Decomposition of the Random Error Vector of a General Linear Model

By Jaesung Choi

Keimyung Univerity

Abstract- This paper deals with the decomposition of an error vector to identify how the error vector is related to the expected value of an observation vector under a general linear sample model since the error vector is defined as the deviance of observation vector from the expected value. The main idea of the paper is in that a random error vector can be decomposed into two orthogonal components vectors; i.e., one is in a vector space generated by the coefficient matrix of the unknown parameter vector and the other is in orthogonal complement of it. As related topics to the decomposition, two things are discussed: partitioning an observation vector and constructing the covariance structure of it. It also shows the reason why a projection method would be preferred rather than a least squares method.
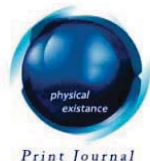
Keywords: coefficient matrix; decomposition; least squares; orthogonal complement; projection; vector space.

GJSFR-F Classification: MSC 2010: 15A03

DECOMPOSITIONOFTHERANDOMERRORVECTOROFAGENERALLINEARMODEL

Strictly as per the compliance and regulations of:

Decomposition of the Random Error Vector of a General Linear Model

Jaesung Choi

Abstract- This paper deals with the decomposition of an error vector to identify how the error vector is related to the expected value of an observation vector under a general linear sample model since the error vector is defined as the deviance of observation vector from the expected value. The main idea of the paper is in that a random error vector can be decomposed into two orthogonal components vectors; i.e., one is in a vector space generated by the coefficient matrix of the unknown parameter vector and the other is in orthogonal complement of it. As related topics to the decomposition, two things are discussed: partitioning an observation vector and constructing the covariance structure of it. It also shows the reason why a projection method would be preferred rather than a least squares method.

Keywords: coefficient matrix; decomposition; least squares; orthogonal complement; projection; vector space.

I. Introduction

When there are some nonrandom quantities affecting a response variable, the analysis of data can be done by using a general linear model. There are some notable references about linear models such as [1-4]. In a general linear model, the response variable is composed of two parts in general; one part is the deterministic portion as a linear function of the unknown parameters of the independent or predictor variables and the other is the random portion. When data are collected, the sample general linear model in matrix form is applicable to the data. However, the matrix equation seems not to be useful for catching any idea about the relationship between the matrix of predictors and error vector only with the assumptions about the error vector; i.e., $E(\epsilon) = 0$ and $var(\epsilon) = \sigma^2 I$. This requires the error vector be the line segment from the origin, which is the point 0, to the point ϵ. The general linear sample model in matrix form is

$$y = X\beta + \epsilon \tag{1}$$

where y denotes the $n \times 1$ vector of observations, X denotes the $n \times p$ matrix of known values, β denotes the $p \times 1$ vector of unknown parameters and ϵ denotes the $n \times 1$ vector of unobserved random errors. The detailed discussion on random errors can be seen in Searle[1]. The matrix equation (1) shows that the vector of observations is composed of mean vector and error vector. When $E(y)$ is a 0 or a mean vector acting like an origin, the error vector satisfies the conditions of mean 0 and $var(\epsilon) = \sigma^2 I$. However, if $E(y) = X\beta$ then we might be interested in how to minimize the deviation vector, $y = X\beta$. The least square method can be used as one of available methods for minimizing the error vector to estimate the parameter vector β. However, the decomposition of an random error vector can be a little bit more comfortable and effective than the minimization of error sum of

Author: Department of Statistics, Keimyung University, 1095 Dalgubeoldaero, Dalseogu, Daegu 42601, Korea.
e-mail: jschoi@kmu.ac.kr

squares by the method of least squares in perspective that the structure of an random error vector is primarily considered. This idea is applied to breaking up an error vector in model (1). Since $X\beta$ is in a specific vector subspace generated by the column space of X, and $\epsilon = y - X\beta$ is nonzero, ϵ is not in the column space of X. The related topics about these can be seen in Graybill[5], Johnson[6] and so forth. Thus, when $X\beta$ is used as a base for getting a error vector of y that is required to have one with the shortest distance from the origin among all the error vectors, we need to break up the error vector into a few component error vectors. This is the main idea of this paper. First, we discuss how to decompose the error vector. Secondly, we study the structure of error vector adjusted for the mean vector. Thirdly, we find the structure of $var(y)$ that is related to the error structure. Finally, we discuss a method that is useful to calculate the sum of squares associated with the error components.

