

Araştırma makalesi/Research Article (Original Paper)

Yumurta İç Kalite Özelliklerinin Kısmi En küçük Kareler Regresyonu Kullanılarak Tahmin Edilmesi

Seda AKYÜREK¹ Suna AKKOL^{2*}

¹Van Yüzüncü Yıl Üniversitesi, Fen Bilimleri Enstitüsü, Zootekni ABD, Van, Türkiye

²Van Yüzüncü Yıl Üniversitesi, Ziraat Fakültesi, Zootekni Bölümü, Van, Türkiye

*e-posta: sgakkol@yyu.edu.tr

Öz: Bu çalışma, yumurta dış kalite özellikleri kullanılarak iç kalite özellikleri olan yumurta ak ve sarı ağırlığı için bir tahmin modeli elde etmek amacıyla yapılmıştır. Bu amaçla, çalışmada kullanılan veri setine kısmi en küçük kareler regresyon yöntemi uygulanmış ve elde edilen sonuçlar temel bileşenler regresyon yöntemi ile karşılaştırılmıştır. Yumurta ak ve sarı ağırlığı için kısmi en küçük kareler regresyon analizinde gizil faktör sayısı bir ve belirleme katsayıları sırasıyla % 68.44 ve % 63.40 olmuştur. Yumurta ak ve sarı ağırlığı için bir faktörlü temel bileşenler regresyonu için belirleme katsayısı sırasıyla % 63.40 ve %53.80 olarak elde edilmiştir. Temel bileşenler regresyonunda faktör sayısı için kısıtlama olmadığı durumda, yumurta ak ağırlığı için gizil faktör sayısı beş ve belirleme katsayısı % 79.77; yumurta sarı ağırlığı için bu değerler sırasıyla iki ve % 75.35 olmuştur. Elde edilen bu sonuçlar, boyut indirgeme konusunda kısmi en küçük kareler regresyon yönteminin temel bileşenler regresyon yönteminden daha etkin olduğunu ve çoklu bağlantıya sahip küçük örnek setlerinde daha güvenilir tahminler elde edilebileceğini ortaya koymuştur.

Anahtar kelimeler: Çoklu bağlantı, En küçük kareler, Kısmi en küçük kareler, Temel bileşenler

Prediction of Inner Quality Characteristics of Eggs Using Partial Least Squares Regression

Abstract: This study was carried out to obtain a prediction model for egg albumen and yolk weight, which are the internal quality characteristics of egg predicted from external quality characteristics of egg. For this purpose partial least squares regression method was applied to the data set used in the study and the results were compared with the principal component regression method. In the partial least squares regression analysis for egg albumen and yolk weight, the number of latent factor was 1 and the determination coefficients were 68.44% and 63.40%, respectively. For the egg albumen weight, the coefficients of determination for the principal component regression with one latent factor were 63.40% and 53.80%. When there is no restriction for the number of factors in the principal component regression, for the egg albumen weight the number of latent factors was five and the coefficients of determination was 79.77%; for the egg yolk weight the values were two and 75.35%, respectively. The results shown that the partial least squares regression method was more effective than the principal component regression method in dimension reduction, and more reliable predictions can be obtained in small sample sets with multicollinearity using the partial least squares regression method.

Anahtar kelimeler: Multicollinearity, Ordinary least square, Partial least square, Principal component

Giriş

Çoklu doğrusal regresyon modelinde parametre tahminleri sıklıkla En Küçük Kareler (EKK) yöntemi kullanılarak elde edilir. Bu yöntemde tahminlerin elde edilebilmesi, açıklayıcı değişkenlerin kovaryans matrisinin tersinin hesaplanması ile mümkündür. Bu nedenle açıklayıcı değişkenler için bağımsızlık varsayımının yerine getirilmesi gerekir. Uygulamada açıklayıcı değişkenler arasında güçlü veya güçlüye yakın ilişki söz konusu olmakta ve bunun bir sonucu olarak çoklu bağlantı sorunu ortaya çıkmaktadır. Açıklayıcı değişkenler arasında çoklu bağlantı sorunu, EKK ile elde edilen tahminlerin işaretinin yanlış ve varyanslarının büyük hesaplanmasına neden olabilir. Modelin anlamlı çıkmasına rağmen, değişkenler istatistiksel olarak anlamlı olmayabilir. Bunun bir sonucu olarak EKK yöntemi kullanılarak elde edilen tahminler güvenilirliğini yitirmiş olur (Montgomery ve ark. 2001; Albayrak 2005).

Çoklu bağlantı sorununun üstesinden gelmek için yanlı tahmin yöntemleri kullanılmaktadır. Ridge Regresyon (RR) (Hoerl ve Kennard 1970), temel bileşenler (PC) (Hotelling 1933), en küçük mutlak büzülme ve seçim operatörü (Lasso) (Tibshirani 1996), elastic net (Zou ve Hastie,2005) ve kısmi en küçük kareler (PLS) (Wold 1966; Wold ve ark. 1984) regresyon yöntemleri bunlardan birkaçıdır. RR sadece tahmin yaparken Lasso ve elastic net RR'den farklı olarak hem tahmin hem değişken seçimini bir arada yapan yöntemlerdir (Tibshirani 1996; Zou and Hastie 2005). PC

