



Çoklu Doğrusal Bağını Varlığında En Küçük Karelere Alternatif Yaklaşım: Ridge Regresyon

Mustafa Çağatay BÜYÜKUYSAL¹, İbrahim İlker ÖZ²

ÖZ

Bu çalışmada çok değişkenli regresyon modellerinde çoklu doğrusal bağını varlığında en küçük kareler yöntemine alternatif olarak önerilen Ridge regresyon yöntemi üzerinde çalışılmış, en küçük kareler ve Ridge regresyon ile elde edilen regresyon katsayıları karşılaştırılmıştır. 2015 yılı Nisan ve Mayıs aylarında Bülent Ecevit Üniversitesi Tıp Fakültesi Hastanesi Radyoloji Anabilim Dalı'na abdomen bilgisayarlı tomografik incelemesi için başvuran hastalardan elde edilen subkutan yağ doku, viseral yağ doku ve pankreas hacimlerinin beden kitle indeksi ile ilişkisi incelenmiştir. Yapılan analizler sonucunda, Ridge regresyon analiziyle elde edilen regresyon katsayıları, kuramsal beklentilere cevap verirken, en küçük kareler yöntemi ile elde edilen regresyon katsayıları beklenenden uzak sonuçlar vermiştir. Sonuç olarak, çoklu doğrusal bağını halinde Ridge regresyonun en küçük kareler yöntemine göre daha doğru sonuçlar verdiği görülmüştür. Bu sebeple, çoklu doğrusal bağını varlığında Ridge regresyon analizinin kullanılması önerilmiştir.

Anahtar Kelimeler: Ridge regresyon; çoklu doğrusal bağını; en küçük kareler.

An Alternative Method to Least Squares in Presence of Multicollinearity: Ridge Regression

ABSTRACT

At this study, we studied on an alternative method suggested to least squares called "Ridge Regression" which is used in presence of multicollinearity between independent variables. Regression coefficients from least squares and Ridge regression are compared. The data are collected from the patients whom applied to Bülent Ecevit University, Faculty of Medicine, Radiology Department for abdomen computerized tomography from April 2015 to May 2015 and body mass index, subcutaneous adipose tissue, viseral adipose tissue and pancreas volumes are measured. The relationships with body mass index and subcutaneous adipose tissue, viseral adipose tissue and pancreas volume are evaluated with both techniques. According to results, Ridge regression coefficients are very close to theoretical expectations where least square coefficients are so far from expectations. As a result, in presence of multicollinearity, it is seen that Ridge regression gives more appropriate results than least squares. Therefore, at problem of multicollinearity we suggest using Ridge regression analysis.

Keywords: Ridge regression; multicollinearity; least squares.

GİRİŞ

Çoklu doğrusal regresyonda temel amaç bağımsız değişkenler yardımıyla bağımlı değişken değerini kestirmek, bağımsız değişkenlerden hangisi ya da hangilerinin bağımlı değişkeni daha çok etkilediğini belirlemek ve diğer değişkenlerin varlığında katsayı kestirimlerinde bulunabilmektir (1). Çoklu doğrusal regresyon modellerinde parametre tahmini en küçük kareler yöntemi ile yapılır. En küçük kareler yöntemi, hatalar kareler toplamını minimum yapan parametre tahmincilerinin seçildiği bir yöntemdir. Ancak en küçük kareler yönteminin uygulanabilmesi için bir takım varsayımların yerine getirilmiş olması gerekmektedir. Bu varsayımlardan biri ya da bir kaç gerçekleşmediğinde elde edilen kestirim sonuçlarının yanlı çıkması kaçınılmazdır. Hata terimlerinin varyansının sabit olması ve birbirinden bağımsız olmaları, hata terimlerinin beklenen değerinin "0" olması ve bağımsız değişkenler arasında bir ilişki olmaması regresyon analizinin varsayımlarıdır ve bu varsayımların gerçekleşmesi regresyon analizinde parametre kestirimlerinin yansız ve doğru bir şekilde elde edilebilmesini sağlar. Ancak sağlık alanında elde edilen verilerde bu varsayımlardan sapmalar görülmektedir (2).

