



doi 10.33188/vetheder.1554543

Araştırma Makalesi / Research Article

Çoklu bağlantı sorununda Ridge, Liu ve LASSO tahmin edicilerinin kullanımı: yumurta tavukçuluğunda bir uygulaması

Ayşe YENER^{1,a*}, Ali Alparslan SAYIM^{1,2,b}, Aytaç AKÇAY^{2,c}¹ Ankara Üniversitesi Sağlık Bilimleri Enstitüsü, Veterinerlik Biyoistatistik Anabilim Dalı, Ankara, Türkiye² Ankara Üniversitesi Veteriner Fakültesi, Biyoistatistik Anabilim Dalı, Ankara, TürkiyeORCID 0009-0008-1226-750X^a; 0009-0009-4299-3315^b; 0000-0001-6263-5181^cMAKALE BİLGİSİ /
ARTICLE INFORMATION:Geliş / Received:
23 Eylül 24
23 September 24Revizyon/Revised:
15 Kasım 24
15 November 24Kabul / Accepted:
26 Kasım 24
26 November 24

ÖZET

Çalışmada, çoklu bağlantı sorununda en küçük kareler yöntemine(EKK) alternatif olarak önerilen Ridge, Liu ve LASSO tahmin edici yöntemlerinin kullanımının ve model başarı kriterlerine göre sonuçların karşılaştırılması amaçlanmıştır. Çalışma materyalini, "Nick Chick" ırkı tavukların 19-100 haftalık dönemlerdeki ortalama yumurta ağırlığı (g) ve canlı ağırlık (kg) verileri ile ortalama yumurta fiyatlarına göre hesaplanan tahmini satış gelirleri oluşturmuştur. Çalışmada yaş (hafta), ortalama yumurta ağırlığı (g) ve canlı ağırlık değişkenleri ile haftalık toplam geliri tahmin eden bir modelin geliştirilmesi amacıyla, çoklu bağlantı sorunu varlığında, en küçük kareler regresyonuna (EKK) alternatif yanlı regresyon tekniklerinden Ridge, Liu ve LASSO tahmin edicileri kullanılmıştır. Haftalık toplam gelirin tahmininde, EKK, Ridge, Liu ve LASSO tahmin edicilerinde oluşturulan modellerin hesaplanan belirlenme katsayısı (R^2) sırasıyla 0.96, 0.95, 0.95, 0.97 olarak hesaplanmıştır. Ayrıca, kurulan modellere ait regresyon hata kare ortalamalarının karekökü (RMSE) EKK, Ridge, Liu ve LASSO tahmin edicilerinde, 132.3, 130.5, 56.2, 129.5 olarak belirlenmiş dolayısıyla en düşük Liu tahmin edicisi yönteminde bulunmuştur. Sonuç olarak; çalışmada uygulanan yanlı tahmin edici yöntemlerin Liu tahmin edicisinin EKK regresyonuna göre, daha düşük standart hatalı, tutarlı ve uygun tahmin sağladığı belirlenmiştir.

Anahtar Sözcükler:
Yumurta tavukçuluğu
Çoklu bağlantı
LASSO
Liu
Ridge
Keywords:
Hen industry
Multicollinearity
LASSO
Liu
Ridge

The use of Ridge, Liu and LASSO estimators in the multicollinearity problem: an application in layer hen industry

ABSTRACT

This study aims to compare the use of Ridge, Liu, and LASSO estimator methods, proposed as alternatives to Ordinary Least Squares (OLS) in the presence of multicollinearity, and to evaluate the results based on model performance criteria. The study material consists of average egg weight (g) and live weight (kg) of "Nick Chick" breed chickens during the 19–100-week period, as well as the estimated sales revenue calculated based on average egg prices. In order to develop a model that predicts weekly total revenue using the variables of age (weeks), average egg weight (g), and live weight, biased regression techniques, such as Ridge, Liu, and LASSO estimators, were employed as alternatives to OLS in the presence of multicollinearity. In predicting weekly total revenue, the calculated coefficients of determination (R^2) for the models generated using OLS, Ridge, Liu, and LASSO estimators were found to be 0.96, 0.95, 0.95, and 0.97, respectively. Additionally, the root mean square errors (RMSE) of the regression models were 132.3, 130.5, 56.2, and 129.5 for OLS, Ridge, Liu, and LASSO estimators, respectively, with the lowest error found in the Liu estimator method. As a result, it was determined that the biased estimator methods applied in the study, particularly the Liu estimator, provided more consistent, accurate, and lower-standard-error predictions compared to the OLS regression method.

©2025 The Authors.
Published by Veteriner Hekimler Derneği. This is an open access article under CC-BY-NC license. (<https://creativecommons.org/licenses/by-nc/4.0>)



How to cite this article: Yener A, Sayim AA, Akçay A. Çoklu bağlantı sorununda Ridge, Liu ve LASSO tahmin edicilerinin kullanımı: yumurta tavukçuluğunda bir uygulaması. Vet Hekim Der Derg. 2025;96(1):14-22.

