



Comparison of Least Squares and Some Bias Estimators in Multicollinearity

Furkan Yılmaz^{1,a}, Lütfi Bayyurt^{1,b}, Samet Hasan Abacı^{2,c}, Yalçın Tahtalı^{1,d,*}

¹Department of Animal Science, Faculty of Agriculture, Tokat Gaziosmanpaşa University, 60250 Tokat, Turkey

²Department of Animal Science, Faculty of Agriculture, Samsun Ondokuz Mayıs University, 55139 Samsun, Turkey

*Corresponding author

ARTICLE INFO	ABSTRACT
<p><i>Research Article</i></p> <p>Received : 27/02/2020 Accepted : 16/03/2020</p> <p>Keywords: Least squares method Ridge regression Principal component regression Saanen Multicollinearity</p>	<p>The aim of this study is to compare the least squares (LS) method that lost its function in the case of multicollinearity in regression methods with Ridge Regression (RR) and Principal Components Regression (PCR) which are bias estimators. For this aim, the effect of some body measurements on body weight (BW), body length (BL), height at withers (HW), height at rump (HR), chest depth (CD), chest girth (CG) and chest width (CW) obtained from 59 Saanen kids at weaning period raised at Research Farm of Tokat Gaziosmanpaşa University. Determination coefficient (R^2) and mean square error (MSE) values were used to evaluate the estimation performance of the methods. The multicollinearity between height at withers (HW) and height at rump (HR) which were used to estimate body weight was eliminated by using RR and PCR. When R^2 and HKO values of the examined methods are compared; It has been shown that RR method have better results of live weight of Saanen goats.</p>

Türk Tarım – Gıda Bilim ve Teknoloji Dergisi, 8(3): 793-799, 2020

Çoklu Doğrusal Bağlantı Durumunda En Küçük Kareler ve Bazı Yanlı Tahmin Edicilerin Karşılaştırılması

MAKALE BİLGİSİ	ÖZ
<p><i>Araştırma Makalesi</i></p> <p>Geliş : 27/02/2020 Kabul : 16/03/2020</p> <p>Anahtar Kelimeler: En küçük kareler Ridge regresyon Temel bileşenler regresyonu Saanen Çoklu doğrusal bağlantı</p>	<p>Bu çalışmanın amacı regresyon modellerinde çoklu doğrusal bağlantı olması durumunda işlevini yitiren en küçük kareler (EKK) metodu ile yanlı tahmin ediciler olan Ridge Regresyon (RR) ve Temel Bileşenler Regresyonu (TBR) yöntemlerini karşılaştırmaktır. Bu amaçla Tokat Gaziosmanpaşa Üniversitesi Araştırma ve Uygulama Çiftliğinde yetiştirilen süten kesim döneminde bulunan 59 baş Saanen oğlaklarının canlı ağırlıkları (CA) ve bazı vücut özelliklerinden vücut uzunluğu (VU), cidago yüksekliği (CY), sağrı yüksekliği (SY), göğüs derinliği (GD), göğüs genişliği (GG) ve göğüs çevresi (GÇ) ölçümleri kullanılmıştır. Metodların tahmin performansını değerlendirmek için belirtme katsayısı (R^2) ve hata kareler ortalaması (HKO) değerleri incelenmiştir. Çalışmada, canlı ağırlık tahmininde kullanılan cidago yüksekliği (CY) ve sağrı yüksekliği (SY) arasında çoklu doğrusal bağlantı olduğu görülmüş ve bu sorun RR ve TBR ile ortadan kaldırılmıştır. İncelenen yöntemlerin R^2 ve HKO değerleri karşılaştırıldığında; Saanen oğlaklarının canlı ağırlık tahmininde RR yaklaşımının diğer yöntemlerden daha iyi sonuç verdiği belirlenmiştir.</p>

^a furkanyilmaz60@hotmail.com

^c shabaci37@gmail.com

^b <http://orcid.org/0000-0003-1448-8185>

^d <http://orcid.org/0000-0002-1341-4056>

^b lufti.bayyurt@gop.edu.tr

^d yalcin.tahtali@gop.edu.tr

^b <http://orcid.org/0000-0003-2613-9302>

^d <http://orcid.org/0000-0003-0012-0611>



Giriş

Regresyon analizi, biyoloji, tıp, fen ve sosyal bilimler gibi bilim dallarında, sıklıkla kullanılan ve bağımlı değişken ile bir veya daha fazla bağımsız değişken arasındaki ilişkilerin tahmininde kullanılan istatistiksel yöntem olup değişkenler arasındaki ilişkileri belirlemek ve bu ilişkilere ait modelleri ortaya koymak amacıyla kullanılmaktadır. Basit regresyon olarak adlandırılan, bir bağımlı değişken ve bir bağımsız değişkenin oluşturduğu regresyon modelinde fonksiyonel ilişki belirlenerek ve bağımsız değişkenden yararlanarak bağımlı değişken tahmin edilebilir. Çoklu regresyon modelinde ise bağımlı değişken birden çok bağımsız değişken tarafından etkilenmektedir (Alpar, 2003).