¹ Bülent Ecevit Üniversitesi, Tıp Fakültesi, Biyoistatistik AD

² Bülent Ecevit Üniversitesi, Tıp Fakültesi, Radyoloji AD

Correspondence: Mustafa Çağatay BÜYÜKUYSAL, e-posta: cbuyukuysal@gmail.com

Geliş Tarihi / Received: 02.11.2015 Kabul Tarihi / Accepted: 05.03.2016

Bu varsayımlardan sapmaların ilki “değişen varyans”tır (2). Doğrusal regresyon modelinde hata terimlerinin $N(0, \sigma^2)$ ile normal dağıldığı varsayılmaktadır ve buna sabit varyans (homoscedasticity) denir.

$$\text{Var}(u_i) = \sigma^2 \quad [1]$$

Ancak hata terimleri sabit bir varyansa sabit olmazsa bu duruma değişen varyans denir. Bu durumda hataların varyansı aşağıdaki gibi olur.

$$\text{Var}(u_i) = \sigma_i^2 \quad [2]$$

Varsayımlardan sapmaların ikincisi, gözlemler arası ardışık bağımlılığın (otokorelasyon) bulunmamasıdır (2). Otokorelasyon, çoklu regresyon modelindeki hata terimlerinin birbirinden bağımsız olduğu varsayımından olan sapmayı belirtir. Eğer hata terimleri birbirleriyle ilişkili ise, yani otokorelasyon varsa, $E(\varepsilon_i \varepsilon_j) \neq 0$ olacaktır (2). Bir diğer varsayımdan sapma açıklayıcı değişkenler arasında çoklu doğrusal bağıntı bulunmamasıdır. Çoklu doğrusal bağıntı, iki ya da daha fazla açıklayıcı değişkenin doğrusal bir ilişki içerisinde olmasıdır. Bu durum sağlık alanına ait verilerde sık karşılaşılan bir durumdur. Bazı durumlarda çoklu doğrusal regresyon çelişkili sonuçlar verebilir. Model anlamlı iken, tüm bağımsız değişkenler tek tek anlamsız olabilir. Bunun anlamı model veri için yeterince uygun ancak bağımsız değişkenler bağımlı değişkeni istatistiksel olarak açıklayamamaktadır. İki değişken arasında yüksek korelasyon olduğunda, her ikisi de aslında aynı bilgiyi aktarmaktadırlar. Bu durumda her iki değişken birlikte modelde olduklarında etkileri çok fazlayken, her ikisi de diğerinden sonra modele eklendiğinde anlamlı derecede bir katkı yapmayabilir. Eğer her iki değişken de modelden çıkartılırsa modelin uyumu düşecektir. Böylece model anlamlı iken, bağımsız değişkenlerin hiçbiri modele eklense bile anlamlı katkı yapmayacaktır. Bu durumda bağımsız değişkenler aralarında doğrusaldırlar ve ortaya çıkan bu duruma çoklu doğrusal bağıntı denir (3).

Çoklu doğrusal bağıntıyı tespit etmek için pek çok yöntem geliştirilmiştir. Bunlardan ilki Frisch’in kavşak çözümlemesine dayalı yöntemdir. Bu yöntemde, bağımlı değişkenin açıklayıcı değişkenlerden her biriyle ayrı ayrı regresyon modeli kurulur ve başlangıç modelleri elde edilir. Sonrasında istatistiksel ve mantıksal olarak en olası modeller seçilir ve diğer bağımsız değişkenler kademe kademe eklenir. Eğer eklenen değişken, katsayıları anlamsız yapmıyor ve R^2 'yi de yükseltiyorsa faydalı kabul edilir ve modele alınır aksi halde modele alınmaz (4).

