



## Kıl Keçilerinin Vücut Ölçülerini Kullanarak Canlı Ağırlıklarını Tahmin Etmede Kısmi En Küçük Kareler ve Temel Bileşenler Regresyon Yöntemlerinin Karşılaştırılması \*

Seda AKYÜREK SAĞIR<sup>1\*\*</sup>, Suna AKKOL<sup>2</sup>

<sup>1</sup>Ankara Üniversitesi, Ziraat Fakültesi, Zootečni Bölümü, 6130, Ankara

<sup>2</sup>Van Yüzüncü Yıl Üniversitesi, Ziraat Fakültesi, Zootečni Bölümü, 65080, Van

<sup>1</sup><https://orcid.org/0009-0000-7493-9553>

<sup>2</sup><https://orcid.org/0000-0001-5123-7516>

\*\*Sorumlu yazar: seda.akyurek@gsb.gov.tr

### Araştırma Makalesi

#### Makale Tarihiçesi:

Geliş tarihi: 21.11.2023

Kabul tarihi:21.01.2024

Online Yayınlanma: 25.06.2024

#### Anahtar Kelimeler:

Canlı ağırlık tahmini

Kıl keçileri

Kısmi en küçük kareler

Temel bileşenler

Vücut ölçüleri

Çoklu bağlantı problemi

### ÖZ

Bu çalışmanın amacı, çoklu bağlantı probleminin varlığında kıl keçilerinde çeşitli vücut ölçüleri kullanılarak canlı ağırlıklarının tahmin edilmesinde Kısmi En Küçük Kareler (KEKK) ve Temel Bileşenler (TB) regresyon yöntemlerinin kullanılması ve çalışma verisi için en iyi tahmin yönteminin bulunarak sonuçların yorumlanması amaçlanmıştır. Bu amaçla, 119 baş dişi kıl keçisinden ölçümü yapılmış canlı ağırlıklar ve çeşitli vücut ölçüleri (cidago yüksekliği, sağrı yüksekliği, sırt yüksekliği, vücut uzunluğu, göğüs derinliği, göğüs genişliği ve göğüs çevresi) kullanılmıştır. 10 katmanlı çapraz doğrulama sonunda her iki yöntem için gizil faktör sayısı iki olmuştur. Açıklanan toplam varyans KEKK ile %82,10, TB ile %80,04 ve Hata Kareler Ortalaması (HKO) sırasıyla 0,213 ve 0,230 olarak elde edilmiştir. Buna göre, kıl keçilerinde çeşitli vücut ölçüleri kullanılarak canlı ağırlığın tahmin edilmesinde toplam açıklanan varyansın daha yüksek ve HKO'sunun daha düşük olması nedeniyle KEKK, TB regresyon yönteminden daha güvenilir olduğu sonucuna ulaşılmıştır. Ancak her iki yöntemde canlı ağırlıktaki değişimi açıklamada en yüksek etkiye sahip olan değişken göğüs çevresi olarak bulunmuştur. Dolayısıyla kıl keçilerinde canlı ağırlık üzerinde yapılacak seleksiyon çalışmalarında göğüs çevresinin önemli bir kriter olduğu bu çalışmada ulaşılan diğer bir sonuçtur.

## Comparison of Partial Least Squares and Principal Component Regression Methods for Estimating Body Weight of Hair Goats Using Body Measurements

### Research Article

#### Article History:

Received: 21.11.2023

Accepted: 21.01.2024

Published online: 25.06.2024

#### Keywords:

Live weight estimation

Hair goats

Partial least squares

Key components

Body measurements

Multiple connection problem

### ABSTRACT

The aim of this study was to use Partial Least Squares (PLS) and Principal Component (PC) regression methods to estimate body weights of hair goats using various body measurements in the presence of multicollinearity problem and to interpret the results by finding the best estimation method for the study data. For this purpose, measured body weights and various body measurements (withers height, rump height, back height, body length, chest depth, chest width and chest girth) of 119 female hair goats were used. After 10-fold cross-validation, the number of latent factors for both methods was two. The total variance explained were 82.10%, 80.04%, and mean square errors

\* Bu çalışma birinci yazarın ikinci yazarın danışmanlığında hazırladığı yüksek lisans tezinden üretilmiştir.

(MSE) were 0.213 and 0.230 for PLS and PC, respectively. Accordingly, it was concluded that PLS was more reliable than PC regression method in estimating body weight by using various body measurements in hair goats because of its higher total explained variance and lower MSE. However, chest girth was found to be the variable with the highest effect in explaining the change in body weight in both methods. Therefore, it is another conclusion reached in this study that chest girth is an important criterion in selection studies on body weight in hair goats.

---

**To Cite:** Akyürek Sağır S., Akkol S. Kıl Keçilerinin Vücut Ölçülerini Kullanarak Canlı Ağırlıklarını Tahmin Etmede Kısmi En Küçük Kareler ve Temel Bileşenler Regresyon Yöntemlerinin Karşılaştırılması. *Osmaniye Korkut Ata Üniversitesi Fen Bilimleri Enstitüsü Dergisi* 2024; 7(3): 1162-1176.

## 1. Giriş

Çoklu doğrusal regresyon analizi bir bağımlı (Y) ve birden çok bağımsız değişken ( $X_1, X_2, \dots, X_n$ ) arasındaki doğrusal ilişkinin belirlenmesi için kullanılır. Birçok araştırmacı regresyon denkleminde yer alan parametre tahminleri için yaygın olarak En Küçük Kareler (EKK) yönteminden faydalanmaktadır. Çoklu doğrusal regresyon analizinde bağımsız değişkenler arasında güçlü veya güçlüye yakın ilişkiler söz konusu olduğunda bağımsızlık varsayımı ihlal edilmiş olur. Bu varsayımın ihlali, çoklu bağlantı probleminin varlığının işareti olabilir. Bu nedenle çalışma verisi çoklu bağlantı belirteçleri dikkate alınarak incelenmektedir. Zira çoklu bağlantı problemi varlığında EKK kullanılarak elde edilen tahminler güvenilirliğini yitirmektedir (Montgomery ve ark., 2001; Albayrak, 2005; Çiftsüren ve Akkol, 2018).

