Augmenting a linear model with preservation of parameter estimates and standard errors

Vregelaar, ten, J.M.

Published: 01/01/1999

Document Version
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the author's version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Citation for published version (APA):


General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 04. Jan. 2019
Augmenting a linear model with preservation of parameter estimates and standard errors

J.M. ten Vregelaar

Department of Agricultural, Environmental and Systems Technology
Sub-department of Mathematics
Dreijenlaan 4, 6703 HA Wageningen, The Netherlands
E-mail: jan.tenvregelaar@wts.wk.wau.nl

Abstract

In order to estimate expected values for missing plots in an analysis of variance design, covariates and zero observations for the missing plots could be added to the model. We discuss the general idea of augmenting a linear model and show that for any (estimable) parameter(combination) in the original model the estimates and standard errors in the two models are identical. Estimates for standard errors belonging to missing plots can be obtained easily.

Keywords: linear model; least squares estimation; missing values
1. Introduction

In this paper we discuss the use of covariates in case missing plots occur in an analysis of variance design. This appears to be the (most important) example of a general technique of augmenting design and vector of observations in the context of a linear model, which is presented by the paper. Estimates for (estimable) parameter combinations are shown to be equal for both models, the model at hand and its augmented version. The main contribution of the paper is that the same is true for (estimated) standard errors of such parameter combinations. Expressions for (pseudo)variance-covariance matrices are given explicitly. Furthermore, some remarks on testing hypotheses are given.

The idea of using covariates for estimating missing values is due to Bartlett (1937). According to Rao and Toutenburg (1995, section 8.2.3), the method is known under the name Bartlett's ANCOVA. Rubin (1972) gives a brief review of missing value algorithms for analysis of variance designs as alternatives to the exact least squares approach. Analysis of covariance is the only non-iterative method for inserting least squares missing values in his review. He presents a method for performing this procedure which only needs routines for finding residuals and inverting a \( mxm \) symmetric matrix where \( m \) is the number of missing values. Examples of the covariance technique to deal with missing data in a fractionally factorial and a split-plot design are described by Coons (1957). Raghavarao and Nageswararao (1979) present the analysis of a \( (n-1) \) lay-out with covariance to be equivalent to a \( n \)-way non-orthogonal design.

How to obtain standard errors for missing values is not discussed by these authors. We will show the relation between these standard errors and those for the estimators corresponding to the parameters for the covariates.

In the next section we present the idea of augmenting a linear model, whereafter we give an application in section 3. In section 4 the augmentation of section 2 is generalized.

2. Augmenting a linear model

We consider \( n \) observations \( y_1, y_2, ..., y_n \) arranged in a column vector \( y \). We assume a linear model for the expected value of \( y \):

\[
E y = X \beta, \tag{2.1a}
\]

where \( X \) is a \( nxk \) model matrix with possibly linearly dependent columns and \( \beta \) is a column vector consisting of \( k \) parameters \( k \leq n-1 \).

The observations are assumed to be uncorrelated and have equal variance \( \sigma^2 \), which is denoted for short as

\[
\text{var } y = \sigma^2 I_n, \tag{2.1b}
\]

where \( \text{var } y \) represents the variance-covariance or dispersion matrix of \( y \) and \( I_n \) denotes the \( nxn \) unity matrix.

\*) A random variable is denoted by a underlined symbol and its realization by the same symbol without underscore.
We are interested in estimating linear combinations of the components of $\beta$ by means of the (ordinary) least squares method. A linear combination $p'\beta$ is said to be *estimable* if there exists a vector $v$ such that $p = X'v$. This is one possible way to define the estimability of a linear combination, cf. Rao (1973).

According to the well-known Gauss-Markov theorem, the best (minimum variance) linear unbiased estimator (blue) for any estimable linear combination $p'\beta$ is given by $p'\hat{b}$, where $\hat{b}$ satisfies the normal equations

$$X'X\hat{b} = X'y.$$  \hfill (2.2)

The usual (unbiased) estimator for $\sigma^2$ is the mean residual sum of squares

$$\hat{\sigma}^2 = (y'y - b'X'y)/(n-\text{rank}(X)).$$  \hfill (2.3)

Let us now augment the model (2.1) by adding

$$E\tilde{y} = Z\beta + \gamma$$  \hfill (2.4)

where $\gamma$ is a column vector of $m$ additional (uncorrelated) observations having variance $\sigma^2$. Again, $\gamma$ is a column vector of $m$ additional parameters $\gamma_1, \gamma_2, \ldots, \gamma_m$ and $Z$ is any $m\times k$ matrix. Introducing the $(n+m) \times (k+m)$ matrix

$$\tilde{X} = \begin{pmatrix} X & O \\ Z & I_m \end{pmatrix}$$  \hfill (2.5a)

($O$ denoting the $k\times m$ zero-matrix and $I_m$ the $m\times m$ unity matrix), the column vector with $k+m$ components

$$\beta = \begin{pmatrix} \beta \\ \gamma \end{pmatrix},$$  \hfill (2.5b)

and the vector of $n+m$ observations

$$\tilde{y} = \begin{pmatrix} y \\ \gamma_a \end{pmatrix}$$  \hfill (2.5c)

we now obtain the augmented linear model

$$E\tilde{y} = \tilde{X}\beta, \quad \text{var} \tilde{y} = \sigma^2 I_{n+m}.$$  \hfill (2.6)

The realizations of the components of $\tilde{y}$ are the $n$ (realizations of) observations $y_1, y_2, \ldots, y_n$. 