* Sorumlu Yazar e-posta adresi / Corresponding Author e-mail address: aysekucuk22@gmail.com

1. Giriş

Çoklu doğrusal regresyon analizi bir bağımlı değişkenin birden fazla bağımsız değişken tarafından açıklamak amacı ile tercih edilen istatistiksel yöntemlerden biridir (1,2). Oluşturulan regresyon modelleri değişkenler arası sebep-sonuç ilişkilerini belirleyerek, geleceğe yönelik tahminlerde bulunabilmektedir.

Çoklu doğrusal regresyon analizinde model parametrelerinin kestiriminde, hata kareler ortalamasını (HKT) minimum yapan tahmin yöntemi olan en küçük kareler (EKK) kullanılmaktadır. Ancak bu yöntem bazı varsayımların (gözlemlenen verilerin normal olması, hataların beklenen değerinin sıfır olması, hataların varyansının sabit olması ve bağımsız değişkenler arasında anlamlı ilişkinin olmaması) sağlandığı durumlarda parametre kestirimlerinin yansız ve doğru bir şekilde elde edilebilmesini sağlar (3,4).

Çoklu regresyon analizinde, bağımsız değişkenler arasında ilişkilerin olması ve bağımsızlık varsayımının bozulması sık karşılaşılan sorunlardan biridir. Çoklu bağlantı sorunu (multicollinearity) olarak da adlandırılan bu durumun varlığında, parametre kestirimlerine ait varyanslar büyüdüğünden model parametreleri için gerçekte olması gerekenden farklı olacak ve tahminler kararsızlaşacaktır. Bağımsız değişkenler arasındaki korelasyonun büyük olması ($r > 0,75$), VIF (Varyans artırıcı faktör) değerinin 10'un üzerinde hesaplanması çoklu bağlantı sorunu olduğunu gösterir. Ayrıca çoklu bağlantı sorununu, değişkenlere ait özdeğerlerin en az birinin 1'den farklı olması veya 0'a yakın olması da gösterir. Bunu belirlemede, koşul indeksi (KI) olarak tanımlanan en büyük özdeğer ile en küçük özdeğer üzerine dayalı bir hesaplama önerilmiştir. Koşul indeksinin 30'dan büyük olması ileri derecede çoklu bağlantı sorunu olduğunu göstermektedir (3,5,6). Bu sorunu gidermek için daha fazla veri toplanması veya bağımsız değişkenlerden bir veya birkaçı modelden çıkarılması önerilmiştir (7,8). Ancak bu yöntemlerin bağımsız değişkenler arasında ciddi çoklu doğrusal bağlantı sorunu durumunda etkili sonuçlar vermediği de bildirilmiştir (9).

Çoklu doğrusal bağlantı sorununu gidermeye yönelik mevcut gözlem ve değişkenlerle, EKK tahmincisine alternatif daha başarılı tahminler yapmayı sağlayan yanlı regresyon tahmin edicileri önerilmiştir (11,12,14). Yanlı regresyon tahmin ediciler ile küçük varyans sahip tahminler yapılarak problemin düzeltilebildiği bildirilmiştir (5). Yansız ve yüksek varyanslı tahmin edici yerine yanlı ama düşük varyanslı bir tahmincinin kullanılmasına hata kareler ortalaması (HKO) değeri ile karar verilebilir. HKO değeri en küçük olan tahmin edicinin seçilmesi önerilir (10). Çoklu bağlantı sorununda modeldeki değişkenleri çıkarmadan regresyon katsayılarını yanlı olarak tahmin etmede en geniş uygulama alanı olan Ridge tahmin edicisi ve buna alternatif olarak geliştirilen Liu ve LASSO tahmin edicileri kullanılabilir.

Ridge tahmin edicisi, çoklu doğrusal bağlantı sorunu varlığında EKK yönteminin yetersiz kalması nedeniyle önerilmiştir (11). Ridge tahmin edicisinde varyans ve kovaryans matrisinin köşegen değerlerine 0'dan büyük olan bir yanlılık sabiti, Ridge parametresi (K) ilave edilmektedir. Bu sayede tahminlerin varyansları azaltılarak, K katsayısı ölçüsünde yanlı tahminler yapılabilmektedir (5). En uygun K değeri, yanlı standartlaştırılmış regresyon katsayıları ve K değerleri arasında çizilen Ridge eğim grafiği ile belirlenir (11). Optimal K değeri, grafikte yanlı standartlaştırılmış regresyon katsayılarının durağan olduğu bölgelerden seçilir (5).

Liu tahmin edicisi, Ridge ile Stein tipi tahmin edicisinin bir kombinasyonu olarak Liu tarafından önerilen yanlı tahmin edici olarak tanımlanmıştır. Liu tahmin edicisi bu iki tahmin edicinin dezavantajlarını ortadan kaldırarak, hem çoklu doğrusal bağlantı sorununda başarılı sonuçlar veren hem de yanlılık parametresi d 'nin doğrusal artan bir fonksiyonudur. Yanlılık parametresi $0 \leq d \leq 1$ arasında bir sabittir. Liu tahmin edicisinin üstünlüğü, d parametresinin doğrusal bir fonksiyonu olduğundan, d 'nin seçimi ridge tahmin edicisi olan K 'nin seçimine oranla oldukça basittir ve birçok yöntem önerilmiştir (12). Bu tahmin ediciler EKK ile karşılaştırılmış ve daha az hata kareler ortalaması verdiği bildirilmiştir. Optimal K değerinin sıfıra yakın seçilmesi önerilse de Ridge ile Liu tahmin edicilerinin eşitliğini sağlayacak ortak bir K ve d değeri elde edilebilir (12).