Çoklu doğrusal bağıntıyı tespit etmek için kullanılan bir diğer yöntem ise varyans artış faktörüdür (VIF – Variance Inflation Factor). Bağımsız değişkenlere ilişkin korelasyon matrisinin tersinin köşegen öğelerine VIF denir. VIF bir bağımsız değişkenin diğer bağımsız değişkenlerle olan ilişkisinin derecesini belirlemek için hesaplanır (5). VIF 10'a eşit veya daha büyük ise, çoklu doğrusal bağıntı problemi mevcuttur (1,6).

Çoklu doğrusal bağıntının belirlenmesinde kullanılan üçüncü yöntemde de koşul indeksinin (CI – Condition Index) kullanılmasıdır. CI'nin 10'dan küçük olması çoklu doğrusal bağıntı olmadığı gösterirken, 10 ile 30 arasında ise orta derecede; 30'un üzerinde ise aşırı derecede çoklu doğrusal bağıntı söz konusudur (7).

Farrar ve Glauber'ın 1967'de geliştirilen çoklu doğrusal bağıntı sınaması 3 ayrı testi (ki-kare, F ve t testleri) içermektedir. Bunlardan ilki çoklu doğrusal bağıntının varlığını test ederken, ikincisi hangi değişkenlerin doğrusal olduklarını test etmekte ve üçüncü test de çoklu doğrusal bağıntının formunu belirlemektedir (8).

ÇOKLU DOĞRUSAL BAĞINTIYI GİDERME YÖNTEMLERİ

Bağımlı ve bağımsız değişkenler arasındaki ilişkiyi belirleyen katsayı kestirimlerinin standart hatasını küçültmek ve daha tutarlı kestirim yapabilmek için çoklu doğrusal bağıntının giderilmesi zorunludur. Bu durumda aşağıdaki yöntemler önerilmektedir.

1. Daha fazla bilgi toplama

Çoklu doğrusal bağıntı sorununa bir çözüm olarak Farrar ve Glauber daha fazla veri toplamayı önermişlerdir. Bu yöntemde gözlem sayısının artırılmasının yanında bu verilerin bilgi kapsamalarının artırılması da söz konusudur. Örneğin, başlangıçta yıllık verilerin kullanılmış ve çoklu doğrusal bağıntı ortaya çıkmış ise, 3 aylık veya aylık veriler denenmelidir. Gözlem sayısının artırılması, parametre kestirimlerine ait varyansların küçülmesini sağlayabilir (2).

2. Bağımsız Değişkenlerden Bir veya Birkaçının Modelden Çıkartılması

Çoklu regresyon modelinde iki veya daha fazla bağımsız değişken arasında önemli ölçüde bir ilişki olduğunda, bu değişkenlere ait regresyon katsayılarının en küçük kareler kestirimleri sapmasız, fakat bu kestirimlerin varyansları büyük olur. Varyansı küçültmek için bağımlı değişken üzerinde daha az etkili olan bağımsız değişkenler regresyon modelinden çıkartılır. Bunun sonucunda regresyon modelinde kestirimlerin varyansları daha küçük, dolayısıyla kestirimlerin güvenilirliği yüksek olur. Bu yöntem uygulandığında, regresyon modelindeki parametrelerin kestirimleri sapmalı olur. Modelden çıkartılacak değişkenin değişimini açıklamada “çok önemli” olmadığı durumlarda bu yolun uygulanması önerilir (2).

3. Modelin Yeniden Belirlenmesi

Çoklu doğrusal bağıntının nedeni, model seçiminden kaynaklanabilir. Örneğin, ilişkili açıklayıcı değişkenlerin kullanılması gibi. Bu durumda ya açıklayıcı değişkenler yeniden tanımlanır, ya da ilişkili açıklayıcı değişkenlerin biri çıkartılır. Ancak açıklayıcı değişkenlerin çıkartılması modelin etkinliğini azaltabilir (9).

