Hacettepe Journal of
Mathematics & Statistics

Hacet. J. Math. Stat.
Volume 51 (4) (2022), 1174 – 1188
DOI : 10.15672/hujms.955559

Research Article

Two-way ANOVA by using Cholesky decomposition and graphical representation

Mustafa Tekin, Haydar Ekelik*

Department of Econometrics, Istanbul University, Istanbul, Turkey

Abstract

In general, the coefficient estimates of linear models are carried out using the ordinary least squares (OLS) method. Since the analysis of variance is also a linear model, the coefficients can be estimated using the least-squares method. In this study, the coefficient estimates in the two-way analysis of variance were performed by using the Cholesky decomposition. The purpose of using the Cholesky decomposition in finding coefficient estimates make variables used in model being orthogonal such that important variables can be easily identified. The sum of squares in two-way analysis of variance (row, column, interaction) were also found by using the coefficient estimates obtained as a result of the Cholesky decomposition. Thus, important variables that affect the sum of squares can be determined more easily because the Cholesky decomposition makes the variables in the model orthogonal. By representing the sum of squares with vectors, how the prediction vector in two-way ANOVA model was created was shown. It was mentioned how the Cholesky decomposition affected the sum of squares. This method was explained in detail on a sample data and shown geometrically.

Mathematics Subject Classification (2020). 15A23, 15A06, 62J10

Keywords. Cholesky decomposition, aalysis of variance and covariance (ANOVA), linear equations

1. Introduction

Statistical methods generally guide scientific learning. As a part of this type of learning, linear statistical methods are widely used. In basic and social sciences as well as in business and engineering, linear models benefit both the planning of the research and the analysis of the data obtained [20]. Linear models are called "analysis of variance (ANOVA)" in conditions where the independent variables are categorical. Analysis of variance can be considered as a special case of linear models [7]. So far, many studies have been conducted with two-way analysis of variance [1–3, 14]. There are studies in which the Cholesky decomposition is used both in social sciences and basic sciences [8, 13, 24]. Studies in which linear models and the Cholesky decomposition are used together are also included in the literature [5, 15, 16, 22, 26]. Moreover, there are studies conducted by applying the Cholesky decomposition to the correlation matrix [11, 12, 18, 19]. In this

*Corresponding Author.
Email addresses: mustafatek@istanbul.edu.tr (M. Tekin), haydar.ekelik@istanbul.edu.tr (H. Ekelik)
Received: 22.06.2021; Accepted: 11.04.2022

study, unlike the previous studies, the two-way analysis of variance (two-way ANOVA) model was obtained using the Cholesky decomposition. In our previous study, the one-way ANOVA model was performed using the Cholesky decomposition [23]. In this study, the advantages of the Cholesky decomposition are shown on the two-way ANOVA model, and the geometric representation of the sum of squares in two-way analysis of variance (row, column, interaction) was performed. The purpose of using Cholesky decomposition is to make the variables used in two-way ANOVA model orthogonal (orthonormal). Thus, variables that affect the sum of squares in two-way ANOVA can be determined more easily. This also makes it easier to identify important variables.

The organization of this article is as follows. In Section 2, ANOVA, two-way ANOVA, and the Cholesky decomposition theory are mentioned. In Section 3, two-way ANOVA design, the design matrix, and the Cholesky decomposition are presented on the sample data set, and the sum of squares and vector representations are shown geometrically in detail. Finally, the conclusion of the study is presented in Section 4.

2. Linear models - Analysis of variance (ANOVA)

Linear models are statistical techniques widely used in behavioral science, medical research, marketing research, and other fields [4]. Generally, a linear model whose mathematical form is shown in Equation (2.1) is used to estimate the values of the Y variable, which is called the dependent variable, from the series of x estimators for p amount of variables called independent variables. The most important feature to be considered in linear models is that the dependent variable Y is obtained as result of measurement. In some cases, values such as age and income obtained from the interval scale result can be also used [4].

$$Y_j = \beta_0 + \beta_1 x_{1j} + \beta_2 x_{2j} + ... + \beta_p x_{pj} + \varepsilon_j, \quad j = 1, 2, ..., n \quad i = 1, 2, ..., m \tag{2.1}$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \tag{2.2}$$

$$Y = X\beta + \varepsilon$$

Equation (2.1) is shown with matrices in Equation (2.2). In linear models, Y refers to the dependent variable column vector, x refers to the independent variable matrix, β refers to the column vector of regression coefficients, and ε refers to the error term column vector [22]. Linear models are called analysis of variance when independent variables are categorical. Analysis of variance can be considered as a special case of linear models. This method was discovered by [7].

Analysis of variance (ANOVA) is used that split an observed total variability inside a data set. Total variability is consists of into two parts: systematic factors and random factors. The systematic factors have a statistical influence on the given data set, while the random factors do not. Analysis-of-variance (ANOVA) models are interested in comparing several populations or conditions. Therefore, ANOVA is a method that compares means of results[20]. In analysis of variance, independent variables are called factors and the categories of each factor are called the level of that factor. For example, the occupation variable is a factor, and worker and manager are the levels of the occupation factor. The use of factor instead of variable emphasizes that the factors cannot be measured with cardinal values. The word variable is used for phenomena that can be measured in this way [21]. If there is only one factor in model, it is called one-way analysis of variance (one-way ANOVA), if there is two factor, it is called two-way analysis of variance (two-way ANOVA).

2.1. Two-way ANOVA

The notation for two-way ANOVA is shown in the Table 1 [7].

Table 1. Two-way ANOVA design.

