Statistics



Original Research Paper

Analysis of Variance Application with Generalized Inverse Matrix

Senol Celik	Bingol University, Agricultural Faculty, Department of Animal Science, Bingol, Turkey
Fahrettin Arslan	Ankara University, Science Faculty, Department of Statistics, An- kara, Turkey
In this study, a linear more model is a non-full rank r The same results were of non-full-rank models, sol	del of Y=X β + ϵ is used. The X'X matrix of the normalequations of(X^' X) β =XYobtained from this natrix. It is solved with generalized G-inverse since it is not the inverse of this non-full-rank matrix. It is when the results of the generalized inverse matrix and variance analysis are compared. In ution with generalized inverse matrix is important in terms of implementation.

KEYWORDS	Non-full rank matrix, generalized inverse, linear model

Introduction

Analysis of varianceis examined while using linear models the statistics of qualitative variables (Ipek, 1980). Extensions of these ideas to general operators have been made (Tseng, 1949). But no systematic study of the subjectwas made until 1955 when (Penrose, 1955), unannounced of the earlier work, redefined the Moore inverse in a different way.

In1920, Moore published the first work on generalized inverses (Moore, 1920). Bott and Duffin (1953) defined what is called a constrained inverse of a square matrix, which is different from a G inverse and is useful in some applications. Chernoff (1953) considered an inverse of a singular nonnegative definite matrix, which is also not a G inverse but is useful in discussing some estimation problems. Mazmanoğlu and Kahramaner (2004) examined by generalized inverse matrices solution of qualitative variable non-full rank models. Nowadays, matrix algebra is used in all branches of mathematics and the sciences and constitutes the basis of most statistical procedures (Abdi and Williams, 2010).

The goal of this research, analysis of variance is carried out by generalized inverse matrix and present the application.

Material and Method

The marks of students entered to postgraduate foreign language exam, in Ankara University Turkish and Foreign Language Research and Application Center, were examined. Whether or significant differences or not among marks of students whom of refer to Graduate School of Natural and Applied Sciences, Graduate School of Health and Applied Sciences, Graduate School of Social and Applied Sciences, Graduate School of Education and Applied Sciences among in the exam was tested. The data were used Senol Celik's (2006) in master thesis.

The model it endeavor is $Y = X\beta + \varepsilon$

where Y is an Nx1 vector observations Y_i, β is a px1 vector of parameters, X is an Nxp matrix of known values (in most cases 0 and 1) and ε is a vector of random error terms (Searle, 1997). Firstly, ε can be considered defined as $\varepsilon = Y - E(Y)$ so that $E(\varepsilon) = 0$ and $E(Y) = X\beta$.

Every element in ε is assumed to have variance σ^2 and zero covariance with every other element.

$$V(\varepsilon) = E(\varepsilon \varepsilon') = \sigma^2 I_N$$
 (Rao, 1973; Searle, 1997).

Thus

 $\varepsilon \sim N(0, \sigma^2 \mathbf{I})$ and $Y \sim N(X\beta, \sigma^2 \mathbf{I})$

with normality being introduced subsequently (Searle, 1997). The normal equations corresponding to the model $Y = X\beta + \varepsilon$ can be derived by least squares (Equation 1).

$$X'X\hat{\beta} = X'Y$$

Linear model of one-way ANOVA are examined.

 $y_{ij} = \mu + \alpha_i + \varepsilon_{ij}$, $j = 1, 2, ..., n_i$ i = 1, 2, ..., k

(2)

(1)

where μ : population mean, α_i : the effect of type i, ε_{ij} : random error term peculiar to the observation y_{ij} (Searle, 1997).

$$E(y_{ij}) = \mu + \alpha_i , \ j = 1, 2, ..., n_i , \ i = 1, 2, ..., k ,$$

$$E(\varepsilon_i) = 0, \ V(\varepsilon_i) = \sigma^2 \text{and} \varepsilon_i \text{'s independent (Akdeniz and Öztürk, 1996).}$$
(3)

They are easily rewritten in the form $Y = X\beta + \varepsilon$ as

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ \vdots \\ \vdots \\ y_{ij} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \cdots & \cdots \\ 1 & 1 & 0 & \cdots & 0 \\ \vdots & 0 & 1 & 0 & \vdots \\ \vdots & \vdots & \vdots & 1 & \vdots \\ \vdots & \vdots & 0 & \vdots & \vdots \\ 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_i \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \vdots \\ \varepsilon_{ij} \end{bmatrix}$$
(4)

where, *Y*: vector of observation, X is the design matrix of 0 and 1, β : the vector of parameters, ε : the vector of error terms (Akdeniz and Öztürk, 1996; Searle 1997). *Y* = *X* β + ε model parameters are given in Table 1.