4. Yanlı Kestirim Yöntemleri

Çoklu doğrusal bağıntı varlığında parametre kestirimi yapabilmek için bir başka seçenek de yanlı regresyon yöntemlerinin kullanılmasıdır. Bu yöntemlerden Shrunken Regresyon, Temel Bileşenler Regresyonu ve Özdeğerler Regresyonunun esas amacı uygun değişken seçimi iken, Ridge Regresyonunda temel amaç çoklu bağıntı varlığında en küçük varyansla parametre kestirimidir (2).

Ridge regresyonun çözüm tekniği, basit en küçük kareler yöntemine benzer bir yöntemdir. Ridge regresyon

yönteminde, regresyon katsayı tahminlerini hesaplamadan önce standart formdaki değişkenlerin oluşturdukları ($X^T X$) matrisinin köşegen elemanlarına küçük ve pozitif bir sabitin (k) eklenmesiyle gerçekleştirilir. Buna göre Ridge regresyon çözümü,

$$\hat{\beta}_{RR} = [(X^T X + kI)]^{-1} X^T Y \quad [3]$$

şeklinde. Açıklayıcı değişkenler korelasyon matrisinin köşegen elemanlarına pozitif sabitlerin eklenmesindeki amaç, matris şartlı sayısının önemli ölçüde küçültülmesidir. $k=0$ için Ridge çözümü en küçük kareler çözümüne eşdeğer olduğundan Ridge tahmini, en küçük kareler tahmininin bir doğrusal dönüşümü olarak da ifade edilebilir (10).

GEREÇ VE YÖNTEMLER

2015 yılı Nisan ve Mayıs aylarında Bülent Ecevit Üniversitesi Tıp Fakültesi Hastanesi Radyoloji Anabilim Dalına abdomen bilgisayarlı tomografik incelemesi için başvuran hastalardan elde edilen subkutan yağ doku (X_1), viseral yağ doku (X_2), ve pankreas hacimlerinin (X_3) beden kitle indeksi (BKI) ile ilişkisi hem en küçük kareler yöntemiyle hem de Ridge regresyon yöntemiyle incelenmiş ve elde edilen sonuçlar karşılaştırılmıştır. Aşağıdaki tablolarda sonuçları verilen analizler için NCSS 2007 (versiyon 7.1.5) programı kullanılmıştır.

BULGULAR

Çalışmaya alınan 27'si (%42.2) erkek 37'si (%57.9) kadın 64 hastanın yaş ortalaması 60.3 ± 12.3 'tür. Hastaların çalışılan değişkenlerine ait tanımlayıcı istatistikler aşağıdaki tabloda özetlenmiştir (Tablo 1).

En küçük kareler yöntemine göre beden kitle indeksi ile açıklayıcı değişkenler arasındaki doğrusal ilişkinin değeri %85.2'dir. Beden ağırlığında meydana gelen değişimlerin yaklaşık %72.5'i bağımsız değişkenler tarafından açıklanmaktadır (Tablo 2).

$\alpha=0.05$ anlamlılık düzeyinde modelin anlamlı olduğu ($F=51.946$ ve $p < 0.001$) görülmektedir (Tablo 3).

Modelin yüksek belirlilik ($r^2 = 0.895$) ve korelasyon katsayısının ($R = 0.895$) aksine katsayılar incelendiğinde hiçbir değişkenin kısmi t testine göre anlamlı olmadığı

Tablo 1. Çalışılan değişkenlere ait tanımlayıcı istatistikler

	n	Ortalama \pm Std. Sapma	Minimum - Maksimum
Beden Kitle İndeksi (BKI)	64	24.7 \pm 1.9	19.9 - 26.7
Subkutan yağ doku	64	250.4 \pm 48.8	160.7 - 160.7
Viseral yağ doku	64	199.9 \pm 77.1	91.6 - 332.0
Pankreas hacmi	64	63.7 \pm 9.0	48.2 - 83.6

Tablo 2. Model Özeti

R	R ²	Düzeltilmiş R ²	Tahminin standart hatası	D-Watson (d)
0.852	0.725	0.711	13.67	2.057

Tablo 3. ANOVA Tablosu

	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	F	P
Regresyon	2102.932	3	700.977	51.946	<0.001
Hata	796.165	59	13.494		
Toplam	2899.097	62			

görülmektedir (Tablo 4). Bu durum olası çoklu doğrusal bağıntı probleminin göstergelerinden biridir.