	C_1	C_2	$\ldots$	C_c	
R_1	μ_{11}	μ_{12}	$\ldots$	μ_{1c}	$\mu_{1.}$
R_2	μ_{21}	μ_{22}	$\ldots$	μ_{2c}	$\mu_{2.}$
$\vdots$	$\vdots$	$\vdots$	$\ldots$	$\vdots$	$\vdots$
R_r	μ_{r1}	μ_{r2}	$\ldots$	μ_{rc}	$\mu_{r.}$
	$\mu_{.1}$	$\mu_{.2}$	$\ldots$	$\mu_{.c}$	$\mu_{..}$

- $\mu_{j.} = \frac{\sum_{k=1}^{c} \mu_{jk}}{c}$ is the marginal mean of dependent variable on row j
- $\mu_{.k} = \frac{\sum_{j=1}^{r} \mu_{jk}}{r}$ is the marginal mean of dependent variable on column k
- $\mu_{..} = \frac{\sum_{j=1}^{r} \sum_{k=1}^{c} \mu_{jk}}{rc} = \frac{\sum_{j=1}^{r} \mu_{j.}}{r} = \frac{\sum_{k=1}^{c} \mu_{.k}}{c}$ is the general mean.

R and C (for the table of meanings, "row" and "column" respectively) are factors. There are categories (levels) of R and C, respectively. Factor categories are denoted by R_j and C_k , $j = 1, 2, ..., r$ and $k = 1, 2, ..., c$. In each cell of the table, there is a cell mean (μ_{jk}) of the dependent variable (response variable) in each category combination of two factors. Using point notation shows the marginal effects of the mean.

If there is no interaction between R and C, the partial relationship between each factor and the dependent variable does not depend on whether the other factor is constant or not. The mean difference $(\mu_{jk} - \mu_{j'k})$ between rows (categories R_j and $R_{j'}$) is constant for all categories of the column. This difference $(\mu_{jk} - \mu_{j'k})$ is the same for all C_k $(k = 1, 2, ..., c)$ categories. As a result, the mean difference between rows equals the difference of marginal row means:

$$\mu_{jk} - \mu_{j'k} = \mu_{jk'} - \mu_{j'k'} = \mu_{j.} - \mu_{j'.} \quad \text{for all } j, j' \text{ and } k, k'$$

In two-way ANOVA, groups are classified across two factors, and each level of a factor appears together with each combination of the levels of other factor. This allow a researcher to investigate both the main effect of each factor on the dependent variable and the modulating of the effects of a factor by the levels of other factor. The second effects known as interactions are of central importance in many studies [25].

$$y_{ijk} = \mu + \alpha_j + \beta_k + \gamma_{jk} + \varepsilon_{ijk}, \tag{2.3}$$

where Y_{ijk} is the i^{th} observation on row j and column k; μ is the general mean of Y; α_j and β_k are main-effect parameters for row effects and column effects, respectively; γ_{jk} are interaction parameters, and ε_{ijk} are errors satisfying the usual linear-model assumptions[7]. Assumptions of the model in Equation (2.3):

(1) $E(\varepsilon_{ijk}) = 0$ for all i, j, k,
(2) $var(\varepsilon_{ijk}) = \sigma^2$,
(3) $cov(\varepsilon_{ijk}, \varepsilon_{rst}) = 0 \quad (i, j, k) \neq (r, s, t)$,

where is assumption 1;

$E(y_{ijk}) = \mu_{ij} = \mu + \alpha_j + \beta_k + \gamma_{jk}$ and with reference to this equation, Equation (2.4) is obtained.

$$y_{ijk} = \mu_{ij} + \varepsilon_{ijk}, \tag{2.4}$$

where $\mu_{ij} = E(y_{ijk})$ (ij) refers to the mean of a random observations in the $(ij)^{th}$ cell [20].

There are three effects in a two-way ANOVA; these are the main effects representing row factor and column factor and the interaction that reflects the position of these factors relative to each other. When it is observed that the measurement difference between levels of row factor is not the same at all levels of column factor, it is said relatively that there is an interaction between the two factors [25].

From a geometric perspective, each of the effects in Equation (2.3) corresponds to a subspace of the observation space, and the effects are measured by taking the projection of Y vector (dependent variable) onto these spaces [22].

2.2. Cholesky decomposition

The Cholesky decomposition is a method found by André-Louis Cholesky and used for real matrices. The Cholesky decomposition is a decomposition method ($A = LL'$) that allows the formation of a real, symmetric ($A' = A$), and positively defined ($x'Ax > 0$) matrix A by multiplying the lower triangular matrix (L) with the transpose of this lower triangular matrix (upper triangular matrix) (L') [10]. The Cholesky decomposition is also called triangular factorization. Equivalently, it is also called square root factorization [6]. For $S = LL'$ (2x2 matrix of order), L matrix is found as described below. Suppose S is the variance-covariance matrix belonging to a data set of 2 variables ($s_{12} = s_{21}$)

$$S = \begin{bmatrix} s_1^2 & s_{12} \\ s_{21} & s_2^2 \end{bmatrix}.$$

The purpose here is to make the matrix S diagonal. For this, elementary operations are performed on the rows and the columns of the matrix S. For the elementary row operation, the first row of the matrix S is added to the second row by multiplying with $-s_{21}/s_1^2$. This elementary row operation is also applied to the identity matrix I and thus the elementary matrix E_1 is obtained [10].

$$I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{matrix} \leftarrow R_1 \\ \leftarrow R_2 \end{matrix}$$

$$(-s_{21}/s_1^2)R_1 + R_2 \to R_2$$

$$E_1 = \begin{bmatrix} 1 & 0 \\ -s_{21}/s_1^2 & 1 \end{bmatrix}$$

Since the matrix S is symmetric, the elementary row operation used to obtain the matrix E_1 is applied to the columns of the identity matrix I and the elementary matrix E_2 is obtained. Also, the transpose of the elementary matrix E_1 equals the elementary matrix E_2.

$$E_1{}' = E_2 = \begin{bmatrix} 1 & -s_{21}/s_1^2 \\ 0 & 1 \end{bmatrix}$$

The matrix S is transformed into the diagonal matrix (D) using the E_1 and E_2 elementary matrices

$$\begin{bmatrix} 1 & 0 \\ -s_{21}/s_1^2 & 1 \end{bmatrix} \begin{bmatrix} s_1^2 & s_{12} \\ s_{21} & s_2^2 \end{bmatrix} \begin{bmatrix} 1 & -s_{21}/s_1^2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} s_1^2 & 0 \\ 0 & s_2^2 - s_{12}^2/s_1^2 \end{bmatrix}$$

$$E_1 S E_2 = D. \tag{2.5}$$

Using the equation of $E_1{}' = E_2$, Equation (2.5) can be written as $E_1 S E_1{}' = D$.