regresyonu (PCR), orjinal değişkenlerden aralarında korelasyon bulunmayan ve daha az sayıda olan orjinal değişkenlerin doğrusal bileşenlerinin elde edilmesi esasına dayanır. Aynı zamanda bu yöntem, değişken sayısının azalmasına bir diğer ifadeyle değişken seçimine imkan vermez (Tatlıdil 1996). PLS regresyonu (PLSR), PCR gibi orjinal değişkenlerden doğrusal bileşenler elde ederken aynı zamanda değişken seçimi yapmaya da izin veren bir yöntemdir (Boulesteix 2004; SAS 2014). 1960'lı yılların sonlarında, Wold tarafından geliştirilen bu yöntem kısa süre içinde diğer çalışma alanlarında kullanılan istatistiksel bir araç haline almıştır (Macciotta ve ark. 2006; Maitra ve Yan 2008; Carrascal ve ark. 2009; Dimauro ve ark. 2011). PLSR temel bileşenler analizi ve çoklu regresyon analizinin özelliklerini birleştiren ve genelleştiren bir tekniktir (Abdi 2003). PLSR yönteminin, özellikle açıklayıcı değişkenler arasında yüksek korelasyon olması ve/veya açıklayıcı değişkenlerin sayısının (p) gözlem sayısına (n) benzer veya çok fazla olması durumunda ($p \gg n$) kullanışlı olduğu gösterilmiştir (Dimauro ve ark. 2011; Acharjee ve ark. 2013; Liu ve Li 2017). Çoklu doğrusal regresyon analizinde bilinen yöntemlerin her bir değişken için 10 katı kadar gözlemin olmasına ihtiyaç duymasından dolayı, bu özelliklere sahip veri setleri için bilinen EKK yöntemi kullanılarak çoklu doğrusal regresyon analizinin yapılması mümkün değildir. Chin ve Newsted (1999) örnek büyüklüğü ile ilgili olarak, gözlem sayısının 20 kadar küçük olması ve değişken sayısının dört veya sekiz olması durumunda PLS yaklaşımı ile tahminlerin elde edilebileceğini ve bu tahminlerin basit eklemeli regresyondan çok "daha iyi" olduğunu bildirmişlerdir. Ayrıca "daha iyi" kelimesini, tahmin edilen değer ile elde gerçek değerlerin birbirine yakınlığı olarak tanımlamışlardır.

Yumurta iç ve dış kalite özelliklerinin incelendiği araştırmalarda, regresyon modelinde yer alan açıklayıcı değişkenler arasında yüksek korelasyonların olduğu ortaya koyulmuştur (Özçelik 2002; Farooq ve ark. 2003; Kul ve Seker 2004; Abanikannda ve ark. 2007; Alkan ve ark. 2010; Üçkardeş ve ark. 2012). Bununla birlikte, yumurta iç kalite özelliklerini tahmin etmek amacıyla yapılan çalışmalarda açıklayıcı değişkenlerde ortaya çıkan çoklu bağlantı sorununu dikkate alan RR (Üçkardeş ve ark. 2012; Sahin ve ark. 2018), PCR (Rathert ve ark. 2011; Feddern ve ark. 2017; Ukwu ve ark. 2017; Sahin ve ark. 2018), Lasso ve elastic net (Çiftsüren ve Akkol 2018) yöntemleri çeşitli araştırmacılar tarafından kullanılmıştır. Ancak bu alanda PLSR yönteminin kullanımına ilişkin bir bilgi mevcut değildir. Bu çalışmada, yumurta dış kalite özellikleri kullanılarak yumurta iç kalite özelliklerinden olan yumurta ak ve sarı ağırlığının PLSR yöntemi ile tahmin edilmesi ve sonuçların PCR yöntemi ile karşılaştırılarak çoklu bağlantı durumunda söz konusu veri kümesi için uygun tahmin modelinin belirlenmesi amaçlanmıştır.

Materyal ve Yöntem

Materyal

Çalışmanın materyalini, Van Yüzüncü Yıl Üniversitesi Araştırma ve Uygulama Çiftliği'nde yetiştirilmekte olan Japon bıldırcımlardan alınmış 20 yumurta oluşturmaktadır.

Yöntem

PLS regresyonu, veri setindeki çoklu bağlantı sorununun üstesinden gelen çok değişkenli bir analiz yöntemidir. PLSR öncelikle birbiri ile ilişkisiz daha küçük gizil değişken seti oluşturan bir boyut indirgeme prosedürü gibi çalışır ve daha sonra regresyon veya sınıflama yöntemi olarak uygulanır (Garthwaite 1994; Abdi 2003).

Doğrusal bir regresyon modeli, aşağıdaki eşitlik ile tanımlanır.

$$y = \beta_0 + X\beta + \varepsilon \quad (2.1)$$

Eşitlikte y , $nx1$ boyutlu bağımlı değişkenlerin vektörü; β_0 , bilinmeyen sabit; X , $n \times p$ boyutlu p adet açıklayıcı değişkene ait gözlem vektörü; β , $p \times 1$ boyutlu bilinmeyen regresyon katsayıları vektörü ve ε , $nx1$ boyutlu ortalaması sıfır ve varyansı σ^2 olan benzer ve bağımsız normal dağılımlı hata vektörüdür. y vektörü merkezleştirildiğinde 2.1 numaralı eşitlik 2.2 numaralı eşitlikteki gibi yazılabilir.

$$y = X\beta + \varepsilon \quad (2.2)$$

Bağımlı değişken sayısının birden fazla olması durumunda 2.2 numaralı eşitlik aşağıdaki gibi ifade edilebilir.

$$Y = XB + E \quad (2.3)$$

Yukarıda verilen 2.3 numaralı eşitlikte, q bağımlı değişken sayısı olmak üzere Y , nxq boyutlu bağımlı değişken matrisi; B , $p \times q$ boyutlu bilinmeyen parametre matrisi ve E nxq boyutlu her bir sırasının benzer, bağımsız ve çok değişkenli normal dağılıma sahip hata matrisidir. Çok değişkenli PLSR yönteminde standardize edilmiş X ve Y matrisleri ayrıştırma genel olarak aşağıda verilen eşitlikler yardımı ile yapılır.