EKK, Ridge ve Liu tahmin edicilerinin çoklu bağlantı sorunu varlığında, tahmin kesinliği yüksek olmasına rağmen büyük varyansa sahip olmaları sebebiyle yetersiz kaldığı durumlar olmaktadır. Ridge ve Liu tahmin edicileri, katsayıları daraltan kararlı bir süreçtir ancak katsayılar 0'a eşitlenemediği için model yorumlama sorunu olmaktadır. Kestirim doğruluğunu elde etmek için bazen değişken seçimi (shrinkage) yapılabilir. Bu sayede, varyanslar azaltılarak genel tahmin doğruluğu artırılabilir (13). Tibshirani tarafından, tahmin ve değişken seçiminin birlikte yapıldığı bir

yöntem olarak LASSO (Least Absolute Shrinkage and Selection Operator) alternatif yanlı tahmin bir edici önerilmiştir (14). Bu yöntem, katsayıların mutlak değerinin toplamını sabitten yaparak ve artık karelerin toplamı düşürülür. LASSO tahmin edicisini diğer yöntemlerden ayıran en önemli fark değişken seçiminin yapılmasıdır (15).

LASSO tahmin edicisinde de β katsayılarının elde edilmesi için yanlılık sabiti modele dahil edilir. Ceza parametresi olarak ifade edilen bu sabit (λ) ile gösterilir, pozitif bir değerdir ve büzülme miktarını belirler. Bu tahmin edicide ceza parametresinin kullanıldığı ceza fonksiyonu modelin başarısı için önemlidir. Bu fonksiyon LASSO tahmin edicisinin çözümünü Ridge'den farklı olarak quadratik forma dönüştürür. LASSO tahmin edicisinde λ değeri artarken β katsayılarının sıfıra doğru yaklaşır, λ yeteri kadar büyük olduğunda katsayılar sıfır olur (16).

Literatürde, çoklu bağlantı sorununu ortadan kaldırabilmek amacıyla önerilen yanlı tahmin edici yöntemlerin karşılaştırıldığı çalışmalar son yıllarda artış göstermiştir. Bu çalışmalar simülasyon veri setlerinde (17-20); sağlık alanında elde edilen biyolojik veri setlerinde (21-25), hayvancılık alanındaki veri setlerinde (3,26,27) uygulanmıştır.

Bu çalışmada uygulamada hayvancılıkta çok önemli bir paya sahip olan, en hızlı gelişen ve modern teknolojinin kullanıldığı tavukçuluk sektörü verileri kullanılmıştır. Türkiye'de yumurta tavukçuluğunda çok sayıda faktöre bağlı olarak değişim gösteren ekonomik göstergeler ile verim özellikleri arasındaki sebep-sonuç ilişkisini ortaya çıkararak yapılacak olan yorumları güçlendirecektir.

Bu amaçla çalışmada, çoklu bağlantı sorununda EKK 'ye alternatif olarak önerilen Ridge, Liu ve LASSO tahmin edici yöntemlerinin yumurta tavukçuluğu verileri üzerinde kullanımının ve model başarı kriterlerine göre sonuçların karşılaştırılması yapılmıştır.

2. Gereç ve Yöntem

Çalışma materyalini, Türkiye'de yumurta tavukçuluğunda yetiştiriciliği yapılan "Nick Chick" ırkı tavukların 19-100 haftalık dönemlerdeki ortalama yumurta ağırlığı (g) ve canlı ağırlık (kg) verileri ile ortalama yumurta fiyatlarına göre hesaplanan tahmini satış gelirleri oluşturmuştur.

Haftalık ortalama yumurta ağırlığı ve canlı ağırlık verileri Nick Chick ırkı beyaz yumurtacı tavuk ırkı bakım kılavuzu kitapçığından elde edilmiştir (28).

Haftalık satış geliri, 82 hafta (19-100. haftalar arası) beyaz yumurta üretimi yaptığı varsayılan 10 bin başlık yumurta tavukçuluğu işletmesinin, haftalık üretilen ortalama yumurta sayısı ve ağırlığına göre yumurta fiyatları baz alınarak hesaplanmış, Haziran 2024 Dolar kuruna çevrilmiştir (Gelir = Yumurta sayısı x Yumurta fiyatı). Haftalık tahmini satış gelirinin belirlenmesinde, 2024 yılı Haziran ayında Yumurta Üreticileri Merkez Birliği tarafından açıklanan yumurta fiyatları kullanılmıştır (29).

Çalışmada haftalık toplam geliri bağımlı değişken (Y) olarak; yaş (hafta), ortalama yumurta ağırlığı (g) ve canlı ağırlık değişkenleri bağımsız değişkenler (X1, X2, X3) olarak alınmıştır. Haftalık toplam geliri tahmin eden bir modelin geliştirilmesi amacıyla, çoklu bağlantı sorununda EKK regresyonuna alternatif yanlı regresyon tekniklerinden Ridge, Liu ve LASSO tahmin edicileri kullanılmış ve karşılaştırmaları yapılmıştır. İstatistik analizlerde SPSS 30, RStudio 2023.03.0 ve Python 3.11.2 programları kullanılmıştır.