Çoklu bağlantı problemi çeşitli çalışma alanlarındaki uygulamada sıklıkla ortaya çıkmaktadır (Ortabaş, 2001; Çankaya ve ark., 2009; Topal ve ark., 2010; Mendeş, 2011; Yılmaz ve ark., 2020). Araştırmacılar tarafından bu problemin üstesinden gelecek güvenilir tahminler elde etmek için kullanılan alternatif yöntemler vardır. Bu yöntemlerden birkaçı temel bileşenler (TB) (Hotelling, 1933) regresyonu, Ridge regresyon (Hoerl ve ark., 1975) ve Kısmi En Küçük Kareler (KEKK) regresyonudur. Çok değişkenli istatistiksel yöntemlerden biri olan TB analizi tahmin yapma, veri indirgeme ve veri setini başka bazı analiz yöntemleri için hazırlama amaçları ile kullanılmaktadır. Bu teknik, başlangıçta birbirine bağımlı olan p değişkeni içeren bir yapının, daha az sayıda k ( $k \leq p$ ) yeni değişkene dönüştürülmesi prensibine dayanmaktadır. Bu dönüşüm sonucunda elde edilen k değişken, doğrusal, ortogonal ve birbirinden bağımsız özelliklere sahip olmaktadır. (Özdamar, 1999; Ersungur ve ark., 2007). KEKK regresyonu, çok değişkenli analizdeki çoklu bağlantı sorunlarına çözüm sunan bir yöntemdir. Bu metodoloji, başlangıçta birbirleriyle ilişkisiz ve daha küçük bir gizil değişken setini oluşturmak amacıyla bir boyut indirgeme prosedürü olarak çalışır. Daha sonra, bu türetilen gizil değişken seti, regresyon ya da sınıflama yöntemi olarak uygulanır. (Garthwaite, 1994; Abdi, 2003).

Çoklu bağlantı probleminin varlığı durumunda TB regresyon yöntemi (Graham, 2003; Albayrak, 2005; Voncina ve ark., 2005; Macciotta ve ark., 2006; Mendes, 2011; Eker, 2011) ve KEKK yöntemi (Korkmazoğlu ve Kemalbey, 2012; Oliveria ve ark., 2013; Ferrand-Calmels ve ark., 2014; Shaibu ve ark., 2015) araştırmacılar tarafından başarılı bir şekilde kullanılmıştır. EKK, TB ve Ridge

regresyonunun (RR) karşılaştırıldığı çalışmalarda TB ve Ridge regresyonun EKK yöntemine göre durağan ve güvenilir olduğu bildirilmiştir (Ortabaş, 2001; Albayrak, 2005; Topal ve ark., 2010; Abdelgadir ve Eledum, 2016; Firinguetti ve ark., 2016). KEKK ve RR yöntemlerini karşılaştırmış ve KEKK yöntemi ile en iyi sonuçlara ulaşıldığını ortaya koymuştur. Höskuldsson (1988), KEKK yönteminin maksimum indirgeme ile minimum değişken sayısını sağladığını göstermiştir. Helland (1988), TB'ye kıyasla KEKK regresyonu kullanıldığında, daha az sayıda bileşenle birlikte daha küçük hata kareleri ortalaması elde edildiğini göstermiştir. Ancak KEKK ve TB regresyon analizini karşılaştıran az sayıda çalışma mevcuttur (Godoy ve ark., 2014; Akyürek ve Akkol, 2018).

Bu çalışmanın amacı çoklu bağlantı probleminin varlığında EKK yöntemine alternatif olan KEKK ve TB regresyon yöntemleri karşılaştırmalı olarak inceleyip kıl keçilerinin canlı ağırlıklarını kullanarak canlı ağırlık tahmin etmede doğru ve güvenilir sonuçları veren yöntemin belirlenmesidir.

## 2. Materyal ve Metot

2013-2014 yılları arasında Denizli ilinde gerçekleştirilen Kıl keçisi ıslahı projesi kapsamında toplanan verilerden 4 yaşında 119 baş dişi kıl keçisine ait canlı ağırlıklar ve bazı vücut ölçüleri kullanılmıştır. Çalışmada kullanılan modellerde canlı ağırlık (CA) bağımlı değişken ve vücut ölçüleri bağımsız değişken olarak incelenmiştir. Bağımsız değişkenler cidago yüksekliği (CY), sağrı yüksekliği (SAY), sırt yüksekliği (SY), vücut uzunluğu (VU), göğüs derinliği (GD), göğüs genişliği (GG) ve göğüs çevresi (GC) ölçümlerini kapsamaktadır. İstatistik analizler için SAS (9.4) paket programından faydalanılmıştır.

### 2.1. Kısmi En Küçük Kareler Yöntemi (KEKK)

Çoklu doğrusal bir regresyon modeli, aşağıdaki eşitlik ile tanımlanır.

$$y = \beta_0 + X\beta + \varepsilon \quad (1)$$

Eşitlikte  $y$ ,  $n \times 1$  boyutlu bağımlı değişkenlerin vektörü;  $\beta_0$ , bilinmeyen sabit;  $X$ ,  $n \times p$  boyutlu  $p$  adet bağımsız değişkenler vektörü;  $\beta$ ,  $p \times 1$  boyutlu regresyon katsayıları vektörü ve  $\varepsilon$ ,  $n \times 1$  boyutlu hata vektörüdür. Hataların ortalaması sıfır ve varyansı  $\sigma^2$  olan benzer ve bağımsız normal dağılıma sahip olduğu varsayılır. Merkezileştirilen bir  $y$  vektörü 2 numaralı eşitlikte verilmiştir.

$$y = X\beta + \varepsilon \quad (2)$$

Bağımlı değişken sayısı iki ve daha fazla olduğunda denklem 3 numaralı eşitlikteki gibi olur.

$$Y = X\beta + E \quad (3)$$

Burada  $q$  bağımlı deęişken sayısıdır ve  $Y$ ,  $n \times q$  boyutlu bağımlı deęişken matrisi;  $B$ ,  $p \times q$  boyutlu bilinmeyen regresyon katsayıları matrisi ve  $E$   $n \times q$  boyutlu hata matrisidir. Çok deęişkenli KEKK regresyon yönteminde ayrıştırma 4 ve 5 numaralı eşitlikte verildięi gibi olur.