II. Decomposition of Random Error Vector

In the matrix equation (1) let the mean vector $X\beta$ be an nonzero vector. Then the equation can be changed into

$$y = X\beta + \epsilon \qquad (2)$$
$$= X\beta + \epsilon_m + \epsilon_r$$

where ϵ_m and ϵ_r are denoting two component vectors of the error vector. Since $E(y)$ is $X\beta$, ϵ can be broken down into two types of error vector; one type of error vector is the one that is in column space of X, and the other has the characteristic orthogonal to an every vector in a vector space generated by the columns of X, which is an orthogonal complement of the column space of X. In matrix equation (2), $y - X\beta$ defines the error vector based on the mean vector $X\beta$ assumed to have mean $\mathbf{0}$ and $\sigma^2 I$ in n dimensional space. Since the error vector is defined as a vector of deviations from the mean vector of $E(y)$, we can decompose it into two component vectors depending on sources where the error vector is coming up; i.e., ϵ_m or ϵ_r. Let's rewrite the matrix equation (2) in terms of error vector from the mean vector. Then,

$$y - X\beta = \epsilon_m + \epsilon_r \qquad (3)$$
$$= XX^-\epsilon + (I - XX^-)\epsilon$$

where $\epsilon_m = XX^-\epsilon$, $\epsilon_r = (I - XX^-)\epsilon$, and $X^- = (X'X)^{-1}X'$ denotes the Moore-Penrose generalized inverse where $(X'X)^{-1}$ exists because X is a full column rank matrix. From the equation (3), we can know that there are two types of error vector when a vector space is decomposed into two orthogonal vector subspaces based on the mean vector, $X\beta$, of y. The above equation shows that $E(y - X\beta) = \mathbf{0}$ and $var(y - X\beta) = \sigma^2 I$. Here, the error vector $y - X\beta$ is assumed as a linear combination of two different types of error vector each of which coming from two orthogonal vector subspaces. This is completely different concept of an error vector from the one we have thought traditionally. A lot of stuffs can be developed by the newly idea of viewing the error vector as the sum of a mean component error vector and a residual component error vector. Here, two specific terms are used to differentiate the types of error component: a mean component error for ϵ_m and a residual component error for ϵ_r.

III. Structure of Random Error Vector

Since the structure of a random error vector in matrix model (1) is changed depending on the structure of the mean vector, we are going to take a look at it with a bit simpler general linear models. Consider a situation where measurements are measured as deviations from the fixed size for products from a routine process. Let y be a random variable taking an observation on a randomly selected product from a population of products. Data collected from a sample of size n from the population can be arrayed in matrix form. The ith observation of y is expressed as $y_i = 0 + \epsilon_i$ for $i = 1, 2, \ldots, n$, where ϵ_i's are assumed to be

independent with $E(\epsilon_i) = 0$ and $var(\epsilon_i) = \sigma^2$. Applying the sample general linear model (1) to the data, the model turns out to be

$$y = 0 + \epsilon \tag{4}$$

where y is in R_n denoting a Euclidean n-space, and $E(y) = 0$. Let V_0 be the vector space consisting only of 0 and let V_1 be the orthogonal complement of V_0. Since y is the sum of two vectors such that one in V_0 and the other in V_1, $0'\epsilon = 0$ which shows the relationship between mean vector and error vector of y; i.e., an orthogonal property. To express the equation (4) as the second expression in (2), ϵ can be divided into two terms, ϵ_0 and ϵ_1 where ϵ_0 denotes the error vector generated by 0 in a basis of V_0, and ϵ_1 denotes the error vector generated by a basis set of V_1. Now, the matrix equation can be expressed as