3. Bulgular

Çalışmada değişkenlere ait tanımlayıcı istatistikler (Tablo1) ve EKK yöntemi sonucunda haftalık toplam geliri (Y) ile bağımsız değişkenler (yaş, ortalama yumurta ağırlığı (g) ve canlı ağırlık) arasındaki doğrusal ilişki için kurulan modelin yüksek belirtme katsayısı ($R^2 = \%96$) bulunmuştur. Katsayılar tablosu incelendiğinde, yaş ve yumurta ağırlığı değişkenlerin kısmi T testine göre anlamlı olduğu, canlı ağırlık değişkeninin anlamlı olmadığı görülmüştür (Tablo 2).

Tablo 1: Değişkenlere ait tanımlayıcı istatistikler**Table 1:** Descriptive statistics of variables

| Değişkenler | n | Ortalama ± Standart Sapma | Minimum - Maksimum |
|-----------------------|----|---------------------------|--------------------|
| Yaş (hafta) | 82 | 59,5 ± 23,8 | 19 -100 |
| Yumurta ağırlığı (gr) | 82 | 61,3 ± 4,7 | 40,8 – 64,2 |
| Canlı ağırlık(kg) | 82 | 1,6 ± 0,1 | 1,4 – 1,7 |
| Gelir (\$) | 82 | 3391,5 ± 615,2 | 236,3 – 3919,3 |

Tablo 2: EKK ile tahmin edilen regresyon katsayıları**Table 2:** Regression Coefficients Estimated by OLS

| Değişkenler | β katsayısı | S(β) | T değeri | P değeri | Güven aralığı (%95) | |
|-----------------------|-------------|--------|----------|----------|---------------------|---------|
| | | | | | Min | Maks |
| Model sabiti | -9848,2 | 1871,9 | -5,3 | 0,000 | -13574,9 | -6121,6 |
| Yaş (hafta) | -28,7 | 2,2 | -13,2 | 0,000 | -33,0 | -24,4 |
| Yumurta ağırlığı (gr) | 157,4 | 22,0 | 7,1 | 0,000 | 113,5 | 201,3 |
| Canlı ağırlık(kg) | 3208,1 | 1993,4 | 1,6 | 0,102 | -760,3 | 7176,6 |
| R ² | | | | 0,96 | | |
| RMSE | | | | 132,3 | | |

β: Regresyon katsayıları, S(β): Regresyon katsayılarının standart hatası, R²: Belirtme katsayısı, RMSE: Hata kareler ortalamasının karekökü

Bağımsız değişkenler aralarındaki korelasyonlar ve hesaplanan VIF değerlerine göre çoklu bağlantı sorununun varlığından bahsedilebilir. Tüm değişkenler için hesaplanan VIF değerlerinin 10'dan büyük olduğu tespit edilmiştir (Tablo 3).

Tablo 3: Bağımsız değişkenler arasındaki korelasyonlar ve VIF değerleri**Table 3:** Correlations between independent variables and VIF values

| Değişkenler | Canlı ağırlık(kg) | Yaş (hafta) | Yumurta ağırlığı (gr) | VIF Değerleri |
|-----------------------|-------------------|-------------|-----------------------|---------------|
| Canlı ağırlık(kg) | 1 | -0,90 | -0,98 | 91,2 |
| Yaş (hafta) | -0,90 | 1 | 0,81 | 12,4 |
| Yumurta ağırlığı (gr) | -0,98 | 0,81 | 1 | 49,7 |

Korelasyon matrisinin özdeğerleri ve koşul indeksine (KI) göre özdeğerlerin sıfıra yakın çıkması, üç ve dördüncü özdeğere ait KI değerinin 30'dan yüksek hesaplanması ciddi çoklu bağlantının sorunun olduğunu göstermiştir (Tablo 4).

Çoklu bağlantı sorununda alternatif yöntemlerinden olan Ridge tahmin edicisinde yanlılık sabiti K değeri, tüm VIF değerlerinin 10'dan küçük olduğu K =0,02 olarak belirlenmiştir. Ridge tahmin edicisi ile değişkenlere ait standart hatalar, EKK regresyonuna göre daha düşük çıkmıştır. Bu tahmin edicide K =0,02 yanlılık sabiti ile regresyon katsayılarının daha durağan hale geldiği görülmüştür. Kurulan modelin belirtme katsayısı R² = 0,95 olarak hesaplanmıştır (Tablo 5).

Tablo 4: Özdeğerler ve Koşul indeksleri**Table 4:** Eigenvalues and condition indices

| No | Özdeğerler | Oran | Kümülatif Oran | KI (Koşul indeksi) |
|----|------------|-------|----------------|--------------------|
| 1 | 3,91 | 97,70 | 97,70 | 1,0 |
| 2 | 0,09 | 2,27 | 99,96 | 6,6 |
| 3 | 0,00 | 0,04 | 100,00 | 52,1 |
| 4 | 0,00 | 0,00 | 100,00 | 542,1 |

Liu tahmin edicisinde, yanlılık parametresi $d=0,01$ seçildiğinde regresyon katsayılarının standart hataları EKK'e ve Ridge'ye göre daha düşük hesaplanmıştır. Modelin belirtme katsayısı $R^2 = 0,95$ olarak hesaplanmıştır (Tablo 6).