With the help of the equation above and by using the matrix inverses, the following equation is obtained.

$$\begin{bmatrix} 1 & 0 \\ s_{21}/s_1^2 & 1 \end{bmatrix} \begin{bmatrix} s_1^2 & 0 \\ 0 & s_2^2 - s_{12}^2/s_1^2 \end{bmatrix} \begin{bmatrix} 1 & s_{21}/s_1^2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} s_1^2 & s_{12} \\ s_{21} & s_2^2 \end{bmatrix}$$

$$E_1^{-1} D (E_1^{-1})' = S \tag{2.6}$$

Here, Equation (2.6) can be rearranged as

$$E_1^{-1} D^{1/2} D^{1/2} (E_1^{-1})' = S.$$

It is seen that the equation below is valid.

$$\begin{bmatrix} s_1 & 0 \\ 0 & \sqrt{s_2^2 - s_{12}^2/s_1^2} \end{bmatrix} \begin{bmatrix} s_1 & 0 \\ 0 & \sqrt{s_2^2 - s_{12}^2/s_1^2} \end{bmatrix} = \begin{bmatrix} s_1^2 & 0 \\ 0 & s_2^2 - s_{12}^2/s_1^2 \end{bmatrix}$$

$$D^{1/2} D^{1/2} = D$$

L and L' matrices are obtained in the way shown below:

$$E_1^{-1} D^{1/2} = L \tag{2.7}$$

$$L = \begin{bmatrix} 1 & 0 \\ s_{21}/s_1^2 & 1 \end{bmatrix} \begin{bmatrix} s_1 & 0 \\ 0 & \sqrt{s_2^2 - s_{12}^2/s_1^2} \end{bmatrix} = \begin{bmatrix} s_1 & 0 \\ s_{21}/s_1 & \sqrt{s_2^2 - s_{12}^2/s_1^2} \end{bmatrix}$$

$$D^{1/2} (E_1^{-1})' = L' \tag{2.8}$$

$$L' = \begin{bmatrix} s_1 & 0 \\ 0 & \sqrt{s_2^2 - s_{12}^2/s_1^2} \end{bmatrix} \begin{bmatrix} 1 & s_{21}/s_1^2 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} s_1 & s_{21}/s_1 \\ 0 & \sqrt{s_2^2 - s_{12}^2/s_1^2} \end{bmatrix},$$

for matrix L

$$L = \begin{bmatrix} s_1 & 0 \\ s_{21}/s_1 & \sqrt{s_2^2 - s_{12}^2/s_1^2} \end{bmatrix} = \begin{bmatrix} l_{11} & 0 \\ l_{21} & l_{22} \end{bmatrix}$$

$$\sqrt{s_2^2 - s_{12}^2/s_1^2} = s_{2.1}.$$

While l_{11} element of matrix L gives the standard deviation of the 1^{st} variable, element l_{22} will give us the conditional standard deviation ($s_{2.1}$) of the 2^{nd} variable when the 1^{st} variable is constant. Looking at this definition, it is better understood why the Cholesky decomposition is called square root factorization. The same method is also applied for matrices in different sizes. For 3-variable variance-covariance matrix ($s_{12} = s_{21}, s_{13} = s_{31}, s_{23} = s_{32}$)

$$S = \begin{bmatrix} s_1^2 & s_{12} & s_{13} \\ s_{21} & s_2^2 & s_{23} \\ s_{31} & s_{32} & s_3^2 \end{bmatrix}.$$

If the Cholesky decomposition is applied, the matrix L is found as follows:

$$L = \begin{bmatrix} s_1 & 0 & 0 \\ s_{21}/s_1 & s_{2.1} & 0 \\ s_{31}/s_1 & s_{32.1} & s_{3.12} \end{bmatrix}$$

If matrix L written in more detail

$$L = \begin{bmatrix} \sqrt{s_1^2} & 0 & 0 \\ s_{21}/\sqrt{s_1^2} & \sqrt{s_2^2 - s_{12}^2/s_1^2} & 0 \\ s_{31}/\sqrt{s_1^2} & \sqrt{s_3^2 - s_{13}^2/s_1^2} & \sqrt{s_3^2 - \left[s_{23}^2 s_1^2 + s_{13}^2 s_2^2 - 2s_{12}s_{13}s_{23}/\left(s_1^2 s_2^2 - s_{12}^2\right)\right]} \end{bmatrix}.$$

$s_{2.1}$ gives the conditional standard deviation of the 2^{nd} variable when the 1^{st} variable is constant, and $s_{32.1}$ gives the conditional covariance between the 2^{nd} and the 3^{rd} variables when the 1^{st} variable is constant. $s_{3.12}$ gives the conditional standard deviation of the 3^{rd} variable when the 1^{st} and 2^{nd} variables are constant [6]. How to calculate these values is given in detail in the matrix L above. Here, the matrix L is the matrix found as a result of the Cholesky decomposition. Depending on this, it can be said that the Cholesky decomposition relatively reduces the calculation load.