$$X = TP' + E \quad (4)$$

$$Y = UQ' + F \quad (5)$$

Eşitliklerde yer alan  $T$  ve  $U$  skor matrislerini,  $P$  ve  $Q$  yük matrislerini,  $E$  ve  $F$  ise hata matrislerini göstermektedir. KEKK regresyonda katsayılar 6 numaralı eşitlikler kullanılarak elde edilir (Phatak ve De Jong, 1997).

$$X = t_1p'_1 + t_2p'_2 + \dots + t_pp'_p = \sum_{i=1}^p t_ip'_i = TP' \quad (6)$$

Burada  $t_i$   $X$ 'in doğrusal bir kombinasyonu ve  $p_i$   $p \times 1$  boyutlu yük vektörüdür KEKK regresyonu için tahmin denklemleri aşağıda verilmiştir.

$$T_m = XR_m \quad (7)$$

$$P_m = X'T_m(T'_mT_m)^{-1} \quad (8)$$

$$R_m = W_m(P'_mW_m)^{-1} \quad (9)$$

Burada  $m$  gizil faktör sayısını gösterir. Tek deęişkenli KEKK regresyon yönteminde tahmin denklemi aşağıda verilen eşitlikteki gibi yazılır.

$$\hat{y}_{PLS}^m = T_m(T'_mT_m)^{-1}T'_m y \quad (10)$$

Burada,  $T_m = XR_m$ ,  $X'y = X'X\hat{\beta}_{OLS}$  ve  $\hat{\beta}_{OLS} = (X'X)^{-1}X'Y$  yazıldığında tahmin denklemi 11 numaralı eşitlikteki gibi olur.

$$\hat{y}_{PLS}^m = XR_m(R'_mX'XR_m)^{-1}R_mX'X\hat{\beta}_{OLS} \quad (11)$$

KEKK regresyonda katsayılar vektörü 12 numaralı eşitlik kullanılarak elde edilir.

$$\hat{\beta}_{PLS} = R_m(R'_mX'XR_m)^{-1}R'_mX'X\hat{\beta}_{OLS} \quad (12)$$

Burada  $P_m = X'XR_m(R'_mX'XR_m)^{-1}$  yazıldığında eşitlik 12 deki denklem aşağıdaki gibi 13 numaralı eşitlik olarak yeniden yazılabilir.



$$\begin{aligned} Var(Y_i) &= t_i' S t_i, & i &= 1, 2, \dots, p \\ Cov(Y_i, Y_k) &= t_i' S t_k & i, k &= 1, 2, \dots, p \end{aligned} \quad (16)$$

olur ve p adet özdeğer için,  $\lambda$ 'nın p, dereceden polinomu elde edilmektedir.

$$|S - \lambda I| = 0 \quad (17)$$

Bu polinom p adet özdeğerleri içerir. Bu özdeğerler büyüklüklerine göre  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$  şeklinde ifade edilir. Bu bileşenlerin varyansı özdeğere eşit olup bunun bir sonucu olarak varyanslar da aynı sıralamaya sahip olmaktadır. Dolayısıyla sıralamada yer alan ilk bileşenin varyansı da en büyüktür ve toplam toplam varyans içinde en büyük orana sahiptir (Tatlídil, 1996; Skrbic ve Onjia, 2007).

### 3.Bulgular ve Tartışma

Çalışmada Kıl keçilerinden alınan canlı ağırlık ve bazı vücut ölçülerine ilişkin tanımlayıcı istatistikler Tablo 1'de ve bu ölçümler arasındaki korelasyonlar, tolerans değeri (TV) ve varyans şişirme faktörü (VIF) Tablo 2'de verilmiştir. Tablo 1'de CA, GG, SAY, CY, SIY, GD, GC ve VU değişkenlerinin ortalama ve standart hataları sırasıyla 53,34±0,90, 20,08±0,18, 78,82±0,43, 81,28±0,46, 76,71±0,43, 33,84±0,20, 89,57±0,53 ve 75,23±0,37 olduğu görülmektedir. CA ve çalışmada kullanılan tüm vücut ölçülerine varyasyon katsayıları %5,41 ila %17,17 arasında değerler almaktadır.

**Tablo 1.** Kıl keçilerinde canlı ağırlık ve vücut ölçülerine ait tanımlayıcı istatistikleri

Özellikler	n	$\bar{X}$	S	$S_{\bar{X}}$	CV	Minimum	Maksimum
CA	119	57,34	9,85	0,90	17,17	38,00	86,45
GG	119	20,08	1,96	0,18	9,77	15,00	25,00
SAY	119	78,82	4,68	0,43	5,94	64,00	89,00
CY	119	81,28	5,00	0,46	6,16	67,00	95,00
SIY	119	76,71	4,70	0,43	6,13	62,00	88,00
GD	119	33,84	2,13	0,20	6,29	28,00	39,00
GC	119	89,57	5,75	0,53	6,42	75,30	106,00
VU	119	75,23	4,07	0,37	5,41	65,00	86,00

CA: Canlı ağırlık, GG: Göğüs genişliği, SAY: Sağrı yüksekliği, CY: Cidago yüksekliği, SIY: sırt yüksekliği, GD: Göğüs derinliği, GC: Göğüs çevresi, VU: vücut uzunluğu, n: gözlem sayısı,  $\bar{X}$ : Ortalama, S: Standart sapma,  $S_{\bar{X}}$ : Standart hata, CV: Varyasyon katsayısı