3
and \( m \) arbitrary realizations for the components of \( \mathbf{y}_a \). Often it is suitable to choose zero realizations, cf. section 3.

We will now compare the two models (2.1) and (2.6) with respect to estimating any linear combination \( p' \beta \) and the variance parameter \( \sigma^2 \).

The normal equations for (2.6) \( \bar{X}' \bar{X} \bar{b} = \bar{X}' \bar{y} \) are written as

\[
\begin{pmatrix}
X'X + Z'Z & Z' \\
Z & I_m
\end{pmatrix}
\begin{pmatrix}
b \\
c
\end{pmatrix} =
\begin{pmatrix}
X'y + Z'y_a \\
y_a
\end{pmatrix}
\]

which is equivalent to

\[
\begin{align*}
X'X b &= X'y \\
Zb + c &= y_a.
\end{align*}
\]

If \( b \) satisfies \( X'X b = X'y \) (the normal equations for model (2.1)) then \( b \) also satisfies (2.8) by the obvious solution \( c = y_a - Zb \). Hence, realizations for the \textit{blue} of any estimable \( p' \beta \) are equal for both models (2.1) and (2.6). Notice the equivalence of the estimability of \( p' \beta \) for the two models: the estimability condition in model (2.1)

\[ p = X'v \text{ for some vector } v \]

is equivalent to the one in model (2.6):

\[ \tilde{p} = \bar{X}' \begin{pmatrix} v \\ 0_m \end{pmatrix}, \]

where \( \tilde{p} = \begin{pmatrix} p \\ 0_m \end{pmatrix} \), so \( \tilde{p}' \beta = p' \beta \).

Furthermore, the corresponding residual sum of squares equals

\[
\bar{y}' \bar{y} - \bar{b}' \bar{X}' \bar{y} = y'y + y'_a y_a - (b' c')(X'y + Z'y_a) = y'y - b'X'y
\]

(due to (2.8)), which results in the same estimator of \( \sigma^2 \) as the one given by (2.3), since \( n+m - \text{rank}(\bar{X}) = n+m - (\text{rank}(X) + m) = n - \text{rank}(X) \).
In order to compare (expressions for) (co)variances of blues for estimable linear combinations of the components of \( \beta \) we distinguish two cases.

2.1 Matrix \( X \) is of full column rank: \( X'X \) is non-singular

In this case it is well-known that in model (2.1a), (2.1b) any \( p'\beta \) is estimable and its blue \( p'h \) (where \( h = (X'X)^{-1}X'y \)) has variance \( \text{var}(p'h) = \sigma^2 p'(X'X)^{-1}p \), since the variance-covariance or dispersion matrix of \( h \) is

\[
\text{var} h = \sigma^2 (X'X)^{-1}. \tag{2.10}
\]

The matrix \( \bar{X} \) has \( k+m \) linearly independent rows and therefore it is of full column rank too. Therefore, any linear combination \( \bar{p}'\bar{\beta} \) in model (2.6) is estimable with blue \( \bar{p}'\bar{h} \) (where \( \bar{h} = (X \bar{X})^{-1}X \bar{Y} \)), which has variance \( \text{var}(\bar{p}'\bar{h}) = \sigma^2 \bar{p}'(X \bar{X})^{-1}\bar{p} \).

It is easy to verify that the inverse matrix of \( X'X \) equals

\[
(X \bar{X})^{-1} = \begin{pmatrix}
(X'X)^{-1} & -(X'X)^{-1}Z' \\
-Z(X'X)^{-1} & I_m + Z(X'X)^{-1}Z'
\end{pmatrix}. \tag{2.11}
\]

Now the result \( \text{var} \bar{h} = \sigma^2 (X \bar{X})^{-1} \) for \( \bar{h} = \begin{pmatrix} h \\ c \end{pmatrix} \) in model (2.6) implies that (2.10) is true for model (2.6) as well.

In addition, this result implies

\[
\text{var} \ c = \sigma^2 [I_m + Z(X'X)^{-1}Z'] \tag{2.12a}
\]

and

\[
\text{cov} (\bar{h}, \bar{c}) = E(\bar{h}'\bar{c}') - (E\bar{h})(E\bar{c})' = -(X'X)^{-1}Z'. \tag{2.12b}
\]

The two models give the same result for the variance of \( p'h \) as blue for an arbitrary chosen linear combination \( p'\beta \).

Remark

As a special case of \( \bar{p}'\bar{\beta} \) any linear combination \( q'\gamma \) is estimable. Referring to (2.8) we can compute its estimate as \( q'c = q'(y_a - Zb) \) provided the \( m \) realizations for the components of \( y_a \). Using that \( y_a \) and \( h = (X'X)^{-1}X'y \) are uncorrelated, we can obtain (2.12a) again as follows:

\[
\text{var} c = \text{var}(y_a - Z\bar{h}) = \text{var} y_a + \text{var}(Z\bar{h}) = \text{var} y_a + Z(\text{var} \ h)Z' = \sigma^2 [I_m + Z(X'X)^{-1}Z']. \tag{2.13}
\]

In applications the vector \( Z\bar{\beta} \) will often be the vector of expected values for missing observations. In this case the estimated variance of the blue for such an expected value equals the estimated variance of the blue for the corresponding component of \( \gamma \) minus the estimate for \( \sigma^2 \). This gives us the possibility to obtain the estimated variances for missing values from

Page 5
computer outputs of ANOVA-modules (like in SPSS) which lack the facility of providing them directly.

