Cell Means Analysis of ANCOVA Designs

B. Nebiyou Bekele

The University of Texas M. D. Anderson Cancer Center

Roger E. Kirk

Baylor University

Waco, Texas 76798-7334

The advantage of using the cell means model approach to the analysis of variance is well documented (Kirk, 1995; Milliken & Johnson, 1987; Searle, 1987; Woodward, Bonett, & Brecht, 1990). The purpose of this paper is to extend the cell means model approach to a variety of analysis of covariance (ANCOVA) designs. A familiarity with the cell means model as described by Kirk (1995) is assumed.

Searle (1987) and Hocking (1985) have described an approach to the cell means model ANCOVA in which an error sum of squares for a reduced model under the restriction that population means are equal is compared to an error sum of squares for the full model under no restrictions. Unfortunately, their approach is not intuitive. An alternative approach to ANCOVA designs is now presented.

Cell Means Model Approach

Completely Randomized Analysis of Covariance Design

For a completely randomized design, the null hypothesis for the cell means model, $\mathbf{C}'_A \boldsymbol{\mu}_y = \mathbf{0}$, can be expressed in a variety of ways where $\mathbf{C}'_A$ is a $(p - 1) \times p$ coefficient matrix of full row rank that defines the null hypothesis and $\hat{\boldsymbol{\mu}}_y$ is a $p \times 1$ vector of response variable cell means. For a design with $p = 4$ treatment levels, one example is

$$\begin{bmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \mu_{\cdot 1} \\ \mu_{\cdot 2} \\ \mu_{\cdot 3} \\ \mu_{\cdot 4} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

The between-groups sum of squares, A_{yy}, for the cell means model is given by

$$A_{yy} = (\mathbf{C}'_A \hat{\boldsymbol{\mu}}_y)' (\mathbf{C}'_A (\mathbf{X}'\mathbf{X})^{-1} \mathbf{C}_A)^{-1} (\mathbf{C}'_A \hat{\boldsymbol{\mu}}_y), \qquad (1)$$

where $\mathbf{X}$ is a $N \times p$ design matrix. Kirk's publication shows the coding of the design matrix

(Kirk, 1995, p. 241). The adjusted between-groups sum of squares, A_{adj}, for a completely randomized analysis of covariance design is given by

$$A_{adj} = (A_{yy} + E_{yy}) - \frac{(A_{zy} + E_{zy})^2}{A_{zz} + E_{zz}} - E_{adj} \tag{2}$$

The sums of squares in the formula can be expressed in matrix notation using a cell means model by defining

$$A_{yy} = (\mathbf{C}'_A\hat{\boldsymbol{\mu}}_y)'(\mathbf{C}'_A(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}_A)^{-1}(\mathbf{C}'_A\hat{\boldsymbol{\mu}}_y)$$

$$A_{zy} = (\mathbf{C}'_A\hat{\boldsymbol{\mu}}_z)'(\mathbf{C}'_A(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}_A)^{-1}(\mathbf{C}'_A\hat{\boldsymbol{\mu}}_y)$$

$$A_{zz} = (\mathbf{C}'_A\hat{\boldsymbol{\mu}}_z)'(\mathbf{C}'_A(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}_A)^{-1}(\mathbf{C}'_A\hat{\boldsymbol{\mu}}_z)$$

$$\hat{\boldsymbol{\mu}}_y = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

$$\hat{\boldsymbol{\mu}}_z = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{z}$$

$$E_{yy} = \mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\mu}}'_y\mathbf{X}'\mathbf{y}$$

$$E_{zy} = \mathbf{z}'\mathbf{y} - \hat{\boldsymbol{\mu}}'_z\mathbf{X}'\mathbf{y}$$

$$E_{zz} = \mathbf{z}'\mathbf{y} - \hat{\boldsymbol{\mu}}'_z\mathbf{X}'\mathbf{z}$$

$$E_{adj} = E_{yy} - \frac{(E_{zy})^2}{E_{zz}}$$

where $\hat{\boldsymbol{\mu}}_z$ is a $p \times 1$ vector of p covariate cell means, $\mathbf{z}$ is an $N \times 1$ vector of covariates, and E_{adj} is the adjusted within-groups sum of squares. The adjusted between- and within-groups mean squares, respectively, are given by the following equations:

$$MSA_{adj} = A_{adj}/(p - 1) \text{ and } MSE_{adj} = E_{adj}/p(n - 1) - 1.$$

The F statistic for the adjusted between-groups treatment is $F = MSA_{adj}/MSE_{adj}$ with $p - 1$ and $p(n - 1) - 1$ degrees of freedom.

