Ankara Hacı Bayram Veli Üniversitesi
İktisadi ve İdari Bilimler Fakültesi Dergisi Özel Sayı (2020) 58-77
e-ISSN 2667-405X

Cholesky Ayrıştırması ile Varyans Analizine (ANOVA) Lineer Regresyon Yaklaşımı ve Excel Uygulaması

Mustafa TEKİN[*] Haydar EKELİK[**]

Geliş Tarihi (Received): 12.03.2020 – Kabul Tarihi (Accepted): 05.05.2020

Öz

Bu çalışmada, istatistik analizin temel testlerinden biri olan varyans analizi (ANOVA), lineer regresyon analiz yöntemleri ve Cholesky ayrışması yöntemi kullanılarak incelenmiş ve bu analiz yöntemlerine ait bazı kavramlar farklı bir bakış açısı ile tanımlanarak açıklanmıştır. Cholesky ayrışması varyans analizine ve lineer regresyon yöntemlerine farklı bir bakış açısı sunmak için kullanılmıştır. Cholesky ayrışması, varyans analizindeki ve lineer regresyondaki kavramların daha anlaşılır olmasına olanak verir. Böylelikle modelin daha iyi yorumlanması için araştırmacıya fayda sağlamaktadır. Daha çok varyans analizinde kareler ayrışımı üzerinde durulan çalışmada, varyans analizi için bulunan gruplar arası, grup içi değişim kareler toplamları, lineer regresyon yöntemiyle ve Cholesky ayrışması kullanılarak bulunmuştur. Cholesky ayrışmasının faydalarından bahsedile çalışmada Excel uygulamasıyla Cholesky ayrışmasının nasıl kullanılacağı örnek bir veri seti üzerinde gösterilmiştir.

***Anahtar Sözcükler**: Varyans Analizi(ANOVA), Cholesky Ayrışımı, Lineer Regresyon*

***JEL Kodu**: C12, C13, C02*

[*]Prof.Dr., İstanbul Üniversitesi, İktisat Fakültesi Ekonometri Bölümü, mustafatek@istanbul.edu.tr , ORCID:orcid.org/0000-0002-1169-1463
[**]Arş.Gör., İstanbul Üniversitesi, İktisat Fakültesi Ekonometri Bölümü, haydar.ekelik@istanbul.edu.tr , ORCID:orcid.org/0000-0002-0661-4164

Linear Regression Approach to Analysis of Variance (ANOVA) With the Cholesky Decomposition and Excel Application

Abscart

In this study, variance analysis (ANOVA), which is one of the basic tests of statistical analysis, was examined using linear regression analysis methods and Cholesky decomposition method, and some concepts of these analysis methods were defined and explained with a different perspective. Cholesky decomposition has been used to offer a different perspective to variance analysis and linear regression methods. Cholesky decomposition allows the concepts in analysis of variance and linear regression to be more understandable, thereby benefiting the researcher to better interpret the model. In the study, which focuses on the separation of squares in the analysis of variance, the within-group, between-group change squares totals found for variance analysis were found by using linear regression method and Cholesky decomposition. In this study, the benefits of Cholesky decomposition are shown on a sample dataset how to use Cholesky decomposition with Excel application.

Keywords: *Analysis of Variance (ANOVA), Linear Regression, Cholesky Decomposition*

JEL Code: *C12, C13, C02*

Giriş

Davranış bilimleri, tıbbi araştırmalar, pazarlama araştırmaları ve diğer bazı alanlarda yaygın bir şekilde kullanılan lineer modellerin alt başlığı olarak karşımıza çıkan varyans analizi, istatistiksel analizlerin en temel yöntemlerinden birisidir. Çeşitli alanlarda birçok uygulaması yapılmış ve günümüzde de hala yapılmaktadır. Varyans analizine farklı bir bakış açısı getirmek için yapılan çalışmada, varyans analizinin lineer modellerin bir alt dalı olduğu ve buradan yola çıkarak lineer denklem şeklinde gösterilebileceği, katsayılarının en küçük kareler yöntemiyle bulunabileceği gösterilmiştir. Daha çok denklem sistemlerinin çözümü için kullanılan Cholesky ayrışmasının da kullanılarak varyans analizinde kullanılabileceği gösterilmiştir. Bu amaçla çalışmada lineer modeller ve Cholesky ayrışması konularına değinilmiştir. Yapılan uygulamada bu yöntemlerin nasıl kullanılacağı, bu yöntemlere ait bulgular gösterilmiştir. Kullanılan yöntemlerin avantajlarından bahsedilip başka hangi analizlerde kullanılacağına değinilmiştir.

1. LİTERATÜR TARAMASI

Literatürde yapılan çalışmalara bakıldığında; Sumiati ve arkadaşları tarafından yapılan "Multiple Linear Regression Using Cholesky Decomposition" adlı çalışmada simülasyonla üretilen 5 bağımsız değişkenli hipotetik veri seti ve 5 bağımsız değişkenli gerçek veri seti (Denver Neighbohoods adlı) kullanılarak iki çoklu lineer regresyon modeli katsayılarını Choleksy ayrışması ile bulmuştur (Sumiati, Handoyo, & Purwani, 2020, s. 12-25). Ghadi Yunois tarafından yapılan "Practical method to solve large least squares problems using Cholesky decomposition" adlı çalışmada en küçük kareler yöntemini Cholesky ayrışması kullanarak geleneksel çözümlere göre daha hızlı çözümler üretildiği gösterilmiştir. (Younis, 2015, s. 113-118)

2. Veri, Yöntem Ve Bulgular

2.1.Varyans Analizi (Analysis Of Variance) (ANOVA)

