Analysis of Covariance For Repeated Measures Design with Missing Observations

AHMAD BIN ALWI and C.J. MONLEZUN

Guthrie Research Chemara,
Seremban, Negeri Sembilan.

Key words: Repeated measures design; subspaces; noncentrality; orthonormal basis; analysis of covariance.

ABSTRACT

Analysis of covariance is considered as one of the most misunderstood and inadequately taught of all applied statistical methods. Many methods books do not deal with it at all (Guttman et. al., 1982), or sparingly (Brownlee, 1965), and most of those that treat it substantially, such as Federer (1955), Snedecor and Cochran (1980), Steel and Torrie (1980), and Winer (1971), concentrate on balanced data, namely those which have equal numbers of observations in the subclasses. What happens if the data are not balanced and moreover if some of the observations are missing? The missing observations complicate computations and affect what is estimable. The analysis of covariance would become more complex. The application of geometry in the analysis of covariance may offer an understanding of the analysis as well as broaden the variety of methods that can be considered. When there are no missing observations on the repeated measures factor(s), computational algorithms can be used (see Henderson and Henderson, 1979).

INTRODUCTION

The design considered here is a two-factor Repeated Measures Design. Let \( Y_{ijk} \) be the measurement made on subject \( i \) (\( 1 \leq i \leq n_j \)) at level \( j \) (\( 1 \leq j \leq a \)) of factor A and level \( k \) (\( 1 \leq k \leq b \)) of factor B. For every \( Y_{ijk} \), there is a concomitant measurement \( X_{ijk} \). If we let \( a = 3, \ b = 4, \ n_1 = 3, \ n_2 = 2, \ n_3 = 4 \); we can tabulate a data table as in Table 1.

We arbitrarily set the observations \( (Y_{113}, X_{113}), (Y_{312}, X_{312}), (Y_{123}, X_{123}), (Y_{232}, X_{232}), (Y_{335}, X_{335}) \) and \( (Y_{335}, X_{334}) \) be missing.

\(^{1}\)Ass. Prof., Dept. of Exp. Statistics, LSU, U.S.A.
AHMAD BIN ALWI AND C.J. MONLEZUN

TABLE 1
Data table for observations

| A1 | Y_{111} \cdot X_{111} | Y_{211} \cdot X_{211} | Y_{311} \cdot X_{311} |
| A2 | Y_{121} \cdot X_{121} | Y_{221} \cdot X_{221} | Y_{321} \cdot X_{321} |
| A3 | Y_{131} \cdot X_{131} | Y_{231} \cdot X_{231} | Y_{331} \cdot X_{331} |
|    | Y_{451} \cdot X_{451} | Y_{141} \cdot X_{141} | Y_{241} \cdot X_{241} |

*The observations assumed missing.

MODEL DESCRIPTION

In the model where there is no covariate, the model used by A. Ahmad and C.J. Monlezun (1984) is given by

\[ Y_{ijk} = U_{jk} + S_{ij} + E_{ijk} \]

where \( U_{jk} \) is the cell mean for the level \( j \) of factor A and the level \( k \) of factor B, \( S_{ij} \) is the effect of subject i in the level \( j \) of factor A, and \( E_{ijk} \) is the random error component. With the additional concomitant measurement \( X_{ijk} \), a model can now be written as

\[ Y_{ijk}(X) = U_{jk}(X) + S_{ij} + E_{ijk} \]

where \( U_{jk}(X) = \beta_0^{(jk)} + \beta_1^{(jk)} (X_{ijk} - \bar{X} \ldots) \) and \( \beta_0^{(jk)} \) is the intercept for group \((j,k)\) and \( \beta_1^{(jk)} \) is the common slope of all lines. Let \( Y_x \) be the observational vector and it can be written as illustrated in Table 2.

An alternative way of writing the model is

\[ Y_X \sim MVN(\mu_Y, C) = C + X, \]

where

\[ \begin{align*}
\sigma^2_E &+ \sigma^2_S, \\
2 &+ 2 \\
\end{align*} \]

Note that the covariance structure remains the same as in the ANOVA case discussed by A. Ahmad and C.J. Monlezun (1984). All subspaces defined in A. Ahmad and C.J. Monlezun (1984) will be used throughout this paper.

HYPOTHESES TESTING

In the ANOVA case, we have shown that there is no exact test for testing no main effect A. A similar situation prevails in the case of analysis of covariance. Therefore we are interested in testing the following hypotheses:

\[ H_B : \cup_k (X) = \cup_k ' (X) \]

\[ H_{AB} : \cup_{jk} (X) - \cup_{jk} ' (X) = \cup_{jk} (X) - \cup_{jk} ' (X) \]

In the analysis of variance model, all observations for the cell \((j,k)\) have the same mean response \( \cup_{jk} \). This is not so with the covariance model, since the mean response here depends on the \((j,k)\) combinations and also on the value of the concomitant variable \( X_{ijk} \) for the experimental unit. Thus the expected response for the \((j,k)\) cell with the covariance model is given by a regression line:

\[ \cup_{jk} (X) = \beta_0^{(jk)} + \beta_1^{(jk)} (X_{ijk} - \bar{X} \ldots) \]

where

\[ \beta_0^{(jk)} = \cup_\ldots + A_j + B_k + (AB)_{jk} \]
ANALYSIS OF COVARIANCE FOR REPEATED MEASURES DESIGN WITH MISSING OBSERVATIONS

