ACT Research Report Series 95-2

A Note on a Relationship Between Covariance Matrices and Consistently Estimated Variance Components

David J. Woodruff

ACT

August 1995

For additional copies write:
ACT Research Report Series
P.O. Box 168
Iowa City, Iowa 52243

A Note on a Relationship Between Covariance Matrices and Consistently Estimated Variance Components

David J. Woodruff

Abstract

The one observation per cell two-way items by examinees random effects ANOVA with all error components zero is considered. The estimated variance components are expressed as functions of the inter-item covariance matrix and the inter-examinee covariance matrix. These expressions show that under the random effects model if the inter-item and inter-examinee covariance matrices are unconstrained then both the number of items and the number of examinees must approach infinity for the estimates of any of the variance components to be consistent. However, if these two covariance matrices are constrained so that each has homogeneous variances and covariances, then consistent estimates of the variance components can be obtained without both the number of items and the number of examinees simultaneously approaching infinity.

Keywords: compound symmetry, consistency, variance components

A Note on a Relationship Between Covariance Matrices and Consistently Estimated Variance Components

In applications of the random effects ANOVA model to measurement data, usually little if any consideration is given to the covariance structure of the data—the focus is on the variance components. The purpose of this note is to explore the relationship between the covariance structure of the data and the variance components and to demonstrate that when the covariance structure of the data has Scheffés (1959, p. 264) compound symmetric form, then consistent estimates of the variance components can be obtained more simply than when the covariance structure of the data is unconstrained. This simplification under compound symmetry has an important advantage for the practical application of the model. Hocking, Green, and Bremer (1989) developed the random effects ANOVA model with the assumption of compound symmetry for the covariance structure, and they presented diagnostics for deciding on the appropriateness of the model under this covariance structure, but they did not motivate the usefulness of the compound symmetry assumption as this note does.

Consider the I by J data matrix $\mathbf{Y} = \left\{y_{ij}\right\}$ that results from administering I items to J examinees. Two covariance matrices can be computed from **Y**: the J by J inter-examinee covariance matrix $S_\theta = \left\{s_i(j,j')\right\}_{j,j'=1}^{J}$ (computed using I–1 as the divisor) and the I by I inter-item covariance matrix $\mathbf{S}_\pi = \left\{s_j(i,i')\right\}_{i,i'=1}^{I}$ (computed using J–1 as the divisor). The i or j subscript associated with the elements of a sample covariance matrix denote the index over which the elements are computed, for example,

$$s_j(i,i') = \sum_{j=1}^{J} \left(y_{ij} - \bar{y}_{i.}\right)\left(y_{i'j} - \bar{y}_{i'.}\right) / (J-1).$$

The data matrix **Y** will be modeled using a one observation per cell two-way random effects ANOVA model. A complete and detailed development of this model may be found in Scheffé (1959, Chap. 7) where the model equation is given, the variance component parameters are defined, and method of moments variance component estimators are developed as functions of observed mean-squares. See also Searle (1971, Chap. 9) or Hocking (1985, Chap. 10). Note, however, that for simplicity and without loss of generality all error component values are assumed to be zero so that the error variance equals zero. Furthermore, no distributional assumptions are made about the observations or the model components.

Using results given by Hocking et al. (1989), the method of moments variance component estimators given by Scheffé (1959 Chap. 7) also can be expressed as functions of the two sample covariance matrices:

$$\hat{\sigma}_A^2 = \frac{\sum_{j=1}^{J} \sum_{j'>j}^{J} s_i(j,j')}{J(J-1)/2} = \bar{s}_i(j,j'), \tag{1}$$

$$\hat{\sigma}_B^2 = \frac{\sum_{i=1}^{I} \sum_{i'>i}^{I} s_j(i,i')}{I(I-1)/2} = \bar{s}_j(i,i'), \tag{2}$$

$$\hat{\sigma}^2_{AB} = \frac{\sum_{j=1}^{J} s_i^2(j)}{J} - \frac{\sum_{j=1}^{J}\sum_{j'>j}^{J} s_i(j,j')}{J(J-1)/2} = \bar{s}_i^2(j) - \bar{s}_i(j,j'), \quad \text{or} \tag{3}$$

$$\hat{\sigma}^2_{AB} = \frac{\sum_{i=1}^{I} s_j^2(i)}{I} - \frac{\sum_{i=1}^{I}\sum_{i'>i}^{I} s_i(i,i')}{I(I-1)/2} = \bar{s}_j^2(i) - \bar{s}_j^2(i,i'). \tag{4}$$

Note that the interaction variance component estimator can be expressed as a function of either $\mathbf{S}_\pi$ or $\mathbf{S}_\theta$.