Lineer modeller, davranış bilimlerinde tıbbi araştırmalarda, pazarlama araştırmaları ve diğer bazı alanlarda çok yaygın olarak kullanılmaktadır (Darlington & Hayes, 2017, s. 9). Genel olarak matematiksel formu denklem 2.1'de gösterilen bir lineer model bağımsız değişkenler olarak adlandırılan p adet değişken için *X* tahmin ediciler dizisinden bağımlı değişken olarak adlandırılan *Y* değişkenin aldığı değerleri tahmin etmede kullanılır. (Darlington & Hayes, 2017, s. 8). Lineer modellerde dikkat edilmesi gereken en önemli özellik bağımlı değişken Y'nin ölçüm sonucu elde edilmiş olmasıdır. Bazı durumlarda yaş ve gelir gibi aralıklı ölçek sonucu elde edilen ölçümlerde kullanılabilir (Darlington & Hayes, 2017, s. 10).

$$Y_j = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \ldots + \beta_p X_{pj} + \varepsilon_j \qquad (2.1)$$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1\, X_{11}\; X_{12} & \cdots & X_{1p} \\ 1\, X_{21}\; X_{22} & \cdots & X_{2p} \\ \vdots\;\; \vdots \;\;\; \vdots & \cdots & \vdots \\ 1\, X_{n1}\; X_{n2} & \cdots & X_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \qquad (2.2)$$

$$\mathbf{Y_{nx1}} = \mathbf{X_{nx(p+1)}} \boldsymbol{\beta}_{\mathbf{(p+1)x1}} + \boldsymbol{\varepsilon}_{\mathbf{nx1}}$$

Denklem 2.1 ile gösterilen matematiksel ifade matrislerle denklem 2.2'de gösterilmiştir. Lineer modellerde **Y;** bağımlı değişken sütun vektörünü, **X;** bağımsız değişken matrisini, **β;** regresyon katsayıları sütun vektörünü, **ε** ise hata terimi sütun vektörünü göstermektedir (Sumiati, Handoyo, & Purwani, 2020, s. 12-25).

Lineer modeller bağımsız değişkenlerin kategorik olması durumunda **"varyans analizi"** adını alırlar. Varyans analizi lineer modellerin özel bir hali olarak düşünülebilir. Bu yöntem 1925 yılında İngiliz istatistikçi Fisher tarafından bulunmuştur (Fox, 2016, s. 153). Varyans analizinde bağımsız değişkenlere **"faktör"**, bağımsız değişkenin aldığı değerlere de **"faktör düzeyleri"** denir. (Benian, 1998). Örnek olarak, cinsiyet faktörken cinsiyetin aldığı değerler olan kadın ve erkek değerler ise faktör düzeyleridir. Varyans analizlerinde tek bir faktör (ya da sınıf) olması durumunda tek-yönlü varyans analizi (*one-way ANOVA*) adını almaktadır. Bu çalışmada da tek yönlü varyans analizi uygulaması üzerinde durulmuştur.

2.2.Cholesky Ayrışması (Cholesky Decomposition)

Cholesky ayrışması, André-Louis Cholesky tarafından bulunan reel matrisler için kullanılan bir yöntemdir. Lineer cebirde, Cholesky ayrışması; reel, simetrik ($\mathbf{A^T = A}$), ve pozitif tanımlı ($\mathbf{x^T.A.x > 0}$) bir A matrisinin, alt üçgen matris (**L**) ve bu alt üçgen matrisin transpozesinin ($\mathbf{L^T}$) çarpımı ile oluşmasını sağlayan ($\mathbf{A=LL^T}$) bir ayrıştırma yöntemidir (Horn & Johnson, 2013, s. 442). Cholesky ayrışması **üçgensel faktorizasyon** olarak da adlandırılmaktadır. Buna eş değer olarak karekök faktorizasyon da denilmektedir (Finn, 1974, s. 37).

S’nin 2 değişkenli bir veri setine ait varyans – kovarynas matrisi olduğunu varsayalım;

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

Cholesky ayrışması sonucu bulunan alt üçgensel matris **L**;

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

L matrisinin L_{11} elemanı 1. Değişkene ait standart sapmayı verirken, L_{22} elemanı bize 1. değişken sabitken 2. değişkenin şartlı standart sapmasını verecektir. Buradan Cholesky ayrışmasına neden karekök faktorizasyon denildiği de daha iyi anlaşılmaktadır.

Cholesky faktörü, diyagonal elemanlar için önceki tüm değişkenler sabit tutulduğunda koşullu standart sapmaları gösterir. Diyagonal olmayan elemanlar, önceki değişkenler verildiğinde koşullu kovaryanslardır. Bu nedenle, koşullu değişimlerin etkisini incelemek istediğimizde Cholesky ayrışması kullanılan yöntemler arasında önemli bir yere sahiptir. (Finn, 1974, s. 40).

Cholesky ayrışması kullanılarak lineer modeldeki β katsayılarına ait farklı bir formülizasyon da elde edilebilir. Bilindiği gibi β’nın ek küçük kareler tahmincisi

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \qquad (2.3)$$

olarak bulunur (Greene, 2003, s. 21).

$\mathbf{X}'\mathbf{X}$ matrisine yapılacak olan Cholesky ayrışması sonucunda;

$$\begin{aligned} \mathbf{X}'\mathbf{X} &= \mathbf{L}\mathbf{L}' \\ (\mathbf{X}'\mathbf{X})^{-1} &= (\mathbf{L}')^{-1}\mathbf{L}^{-1} \end{aligned} \qquad (2.4)$$

Denklem 2.3 ve 2.4 birleştirilirse;

$$\hat{\boldsymbol{\beta}} = \mathbf{L}^{-1}\mathbf{X}'\mathbf{Y} \qquad (2.5)$$

Denklem 2.5 elde edilir (Finn, 1974, s. 141). Böylelikle $\hat{\boldsymbol{\beta}}$ katsayılarının bulunmasına yönelik alternatif bir yol elde edilmiş olur.

Veri ve Bulgular;

Çalışmada örnek veri seti olarak bir fabrikada aynı işi yapmakta olan 3 işçinin rassal olarak belirlenen 5 gün içinde ürettikleri parça sayıları ortalamaları arasında α=0.05 anlamlılık düzeyinde bir fark olup olmadığı araştırılmak istenmiştir. (Aytaç, 2012, s. 415). Uygulamada kullanılacak veri setine varyans analizi uygulanabilmesi için varsayımların sağlandığı kabul edilecektir.