$$X = TP' + E \quad (2.4)$$

$$Y = UQ' + F \quad (2.5)$$

Burada, T ve U skor matrislerini, P ve Q yük matrislerini, E ve F ise hata matrislerini göstermektedir. PLSR katsayılarını elde etmek için aşağıdaki eşitlikler kullanılmaktadır (Phatak ve De Jong 1997).

$$X = t_1 p_1' + t_2 p_2' + \dots + t_p p_p' = \sum_{i=1}^p t_i p_i' = TP' \quad (2.6)$$

Burada t_i X 'in doğrusal bir kombinasyonudur; P_i , $p \times 1$ boyutlu yük vektörüdür. NIPALS algoritmasında t_i , hata matrisinin (E_i) doğrusal kombinasyonu olarak varsayılır. Aşağıda verilen eşitlikler kullanılarak PLSR için tahmin denklemleri elde edilir.

$$T_m = XR_m \quad (2.7)$$

$$P_m = X' T_m (T_m' T_m)^{-1} \quad (2.8)$$

$$R_m = W_m (P_m' W_m)^{-1} \quad (2.9)$$

Gizil faktör sayısı m ile ifade edilir. Tek değişkenli PLSR yöntemine göre,

$$\hat{y}_{PLS}^m = T_m (T_m' T_m)^{-1} T_m' y \quad (2.10)$$

yazılır ve 2.10 numaralı eşitlikte $T_m = XR_m$, $X' y = X' X \hat{\beta}_{OLS}$ yazıldığında aşağıdaki eşitlik elde edilir. Burada $\hat{\beta}_{OLS} = (X' X)^{-1} X' Y$ olmaktadır.

$$\hat{y}_{PLS}^m = XR_m (R_m' X' XR_m)^{-1} R_m' X' X \hat{\beta}_{OLS} \quad (2.11)$$

PLS regresyonda katsayılar vektörünün elde edilmesi için aşağıda verilen eşitlik kullanılır.

$$\hat{\beta}_{PLS} = R_m (R_m' X' XR_m)^{-1} R_m' X' X \hat{\beta}_{OLS} \quad (2.12)$$

Burada $P_m = X' XR_m (R_m' X' XR_m)^{-1}$ yazıldığında 2.12 numaralı eşitliğin denk formu 2.13 numaralı eşitlik aşağıdaki gibi olur.

$$\hat{\beta}_{PLS} = R_m P_m \hat{\beta}_{OLS} = W_m (P_m' W_m)^{-1} P_m' \hat{\beta}_{OLS} \quad (2.13)$$

Çalışmada kullanılan yumurtaların ak ağırlığı (AKAG) ve sarı ağırlığı (SARAG) regresyon modellerinde bağımlı değişken olarak yer almıştır. Açıklayıcı değişkenler olarak yumurta ağırlığı (YUMAG), yumurta eni (YUMEN), yumurta boyu (YUMBOY), yumurtanın yaş kabuk ağırlığı (YKABKAG), kuru kabuk ağırlığı (KKABKAG), küt ucunun kalınlığı (KLNK_kut), ortasının kalınlığı (KLNK_ort) ve sivri ucunun kalınlığı (KLNK_siv) modele dahil edilmiştir. AKAG, SARAG, YUMAG, YKABKAG ve KKABKAG gr, YUMEN ve YUMBOY mm, KLNK_kut, KLNK_ort ve KLNK_siv mikro metre cinsinden ölçülmüştür. Bağımlı değişken olarak incelenen AKAG ve SARAG'da meydana gelen değişimin ayrı ayrı incelenmesi nedeni ile çok değişkenli bir model için tahmin eşitliklerine yer verilmemiştir. Eşitlikler için detaylar, Garthwaite (1994) ve Phatak ve De Jong (1997)'nin çalışmalarında bulunabilir.

İstatistik analizler, SAS/STAT (SAS 2014) istatistik paket programında PLS prosedürü kullanılarak NIPALS (Wold 1975) algoritması ile yapılmıştır. Uygun gizil faktör sayısı 5 katmanlı çapraz doğrulama ile belirlenmiştir.

Bulgular ve Tartışma

Çalışmada kullanılan yumurta iç ve dış kalite özelliklerine ilişkin korelasyon matrisi, ortalama ve standart hata değerleri Tablo 1'de sunulmuştur. Tablo 1'de açıklayıcı değişkenler arasında özellikle YUMAG ile YUMEN ve YUMBOY arasındaki korelasyonlar (sırasıyla 0.86 ve 0.91) ve YUMEN ile YUMBOY arasındaki korelasyonun (0.67) yüksek tahmin edildiği görülmektedir. Bu nedenle açıklayıcı değişkenler arasındaki bağımsızlık varsayımı, varyans şişirme faktörü (VIF) ve koşul indeksi (CI) dikkate alınarak araştırılmıştır. VIF değerlerinin 1.76-28.42 arasında değiştiği ve en yüksek CI değerinin 445.37 olduğu belirlenmiştir. Modelde açıklayıcı değişken olarak yer alan yumurta dış kalite özellikleri arasında VIF>10 ve CI> 30 olması, açıklayıcı değişkenler arasında çoklu bağlantı sorununun olduğunu göstermektedir (Marquardt ve Snee 1975; Belsley 1991). Rathert ve ark. (2011), Çiftsüren ve Akkol (2018) ve Sahin ve ark. (2018) tarafından yapılan çalışmalarda da, yumurta dış kalite özelliklerine ilişkin değişkenlerde çoklu bağlantı sorunu olduğu ortaya konulmuştur.