A parameterization of the random effects ANOVA model in terms of covariance matrices gives an interesting perspective on this model. Two underlying countably infinite dimensional parametric covariance matrices, $\boldsymbol{\Sigma}_\theta = \{\sigma_\theta(j,j')\}_{j,j'=1}^{\infty}$ and $\boldsymbol{\Sigma}_\pi = \{\sigma_\pi(i,i')\}_{i,i'=1}^{\infty}$, can be postulated. Conditional on any random sample of I items, let $\boldsymbol{\Sigma}_{\pi I}$ denote an I by I inter-item parametric covariance matrix whose elements are a subset of the elements of $\boldsymbol{\Sigma}_\pi$. Conditional on any random sample of J examinees, let $\boldsymbol{\Sigma}_{\theta J}$ denote a J by J inter-examinee parametric covariance matrix whose elements are a subset of the elements of $\boldsymbol{\Sigma}_\theta$. Assume that $\boldsymbol{\Sigma}_\pi$ and $\boldsymbol{\Sigma}_\theta$ satisfy the compound symmetry structure of Scheffé (1959, p. 264) so that

$$\boldsymbol{\Sigma}_\pi = \left(\sigma_\pi^2 - \gamma_\pi\right)\boldsymbol{I} + \gamma_\pi \boldsymbol{J} \quad \text{and} \tag{5}$$

$$\boldsymbol{\Sigma}_\theta = \left(\sigma_\theta^2 - \gamma_\theta\right)\boldsymbol{I} + \gamma_\theta \boldsymbol{J}, \tag{6}$$

where $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{J}$ is a matrix of all ones and the dimensions of $\boldsymbol{I}$ and $\boldsymbol{J}$ are countably infinite. It then follows that $\boldsymbol{\Sigma}_{\pi I}$ is completely determined by the two parameters σ_π^2 and γ_π which are the same for any selection of I items; hence, $\Sigma_{\pi I}$ is invariant under any selection of I items. Similarly, $\Sigma_{\theta J}$ is completely determined by the two parameters σ_θ^2 and γ_θ which are the same for any selection of J examinees; hence $\Sigma_{\theta J}$ is invariant under any selection of J examinees. This invariance under compound symmetry has an important effect on the consistency of the variance component estimators.

The situation is similar but not identical to joint estimation of IRT models where distinct parameters are associated with each item and each examinee, and both the number of items and the number of examinees must simultaneously approach infinity to obtain consistent estimates of the model parameters (Haberman, 1977). In the unconstrained random effects ANOVA model, a set of parameters (a variance and covariances) is associated with each item and each examinee, and both the number of items and the number of examinees must simultaneously approach infinity to obtain consistent estimates of the variance component parameters which are functions of the item and examinee variances and covariances. However, if the inter-item and inter-examinee covariance matrices are constrained to have the compound symmetric form, then the number of variances and covariances is fixed at four, and consistent estimates of the variance component parameters can be obtained without both I and J simultaneously approaching infinity. This will now be demonstrated.

As already noted, Scheffé (1959, Chap. 7) defines the variance component parameters and develops method of moments estimators for these parameters. The variance component estimators given in equations (1) through (4) are algebraically equivalent to Scheffé's method of moments estimators and so are consistent estimators because they are method of moment estimators (Serfling, 1980). Because Σ_π and Σ_θ are of countably infinite dimension, when no constraints are placed on the structures of Σ_π and Σ_θ there are a countably infinite number of moments (variances and covariances) to be estimated, and so both I and J must go to infinity to obtain consistency. More specifically,

$$\sigma_A^2 = \lim_{I,J\to\infty} \hat{\sigma}_A^2 = \lim_{J\to\infty} \frac{\sum_{j=1}^{J} \sum_{j'>j}^{J} \lim_{I\to\infty} s_i(j,j')}{J(J-1)/2} = \lim_{J\to\infty} \frac{\sum_{j=1}^{J} \sum_{j'>j}^{J} \sigma_\theta(j,j')}{J(J-1)/2}, \tag{7}$$

$$\sigma_B^2 = \lim_{I,J\to\infty} \hat{\sigma}_B^2 = \lim_{I\to\infty} \frac{\sum_{i=1}^{I} \sum_{i'>i}^{I} \lim_{J\to\infty} s_j(i,i')}{I(I-1)/2} = \lim_{I\to\infty} \frac{\sum_{i=1}^{I} \sum_{i'>i}^{I} \sigma_\pi(i,i')}{I(I-1)/2}, \text{ and} \tag{8}$$