Tablo 1. Veri Seti

	İşçiler		
Günler	**1.İşçi**	**2.İşçi**	**3.İşçi**
1	79	74	72
2	74	69	71
3	92	87	81
4	67	81	61
5	85	64	63
Gözlem	**5**	**5**	**5**
Ortalama	**79,40**	**75,00**	**69,60**
Varyans	**93,30**	**84,50**	**63,80**
Std.Sapma	**9,66**	**9,19**	**7,99**

Klasik olarak hipotezleri yazdığımızda;

μ_1, μ_2, μ_3 işçilere ait ortalamaları göstermek üzere;

$$H_0 : \mu_1 = \mu_2 = \mu_3$$
$$H_1 : \mu_i \neq \mu_j, \quad i \neq j = 1, 2, 3$$

Hipotezler bu şekilde kurulabilir.

Tablo 2. Klasik varyans analizi tablosu (ANOVA)

ANOVA

Varyans Kaynağı	SS	df	MS	F-Hesap	P-değeri	F ölçütü
Gruplar Arasında	240,93	2,00	120,47	1,50	**0,26**	3,89
Gruplar İçinde	966,40	12,00	80,53			
Toplam	1207,33	14,00				

Excel'de veri çözümleme eklentisi veya herhangi başka bir programda yapılan varyans analizlerinden **Tablo 2**'ye benzer çıktılar elde edilmektedir. Amaç bu tablodaki ***gruplar arası* ve *grup içi*** değişimleri gösteren değerleri regresyon yöntemiyle ve **Cholesky ayrışması** kullanılarak göstermek olacaktır. Tabloya bakıldığında hesaplanan F değeri, (2,12) serbestlik dereceli F tablo değerinden küçük olduğu için 3 ortalama değeri için 0,05 anlamlılık düzeyinde aralarında istatistiksel olarak bir fark olmadığı da **(p=0.26>0.05)** söylenebilir.

Tasarım matrisi ve Kareler Toplamının Bulunması

Lineer regresyon yöntemine geçmeden önce bağımlı ve bağımsız değişkenleri oluşturmamız gerekmektedir. Bağımlı değişken (***Y***) veri setinde de görüldüğü gibi üretilen parça sayısı olacaktır. Bağımsız değişkenler öncelikle **bağımlı değişkenin ortalamasını temsil etmek üzere 1 sayısından oluşan X_0 bağımsız değişkeni** ile grup üyeliklerini gösterecek olan $\mathbf{X_1}$ ve $\mathbf{X_2}$ bağımsız değişkenleri olacaktır.

$$x_1 = \begin{cases} 1, & \text{1.gruba aitse} \\ 0, & \text{diğer grublarda ise} \end{cases} \qquad x_2 = \begin{cases} 1, & \text{2.gruba aitse} \\ 0, & \text{diğer grublarda ise} \end{cases}$$

$\mathbf{X_1}$ ve $\mathbf{X_2}$ değişkenlerinin (vektörlerinin) aldığı değerler yukarıda açıklandığı gibi kodlanıp **Tablo 3.**' te gösterilmiştir**.**

Tablo 3. Tasarım Matrisi ile verilerin gösterilmesi

X Veri(Tasarım) Matrisi			Bağımlı Değişken	
x_0	**x_1**	**x_2**	**Y**	
1	1	0	**79**	
1	1	0	**74**	
1	1	0	**92**	
1	1	0	**67**	**1.grup ort**
1	1	0	**85**	**79.4**
1	0	1	**74**	
1	0	1	**69**	
1	0	1	**87**	
1	0	1	**81**	**2.grup ort**
1	0	1	**64**	**75**
1	0	0	**72**	
1	0	0	**71**	
1	0	0	**81**	
1	0	0	**61**	**3.grup ort**
1	0	0	**63**	**69,60**
		Genel Ortalama=	**74,66**	

Lineer regresyon yöntemine ait katsayılar $\hat{\boldsymbol{\beta}} = (\mathbf{X'X})^{-1}\,\mathbf{X'Y}$ formülü ile

Tablo 4. Katsayılar ve anlamları

	$\hat{\beta}$ **Katsayılar Vektörü**	**Anlamı**		
		Açıklaması	**Sembolik Gösterimi**	**Değeri**
X_0	**69,60**	**3.grup Ortalaması**	$\bar{X}_3$	**= 69,60**
X_1	**9,80**	**1.grup - 3.grup Ortalaması**	$\bar{X}_1 - \bar{X}_3$	**79,4 – 69,6 = 9,80**
X_2	**5,40**	**2.grup - 3.grup Ortalaması**	$\bar{X}_2 - \bar{X}_3$	**75,0 – 69,6 = 5,40**

bulunur. **X** katsayılar vektörü kullanılarak bağımlı değişkenin tahmin değerleri $\mathbf{Y_{nx1}} = \mathbf{X_{nx(p+1)}}\hat{\boldsymbol{\beta}}_{\mathbf{(p+1)x1}}$ ve bu tahminlere bağlı olarak hata($\varepsilon = Y - \hat{Y}$) değerleri bulunur. Bulunan değerler Tablo 5' de gösterilmiştir.