$$y = 0 + \epsilon_0 + \epsilon_1 \tag{5}$$

where ϵ_0 is in V_0 of dimension 0, while ϵ_1 is in V_1 of dimension n. Adding the information about the dimension of error component vectors, the equation (5) can be transformed into

$$y - 0 = O\epsilon + (I - O)\epsilon \tag{6}$$

where O represents $n \times n$ zero matrix and I is $n \times n$ identity matrix. The equation shows that ϵ from the origin of rank 0 can be decomposed into two orthogonal vectors one of which being in the orthogonal complement of a vector space. In other words, this means that ϵ is actually composed of a linear combination of two component vectors; i.e. ϵ_0 and ϵ_1. It seems such valuable to grasp the structure of a random error vector for finding the covariance structure of an observation vector. Now, we can study the error structure further with a little bit general but still simpler model having just one quantitative variable as an predictor. Consider the simple linear model with only one nonrandom independent variable in addition to an intercept term; i.e., $y_i = \beta_0 + \beta_1 X_{1i} + \epsilon_i$, for $i = 1, 2, \cdots, n$. We rewrite this in vector and matrix form as

$$y = j\beta_0 + X_1\beta_1 + \epsilon \tag{7}$$

$$= X\beta + \epsilon$$

where $X = (j, X_1)$ is an $n \times 2$ coefficient matrix of β, $\beta = (\beta_0, \beta_1)'$ is an $n \times 2$ parameter vector, j is an $n \times 1$ vector of ones, and X_1 is an $n \times 1$ vector of quantitative values. ϵ is an error vector assumed to have $E(\epsilon) = 0$ and $var(\epsilon) = \sigma^2 I$. The equation (7) is different from the one in (6) in that $E(y) \neq 0$. This is not a surprising thing in a general linear model other than that the mean vector $X\beta$ belongs to the column space of X, V_m of dimension 2 which is thought to be a vector subspace in a Euclidean n-space, R_n. Since the mean vector of y, $X\beta$, is in V_m, ϵ should be divided into two components: one in V_m and the other in $V_m{}^\perp$ denoting an orthogonal complement of V_m in R_n; $V_m \oplus V_m{}^\perp = R_n$. The set of two vectors in the matrix of X can be regarded as a basis set of V_m, which implies that $X\beta$ is in V_m. Hence, the error vector can be divided into two component vectors such that one component in V_m and the other in $V_m{}^\perp$; that is, $\epsilon = \epsilon_m + \epsilon_r$. When we add this kind of information to the equation, the model will be

$$y - X\beta = \epsilon_m + (I - \epsilon_m)\epsilon \tag{8}$$

where $y - X\beta \in R_n$, $\epsilon \in V_m$ and $(I - \epsilon_m)\epsilon$ in $V_m{}^\perp$. Both the set of the columns of a matrix X of rank 2 equivalent to a basis for V_m and the set of the columns of a matrix XX^- can generate the same space, V_m. Hence, the equation (8) can be changed into

$$y - X\beta - XX^-\epsilon = (I - XX^-)\epsilon \tag{9}$$

where $XX^-\epsilon$ replaces ϵ_m and denotes the projection of ϵ_m onto a vector space, V_m, generated by two vectors j and x. The matrix equation (9) turns out to be

$$(I - XX^-)\epsilon = y - (X\beta + XX^-\epsilon) \qquad (10)$$

$$= (I - XX^-)y$$

where $X\beta + XX^-\epsilon = XX^-y$.