Bağımsız değişken ile bağımlı değişken arasındaki ilişki, doğrusal ve doğrusal olmayan ilişki olarak farklılık göstermekte olup, değişkenler arasında doğrusal ilişki olması, bağımsız değişkenin bağımlı değişkene etkisinin doğrusal ve toplanabilir olduğunu ifade ederken, doğrusal olmayan ilişkide ise, bağımsız değişkenin bağımlı değişkene etkisi toplanabilir yapıda değildir (Tunç, 2018).

Basit veya çoklu doğrusal regresyon analizinde, modele ait parametre tahminlerinin güvenilir olması modelle bağlantılı bazı varsayımların sağlanmasını gerektirmektedir (Arı ve Onder, 2012). Bu varsayımlardan birisi bağımsız değişkenler arasında ilişkinin minimum olmasıdır. Yani bağımsız değişkenler arasındaki korelasyon katsayılarının sıfır veya sıfıra çok yakın olması anlamına gelmekte olup, aynı zamanda çoklu doğrusal bağlantı olmama şeklinde açıklanmaktadır (Orhunbilge, 2002). Çoklu doğrusal bağlantı olması durumunda ise elde edilen regresyon modelinde; regresyon katsayılarının tahminin hatalı olması, standart hata değerlerinin yüksek elde edilmesi, güven aralıklarının artması ve t test değerlerinin küçülmesi gibi problemler ile karşılaşmaktadır (Topal ve ark., 2010).

Çoklu doğrusal bağlantının problemini ortadan kaldırmak için çeşitli teknikler bulunmaktadır. Bunlardan biri mevcut veriye yeni bağımsız değişkenler eklenmesi veya yüksek ilişkili olan bağımsız değişkenlerden bazılarının modelden çıkarılmasıdır. Fakat bu durum model için önemli olan bir veya birkaç değişkenin modelden çıkarılmasına yani modelde bilgi kaybına neden olmaktadır. Öne sürülen diğer teknikler ise çoklu bağlantı olması durumunda modelde ki bağımsız değişkenleri çıkarmadan regresyon katsayılarını tahmin etmeyi amaçlamaktadır. Çoklu doğrusal regresyonda; parametre tahminleri için en küçük kareler (EKK) yönteminin kullanılması durumunda, hata kareler toplamının minimum yapılabilmesi için gerekli olan parametre tahmin edicileri seçilmektedir. Ancak bu yöntemin kullanılabilmesi için gerekli varsayımların yerine getirilmesi gerekmektedir. Bu varsayımların gerçekleşmediği durumda, tahmin sonuçlarının yanlış çıkması gibi problemler ile karşılaşmaktadır. Hata varyansının sabit ve bağımsız olması, hataya ait beklenen değerlerin sıfır olması, bağımsız değişkenler arasındaki ilişkinin olmaması gibi durumlar regresyon analizine ait varsayımlar olup, bunların gerçekleşmesi durumunda parametre tahminleri yanlış elde edilebilmektedir.

Çoklu bağlantı olması durumunda, işlevini yitiren EKK yöntemi yerine yanlış kestirim tekniklerinin kullanılması

daha güvenilir tahmin modellerinin oluşmasını sağlamaktadır. EKK yerine yanlış kestirim tekniklerinden yaygın olarak, modelde korelasyon matrisinin köşegen elemanlarına k yanlışlık parametresi eklenerek kestirim varyanslarının küçültüldüğü Ridge Regresyon (RR) ve gerçek bağımsız değişkenler yerine bunların dik açıklayıcı dönüşüm kümesinin kullanıldığı Temel Bileşenler Regresyonu (TBR) kullanılmaktadır (Polat, 2009).

Bu çalışmada, süttan kesim döneminde bulunan 59 baş Saanen oğlaklarından elde edilen bazı vücut özellikleri ve canlı ağırlıklarına ait veriler kullanılarak model tespiti amaçlanmış olup, değişkenler arasında çoklu bağlantı problemi olmasından dolayı EKK yöntemine alternatif olarak RR ve TBR analiz yöntemleri kullanılmış ve yöntemlerin etkinliği karşılaştırılmıştır.

Materyal ve Metot

Materyal

Bu çalışmada Tokat Gaziosmanpaşa Üniversitesi araştırma uygulama çiftliğinde yetiştirilen, 59 baş Saanen oğlağının vücut uzunluğu (VU), cidago yüksekliği (CY), sağrı yüksekliği (SY), göğüs derinliği (GD), göğüs genişliği (GG) ve göğüs çevresi (GÇ) ölçülerine ait 56. gün verileri kullanılmıştır. Denemede her 15 günde bir kayıt altına alınan vücut ölçüleri Ertuğrul (1996) tarafından bildirilen ölçü alma yöntemlerine göre ve yine bu araştırmacı tarafından bildirilen ölçü bastonu ve ölçü şeridi kullanılarak ölçme işlemi yapılmış ve veriler kaydedilmiştir. Verilerin istatistik analizleri için SPSS ve NCSS paket programlarından yararlanılmıştır.