Tablo 5. Katsayılar, tahmin değerleri ve hatanın bulunması

	Y Tahmin ($\hat{Y}$)	Hata ($\varepsilon = Y - \hat{Y}$)		Y Tahmin -Yort ($\hat{Y} - \bar{Y}$)	Hata-Hata ort ($\varepsilon - \bar{\varepsilon}$)
	79,40	-0,40		4,73	-0,40
	79,40	-5,40		4,73	-5,40
	79,40	12,60		4,73	12,60
	79,40	-12,40		4,73	-12,40
	79,40	5,60		4,73	5,60
	75,00	-1,00		0,33	-1,00
	75,00	-6,00		0,33	-6,00
	75,00	12,00		0,33	12,00
	75,00	6,00		0,33	6,00
	75,00	-11,00		0,33	-11,00
	69,60	2,40		-5,07	2,40
	69,60	1,40		-5,07	1,40
	69,60	11,40		-5,07	11,40
	69,60	-8,60		-5,07	-8,60
	69,60	-6,60		-5,07	-6,60
Ortalama =	74,66	**0**	**Kareler Toplamı=**	**240,93**	**966,4**

Tablo 6. Regresyon Yöntemiyle elde edilen ANOVA Tablosu

Değişkenlik Kaynağı	Kareler Ayrışımı	Serbestlik Derecesi	Ortalama Kare	F Değeri	F tablo	P-Value
Gruplar Arası (Regresyon)	**240,93**	**2,00**	**120,47**	**1,50**	**3,89**	**0,26**
Grup İçi (Hata)	**966,40**	**12,00**	**80,53**			
Toplam	**1207,33**	**14,00**	**86,24**			

Tablo 5 ve Tablo 6'deki değerlere bakıldığında ANOVA tablosunda bulunan **gruplar arası değişimin** lineer regresyon yöntemiyle elde edilen tahmin değerlerinin ortalamadan sapmalarının karesi oluğu görülmektedir. ANOVA tablosundaki **grup içi değişimin lineer regresyon yöntemiyle elde edilen hatanın ortalamadan sapmalarının karesi olduğu** görülmektedir. Toplam değişiminde bağımlı değişkenin gerçek değerlerinin genel ortalamadan sapmalarının karesi olduğu da görülmektedir.

Cholesky Ayrışması

Tasarım matrisi oluşturulduktan sonra elde edilen **X′X** matrisine Cholesky ayrışması uygulanır. Tablo 3'de elde edilen **X** tasarım matrisine ait **X′X** matrisi ve Cholesky ayrışması sonucu oluşan matrisler aşağıdaki yer almaktadır.

X′X matrisi

15	**5**	**5**
5	**5**	**0**
5	**0**	**5**

X′X matrisine Cholesky ayrışması adım adım uygulanırsa

1.aşama

Matrisi üst üçgensel hale getiren elementer işlemler için gerekli olan ilk matris birim matris üzerinden hareketle elde edilen E_1 matrisi; yapılan elemanter işlemin ters işaretlisi olarak alınarak aşağıda verilmiştir.

E_1 matrisi

1	**0**	**0**
-5/15	**1**	**0**
-5/15	**0**	**1**

E1 matrisi ile **X′X** matrisinin çarpımı aşağıdaki gibi olacaktır.

E_1.(X′X) matrisi

15	**5**	**5**
0	**10/3**	**-5/3**
0	**-5/3**	**10/3**

İkinci matris (E_2) birim matris üzerinden hareketle elde edilen E_1 matrisi üzerinde elemanter işlemin ters işaretlisi olarak aşağıda verilmiştir

E_2 matrisi

1	**0**	**0**
0	**1**	**0**
0	**(-5/3)/(10/3) = -0.5**	**1**

E_2 matrisi ile **E_1. X′X** matrisinin çarpımı aşağıdaki gibi olacaktır.

V=E_2E_1.(X′X) matrisi

15	**5**	**5**
0	**10/3**	**-5/3**
0	**0**	**5/2**

Buradan da görüldüğü gibi (**X′X**) matrisi Üst üçgensel hale getirilmiştir. Bu sonuca tek bir aşamada ulaşılması istenirse **E_2.E_1=E** matrisi aşağıdaki gibi olacaktır.

E_2			E_1				$E_2.E_1$=E matrisi		
1	0	0	1	0	0		1	0	0
0	1	0	-5/15	1	0	=	-5/15	1	0
0	-0.5	1	-5/15	0	1		-1/2	1/2	1

E matrisi ile **X′X** matrisi çarpılırsa üç üçgensel matris tek aşamada elde edilir

E= $E_2.E_1$			X′X				V=E.(X′X)		
1	0	0	15	5	5		15	5	5
-5/15	1	0	5	5	0	=	0	10/3	-5/3
-1/2	1/2	1	5	0	5		0	0	5/2

V matrisini elde ederken oluşturulan **E** matrisinin transpozesi ile V matrisi sağdan çarpılarak **D=Diag(V)** elde edilir

V			E′				D=V.E′=Diag(V)		
15	5	5	1	-5/15	-1/2		15	0	0
0	10/3	-5/3	0	1	1/2	=	0	10/3	0
0	0	5/2	0	0	1		0	0	0

Yukarıda yapılan işlemler matrislerle ifade edilirse;

$$\mathbf{E(X'X) = V}$$
$$\mathbf{VE' = D = Diag(V)}$$
$$\mathbf{E(X'X)E' = D}$$

Olacaktır. Buradan **X′X** matrisi yalnız bırakılırsa;

$$\mathbf{X'X = E^{-1} D E'^{-1}}$$

Elde edilir. Bu eşitliği basitleştirmek adına ;

$$\mathbf{A = E^{-1}}$$
$$\mathbf{B = E'^{-1}}$$

İfadeleri kullanılırsa denklem 3.6 elde edilir.

$$\mathbf{X'X = A \cdot D \cdot B} \qquad (2.6)$$

Buradan yola çıkarak;

$$\mathbf{X'X} = \mathbf{A} \cdot \mathbf{D}^{1/2} \cdot \mathbf{D}^{1/2} \cdot \mathbf{B}$$
$$\mathbf{A} \cdot \mathbf{D}^{1/2} = \mathbf{L} \qquad (2.7)$$
$$\mathbf{D}^{1/2} \cdot \mathbf{B} = \mathbf{L'}$$

Dönüşümleri yapılarak **L** ve **L'** matrisleri elde edilebilir.