IV. COVARIANCE STRUCTURE OF VECTOR OF OBSERVATIONS

From the decomposition of a random error vector of a general linear sample model in matrix form, we can identify that the matrix model (1) can be transformed into

$$y = X\beta + \epsilon \qquad (11)$$

$$= XX^-y + (I - XX^-)y$$

where y is composed of two orthogonal vectors: i.e., $(XX^-y)'(I - XX^-)y = 0$. The model equation (11) implies that all types of a general linear model can be represented by a sum of two orthogonal vectors where one vector belongs to an vector subspace and the other is in the orthogonal complement of the vector space generated by the coefficient matrix of β: i.e., $XX^-y \in V_m$, and $(I - XX^-)y \in V_m^\perp$. Here, the primary concern is actually in structural aspects of an assumed linear model while the least square method focuses only on getting the best approximate solution from a system of inconsistent equations such that $X\beta - y = \epsilon$ by the method of minimizing the error sum of squares. Hence, they are different approaches developed from different view of points. Now, consider the calculation of $var(y)$. The covariance matrix of y is

$$var(y) = var(XX^-y + (I - XX^-)y) \qquad (12)$$

$$= \sigma^2 XX^- + \sigma^2(I - XX^-)$$

$$= \sigma^2 I$$

From the above equation (12), $var(y)$ can be obtained by identifying the linear transformations of y; i.e., the covariance matrix of y can be partitioned as the sum of component covariance matrices, which can be done by ascertaining transformation matrices for component vectors of y. There are some referable literature related to covariance matrix such as Milliken and Johnson[7], Hill[8], and Searle[9]. Hence, it is essential to figure out the coefficient matrices of component error vectors to find the projections of y onto the vector subspaces generated by the orthogonal coefficient matrices. Discussions on coefficient matrices are seen in Choi[10-12], where they are related to get nonnegative variance estimates.

V. PROJECTION METHOD

As a result of the decomposition of e, we see y can be represented by the sum of two orthogonal component vectors such as (11) where one is in a vector space covering the $E(y)$ and the other is in the orthogonal complement of it. This means that XX^-y actually defines a projection of y onto the vector space spanned by the XX^- where X is coefficient matrix of β and given as $X = (j, X_1)$. For the estimation of parameter vector β we can use the mean part of the model in matrix form of (11). From the concept of a projection in a vector space the projection of y onto a column space of X is as follows:

$$X\beta = XX^-y \qquad (13)$$

where $E(y) = X\beta$. When XX^-y is viewed as the orthogonal projection of y onto a column space of X we can take $\hat{\beta}_p = X^-y$ as the value of β where $\hat{\beta}_p$ is an notation for differentiating from $\hat{\beta}$ obtained from the normal equations. When the expression in (13) is viewed as the system of equations, the best approximate solution to the system can be

Notes

obtained as $\hat{\boldsymbol{\beta}} = \boldsymbol{X}^{-}\boldsymbol{y}$ because the system of equations is inconsistent and $\boldsymbol{X}$ is $n \times 2$ matrix of rank 2. Although solutions of $\boldsymbol{\beta}$ can be obtained in different approaches, the results are actually same. In a similar way that $\boldsymbol{X}\boldsymbol{X}^{-}\boldsymbol{y}$ can be used for the estimation of $\boldsymbol{\beta}$, a quadratic form in $\boldsymbol{y}$ can also be used for the estimation of σ^2. Here, the required quadratic form is given as

$$Q_r = \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^{-})\boldsymbol{y} \tag{14}$$

where $(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^{-})$ is a symmetric and idempotent matrix of rank $n-2$. Since $(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^{-})\boldsymbol{y}$ is regarded as a linear transformation of $\boldsymbol{y}$, it has all the information about the residual random error component, ϵ_r. Hence, the quadratic form Q_r in $\boldsymbol{y}$ can be used to estimate the variance σ^2. Taking the expectation of Q_r is given as

$$E(Q_r) = E(\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^{-})\boldsymbol{y}) \tag{15}$$
$$= \sigma^2 tr(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^{-}) + (\boldsymbol{X}\boldsymbol{\beta})'(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^{-})\boldsymbol{X}\boldsymbol{\beta}$$
$$= \sigma^2(n-2)$$

where $tr(\cdot)$ means trace of a square matrix denoted by $(\cdot)$, which is defined to be the sum of the diagonal elements of the square matrix. Some theorems and properties of trace can be seen in Graybill[2]. As an estimate of σ^2 from the equation (15), $\hat{\sigma}_p^2$ can be taken as $Q_r/(n-2)$ which can also be obtained by the least square method when there is no normality assumption for ϵ. Even though those two procedures have the same result, it should be noticed they are basically approaching from different view of point; that is, one is from the decomposition of an error vector, and the other is from the minimization of error sum of squares.