	Param	eters				
Observations	μ	α_1	α_2	α_{3}		α_{i}
<i>Y</i> ₁₁	1	1	0	0	:	0
y_{12}	1	1	0	:	:	0
:	:	0	1	:	:	:
:	:	:	:	:	:	:
•	:	:	0	1	:	:
${\cal Y}_{ij}$	1		:		:	1

Table 1.Parameters of $Y = X\beta + \varepsilon$ model

In Table 1, as in equation (4), it is clear that the sum the last i columns equals the first column. The first column of X is all 1's; and every Y also contains just one α and so the sum the last i columns is also 1's. Thus X not of full column rank. The consider the normal equations (1). XX' is square and symmetric (Equation 5). XX' is not of full column rank.

$$X'X = \begin{bmatrix} n & n_1 & n_2 & \cdots & n_i \\ n_1 & n_1 & \vdots & \vdots & 0 \\ n_2 & 0 & n_2 & \vdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ n_i & 0 & \cdots & \cdots & n_i \end{bmatrix}$$
(5)

Normal equations involve the vector X'Y, its elements are the inner products of the columns of X with the vector Y. The elements of X'Y are the sums of elements of Y from (4).

$$X'Y = \begin{bmatrix} 1 & 1 & \cdots & \cdots & 1 \\ 1 & 1 & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & 1 & \cdots & 0 & \vdots \\ \vdots & \vdots & \vdots & 1 & \vdots & \vdots \\ 0 & \cdots & \cdots & 1 & 1 \end{bmatrix} \begin{bmatrix} Y_{11} \\ Y_{12} \\ \vdots \\ \vdots \\ Y_{ij} \end{bmatrix} = \begin{bmatrix} Y_{11} + Y_{12} + \dots + Y_{ij} \\ Y_{11} + \dots + Y_{1j} \\ Y_{21} + \dots + Y_{2j} \\ \vdots \\ Y_{ij} \end{bmatrix} = \begin{bmatrix} Y \\ Y_1 \\ Y_1 \\ Y_2 \\ \vdots \\ Y_{ij} \end{bmatrix}$$
(6)

This is nature of X'Y in linear models a vector various subtotals o the Y observations (Equation 6). XX' is not of full rank, as in (5), the normal equations (1) cannot be solved with one solution $\hat{\beta} = (X'X)^{-1}X'Y$. To underline this it is written the normal equations as

$$XX\beta^0 = XY \tag{7}$$

using the symbol to distinguish the many solutions of (7) from the solution that exist when X'X has full rank. β^0 is used a solution GXY to (7), where G is a generalized inverse of X'X.

[n	n_1	n_2	•••	n_i	$\left[\mu^{0} \right]$		$\left\lceil Y \right\rceil$
n_1	n_1	:	÷	0	α_1^0		Y_1
$ n_2 $	0	n_2	÷	0	α_2^0	=	Y_2
:	÷	÷	۰.	:			:
n_i	0	•••	•••	n_i	α_i^0	i	Y_{ij}

X'X does not have full column and X'X has no inverse. Therefore, the normal equations (7) have no unique solution. To get any one of them it is found any generalized inverse G of X'X and write the corresponding solution as

$$\beta^0 = GX'Y \tag{9}$$

The notation β^0 in equation (9) for a solution to the normal equations (7) emphasizes that what is derived by solving (7) is only a solution to the equations (Searle, 1997).

Let A be a matrix of order mxn. Then, generalized inverse (G-inverse) any nxm matrix A, denoted by $A^- = G$, is a matrix of order nxm that satisfies the relation AGA=A (Khuri, 2003; Kabe and Gupta, 2007).

Results

Whether or significant differences or not among marks of students whom of refer to postgraduate exam was investigated. Exam marks of total 380 students whom of refer to Graduate School of Natural and Applied Sciences, Graduate School of Health Sciences, Graduate School of Social Sciences, Graduate School of Education Sciences was researched. This problem is solved according to One-Way Analysis of Variance (ANOVA) It is reached to solution by Generalized Inverse. The model for analysis is the fixed-effects and the null hypothesis is (Chiang, 2003).