**Tablo 2.** Kıl keçilerinde canlı ağırlık ve vücut ölçüleri arasındaki korelasyonlar, TV ve VIF değerleri

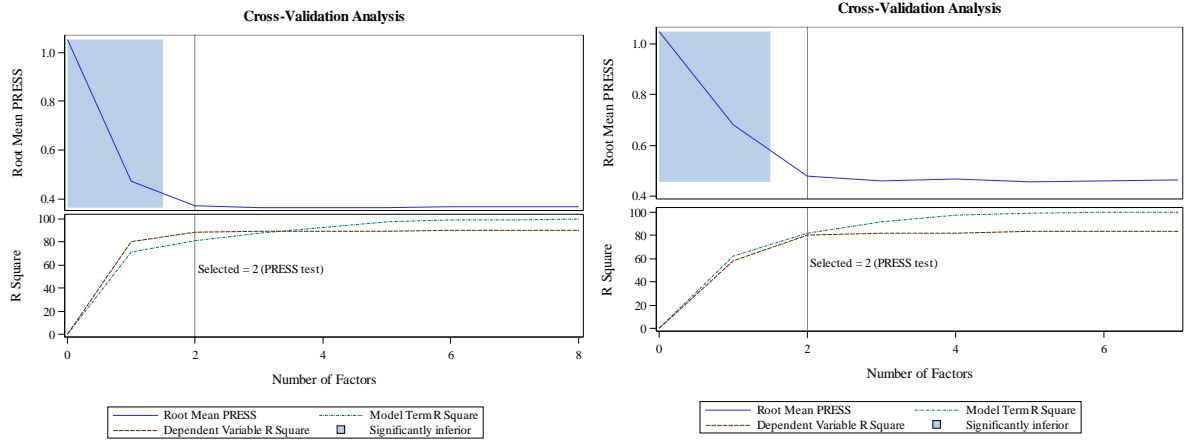
	CA	GG	SAY	CY	SIY	GD	GC	VU	TV	VIF
CA	1									
GG	0,543 ***	1							0,641	1,561
SAY	0,455 ***	0,318 ***	1						0,033	30,410
CY	0,474 ***	0,309 ***	0,955 ***	1					0,074	13,502
SIY	0,428 ***	0,311 ***	0,980 ***	0,949 ***	1				0,035	28,944
GD	0,815 ***	0,458 ***	0,466 ***	0,475 ***	0,443 ***	1			0,248	4,038
GC	0,883 ***	0,591 ***	0,484 ***	0,495 ***	0,474 ***	0,854 ***	1		0,211	4,733
VU	0,722 ***	0,328 ***	0,497 ***	0,546 ***	0,454 ***	0,638 ***	0,617 ***	1	0,461	2,169

CA: Canlı ağırlık, GG: Göğüs genişliği, SAY: Sağrı yüksekliği, CY: Cidago yüksekliği, SIY: sırt yüksekliği, GD: Göğüs derinliği, GC: Göğüs çevresi, VU: vücut uzunluğu, TV: Tolerans değeri, VIF: Varyans şişirme faktörü. \*\*\*: P<0,01

Tablo 2’de, CA ve incelenen tüm özellikler arasındaki Pearson korelasyon katsayılarının pozitif ve anlamlı olduğu ( $p<0,01$ ) görülmektedir. Bu sonuçlar literatür ile uyumlu olmuştur (Ünalın ve Ceyhan, 2017; Karadağ ve Soysal, 2018). CA ile incelenen vücut ölçümleri arasındaki en yüksek korelasyon GC arasında olup bu değer %88,3 olarak bulunmuştur. Bu bulgu, daha önce yapılmış çalışmaların sonuçları ile tutarlı olmuştur (Pesmen ve Yardımcı, 2008; Cam ve ark., 2010; Tsegaye ve ark., 2013; Das ve Yadav, 2015; Sam ve ark., 2016; Akyürek ve Akkol, 2018). Tüm vücut ölçüleri arasında pozitif ve anlamlı korelasyonlar tespit edilmiştir. Özellikle CY ile SAY, SIY ile SAY, SIY ile CY ve CG ile GD arasındaki ilişki miktarı sırasıyla %95,5, %98,0, %94,9 ve %85,4 olup güçlü korelasyonları işaret etmektedir. Bu güçlü ilişkiler nedeniyle cevap değişkenlerinin bağımsızlığı varsayımı tartışılmalı hale geldiği için (Montgomery ve ark., 2001; Albayrak, 2005; Akyürek ve Akkol, 2018) çalışma verisinde çoklu bağlantı varlığı araştırılmıştır. Bu amaçla çoklu bağlantı belirleyicilerinden VIF ve TV değerleri Tablo 2’de sunulmuştur. Buna göre bazı değişkenlerin 10’ dan büyük VIF değerlerine sahip olduğu tespit edilmiştir. Bu değişkenler ve VIF değerleri, SAY için VIF=30,410 CY için VIF= 13,502 ve SIY için VIF=28,944 elde edilmiştir. Aynı değişkenler için TV değerleri sırasıyla 0,033, 0,074, 0,034 olup sıfıra yakın olduğu tespit edilmiştir. Morfolojik özelliklerden SAY, CY ve SIY ve için VIF>10 ve TV<0,10 olması nedeniyle çalışma verisinde çoklu bağlantı sorunu olduğu sonucuna ulaşılmıştır (Marquardt ve

Snee, 1975; Belsley, 1991; Alpar, 2011; Akyürek ve Akkol, 2018; Sahin ve ark., 2018; Iqbal ve ark., 2019).

Çalışma verisinde çoklu bağlantı varlığı uygun analiz tekniklerinin kullanılmasını zorunlu hale getirmektedir. Zira çoklu bağlantı varlığında regresyon katsayılarının standart hataları büyüdüğü ve sonuçların güvenilir olmadığı bildirilmektedir (Montgomery ve ark., 2001; Albayrak, 2005; Topal ve ark., 2010). Bu çalışmada çoklu bağlantı probleminin üstesinden gelebilen yöntemlerden KEKK ve TB regresyon yöntemleri kullanılmıştır. 10 katmanlı çapraz doğrulama yapılarak elde edilen optimum gizil faktör sayılarının seçimine ilişkin görsel Şekil 1’de gösterilmiştir. Görselde KEKK ve TB için gizil faktör sayısının 2 olarak seçildiği izlenmektedir. Literatürde boyut indirgeme konusunda KEKK’nin TB’den daha etkin olduğu bildirilmiş olsa da (Maitra ve Yan, 2008; Akyürek ve Akkol, 2018) çalışmamızda her iki yöntem boyut indirgeme konusunda benzer sonuç üretmiştir. Bu bulgu, Helland (1988)’ın yaptığı çalışmanın sonuçları ile uyumlu olmuştur. Zira Helland KEKK regresyon yönteminin TB’den daha az sayıda gizil faktöre sahip olacağını bildirmiştir.



