tilde\sigma^2_\beta) + 2c'_{31}c'_{32}\mathrm{cov}(\tilde\sigma^2_e, \tilde\sigma^2_\beta) \\ \mathrm{var}(Y_a) &= \mathrm{var}(\mathrm{SS}''_a) + (c''_{31})^2\mathrm{var}(\tilde\sigma^2_e) + (c''_{32})^2\mathrm{var}(\tilde\sigma^2_\beta) \\ &\quad - 2c''_{32}\mathrm{cov}(\mathrm{SS}''_a, \tilde\sigma^2_\beta) + 2c''_{31}c''_{32}\mathrm{cov}(\tilde\sigma^2_e, \tilde\sigma^2_\beta) \\ \mathrm{cov}(X_a, Y_a) &= -c''_{32}\mathrm{cov}(\mathrm{SS}'_a, \tilde\sigma^2_\beta) + (c'_{31}c''_{32} + c''_{31}c'_{32}) \\ &\quad \mathrm{cov}(\tilde\sigma^2_e, \tilde\sigma^2_\beta) - c'_{32}\mathrm{cov}(\mathrm{SS}''_a, \tilde\sigma^2_\beta) \\ &\quad + c'_{31}c''_{31}\mathrm{var}(\tilde\sigma^2_e) + c'_{32}c''_{32}\mathrm{var}(\tilde\sigma^2_\beta). \end{aligned}$$

[Received September 1993. Revised April 1996.]

REFERENCES

Burdick, R. K., and Graybill, F. A. (1988), "The Present Status of Confidence Interval Estimation on Variance Components in Balanced and Unbalanced Random Models," *Communications in Statistics—Theory and Methods*, 17, 1165–1195.

——— (1992), *Confidence Interval on Variance Components*, New York: Marcel Dekker.

Little, R. J. A. (1992), "Regression With Missing X's: A Review," *Journal of the American Statistical Association*, 87, 1227–1237.

Little, R. J. A., and Rubin, D. B. (1987), *Statistical Analysis With Missing Data*, New York: Wiley.

Searle, S. R. (1961), "Variance Components in the Unbalanced 2-Way Nested Classification," *The Annals of Mathematical Statistics*, 32, 1161–1166.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992), *Variance Components*, New York: Wiley.