2.2 Matrix \( X \) is not of full column rank: \( X'X \) is singular

The blue for any estimable \( p' \beta \) in model (2.1) is \( p'b \) (\( b \) satisfying the normal equations \( X'Xb = X'y \)), which has variance

\[
\text{var}(p'b) = \sigma^2 p'(X'X)^{-}p
\]

(2.14)

where \((X'X)^{-}\) is any generalized inverse of \(X'X\), i.e. \((X'X)^{-}\) satisfies \(X'X(X'X)^{-}X'X = X'X\). The matrix \(\sigma^2(X'X)^{-}\) is called a pseudo variance-covariance matrix of \(b\).

Since \(\text{rank}(X) < k\), \(\hat{X}\) has less than \(k+m\) linearly independent rows and therefore \(\hat{X}\) is not of full column rank either. The blue for any estimable \( (p' q') \) \( (\beta, \gamma) \) = \( \tilde{p}' \beta \) in model (2.6) is \( \tilde{p}' \tilde{b} \) where \( \tilde{b} \) satisfies the normal equations \( \hat{X}' \hat{X} \tilde{b} = \hat{X}' \hat{Y} \). The variance of this blue is

\[
\text{var}(\tilde{p}' \tilde{b}) = \sigma^2 \tilde{p}' (\hat{X}'\hat{X})^{-} \tilde{p},
\]

(2.15)

where \((\hat{X}'\hat{X})^{-}\) is any generalized inverse of \(\hat{X}'\hat{X}\). By verifying \( \hat{X}' \hat{X}(\hat{X}'\hat{X})^{-} \hat{X}' \hat{X} = \hat{X}' \hat{X} \) we obtain as a generalization of (2.11) that \((\hat{X}'\hat{X})^{-}\) can be chosen as

\[
(\hat{X}'\hat{X})^{-} = \begin{pmatrix} (X'X)^{-} & -(X'X)^{-} Z' \\ -Z(X'X)^{-} & I_m + Z(X'X)^{-} Z' \end{pmatrix}
\]

(2.16)

for arbitrary \((X'X)^{-}\). Recall that the estimability of any \( p' \beta \) does not depend on which of the models (2.1) or (2.6) is considered. It follows that for model (2.6) the variance of the blue of any estimable \( p' \beta \) can be written as \( \text{var}(p'b) = \sigma^2 p'(X'X)^{-}p \), which again coincides with expression (2.14) for model (2.1). The variance of the blue \( q'\zeta \) for any estimable \( q'\gamma \) is

\[
\text{var}(q'\zeta) = \sigma^2 q' [I_m + Z(X'X)^{-} Z'] q.
\]

We notice that so far there are no restrictions upon the values of \( m \) (the number of \( \gamma \)-parameters), the \( m \times k \) matrix \( Z \) and the vector of additional realizations \( \gamma \). From \( E\gamma = Z\beta + \gamma \) it is obvious that the vector \( \gamma \) is estimable if and only if the vector \( Z\beta \) is estimable. This can also be seen as follows:

\[
\gamma = (O I_m) \begin{pmatrix} \beta \\ \gamma \end{pmatrix} = (O I_m) \hat{X}' \hat{X} = (O I_m) \text{ or (via the (}\hat{X}'\hat{X})^{-} \text{ given by (2.16)) to } Z(X'X)^{-} X'X = Z, \text{ which is the condition for estimability of } Z\beta \text{ in either of the two models (2.1) and (2.6); } O \text{ denotes the } m \times k \text{ zero matrix.}
\]

In general, the estimability condition in model (2.1) for (the components of) the vector \( L\beta \) where \( L \) is any \( l \times k \) matrix reads
\[ L(X'X)^{-1}X'X = L \]

and is equivalent to the one in model (2.6)

\[ L(X'X)^{-1}X'X = \hat{L} \]

where \( \hat{L} = (L \ O) \), so \( \hat{L} \hat{\beta} = L \hat{\beta} \) (now \( O \) denotes the \( l \times m \) zero matrix). The equivalence of the two conditions follows easily from expression (2.16). For \( l = 1 \) we proved this result in the first part of this section without using generalized inverse matrices.

**Remark**

As in section 2.1 the vector \( Z\hat{\beta} \) will often be the vector of expected values for missing observations. We will meet an example in section 3. If \( Z\hat{\beta} \) and hence \( \gamma \) are estimable then (2.13) holds again with \( (X'X)^{-1} \) instead of \( (X'X)^{-1} \). If for some reason the estimated standard error for the missing (expected) value cannot be obtained directly, one can compute it by taking the square root of the difference between estimated variance of the corresponding component of \( \xi \) and the estimate for \( \sigma^2 \).

### 2.3 Hypothesis testing

Let us assume now the *multivariate normal distribution* for the vectors \( \gamma \) and \( \tilde{\gamma} \) in models (2.1) and (2.6) respectively. The matrix \( X \) may or may not be of full column rank. We consider any testable hypothesis (i.e. all components of \( L\hat{\beta} \) are estimable)

\[ H_0: L\hat{\beta} = (L\hat{\beta})_0, \]

\( L \) denoting a matrix of full row rank \( l \). For both models (2.1) and (2.6), the corresponding test statistic can be written as

\[
F = (L\hat{\beta} - (L\hat{\beta})_0)'(\text{var}(L\hat{\beta}))^{-1}(L\hat{\beta} - (L\hat{\beta})_0)/l = \]

\[
= (L\hat{\beta} - (L\hat{\beta})_0)'(L(X'X)^{-1}L')^{-1}(L\hat{\beta} - (L\hat{\beta})_0)/(l\tilde{\sigma}^2) \tag{2.17}
\]

which has under \( H_0 \) an \( F \) distribution with \( l \) and \( n - \text{rank}(X) \) degrees of freedom. The test can be performed in either of the two models. The results are the same. If \( X \) is of full column rank we replace \((X'X)^{-1}\) by \((X'X)^{-1}\).