$$\mathbf{X'X} = \mathbf{LL'}$$

$\mathbf{L} = \mathbf{A} \cdot \mathbf{D}^{1/2}$

3,87	**0**	**0**
1,29	**1,83**	**0**
1,29	**-0,91**	**1,58**

$\mathbf{L'} = \mathbf{D}^{1/2} \cdot \mathbf{B}$

3,87	**1,29**	**1,29**
0	**1,83**	**-0,91**
0	**0**	**1,58**

Yukarıda yapılan Cholesky ayrışması için kullanılan işlemleri kısaca açıklamak gerekirse; **X′X** matrisi önce **V** ve sonrasında **D** matrisine dönüştüren elementer işlemler yapılır. Bu işlemler matrislerle gösterilirse; Burada **A** ve **B** matrisleri yapılan elementer işlemlerin matrislerle gösterimidir. **A** matrisi **X′X** matrisinden **V** matrisini elde etmek için yapılan elementer işlemleri göstermektedir. **B** matrisi ise **V** matrisini **D** matrisine dönüştürmek için yapılan elementer işlemleri gösteren matrisleridir. **X′X** matrisi simetrik bir matris olduğundan **A** ve **B** matrisleri bir birlerinin transpozesi olurlar**.**

L ve **L′** matrisleri çeşitli program ve eklentiler kullanılarak bulunabilir. Bu çalışmada **X′X** matrisi önce üst üçgensel matrise indirgenmiş sonrasında üst üçgensel matrise çevirmek için yapılan işlemler sütunlara da uygulanmış ve sonucunda **L** ve **L'** matrisleri bulunmuştur.

Yukarıda nasıl bulunduğu izah edilen **L** matrisini kullanarak **X** katsayılarının elde edilmesi denklem 2.5'deki formülün uygulanmasıyla $\hat{\boldsymbol{\beta}} = \mathbf{L}^{-1}\mathbf{X'Y}$ yeni regresyon katsayıları elde edilir. Bu katsayılar **U** matrisi ile de gösterilmektedir. Bu yöntemle elde edilen katsayıların özelliği ilgili oldukları bağımsız değişkenlerin uzunluklarının (boylarının) 1 olması ve aynı zamanda birbirlerine dik olmalarıdır. Kısaca bağımsız değişkenler ortonormal bir baz oluşturmaktadır. Cholesky ayrışması ile elde edilen regresyon katsayıları ortonormal olan vektörlere (bağımsız değişkenlere) ait olan katsayılardır.

$$\hat{Y}'\hat{Y} = \beta_0^2 \left\| X_0 \right\|^2 + \beta_1^2 \left\| X_1 \right\|^2 + \beta_2^2 \left\| X_2 \right\|^2 \qquad (2.8)$$

Denklem 2.8’de görüldüğü gibi lineer regresyon yöntemi sonucu oluşan bağımlı değişkenin tahmin değerleri($\hat{Y}$) ile transpozesinin çarpımı yani tahmin vektörünün uzunluğu, 3 vektörün (bağımsız değişkenin) uzunlukları ile onlara karşılık gelen katsayıların karelerinin çarpılıp toplamlarıyla elde edilmesine dayanmaktadır. Bu sistemdeki **bağımsız değişkenlerin (vektörlerin) uzunlukları 1’dir. Yani birim vektörlerdir. Aynı zamanda bu vektörler birbirleri üzerine izdüşüm alınarak birbirlerine dik hale getirilmektedir. Böylelikle bağımlı değişkenin tahmin edilmesinde her bir bağımsız değişkenin bireysel katkısı daha net görülmektedir**. **Cholesky yönteminin kullanılması bu avantajı sağlamaktadır.**

Tablo 7. Kare ayrışımlarının yeniden oluşturulması

	$\mathbf{U = \hat{\beta} = L^{-1}X'Y}$
u_1=	**289,182**
u_2=	**12,962**
u_3=	**8,538**

Y'Y=	**84834,00**
U'U=$\hat{Y}'\hat{Y}$ =	**83867,6**
u_1^2=	**83626,7**
u_2^2=	**168,033**
u_3^2=	**72,9**

u_2^2 +u_3^2=	**240,93**	**=Gruplar Arası**
Y'Y - U'U=	**966,40**	**=Grup İçi (Hata)**
Y'Y -u_1^2 =	**1207,33**	**=Toplam**