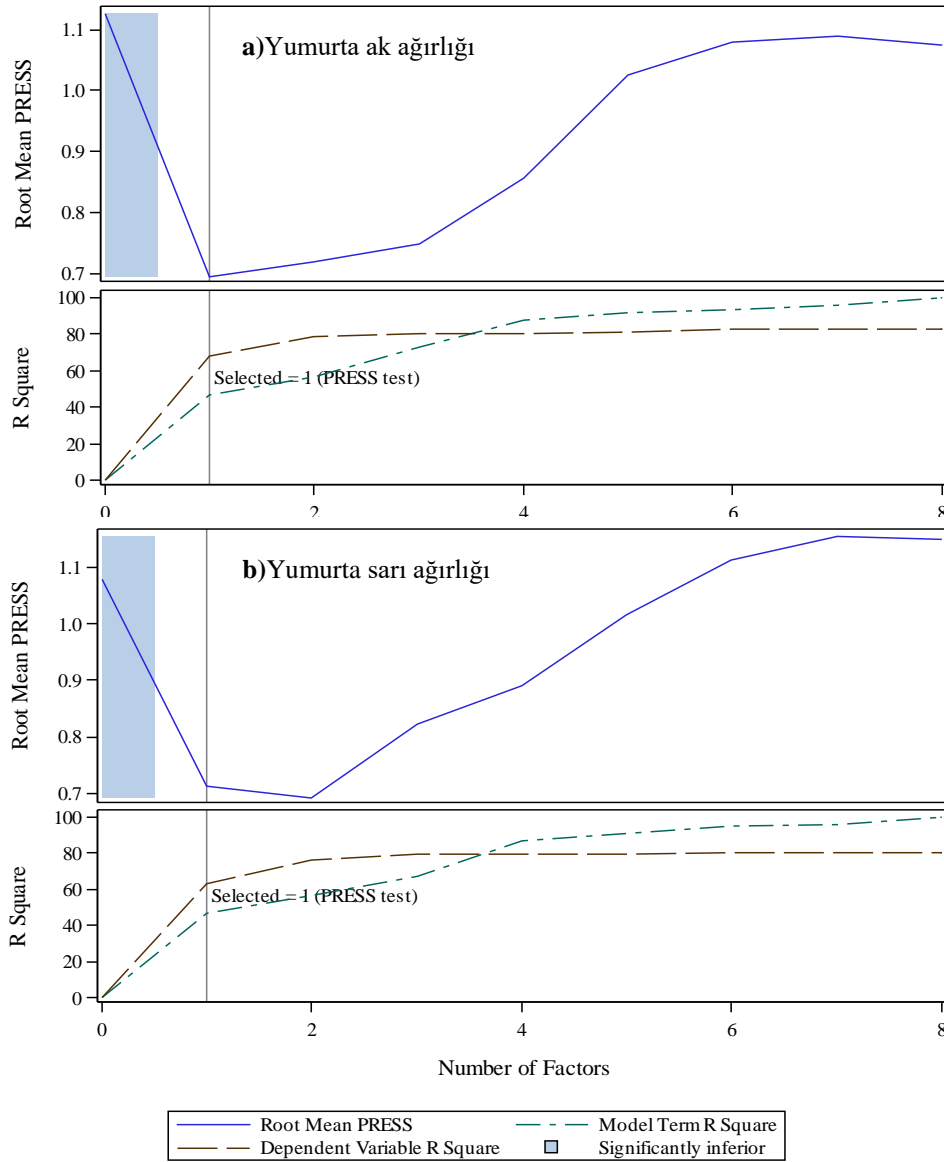
Çizelge 1. Değişkenlere ilişkin korelasyon matrisi, ortalama (\bar{X}) ve standart hata ($S_{\bar{x}}$) değerleri

Değişkenler	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	\bar{X}	$S_{\bar{x}}$
AKAG	0.86**	0.77**	0.77**	0.35	0.60*	-0.20	-0.22	0.00	5.92	0.62
SARIAG	0.84**	0.73**	0.78**	0.34	0.35	-0.28	-0.22	0.27	4.00	0.58
YUMAG(X ₁)		0.86**	0.91**	0.51*	0.61*	-0.27	-0.26	0.19	11.49	1.09
YUMEN (X ₂)			0.67**	0.43	0.41	-0.25	-0.08	0.09	32.69	1.26
YUMBOY (X ₃)				0.55*	0.58*	-0.28	-0.20	0.22	25.53	0.88
YKABKAG (X ₄)					0.61*	-0.08	-0.18	0.35	1.57	0.15
KKABKAG (X ₅)						-0.11	-0.30	0.21	0.87	0.08
KLNK_kut (X ₆)							0.48*	0.49*	0.22	0.01
KLNK_ort (X ₇)								0.46*	0.21	0.01
KLNK_siv (X ₈)									0.20	0.01

*: p<0.05, **: p<0.01.

Çalışmada yumurta ak ve sarı ağırlığının her biri için birbirinden bağımsız olacak şekilde açıklayıcı değişken seçilerek performanslarının değerlendirilmesi planlanmıştır. Bu nedenle yumurta ak ve sarı ağırlığı için veriye tek değişkenli PLSR analizi uygulanmıştır. Modellerdeki uygun gizil faktör sayısı belirlenirken 5 katmanlı çapraz doğrulama yapılmıştır. Hem yumurta ak ağırlığı hem de sarı ağırlığı için gizil faktör sayısı bir olduğu, bir diğer ifade ile bir doğrusal kombinasyona sahip olduğu belirlenmiş ve bu bilgiye ait görsel sırasıyla Şekil 1a ve 1b'de verilmiştir. Şekilde, tahminlerin hata kareler ortalama (Root Mean PRESS) değerinin en küçük olduğu nokta, uygun gizil faktör sayısını göstermektedir.