$R_{(1:23)}^2, R_{(1:13)}^2, R_{(1:12)}^2$ değerlerinin yüksek olması, X_1, X_2 ve X_3 bağımsız değişkenleri için hesaplanan varyans artış faktörlerinin (VIF) 10'dan büyük olması ve koşul indeksinin (CI) de 30'un üzerinde olması çoklu doğrusal bağıntının işaretlerinden biridir (Tablo 5).

Tablo 4. Tahmin Edilen Regresyon Katsayıları ve Güven Aralıkları

	\hat{b}	Std. Hata	$\hat{\beta}$	t	p	\hat{b} için %95 güven aralığı	
						Alt	Üst
Sabit	54.569	22.337		1.955	0.055	-0.107	9.245
X_1	22.828	9.032	2.062	1.906	0.394	-0.037	0.092
X_2	10.239	6.011	1.079	0.901	0.381	-0.025	0.059
X_3	-6.477	3.014	1.059	-1.001	0.676	-0.022	0.024

Std: Standart

Tablo 5. En Küçük Kareler ve Çoklu Doğrusal Bağıntının Saptanması

	Korelasyon Matrisi			VIF	CI	
	X_1	X_2	X_3			
X_1	1			11.07	1.07	$R_{1,23}^2 = 0.872$
X_2	0.310	1		22.55	42.21	$R_{2,13}^2 = 0.791$
X_3	0.303	0.846	1	22.73	24.45	$R_{3,12}^2 = 0.953$

VIF: Varyans Artış Faktörü, CI: Koşul İndeksi

Ridge regresyon analizi sonuçlarının birinci kısmında belirli yanlılık katsayılarına göre R^2 , standartlaştırılmış Ridge regresyon katsayıları ile bu katsayılar ait VIF değerlerindeki değişimler verilmiştir (Tablo 6). Bu verilerden yararlanarak optimum k sabiti için, standartlaştırılmış Ridge regresyon katsayılarının durağanlaşmaya başladığı ve bu katsayılar ait VIF değerlerinin 1'e yaklaştığı nokta ($k=0.02$) seçilir. Yanlılık sabiti k ile modelin R^2 değerleri arasındaki ilişki incelendiğinde $k, 0$ ile 1 aralığında bir değişim gösterirken; $R^2, 0.544$ ile 0.725 aralığında değişmektedir. $k=0.02$ yanlılık sabitiyle elde edilen Ridge regresyon sonuçlarında modelin açıklayıcılığı %72.5'den %80.2'ye yükselmiştir. $k=0.02$ için elde edilen ANOVA tablosunda model anlamlı bulunmuştur (Tablo 7). $k=0.02$ alınan Ridge regresyon modelinde bağımsız değişkenler için hesaplanan varyans artış faktörü (VIF) değerleri sırasıyla 1.030, 1.026 ve 1.009 olarak elde edilmiştir (Tablo 8). Bu değerler kritik VIF = 10 değerinin oldukça altında değerlerdir. Bu durumda yeni modelde çoklu doğrusal bağıntının ortadan kalktığı söylenebilir. Logaritmik ölçekli grafikte standartlaştırılmış Ridge regresyon katsayılarının nasıl durağanlaştığı çok açık bir biçimde görülmektedir (Şekil 1).