 $H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$ and

 H_1 : (At least two institutions of student averages achievement are different).

 $(\mu_i \neq \mu_j) \ i, j = 1, 2, 3, 4$

Where,

 μ_1 = Institute of Science and Technology their students average achievement from foreign language exam

 μ_2 = Institute of Health Sciences their students average achievement from foreign language exam

 μ_3 = Institute of Social Sciences their students average achievement from foreign language exam

 μ_4 = Institute of Education Sciences their students average achievement from foreign language exam.

 $n_1 = n_2 = n_3 = n_4 = 95$ $n = n_1 + n_2 + n_3 + n_4 = 380$ (Total observation) k = 4 (Number of group).

Table 7. The Iolal and average number of foreign language exam grade	Tal-1. 1	The total	and arranges		offorming	1	ma anada
Twole 2. The form who we ended of foreight fully and the firm	Table 2.	The total	and average	number	of foreign	language exa	m grade.

Institute	Total success notes	The average success
Science and Technology	$y_1 = \sum_{i=1}^{n_1} y_{1i} = \sum_{j=1}^{95} y_{1j} = 5398$	$\overline{y}_{1} = \frac{\sum_{j=1}^{n_{1}} y_{1j}}{n_{1}} = \frac{5398}{95} = 56,821$
Health Sciences	$y_2 = \sum_{j=1}^{n_2} y_{2j} = \sum_{i=1}^{95} y_{2j} = 5748$	$\overline{y}_2 = \frac{\sum_{j=1}^{n_2} y_{2j}}{n_2} = \frac{5748}{95} = 60.505$
Social Science	$y_3 = \sum_{j=1}^{n_3} y_{3j} = \sum_{j=1}^{95} y_{3j} = 6001$	$\overline{y}_3 = \frac{y_3}{n_3} = \frac{6001}{95} = 63.168$
Education Science	$y_4 = \sum_{i=1}^{n_4} y_{4j} = \sum_{j=1}^{95} y_{4j} = 5673$	$\overline{y}_4 = \frac{y_4}{n_4} = \frac{5673}{95} = 59.716$
Overall Success	$Y = \sum_{i=1}^{k} y_i = \sum_{i=1}^{k} \sum_{j=1}^{n_j} y_{ij} = \sum_{i=1}^{4} \sum_{j=1}^{95} y_{ij} = 22820$	$\overline{Y} = \frac{\sum_{i=1}^{k} n_i \overline{y}_i}{n} = \frac{\sum_{i=1}^{4} \overline{y}_i}{4} = 60.0526$

Matrix representation according to the model $Y = X\beta + \varepsilon$

$$Y = \begin{bmatrix} 64\\47\\\vdots\\\vdots\\77\\36 \end{bmatrix}_{(380x1)} \qquad X = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots\\1 & \vdots & \vdots & \vdots & \vdots\\\vdots & 0 & 1 & \vdots & \vdots\\\vdots & \vdots & 0 & 1 & \vdots\\1 & \vdots & \vdots & 0 & 1\\1 & 0 & 0 & \cdots & \vdots \end{bmatrix} \beta = \begin{bmatrix} \mu\\\alpha_1\\\alpha_2\\\vdots\\\alpha_i \end{bmatrix}_{(i+1)x1} \varepsilon = \begin{bmatrix} \varepsilon_{11}\\\varepsilon_{12}\\\vdots\\\varepsilon_{i1}\\\varepsilon_{ij} \end{bmatrix}$$

is shaped.

$$Y = \begin{bmatrix} 64 \\ 47 \\ \vdots \\ 77 \\ 36 \end{bmatrix}_{(380x1)} = \begin{bmatrix} 1 & 1 & 0 & 0 & \cdots \\ 1 & \vdots & \vdots & \vdots & \vdots \\ \vdots & 0 & 1 & \vdots & \vdots \\ \vdots & \vdots & 0 & 1 & \vdots \\ 1 & \vdots & \vdots & 0 & 1 \\ 1 & 0 & 0 & \cdots & \vdots \end{bmatrix}_{380x5} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_k \end{bmatrix}_{5x1} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \vdots \\ \varepsilon_{i1} \\ \varepsilon_{ij} \end{bmatrix}$$
$$Y = X \qquad \beta + \varepsilon$$

X: Design matrix with values 0 and 1 design matrix. Rank of the X design matrix is 4 and number of parameter is 5. X', X'X and X'Y matrix are created to determine the normal equation

	1	1	•••	•••	•••	•••	•••	1
	1	÷	0	•••	•••	•••	•••	
X' =	0	÷	1	•••	0	•••	•••	0
	÷	÷	÷	•••	1	•••	0	
	:	÷	•••	•••	•••	•••	1] _{5x380}

is obtained.

as calculated.