$$\sigma_{AB}^2 = \lim_{I,J\to\infty} \hat{\sigma}_{AB}^2 = \lim_{I,J\to\infty} \left(\bar{s}_i^2(j) - \bar{s}_i(j,j')\right) = \lim_{I,J\to\infty} \left(\bar{s}_j^2(i) - \bar{s}_j(i,i')\right)$$

$$= \lim_{J\to\infty} \left(\frac{\sum_{j=1}^{J} \sigma_\theta^2(j)}{J} - \frac{\sum_{j=1}^{J} \sum_{j'>j}^{J} \sigma_\theta(j,j')}{J(J-1)/2} \right)$$

$$= \lim_{I\to\infty} \left(\frac{\sum_{i=1}^{I} \sigma_\pi^2(i)}{I} - \frac{\sum_{i=1}^{I} \sum_{i'>i}^{I} \sigma_\pi(i,i')}{I(I-1)/2} \right). \tag{9}$$

The purpose of this note is to show that with the compound symmetry structure imposed on Σ_π and Σ_θ the consistency of the variance component estimators does not depend on both I and J simultaneously approaching infinity. Because of the invariance of $\Sigma_{\theta J}$ for any selection of J examinees under compound symmetry for Σ_θ, it follows that for any random sample of J examinees:

$$\sigma_A^2 = \lim_{I\to\infty} \hat{\sigma}_A^2 = \lim_{I\to\infty} \overline{s}_i(j,j') = \gamma_\theta \quad \text{and} \tag{10}$$

$$\sigma_{AB}^2 = \lim_{I\to\infty} \hat{\sigma}_{AB}^2 = \lim_{I\to\infty}\left(\overline{s}_i^2(j) - \overline{s}_i(j,j')\right) = \sigma_\theta^2 - \gamma_\theta, \tag{11}$$

and so only I need increase without bound to obtain consistent estimates of these two variance components. Because of the invariance of $\Sigma_{\pi I}$ for any selection I items under compound symmetry for Σ_π, it follows that for any random sample of I items:

$$\sigma_B^2 = \lim_{J\to\infty} \hat{\sigma}_B^2 = \lim_{J\to\infty} \overline{s}_j(i,i') = \gamma_\pi, \quad \text{and} \tag{12}$$

$$\sigma_{AB}^2 = \lim_{J\to\infty} \hat{\sigma}_{AB}^2 = \lim_{J\to\infty}\left(\overline{s}_j^2(i) - \overline{s}_j(i,i')\right) = \sigma_\pi^2 - \gamma_\pi, \tag{13}$$

and so only J need increase without bound to obtain consistent estimates of these two variance components. Note that the interaction variance component can be consistently estimated as either I or J approach infinity under compound symmetry.

In the practical application of the two-way random effects ANOVA model (including the items by examinees situation), it is sometimes the case that it is expensive or impractical to increase the number of levels measured on the first factor but relatively inexpensive to increase the number of levels measured on the second factor, and the purpose of the study is to get an accurate estimate of the variance due to the second factor or to the interaction. If the compound symmetry assumption is not satisfied, then increasing the number of levels measured on the second factor is insufficient to ensure a more accurate estimate of the second factor variance component or the interaction variance component. The number of levels of the first factor would also have to be increased, at greater expense, to ensure more accurate estimates of the second factor variance component or the interaction variance component.

Hocking (1985, Chap. 10) can be consulted for the compound symmetric covariance structure formulation of the two-way random effects ANOVA model when the error variance is non-zero and there is more than one observation per cell. Woodruff (1993) discusses reliability estimation and other psychometric issues under this model.

References

Haberman, S. J. (1977). Maximum likelihood estimates in exponential response models. *The Annals of Statistics, 5*, 815-841.

Hocking, R. R. (1985). *The analysis of linear models.* Monterey, CA: Brooks Cole.

Hocking, R. R., Green, J. W., & Bremer, R. H. (1989). Variance-component estimation with model-based diagnostics. *Technometrics, 31*, 227-239.

Scheffé, H. (1959). *The analysis of variance.* New York, NY: Wiley.

Searle, S. R. (1971). *Linear models.* New York, NY: Wiley.

Serfling, R. J. (1980). *Approximation theorems of mathematical statistics.* New York, NY: Wiley.

Woodruff, D. J. (1993). *Linear models for item scores: Reliability, covariance structure, and psychometric inference.* ACT Research Report No. 93-4. Iowa City, IA: American College Testing.