Randomized Block Analysis of Covariance Design

The extension of the cell means model to a randomized block ANCOVA design is as follows. The adjusted treatment sum of squares, A_{adj}, is given by

$$A_{adj} = (A_{yy} + R_{yy}) - \frac{(A_{zy} + R_{zy})^2}{A_{zz} + R_{zz}} - R_{adj} \tag{3}$$

where

$$A_{yy} = (\mathbf{C}'_A\mathbf{y})'(\mathbf{C}'_A\mathbf{C}_A)^{-1}(\mathbf{C}'_A\mathbf{y})$$

$$A_{zy} = (\mathbf{C}'_A\mathbf{z})'(\mathbf{C}'_A\mathbf{C}_A)^{-1}(\mathbf{C}'_A\mathbf{y})$$

$$A_{zz} = (\mathbf{C}'_A\mathbf{z})'(\mathbf{C}'_A\mathbf{C}_A)^{-1}(\mathbf{C}'_A\mathbf{z})$$

$$R_{yy} = (\mathbf{R}'\mathbf{y})'(\mathbf{R}'\mathbf{R})^{-1}(\mathbf{R}'\mathbf{y})$$

$$R_{zy} = (\mathbf{R}'\mathbf{z})'(\mathbf{R}'\mathbf{R})^{-1}(\mathbf{R}'\mathbf{y})$$

$$R_{zz} = (\mathbf{R}'\mathbf{z})'(\mathbf{R}'\mathbf{R})^{-1}(\mathbf{R}'\mathbf{z})$$

$$R_{adj} = R_{yy} - \frac{(R_{zy})^2}{R_{zz}}$$

and $\mathbf{R}'$ is a $(p - 1)(n - 1) \times np$ coefficient matrix of full row rank that defines the hypothesis of no interaction between the p levels of treatment A and the n blocks. The latter hypothesis can be expressed as $\mathbf{R}'\boldsymbol{\mu} = \boldsymbol{\theta}$, where $\boldsymbol{\mu}$ is an $np \times 1$ vector of population means, and $\boldsymbol{\theta}$ is a $(p - 1)(n - 1) \times 1$ vector of zeros. An example of this hypothesis for a design with three blocks and three treatment levels is provided as follows:

$$\begin{bmatrix} 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 & 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} \mu_{11} \\ \mu_{21} \\ \mu_{31} \\ \mu_{12} \\ \mu_{22} \\ \vdots \\ \mu_{33} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

The adjusted treatment and residual mean squares, respectively, are given in the following equations:

$$MSA_{adj} = A_{adj}/(p - 1) \qquad \text{and} \qquad MSR_{adj} = R_{adj}/(p - 1)(n - 1) - 1.$$

The F statistic for a mixed model with A fixed and Blocks random is $F = MSA_{adj}/MSR_{adj}$ with $p - 1$ and $(p - 1)(n - 1) - 1$ degrees of freedom.

The adjusted block sum of squares, S_{adj}, is given by

$$S_{adj} = (S_{yy} + R_{yy}) - \frac{(S_{zy} + R_{zy})^2}{(S_{zz} + R_{zz})} - R_{adj}$$

where

$$S_{yy} = (\mathbf{C}'_{BL}\mathbf{y})'(\mathbf{C}'_{BL}\mathbf{C}_{BL})^{-1}(\mathbf{C}'_{BL}\mathbf{y})$$

$$S_{zy} = (\mathbf{C}'_{BL}\mathbf{z})'(\mathbf{C}'_{BL}\mathbf{C}_{BL})^{-1}(\mathbf{C}'_{BL}\mathbf{y})$$

$$S_{zz} = (\mathbf{C}'_{BL}\mathbf{z})'(\mathbf{C}'_{BL}\mathbf{C}_{BL})^{-1}(\mathbf{C}'_{BL}\mathbf{z})$$

and $\mathbf{C}'_{BL}$ is an $(n - 1) \times np$ coefficient matrix of full row rank that defines the block null hypothesis. The adjusted block mean square is given by $MSS_{adj} = S_{adj}/(n - 1)$. The F statistic for Blocks is $F = MSS_{adj}/MSR_{adj}$ with $n - 1$ and $(p - 1)(n - 1) - 1$ degrees of freedom.

The adjusted sums of squares are a linear function of the covariate(s). One of the advantages of these formulations of adjusted sums of squares is that the formulas have only two forms for most designs. Experimental designs that take the form of equation (2) include the completely randomized design, completely randomized factorial design, completely randomized fractional factorial design, generalized randomized block design, and Latin square design. Designs that take the form of equation (3) include the confounded factorial design, randomized block design, randomized block factorial design, and split-plot factorial design.

When one observes formulas (2) and (3) for adjusted sums of squares, one does not ordinarily think of these quantities in terms of the cell means model, although there is no reason not to do so. The main result of this paper (shown in the appendix) stems from the recognition that formulas (2) and (3) can be expressed using the general linear model approach represented by formula (1).