Using the Cholesky decomposition, a different formulation of the coefficients β in the linear model can also be obtained. As known, for linear model $Y = X\beta + \varepsilon$, the least-squares estimator of β is found as [9]

$$\hat{\beta} = \left(X'X\right)^{-1} X'Y. \tag{2.9}$$

As a result of the Cholesky decomposition to be applied to the matrix $X'X$, by combining Equations (2.9) and (2.10) and by using the $U = L'\hat{\beta}$ equation, Equation (2.11) is obtained [6]:

$$\begin{gathered} X'X = LL' \\ \left(X'X\right)^{-1} = \left(L'\right)^{-1} L^{-1} \end{gathered} \tag{2.10}$$

$$U = L'\hat{\beta} = L'\left(X'X\right)^{-1} X'Y = L'\left(L'\right)^{-1} L^{-1}X'Y = IL^{-1}X'Y = L^{-1}X'Y \tag{2.11}$$

The coefficients in Equation (2.11) are also called orthogonal regression coefficients. Thus, an alternative way to find $\hat{\beta}$ coefficients is obtained. With this transformation in Equation (2.11), the independent variables corresponding to $\hat{\beta}$ coefficients become orthogonal to each other. The transformation may be accomplished through the Gram-Schmidt orthonormalization of matrix X. The Gram-Schmidt will create columns of X that at each stage are orthogonal to (and uncorrelated with) all preceding columns. An alternative method to the Gram-Schmidt process is shown in Equation (2.12) [6].

$$X^* = X\left(L'\right)^{-1} = X\left(L^{-1}\right)' \tag{2.12}$$

Equation (2.12) transforms to orthonormal vectors the vectors in the matrix X.

$$\begin{gathered} X^{*\prime} = L^{-1}X' \\ X^*X^* = L^{-1}X'X\left(L'\right)^{-1} = L^{-1}LL'\left(L'\right)^{-1} = I \end{gathered}$$

Thus, it is seen that the matrix X^* is orthonormal.

$$U = X^{*\prime}Y \tag{2.13}$$

Equation (2.13) can also be used as an alternative to Equation (2.11). Unlike $\hat{\beta}$, each row of U contains the regression coefficients for the corresponding predictor variable, eliminating all preceding predictors. The values of the conditional predictor variables themselves are the columns of X^*. Thus, variables that affect the sum of squares in two-way ANOVA can be determined more easily. This also makes it easier to identify important variables.

3. Applications

In the study, the sample data set in Table 2 was used. To be able to apply the analysis of variance on the data set to be used in practice, it will be assumed that the assumptions are fulfilled ($\alpha = 0.05$)

Table 2. Sample data set.

	$Column(1) = C_1$	$Column(2) = C_2$	$Column(2) = C_2$	
$Row(1) = R_1 =$	128	166	151	$= R_1(Mean) = 150.8$
	150	178	125	
	174	187	117	
	116	153	155	
	109	195	158	
$Row(2) = R_2 =$	175	140	167	$= R_2(Mean) = 155.1$
	132	145	183	
	120	159	142	
	187	131	167	
	184	126	168	
	$C_1(Mean) = 147.5$	$C_2(Mean) = 158$	$C_3(Mean) = 153.3$	$General\ Mean(GM) = 152.9$

Table 3 was created by taking the cell means in Table 2. Table 3 contains only the means values.

Table 3. Summary data set.

	$Column(1) = C_1$	$Column(2) = C_2$	$Column(2) = C_2$	**Row Means**
$Row(1) = R_1$	$Cell_{R_1C_1}(Mean) = 135.4$	$Cell_{R_1C_2}(Mean) = 175.8$	$Cell_{R_1C_3}(Mean) = 141.2$	$R_1(Mean) = 150.8$
$Row(2) = R_2$	$Cell_{R_2C_1}(Mean) = 159.6$	$Cell_{R_2C_2}(Mean) = 140.2$	$Cell_{R_2C_3}(Mean) = 165.3$	$R_1(Mean) = 155.1$
Column Means	$C_1(Mean) = 147.5$	$C_2(Mean) = 158$	$C_3(Mean) = 153.3$	$General\ Mean(GM) = 152.9$

Hypotheses

Hypotheses can be proposed as follows:

- If α_1 and α_2 show the row means ($\alpha_1 = R_1(Mean)$ and $\alpha_2 = R_2(Mean)$),

$$H_0 : \alpha_1 = \alpha_2$$

$$H_1 : \alpha_1 \neq \alpha_2$$

- If β_1, β_2, β_3 show the column means ($\beta_1 = C_1(Mean)$, $\beta_2 = C_2(Mean)$, $\beta_3 = C_3(Mean)$),

$$H_0 : \beta_1 = \beta_2 = \beta_3$$

$$H_1 : \beta_i \neq \beta_j \quad i \neq j = 1, 2, 3$$

- If γ_{ij} shows cell mean values to test the presence of interaction between rows and columns,

$$H_0 : \gamma_{11} = \gamma_{12} = \cdots = \gamma_{23}$$

$$H_1 : \gamma_{ij} \neq \gamma_{ik} \quad i = 1, 2 \quad j, k = 1, 2, 3$$

Interaction Graph

Before starting the analysis, a preliminary information about whether there is any interaction is obtained by examining the graph.

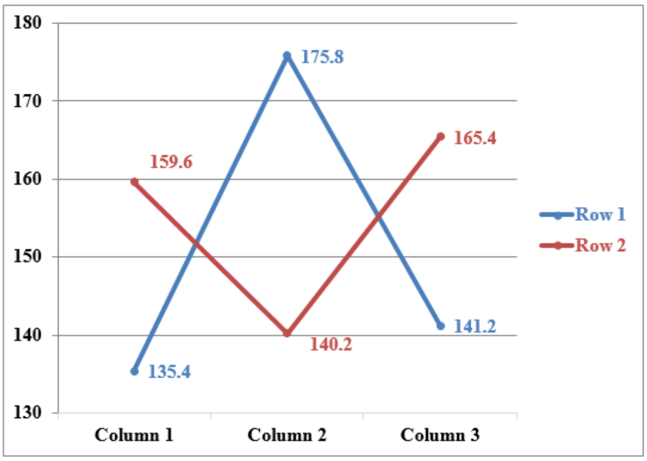

Figure 1. Plot of cell means (horizontal axis columns).

As seen in Figure 1, there is no parallelism. Differences are observed in the row means depending on the different values of the column mean. This difference is called interaction [21].