Şekil 1. KEKK ve TB regresyon analizinde çapraz doğrulama ve gizil faktör seçimi

Gizil faktör sayısının seçimine ilişkin HKO ve açıklanan varyans oranları Tablo 3’te sunulmuştur. Buna göre iki gizil faktörlü KEKK regresyon yönteminde hata kareler ortalaması 0,213 ve iki gizil faktörlü TB’de 0,230 olmuştur. HKO bakımından KEKK regresyonda TB’e göre bir miktar daha az olduğu söylenebilir. Gizil faktör sayısında herhangi bir kısıtlama yapılmadığı durumlarda KEKK ve TB tarafından elde edilen tahmin hatalarının birbirlerinden çok farklı olmadığı bildirilen literatür ile bu çalışmanın bulguları uyumlu olmuştur (Wentzell ve Montoto, 2003). KEKK yönteminde birinci gizil faktör ile toplam varyans büyük oranda açıklanmış olup bu miktar %70,12 olup ikinci gizil faktör toplam varyansın %11,38’ini açıklamaktadır. Dolayısıyla iki gizil faktörlü KEKK yöntemi tarafından açıklanan toplam varyans %82,10 olmuştur (Tablo 3). TB regresyon analizi sonuçlarına göre birinci gizil faktör ile toplam varyansın % 58,54’ü, ikinci gizil faktör ile % 21,50’si ve toplamda %80,04’ü açıklanmıştır. İki yöntem HKO ve varyans açıklama oranları bakımından karşılaştırıldığında daha küçük HKO ve daha yüksek varyans açıklama oranına sahip KEKK yönteminin TB’den daha güvenilir sonuçlar ürettiği



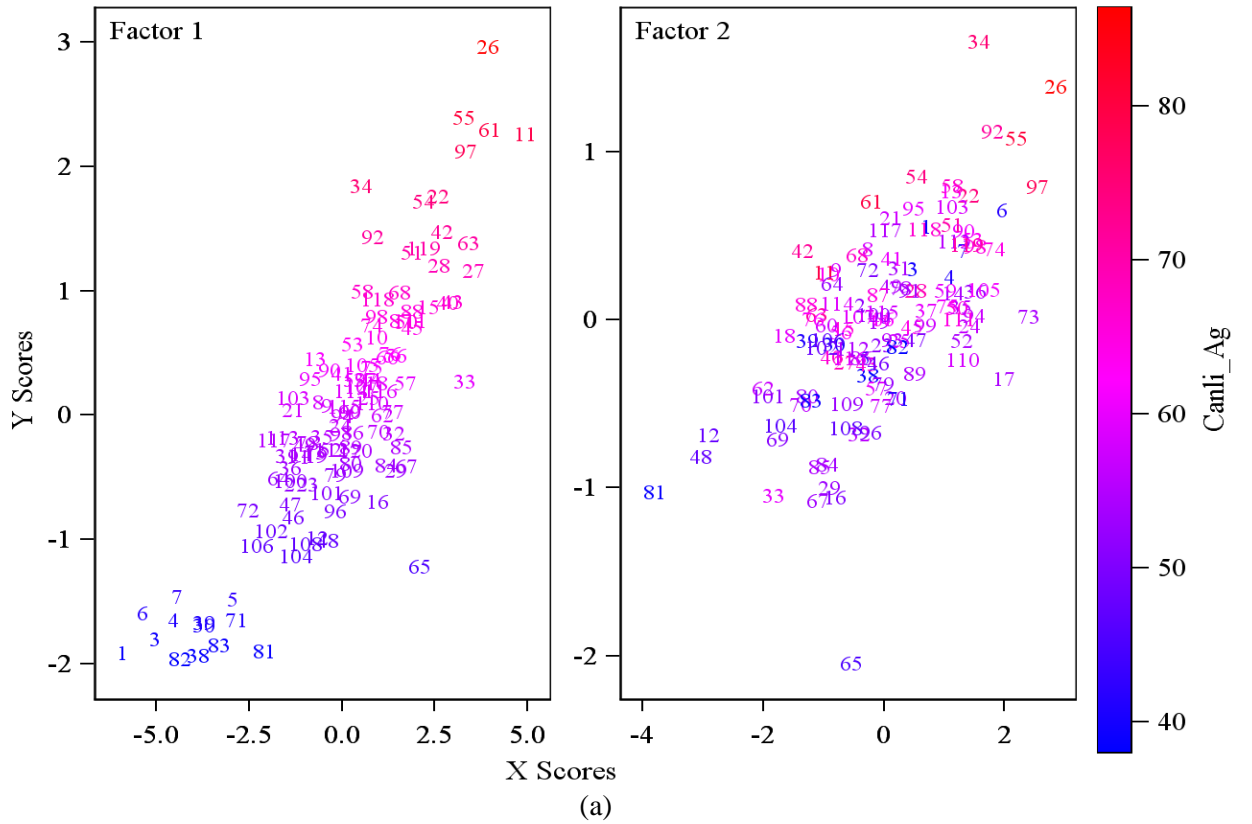
sonucuna varılmıştır. Bu bulgu literatürle uyumlu bulunmuştur (Helland, 1988; Godoy ve ark., 2014; Akyürek ve Akkol, 2018).