Tablo 5: Ridge Tahmin Sonuçları ($K=0,02$)**Table 5:** Ridge Estimation Results ($K=0.02$)

| Değişkenler | β katsayısı | S(β) | t değeri | P değeri | Güven aralığı (%95) | |
|-----------------------|-------------------|--------------|----------|----------|---------------------|--------|
| | | | | | Min | Maks |
| Model sabiti | -9657,0 | | | | | |
| Yaş (hafta) | -26,9 | 228,3 | 25,2 | < 0,01 | -474,3 | 420,5 |
| Yumurta ağırlığı (gr) | 142,0 | 282,6 | 21,3 | < 0,01 | -411,8 | 695,8 |
| Canlı ağırlık(kg) | 3599,0 | 332,7 | 6,9 | < 0,01 | 2946,8 | 4251,2 |
| R^2 | | | | 0,95 | | |
| RMSE | | | | 130,5 | | |

β : Regresyon katsayıları, S(β): Regresyon katsayılarının standart hatası, R^2 : Belirtme katsayısı, RMSE: Hata kareler ortalamasının karekökü

LASSO tahmin edicisi sonuçlarına göre modelde değişken seçimi yapılmış, canlı ağırlık değişkeni model dışı bırakılmış, yaş ve yumurta ağırlığı değişkenlerinin geliri tahmin etmede en yüksek açıklama miktarına sahip değişkenler olduğu belirlenmiştir. Kurulan modelin belirtme katsayısı $R^2 = 0,97$ olarak hesaplanmıştır (Tablo 7).

Tablo 6: Liu Tahmin Sonuçları ($d=0,01$)**Table 6:** Liu Estimation Results ($d=0.01$)

| Bağımsız Değişkenler | β | S(β) | T değeri | P değeri | %95 lik Güven aralığı | |
|-----------------------|---------|--------------|----------|----------|-----------------------|-----------|
| | | | | | Alt Sınır | Üst Sınır |
| Model sabiti | -6894,6 | 307,9 | -22,4 | <0,01 | -7498,1 | -6291,2 |
| Yaş (hafta) | -25,6 | 0,9 | -27,3 | <0,01 | -27,4 | -23,7 |
| Yumurta ağırlığı (gr) | 191,3 | 4,8 | 40,3 | <0,01 | 182,0 | 200,6 |
| Canlı ağırlık(kg) | 48,0 | 28,7 | 1,7 | 0,09 | -8,2 | 104,3 |
| R^2 | | | | 0,95 | | |
| RMSE | | | | 56,2 | | |

β : Regresyon katsayıları, S(β): Regresyon katsayılarının standart hatası, R^2 : Belirtme katsayısı, RMSE: Hata kareler ortalamasının karekökü

Tablo 7: LASSO Tahmin Sonuçları**Table 7:** LASSO Estimation Results

| Bağımsız Değişkenler | β | S(β) | T değeri | P değeri | %95 lik Güven aralığı | |
|-----------------------|---------|--------------|----------|----------|-----------------------|-----------|
| | | | | | Alt Sınır | Üst Sınır |
| Model sabiti | -6258,1 | | | | | |
| Yaş (hafta) | -24,3 | 1,4 | -17,5 | <0,001 | -25,8 | -20,5 |
| Yumurta ağırlığı (gr) | 179,9 | 9,5 | 19,0 | <0,001 | 161,2 | 198,5 |
| Canlı ağırlık(kg) | | | | | | |
| R ² | | | | 0,97 | | |
| RMSE | | | | 129,5 | | |

β : Regresyon katsayıları, S(β): Regresyon katsayılarının standart hatası, R²: Belirtme katsayısı, RMSE: Hata kareler ortalamasının karekökü

Çalışma sonucunda, bir yumurta tavukçuluğu işletmesinde haftalık yumurta satışı gelirini (\$) tahmin eden değişkenlere ilişkin;

EKK regresyon modeli;

$$\text{Gelir (\$)} = -9848,2 - 28,7 * \text{yaş(hafta)} + 157,4 * \text{yumurta ağırlığı(gr)} + 3208,1 * \text{canlı ağırlık(kg)}$$

Ridge regresyon modeli;

$$\text{Gelir (\$)} = -9657 - 26,9 * \text{yaş(hafta)} + 142 * \text{yumurta ağırlığı(gr)} + 3599 * \text{canlı ağırlık(kg)}$$

Liu tahmin edicisi modeli;

$$\text{Gelir (\$)} = -6894,6 - 25,6 * \text{yaş(hafta)} + 191,3 * \text{yumurta ağırlığı(gr)} + 48 * \text{canlı ağırlık(kg)}$$

LASSO tahmin edicisi modeli;

$$\text{Gelir (\$)} = -6258,1 - 24,3 * \text{yaş(hafta)} + 179,9 * \text{yumurta ağırlığı(gr)}$$

şeklinde kurulmuştur.