Randomized Block ANCOVA Design with Missing Observations

When one or more observations are missing, several modifications to the analytic procedures just described for a randomized block ANCOVA are necessary because the rows of the treatment coefficient matrix $\mathbf{C}'_A$ and the rows of the block × treatment coefficient matrix $\mathbf{R}'$ are usually not orthogonal. A researcher wants to test a null hypothesis for treatment A subject to the restrictions on the observation parameters, μ_{ij}, that all block × treatment population effects are zero (see Kirk, 1995, pp. 289–301). This is accomplished by computing the sums of squares for A_{yy}, A_{zy}, and A_{zz} using formulas that have the form of

$$SS = (\mathbf{Q}'\mathbf{y})'(\mathbf{Q}'\mathbf{Q})^{-1}(\mathbf{Q}'\mathbf{y}) - (\mathbf{R}'\mathbf{y})'(\mathbf{R}'\mathbf{R})^{-1}(\mathbf{R}'\mathbf{y})$$

instead of

$$SS = (\mathbf{C}'\mathbf{y})'(\mathbf{C}'\mathbf{C})^{-1}(\mathbf{C}'\mathbf{y}),$$

where $\mathbf{Q}'$ is an augmented matrix. To compute A_{yy}, A_{zy}, and A_{zz}, the $\mathbf{Q}'$ augmented matrix is

$$\mathbf{Q}'_A = \begin{bmatrix} \mathbf{R}' \\ \mathbf{C}'_A \end{bmatrix}.$$

The formula for A_{yy}, for example, is

$$A_{yy} = (\mathbf{Q}'_A\mathbf{y})'(\mathbf{Q}'_A\mathbf{Q}_A)^{-1}(\mathbf{Q}'_A\mathbf{y}) - (\mathbf{R}'\mathbf{y})'(\mathbf{R}'\mathbf{R})^{-1}(\mathbf{R}'\mathbf{y}).$$

The formulas for A_{zy}, and A_{zz} are modified in the same manner.

Split-Plot Factorial ANCOVA Design

The extension of the cell means model to a two-treatment split-plot factorial ANCOVA design where treatments *A* and *B* are fixed effects and blocks are random effects is as follows. Consider first the between-blocks sums of squares. The adjusted treatment *A* sum of squares, A_{adj}, is given by

$$A_{adj} = (A_{yy} + E_{yy}) - \frac{(A_{zy} + E_{zy})^2}{A_{zz} + E_{zz}} - E_{adj}$$

where

$$A_{yy} = (\mathbf{C}'_A\mathbf{y})'(\mathbf{C}'_A\mathbf{C}_A)^{-1}(\mathbf{C}'_A\mathbf{y})$$

$$A_{zy} = (\mathbf{C}'_A\mathbf{z})'(\mathbf{C}'_A\mathbf{C}_A)^{-1}(\mathbf{C}'_A\mathbf{y})$$

$$A_{zz} = (\mathbf{C}'_A\mathbf{z})'(\mathbf{C}'_A\mathbf{C}_A)^{-1}(\mathbf{C}'_A\mathbf{z})$$

$$E_{yy} = (\mathbf{C}'_{BL(A)}\mathbf{y})'(\mathbf{C}'_{BL(A)}\mathbf{C}_{BL(A)})^{-1}(\mathbf{C}'_{BL(A)}\mathbf{y})$$

$$E_{zy} = (\mathbf{C}'_{BL(A)}\mathbf{z})'(\mathbf{C}'_{BL(A)}\mathbf{C}_{BL(A)})^{-1}(\mathbf{C}'_{BL(A)}\mathbf{y})$$

$$E_{zz} = (\mathbf{C}'_{BL(A)}\mathbf{z})'(\mathbf{C}'_{BL(A)}\mathbf{C}_{BL(A)})^{-1}(\mathbf{C}'_{BL(A)}\mathbf{z})$$

$$E_{adj} = E_{yy} - \frac{(E_{zy})^2}{E_{zz}}$$

and $\mathbf{C}'_A$ is a $(p - 1) \times npq$ coefficient matrix of full row rank that defines the treatment *A* null hypothesis, $\mathbf{C}'_{BL(A)}$ is a $p(n - 1) \times npq$ coefficient matrix of full row rank that defines the hypothesis that the *n* block population means within each level of treatment *A* are equal, **y** is an $npq \times 1$ response vector, *n* is the number of blocks within each level of treatment *A*, *p* is the number of levels of treatment *A*, and **z** is an $npq \times 1$ covariate vector. The adjusted treatment *A* mean square and adjusted Blocks within A mean square, MSE_{adj}, respectively, are given in the following equations:

$$MSA_{adj} = A_{adj}/(p - 1) \text{ and } MSE_{adj} = E_{adj}/[p(n - 1) - 1].$$

The F statistic is $F = MSA_{adj}/MSE_{adj}$ with $p - 1$ and $p(n - 1) - 1$ degrees of freedom.

Treatment B and the AB interaction are within-blocks sums of squares. The adjusted treatment B sum of squares, B_{adj}, is given by

$$B_{adj} = (B_{yy} + R_{yy}) - \frac{(B_{zy} + R_{zy})^2}{B_{zz} + R_{zz}} - R_{adj}$$

where

$$B_{yy} = (\mathbf{C}'_B\mathbf{y})'(\mathbf{C}'_B\mathbf{C}_B)^{-1}(\mathbf{C}'_B\mathbf{y})$$

$$B_{zy} = (\mathbf{C}'_B\mathbf{z})'(\mathbf{C}'_B\mathbf{C}_B)^{-1}(\mathbf{C}'_B\mathbf{y})$$

$$B_{zz} = (\mathbf{C}'_B\mathbf{z})'(\mathbf{C}'_B\mathbf{C}_{B)})^{-1}(\mathbf{C}'_B\mathbf{z})$$