Table 2
Observational vector

\[
Y_{ij} = \begin{bmatrix}
Y_{ij1} & X_{ij1} \\
Y_{ij2} & X_{ij2} \\
Y_{ij3} & X_{ij3} \\
Y_{ij4} & X_{ij4}
\end{bmatrix}
\]

for \( ij = 21, 22, 13, 43 \)

\[
Y_{11} = \begin{bmatrix}
Y_{111} & X_{111} \\
Y_{112} & X_{112} \\
Y_{114} & X_{114}
\end{bmatrix}
\]

\[
Y_{12} = \begin{bmatrix}
Y_{121} & X_{121} \\
Y_{122} & X_{122} \\
Y_{124} & X_{124}
\end{bmatrix}
\]

\[
Y_{31} = \begin{bmatrix}
Y_{311} & X_{311} \\
Y_{313} & X_{313} \\
Y_{314} & X_{314}
\end{bmatrix}
\]

\[
Y_{33} = \begin{bmatrix}
Y_{331} & X_{331} \\
Y_{332} & X_{332} \\
Y_{333} & X_{333}
\end{bmatrix}
\]

\[
Y_{23} = \begin{bmatrix}
Y_{231} & X_{231} \\
Y_{234} & X_{234}
\end{bmatrix}
\]

\[
Y_X = \begin{bmatrix}
Y_{11} & Y_{21} & Y_{31} & Y_{12} & Y_{22} & Y_{13} & Y_{23}
\end{bmatrix}
\]

\[
Y_{33} & Y_{43}
\]

Note that \( \beta_0^{(k)} = \sum_{j=1}^n + B_k \) for \( \Sigma A_j = \Sigma (AB)_{jk} = 0 \) if we want to measure the difference at any convenient point \( X_{ijk} \), say \( X_{ijk} = X \ldots \), then \( (U_{ijk} + B_k) - (U_{ijk} + B_2) = B_1 - B_2 \).

Thus \( B_1 - B_2 \) measures how much higher the mean response is with \( B_1 \) than with \( B_2 \) for any value of \( X_{ijk} \). Therefore for testing no main effect \( B \) in \( U_{jk}(X) \), the regression lines must have equal slopes and the test for main effect \( B \) is \( B_1 - B_2 = 0 \). In other words all of the \( B_k \)'s have to be equal. Fig. 1 illustrates an experiment with four levels of factor \( B \), and how these regression lines might appear.

In constructing the statistics, we first need to define subspaces for the numerator space and the error space. For the error space, we need a space that is orthogonal to \( C_X \) and \( W_s \). Recall from ANOVA case that the smallest subspace containing both \( C \) and \( W_s \) is \( W_s \oplus B \oplus AB \), and thus the smallest subspace that contains only \( X, C, \) and \( W_s \) is

\[
P_{s}X_{11}(W_s \oplus B \oplus AB).
\]

Therefore the error space that we desire must be

\[
E_X = (P_{s}X_{11}(W_s \oplus B \oplus AB))\downarrow = E \cap P_{s}X
\]
Fig. 1: Regression lines of an experiment with four levels of factor B.

From the ANOVA case, we have also defined the following subspaces:

\[ T = N_B \oplus (W_S \oplus AB) \quad \text{for } H_B \]
\[ T = N_{AB} \oplus (W_S \oplus B) \quad \text{for } H_{AB} \]

Let \( M_B = (W_S \oplus AB) \) and \( M_{AB} = (W_S \oplus B) \).

The smallest subspace containing \( W_S \) and \( W_{D,X} \) is

\[ (P_{ND,X} + P_E X) \oplus M_D \quad \text{for } D = B, AB \]

The numerator space for testing \( H_D \) has to be orthogonal to \( E_X, W_S \) and \( W_{D,X} \). Therefore we can define the numerator space by

\[ N_{D,X} = (P_E X + N_D + M_D) \ominus ((P_{ND,X} + P_E) \oplus M_D) = N_{D,X} \oplus \tau_{D,X} \]

where \( N_{D,X} = N_D \ominus P_{ND,X} \) and

\[ \tau_{D,X} = \frac{P_{ND,X}}{X'P_{ND,X}} - \frac{P_E X}{X'P_E X} \]

\section*{DISTRIBUTION OF SUM OF SQUARES/TEST STATISTICS}

Here we are interested to determine the distribution of the numerator and the error sum of squares and hence derive the test statistics for testing \( H_B \) and \( H_{AB} \). For simplicity, we let \( D = B, AB \). The sum of squares for effect \( D \) is defined as

\[ SSD = \sum_{k=1}^{d} [N(\, E(Y_{X_k}) - \, v_k^{\text{null}})]^2 \]

\[ = \sum_{k=1}^{d} \chi^2 \quad \text{(d; noncentral)} \]

where \( d = \begin{cases} (b - 1) & \text{if } D = B \\ (a - 1)(b - 1) & \text{if } D = AB \end{cases} \)
where the noncentrality parameter is given by

\[
\frac{E(Y_{X})' P_{n, X} E(Y_{X})}{2 \sigma^2_E}
\]

Similarly, the distribution of the sum of squares error can be derived as

\[
\text{SSE} = \sum \left[ N(0, \sigma^2_E) \right]^2
\]

\[= \sigma^2_E \chi^2 (t-\alpha-\beta+n; \text{central})
\]

After knowing the distribution of the sum of squares, we can now write a test statistic for testing no effect \( D \) in \( U_{jk}(X) \) as

\[
\frac{\text{SSD} / d}{\text{SSE} / (t-\alpha-\beta+n)} = \frac{\chi^2 (d; \text{noncentral})}{\chi^2 (t-\alpha-\beta+n; \text{central})}
\]

The test statistic above is distributed as a noncentral \( F \) distribution and when the null hypothesis is true, the test statistic becomes central \( F \) distribution.

REFERENCES


(Received 7 February 1985)