Metot

Regresyon analizi, bir bağımlı ve bir veya birden fazla bağımsız değişken arasındaki sebep-sonuç ilişkisini ortaya koyan bir yöntem olup modele yönelik tahmin yapılmasına imkân sağlamaktadır. Birden fazla bağımsız değişkenin olduğu çoklu regresyon modeline ait fonksiyonun matris notasyonu aşağıda gösterilmiştir (Alpar, 2011).

$$Y = X\beta + \varepsilon \quad (1)$$

Bu denklemde

$$\begin{aligned} Y &= n \times 1 \text{ boyutlu bağımlı değişken vektörü} \\ X &= n \times (p+1) \text{ boyutlu bağımsız değişkenler matrisi} \\ \beta &= (p+1) \times 1 \text{ boyutlu katsayılar vektörü} \\ \varepsilon &= n \times 1 \text{ boyutlu hata vektörünü göstermektedir.} \end{aligned}$$

Matris notasyonunda elde edilen tahmin denklemi ise

$$\hat{Y} = X\hat{\beta} \quad (2)$$

şeklinde gösterilmektedir.

Çoklu regresyon modelinde yer alan bağımsız değişkenler arasındaki yüksek doğrusal ilişkinin varlığı regresyon modeli için çoklu doğrusal bağlantı problemini ortaya çıkarmaktadır. Bu problemin varlığı oluşturulan modelin yanlış tahminde bulunmasına neden olmaktadır. Bu nedenle çoklu bağlantı probleminin tespiti ve giderilmesi çoklu regresyon modeli için oldukça önemlidir.

Çoklu doğrusal bağlantıyı belirlemek amacıyla, birçok yöntem kullanılmakta olup, bunlardan birisi bağımlı değişken ile ilk bağımsız değişken arasındaki regresyon modeli belirlenir ve diğer bağımsız değişkenlerin ayrı ayrı eklenmesiyle en iyi modelin tespiti amaçlanır (Koutsoyiannis, 1989). Diğer bir yöntem ise varyans artış faktörünün (VIF: Variance Inflation Factor) kullanılmasıdır. Bağımsız değişkenlere ait VIF değerleri 10'dan büyük olarak elde edilmiş ise bu durumda çoklu bağlantı probleminde söz etmek mümkündür. Çoklu bağlantı olması durumunda, tahmin modellerinin elde edilmesi amacıyla EKK yöntemi yerine RR ve TBR yöntemlerinden yararlanılmaktadır.

En Küçük Kareler Yöntemi

En küçük kareler yöntemi, veri noktaları arasındaki ilişkinin görsel olarak gösterilmesini sağlayan, bir veri kümesi için en uygun çizgiyi belirlemek için kullanılan yöntemdir. Bu yöntemin amacı normal dağılım ve homojen varyansa sahip hata terimlerinin karelerinin toplamını minimum yaparak oluşturulan modeli optimize etmektir (Çankaya ve ark., 2019).

Çoklu regresyon analizinde, EKK yöntemi ile katsayılar vektörünün tahmininde aşağıdaki eşitlik kullanılır (Alpar, 2011).

$$\hat{\beta}_{OLS}=(X'X)^{-1}X'Y \quad (3)$$

$\hat{\beta}_{OLS}$, $\hat{\beta}$ parametresinin var olan tüm yansız tahmin edicileri arasında en küçük varyansa sahip olanıdır. En küçük varyansa sahip olma özelliği teoride tatmin edici görülmüş bağımsız değişkenler arasında lineer bağımlılığa yakın bir ilişki varsa daha küçük varyansa sahip yanlış bir tahmin edici bulunabileceğinden EKK tahmin edicisinin pratikte kullanılabilirliği azalmaktadır (Kurnaz, 2011).

Ridge Regresyon

Ridge regresyon, (Hoerl ve Kennard, 1970) tarafından regresyon katsayılarının yanlış tahmin edicilerine izin veren EKK yönteminin değiştirilmesi ile geliştirilmiş bir yöntemdir. Yanlış tahmin edicilerin yansız tahmin edicilerden küçük olmasından dolayı, yanlış tahmin edicilerin gerçek parametre değerlerine yakın olma olasılıkları daha yüksek olacağından, yansız olanlara göre tercih edilmektedirler.