									[64]			
	[1	1	•••	•••	•••	•••	•••	1]	47		22820	
	1	÷	0	•••	•••	•••	•••		4/		5398	
X'Y =	0	÷	1	•••	0	•••	•••	0		=	5748	
	:	÷	÷	•••	1	•••	0				6001	
	Ŀ	÷	•••	•••	•••	•••	1]	26		5673	
									[30]			

as calculated.

 $(X'X)\hat{\beta} = X'Y$

380	95	95	95	95]	Γµ ⁻		22820
95	95	0	0	0	$\hat{\alpha}_1$		5398
95	0	95	0	0	$\hat{\alpha}_2$	=	5748
95	0	0	95	0	$\hat{\alpha}_3$		6001
95	0	0	0	95	$\lfloor \hat{\alpha}_4 \rfloor$		5673

is found. $\hat{\beta}$ vector system is estimated from this equation. X'X matrix is not full rank. $\hat{\beta}$ vector is estimated using generalized inverse. Generalized G inverse of X'X

	0	0	0	0	0
	0	1/95	0	0	0
G =	0	0	1/95	0	0
	0	0	0	1/95	0
	0	0	0	0	1/95

as calculated. Equality from $\hat{\beta}^0 = GXY'$

		0	0		0	0	0	22820
		0	1/9	95	0	0	0	5398
$\hat{\beta}$	0 =	0	0		1/95	0	0	5748
		0	0		0	1/95	0	6001
		0	0		0	0	1/95	5673
	[0	_		0]		
	53	98/	95		56.82	1		
=	57	48/	95	=	60.50	5		
	60	01/	95		63.16	8		
	56	73/	95		59.71	6		

is obtained.

Sum of Squares Between Groups (SSB), Sum of Squares Within Groups (SSW) and General Sum of Squares (SST) are calculated to F test created.

 $SSB = \beta^{0'}X'Y - ny^{-2}$

n = 380 (Number of student) and $\overline{y} = 60,0526$ (average mark).

$$SSB = \begin{bmatrix} 0 & 56.821 & 60.505 & 63.168 & 59.168 \end{bmatrix} \begin{bmatrix} 22820 \\ 5398 \\ 5748 \\ 6001 \\ 5673 \end{bmatrix} - 380(60.0526)^2$$

= 1372345.663 - 1370399.611=1946.663

is found or

$$SSB = \sum_{i=1}^{k} n_i (\overline{y}_i - \overline{y})^2 = \sum_{i=1}^{4} n_i (\overline{y}_i - \overline{y})^2$$

= 95x(56,82-60,05)² + 95x(60,51-60,05)² + 95x(63,17-60,05)²
+95x(59,72-60,05)² = 1946.663

as calculated.

Mean Square of Between Groups (MSB),

$$MSB = \frac{SSB}{k-1} = 1946.663 / 3 = 648.888$$

Mean Square Between Groups $MSB = Y'Y - \beta^{0'}X'Y$ is shaped.

$$SSW = \begin{bmatrix} 64 & 47 & \cdots & 77 & 36 \end{bmatrix} \begin{bmatrix} 64 \\ 47 \\ \vdots \\ .77 \\ .36 \end{bmatrix} - \begin{bmatrix} 0 & 56.821 & 60.505 & 63.168 & 59.716 \end{bmatrix} \begin{bmatrix} 22820 \\ 5398 \\ 5748 \\ 6001 \\ 5673 \end{bmatrix}$$

= 1464054 - 1372345.663 = 91708.337

as calculated or

$$SSW = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \overline{y}_j)^2 = \sum_{i=1}^{4} \sum_{j=1}^{95} (y_{ij} - \overline{y}_j)^2$$

= $(64 - 56, 82)^2 + (47 - 56, 82)^2 + ... + (73 - 56, 82)^2 + (67 - 60, 51)^2 + ... + (50 - 60, 51)^2 + (69 - 63, 17)^2 + ... + (64 - 63, 17)^2 + (42 - 59, 72)^2 + ... + (36 - 59, 72)^2 = 91708.337$

as found.