If the vector \( Z\beta \) is testable we obtain the special case with \( L = Z \) and \( l = m \) (assuming independent rows for the matrix \( Z \)).

If \( \gamma \) is testable (equivalent to \( Z\beta \) is testable cf. (2.4)) we use as test statistic for \( H_0: \gamma = \gamma_0: \)

\[
F = (c - \gamma_0)'(\text{var}(c))^{-1}(c - \gamma_0)/m = (c - \gamma_0)'(I_m + Q)^{-1}(c - \gamma_0)/(m\tilde{\sigma}^2) \tag{2.18}
\]
where $Q$ is the matrix $Q = Z(X'X)^{-1}Z'$ which is non-singular in case $\gamma$ is testable. If we impose now zero additional realizations (all components of $y_a$ are zero) then $c = -Zb$ (cf. (2.8)) and $(I_m + Q)^{-1} = Q^{-1} - Q^{-1}(I_m + Q^{-1})^{-1}Q^{-1}$ imply for the realization of $F$:

$$F = (Zb + \gamma_b)'(I_m + Q)^{-1}(Zb + \gamma_b)/(ms^2) = (Zb + \gamma_b)'Q^{-1}(Zb + \gamma_b)/(ms^2) - (Zb + \gamma_b)'Q^{-1}(I_m + Q^{-1})^{-1}Q^{-1}(Zb + \gamma_b)/(ms^2).$$

The first expression in the right-hand-side of (2.19) is the realization for the $F$-statistic corresponding to $H_0: \beta = -\gamma_b$. The second one, which is not positive, might also be written as $-(Zb + \gamma_b)'(Q + Q^2)^{-1}(Zb + \gamma_b)/(ms^2)$.

Finally it is noticed that the vector $E_{1a} = Z\beta + \gamma = (Z I_m)\beta$ is testable in model (2.6). The value of the $F$-statistic corresponding to $H_0: Z\beta + \gamma = \delta_0$ where $\delta_0$ is any vector with $m$ components is:

$$F = (Zb + c - \delta_0)' [\text{var}(Zb + c)]^{-1} (Zb + c - \delta_0)/m$$

$$= (y_a - \delta_0)' [ (Z I_m) (X'X)^{-1} \begin{pmatrix} Z' \\ I_m \end{pmatrix} ]^{-1} (y_a - \delta_0)/(ms^2)$$

$$= (y_a - \delta_0)' I_m (y_a - \delta_0)/(ms^2) = (y_a - \delta_0)'(y_a - \delta_0)/(ms^2).$$

Summarizing we have seen that the two models (2.1) and (2.6) are equivalent as far as estimating and testing of $\beta$-parameter combinations and $\sigma^2$ is concerned.

The vector $\gamma$ is estimable or testable if and only if $Z\beta$ is estimable in which case

$$\text{var} \gamma = \sigma^2 I_m + \text{var}(Zb),$$

according to (2.11) or (2.16).

3. Application: example with missing data

Consider the data set which was presented by Cochran and Cox (1957), page 111:

<table>
<thead>
<tr>
<th></th>
<th>$B_1$</th>
<th>$B_2$</th>
<th>$B_3$</th>
<th>totals</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_1$</td>
<td>8.00</td>
<td>7.93</td>
<td></td>
<td>15.93</td>
</tr>
<tr>
<td>$A_2$</td>
<td>8.14</td>
<td>8.15</td>
<td>7.87</td>
<td>24.16</td>
</tr>
<tr>
<td>$A_3$</td>
<td>7.76</td>
<td></td>
<td>7.74</td>
<td>15.50</td>
</tr>
<tr>
<td>$A_4$</td>
<td>7.17</td>
<td>7.57</td>
<td>7.80</td>
<td>22.54</td>
</tr>
<tr>
<td>$A_5$</td>
<td>7.46</td>
<td>7.68</td>
<td>7.21</td>
<td>22.35</td>
</tr>
<tr>
<td>totals</td>
<td>30.53</td>
<td>31.40</td>
<td>38.55</td>
<td>100.48</td>
</tr>
</tbody>
</table>
The data represent the breaking strengths of cotton for 5 levels of application of potash, supplying respectively 36, 54, 72, 108 and 144 lb. K₂O per acre \( A_1, \ldots, A_5 \). Since \( B_1, B_2 \), and \( B_3 \) are blocks, an additive model with parametrization

\[
EY_{ij} = \mu + \alpha_i + \beta_j
\]

for the \((A_i, B_j)\)-observation \( Y_{ij} \) is introduced. All observations are by assumption uncorrelated and have equal variance \( \sigma^2 \).