VI.   EXAMPLE

As for an example of a simple linear model, we consider following data from Krumbein and Graybill[13]. The data are assumed to satisfy the model $y_i = \beta_0 + \beta X_{1i} + \epsilon_i$, for $i = 1, 2, \cdots, 10$, where ϵ_i are independent and identically distributed $N(0, \sigma^2)$.

Table 1: Krumbein and Graybill's Data[13].

x_i	550	200	280	340	410	475	160	380	510	510
y_i	200	50	60	140	130	180	20	120	190	160

For the estimation of two unknown parameters, β_0 and β, we can get $\hat{\boldsymbol{\beta}}_p$ by multiplying $(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ on both sides of the equation (13), which is given as:

$$(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{X}^{-}\boldsymbol{y} \tag{16}$$

$$\hat{\boldsymbol{\beta}}_p = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

$$= \begin{pmatrix} 0.982060878 & -2.312086e-03 \\ -0.002312086 & 6.060514e-06 \end{pmatrix} \begin{pmatrix} 1250 \\ 550500 \end{pmatrix} = \begin{pmatrix} -45.2273450 \\ 0.4462054 \end{pmatrix}$$

Denoting $\hat{\beta_{0p}}$, and $\hat{\beta}_p$ as estimates of β_0 and β respectively, $\hat{\beta_{0p}} = -45.2273450$ and $\hat{\beta}_p = 0.4462054$. Least squares estimates are given as $\hat{\beta}_0 = -45.227$ and $\hat{\beta} = 0.446$. For the estimation of σ^2, we can get an estimate as:

$$\hat{\sigma}_p^2 = \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^{-})\boldsymbol{y}/8 = 2398.13/8 = 299.7663 \tag{17}$$

where $(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^{-})$ is the 10×10 matrix of rank 8. As an estimate of σ^2 denoted by $\hat{\sigma}_p^2$ is given as 299.7663 which is the same as the least squares estimate $\hat{\sigma}^2 = 299.766$; hence, $\hat{y} = -45.2273450 + 0.446205X_1$.

VII. Discussion

The primary concern of the study is on the decomposition of an error vector in matrix form of a general linear model. When ϵ is $n \times 1$ vector, the usual assumptions for error vector are sometimes given as $E(\epsilon) = \mathbf{0}$ and $Var(\epsilon) = \sigma^2 \mathbf{I}$. Under these assumptions, an idea for breaking up the error vector lies on the thought of which the mean vector is related to the error vector because the error vector is defined to be the deviation vector from the mean of the model. When the error vector is decomposed into two orthogonal components, it is shown that a projection can be defined from the decomposition of the error vector. Hence, a partition of the vector of observations can be seen as the sum of vectors which are orthogonal projections each other. The covariance matrix of $\boldsymbol{y}$ is partitioned into two covariance matrices; that is, one for $(\boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{y}$, and the other for $(\boldsymbol{I} - \boldsymbol{X}\boldsymbol{X}^-)\boldsymbol{y}$. This implies that the covariance matrix of a vector of observations can always be partitioned into component matrices each of which corresponding to an orthogonal projection of $\boldsymbol{y}$ respectively. From the decomposition of an error vector of a general linear model, we derived two types of estimators; one is linear transformation of $\boldsymbol{y}$ for $\boldsymbol{X}\boldsymbol{\beta}$ to estimate $\boldsymbol{\beta}$ and the other is quadratic form in $\boldsymbol{y}$ for σ^2. Partitioning of the covariance matrix can be useful to ascertain the covariance matrix of each component projection. It is worth to note that decomposition of an error vector is actually defines a projection of $\boldsymbol{y}$ onto a column space of $\boldsymbol{X}$ and which is quite different approach from the least square method in a point of view for an error vector.