$$R_{yy} = (\mathbf{C}'_{B\times BL(A)}\mathbf{y})'(\mathbf{C}'_{B\times BL(A)}\mathbf{C}_{B\times BL(A)})^{-1}(\mathbf{C}'_{B\times BL(A)}\mathbf{y})$$

$$R_{zy} = (\mathbf{C}'_{B\times BL(A)}\mathbf{z})'(\mathbf{C}'_{B\times BL(A)}\mathbf{C}_{B\times BL(A)})^{-1}(\mathbf{C}'_{B\times BL(A)}\mathbf{y})$$

$$R_{zz} = (\mathbf{C}'_{B\times BL(A)}\mathbf{z})'(\mathbf{C}'_{B\times BL(A)}\mathbf{C}_{B\times BL(A)})^{-1}(\mathbf{C}'_{B\times BL(A)}\mathbf{z})$$

$$R_{adj} = R_{yy} - \frac{(R_{zy})^2}{R_{zz}}$$

and $\mathbf{C}'_B$ is a $(q - 1) \times npq$ coefficient matrix of full row rank that defines the treatment B null hypothesis and $\mathbf{C}'_{B\times BL(A)}$ is a $p(n - 1)(q - 1) \times npq$ coefficient matrix of full row rank that defines the hypothesis of no interaction between the q levels of treatment B and the n blocks within each level of treatment A. The adjusted treatment B mean square and adjusted B times blocks-within-treatment A mean square, MSR_{adj}, respectively, are given in the following equations:

$$MSB_{adj} = B_{adj}/(q - 1) \text{ and } MSR_{adj} = R_{adj}/[p(n - 1)(q - 1) - 1].$$

The F statistic is $F = MSB_{adj}/MSR_{adj}$ with $q - 1$ and $p(n - 1)(q - 1) - 1$ degrees of freedom. The computation of $MSAB_{adj}$ follows the pattern illustrated for MSB_{adj}.

Split-Plot Factorial ANCOVA Design with Missing Observations

When one or more observations are missing, several modifications to the analytic procedures just described for a split-plot factorial ANCOVA are necessary because the rows of the within-blocks coefficient matrices $\mathbf{C}'_B$ and $\mathbf{C}'_{B\times BL(A)}$ and the rows of $\mathbf{C}'_{AB}$ and $\mathbf{C}'_{B\times BL(A)}$ are usually not orthogonal. A researcher wants to test the null hypotheses for treatment B and the AB interaction subject to the restrictions on the observation parameters, μ_{ijk}, that all $B \times BL(A)$ population effects are zero (see Kirk, 1995, pp. 576–580). This is accomplished by computing the sums of squares for B_{yy}, B_{zy}, B_{zz}, AB_{yy}, AB_{zy}, and AB_{zz} using formulas that have the form

$$SS = (\mathbf{Q}'\mathbf{y})'(\mathbf{Q}'\mathbf{Q})^{-1}(\mathbf{Q}'\mathbf{y}) - (\mathbf{C}'_{B\times BL(A)}\mathbf{y})'(\mathbf{C}'_{B\times BL(A)}\mathbf{C}_{B\times BL(A)})^{-1}(\mathbf{C}'_{B\times BL(A)}\mathbf{y})$$

instead of

$$SS = (\mathbf{C}'\mathbf{y})'(\mathbf{C}'\mathbf{C})^{-1}(\mathbf{C}'\mathbf{y}),$$

where $\mathbf{Q}'$ is an augmented matrix. To compute B_{yy}, B_{zy}, and B_{zz}, the $\mathbf{Q}'$ augmented matrix is

$$\mathbf{Q}'_B = \begin{bmatrix} \mathbf{C}'_{B\times BL(A)} \\ \mathbf{C}'_B \end{bmatrix}.$$

For example, the formula for B_{yy} is

$$B_{yy} = (\mathbf{Q}'_B\mathbf{y})'(\mathbf{Q}'_B\mathbf{Q}_B)^{-1}(\mathbf{Q}'_B\mathbf{y}) - (\mathbf{C}'_{B\times BL(A)}\mathbf{y})'(\mathbf{C}'_{B\times BL(A)}\mathbf{C}_{B\times BL(A)})^{-1}(\mathbf{C}'_{B\times BL(A)}\mathbf{y}).$$

The formulas for B_{zy}, and B_{zz} are modified in the same manner. To compute AB_{yy}, AB_{zy}, and AB_{zz}, the $\mathbf{Q}'$ augmented matrix is

$$\mathbf{Q}'_{AB} = \begin{bmatrix} \mathbf{C}'_{B\times BL(A)} \\ \mathbf{C}'_{AB} \end{bmatrix}.$$

Confounded Factorial ANCOVA Design

The extension of the cell means model to a two-treatment (each treatment with 2 levels) randomized block confounded factorial ANCOVA design (RBCFAC-2^2 design) is as follows. In this design, treatments *A* and *B* are fixed effects, blocks are random effects and the *AB* interaction is completely confounded with Groups.