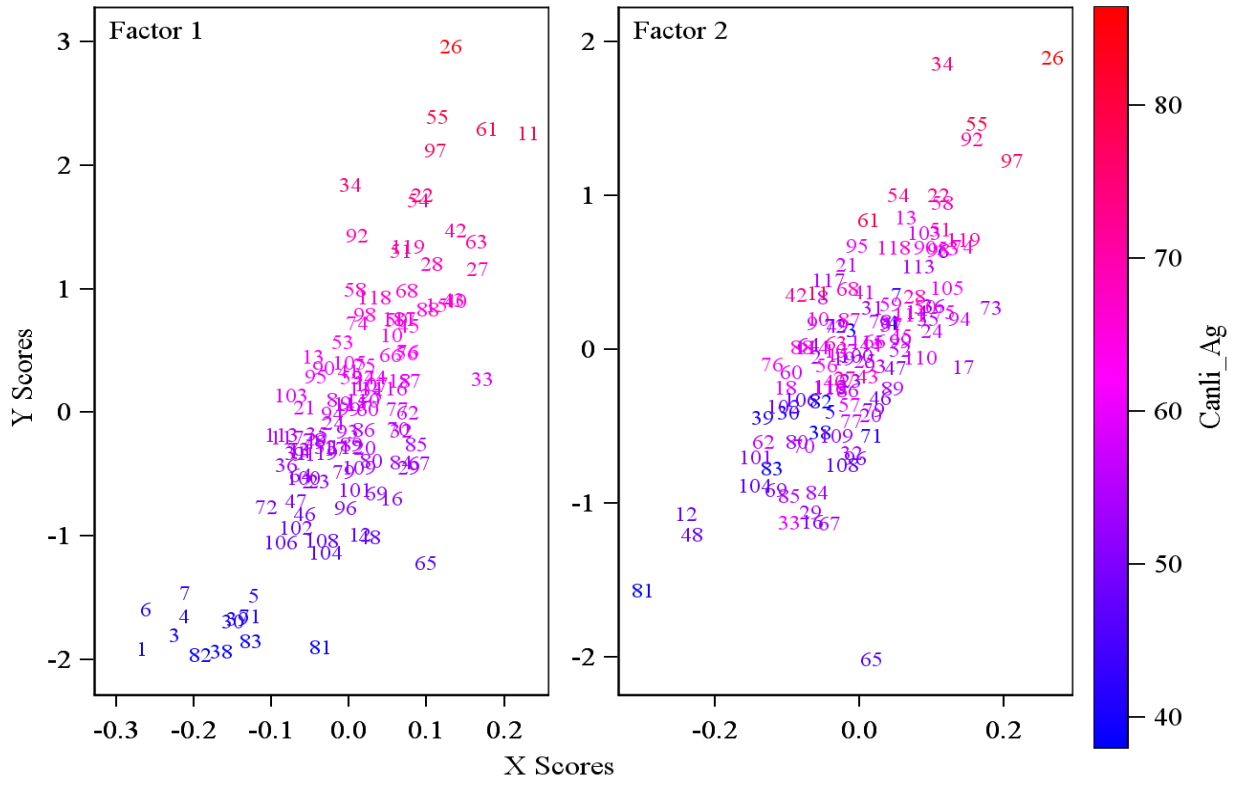
**Tablo 3.** KEKK ve TB regresyon için model uyum kriteri ve gizil faktör sayıları

Yöntem	Gizil faktör sayısı	HKO	$\sqrt{HKO}$	Açıklanan varyans	Toplam açıklanan varyans
KEKK	1	0,332	0,576	70,715	70,715
	2	0,213	0,462	11,384	82,099
TB	1	0,463	0,680	58,540	58,540
	2	0,230	0,479	21,498	80,038

*HKO*: Hata kareleri ortalaması,  $\sqrt{HKO}$ : Hata kareleri ortalamasının karekökü

CA ile gizil faktörde yer alan açıklayıcı değişkenler arasındaki korelasyon görselleri KEKK ve TB regresyonu için sırasıyla Şekil 2 (a) ve (b)'de verildi. Her iki yöntemde de birinci gizil faktörde CA ile modeldeki açıklayıcı değişkenler arasındaki ilişkinin daha güçlü olup ikinci gizil faktörde bu ilişkinin azaldığı görülmüştür. Zira her iki yöntemin birinci gizil faktörü tarafından açıklanan varyans oranları ikinci gizil faktörlerden daha yüksektir (Tablo 3). Şekilde TB analizine ait ikinci gizil faktördeki korelasyon KEKK'ye göre daha büyük olduğu gözlenmiştir. Bu durum TB'de ikinci gizil faktörün KEKK yönteminden daha fazla varyasyonu açıklaması bulgusuyla uyumludur.





(b)

**Şekil 2.** KEKK (a) ve TB (b) regresyonu için gizil faktörlerin korelasyon grafiği

KEKK ve TB regresyon yöntemlerinin gerçek gözlemlere yakınlığını değerlendirmek amacıyla Kıl keçilerinin ölçülen CA değerleri ile her iki yöntemden elde edilen tahmin değerleri arasındaki korelasyonlar ve tamamlayıcı istatistiklerle birlikte Tablo 4’de sunulmuştur. Her iki yöntemde ortalama CA değerinin aynı olup gözlenen CA değerine eşit olduğu görülmektedir. KEKK’ de standart sapma değeri TB’den daha yüksek ve gözlenen değerlerden daha düşük olmuştur. Gözlenen değerler ile KEKK yöntemi kullanılarak yapılan tahmin değerleri arasındaki korelasyon %90,6 iken TB ile olan korelasyon %89,5 olmuştur. Bu sonuç, çalışma verisini KEKK regresyon yönteminin TB’den daha iyi açıkladığını gösteren diğer bir bulgudur.

**Tablo 4.** Gözlenen CA ile KEKK ve TB ile tahmin edilen CA arasındaki korelasyonlar

	N	Ortalama	Std. sapma	Minimum	Maksimum	CA	p
CA	119	57,345	9,847	38,000	86,450	1,000	
CA ile KEKK	119	57,345	8,922	34,740	80,598	0,906	***
CA ile TB	119	57,345	8,810	35,229	80,889	0,895	***

\*\*\*:P<0,001

Çalışma verisine KEKK ve TB regresyon analizleri uygulandıktan sonra tahmin edilen katsayılar ile standardize edilmiş katsayılar Tablo 4’de sunulmuştur. Her iki yöntemde de CA’daki değişimi açıklamada en yüksek katkıya sahip vücut ölçüsünün göğüs çevresi (GC) olduğu sonucuna varılmıştır. Elde edilen bu bulgu önceki çalışmalarla uyumlu olmuştur (Pesmen and Yardımcı, 2008; Cam ve ark.,

2010; Tsegaye ve ark., 2013; Das ve Yadav, 2015; Sam ve ark., 2016; Akkol, 2018). KEKK’de en düşük etkiye sahip vücut ölçüsü CY olurken TB’de SAY olmuştur.

**Tablo 5.** KEKK ve TB regresyon için elde edilen regresyon katsayıları

Özellikler	KEKK		TB	
	Std. $\beta$	$\beta$	Std. $\beta$	$\beta$
Sabit	0,0	-101,441	0,0	-98,670
GG	0,173	0,867	0,260	1,303
SAY	-0,015	-0,031	-0,003	-0,007
CY	-0,003	-0,005	0,006	0,011
SIY	-0,031	-0,065	-0,013	-0,027
GD	0,304	1,408	0,297	1,374
GC	0,356	0,610	0,308	0,527
VU	0,258	0,624	0,208	0,503

Std.  $\beta$ : Standardize edilmiş regresyon katsayısı;  $\beta$ : regresyon katsayısı.