Yumurta ak ve sarı ağırlığı için bir faktörlü PLS regresyon analizi kullanılarak tahmin edilen standardize edilmiş regresyon, gerçek regresyon katsayıları ve Wold (1994) istatistiği (VIP) Tablo 2'de verilmiştir. VIP istatistiği her bir açıklayıcı değişkenin modele yaptığı katkıyı özetlemektedir. Bir açıklayıcı değişkenin modelden çıkarılması için VIP istatistiği ile birlikte standardize edilmiş katsayılar mutlak değer bakımından değerlendirilir. Mutlak olarak küçük katsayılar aynı zamanda VIP<0.8 koşulunu sağladığında modelden çıkarılacak aday değişken olarak ifade edilmektedir (SAS 2014). Yumurta ak ağırlığı için Tablo 2 incelendiğinde, YKABKAG, KLNK_kut, KLNK_ort ve KLNK_siv değişkenlerine ait VIP değerinin 0.8'den küçük olduğu görülmektedir. Bunlara ilişkin standardize edilmiş katsayılar negatif işaretli olup bu değerler sırasıyla 0.096, 0.056, 0.060 ve 0.001 elde edilmiştir. Buna göre değişken seçimi yaparak daha az karmaşık bir model üzerinde çalışmak söz konusu olduğunda, KLNK_kut, KLNK_ort ve KLNK_siv için elde edilen katsayıların da küçük olması nedeni ile modelden çıkarılabilirken YKABAG'nin katsayısının mutlak olarak diğerlerinden daha büyük olması nedeni ile modelde tutulabilir.

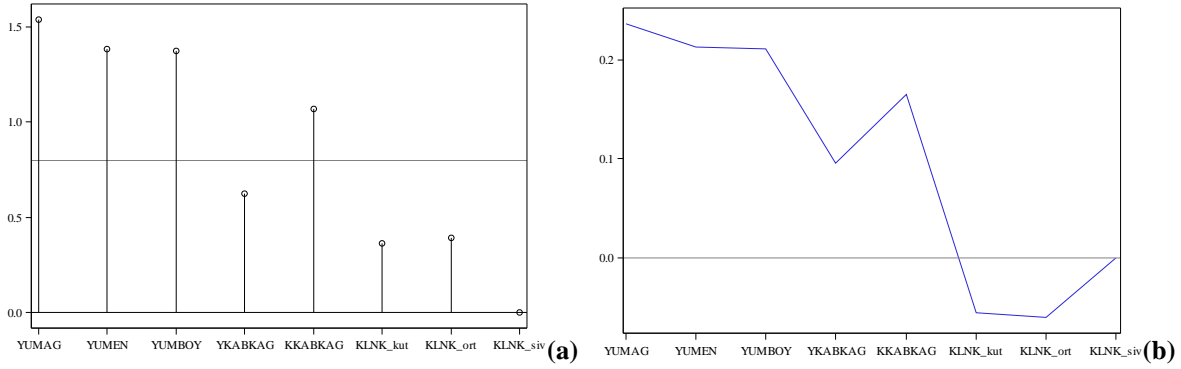


Şekil 1. Çapraz doğrulama analizi ile a) yumurta ak ağırlığı ve b) yumurta sarı ağırlığı için seçilen gizil faktör sayısının belirlenmesi

Çizelge 2. Yumurta ak ve sarı ağırlığı için tahmin edilen regresyon katsayıları ve VIP değerleri

Değişkenler	Yumurta ak ağırlığı			Yumurta sarı ağırlığı				
	Std. katsayıları	regresyon	Regresyon katsayıları	VIP	Std. katsayıları	regresyon	Regresyon katsayıları	VIP
Sabit	0.000		-2.586	-	0.000		-4.174	-
YUMAG	0.236		0.134	>0.8	0.234		0.124	>0.8
YUMEN	0.212		0.104	>0.8	0.203		0.093	>0.8
YUMBOY	0.211		0.149	>0.8	0.218		0.143	>0.8
YKABKAG	0.096		0.406	<0.8	0.094		0.372	<0.8
KKABKAG	0.165		1.203	>0.8	0.097		0.662	<0.8
KLNK_kut	-0.056		-4.266	<0.8	-0.077		-5.584	<0.8
KLNK_ort	-0.060		-4.662	<0.8	-0.062		-4.451	<0.8
KNLK_siv	-0.0001		-0.005	<0.8	0.076		5.396	<0.8

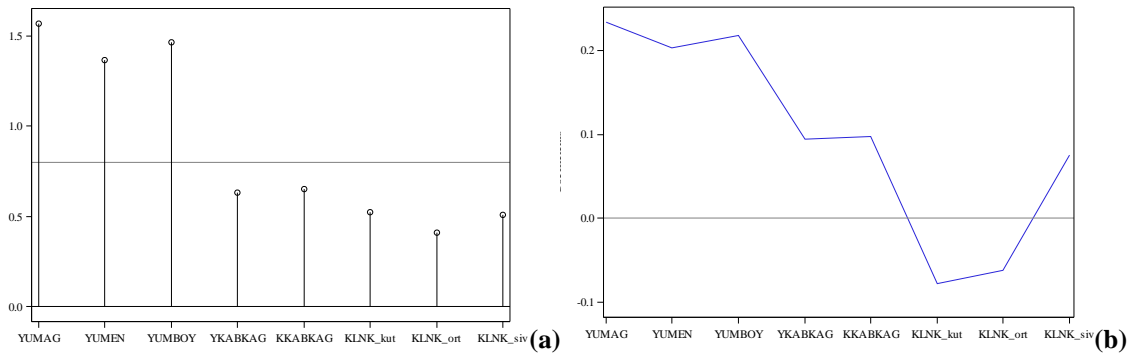
Std.: Standardize edilmiş, YUMAG: Yumurta ağırlığı, YUMEN: Yumurta eni, YUMBOY: Yumurta boyu, YKABKAG: Yaş kabuk ağırlığı, KKABKAG: Kuru kabuk ağırlığı, KLNK_kut: Küt uç kabuk kalınlığı, KLNK_ort: Orta kabuk kalınlığı, KLNK_siv: Sivri uç kabuk kalınlığı



Şekil 2. Yumurta ak ağırlığı için, **a)** açıklayıcı değişkenlerin önem grafiği (VIP) ve **b)** standardize edilmiş katsayıların profili.