Consider first the between-blocks sums of squares. The adjusted Groups (*AB* interaction) sum of squares, G_{adj}, is given by

$$G_{adj} = (G_{yy} + E_{yy}) - \frac{(G_{zy} + E_{zy})^2}{G_{zz} + E_{zz}} - E_{adj}$$

where

$$G_{yy} = (\mathbf{C}'_G\mathbf{y})'(\mathbf{C}'_G\mathbf{C}_G)^{-1}(\mathbf{C}'_G\mathbf{y})$$

$$G_{zy} = (\mathbf{C}'_G\mathbf{z})'(\mathbf{C}'_G\mathbf{C}_G)^{-1}(\mathbf{C}'_G\mathbf{y})$$

$$G_{zz} = (\mathbf{C}'_G\mathbf{z})'(\mathbf{C}'_G\mathbf{C}_G)^{-1}(\mathbf{C}'_G\mathbf{z})$$

$$E_{yy} = (\mathbf{C}'_{BL(G)}\mathbf{y})'(\mathbf{C}'_{BL(G)}\mathbf{C}_{BL(G)})^{-1}(\mathbf{C}'_{BL(G)}\mathbf{y})$$

$$E_{zy} = (\mathbf{C}'_{BL(G)}\mathbf{z})'(\mathbf{C}'_{BL(G)}\mathbf{C}_{BL(G)})^{-1}(\mathbf{C}'_{BL(G)}\mathbf{y})$$

$$E_{zz} = (\mathbf{C}'_{BL(G)}\mathbf{z})'(\mathbf{C}'_{BL(G)}\mathbf{C}_{BL(G)})^{-1}(\mathbf{C}'_{BL(G)}\mathbf{z})$$

$$E_{adj} = E_{yy} - \frac{(E_{zy})^2}{E_{zz}}$$

and $\mathbf{C}'_G$ is a $(w - 1) \times nvw$ coefficient matrix of full row rank that defines the Group null hypothesis, $\mathbf{C}'_{BL(G)}$ is a $w(n - 1) \times nvw$ coefficient matrix of full row rank that defines the hypothesis that the population block means within each level of Groups are equal, **y** is an is an $nvw \times 1$ response vector, *n* is the number of blocks within each level of treatment *A*, *w* is the number of levels of Groups, v is the number of combinations of treatments A and B within each block, and **z** an $nvw \times 1$ covariate vector. The adjusted Group mean square, MSG_{adj}, and adjusted Blocks within Groups mean square, MSE_{adj}, respectively, are given by the following equations:

$$MSG_{adj} = G_{adj}/(w - 1) \text{ and } MSE_{adj} = E_{adj}/[w(n - 1) - 1].$$

The F statistic is $F = MSG_{adj}/MSE_{adj}$ with $w - 1$ and $w(n - 1) - 1$ degrees of freedom.

Consider next the within-blocks sums of squares. The adjusted treatment A sum of squares, A_{adj}, is given by

$$A_{adj} = (A_{yy} + R_{yy}) - \frac{(A_{zy} + R_{zy})^2}{A_{zz} + R_{zz}} - R_{adj}$$

where

$$A_{yy} = (\mathbf{C}'_A\mathbf{y})'(\mathbf{C}'_A\mathbf{C}_A)^{-1}(\mathbf{C}'_A\mathbf{y})$$

$$A_{zy} = (\mathbf{C}'_A\mathbf{z})'(\mathbf{C}'_A\mathbf{C}_A)^{-1}(\mathbf{C}'_A\mathbf{y})$$

$$A_{zz} = (\mathbf{C}'_A\mathbf{z})'(\mathbf{C}'_A\mathbf{C}_A)^{-1}(\mathbf{C}'_A\mathbf{z})$$

$$R_{yy} = (\mathbf{C}'_{AB\times BL(G)}\mathbf{y})'(\mathbf{C}'_{AB\times BL(G)}\mathbf{C}_{AB\times BL(G)})^{-1}(\mathbf{C}'_{AB\times BL(G)}\mathbf{y})$$

$$R_{zy} = (\mathbf{C}'_{AB\times BL(G)}\mathbf{z})'(\mathbf{C}'_{AB\times BL(G)}\mathbf{C}_{AB\times BL(G)})^{-1}(\mathbf{C}'_{AB\times BL(G)}\mathbf{y})$$

$$R_{zz} = (\mathbf{C}'_{AB\times BL(G)}\mathbf{z})'(\mathbf{C}'_{AB\times BL(G)}\mathbf{C}_{AB\times BL(G)})^{-1}(\mathbf{C}'_{AB\times BL(G)}\mathbf{z})$$

$$R_{adj} = R_{yy} - \frac{(R_{zy})^2}{R_{zz}}$$

and $\mathbf{C}'_A$ is a $(p - 1) \times nvw$ coefficient matrix of full row rank that defines the treatment A null hypothesis, $\mathbf{C}'_{AB\times BL(G)}$ is a $w(n - 1)(v - 1) \times nvw$ coefficient matrix of full row rank that defines the hypothesis of no interaction between the v combinations of treatments A and B and the n blocks within each level of Groups. The adjusted treatment A mean square and the adjusted

$AB \times BL(G)$ mean of squares, respectively, are given by the following equations:

$$MSA_{adj} = A_{adj}/(p - 1) \text{ and } MSR_{adj} = R_{adj}/[w(n - 1)(v - 1) - 1].$$

The F statistic is $F = MSA_{adj}/MSR_{adj}$ with $p - 1$ and $w(n - 1)(v - 1) - 1$ degrees of freedom. The computation of MSB_{adj} follows the pattern illustrated for MSA_{adj}.