#### 4.Sonuç

Çoklu doğrusal regresyon analizinde EKK yöntemi uygulama ve yorumlamadaki kolaylığı nedeniyle tercih edilen bir yöntem olmuştur. EKK yönteminin uygulanabilmesi için geçerli olan varsayımlardan biri açıklayıcı değişkenlerin bağımsızlığı varsayımının sağlanamaması durumunda çoklu bağlantı problemi ortaya çıkmaktadır. Güvenilir sonuçların elde edilebilmesi için çoklu bağlantı problemini dikkate alan yöntemlerin kullanılması önerilmektedir. Bu araştırmada kıl keçilerinde çeşitli vücut ölçüleri kullanılarak canlı ağırlık tahmini yapılmıştır. Çoklu bağlantı problemi tespit edilen çalışma verisi için bu problemi dikkate alan KEKK ve TB regresyon yöntemleri kullanılmıştır. HKO, varyans açıklama oranları ve gerçek ile tahmin edilen canlı ağırlıklar arasındaki korelasyonlar dikkate alındığında KEKK yönteminin TB regresyon yönteminden daha güvenilir sonuçlar ürettiği sonucuna varılmıştır. Bu çalışmada hayvancılık alanında çoklu bağlantı probleminin varlığı tespit edilen verilerin analiz edilmesinde KEKK regresyon yönteminin sıklıkla kullanılan TB regresyon analizine iyi bir alternatif olduğu ortaya koyulmuştur. Bununla birlikte canlı ağırlıktaki değişimi açıklamada en yüksek etkiye sahip olan değişken her iki yöntemde de göğüs çevresi (GC) olarak belirlenmiştir. Dolayısıyla kıl keçilerinde canlı ağırlık üzerinde yapılacak seleksiyon çalışmalarında GC'nin önemli bir kriter olduğu bu çalışmada ulaşılan bir diğer sonuçtur.

#### Teşekkür

Bu çalışmayı FYL-2018- 6792 proje numarası ile destekleyen Van Yüzüncü Yıl Üniversitesi Bilimsel Araştırma Projeleri Başkanlığı'na desteklerinden dolayı teşekkür ederiz.

#### Çıkar Çatışması Beyanı

Makale yazarları aralarında herhangi bir çıkar çatışması olmadığını beyan eder.

## **Arařtırmacıların Katkı Oranı Beyan Özeti**

Yazarlar makaleye eşit oranda katkı sağlamış olduklarını beyan eder.

### **Kaynakça**

- Abdelgadir GA., Eledum HA. Comparison study of ridge regression and principal component regression with application. *International Journal of Research* 2016; 3(8): 283.
- Abdi H. Partial least square regression (PLS) regression. *Encyclopedia for Research Methods For The Social Sciences* 2003; 6(4): 792-795.
- Akyürek S., Akkol S. Yumurta iç kalite özelliklerinin kısmi en küçük kareler regresyonu kullanılarak tahmin edilmesi. *Yuzuncu Yıl University Journal of Agricultural Sciences* 2018; 28(4): 473-481.
- Akkol S. The prediction of live weight of hair goats through penalized regression methods: LASSO and adaptive LASSO. *Archives Animal Breeding* 2018; 61(4): 451-458
- Albayrak AS. Çoklu doğrusal bağlantı halinde en küçük kareler tekniğinin alternatifi yanlı tahmin teknikleri ve bir uygulama. *Sosyal Bilimler Dergisi* 2005; 1(1): 106-126.
- Alpar R. Uygulamalı çok değişkenli istatistiksel yöntemler. Ankara: Detay yayıncılık; 2011.
- Alpar R. Uygulamalı çok değişkenli istatistiksel yöntemlere giriş. Ankara: Bağırhan yayınevi; 1997.
- Belsley DA. Conditioning diagnostics: Collinearity and weak data in regression. Wiley-interscience. 1st edition; 1991.
- Cam MA., Olfaz M., Soydan E. Body measurements reflect body weights and carcass yields in karayaka sheep. *Asian Journal of Animal and Veterinary Advances* 2010; 5(2): 120-127.
- Çankaya S., Altop A., Kul E., Erener G. Faktör analiz skorları kullanılarak karayaka kuzuları da canlı ağırlık tahmini. *Anadolu Tarım Bilimleri Dergisi* 2009; 24(2): 98-102.
- Çiftsüren MN., Akkol S. Prediction of internal egg quality characteristics and variable selection using regularization methods: Ridge, LASSO and elastic net. *Archives Animal Breeding* 2018; 61(3): 279-284.
- Das SK., Yadav BPS. Productive performance of Jamunapari goat under semi intensive system of management. *Indian Journal of Animal Research*. 2015; 49(2): 254-257.
- Eker S. Regresyon analizinde çoklu bağlantı durumunda en küçük kareler: Ridge regresyon ve temel bileşenler yaklaşımlarının karşılaştırılması. Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü Yüksek Lisans Tezi sayfa: 11-36 Samsun, Türkiye, 2011.
- Ersungur Ş., Kızıltan M., Özgür AP. Türkiye'de bölgelerin sosyo-ekonomik gelişmişlik sıralaması: Temel bileşenler analizi. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi* 2007; 21(2): 55-66.
- Ferrand-Calmels M., Palhière I., Brochard M., Leray O., Astruc JM., Aurel MR., Douguet M. Prediction of fatty acid profiles in cow, ewe, and goat milk by mid-infrared spectrometry. *Journal of Dairy Science* 2014; 97(1): 17-35.