4. Tartışma ve Sonuç

Hayvancılık sektöründe verim özellikleri ve ekonomik göstergeler çok sayıda faktöre bağlı olarak değişim gösterir. Bu özellikler arasındaki sebep-sonuç ilişkisini ortaya çıkarabilmek adına ekonometrik yaklaşımların kullanıldığı çalışmalar son yıllarda önem kazanmıştır. Maliyet unsurlarının belirlenerek gelir veya kar tahminlerinin yapılması, değişkenler arasındaki ilişkilerin doğru olarak modellenmesi, üreticiye üretim planlaması aşamasında katkı sağlayacak ve yapılacak olan yorumları güçlendirecektir. Hayvancılık verilerinde uygulanan modelleme çalışmalarında çoklu bağlantı sorunu ile sıklıkla karşılaşılabilir. Modelin tahmin performansını iyileştirmek ve çoklu bağlantı sorunun olumsuz etkilerini azaltmak amacıyla literatürde en yaygın olarak kullanılan yöntemlerin başında Ridge, Liu ve LASSO tahmin edicileri gelmektedir. Bunların yanında Kısmi en küçük kareler, Temel bileşenler regresyonu, Elastik net, Genelleştirilmiş maksimum entropi (GME) kullanılması da önerilmiştir (30-33).

Bu çalışmada, EKK yöntemi ile kurulan modelde bağımsız değişkenler arasında yüksek derecede korelasyon olmasına bağlı ortaya çıkan çoklu bağlantı sorununun modelin tahmin yeteneğini olumsuz etkileyeceği ve bağımsız değişkenlerin katsayılarının belirsiz hale gelmesine yol açacağı düşünülerek alternatif olarak önerilen Ridge, Liu ve LASSO tahmin edici yöntemleri kullanılmış ve model başarı kriterleri sonuçları karşılaştırılmıştır.

Çalışmada EKK yöntemi ile yapılan parametre tahmininde bağımsız değişkenler arasında yüksek korelasyon nedeniyle, VIF değerleri 10'dan ve KI değerlerinin de 30'dan büyük olması ve modelin R² değerinin yüksek hesaplanması çoklu bağlantı sorunu olduğunu göstermiştir. Bu nedenle, Ridge tahmin edicisini kullanılmış ve hesaplanan belirtme katsayısının, EKK yöntemi ile elde edilene benzer olduğu belirlenmiştir. Ancak, Ridge tahmin edicisinde hesaplanan RMSE değerlerinin EKK'ye göre daha küçük olduğu görülmüştür. Hesaplanan katsayılar göre elde edilen sonuçlar EKK ve Ridge yöntemlerinin karşılaştırıldığı diğer çalışmalar ile uyumlu bulunmuştur (34,35). Çalışmanın uygulanan Liu tahmin edicisinde de hesaplanan belirtme katsayısı EKK ve Ridge tahmin edicisi ile benzer, RMSE değeri ise daha küçük hesaplanmıştır. Elde edilen bu sonuçlar üç yöntemin karşılaştırıldığı çalışmalar ile uyumlu

bulunmuştur (36,37). Diğer bir alternatif olarak uygulanan LASSO tahmin edicisinde modelin belirtme katsayısı en yüksek ($R^2 = 0,97$) hesaplanmıştır. LASSO tahmin edicisinde veri setinden 3 değişken (yaş, yumurta ağırlığı ve Canlı ağırlık) modele dahil edilmiş, sonrasında elde edilen iterasyonlarda değişken seçimi yapılmış ve 2 değişken ile (Yaş ve Yumurta ağırlığı) model oluşturulmuştur. LASSO tahmin edicisinin RMSE değeri ise EKK ve Ridge tahmin edicisinden daha düşük, ancak Liu tahmin edicisinden daha büyük bulunmuştur. Bu kriterlere göre uygulamada kullanılan veri seti için elde edilen sonuçlar doğrultusunda Ridge, Liu ve LASSO tahminleri birbirlerine yakın bulunmuş ancak Liu tahmin edicisinin nispeten daha iyi sonuçlar verdiği görülmüştür. Çalışmada uygulanan tüm yanlış tahmin yöntemleri, EKK yöntemine göre daha sağlıklı ve güvenilir tahminler sağlamıştır. Bu durum literatürde yapılan çalışmalar ile benzer özellik göstermiştir (38, 39).

Çalışma sonucunda yumurta tavukçuluğunda satış gelirinin tahmininde, bağımsız değişkenler arasında belirlenen güçlü çoklu doğrusal bağlantı sorunundan dolayı önerilen Ridge, Liu ve LASSO tahmin edicileri, EKK yönteminden daha geçerli tahminler sağlamıştır. Elde edilen bu tahminlerden en küçük RMSE değerini veren tahminin ise Liu tahmin edicisine ait olduğu belirlenmiştir. Bu çalışmada üç değişken ile kurulan model, LASSO tahmin edici yönteminde iki değişkene indirilmiş ancak modelin RMSE değeri yüksek hesaplanmıştır. LASSO tahmin edicisinin değişken sayısının fazla olduğu durumlarda kullanılmasının daha uygun olduğu bildirilmiştir (32).