RR yöntemi çoklu doğrusal bağlantı olduğu durumda gerekli olan tüm değişkenlerin modele alınmasına olanak sağlamaktadır. Çoklu doğrusal bağlantı olduğunda RR, EKK yönteminden daha küçük varyanslı parametre kestirimlerinin elde edilmesini ve model için gereksiz görülen değişkenlerin modelden çıkarılmasını amaçlamaktadır (Arı ve Onder, 2012). RR yöntemi, 3 nolu eşitliğin köşegen $X'X$ matrisine sabit bir k yanlışlığı ile uygun parametre değerlerini bulmayı amaçlamaktadır (Düzgüneş ve ark., 1987).

$$\hat{\beta}(k)=(X'X+kI)^{-1}X'Y, \dots k \geq 0 \quad (4)$$

RR yönteminde k 'nin değeri artırılması ile yanlış bir tahmin elde edilmekle birlikte varyansın değerinde önemli bir azalma meydana gelmektedir. RR modelinde k değeri matris özdeğerlerine göre belirlenmektedir. k değerini belirlemek için bir çok araştırmacı tarafından çeşitli

yöntemler kullanılmıştır (Hoerl ve Kennard, 1970; Lee ve Campbell, 1985; Akdeniz ve Erol, 2003; Sahin ve ark., 2018). Bu çalışmada ise Ridge tahmin edici için k değerinin belirlenmesinde Kurtuluş (2001) tarafından önerilen aşağıdaki 5 nolu eşitlik verilmiştir.

$$\hat{\beta}(k)=(X'X+kI)^{-1}X'Y, \dots k \geq 0, k \leq \frac{\lambda_{\max}-100\lambda_{\min}}{99}, k \neq 0 \quad (5)$$

Temel Bileşen Regresyonu

Temel bileşen regresyonu, Massy (1965) tarafından model kararsızlığını ortadan kaldırarak ve regresyon katsayılarının varyanslarını azaltarak çoklu bağlantı problemini ele almak için oluşturulmuş bir yöntemdir. TBR her doğrusal regresyon modelinin bir dik bağımsız değişkenler kümesine bağlı olarak yeniden açıklanması temeli üzerine oluşturulmuş olup, bağımsız değişkenler arasında yüksek korelasyon olduğu durumlarda uygulanmaktadır (Özkan, 2009).

Çoklu doğrusal regresyon modeli için bağımsız değişken katsayıları matris notasyonunda;

$$\hat{\beta}=(X'X)^{-1}X'Y \quad (6)$$

olarak tahmin edilebilir. Bu eşitlikte X bağımsız değişkenler matrisini, Y ise bağımlı değişken vektörünü belirtmektedir. TBR'de ise bağımsız değişken katsayıları;

$$\hat{\beta}_{PC}=D_q\Lambda_q^{-1}D_q'X'Y \quad (7)$$

olarak tahmin edilmektedir. Bu eşitlikte, D_q , $X'X$ ye ait ilk q adet öz vektör matrisini; Λ_q , $X'X$ ye ait ilk q adet öz vektör için köşegen matrisi simgelemektedir (Al-Hassan ve Al-Kassab, 2009). TBR sonucunda oluşturulan tahminler yanlış olur. Ancak varyansın düşürülmesi ile bu yanlış tahminde ki büyüklük dengelenmiş olur (Arı ve Onder, 2012).

Bulgular ve Tartışma

Analiz sonucunda, süttan kesim döneminde bulunan 59 baş Saanen oğlağından elde edilen canlı ağırlık ve bazı vücut ölçümlerine ait tanımlayıcı istatistikler Tablo 1'de verilmiştir.

İncelenen özelliklerin normallik varsayımı Kolmogorow-Smirnov testi ile test edilmiş ve verilerin hata terimlerinin normal dağıldığı bulunmuştur ($P>0,05$). Saanen oğlaklarının canlı ağırlıklar ile bazı vücut ölçümleri arasındaki Pearson korelasyon katsayıları ve anlamlılık testi sonuçları Tablo 2'de verilmiştir.

Saanen oğlaklarının süttan kesim dönemindeki canlı ağırlıkları ile incelenen vücut ölçümleri arasında pozitif yönlü ilişki olup, en yüksek korelasyon CY ile SY arasında ($r=0,948$, $P<0,01$) bulunurken, en düşük korelasyon GD ile GG ($r=0,376$, $P<0,01$) ve CA ile GD arasında ($r=0,376$, $P<0,01$) bulunmuştur. İncelenen değişkenler arasındaki korelasyon katsayılarının 1'e yakın olması durumunda, çoklu bağlantı probleminin varlığından bahsedilebilir. Çoklu bağlantı probleminin olup olmadığının tespiti amacıyla kullanılan diğer bir yöntem VIF değerlerinin kullanılmasıdır. Bu amaçla, incelenen değişkenlerin çoklu regresyon analizi sonuçları, EKK, TBR ve RR yöntemlerine göre sırasıyla verilmiştir.