The observations for \((A_1, B_1)\) and \((A_3, B_2)\) are missing. The model matrix \( X \) which belongs to the vectors \( y = (Y_{12}, Y_{13}, \ldots, Y_{53})' \) and \( \beta = (\mu, \alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \beta_1, \beta_2, \beta_3)' \) is

\[
X = \begin{pmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 
\end{pmatrix} = (x_0, x_1, \ldots, x_9),
\]

and has column rank = 9 - 2 = 7 due to the linear dependencies \( x_0 = x_1 + x_2 + \ldots + x_5 \) and \( x_0 = x_6 + x_7 + x_8 \).

A solution of the normal equations \( X'Xb = X'y \) is given by

\[
b = (7.4096, 0.4783, 0.6033, 0.3569, 0.0633, 0, -0.0330, 0.1541, 0)',
\]

cf. the appendix containing SAS-output. The intersection of the linear spaces generated by \( x_1, \ldots, x_5 \) on the one hand and \( x_6, x_7 \) and \( x_8 \) on the other hand has rank one. This means that the design is connected and as a consequence the parameter combinations \( \mu + \alpha_1 + \beta_1 \) and \( \mu + \alpha_3 + \beta_2 \) (the expected values of the two missing observations) are estimable. Their bluehs have realizations 7.8549 (standard error 0.2108) and 7.9206 (standard error 0.2108) respectively, which slightly differ (due to rounding errors) from the estimates 7.86 and 7.92 given by Cochran and Cox who inserted values for the missing ones after some iterative procedure using a formula due to Yates. They only give approximate values for standard errors.

The usual estimate for \( \sigma^2 \) is 0.0491 (mean square error in the SAS output).

Let us now introduce the augmented model

\[
EY_{ij} = \mu + \alpha_i + \beta_j + \gamma_1(z_1)_{ij} + \gamma_2(z_2)_{ij}
\]

where \((z_1)_{11} = 1\) and \((z_1)_{ij} = 0\) otherwise,

\((z_2)_{32} = 1\) and \((z_2)_{ij} = 0\) otherwise.

The covariates \( z_1 \) and \( z_2 \) correspond to the missing values. The matrix \( Z \) of (2.4) reads
\[ Z = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} . \]

We choose zero observations for the missing ones. According to the SAS output belonging to this model with \( y_{11} = y_{32} = 0 \), a solution of the normal equations \( \tilde{X}'\tilde{X} \tilde{b} = \tilde{X}'\tilde{y} \) is given by:

\[ \tilde{b} = \begin{pmatrix} b \\ c \end{pmatrix} , \text{ where } \begin{pmatrix} -7.8549 \\ -7.9206 \end{pmatrix} \text{ is the outcome of the blue for } \gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \end{pmatrix} . \]

Again the variance parameter \( \sigma^2 \) is estimated by \( s^2 = 0.0491 \). The vectors \( Z\tilde{\beta} = \begin{pmatrix} \mu + \alpha_1 + \beta_1 \\ \mu + \alpha_1 + \beta_2 \end{pmatrix} \) and \( \gamma \) are estimable with realizations of blues \( \begin{pmatrix} 7.8549 \\ 7.9206 \end{pmatrix} \) and \( \begin{pmatrix} -7.8549 \\ -7.9206 \end{pmatrix} \) respectively. The estimates 7.8549 and 7.9206 and the corresponding standard errors (both equal to 0.2108) of the expected values of the missing values coincide with the ones given above for the model without covariates, cf. the SAS output from the second PROC GLM.

It is also easily checked from the SAS-output (Std Error of Estimate) that for the estimated variance of \( \xi_i \) (as blue for \( \gamma_i \)), \( \text{vár } \xi_i = \text{vár}(Zb)_i + s^2 \) \( (i = 1, 2) \) holds:

\[ 0.3059^2 = 0.2108^2 + 0.0491 , \]

where \((Zb)_i\) is the blue for \( \mu + \alpha_1 + \beta_1 \) and \((Zb)_2\) is the blue for \( \mu + \alpha_1 + \beta_2 \).

Consequently, the outcomes 37.26 and 37.57 of the t-tests corresponding to \( H_0: (Z\tilde{\beta})_i = 0, \) \( i = 1, 2 \) are in absolute value larger than the ones corresponding to \( H_0: \gamma_i = 0, \) \( i = 1, 2 \) which are equal to \(-25.68\) and \(-25.90\) respectively according to the SAS output.

The outcomes of the F-tests corresponding to \( H_0: \gamma = 0 \) and \( H_0: Z\tilde{\beta} = 0 \) are 760.06 and 1900.14 respectively. The latter being the largest of the two is in agreement with section 2.3.

In some ANOVA-modules (e.g. of SPSS) it is not possible to obtain the estimated variance of the blue for \( \mu + \alpha_1 + \beta_1 \) directly. However, it can be computed now as the estimated variance of \( \xi_i \) minus the estimate \( s^2 \) for \( \sigma^2 \), which are provided in the output.

4. A more general augmentation

Intuitively it seems obvious that the augmentation of model (2.1) by adding (2.4) does not affect parameter estimates and standard errors in model (2.1), since for any additional observation (component of \( \gamma \)) we introduce a new parameter (component of \( \gamma \)). Let us now augment the model (2.1) by adding:

\[ E\xi = Z\tilde{\beta} + W\gamma \]  \hspace{1cm} (4.1)

where \( \xi \) is a column vector of \( l \) additional (uncorrelated) observations having variance \( \sigma^2 \) again, \( \gamma \) is a column vector of \( m \) additional parameters \( \gamma_1, \gamma_2, \ldots, \gamma_m \) and \( Z \) and \( W \) are \( l \times k \) and \( m \times k \), respectively.
\( l \times m \) matrices respectively. We assume that \( m \leq l \) and that \( W \) is of full column rank. Analogous to (2.5) we introduce the \((n+l) \times (k+m)\) matrix

\[
\bar{X} = \begin{pmatrix} X & O \\ Z & W \end{pmatrix}
\]

(4.2)

and the vectors \( \bar{\beta} \) and \( \bar{y} \) (the latter with \( n+l \) observations) again to obtain the augmented model (2.6).