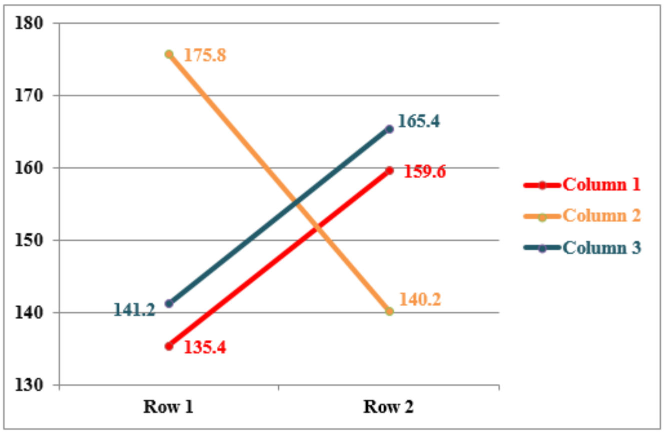

Figure 2. Plot of cell means (horizontal axis rows).

As seen in Figure 2, There is no parallel between the lines. Differences are observed in the column means depending on the different values of the row mean. This difference is called interaction [21].

The following outputs in Table 4 can be obtained using different programs. Here, the following model output was obtained using SPSS.

Table 4. SPSS results of two-way ANOVA table.

Source	Type III Sum of Squares	Degree of Freedom	Mean Square	F-Value	P-Value
Corrected Model	6649.87	5	1329.97	2.95	0.03
Intercept	701658.137	1	701658.13	1558.08	0.00
Column	553.27	2	276.63	0.61	0.55
Row	136.53	1	136.53	0.30	0.59
Interaction	5960.07	2	2980.03	6.61	0.01
Error	10808	24	450.33	2.95	0.03
Total	719116	30			
Corrected Total	17457.88	29			

When looking at the table, it is seen that F values were calculated for the sum of three squares. The null hypothesis, which assumes that there is no difference between the column means, cannot be rejected because the F-value (0.61) calculated for the sum of column squares is less than 3.41 which is the value of the F table with (2,24) degrees of freedom ($p = 0.55 > 0.05$). The null hypothesis, which assumes that there is no difference between the row means, cannot be rejected because the F value (0.30) calculated for the sum of row squares is less than 4.26 which is the value of the F table with (1,24) degrees of freedom ($p = 0.59 > 0.05$). The null hypothesis, which assumes that there is no interaction between the row and the column, is rejected because the F value (6.61) calculated for the sum of interaction squares is greater than 3.40 which is the value of the F table with (2,24) degrees of freedom ($p = 0.01 < 0.05$).

The sum of rows, columns, interactions, and error squares in the table will be found using the Cholesky decomposition. Also, it will be easier to understand what the values in the SPSS output. The advantages of using this method will be mentioned.

Table 5. Representation of the data with the design matrix.

Design Matrix X						
Constant Variable	**Row Variable**	**Column Variables**		**Interaction Variables**		**Dependent Variable**
x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	Y
1	1	1	0	1	0	128
1	1	1	0	1	0	150
1	1	1	0	1	0	174
1	1	1	0	1	0	116
1	1	1	0	1	0	109
1	0	1	0	0	0	175
1	0	1	0	0	0	132
1	0	1	0	0	0	120
1	0	1	0	0	0	187
1	0	1	0	0	0	184
1	1	0	1	0	1	166
1	1	0	1	0	1	178
1	1	0	1	0	1	187
1	1	0	1	0	1	153
1	1	0	1	0	1	195
1	0	0	1	0	0	140
1	0	0	1	0	0	145
1	0	0	1	0	0	159
1	0	0	1	0	0	131
1	0	0	1	0	0	126
1	1	0	0	0	0	151
1	1	0	0	0	0	125
1	1	0	0	0	0	117
1	1	0	0	0	0	155
1	1	0	0	0	0	158
1	0	0	0	0	0	167
1	0	0	0	0	0	183
1	0	0	0	0	0	142
1	0	0	0	0	0	167
1	0	0	0	0	0	168

3.1. Finding the design matrix and the sum of squares

Before proceeding to Cholesky method, we need to create design matrix X, dependent and independent variables. As seen in the data set, the dependent variable Y consists of the numbers in the cells. In the independent variables, there will be x_0 variable consisting of the number $1's$ to represent the mean of the dependent variable and $x_1, x_2,$ and x_3 variable that will show the factor memberships. These variables are also called dummy variables [25].

Dummy variables have two properties. First, a dummy variable has the same value for every member of a group. Second, this value is not the same for other group memberships. For number of groups g, distinguishing can be made between groups using the linearly independent variables (vector) at any number of $g-1$ [25].

$$x_1 = \begin{cases} 1, & Row(1) \\ 0, & \text{Other} \end{cases} \quad x_2 = \begin{cases} 1, & Column(1) \\ 0, & \text{Other} \end{cases} \quad x_3 = \begin{cases} 1, & Column(2) \\ 0, & \text{Other} \end{cases}$$

The values of variables (vectors) $x_1, x_2,$ and x_3 are coded as described above. x_1x_2 was formed as a result of the mutual multiplication of x_1 and x_2 variables. Likewise, x_1x_3 variable is also formed as a result of the multiplication of x_1 and x_3 variables. Thus, the design matrix X is created and it is shown in Table 5.

The data set shown in Table 5 above corresponds to Table 2 with 2 row factors and 3 column factors. Here, x_1 variable was created for the row factors, while x_2 and x_3 variables were created for the column factors. The variables x_1x_2 and x_1x_3 were created for interaction.