Tablo 1. Saanen oğlaklarının bazı vücut özelliklerine ait tanımlayıcı istatistikler.

Table 1. Descriptive statistics of the some body traits of Saanen kids.

Değişkenler	n	Ortalama	Standart Sapma	Varyasyon Katsayısı
Canlı ağırlık (CA)	59	14,266	2,317	16,2
Vücut uzunluğu (VU)	59	60,776	4,981	8,2
Cidago yüksekliği (CY)	59	46,800	2,862	6,1
Sağrı yüksekliği (SY)	59	46,778	3,011	6,4
Göğüs derinliği (GD)	59	19,550	2,054	10,5
Göğüs genişliği (GG)	59	15,676	1,521	9,7
Göğüs çevresi (GÇ)	59	61,253	4,463	7,3

Tablo 2. İncelenen özellikler arasındaki korelasyon katsayıları

Table 2. Correlation coefficients among examined traits.

Değişkenler	CA	VU	CY	SY	GD	GG
VU	0,515**					
CY	0,726**	0,651**				
SY	0,689**	0,652**	0,948**			
GD	0,376**	0,624**	0,637**	0,627**		
GG	0,537**	0,472**	0,407**	0,413**	0,376**	
GÇ	0,647**	0,636**	0,682**	0,677**	0,590**	0,440**

*: P<0,05; **: P<0,01

Tablo 3. En Küçük Kareler metoduna göre regresyon analiz sonuçları

Table 3. Regression analysis results according to Least Squares Method

Değişkenler	Regresyon Katsayısı	Standart Hata	Standardize Katsayılar	t	P	VIF
Sabit	-17,768	3,238		-5,487	**	
VU	-0,010	0,057	-0,022	-0,182	-	2,273
CY	0,547	0,210	0,676	2,602	*	10,308
SY	-0,063	0,197	-0,081	-0,316	-	10,131
GD	-0,298	0,129	-0,264	-2,314	*	1,994
GG	0,424	0,143	0,278	2,956	**	1,353
GÇ	0,150	0,063	0,289	2,380	*	2,250

*: P<0,05; **: P<0,01 ; P>0,05

Tablo 4. Korelasyon öz değer ve koşul indeksi değerleri

Table 4. Correlation eigenvalue and the number of conditions

Sıra No	Öz değer	Koşul İndeksi
1	3,996	1,000
2	0,734	5,450
3	0,482	8,290
4	0,387	10,330
5	0,350	11,410
6	0,052	76,630

EKK Yöntemi Analiz Sonuçları

Sütten kesim dönemindeki Saanen oğlaklarının CA'ları ile bağımsız değişkenler (VU, CY, SY, GD, GG, GÇ) arasındaki EKK yöntemine ait regresyon analizi sonuçları Tablo 3'de gösterilmiştir. EKK yöntemi kullanılarak yapılan çoklu regresyon analizi sonuçlarına göre, CA tahmininde kullanılan VU ile SY ölçüleri istatistiksel olarak önemsiz bulunmuştur. Ayrıca Tablo 3 incelendiğinde, bağımsız değişkenlerden SY ile CY'ye ait VIF değerlerinin 10'dan büyük olduğu görülmekte olup, bu sonuç ile değişkenler arasında çoklu bağlantı olduğu gözlenmiştir. Bu metoda göre, CA tahmininde, standartlaştırılmış regresyon katsayısı dikkate alındığında, modele en fazla katkıyı CY'nin, en düşük katkıyı ise VU'nun yaptığı gözlenmiştir. Ayrıca, VU ile birlikte SY ve GD'nin katkısı negatif yönde elde edilmiştir.

RR Analiz Sonuçları

Yapılan analiz sonucunda, CA'yı tahmin etmede vücut özelliklerinin açıklama gücü %65,96 olarak bulunmuştur (P<0,001). Tablo 4'de koşul indeksi 10'dan büyük olduğu için çoklu bağlantı sorunu görülmektedir.

Tablo 5'de k yanlılık tahmin edicisinin seçim tablosu görülmektedir. RR yönteminde k sabitinin tespiti için standartlaştırılan regresyon katsayılarının durağan hale geldiği ve bu katsayılara ait VIF değerlerinin ise 1'e yaklaştığı k sabiti için seçim değeri olarak belirlenmektedir. Yapılan analiz sonucunda k=0,000878 olarak tespit edilmiştir. Ayrıca, Tablo 5 incelendiğinde R² değerinin %65,96 ile %66,03 arasında değiştiği görülmektedir.

Tablo 6 da ise VIF değerleri görülmektedir. Tablo 5 ve 6'ya bakıldığında VIF değerlerinin 10'dan küçük olduğu yani en yüksek R² değerini veren k değerinin 0,000878 olarak seçildiğinde, CA tahmini için kullanılan vücut

ölçümleri olan SY ve CY ölçümleri arasındaki çoklu bağlantı probleminin RR yöntemi ile ortadan kaldırıldığı görülmektedir (Tablo 7).