Generalization to More than One Covariate

The adjusted sum of squares formula (2) can be generalized to more than one covariate as follows.

$$A_{adj} = (A_{yy} + E_{yy}) - (\mathbf{a}_{zy} + \mathbf{e}_{zy})'(\mathbf{A}_{zz} + \mathbf{E}_{zz})^{-1}(\mathbf{a}_{zy} + \mathbf{e}_{zy}) - (E_{yy} - \mathbf{e}'_{zy}(\mathbf{E}_{zz})^{-1}\mathbf{e}_{zy})$$

where $\mathbf{a}_{zy}$ is a $p' \times 1$ vector

$$\mathbf{a}_{zy} = (\mathbf{C}'_A\hat{\boldsymbol{\mu}}_z)'(\mathbf{C}'_A(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}_A)^{-1}(\mathbf{C}'_A\hat{\boldsymbol{\mu}}_y)$$

$$= (\mathbf{C}'_A(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z})'(\mathbf{C}'_A(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}_A)^{-1}(\mathbf{C}'_A(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y})$$

$$= (\mathbf{C}'_A(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z})'(\mathbf{C}'_A(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}_A)^{-1}(\mathbf{C}'_A(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y})$$

$$= \mathbf{Z}'(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}_A)(\mathbf{C}'_A(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}_A)^{-1}(\mathbf{C}'_A(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}$$

and $\mathbf{Z}$ is a $np \times p'$ matrix of p' covariates. Following the same pattern for $\mathbf{A}_{zz}$ and $\boldsymbol{A}_{yy}$ we have

$$\mathbf{A}_{zz} = \mathbf{Z}'(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}_A)(\mathbf{C}'_A(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}_A)^{-1}(\mathbf{C}'_A(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Z}$$

$$A_{yy} = \mathbf{y}'(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}_A)(\mathbf{C}'_A(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}_A)^{-1}(\mathbf{C}'_A(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}$$

where $\mathbf{A}_{zz}$ is a $p' \times p'$ matrix.

We can express $\mathbf{e}_{zy}$ as

$$\mathbf{e}_{zy} = \mathbf{Z}'\mathbf{y} - \hat{\boldsymbol{\mu}}'_z\mathbf{X}'\mathbf{y}$$

$$= \mathbf{Z}'\mathbf{y} - \mathbf{Z}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

$$= \mathbf{Z}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}.$$

Following the same pattern for $\mathbf{E}_{zz}$ and E_{yy} we have

$$\mathbf{E}_{zz} = \mathbf{Z}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{Z}$$

$$E_{yy} = \mathbf{y}'(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}.$$

This generalization allows for a smooth transition from the one-covariate case. For some designs the error term represented by E_{yy} may not be appropriate. For example, E_{yy} is not appropriate for a randomized block ANCOVA.

Similarly, the adjusted sum of squares in equation (3) for randomized block designs can generalized to

$$A_{adj} = (A_{yy} + R_{yy}) - (\mathbf{a}_{zy} + \mathbf{r}_{zy})'(\mathbf{A}_{zz} + \mathbf{R}_{zz})^{-1}(\mathbf{a}_{zy} + \mathbf{r}_{zy}) - (R_{yy} - \mathbf{r}'_{zy}(\mathbf{R}_{zz})^{-1}\mathbf{r}_{zy})$$

where $\mathbf{a}_{zy}$, $\mathbf{A}_{zz}$, and $\mathbf{A}_{yy}$ are defined as in the equation above except that we replace $\mathbf{X}'\mathbf{X}$ with the identity matrix. Moreover $\mathbf{r}_{zy}$ can be expressed as

$$\mathbf{r}_{zy} = (\mathbf{R}'\mathbf{z})'(\mathbf{R}'\mathbf{R})^{-1}(\mathbf{R}'\mathbf{y})$$

and $\mathbf{R}$ is defined in equation (3). The quantities $\mathbf{R}_{zz}$, and R_{yy} follow the same pattern as $\mathbf{r}_{zy}$.

Conclusions

The procedures described here provide an intuitive extension of the cell means model to a variety of ANCOVA designs. Examples for the completely randomized design, randomized block design, split-plot factorial design, and completely confounded factorial design were provided. The simplicity of the approach given here can easily be extended to other ANCOVA designs and applications, such as testing the significance of linear and other trend components. The approach allows a researcher to perform an analysis of covariance that is consistent with the cell means model approach to analysis of variance. Defining the research hypothesis in terms of

$\mathbf{C}'\boldsymbol{\mu} = \mathbf{0}$ and then adjusting the sum of squares is very appealing. Furthermore, the formulas are minor modifications of the cell means sum of squares formulas used for ANOVA.