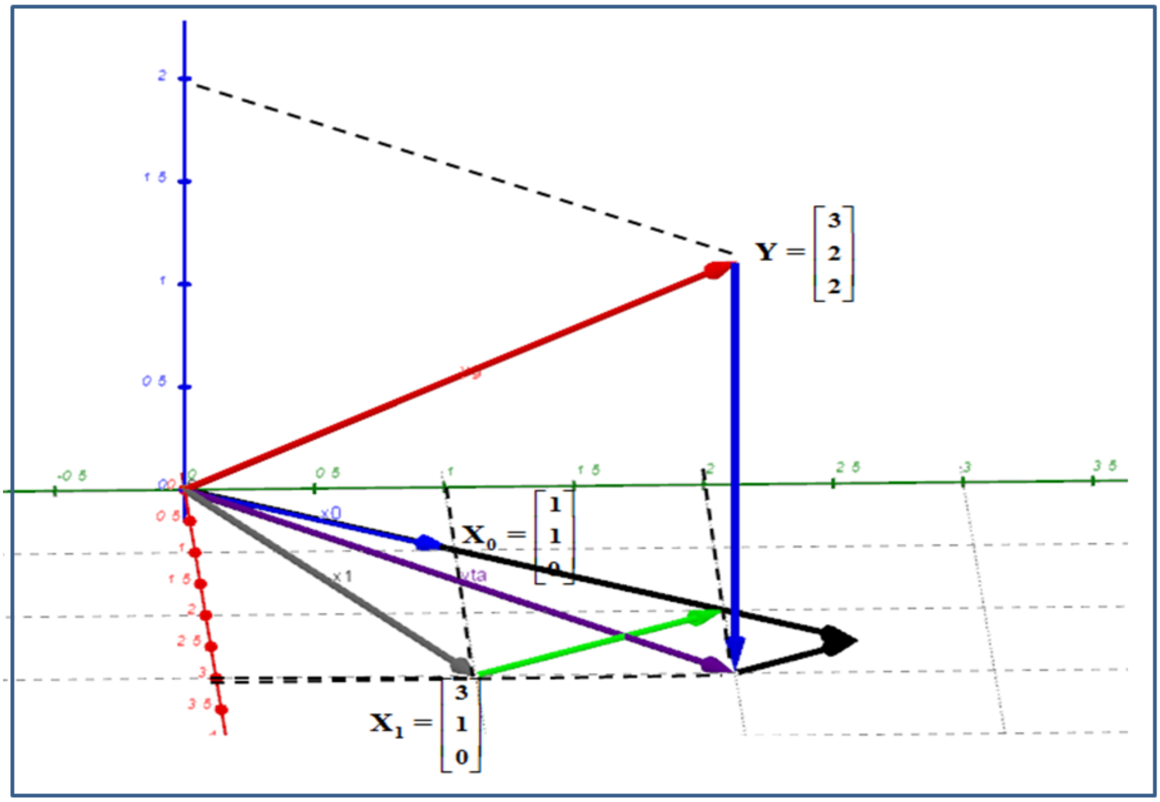
Yukarıda Tablo 7’den de görüldüğü gibi kareler ayrışımı *Gruplar Arası (regresyon*) ve *Grup içi (Hata)* kare ayrışımları **Cholesky ayrışması kullanılarak daha önce bulunan değerlerle aynı bulunmuştur.** Cholesky ayrışmasının kullanılmasının avantajı bağımsız değişkenlerden oluşan tahmin uzayındaki vektörlerin (X_0,X_1,X_2) birbirine dik olmaları ve aynı zamanda uzunluklarının bir olmasıdır. Bir başka deyişle bu vektörlerden ortonormal bir baz oluşmasını sağlamaktır. Tablodan da görüldüğü gibi **u_1^2 ,u_2^2 ,u_3^2** değerleri denklem 2.8’deki $\beta_0^2, \beta_1^2, \beta_2^2$ regresyon katsayılarının karesine denk gelmektedir. **u_1^2 ,u_2^2 ,u_3^2** bu değerlerin toplamı ise **U'U=**$\hat{Y}'\hat{Y}$ değerini başka bir değişle tahmin vektörünün uzunluğunun karesini vermektedir. u_2^2,u_3^2 değerlerinin toplamı **gruplar arası** kare toplamını verirken, **Y'Y - U'U** farkı ise **grup içi** kareler toplamını vermektedir.

Yukarıda anlatıları daha iyi anlamak ve daha rahat görselleştirmek adına tek bağımsız değişkenli bir örnek ele alınırsa aşağıdaki grafik çizilebilir.