Özellikle üretim sektöründe tahminler içeren çalışmalarda sıklıkla karşılaşılan çoklu bağlantı sorununda alternatif olarak Ridge, Liu ve LASSO tahmin edicilerinin kullanılması tahmin gücünü artıracak, üreticiye üretim planlaması aşamasında daha geçerli ve güvenilir sonuçlar sunacaktır. Bu çalışmada da verilerinin kullanıldığı, hayvancılıkta çok önemli ekonomik bir faaliyet olan yumurta tavukçuluğunda, gelir ile ilişkili değişkenlerin doğru olarak belirlenmesi ve modellenmesi, üreticilere üretim planlaması aşamasında katkı sağlayacaktır.

Çıkar Çatışması Beyanı

Makalenin yazarları arasında bu çalışma kapsamında herhangi bir kişisel ve finansal çıkar çatışması bulunmamaktadır.

Finansal Kaynak Beyanı

Bu çalışma sırasında, yapılan araştırma konusu ile ilgili doğrudan bağlantısı bulunan herhangi bir ilaç firmasından, tıbbi alet, gereç ve malzeme sağlayan ve/veya üreten bir firma veya herhangi bir ticari firmadan, çalışmanın değerlendirme sürecinde, çalışma ile ilgili verilecek kararı olumsuz etkileyebilecek maddi ve/veya manevi herhangi bir destek alınmamıştır.

Yazar Katkısı Beyanı (Makale yazım dili İngilizce ise: Authors' Contributions)

Fikir/kavram: Ayşe YENER, Ali Alparslan SAYIM, Aytaç AKÇAY
Deney tasarımı: Ayşe YENER, Ali Alparslan SAYIM, Aytaç AKÇAY
Denetleme/Danışmanlık: Ayşe YENER, Aytaç AKÇAY
Veri toplama: Ayşe YENER, Aytaç AKÇAY
Veri analizi ve yorum: Ayşe YENER, Aytaç AKÇAY
Kaynak taraması: Ayşe YENER, Ali Alparslan SAYIM
Makalenin yazımı: Ayşe YENER, Ali Alparslan SAYIM, Aytaç AKÇAY
Eleştirel inceleme: Aytaç AKÇAY

Etik Onay

Bu makaledeki sunulan verilerin, bilgilerin ve dokümanların akademik ve etik kurallar çerçevesinde elde edildiği, tüm bilgi, belge, değerlendirme ve sonuçlarının bilimsel etik ve ahlak kurallarına uygun olarak sunulduğuna dair yazarlardan etik beyan alınmıştır.

Kaynaklar

1. Alpar R. Uygulamalı çok değişkenli istatistiksel yöntemler. Ankara: Detay Yayıncılık; 2013.
2. Eberly LE. Multiple linear regression. In: Topics in biostatistics. Springer; 2007. p. 165–87.
3. Akçay A, Sariözkan S. Yumurta tavukçuluğunda gelirin Ridge Regresyon analizi ile tahmini. Ankara Üniv Vet Fak Derg. 2015;62:69–74.
4. Bansal S, Singh G. Multiple linear regression-based analysis of weather data: assumptions and limitations. In: Shaw RN, Paprzycki M, Ghosh A, editors. Advanced Communication and Intelligent Systems. Cham: Springer Nature Switzerland; 2023. p. 221–38.
5. Albayrak AS. Çoklu doğrusal bağlantı halinde en küçük kareler tekniğinin alternatifi yanlı tahmin teknikleri ve bir uygulama. Uluslararası Yönetim İktisat ve İşletme Dergisi. 2005;1(1):105–26.
6. Topal M, Eydurhan E, Yağanoğlu M, Sönmez AY, Keskin S. Çoklu doğrusal bağlantı durumunda ridge ve temel bileşenler regresyon analiz yöntemlerinin kullanımı. Atatürk Ü Zir Fak Der. 2010;41(1):53–7.
7. Farrar DE, Glauber RR. Multicollinearity in regression analysis: the problem revisited. Rev Econ Stat. 1967;92:107.
8. Draper RN, Smith H. Applied Regression Analysis. US: Wiley Series in Probability & Statistics; 1998. p. 327–8.
9. Marquardt DW, Snee RD. Ridge regression in practice. Am Stat. 1975;29(1):3–20.
10. Montgomery DC, Peck EA. Introduction to Linear Regression Analysis. US: Wiley Series in Probability & Statistics; 1992.
11. Hoerl AE, Kennard RW. Ridge regression: Biased estimation for nonorthogonal problems. Technometrics. 1970;12(1):55–67.
12. Kejian L. A new class of biased estimate in linear regression. Commun Stat Theory Methods. 1993;22(2):393–402.
13. Acharjee A, Finkers R, Visser R, Maliepaard C. Comparison of regularized regression methods for omics data. Metabolomics. 2013;3(3):1–9.
14. Tibshirani R. Regression shrinkage and selection via the LASSO. J R Stat Soc Series B Stat Methodol. 1996;58(1):267–88.
15. Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. Ann Stat. 2004;32:407–99.
16. Hastie T, Tibshirani R, Friedman JH, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. Springer; 2009.
17. Erişoğlu M, Yaman N. Ridge tahminine dayalı kantil regresyon analizinde yanlılık parametresi tahminlerinin performanslarının karşılaştırılması. Necmettin Erbakan Üniversitesi Fen ve Mühendislik Bilimleri Dergisi. 2019;1(2):103–11.
18. Altelbany S. Evaluation of ridge, elastic net and LASSO regression methods in precedence of multicollinearity problem: a simulation study. J Appl Econ Bus Stud. 2021;5(1):131–42.
19. Riley RD, Snell KI, Martin GP, Whittle R, Archer L, Sperrin M, et al. Penalization and shrinkage methods produced unreliable clinical prediction models especially when sample size was small. J Clin Epidemiol. 2021;132:88–96.
20. Bastiaan RM, Salaki DT, Hatidja D. Comparing the performance of prediction model of Ridge and Elastic Net in correlated dataset. Oper Res Int Conf Ser. 2022;3(1):8–13.
21. Cho S, Kim H, Oh S, Kim K, Park T, editors. Elastic-net regularization approaches for genome-wide association studies of rheumatoid arthritis. BMC Proc. 2009. Springer.