The normal equations for the augmented model are now

\[
\begin{pmatrix} X'X + Z'Z & Z'W \\ W'Z & WW \end{pmatrix} \begin{pmatrix} b \\ c \end{pmatrix} = \begin{pmatrix} X'y + Z'y_a \\ W'y_a \end{pmatrix}
\]

(4.3)

which is equivalent to

\[
\begin{cases}
(X'X + Z'PZ)b = X'y + Z'Py_a \\
\quad c = (W'W)^{-1}W'(y_a - Zb)
\end{cases}
\]

(4.4)

where \( P = I_l - W(W'W)^{-1}W' \) is the orthogonal projection matrix corresponding to projection on the orthogonal complement of the column space of \( W \).

If \( X \) is of full column rank then \( \bar{X} \) is of full column rank too. In this case expression (2.11) can be generalized to

\[
(\bar{X}'\bar{X})^{-1} = \begin{pmatrix} R & -RZ'W(W'W)^{-1} \\ -(W'W)^{-1}W'ZR & (W'W)^{-1}U_m + W'ZRW(W'W)^{-1} \end{pmatrix}
\]

(4.5)

where \( R = (X'X + Z'PZ)^{-1} \).

Let us now consider two special cases.

4.1 Assume \( l = m \)

This assumption implies that the matrix \( W \) is nonsingular and \( P \) equals the \( l \times l \) zero matrix. Therefore the normal equations (4.4) reduce to

\[
\begin{cases}
X'X b = X'y \\
c = W^{-1}(y_a - Zb).
\end{cases}
\]

(4.6)
Analogous to (2.9) and using $Wc + Zb = y_a$ we obtain for the corresponding residual sum of squares

$$\bar{y}'\bar{y} - \bar{b}'\bar{X}'\bar{y} = y'y + y'_a'y_a - (b'c') \left( X'y + Z'y_a \right) = y'y - b'X'y \quad (4.7)$$

with degrees of freedom number equal to $n - \text{rank}(X)$ again, which results in the same estimator (2.3) for $\sigma^2$ as in model (2.1). The (estimated) standard errors in model (2.1) and augmented model (2.6), (4.1) coincide for any (estimable) $p'\beta$, since $R$ in (4.5) reduces to $(X'X)^{-1}$ if $X$ is of full column rank and to $(X'X)^{-}$ if rank($X$) $< k$. For the latter case, which is equivalent to rank($\bar{X}$) $< m+k$, it can be verified that the right-hand-side of (4.5) with $R = (X'X)^{-}$, is a generalized inverse of the singular matrix $\bar{X}'\bar{X}$.

The augmentation (2.4), (2.6) in section 2 is the special case with $W = I_m$.

### 4.2 The column space of $Z$ and $y_a$ are contained in the column space of $W$ (for $l > m$)

This implies $PZ$ equals the zero-matrix and the normal equations (4.4) reduce to

$$\begin{cases} 
X'Xb = X'y \\
= (W'W)^{-1}W'(y_a - Zb) \quad (4.8) 
\end{cases}$$

The expression (4.7) remains true for this case, since $Wc + Zb = y_a$ now follows from (4.8) and $y_a$ being contained in the column space of $W$. However, the corresponding degrees of freedom number equals $(n+l) - \text{rank}(\bar{X}) = (n - \text{rank}(X)) + l - m$ and therefore exceeds the one in model (2.1).

As in section 4.1, expressions for inverse and generalized inverse of $\bar{X}'\bar{X}$ follow from the right-hand-side of (4.5) by taking $R$ equal to $(X'X)^{-1}$ and $(X'X)^{-}$ respectively.

We conclude that the statements on parameter estimates and standard errors in section 2 remain true for the generalization (4.1) if $l = m$ or the column space of $(Z, y_a)$ is contained in the one of $W$. If $l > m$ then the numbers of degrees of freedom for error differ for the models (2.1) and (2.6), (4.1). Therefore we have to multiply the standard error of the estimator of any estimable $p'\beta$ by $\sqrt{[(n - \text{rank}(X)+l-m)/(n-\text{rank}(X))]}$ in order to obtain the corresponding standard error for model (2.1).

Let us illustrate the generalization by considering the example of section 3 again. Instead of adding one observation and one parameter for each missing value we add now three observations and one parameter for each missing value. This corresponds to the following $6 \times 9$ and $6 \times 2$ matrices $Z$ and $W$:
\[
Z = \begin{pmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0
\end{pmatrix}.
\]

\[
W = \begin{pmatrix}
1 \\
1 \\
0 \\
1 \\
0 \\
1
\end{pmatrix}.
\]

We may choose for \(y_a\) any linear combination of the 2 columns of \(W\), e.g. \(y_a = (2 2 2 3 3)'\). It is obvious that \(W\) is of full column rank and that both the column space of \(Z\) and \(y_a\) are contained in the column space of \(W\).