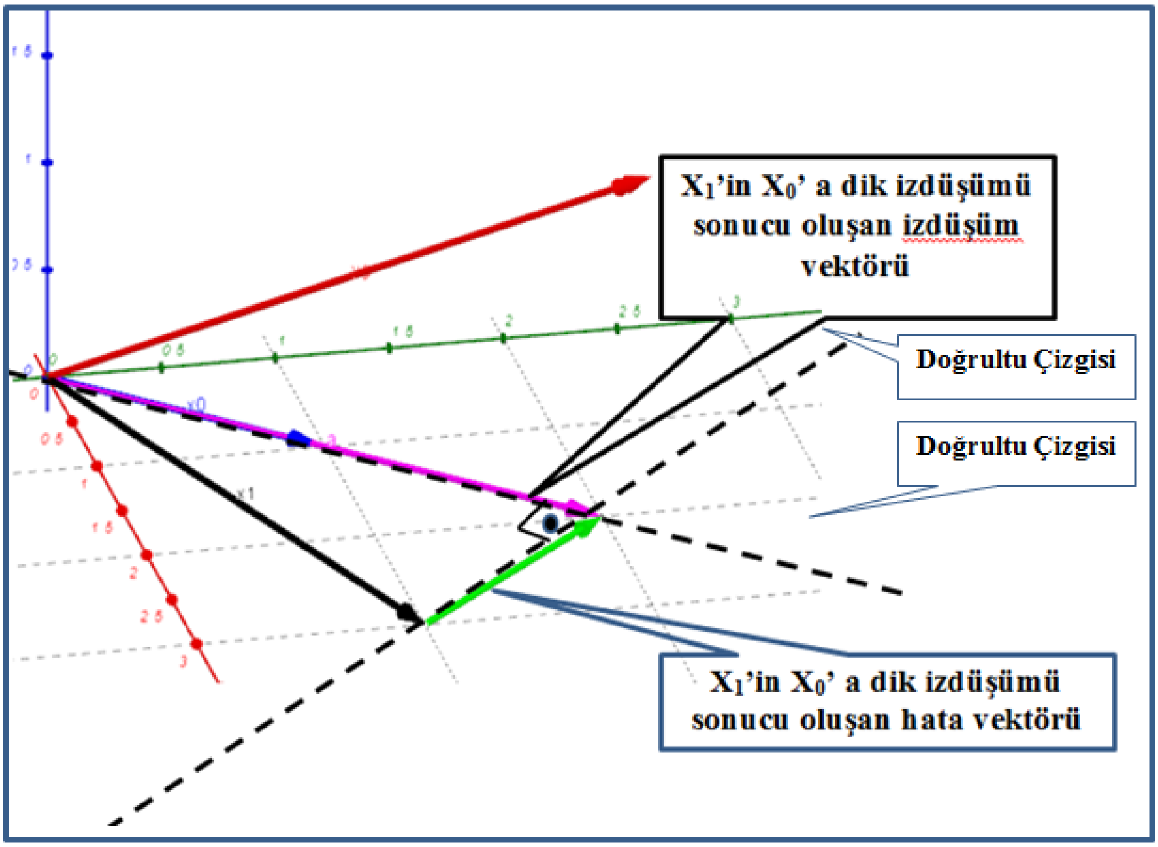
Tablo 8: Örnek veri seti

X_0	X_1	Y
1	3	3
1	1	2
0	0	2

Şekil 1. Ele alınan vektörlerin üç boyutlu uzayda gösterilmesi



Şekil 2. Ele alınan vektörlerin iz düşümlerinin üç boyutlu uzayda gösterilmesi

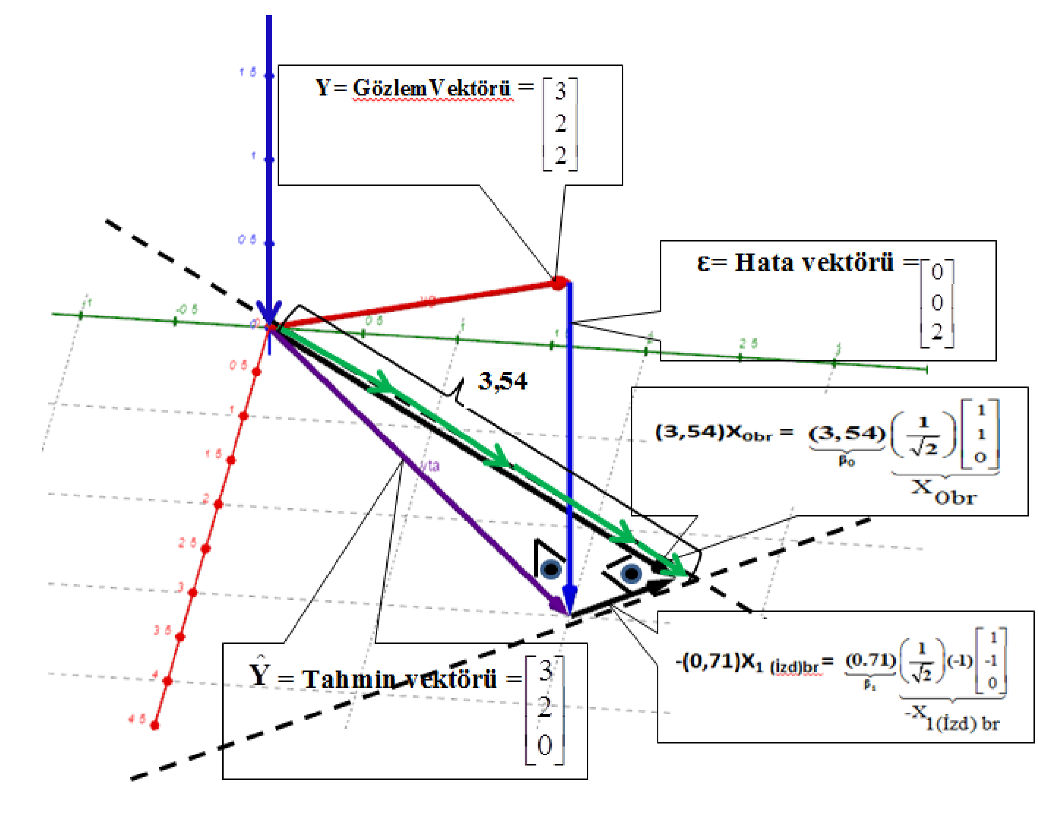


Şekil 2'de bağımsız değişkenlere ait iz düşümler gösterilmiştir. Burada X_1 vektörü X_0 üzerine iz düşümü alınarak, iz düşüm vektörü ve iz düşüm sonucu oluşan hata vektörü gösterilmiştir. Cholesky ayrışması sonucu bulunan **L** matrisi, iz düşümleri oluşturan vektörlerin uzunluklarını vermektedir. Ele alınan örnek için ilgili vektörler aşağıda tabloda verilmiştir. Bu tablodaki vektörler üzerinden **X′X** ve bunun Cholesky ayrıştırması aşağıdaki gibidir.

X′X			**L**		**L′**	
2	**4**		**1.41**	**0**	**1.41**	**2.82**
4	**2**	=	**2.82**	**1.41**	**0**	**1.41**

Şekilde verilen örneğe ait **X′X** matrisine Cholesky ayrışması uygulandıktan sonra bulunan **L** matrisinin L_{11} elemanı **Xo vektörünün uzunluğunu**, L_{21} elemanı **iz düşüm vektör uzunluğunu**, L_{22} elemanı ise **iz düşüm sonucu oluşan hata vektör uzunluğuna** karşılık gelmektedir.