- Firinguetti L., Kibria G., Araya R. Study of partial least squares and ridge regression methods. *Journal Communications in Statistics Simulation and Computation* 2016; 46(8): 6631-6644.
- Garthwaite PH. An interpretation of partial least squares. *Journal of the American Statistical Association* 1994; 89(425): 122-127.
- Godoy JL., Vega JR., Marchetti JL. Relationships between PCA and PLS regression. *Chemometrics and Intelligent Laboratory Systems* 2014;130: 182-191.
- Graham MH. Confronting multicollinearity in ecological multiple regression. *Ecology* 2003; 84(11): 2809–2815.
- Helland IS. On the structure of partial least squares regression. *Journal Commun Statist Simula* 1988;17(2): 581-607.
- Hoerl AE., Kennard RW., Baldwin KF. Ridge regression: Some simulations. *Communication in Statistics* 1975; 4(2): 105-123.
- Hotelling H. Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology* 1933; 24(7): 498-520.
- Höskuldsson A. PLS regression methods, *Journal of Chemometrics* 1988; 2(3): 211- 228.
- Iqbal F., Ali M., Huma ZE., Raziq A. Predicting live body weight of Harnai sheep through penalized regression models. *Journal of Animal and Plant Sciences* 2019; 29(6).
- Karadağ O., Soysal Mİ. Honamlı keçilerinin bazı, döl verimi, büyüme ve morfolojik özelliklerinin belirlenmesi. *Tekirdağ Ziraat Fakültesi Dergisi* 2018; 15(1): 135-142.
- Kemalbay G., Korkmazoğlu OB. Effect of multicollinearity on electricity consumption forecasting using partial least squares regression. *Procedia- Social and Behavioral Sciences* 2012; 62: 1150-1154.
- Ludwig JA., Reynolds JF. *Statistical ecology*. USA A Wiley-Interscience publication. 1988.
- Macciotta NP., Dimauro C., Bacciu N., Fresi P., Cappio-Borlino A. Use of a partial least-squares regression model to predict test day of milk, fat and protein yields in dairy goats. *Animal Science* 2006; 82(04): 463-468.
- Maitra S., Yan J. Principal component analysis and partial least squares: Two dimension reduction techniques for regression. *Casualty Actuarial Society Discussion Paper Program* 2008; 79-90.
- Marquardt DW., Snee RD. Ridge regression in practice. *The American Statistician* 1975; 29(1): 3-20.
- Mendeş M. Multivariate multiple regression analysis based on principal component scores to study relationships between some pre- and post-slaughter traits of broilers. *Journal of Agricultural Sciences* 2011; 17(1): 77-83.
- Montgomery DC., Peck EA., Vining GG. *Introduction to linear regression analysis*. New York. John Wiley-Sons 3rd Edition; 2001.
- Morrison DF. *Multivariate Statistical Methods*. New York. McGraw-Hill Book Company; 1967.
- Oliveira LZ., Arruda RP., Andrade AFC., Celeghini ECC., Reeb PD., Martins JPN., Santos RFGP., Beletti ME., Peres GFR., Monteiro FM. Assessment of in vitro sperm characteristics and their

- importance in the prediction of conception rate in a bovine timed-AI program. *Animal Reproduction Science* 2013; 137(3-4): 145-155.
- Ortabaş N. Principal components in the problem of multicollinearity. Dokuz Eylül Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi 45s. İzmir, Türkiye, 2001.
- Özdamar K. Paket Program ile İstatistiksel Veri Analizi. Eskişehir: Kaan Kitapevi; 1999.
- Pesmen G., Yardımcı ME., Kocatepe A. Estimating the live weight using some body measurements in Saanen goats. *Archiva Zootechnica* 2008; 11(4): 30-40.
- Phatak A., De Jong S. The geometry of partial least squares. *Journal of Chemometrics* 1997; 11(4): 311–338.
- Sahin M., Yavuz E., Uckardes F. Multicollinearity problem and bias estimates in japanese quail. *Pakistan Journal of Zoology* 2018; 50(2).
- Sam I., Ekpo J., Ukpanah U., Eyoh G., Warri M. Relationship between linear body measurement and live body weight in west african dwarf goats in obio akpa. *Journal of Biology, Agriculture and Healthcare* 2016; 6(16): 118.
- Shaibu AS., Adnan AA. Predicting grain yield of maize using drought tolerance traits. *African Journal of Agricultural Research* 2015; 10(33): 3332-3337.
- Shrestha S., Kazama F. Assessment of surface water quality using multivariate statistical techniques a case study of the fuji river basin japan. *Environmental Modelling and Software* 2007; 22: 464-475.
- Skirbić B., Onje A. Multivariate analysis of microelement content in wheat cultivated in Serbia. *Food Control* 2007;18: 338-345.
- Tatlıdil H. Uygulamalı çok değişkenli istatistiksel analiz. Ankara: Akademi Matbaası; 1996.
- Topal M., Eydurhan E., Yağanoğlu AM., Sönmez AY., Keskin S. Çoklu doğrusal bağlantı durumunda ridge ve temel bileşenler regresyon analiz yöntemlerinin kullanımı. *Atatürk Üniversitesi Ziraat Fakültesi Dergisi* 2010; 41(1): 53-57.
- Tsegaye S., Melesse A., Banerjee S., Lakew A., Mersha F., Hailemariam F., Makebo T. Variations in linear body measurements and establishing prediction equations for live weight of indigenous sheep populations of southern Ethiopia. *Scientific Journal of Animal Science* 2013; 2(1): 15-25.
- Ünalın A., Ceyhan A. Kilis keçilerinin canlı ağırlık ve bazı vücut ölçüleri üzerinde cinsiyet etkisinin belirlenmesi. *Harran Tarım ve Gıda Bilimleri Dergisi* 2017; 21(2) : 219-226.
- Voncina DB., Kodba ZC., Novic M. Multivariate data analysis in classification of vegetable oils characterized by the content of fatty acids. *Chemometrics and Intelligent Laboratory Systems* 2005; 75(1): 31-43.
- Wentzell., PD., Montoto LV. Comparison of principal components regression and partial least squares regression through generic simulations of complex mixtures. *Chemometrics and Intelligent Laboratory Systems* 2003; 65(2): 257-270.

Widodo A., Yang BS., Han T. Combination of independent analysis and support vector machines for intelligent faults diagnosis of induction motors. *Expert Systems with Applications* 2007; (32): 299-312.

Yılmaz F., Bayyurt L., Abacı SH., Tahtalı Y. Comparison of least squares and some bias estimators in multicollinearity. *Türk Tarım Gıda Bilim ve Teknoloji Dergisi* 2020; 8(3): 793-799.