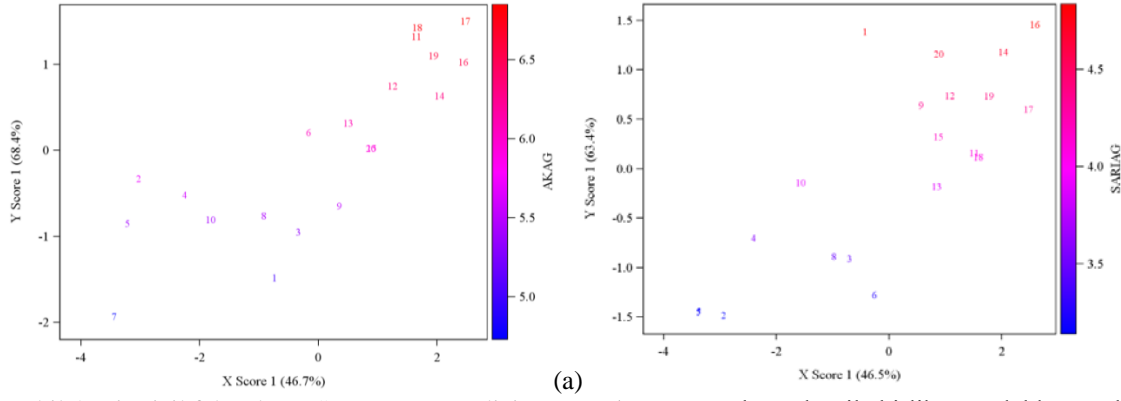
Yumurta ak ağırlığı için Şekil 2a'da VIP değerleri ve 2b'de standardize edilmiş regresyon katsayıları yer almaktadır. Şekil 2a'da X eksenine paralel olan çizgi VIP=0.8 olduğu noktadır. Buna göre Şekil 2a'da bu kritik değerinin altında ve üstünde olan açıklayıcı değişkenler görülürken Şekil 2b'de katsayıların mutlak olarak büyüklükleri görülmektedir.

Tablo 2, yumurta sarı ağırlığı için VIP istatistiği bakımından değerlendirildiğinde modelden çıkarılabilecek değişkenlerin YKABKAG, YKABKAG, KLNK_kut, KLNK_ort ve KLNK_siv olduğu söylenebilir. Bu bulguya ilişkin görsel Şekil 3a'da verilmiştir. YUMAG, YUMEN ve YUMBOY dışındaki tüm değişkenler 0.8 değerinin altında kalmıştır. Modelden çıkarılacak değişkenlere karar vermeden önce Tablo 2'deki standardize edilmiş katsayılar incelenmiş ve en küçük katsayıların (mutlak olarak) KLNK_kut, KLNK_ort ve KLNK_siv değişkenlerine ait olduğu belirlenmiştir. Yumurta ak ağırlığında olduğu gibi, KLNK_kut, KLNK_ort ve KLNK_siv değişkenlerinin modele katkısının çok küçük olduğu ve değişken seçimi yapılmak istenildiği zaman modelden ilk çıkarılması gereken değişkenler olduğu sonucuna varılmıştır. Şekil 3b'de katsayıların mutlak büyüklüklerini gösteren grafik verilmiştir. Modelden uzaklaştırılmasının uygun olabileceği düşünülen bu değişkenlerin, sıfıra en yakın katsayıya sahip olan değişkenler olduğu görülmektedir (Şekil 3b).



Şekil 3. Yumurta sarı ağırlığı için **a)** açıklayıcı değişkenlerin önem grafiği (VIP) ve **b)** standardize edilmiş katsayıların profili.

Bir gizil faktör için X ve Y skorlarının birbirleri ile ilişkisi, açıklayıcı (X eksen) ve bağımlı değişkenlerdeki (Y eksen) açıklama miktarları sırasıyla Şekil 4a ve 4b'de verilmiştir. Buna göre, bir faktörlü PLSR analizi ile yumurta ak ağırlığında meydana gelen değişimin % 68.4'ü açıklanırken, sarı ağırlığındaki değişimin %63.4'ü açıklanmaktadır. Açıklayıcı değişkenlerdeki açıklama miktarı ise yumurta ak ve sarı ağırlığı için birbirlerine oldukça yakın bulunmuştur, sırasıyla % 46.7 ve %46.5.



Şekil 4. Bir gizil faktörlü PLS regresyon analizine göre **a**) yumurta ak ağırlığı ile birlikte modelde yer alan açıklayıcı değişkenlerde meydana gelen açıklama miktarı ve **b**) yumurta sarı ağırlığı ile birlikte modelde yer alan açıklayıcı değişkenlerde meydana gelen açıklama miktarı.

PLSR analizinden sonra verilere PCR analizi uygulanmıştır. Çapraz doğrulama yardımı ile PLSR modelinde yumurta ak ve sarı ağırlığı için bir gizil faktör seçilirken PCR modelinde yumurta ak ağırlığı için beş, yumurta sarı ağırlığı için iki gizil faktör seçilmiştir (Tablo 3). Özellikle yumurta ak ağırlığında PLSR yönteminin PCR yönteminden daha az sayıda gizil faktör seçmesi, PLSR yönteminin boyut indirgeme konusunda PCR yönteminden daha etkin olduğunu ortaya koymaktadır. Elde edilen bu bulgu, Maitra ve Yan (2008) yaptıkları çalışmanın sonuçları ile uyumlu olmuştur.