Mean Square Within Groups

$$MSW = \frac{SSW}{n-k} = \frac{91708.337}{380-4} = 243.905$$

General Sum of Squares

 $SST = Y'Y - n\overline{y}^2 = 1464054 - 1370399.611 = 93654.389$ as found.

F statistics,

 $F = \frac{SSB/(k-1)}{SSW/(N-k)} = \frac{MSB}{MSW} = \frac{648.888}{243.905} = 2.66$

as obtained. Analysis of variance results the obtained are shown in Table 3.

Source of	Sum of squares	Degrees of	Mean square	F statistics
Variation		freedom		
Between group	$\hat{\beta}^{0'}X'Y - n\overline{y}^{-2} = 1946.052$	4-1=3	<i>MSB</i> = 648.888	F = 2.66
Within group	$Y'Y - \hat{\beta}^{0'}X'Y = 91708.337$	380-4=376	<i>MSW</i> = 243.905	
General	$Y'Y - n\overline{y}^{-2} = 93654.389$	380-1=379		

Table 3. Table for one-way classification

The F statistics 2,66 is significant. Accordingly 3 and 376 degrees of freedom, the ANOVA table critical value is 2.60. Null hypothesis is reject. The difference among the achievement of students in a foreign language exam is statistically significant.

Conclusion

In this study, $Y = X\beta + \varepsilon$ model is created. β^0 vector as calculated

 $\beta^{0'} = [0\ 56.821\ 60.505\ 63.168\ 59.716]$

SST, SSB and SSW values are found 93654.389, 1946.663 and 91708.337, respectively. MSB and MSW values are found 648.888 and 243.905, respectively. F statistics 2,66 is statistical significant. The system consisting of qualitative variable normal equationssystem is not inverses ince it is not a full rank model. The solution was made with the generalized G inverse.

References

- Abdi H., Williams, L. F. 2010. Matrix Algebra. In Neil Salkind (Ed.), Encyclopedia of Research Design. Thousand Oaks, CA: Sage.
- Akdeniz, F., Öztürk, F. Lineer Modeller. 1996. A.Ü. F. F. Döner Sermaye letmesi Yayınları No:38, Ankara.
- Bott, R., Duffin, R. J. 1953.On the algebra of networks, Trans. Amer. Math. Soc.. 74: 99-109.
- Celik, S. 2006. Normal Distribution and Related Obtainment for the Normal Distribution. Ankara University Graduate School of Natural and Applied Sciences Department of Statistics, Ankara, Master Thesis.
- Chernoff, H. 1953. Locally optimal designs for estimating parameters, Ann. Math.Statist.24: 586-602.
- Chiang, C. L. 2003. Statistical Methods of Analysis. World Scientific Publishing Co. Pte. Ltd, 107-19.
- Ipek, M. 1980. Genelle tirilmi Ters Matrisler ve Rankı Tam Olmayan Modellere Uygulama. stanbul Üniversitesi ktisat Fakültesi Doçentlik Tezi, stanbul.
- Kabe, D. G. Experimental Design: Exercises and Solutions. Springer Science Business Media, LLC, USA.
- Khuri, A. I. 2003. Advanced Calculus with Applications in Statistics. John Wiley Sons, Inc, Hoboken, New Jersey.

- Mazmano lu, A., Kahramaner, Y. 2004. Nitel De i kenli Rankı Tam Olmayan Modellerin Genelle tirilmi Ters Matrislerle Çözümü ve Sosyal Bilimlerdeki Matematik Ö retim Yöntemleri Üzerine Bir Uygulama. stanbul Ticaret Üniversitesi Dergisi, 6: 95-116.
- 11. Moore,E. H. 1935. General Analysis, Philadelphia.American Philosophical Society.
- 12. Rao, C. R. 1973. Linear Statistical Inference and its Applications. Wiley Series in Probability and Mathematical Statistics, United States of America.
- Penrose, R. 1955. A generalized inverse for matrices. Proc. Cambridge Philos. Soc., 51: 406-413.
- 14. Searle, S.R. 1997. Linear Models. N. Y. State College of Agriculture Cornell University, Ithaca, N.Y, 175-177.
- Tseng, Y. Y. 1949. Generalized inverses of unbounded operators between two unitary spaces, Dokl. Akad. Nauk. SSSR., 67: 431-434