En küçük kareler ve Ridge regresyon yöntemi ile elde edilen regresyon katsayıları, standart hataları, varyans artış faktörleri ve koşul indeksi değerleri birlikte incelendiğinde, Ridge regresyon yöntemi ile standart hataların küçüldüğü, en küçük kareler yöntemi ile 11.07 ile 32.73 arasında değerler alan varyans artış faktörü değerlerinin de 1'e yaklaştığı benzer şekilde koşul indeksi değerlerinin de 10'un altına indiği görülmektedir (Tablo 9).

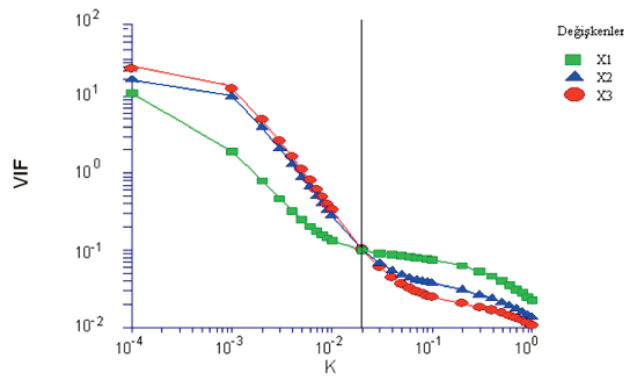
En küçük kareler yöntemi ile Ridge regresyon sonuçları arasındaki en önemli çelişki ise, pankreas hacmi (X_3) değişkenine ait regresyon katsayısının işaretinde görülmektedir. Yapılan çalışmalarda subkutan yağ doku,

Tablo 6. Ridge parametresi k, R2, Standartlaştırılmış Ridge Regresyon Katsayılar ve Varyans Artış Faktörü (VIF) değerleri

k	R ²	Standartlaştırılmış			VIF		
		Ridge Regresyon Katsayıları			X ₁	X ₂	X ₃
		X ₁	X ₂	X ₃			
0.000	0.725	2.828	1.239	-1.446	11.07	22.55	22.735
0.001	0.729	2.771	1.202	-0.703	10.27	18.101	20.033
0.004	0.732	2.339	1.112	-0.370	9.831	15.681	13.881
0.005	0.744	1.994	0.913	0.611	8.190	11.068	4.503
0.006	0.759	1.801	0.888	0.902	5.713	7.177	2.007
0.007	0.766	1.542	0.841	0.980	3.992	5.319	1.906
0.008	0.771	1.349	0.783	1.053	2.347	4.176	1.741
0.009	0.791	1.201	0.752	1.166	2.010	2.563	1.543
0.010	0.801	0.904	0.691	1.184	1.765	1.721	1.308
0.020	0.802	0.882	0.680	1.901	1.030	1.026	1.009
0.030	0.790	0.867	0.678	1.997	0.803	0.773	0.960
0.040	0.779	0.842	0.622	2.014	0.502	0.601	0.871
0.050	0.779	0.756	0.607	2.187	0.414	0.419	0.802
0.080	0.762	0.749	0.574	2.192	0.365	0.388	0.754
0.090	0.750	0.707	0.548	2.201	0.311	0.352	0.699
0.100	0.733	0.698	0.519	2.247	0.272	0.337	0.603
0.200	0.701	0.640	0.472	2.501	0.249	0.287	0.526
0.300	0.681	0.567	0.471	2.577	0.177	0.222	0.491
0.400	0.658	0.513	0.455	2.578	0.151	0.206	0.456
0.500	0.622	0.431	0.341	2.608	0.131	0.198	0.311
0.800	0.581	0.348	0.286	2.671	0.129	0.159	0.278
1.000	0.544	0.281	0.231	2.756	0.113	0.144	0.210

Tablo 7. k=0.02 için ANOVA Tablosu

	Kareler Toplamı	Serbestlik Derecesi	Kareler Ort.	F	P
Regresyon	2003.08	3	667.69	48.454	<0.001
Hata	799.24	58	13.78		
Toplam	2802.33	61	45.940		