References

Hocking R. R. (1985). *The Analysis of Linear Models.* Pacific Grove, CA: Brooks/Cole.

Kirk, R. E. (1995). *Experimental Design: Procedures for the Behavioral Sciences*. Pacific Grove, CA: Brooks/Cole.

Milliken, G. A., & Johnson, D. E. (1984). *Analysis of Messy Data, Vol. 1: Designed Experiments.* Belmont, CA: Wadsworth.

Searle, S. R. (1987). *Linear Models for Unbalanced Data*. New York, NY: John Wiley.

Seber G. A. F. (1977). *Linear Regression Analysis*. New York, NY: John Wiley & Sons.

Woodward, J. A., Bonett, D. G., & Brecht, M. (1990). *Introduction to Linear Models and Experimental Design.* San Diego, CA: Harcourt Brace Jovanovich.

Appendix

To show that formula (3) can be expressed in terms of the matrix formulation of (1) we use a technique suggested by Seber (1977) involving the addition of one or more regressors to a regression model. It can be shown that the error sum of squares for the regression model with a new regressor can be obtained by minimizing the normal equations

$$\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{W}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\gamma})'(\mathbf{y} - \mathbf{W}\boldsymbol{\beta} - \mathbf{Z}\boldsymbol{\gamma}) \quad (4)$$

where **W** is a matrix of original covariates and **Z** is a new covariate vector (or matrix). We minimize the normal equations by setting $\partial\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} / \partial\boldsymbol{\beta} = \mathbf{0}$ and $\partial\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} / \partial\boldsymbol{\gamma} = \mathbf{0}$. Seber shows that the error sum of squares can be expressed as

$$\begin{aligned}\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} &= (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\mu}} - \mathbf{Z}\hat{\boldsymbol{\gamma}})'(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\mu}} - \mathbf{Z}\hat{\boldsymbol{\gamma}}) \\ &= \mathbf{y}'\mathbf{R}\mathbf{y} - \hat{\boldsymbol{\gamma}}'\mathbf{Z}'\mathbf{R}\mathbf{y}\end{aligned}$$

where

$$\hat{\boldsymbol{\gamma}} = (\mathbf{Z}'\mathbf{R}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{R}\mathbf{y} \text{ and } \mathbf{R} = (\mathbf{I} - \mathbf{W}(\mathbf{W}'\mathbf{W})\mathbf{W}')$$

To obtain the adjusted sum of squares formula in expression (3,) we first obtain the normal equations for the cell means model

$$\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\mu} - \mathbf{Z}\boldsymbol{\gamma})'(\mathbf{y} - \mathbf{X}\boldsymbol{\mu} - \mathbf{Z}\boldsymbol{\gamma})$$

where **X** is a cell means design matrix. The error sum of squares for this model has the same form as equation (4) above, except that **W** is replaced by **X**.

To obtain expression (3) we solve the normal equations $\partial\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} / \partial\boldsymbol{\mu} = \mathbf{0}$ and $\partial\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} / \partial\boldsymbol{\gamma} = \mathbf{0}$ subject to the restriction that $\mathbf{C}'\boldsymbol{\mu} = \mathbf{0}$. Using Lagrangian multipliers, we may obtain the error sum of squares subject to the restriction $\mathbf{C}'\boldsymbol{\mu} = \mathbf{0}$. The cell means estimates under the restriction is

$$\hat{\mu}_r = \mu_y - \mu_z\hat{\gamma}_r - (\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}(\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C})^{-1}\mathbf{C}'(\mu_y - \mu_z\hat{\gamma})$$

and the estimate of γ under the restriction is

$$\hat{\gamma}_r = (\mathbf{Z}'(\mathbf{R}+\mathbf{K})\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{R}+\mathbf{K})\mathbf{y}$$

Thus, the error sum of squares for the reduced model is

$$\begin{aligned} \varepsilon_r'\varepsilon_r &= (\mathbf{y} - \mathbf{X}\hat{\mu}_r - \mathbf{Z}\hat{\gamma}_r)'(\mathbf{y} - \mathbf{X}\hat{\mu}_r - \mathbf{Z}\hat{\gamma}_r) \\ &= \mathbf{y}'(\mathbf{R}+\mathbf{K})\mathbf{y} - \hat{\gamma}_r'\mathbf{Z}'(\mathbf{R}+\mathbf{K})\mathbf{y} \end{aligned}$$

where $\mathbf{K} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C}(\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{C})^{-1}\mathbf{C}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$.

It is a well-known result in the theory of the general linear model that to test the significance of a particular restriction one must take the difference between the restricted and full models. This difference, after much tedious matrix algebra, can be expressed as

$$\varepsilon_r'\varepsilon_r - \varepsilon'\varepsilon = \mathbf{y}'(\mathbf{R}+\mathbf{K})\mathbf{y} - \hat{\gamma}_r'\mathbf{Z}'(\mathbf{R}+\mathbf{K})\mathbf{y} - \mathbf{y}'\mathbf{R}\mathbf{y} + \hat{\gamma}_r'\mathbf{Z}'\mathbf{R}\mathbf{y},$$

which in the case of one covariate may be re-expressed as expression (2).