From the SAS-output in the appendix we see that the Parameter Estimates and Error Sums of Squares for the first and third PROC GLM are equal. The Std Errors of Estimate differ due to the difference in degrees of freedom for Error. For the parameter "E(y11)" for instance we verify that the standard error in model (2.1) equals the standard error in model (2.6), (4.1) multiplied by the square root of 10/6 (6 and 10 are the degrees of freedom for Error in models (2.1) and (2.6), (4.1) respectively): 0.2108 = 0.1633*\(\sqrt{10/6}\).

Acknowledgement

I would like to thank L.R. Verdooren for calling my attention to a problem which resulted in this paper.

References


APPENDIX: COCHRAN and COX data set

*** SAS-program ***

options linesize=75 nodate nocenter nonumber;
data one;
  input a b y;
cards;
  1 1 .
  1 2 8.00
  1 3 7.93
  2 1 8.34
  2 2 8.15
  2 3 7.87
  3 1 7.76
  3 2 .
  3 3 7.74
  4 1 7.17
  4 2 7.57
  4 3 7.80
  5 1 7.46
  5 2 7.68
  5 3 7.21;
run;
proc glm data=one;
  class a b;
  model y=a b/solution;
  estimate 'E(y11)' intercept 1 a 1 0 0 0 0 b 1 0 0;
  estimate 'E(y32)' intercept 1 a 0 0 1 0 0 b 0 1 0;
  contrast 'Z*beta' intercept 1 a 1 0 0 0 0 b 1 0 0,
        intercept 1 a 0 0 1 0 0 b 0 1 0;
run;
data two;
  input a b x11 x32 y;
cards;
  1 1 1 0 0
  1 2 0 0 8.00
  1 3 0 0 7.93
  2 1 0 0 8.14
  2 2 0 0 8.15
  2 3 0 0 7.87
  3 1 0 0 7.76
  3 2 0 1 0
  3 3 0 0 7.74
  4 1 0 0 7.17
  4 2 0 0 7.57
  4 3 0 0 7.80
  5 1 0 0 7.46
  5 2 0 0 7.68
  5 3 0 0 7.21;
run;
proc glm data=two;
  class a b;
  model y=a b x11 x32/solution;
  estimate 'E(y11)' intercept 1 a 1 0 0 0 0 b 1 0 0;
  estimate 'E(y32)' intercept 1 a 0 0 1 0 0 b 0 1 0;
  estimate 'E(ya_1)' intercept 1 a 1 0 0 0 0 b 1 0 0 x11 1 x32 0;
  estimate 'E(ya_2)' intercept 1 a 0 0 1 0 0 b 0 1 0 x11 0 x32 1;
  contrast 'Z*beta' intercept 1 a 1 0 0 0 0 b 1 0 0,
        intercept 1 a 0 0 1 0 0 b 0 1 0,
        intercept 1 a 0 0 1 0 0 b 0 1 0 x11 0 x32 1;
  contrast 'E(ya)'  intercept 1 a 1 0 0 0 0 b 1 0 0 x11 1 x32 0,
               intercept 1 a 0 0 1 0 0 b 0 1 0 x11 0 x32 1;
  contrast 'gamma' x11 1,x32 1;
run;
run;

data three;
  input a b x11 x32 y;
  cards;
  1 1 1 0 2.0
  1 1 1 0 2.0
  1 1 1 0 2.0
  1 2 0 0 8.00
  1 3 0 0 7.93
  2 1 0 0 8.14
  2 2 0 0 8.15
  2 3 0 0 7.87
  3 1 0 0 7.76
  3 2 0 1 3.0
  3 2 0 1 3.0
  3 2 0 1 3.0
  3 3 0 0 7.74
  4 1 0 0 7.17
  4 2 0 0 7.57
  4 3 0 0 7.80
  5 1 0 0 7.46
  5 2 0 0 7.68
  5 3 0 0 7.21
;
run;

proc glm data=three;
  class a b;
  model y=a b x11 x32/solution;
  estimate 'B(y11)' intercept 1 a 1 0 0 0 0 b 1 0 0;
  estimate 'B(y32)' intercept 1 a 0 0 1 0 0 b 0 1 0;
run;

*** SAS-output (not complete) from first PROC GLM ***

General Linear Models Procedure
Class Level Information

Class  Levels  Values
  A      5  1 2 3 4 5
  B      3  1 2 3

Number of observations in data set = 15

NOTE: Due to missing values, only 13 observations can be used in this analysis.

Dependent Variable: Y

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>6</td>
<td>0.87319919</td>
<td>0.14553320</td>
<td>2.96</td>
<td>0.1060</td>
</tr>
<tr>
<td>Error</td>
<td>6</td>
<td>0.29469312</td>
<td>0.04911552</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>12</td>
<td>1.16789231</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R-Square</th>
<th>C.V.</th>
<th>Root MSE</th>
<th>Y Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.747671</td>
<td>2.867300</td>
<td>0.22162</td>
<td>7.72923</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Contrast</th>
<th>DF</th>
<th>Contrast SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z*beta</td>
<td>2</td>
<td>186.653004</td>
<td>93.326502</td>
<td>1900.14</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

| Parameter | Estimate | T for H0: Parameter=0 | Pr > |T| | Std Error of Estimate |
|-----------|----------|-----------------------|------|---|-----------------------|
| B(y11)    | 7.85492063 | 37.26 | 0.0001 | 0.21080287 |
| B(y32)    | 7.92063492 | 37.57 | 0.0001 | 0.21080287 |