**Şekil 1.** Varyans Artış Faktörü (VIF) Grafiği**Tablo 8.** Tahmin Edilen Ridge Regresyon Katsayıları ve VIF değerleri

	\bar{b}	Std. Hata	$\bar{\beta}$	t	VIF
Sabit	17.910	4.329		2.175	
X ₁	0.882	0.201	0.314	2.498	1.030
X ₂	0.680	0.154	0.126	3.143	1.026
X ₃	1.901	0.492	0.507	2.001	1.009

Std: Standart, VIF: Varyans Artış Faktörü

Tablo 9. En Küçük Kareler (EKK) ve Ridge Regresyon (RR) Yöntemleriyle Elde Edilen Sonuçların Karşılaştırılması

	\bar{b}		$\bar{\beta}$		Standart Hata		VIF		CI	
	EKK	RR	EKK	RR	EKK	RR	EKK	RR	EKK	RR
Sabit	54.569	17.910			22.337	4.329				
X ₁	22.828	0.882	2.062	0.314	9.032	0.201	11.07	1.003	1.07	1.02
X ₂	10.239	0.680	1.079	0.126	6.011	0.154	22.55	1.026	42.21	3.54
X ₃	-6.477	1.901	1.059	0.507	3.014	0.492	32.73	1.009	24.45	1.98

VIF: Varyans Artış Faktörü, CI: Koşul İndeksi, EKK: En Küçük Kareler, RR: Ridge Regresyon

EKK= $y_i' = 54.569 + 22.828X_1 + 10.239X_2 - 6.477X_3$ EKK= $y_i' = 17.910 + 0.882X_1 - 0.680X_2 + 1.901X_3$

viseral yağ doku ve pankreas hacmi ile beden kitle indeksi arasında pozitif yönlü korelasyon olduğu saptanmıştır (11,12). En küçük kareler ile elde edilen sonuçları incelediğimizde pankreas hacmi ile BKİ arasında negatif yönlü bir ilişki söz konusudur ve bu da literatür bilgisiyle çakışmaktadır. Ridge regresyon analizinin sonuçlarına bakıldığında ise pankreas hacmi (X3) değişkeninin katsayısı klinik beklentilerle uyumlu şekilde pozitif çıkmıştır.

TARTIŞMA

Bilindiği gibi çoklu doğrusal regresyonda temel amaç en küçük kareler yöntemi ile bağımsız değişkenler yardımıyla bağımlı değişkeni tahmin etmeye çalışmaktır. Ancak en küçük kareler yönteminin uygulanabilmesi için bir takım varsayımların yerine getirilmiş olması gerekmektedir. Bu varsayımlardan biri ya da bir kaç gerçekleşmediğinde elde edilen kestirim sonuçlarının yanlı çıkması kaçınılmazdır. Açıklayıcı değişkenler arasında çoklu doğrusal bağıntı olması durumunda yanlı regresyon yöntemleri ile elde edilen sonuçlar en küçük kareler yöntemine göre daha durağan ve klinik beklentilere uygun bulunmuştur. Tablo 9'da verilen sonuçlara bakıldığında Ridge regresyon yöntemi ile elde edilen VIF ve CI değerlerinin en küçük kareler yönteminde elde edilen değerlerden çok düşük olduğu görülecektir. En küçük kareler yönteminde 3 değişken için sırasıyla VIF değerleri 754.92, 601.33 ve 111.10 olarak elde edilirken, Ridge regresyon analizi sonucunda VIF değerleri sırasıyla 1.05, 1.04 ve 1 olarak bulunmuş benzer şekilde en küçük kareler yönteminde 3 değişken için bulunan CI değerleri sırasıyla 1.07, 42.21 ve 24.45 olarak elde edilirken, Ridge regresyon analizi sonucunda CI değerleri sırasıyla 1.01, 3.54 ve 1.98 bulunmuştur. Hem VIF değerleri hem de CI değerleri Ridge regresyon yöntemi ile kritik değerlerinin altına inmiştir (VIF >10 ve CI >10) ve bu da çoklu doğrusal bağıntının ortadan kalktığını göstermektedir. Yine Tablo 9'u incelendiğinde Ridge regresyon yöntemiyle elde edilen standart hatanın en küçük kareler yöntemindekine göre oldukça azaldığı görülmektedir. En küçük kareler yöntemiyle Ridge regresyon sonuçları arasındaki en önemli çelişki pankreas hacmi (X3) değişkenine ait regresyon katsayısının işaretinde ortaya çıkmıştır. En küçük kareler yönteminde pankreas hacminin regresyon katsayısı işareti negatif elde edilmiştir. Bunun anlamı, hastaların pankreas hacimleri arttıkça beden kitle indeksleri azaldığı şeklindedir. Bu da klinik beklentilerin tam tersi bir sonuçtur.