22. Kohannim O, Hibar DP, Stein JL, Jahanshad N, Hua X, Rajagopalan P, et al. Discovery and replication of gene influences on brain structure using LASSO regression. *Front Neurosci.* 2012;6:115.
23. de Vlaming R, Groenen PJ. The current and future use of ridge regression for prediction in quantitative genetics. *Biomed Res Int.* 2015;2015:143712.
24. Pripp AH, Stanišić M. Association between biomarkers and clinical characteristics in chronic subdural hematoma patients assessed with LASSO regression. *PLoS One.* 2017;12(11):e0186838.
25. Usman M, Doguwa S, Alhaji B. Comparing the prediction accuracy of ridge, lasso and elastic net regression models with linear regression using breast cancer data. *Bayero J Pure Appl Sci.* 2021;14(2):134–49.
26. Üçkardeş F, Efe E, Nariñç D, Aksoy T. Japon bildircinlarında yumurta ak indeksinin ridge regresyon yöntemiyle tahmin edilmesi. *Akad Ziraat Derg.* 2012;1(1):11–20.
27. Çiftsüren MN, Akkol S. Prediction of internal egg quality characteristics and variable selection using regularization methods: ridge, LASSO and elastic net. *Arch Anim Breed.* 2018;61(3):279–84.
28. Bovans White product performance [Internet]. Available from: <http://www.isapoultry.com/en/products/bovans/bovans-white/>. Cited 2014 Jan 2.
29. Yumurta Üreticileri Merkez Birliği [Internet]. Available from: <https://www.yum-bir.org/Yumurta/FiyatListele.aspx>. Cited 2024 Jun 3.
30. Özel SÖ, Gezer F. Çoklu doğrusal bağlantı problemi altında tahmin edicilerin karşılaştırılması: Genelleştirilmiş Maksimum Entropi, Ridge, Liu. *Dumlupınar Univ Soc Sci J.* 2020;(66):82–96.
31. Çabuk HA, Özel SÖ. Çoklu iç ilişki sorunu olan regresyon modelinin HKO ölçütü ile bir etkin tahmin edicisi. *Çukurova Univ Soc Sci Inst J.* 2017;26(3):13–25.
32. Orhan H, Vergili M. Genomik veri setlerinin LASSO ve elastik net regresyon yöntemleri ile analizi. *SDU J Health Sci Inst.* 2022;13(3):485–96.
33. Çabuk A, Akdeniz F. İçilişki ve genelleştirilmiş maksimum entropi tahmin edicileri. *İstatistik Araştırma Derg.* 2007;5(2):1–19.
34. Büyükuysal MÇ, Öz İİ. Çoklu doğrusal bağlantı varlığında en küçük karelere alternatif yaklaşım: Ridge regresyon. *Düzce Univ Sağlık Bilim Enst Derg.* 2016;6(2):110–4.
35. Yüzbaşı B, Pala M. Ridge regresyon parametre seçimi: Türkiye'nin doğrudan yabancı yatırım örneği. *İstatistikçiler Derg İstatistik Aktüerya.* 2022;15(1):1–18.
36. Çoşkuntuncel O. Sosyal bilimlerde yanlı regresyon tahmin edicilerinin kullanılması. *J Meas Eval Educ Psychol.* 2010;1(2):100–8.
37. Gökçe A, Evren A. Çoklu bağlantı ve Ridge, Liu, zaman serisi kestiricisi ile Türkiye'nin ihracat modeli için bir uygulama. *Proceedings of the 13th Uluslararası Bilimsel Araştırmalar Kongresi; 2022 Mar; Elazığ. Elazığ: Asos Yayın Evi; 2022. p. 73–81.*
38. Melkumova L, Shatskikh SY. Comparing Ridge and LASSO estimators for data analysis. *Procedia Eng.* 2017;201:746–55.
39. Küçük A. Doğrusal regresyonda Ridge, Liu ve LASSO tahmin edicileri üzerine bir çalışma. *Ankara: Hacettepe Univ Fen Bilimleri Enst; 2019.*