TBR Analiz Sonuçları

Saenen oğlaklarının sütten kesim döneminde ölçülen bazı vücut özelliklerinin TBR analizi sonuçları sırasıyla Tablo 8 ve 9’da gösterilmektedir. Tablo 8 incelendiğinde, 6 temel bileşene ait öz değerler görülmektedir. TBR analiz sonuçlarına göre VIF değerleri Tablo 9 da verilmiştir. Bu değerler incelendiğinde, sütten kesim dönemindeki Saenen

oğlaklarının CA tahmininde kullanılan ve aralarında çoklu bağlantı tespit edilen, bağımsız değişkenler (SY, CY) arasındaki çoklu bağlantı probleminin giderildiği görülmektedir.

Ayrıca, TBR yönteminin uygulanması ile bu değişkenlere ait standart hatalar, EKK yöntemine göre elde edilen standart hatalara göre düşme eğilimi göstermiştir. Elde edilen sonuçlara göre, SY, CY, GD, GG ve GÇ ölçümlerinin istatistiksel olarak anlamlı olduğu görülmektedir.

Tablo 5. k parametre seçimine ait değerler

Table 5. Values of k parameter selection

k	R ²	Sigma	B'B	Ortalama VIF	Max VIF
0,000000	0,6603	1,4262	0,6942	4,7182	10,3085
0,000100	0,6602	1,4263	0,6930	4,7053	10,2712
0,000200	0,6601	1,4265	0,6917	4,6925	10,2341
0,000300	0,6600	1,4266	0,6905	4,6797	10,1973
.
.
0,000700	0,6598	1,4272	0,6857	4,6292	10,0519
0,000800	0,6597	1,4274	0,6845	4,6168	10,0160
0,000878	0,6596	1,4275	0,6836	4,6071	9,9882
0,000900	0,6596	1,4275	0,6833	4,6044	9,9804

Tablo 6. Varyan Artış Faktör (VIF:Variance Inflation Factor) değerleri

Table 6. Variance Inflation Factor (VIF) values

k	VU	CY	SY	GD	GG	GÇ
0,000000	2,2728	10,3085	10,1312	1,9937	1,3531	2,2501
0,000100	2,2716	10,2712	10,0947	1,9928	1,3526	2,2489
0,000200	2,2704	10,2341	10,0584	1,9919	1,3522	2,2477
0,000300	2,2692	10,1973	10,0223	1,9909	1,3518	2,2465
.
.
0,000700	2,2643	10,0519	9,8799	1,9872	1,3502	2,2418
0,000800	2,2631	10,0160	9,8448	1,9862	1,3498	2,2406
0,000878	2,2621	9,9882	9,8176	1,9855	1,3495	2,2397
0,000900	2,2619	9,9804	9,8099	1,9853	1,3494	2,2395

Tablo 7. k = 0,000878 seçimine bağlı Ridge regresyon analizi sonuçları

Table 7. Ridge regression analysis results according to k = 0,000878 bias constant

Değişkenler	Regresyon Katsayısı	Standart Hata	t-değeri	P	VIF
Sabit	-17,749	-	-	-	-
VU	-0,010	0,056	-0,179	-	2,262
CY	0,542	0,207	2,618	**	9,988
SY	-0,058	0,195	-0,297	-	9,818
GD	-0,297	0,129	2,302	*	1,986
GG	0,423	0,143	2,958	**	1,350
GÇ	0,149	0,063	2,365	*	2,240

** : P<0,01. * : P<0,05. - : P>0,05

Tablo 8. Temel Bileşenler Regresyona ait tanımlayıcı istatistikler

Table 8. Descriptive statistics of Principal Component Regression analysis

Temel Bileşenler	Temel Bileşenler Katsayısı	Bireysel R ²	Özdeğerler
PC1	-0,830	0,513	3,996
PC2	-0,254	0,009	0,734
PC3	1,110	0,111	0,482
PC4	0,384	0,011	0,387
PC5	-0,158	0,002	0,350
PC6	1,252	0,015	0,052

Tablo 9. Temel Bileşenler Regresyon Analiz Sonuçları
Table 9. Principal component regression analysis results

Değişkenler	Regresyon Katsayısı	Standart Hata	t-değeri	P	VIF
Sabit	-17,016				
VU	-0,012	0,058	0,207	-	2,272
CY	0,236	0,053	4,452	**	0,619
SY	0,229	0,051	4,490	**	0,649
GD	-0,289	0,131	-2,206	*	1,990
GG	0,418	0,146	2,863	**	1,352
GÇ	0,152	0,064	2,375	*	2,249

Tablo 10. EKK, RR ve TBR analiz sonuçlarının karşılaştırılması
Table 10. Comparison of LS, RR and PCR analysis results

Yöntemler	HKO	R ²	% VK	Önem düzeyleri
EKK	2,036	0,660	0,100	<0,001
RR	2,038	0,659	0,100	<0,001
TBR	2,125	0,645	0,102	<0,001

Yöntemlerin Karşılaştırılması

EKK, RR ve TBR yöntemleri ile elde edilmiş tahminlere ait hata kareler ortalaması (HKO), belirtme katsayısı (R²) ve varyasyon katsayısı (%VK) Tablo 10'da verilmiştir.