15
### General Linear Models Procedure

#### Class Level Information

<table>
<thead>
<tr>
<th>Class</th>
<th>Levels</th>
<th>Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>5</td>
<td>1 2 3 4 5</td>
</tr>
<tr>
<td>B</td>
<td>3</td>
<td>1 2 3</td>
</tr>
</tbody>
</table>

Number of observations in data set = 15

#### Dependent Variable: Y

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>104.424280</td>
<td>13.053035</td>
<td>265.76</td>
<td>0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>0.294693</td>
<td>0.049116</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>104.718973</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>R-Square</th>
<th>C.V.</th>
<th>Root MSE</th>
<th>Y Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.997186</td>
<td>3.308423</td>
<td>0.22162</td>
<td>6.69867</td>
</tr>
</tbody>
</table>

#### Contrasts

<table>
<thead>
<tr>
<th>Contrast</th>
<th>DF</th>
<th>Contrast SS</th>
<th>Mean Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Z*beta</td>
<td>2</td>
<td>186.653004</td>
<td>93.326502</td>
<td>1900.14</td>
<td>0.0001</td>
</tr>
<tr>
<td>E(ya)</td>
<td>2</td>
<td>0.000000</td>
<td>0.000000</td>
<td>0.00</td>
<td>1.0000</td>
</tr>
<tr>
<td>gamma</td>
<td>2</td>
<td>74.661454</td>
<td>37.330727</td>
<td>760.06</td>
<td>0.0001</td>
</tr>
</tbody>
</table>

#### Parameter Estimates

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimate</th>
<th>T for H0:</th>
<th>Pr &gt;</th>
<th>Std Error of Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>7.409629630  B</td>
<td>48.18</td>
<td>0.0001</td>
<td>0.15377976</td>
</tr>
<tr>
<td>A1</td>
<td>0.478306878  B</td>
<td>2.30</td>
<td>0.0609</td>
<td>0.20769810</td>
</tr>
<tr>
<td>A2</td>
<td>0.603333333  B</td>
<td>3.33</td>
<td>0.0157</td>
<td>0.18095215</td>
</tr>
<tr>
<td>A3</td>
<td>0.356878307  B</td>
<td>1.72</td>
<td>0.1366</td>
<td>0.20769810</td>
</tr>
<tr>
<td>A4</td>
<td>0.063333333  B</td>
<td>0.35</td>
<td>0.7383</td>
<td>0.18095215</td>
</tr>
<tr>
<td>B1</td>
<td>-0.033015873  B</td>
<td>-0.22</td>
<td>0.8362</td>
<td>0.15293248</td>
</tr>
<tr>
<td>B2</td>
<td>0.154126984  B</td>
<td>1.01</td>
<td>0.3524</td>
<td>0.15293248</td>
</tr>
<tr>
<td>B3</td>
<td>0.000000000  B</td>
<td>.</td>
<td>.</td>
<td>.</td>
</tr>
<tr>
<td>X11</td>
<td>-7.854920635</td>
<td>-25.68</td>
<td>0.0001</td>
<td>0.30586496</td>
</tr>
<tr>
<td>X32</td>
<td>-7.920634921</td>
<td>-25.90</td>
<td>0.0001</td>
<td>0.30586496</td>
</tr>
</tbody>
</table>

**NOTE:** The X'X matrix has been found to be singular and a generalized inverse was used to solve the normal equations. Estimates followed by the letter 'B' are biased, and are not unique estimators of the parameters.

*** output from second PROC GLM ***
The $X'X$ matrix has been found to be singular and a generalized inverse was used to solve the normal equations. Estimates followed by the letter 'B' are biased, and are not unique estimators of the parameters.

*** output from third PROC GLM ***

General Linear Models Procedure
Class Level Information
Class Levels Values
A 5 1 2 3 4 5
B 3 1 2 3

Number of observations in data set = 19

Dependent Variable: Y

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Square</th>
<th>F Value</th>
<th>Pr &gt; F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model</td>
<td>8</td>
<td>114.631023</td>
<td>14.328878</td>
<td>486.23</td>
<td>0.0001</td>
</tr>
<tr>
<td>Error</td>
<td>10</td>
<td>0.294693</td>
<td>0.029469</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corrected Total</td>
<td>18</td>
<td>114.925716</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

R-Square 0.997436  C.V. 2.824437  Root MSE 0.17167  Y Mean 6.07789

Parameter T for H0: Pr > |T| Std Error of
Estimate Parameter=0 Estimate
E(y11) 7.85492063 48.10 0.0001 0.16328720
E(y32) 7.52063492 48.51 0.0001 0.16328720

Parameter T for H0: Pr > |T| Std Error of
Estimate Parameter=0 Estimate
INTERCEPT 7.409629630 B 62.20 0.0001 0.11911729
A 1 0.478306878 B 2.97 0.0140 0.16088225
2 0.603333333 B 4.30 0.0016 0.14016493
3 0.356978307 B 2.22 0.0508 0.16088225
4 0.063333333 B 0.45 0.5610 0.14016493
5 0.000000000 B . . .
B 1 -0.033015873 B -0.28 0.7861 0.11846099
2 0.154126984 B 1.30 0.2224 0.11846099
3 0.000000000 B . . .
X11 -5.854920635 -30.65 0.0001 0.19101260
X32 -4.920634921 -25.76 0.0001 0.19101260

NOTE: The $X'X$ matrix has been found to be singular and a generalized inverse was used to solve the normal equations. Estimates followed by the letter 'B' are biased, and are not unique estimators of the parameters.