3.2. Cholesky decomposition

The Cholesky decomposition is applied to the matrix $X'X$ obtained after the design matrix X is created. The matrix $X'X$ formed by taking the cross product of the design matrix X obtained in Table 5 and the matrices formed as a result of the Cholesky decomposition are given below. The Cholesky decomposition can be performed by using various programs. How to perform the decomposition in MS Excel is included in our previous article in detail [23]. The matrix $X'X$ resulting from the cross product of the design matrix X is given below.

$$X'X = \begin{bmatrix} 30 & 15 & 10 & 10 & 5 & 5 \\ 15 & 15 & 5 & 5 & 5 & 5 \\ 10 & 5 & 10 & 0 & 5 & 0 \\ 10 & 5 & 0 & 10 & 0 & 5 \\ 5 & 5 & 5 & 0 & 5 & 0 \\ 5 & 5 & 0 & 5 & 0 & 5 \end{bmatrix}$$

The matrix L obtained by applying the Cholesky decomposition to matrix $X'X$ is given below.

$$L = \begin{bmatrix} 5.48 & 0 & 0 & 0 & 0 & 0 \\ 2.74 & 2.74 & 0 & 0 & 0 & 0 \\ 1.83 & 0 & 2.58 & 0 & 0 & 0 \\ 1.83 & 0 & -1.29 & 2.24 & 0 & 0 \\ 0.91 & 0.91 & 1.29 & 0 & 1.29 & 0 \\ 0.91 & 0.91 & -0.65 & 1.12 & -0.65 & 1.12 \end{bmatrix}$$

Using Equation (2.11), U coefficients are obtained.

$$U = L^{-1}X'Y = \begin{array}{c} u_0 \rightarrow \\ u_1 \rightarrow \\ u_2 \rightarrow \\ u_3 \rightarrow \\ u_4 \rightarrow \\ u_5 \rightarrow \end{array} \begin{bmatrix} 837.65 \\ -11.68 \\ -21.04 \\ 10.51 \\ -38.60 \\ 66.86 \end{bmatrix}$$

When the Cholesky decomposition is applied to the matrix $X'X$, the variables in design matrix X are transformed, so new variables are obtained as follows.

$$X^* = X\left(L'\right)^{-1} = X\left(L^{-1}\right)'$$

$$X^{*\prime} = L^{-1}X' = \begin{array}{r}x_0 \rightarrow x_0^* \\ x_1 \rightarrow x_1^* \\ x_2 \rightarrow x_2^* \\ x_3 \rightarrow x_3^* \\ x_1x_2 \rightarrow x_4^* \\ x_1x_3 \rightarrow x_5^*\end{array}$$

The properties of x^* independent variables is that they are perpendicular to each other and their lengths is $1's$. Therefore, the coefficient matrix U is called the orthogonal (orthonormal) regression coefficient matrix [6].

$$\hat{Y} = u_0x_0^* + u_1x_1^* + u_2x_2^* + u_3x_3^* + u_4x_4^* + u_5x_5^*$$

The values in Table 6 are obtained by using the coefficient matrix U elements.

Table 6. Vector lengths.

Vector	**Lengths**
$Y'Y$	719116
$\hat{Y}'\hat{Y}$	708308
$U'U$	708308
u_0^2	701658.13
u_1^2	136.53
u_2^2	442.82
u_3^2	110.45
u_4^2	1490.02
u_5^2	4470.04

The values in Table 6 can be considered as the vector lengths. $Y'Y$ refers to the vector length of the observation values, $U'U$ refers to the vector length of the prediction values, and u_i^2 $(i = 0, 1, 2, 3, 4, 5)$ refers to the vector lengths of the independent variables to which they belong. u_0^2 gives the length of the x_0 independent variable (vector), u_1^2 gives the length of the x_1 independent variable (vector), u_2^2 gives the length of the x_3 independent variable (vector), u_4^2 gives the length of the x_1x_2 independent variable (vector), u_5^2 gives the length of the x_1x_3 independent variable (vector) in design matrix X. Also, $U'U = \hat{Y}'\hat{Y}$ equation is also valid as seen below:

$$\hat{Y} = X\hat{\beta}$$
$$X'X = LL'$$
$$U = L'\hat{\beta}$$
$$\hat{Y}'\hat{Y} = \left(X\hat{\beta}\right)' X\hat{\beta} = \hat{\beta}'X'X\hat{\beta} = \hat{\beta}'LL'\hat{\beta} = U'U$$

Based on this, $\hat{Y}'\hat{Y}$ showing the prediction vector length is rewritten with the following equation.

$$\hat{Y}'\hat{Y} = u_0^2\left\|x_0^*\right\|^2 + u_1^2\left\|x_1^*\right\|^2 + u_2^2\left\|x_2^*\right\|^2 + u_3^2\left\|x_3^*\right\|^2 + u_4^2\left\|x_4^*\right\|^2 + u_5^2\left\|x_5^*\right\|^2 \quad (3.1)$$
$$\left\|x_i^*\right\|^2 = 1 \quad i = 0, 1, 2, 3, 4, 5$$

As previously explained, the properties of x^* independent variables is that they are perpendicular to each other and their lengths is $1's$.

As seen in Table 7, the sum of squares required for two-way ANOVA was obtained by using U variables. Thus, the variables (vectors) that make up the sum of squares became orthogonal (orthonormal). This gives the individual contribution of each variable in the sum of squares. The sum of column squares 553.27 is found using the $u_2^2 + u_3^2$ equation. Here, the individual contribution of u_2^2 was found to be 442.82 as seen in Table 6, while the individual contribution of u_3^2 was found as 110.45. Based on this, it can be said that the variable x_2^* corresponding to the coefficient u_2^2 is a more important variable in creating sum of column squares. The sum of interaction squares 5960.07 is found using the $u_4^2 + u_5^2$ equation. Here, the individual contribution of u_4^2 was found to be 1490.02 as seen in Table 6, while the individual contribution of u_5^2 was found as 4470.04. it can be said that the variable x_5^* corresponding to the coefficient u_5^2 is a more important variable in creating sum of interaction squares. The sum of corrected model squares 6649.87 is found using the $u_1^2 + u_2^2 + u_3^2 + u_4^2 + u_5^2$ equation. Here, the individual contribution of u_4^2 and u_5^2 are more than other as seen in Table 6. It can be said that the variables x_4^* and x_5^* corresponding to the coefficients u_4^2 and u_5^2 respectively is a more important variables in creating sum of corrected model squares. Also, it is seen that the variables x_4^* and x_5^* corresponding to the interaction are important for the two-way ANOVA model. Thus, variables that affect the sum of squares in two-way ANOVA can be determined more easily. This makes it easier to identify important variables. Also in here, it is easier to understand what the values in the SPSS output.

Table 7. Reconstructing the ANOVA table (sum of squares).