Tablo 10 incelendiğinde, üç farklı yöntemle elde edilen tüm modellerin istatistiksel olarak anlamlı olduğu görülmektedir (P<0,001). HKO; EKK, RR ve TBR yöntemlerine göre sırasıyla 2,036, 2,038 ve 2,125 olarak bulunmuş olup, en küçük HKO değeri EKK yönteminde elde edilmiştir. Bununla birlikte, EKK, RR ve TBR için R² değerleri sırasıyla 0,660, 0,659 ve 0,645 olarak bulunmuştur. Bu sonuçlar incelendiğinde, RR ve TBR analizine göre VIF değerlerindeki düşmeye rağmen R² değerlerinde çok büyük bir değişiklik gözlenmemiştir.

Bu sonuçlara göre, özellikle RR yöntemi ile aralarında doğrusal bağlantı bulunan değişkenlere ait VIF değerlerinde önemli bir düşme olmasına rağmen R² değerlerinde çok büyük değişime yol açmadığından diğer yöntemlere göre daha güvenilir bir model ortaya koymuştur.

Çoklu doğrusal regresyon modeli ile esas amaç, EKK yöntemi ile bağımsız değişkenler kullanılarak bağımlı değişkenler tahmin etmektir. Ancak EKK yönteminin kullanılması için bazı varsayımların sağlanmış olması gerekmektedir. Bu çalışmada, EKK yöntemine göre bağımsız değişkenler arasında çoklu bağlantı olması durumunda, bu yönteme alternatif olarak TBR ve RR yöntemleri ele alınmıştır.

Bu konu ile ilgili olarak bazı araştırmacılar tarafından benzer çalışmalar yapılmış olup; Çankaya ve ark. (2019)'nın Karayaka kuzuları ile yaptıkları çalışmada, canlı ağırlık tahmininde EKK, RR ve TBR yöntemleri için R² değerlerini sırasıyla 0,634, 0,627 ve 0,623 olarak bildirmişlerdir. Bu değerler bu çalışmadan elde edilen R² değerlerinden daha düşüktür. Ayrıca Topal ve ark. (2010)'nın Sazan balıklarında yaptıkları çalışmada, karkas ağırlığı tahmini için EKK, RR ve TBR yöntemlerini kullanmışlar ve R² değerlerini sırasıyla 0,905, 0,898 ve 0,878 olarak bulmuşlardır. Bu çalışmadan elde edilen R² değerleri daha yüksek bulunmuştur. Üçkardeş ve ark. (2012)'nin yaptıkları çalışmada Japon bildircin yumurtalarının iç kalite özelliklerinden olan ak yüksekliğini tahmin etmek ve çoklu bağlantı problemini

gidermek için RR yöntemini kullanmışlardır. Çalışmadan elde edilen sonuçlar bu çalışma sonuçlarıyla benzerlik göstermektedir. Iqbal ve ark. (2019) yaptıkları çalışmada Harnai koyunlarında canlı ağırlık tahmini için RR yöntemini kullanmışlardır. Çalışma sonucunda elde ettikleri R² değeri bu çalışma sonucunda elde edilen R² değerinden daha yüksek bulunmuştur.

Sonuç

Bu çalışma sonucunda Saanen oğlaklarının sütten kesim dönemindeki canlı ağırlıklarının tahmini için EKK, RR ve TBR yöntemleri kullanılmıştır. Öncelikle EKK yöntemi ile Saanen oğlaklarının canlı ağırlık tahmini için kullanılan vücut özelliklerinin birbirleriyle yüksek korelasyon gösterdiği tespit edilmiştir. Bu durum çoklu bağlantı probleminin varlığını göstermiştir. Çoklu bağlantı probleminin varlığı durumunda EKK yöntemi, bağımsız değişkenler arasında ilişki olmama varsayımından dolayı işlevselliğini yitirmektedir. Değişkenler arasında çoklu bağlantı olduğu durumda, RR ve TBR yöntemleri, EKK yöntemine göre daha iyi sonuçların elde edilmesini sağlamaktadır. EKK yöntemi sonucunda VU, CY ölçüleri için elde edilen VIF değeri, RR ve TBR'den daha yüksek bulunmuştur. Ayrıca, EKK yöntemi ile yapılan analiz sonucunda elde edilen R² değeri RR ve TBR'den yüksek, HKO değeri ise daha düşük bulunmuştur. Bağımsız değişkenler arasındaki çoklu bağlantıların etkilerini ortadan kaldırmak için EKK yöntemi yerine yanlış tahmin edici olan RR ve TBR kullanılması, elde edilen sonuçların daha doğru yorumlanmasına katkıda bulunacaktır. Ayrıca, bu yöntemlerin hassasiyet ve kullanılabilirliğinin araştırıldığı farklı çalışmalar sonucunda elde edilecek sonuçların bu alanda çalışan araştırmacılara katkı sağlayacağı düşünülmektedir.