PLSR ve PCR yöntemlerinin karşılaştırılması amacı ile PCR modelinde faktör sayısı bir olacak şekilde kısıtlama yapılmış ve yapılan tüm analizlere ilişkin belirleme katsayıları ve gerçek değerler ile tahmin edilen değerler arasındaki korelasyonlar Tablo 3'te verilmiştir. Yumurta ak ağırlığı için bir gizil faktörlü PLSR modeline ait belirleme katsayısı %68.44 ve korelasyon değeri yaklaşık olarak 0.83 (%82.73) iken PCR modeli için bu değerler sırasıyla %63.40 ve 0.80 (%79.62) olmuştur. Yumurta sarı ağırlığı için bir gizil faktörlü PLSR modeli için belirleme katsayısı %63.40 ve korelasyon değeri 0.80 (%79.62) iken PCR için sırasıyla %53.80 ve 0.73 (%73.35) olarak bulunmuştur. Hem yumurta sarı hem de ak ağırlığı için PLSR ve PCR modellerinden elde edilen sonuçlar birbirlerine benzer olmuştur. Daha açık bir ifade ile, her iki bağımlı değişken için PLSR modeli, PCR modelinden hem daha yüksek belirleme katsayısına hem de daha yüksek korelasyon katsayısına sahip olmuştur. Sonuçlar, çalışmada kullanılan veri için PLSR modelinin PCR modelinden daha etkin ve güvenilir tahmin yaptığını göstermektedir. Bu sonuç, PLSR'in PCR ile karşılaştırıldığında daha yüksek bir tahmin etkinliğine sahip olduğunu söyleyen Macciotta ve arkadaşlarının (2006) çalışması ile uyumlu olmuştur. El-Fallah ve El-Salam (2014) yaptıkları çalışmada, çoklu bağlantı probleminin varlığında, regresyon modelindeki değişken sayısının artmasıyla (dörtten daha fazla) birlikte, hem küçük örneklerde (n=20) hem de büyük örneklerde (n=60), PLSR ile elde edilen sonuçların, RR ve PCR ile elde edilenlerden "daha iyi" olduğunu ortaya koymuşlardır. Bu bulgu, örnek sayısının 20 olduğu bu çalışmada elde edilen PLSR analiz sonuçlarının, PCR regresyondan daha iyi olduğu sonucuyla tutarlı olmuştur.

Tablo 3. Yumurta ak ve sarı ağırlığının tahmininde PLSR ve PCR modellerine ait uyum ölçütleri

Yumurta iç kalite özellikleri	Uyum Ölçütü	PLSR	PCR	PCR
Yumurta ak ağırlığı	Gizil faktör sayısı	1	1	5
	Belirleme katsayısı (%)	68.44	63.40	79.77
	Korelasyon (%)	82.73	79.62	89.32
Yumurta sarı ağırlığı	Gizil faktör sayısı	1	1	2
	Belirleme katsayısı (%)	63.40	53.80	56.77
	Korelasyon (%)	79.62	73.35	75.35

Sonuç

Çoklu bağlantı, çoklu doğrusal regresyonda sıklıkla karşılaşılan önemli bir sorundur. Bu sorunun üstesinden gelen çeşitli yöntemler geliştirilmiştir. Bunlardan biri PLSR yöntemidir. Bu çalışmada, çoklu bağlantı durumunda PLSR yöntemi (NIPALS algoritması) kullanılmıştır. Ayrıca sonuçlar PCR ile karşılaştırılmış ve sonuçlar değerlendirilmiştir. PLSR yönteminin, boyut indirgemedeki PCR yönteminden daha başarılı olduğu ve daha etkin tahminler ürettiği ortaya koyulmuştur. Günümüzde, hayvancılık alanında yapılan çalışmalarda özellikle büyük (gözlem sayısının açıklayıcı değişken sayısından çok daha fazla olduğu) veri kümeleri, karşılaşılan bir sorun haline gelmektedir. Bu sorunun üstesinden gelebilecek yöntemlerden birinin PLSR olması, ileride yapılacak çalışmalarda bu yöntemin hem tahmin hem de boyut indirgeme konusunda kullanımının aratacağını göstermektedir. Dolayısıyla çalışmanın hayvancılık

alanında karşılaştırılması muhtemel aşırı doygun modeller ve/veya çoklu bağlantı sorunu olan veri kümeleri için RR ve PCR yöntemlerine alternatif olarak PLSR yönteminin yaygın kullanımı konusunda literatüre önemli katkı sağlayacağı düşünülmektedir.

Teşekkür

Bu çalışma Seda Akyürek'in yüksek lisans tez çalışmasının bir kısmı kullanılarak yapılmıştır. Ayrıca, FYL-2018-6792 proje numarası ile çalışmayı destekleyen Van Yüzüncü Yıl Üniversitesi Bilimsel Araştırma Projeleri Başkanlığı'na teşekkür ederiz.