Source	**Sum of Squares**	**Reconstructing**
Corrected Model	6649.87	$u_1^2 + u_2^2 + u_3^2 + u_4^2 + u_5^2$
Intercept	701658.13	u_0^2
Column	553.27	$u_2^2 + u_3^2$
Row	136.53	u_1^2
Interaction	5960.07	$u_4^2 + u_5^2$
Error	10808	$Y'Y - U'U$
Total	719116	$Y'Y$
Corrected Total	17457.88	$Y'Y - u_0^2$

3.3. Graphical display

The corrected model is the sum of squares of cross product of the $\hat{Y}$ values, the total is the sum of cross product of Y values, and intercept corresponds to the u_0^2 value in Table 7. Other values in Table 7 are seen in Figure 3. The findings obtained in practice can also be displayed on Figure 3.

In the Figure 3, the sum of squares can be thought as the lengths of the related vectors. When looking at Figure 3, it is seen that the length of the prediction vector is formed by summing the lengths of the row, column, and interaction vectors. When calculating the prediction vector, it is seen that the interaction vector is longer. Based on this, it can be said that interaction is important. As seen in Figure 3, vectors are not perpendicular to each other (anyway, it is also impossible to be). It is unimaginable that 3 different vectors (row, column, and interaction) in the plane are perpendicular to each other. The figure was drawn for just information. By making these 3 different vectors perpendicular to each other, the Cholesky decomposition reveals the individual contributions of vectors in the formation of the prediction vector. This also makes it easier to identify important variables.

Looking at Figure 4, it is seen that the prediction vector overlaps the real vector. Because the error vector was not visible due to the point of view, it was shown in red in the figure to indicate its location. Based on this, it is also said that there is no error in the prediction space. In the Figure 4, it is seen that the prediction vector is obtained by summing 3 different (row, column, and interaction) vectors. By making these 3 different vectors orthogonal, the Cholesky decomposition reveals the individual contributions of the vectors in the formation of the prediction vector. Thus, variables that affect the sum of squares in two-way ANOVA can be determined more easily.

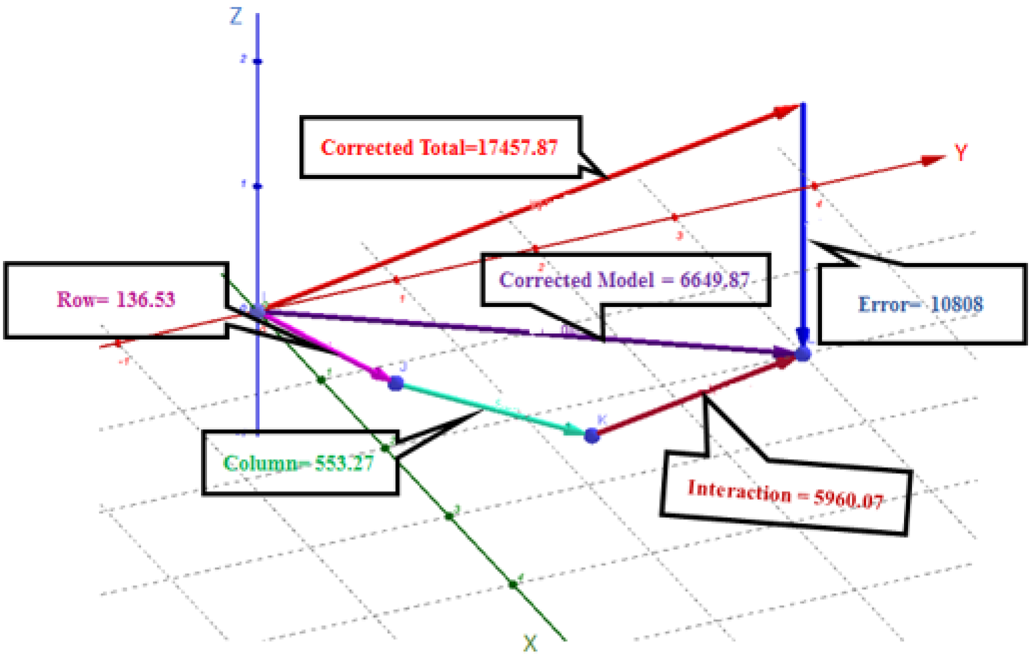

Figure 3. Representation of the decomposition of squares as vector lengths.

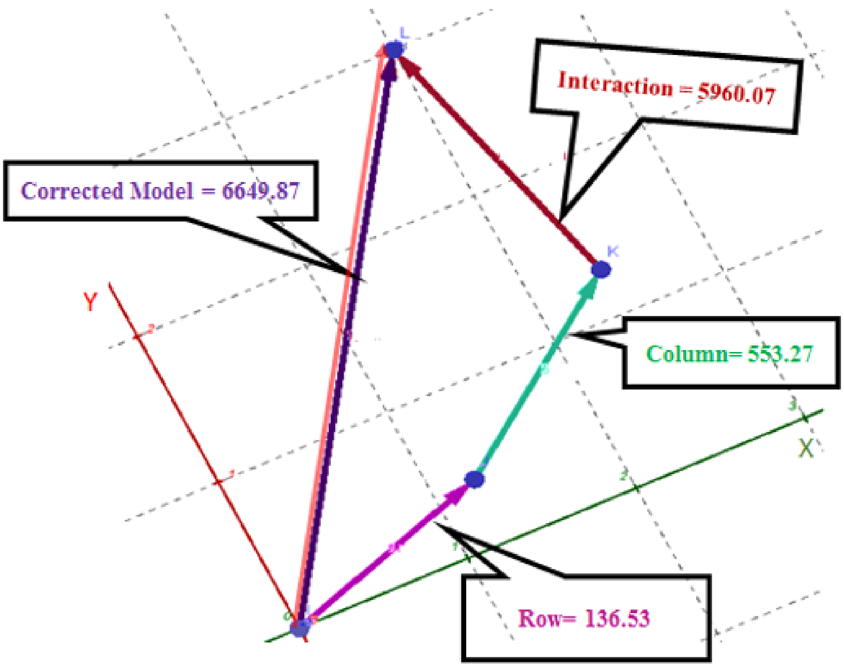

Figure 4. An overview of the prediction space.