References/Kaynaklar

- Akdeniz F, Erol H. 2003. Mean squared error matrix comparisons of some biased estimators in linear regression. Commun. Stat. Theor. Meth., 32: 2389-2413. <https://doi.org/10.1081/STA120025385>
- Al-Hassan YM, Al-Kassab MM. 2009. A monte carlo comparison between ridge and principal components regression methods. Appl. Math. Sci., 3: 2085-2098

- Alpar R. 2003. Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş. 1. Nobel Yayın Dağıtım, Ankara, 404s.
- Alpar R. 2011. Uygulamalı Çok Değişkenli İstatistiksel Yöntemler. 3. Baskı. Kızılay/Ankara. Detay Yayıncılık. ISBN:978-605- 5437-42-8
- Arı A, Onder H. 2013. Farklı Veri Yapılarında Kullanılabilecek Regresyon Yöntemleri. Anadolu Tarım Bilimleri Dergisi, 28(3): 168-174.
- Çankaya S, Eker S, Abacı SH. 2019. Comparison of Least Squares, Ridge Regression and Principal Component Approaches in the Presence of Multicollinearity in Regression Analysis. Turkish Journal of Agriculture-Food Science and Technology, 7(8), 1166-1172.
- Düzgüneş O, Kesici T, Kavuncu O, Gürbüz F. 1987. Experimental design II. Agriculture Faculty Publications, Ankara University, Ankara, Turkey.
- Ertuğrul M. 1996. Küçükbaş Hayvan Yetiştirme Uygulamaları. II. Baskı. Ankara Üniversitesi, Ziraat Fak. Yayın no:1446, Ders Kitabı: 426, Ankara.
- Hoerl AE, Kennard RW. 1970. Ridge Regression: Biased Estimation to Nonorthogonal Problems, Technometrics, 12: 56 – 67.
- Iqbal F, Ali M, Huma ZE, Raziq A. 2019. Predicting Live Body Weight Of Harnaı Sheep Through Penalized Regression Models. Journal of Animal and Plant Sciences, 29(6): 1541-1548.
- Kurnaz FS. 2011. Bazı Yanlı Tahmin Edicilerde Yanlılık Parametresinin Tahmin Edilmesi. İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi.
- Kurtuluş M. 2001. A Study on ridge regression. MSc thesis, Gazi University, Ankara, Turkey
- Lee TZ, Campbell DB. 1985. Selecting the optimum k in ridge regression. Commun. Stat. Theor. Meth., 14: 1589-1604. <https://doi.org/10.1080/03610928508828997>
- Massy WF. 1965. Principal Components Regression in Exploratory Statistical Research, Journal of the American Statistical Association, 60: 234 – 246.
- Orhunbilge N. 2002. Uygulamalı Regresyon ve Korelasyon Analizi. İstanbul Üniversitesi, İşletme Fakültesi Yayınları, İstanbul.
- Özkan K. 2009. Modeling of variation of soil capacity of soil by basic components of soil tiller by regression analysis. Süleyman Demirel Univ. J. Forest., 2: 1-9.
- Polat E. 2009. Kısmi En Küçük Kareler Regresyonu. Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi.
- Sahin M, Yavuz E, Uckardes F. 2018. Multicollinearity Problem and Bias Estimates in Japanese Quail. Pakistan Journal of Zoology, 50(2).
- Topal M, Eydurhan E, Yağanoğlu AM, Sönmez A, Keskin S. 2010. Çoklu Doğrusal Bağlantı Durumunda Ridge ve Temel Bileşenler Regresyon Analiz Yöntemlerinin Kullanımı. Atatürk Üniversitesi, Ziraat Fakültesi Dergisi, 41(1): 53-57.
- Tunç Z. 2018. En Küçük Kareler ve Temel Bileşenler Regresyon Analizlerinin Karşılaştırılması. Yüksek Lisans Tezi, İnönü Üniversitesi, Sağlık Bilimleri Enstitüsü, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Malatya.
- Ertuğrul M. 1996. Küçükbaş Hayvan Yetiştirme Uygulamaları, 2. Baskı, Ankara Üniv.
- Üçkardeş F, Ercan E, Narinç D, Aksoy T. 2012. Japon Bildircinlarında Yumurta Ak İndeksinin Ridge Tahmin Edilmesi. Akademik Ziraat Dergisi, 1(1): 11-20.