Kaynaklar

- Abanikannda OTF, Olutogun O, Leigh AO, Ajayi LA (2007). Statistical modeling of egg weight and egg dimensions in commercial layers. *International Journal of Poultry Science* 6(1): 59-63.
- Abdi H (2003). Partial least square regression (PLS regression). *Encyclopedia for research methods for the social sciences* 6(4): 792-795.
- Acharjee A, Finkers R, Visser RGF, Maliepaard C (2013), Comparison of Regularized Regression Methods for ~Omics Data. *Metabolomics* vol.3, pp.126.
- Albayrak SA (2005). Çoklu bağlantı halinde en küçük kareler teknikleri ve bir uygulama, Zonguldak Kara Elmas Üniversitesi, Sosyal Bilimler Dergisi 1: 105-126.
- Alkan S, Karabağ K, Galiç A, Karşlı T, Balcıoğlu MS (2010). Effects of selection for body weight and egg production on egg quality traits in Japanese quails (*Coturnix coturnix japonica*) of different lines and relationships between these traits. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi* 16(2): 239-244.
- Belsley, DA (1991). *Conditioning Diagnostics, Collinearity and Weak Data in Regression*. John Wiley and Sons, New York, NY, USA.
- Boulesteix AL (2004). PLS dimension reduction for classification with microarray data. *Stat Appl Genet Mol Biol*, 3(1):33.
- Carrascal LM, Galván I, Gordo O (2009). Partial least squares regression as an alternative to current regression methods used in ecology. *Oikos*, 118(5), 681-690.
- Chin WW, Newsted PR (1999). Structural Equation Modeling analysis with Small Samples Using Partial Least Squares. In Rick Hoyle (Ed.), *Statistical Strategies for Small Sample Research*, Sage Publications, pp. 307-341.
- Çiftsüren MN, Akkol S (2018). Prediction of internal egg quality characteristics and variable selection using regularization methods: ridge, LASSO and elastic net. *Archives Animal Breeding*, 61(3): 279-284.
- Dimauro C, Steri R, Pintus M A, Gaspa G, Macciotta NPP (2011). Use of partial least squares regression to predict single nucleotide polymorphism marker genotypes when some animals are genotyped with a low-density panel. *Animal*, 5(6):833-837.
- El-Fallah M., El-Salam A (2014). A Note on Partial Least Squares Regression for Multicollinearity (A Comparative Study). *International Journal of Applied Science and Technology*. 4(1):163-169.
- Farooq M, Durrani FR, Sarbiland K, Chand, N (2003). Predicting egg weight, shell weight, shell thickness and hatching chick weight of Japanese quails using various egg traits as regressors. *International Journal of Poultry Science*.
- Feddern V, Prá MCD, Mores R, Nicoloso RDS, Coldebella A, Abreu PGD (2017). Egg quality assessment at different storage conditions, seasons and laying hen strains. *Ciência e Agrotecnologia*, 41(3): 322-333.
- Garthwaite PH (1994). An interpretation of partial least squares, *Journal of the American Statistical Association*, 89:122-127.
- Hoerl, AE, Kennard RW (1970). Ridge regression: biased estimation for non-orthogonal problems, *Technometrics*, 12 :55-82.
- Hotelling H (1933). Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6): 417.
- Kul S, Seker İ (2004). Phenotypic correlations between some external and internal egg quality traits in the japanese quail (*Coturnix coturnix japonica*) *International Journal of Poultry Science*3 (6): 400-405.
- Liu W, Li Q (2017). An efficient elastic net with regression coefficients method for variable selection of spectrum data. *PloS one*,12(2), e0171122.
- Macciotta NPP, Dimauro C, Bacciu N, Fresi P, Cappio-Borlino A (2006). Use of a partial least-squares regression model to predict test day of milk, fat and protein yields in dairy goats. *Animal Science*, 82(4), 463-468.
- Maitra S, Yan J (2008). *Principal Component Analysis and Partial Least Squares: Two Dimension Reduction Techniques for Regression*. Casualty Actuarial Society Discussion paper Program. 79-90
- Marquardt DW, Snee RD (1975). Ridge Regression in Praticce. *The American Statistician*, 29: 3-20.
- Montgomery DC, Peck EA, Vining GG (2001). *Introduction to Linear Regression Analysis*, 3rd Edition, John Wiley&Sons, New York.

- Özçelik M (2002). Japon bildircını yumurtalarında bazı iç ve dış kalite özellikleri arasındaki fenotipik korelasyonlar. Ankara Üniversitesi Veterinerlik Fakültesi, 49: 67-62.
- Phatak A, De Jong S (1997). The geometry of partial least squares. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 11(4) : 311-338.
- Rathert TÇ, ÜçkardeşF, Nariç D, Aksoy T (2011). Comparison of Principal Component Regression with the Least Square Method in Prediction of Internal Egg Quality Characteristics in Japanese Quails. *Kafkas Üniversitesi Veteriner Fakültesi Dergisi*, 17: 687-692.
- Sahin M, Yavuz E, Uckardes F (2018). Multicollinearity Problem and Bias Estimates in Japanese Quail. *Pakistan J. Zool.*, 50(2): 757-761.
- SAS (2014) SAS/STAT User's Guide: Version 9.4, SAS Institute Inc., Cary, NC, USA, 64.
- Tatlıdil H (1996). Uygulamalı Çok Değişkenli İstatistiksel Analiz, Cem Web Ofset Ltd. Şti., Ankara.
- Tibshirani R (1996). Regression shrinkage and selection via the lasso, *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 58: 267–288.
- Üçkardeş F, Efe E, Nariç D, Aksoy, T (2012). Japon bildircınlarında yumurta ak indeksinin ridge regresyon yöntemiyle tahmin edilmesi. *Akademik Ziraat Dergisi* 1(1): 11-20.
- Ukwu HO, Abari PO, Kuusu DJ (2017). Principal Component Analysis of Egg Quality Characteristics of Isa Brown Layer Chickens in Nigeria. *World Scientific News*, 70(2): 304-311.
- Wold H (1975). Perspectives in Probability and Statistics. In Gani J (ed). *Soft modeling by latent variables: the nonlinear iterative partial least squares approach*. London, UK: Academic Press. p. 520–540.
- Wold H (1966). Estimation of principal components and related models by iterative least squares. In: Krishnaiah, P. R., ed., *Multivariate Analysis*. New York: Academic Press.
- Wold S (1994). "PLS for Multivariate Linear Modeling," *QSAR: Chemometric Methods in Molecular Design. Methods and Principles in Medicinal Chemistry*.
- Wold S, Ruhe A, Wold H, Dunn III W (1984). The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Journal of Scientific Statistical Computing* 5:735–744.
- Zou H, Hastie T (2005). Regularization and variable selection via the elastic net. *Statistical Society: Series B*, 67: 301–320.