VIII. Conclusions

Although the least squares method is very useful and accepted as one of well-known methods for estimating the unknown parameters included in a linear model, it seems not to be right for finding out whether there is any orthogonal property exists among errors. Since the least squares method concentrates only on minimizing the sum of squares of deviations of the observations from the expected values, it is not an appropriate method as a tool for getting the information on an orthogonal property between the groups of errors. The orthogonal property is extremely important in statistics especially in the analysis of variance for getting nonnegative estimates for variance components. There are lots of interesting papers[14-20] related to the negative estimates of variance components seemed to be caused by overlooking the orthogonality. So, it is emphasized that the orthogonal property can be found by the decomposition of the random error vector. Hence, the procedure discussed on this paper is distinct from any other methods for estimating the unknown parameters in a general linear model.

References Références Referencias

1. Searle, S. R. *Linear models;* John Wiley and Sons: New York, USA, 1971.
2. Graybill, F.A. *Theory and Application of the Linear Model;* Wadsworth: Belmont, CA, USA, 1976.

Notes

3. Stapleton, J.H. *Linear Statistical Models;* John Wiley and Sons: New York, USA, 1995.
4. Draper, N.R.; Smith, *H. Applied Regression Analysis;* John Wiley and Sons: New York, USA, 1981.
5. Graybill, F.A. *Matrices with Applications in Statistics;* Wadsworth: Belmont, CA, USA, 1983.
6. Johnson, D.W.; Wichern, R.A. *Applied multivariate statistical analysis;* Prentice hall: Upper Saddle River, NJ, USA, 2014.
7. Milliken, G.A.; Johnson, D. E. *Analysis of messy data volume 1:* designed experiments; Van Nostrand Reinhold: New York, USA, 1984.
8. Hill, B.M. Inference about variance components in the one-way model, *J. Am. Stat. Assoc.* 1965, 60, pp. 806–825.
9. Searle, S.R.; Casella G.; McCulloch, C.E. *Variance components*; John Wiley and Sons, New York, USA, 2009.
10. Choi, J. Nonnegative estimates of *variance components* in a two-way random model. *Communications for Statistical Applications and 127 Methods* 2019, 26, pp. 1–10.
11. Choi, J. Nonnegative variance component estimation for mixed-effects models. *Communications for Statistical Applications and Methods* 2020, 27, pp.523–533.
12. Choi, J. Nonnegative estimation of variance components for a nested three-way random model. *symmetry 2022, 14,* 1210. https://doi.org/10.3390/sym 14061210.
13. Krumbein, W.C.; Graybill, F.A. *An introduction to statistical models in geology*; McGraw-Hill, New York, USA, 1965.
14. Thompson,W.A. Negative estimates of variance components: an introduction; Bulletin, *International Institute of Statistics* 1961,34, 134pp. 1–4.
15. Thompson,W.A. The problem of negative estimates of variance components, *Ann.Math.Stat.* 1962,33, pp. 273–289.
16. Thompson W.A.; Moore, J.R. Non-negative estimates of variance components, *Technometrics* 1963,5, pp. 441–449.
17. Nelder, J.A. The interpretation of negative components of variance, *Biometrika* 1954, 41, pp. 544–548.
18. Henderson, C.R. Estimation of variance and covariance components, *Biometrics* 1953, 9, pp. 226–252.
19. Hartley, H.O. Expectations, variances and covariances of ANOVA means squares by "synthesis", *Biometrics* 1967, 23, pp.105–114.
20. Satterthwaite, F.E. An approximate distribution of estimates of variance components, *Biometrics Bulletin* 1946, 2, pp.110–114.

Notes