Alternatif yöntem olarak Ridge regresyon uygulandığında ise pankreas hacmi değişkeninin işareti pozitif elde edilmiş ve bu da klinik beklentilerle örtüşmektedir.

SONUÇ

Sonuç olarak, çoklu regresyon analizinde eğer çoklu doğrusal bağıntı söz konusu ise, en küçük kareler yöntemiyle parametre tahmininde bulunmak yanlış sonuçlar almamıza ve yorumlamamıza neden olabilir. Bu yanlışlığın önüne geçebilmek için çoklu doğrusal bağıntı varlığında, en küçük kareler yöntemi yerine Ridge regresyon gibi yanlış regresyon yöntemleri tercih edilmelidir.

KAYNAKLAR

1. Alpar R. Spor, Sağlık ve eğitim bilimlerinde örneklerle uygulamalı istatistik ve geçerlilik-güvenirlilik. Ankara: Detay Yayıncılık; 2000.
2. Emel İ. Çoklu bağıntılı doğrusal modellerde Ridge regresyon yöntemiyle parametre kestirimi: Türkiye'de, 1963-1983 enflasyon analizi. Eskişehir: Anadolu Üniversitesi Yayınevi; 1986.
3. Paul RK. Multicollinearity - causes, effects and remedies. Indian: Indian Agricultural Statistics Research Institute (IASRI); 2006.
4. Koutsoyiannis A. Ekonometri kuramı. Çeviren: Şenesen Ü. 1. Baskı. Ankara: Verso Yayıncılık; 1989.
5. Salkin NJ, Rasmussen K. Encyclopedia of measurements and statistics. Volume 1. California: Sage; 2007.
6. Albayrak AS. Çoklu doğrusal bağlantı halinde en küçük kareler tekniğinin alternatifi yanlış tahmin teknikleri ve bir uygulama. Zonguldak Karaelmas Üniversitesi Sosyal Bilimler Dergisi. 2005; 1(1): 105-26.
7. Paulson DS. Handbook of regression and modeling. Florida: Chapman & Hall/CRC; 2007.
8. Adeboye NO, Fagoyinbo IS, Olatayo TO. Estimation of the effect of multicollinearity on the standard error for regression coefficients. Journal of Mathematics. 2014; 10(4): 19.
9. Akdeniz F, Çabuk A. Ridge regresyon teorisinde 1970-2001 arasındaki gelişmeler. 5. Ulusal Ekonometri ve İstatistik Sempozyumu; 19- 21 Eylül 2001; Adana.
10. Sakallıoğlu S, Kaçıranlar S. A new biased estimator based on ridge estimation. Statistical Papers. 2008; 49(4): 669-89.
11. Shuster A, Patlas M, Pinthus JH, Mourtzakis M. The clinical importance of visceral adiposity: a critical review of methods for visceral adipose tissue analysis. The British Journal of Radiology. 2012; 85(1009): 1-10.
12. Camhi SM, Bray GA, Bouchard C, Greenway FL, Johnson WD, Newton RL, et al. The relationship of waist circumference and BMI to visceral, subcutaneous, and total body fat: sex and race differences. Obesity. 2011; 19(2): 402-8.