4. Conclusion

In this study, it was tried to obtain the decomposition of squares in the two-way analysis of variance, which is a type of linear models, by using the Cholesky decomposition. Coefficients of variables in the model were predicted with the Cholesky decomposition and the sum of squares (rows columns and interactions) were obtained from these coefficients. The purpose here is to make the variables in model orthogonal. Thus, variables that affect the sum of squares in two-way ANOVA can be determined more easily. This makes it easier to identify important variables. Also, with the method used in this study, it is easier to understand what the values in the SPSS output. In the study in which graphical representations were made too, the sum of squares is represented by the vector lengths.Thus, advantages of Cholesky decomposition was shown in the formation of the prediction vector. It was determined which sums of squares in two-way ANOVA (row, column, interaction) were more important while finding the length of the prediction vector. All of these were carried out by taking advantage of the Cholesky decomposition.

In this study, the Cholesky decomposition was applied to matrix $X'X$ the cross product of the design matrix X. As another way, QR decomposition could be applied directly to the design matrix X. The fact that the average calculation load for the QR decomposition is two to four times longer than the Cholesky decomposition has been shown by [17]. The methods used in this study can be applied in multiple regression analysis, multivariate regression analysis, factorial ANOVA, MANOVA, MANCOVA, etc., and the advantages said of the Cholesky decomposition can be exploited.

Acknowledgment. The authors would like to thank the editor and anonymous reviewers for their constructive comments which led to the improvement of the paper

References

[1] D. Bradu and K.R. Gabriel, *Simultaneous statistical inference on interactions in two-way analysis of variance*, J. Amer. Statist. Assoc. **69** (346), 428-436, 1974.

[2] J.A. Cano, C. Carazo and D. Salmerón, *Objective Bayesian model selection approach to the two way analysis of variance*, Comput. Stat. **33** (1), 235-248, 2018.

[3] A. Cichocki, D. Mandic, L. De Lathauwer, G. Zhou, Q. Zhao, C. Caiafa and H.A. Phan, *Tensor decompositions for signal processing applications from two-way to multiway component analysis*, IEEE Signal Process. Mag. **32** (2), 145-163, 2015.

[4] R.B. Darlington and A.F. Hayes, *Regression Analysis and Linear Models Concepts, Applications, and Implementation*, The Guilford Press, New York, 2017.

[5] C.R. Dietrich, *Computationally efficient Cholesky factorization of a covariance matrix with block toeplitz structure*, J. Stat. Comput. Simul. **45** (3-4), 203-218, 1993.

[6] J.D. Finn, *A General Model for Multivariate Snalysis*, Rinehart and Winston, New York: Holt, 1974.

[7] J. Fox, *Applied Regression Analysis and Generalized Linear Models*, SAGE, Los Angeles, 2016.

[8] L. Freitag, S. Knecht, C. Angeli and M. Reiher, *Multireference perturbation theory with Cholesky decomposition for the density matrix renormalization group*, J. Chem. Theory Comput. **13** (2), 451-459, 2017.

[9] W.H. Greene, *Econometric Analysis*, EUA: Prentice-Hall, Upper Saddle River, 2003.

[10] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, New York, 2018.

[11] X. Kang and X. Deng, *On variable ordination of Cholesky-based estimation for a sparse covariance matrix*, Canad. J. Statist. **49** (2), 2020.

[12] P. Kohli, T.P. Garcia and M. Pourahmadi, *Modeling the Cholesky factors of covariance matrices of multivariate longitudinal data*, J. Multivariate Anal. **145**, 87-100, 2016.

[13] A. Krishnamoorthy and D. Menon, *Matrix inversion using Cholesky decomposition*, Signal Processing: Algorithms, Architectures, Arrangements, and Applications, 70-72, 2013.

[14] R.S. Krutchkoff, *Two-way fixed effects analysis of variance when the error variances may be unequal*, J. Stat. Comput. Simul. **32** (3), 177-183, 1989.

[15] K. Lee, C. Baek and M.J. Daniels, *ARMA Cholesky factor models for the covariance matrix of linear models*, Comput. Statist. Data Anal. **115**, 267-280, 2017.

[16] K. Lee and J.K. Yoo, *Bayesian Cholesky factor models in random effects covariance matrix for generalized linear mixed models*, Comput. Statist. Data Anal. **80**, 111-116, 2014.

[17] R.I.M. Lira, A.A. Trindade and V. Howle, *QR versus Cholesky: A probabilistic analysis*, Int. J. Numer. Anal. Model. **13** (1), 114-121, 2016.

[18] J.H. Maindonald, *Least squares computations based on the Cholesky decomposition of the correlation matrix*, J. Stat. Comput. Simul. **5** (4), 247-258, 1977.

[19] M. Pourahmadi, *Cholesky decompositions and estimation of a covariance matrix: orthogonality of variancecorrelation parameters*, Biometrika **94** (4), 1006-1013, 2007.

[20] A.C. Rencher and G.B. Schaalje, *Linear Models in Statistics*, Wiley-Interscience, Hoboken, 2008.

[21] S.R. Searle, *Linear Models*, Wiley, New York, 1971.

[22] I. Sumiati, F. Handoyo and S. Purwani, *Multiple linear regression using Cholesky decomposition*, World Sci. News **140**, 12-25, 2020.

[23] M. Tekin and H. Ekelik, *Linear regression approach to analysis of cariance (ANOVA) with the Cholesky decomposition and excel application*, Ankara Haci Bayram Veli University Journal of the Faculty of Economics and Administrative Sciences **20**, 58-77, 2020.

[24] T. Terao, K. Ozaki and T. Ogita, *LU-Cholesky QR algorithms for thin QR decomposition*, Parallel Comput. **92**, 102571, 2020.

[25] T.D. Wickens, *The Geometry of Multivariate Statistics*, Pychology Press, New York, 2014.

[26] G. Younis, *Practical method to solve large least squares problems using Cholesky decomposition*, Geod. Cartogr. **41** (3), 113-118, 2015.