For standard experimental designs involving blocks, there is only one observation per cell. For this type of design, the error sum of squares for the model is calculated under a restriction (usually denoted as the residual sum of squares). For example, the residual sum of squares for the randomized block design is the error sum of squares under the restriction that all interactions are equal to zero. Moreover, because there is only one observation per cell, we have that $\mathbf{X}'\mathbf{X} = \mathbf{I}_p$ and $\mathbf{X}\mathbf{X}' = \mathbf{I}_n$. For these situations, the residual sum of squares for the model with one covariate is

$$\varepsilon'\varepsilon = \mathbf{y}'\mathbf{K}_1\mathbf{y} - \hat{\gamma}_r'\mathbf{Z}'\mathbf{K}_1\mathbf{y}$$

where

$$\hat{\gamma}_r = (\mathbf{Z}'\mathbf{K}_1\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{K}_1\mathbf{y}$$

and $\mathbf{C}_1'\mu_y$ is a restriction on the cell means that defines the residual sum of squares for the block design under consideration (e.g., for the randomized block model, $\mathbf{C}_1'\boldsymbol{\mu}_{\text{y}}$ is the restriction that all interaction terms are zero), and

$$\mathbf{K}_1 = \mathbf{X}\mathbf{C}_1(\mathbf{C}_1'\mathbf{C}_1)^{-1}\mathbf{C}_1'\mathbf{X}'.$$

To obtain expression (4) we solve the normal equations $\partial\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}/\partial\boldsymbol{\mu} = \mathbf{0}$ and $\partial\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}/\partial\boldsymbol{\gamma} = \mathbf{0}$ subject to the restriction that $\mathbf{C}_1'\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{C}_2'\boldsymbol{\mu} = \mathbf{0}$. Furthermore, we must have $\mathbf{C}_1'\mathbf{C}_2 = \mathbf{C}_2'\mathbf{C}_1 = \mathbf{0}$ because these contrasts define the orthogonal partitions of the total sum of squares. Using Lagrangian multipliers, we may obtain the error sum of squares subject to the restrictions that $\mathbf{C}_1'\boldsymbol{\mu} = \mathbf{0}$ and $\mathbf{C}_2'\boldsymbol{\mu} = \mathbf{0}$. The cell means estimates under these restrictions are

$$\hat{\boldsymbol{\mu}}_r = \hat{\boldsymbol{\mu}}_{\text{y}} - \hat{\boldsymbol{\mu}}_{\text{z}}\hat{\boldsymbol{\gamma}} - \mathbf{C}_1(\mathbf{C}_1'\mathbf{C}_1)^{-1}\mathbf{C}_1'(\hat{\boldsymbol{\mu}}_{\text{y}} - \hat{\boldsymbol{\mu}}_{\text{z}}\hat{\boldsymbol{\gamma}}) - \mathbf{C}_2(\mathbf{C}_2'\mathbf{C}_2)^{-1}\mathbf{C}_2'(\hat{\boldsymbol{\mu}}_{\text{y}} - \hat{\boldsymbol{\mu}}_{\text{z}}\hat{\boldsymbol{\gamma}})$$

and the estimate of $\boldsymbol{\gamma}$ under these restrictions is

$$\hat{\gamma}_r = (\mathbf{Z}'(\mathbf{K}_1 + \mathbf{K}_2)\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{K}_1 + \mathbf{K}_2)\mathbf{y}$$

Thus, the error sum of squares for the reduced model for this block design is

$$\boldsymbol{\varepsilon}_r'\boldsymbol{\varepsilon}_r = \mathbf{y}'(\mathbf{K}_1 + \mathbf{K}_2)\mathbf{y} - \hat{\boldsymbol{\gamma}}_r'\mathbf{Z}'(\mathbf{K}_1 + \mathbf{K}_2)\mathbf{y}$$

where $\mathbf{K}_1 = \mathbf{X}\mathbf{C}_1(\mathbf{C}_1'\mathbf{C}_1)^{-1}\mathbf{C}_1'\mathbf{X}'$ and $\mathbf{K}_2 = \mathbf{X}\mathbf{C}_2(\mathbf{C}_2'\mathbf{C}_2)^{-1}\mathbf{C}_2'\mathbf{X}'$

Thus, the treatment sum of squares is (again, after much tedious matrix algebra)

$$\boldsymbol{\varepsilon}_r'\boldsymbol{\varepsilon}_r - \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = \mathbf{y}'(\mathbf{K}_1 + \mathbf{K}_2)\mathbf{y} - \hat{\boldsymbol{\gamma}}_r'\mathbf{Z}'(\mathbf{K}_1 + \mathbf{K}_2)\mathbf{y} - \mathbf{y}'\mathbf{K}_1\mathbf{y} + \hat{\boldsymbol{\gamma}}_r'\mathbf{Z}'\mathbf{K}_1\mathbf{y},$$

which in the case of one covariate may be re-expressed as equation (3).

The same steps may be used to obtain the adjusted sum of squares for the split plot and confounded factorial designs.