Şekil 3. Tahmin vektörünün dik bileşenler cinsinden gösterilmesi



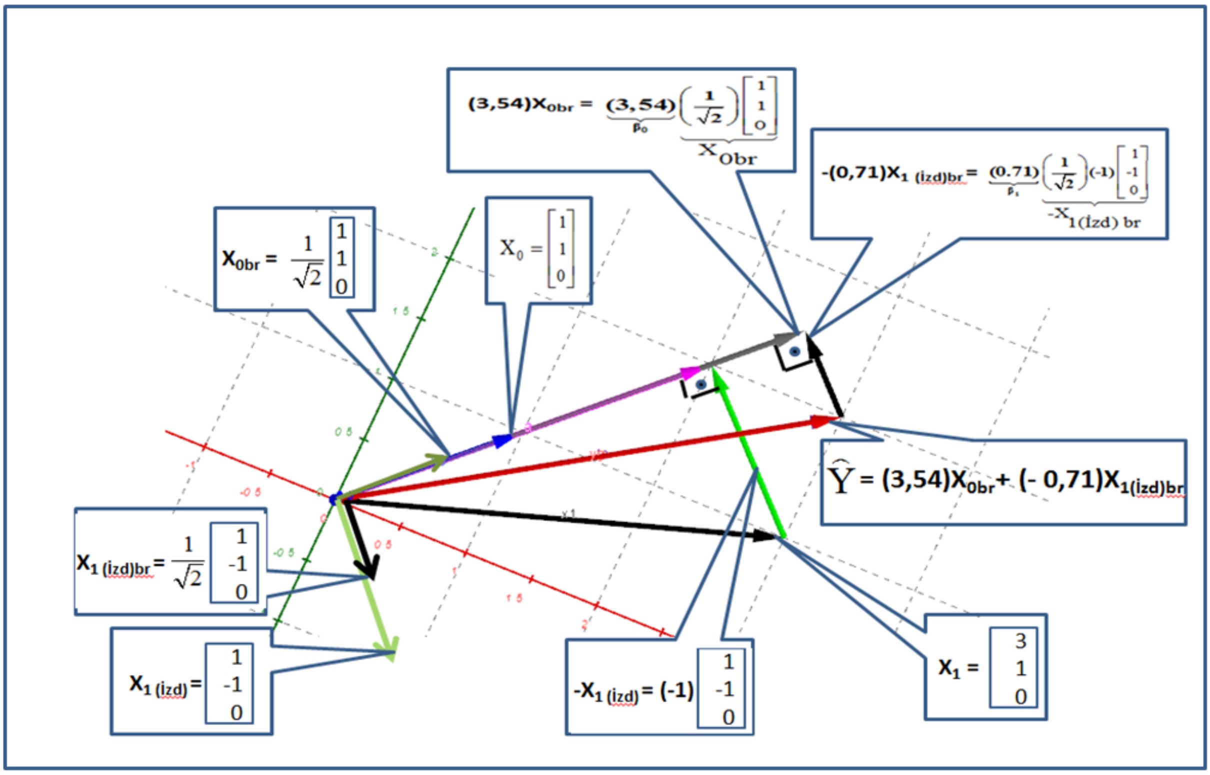
$U'U = \hat{Y}'\hat{Y}$ = **Tahmin vektörünün Boyunu vermektedir.**

$Y'Y$ = **Gözlem Vektörünün Boyunu vermektedir.**

$Y'Y - U'U$ = **Hata vektörünün boyunu vermektedir.**

Şekil 4. 3 boyutlu uzaya kuş bakışı

TAHMİN UZAYINA KUŞ BAKIŞI



Şekil 4'de görüldüğü gibi üç boyutlu uzaya kuş bakışı bakılarak, vektörlerin dik iz düşümleri, birbirlerine göre dik olmaları ve bu dik vektörlerin tahmin vektörünü nasıl oluşturdukları görülmektedir.

Sonuç ve Öneriler

Davranış bilimleri, tıbbi araştırmalar, pazarlama araştırmaları ve diğer bazı alanlarda yaygın bir şekilde kullanılan lineer modellerin alt başlığı olarak karşımıza çıkan varyans analizi, çokça kullanılan istatistiksel bir yöntemdir. Çeşitli alanlarda birçok uygulaması yapılmıştır. Bu çalışmada varyans analizi lineer regresyon yöntemiyle nasıl yapılacağı ve buradan hareketle Cholesky ayrışmasının da nasıl kullanılacağı üzerinde durulmuştur. Varyans analizlerinde kullanılan *gruplar arası* ve *grup içi* kareler toplamının lineer regresyon yöntemiyle ve Cholesky ayrışması kullanılarak nasıl hesaplanacağı gösterilmiştir. Lineer regresyon yönteminde gruplar arası kare toplamının regresyon kare toplamı olduğu, grup içi kareler toplamının ise hata kare toplamı olduğu gösterilmiştir. Cholesky ayrışmasıyla bağımsız değişkenlerden oluşan tahmin uzayının ortonormal bir baz oluşmasını sağlanmış ve böylelikle bağımsız değişkenlerin(açıklayıcı değişkenlerin) birbirlerinden bağımsız hale gelmesi sağlanarak, bağımlı değişkenin tahmin edilmesinde her bir bağımsız değişkenin tekil katkısı daha net görülebilmiştir. Yine Choleksy ayrışmasıyla grupla arası ve grup içi kareler toplamı bulunmuştur. Yapılan bu çalışma ile varyans analizine farklı bir bakış açısı getirilmiştir. Kullanılan bu yöntemler varyans analizini farklı versiyonlarında (iki-yönlü varyans analizi, faktöriyel varyans analizi) özellikle etkileşimin olduğu varyans analizlerinde ve çok değişkenli varyans analizlerinde (MANOVA) kullanılarak lineer regresyon ve Choleksy yöntemlerinin faydaları daha net bir şekilde görülebilir.

Kaynakça

Aytaç, M. (2012). *Matematiksel İstatistik* (4.Baskı b.). Bursa: Ezgi Kitabevi Yayınları.

Benian, T. (1998). *Varyans Analizinin Önşartları ve Transformansyonlar.* Doktora Tezi: Ankara Üniversitesi Fen Bilimleri Enstitüsü .

Darlington, R., & Hayes, A. (2017). *Regression Analysis and Linear Models Concepts, Applications, and Implementation.* New York: The Guilford Press.

Finn, J. (1974). *A General Model for Multivariate Analysis.* International Series in Decision Processes.

Fox, J. (2016). *Applied Regression Analysis and Generalized Linear Models* (3. Edition b.). California: SAGE Publications.

Greene, W. H. (2003). *Econometric Analysis* (5.Edition b.). New Jersey: Prentice Hall.

Horn, R., & Johnson, C. (2013). *Matrix Analysis.* New York: Cambridge University Press.

Sumiati, I., Handoyo, F., & Purwani, S. (2020). Multiple Linear Regression Using Cholesky Decomposition. *World Scientific News 140*, 12-25. http://www.worldscientificnews.com/article-in-press/2020-2/140-2020/ adresinden alındı

Younis, G. (2015). Practical Method to Solve Large Least Squares Problems Using Cholesky. *Geodesy and Cartography, 3*(41), 113-118. doi:10.3846/20296991.2015.1086118