Hayvansal Üretim Verilerinde Çoklu Bağlantı Probleminin Yanlı Regresyon Yöntemi İle Çözümlenmesi

İrfan ÖZTÜRK
HÜ, Ziraat Fakültesi, Zootekni Bölümü, Şanlıurfa

Geliş (Received): **12.09.2014** **Kabul (**Accepted): **08.10.2014**

Özet: Bu çalışmanın amacı, küçük örneklemlerde karkas ağırlığının tahmin edilmesinde yanlı tahmin tekniklerinden Ridge Regression (RR) yönteminin en küçük kareler EKK yöntemine karşı etkinliğini araştırmaktır. Bu amaçla broilerde karkas ağırlığı ile açıklayıcı değişkenler arasındaki doğrusal ilişkinin tahmininde EKK ve yanlı tahmin tekniklerinden Ridge Regression yöntemi karşılaştırılmaktadır. Araştırmada, bağımsız değişkenler arasındaki yüksek çoklu doğrusal bağlantı problemine dayanarak RR yönteminin EKK yöntemine göre daha küçük standart hatalı, durağan ve kuramsal beklentilere uygun tahminler sağlayacağı beklenmiştir.
Anahtar Kelimeler: En küçük Kareler Yöntemi, Ridge Regresyon, Çoklu Doğrusal Bağlantı

The Solution of Multicolinearity Problem via Biased Regression Analysis in Animal Production Data

Abstract: The aim of this study is to investigate the effectiveness of Ridge Regression (RR) applying biased estimation techniques over the method of Least Squares (LS) technique on the estimation of carcass weight. For this purpose, the Ridge Regression method biased estimation techniques are compared with LS to estimate linear relationship between carcass weight and explanatory variables in broiler. In this study, based the problem of high multiple linear connection between the independent variables, it was hypothesized that RR method has smaller standard errors and estimates in accordance with theoretical expectations according to the method of LS.
Keywords: Least Squares Technique, Ridge Regression, Multicolinearity

GİRİŞ

Aralarında sebep–sonuç ilişkisi bulunan iki veya daha fazla değişken arasındaki ilişkiyi ortaya koyabilmek ve o konu ile ilgili tahminler (estimation) yada kestirimler (prediction) yapabilmek amacıyla regresyon olarak adlandırılan matematiksel bir modelden yararlanılmaktadır(Şahinler, 2000)

Matematiksel olarak modellenen bu ilişki, iki ya da daha çok değişken arasındaki fonksiyonel ilişkiyi göstermekle kalmaz, değişkenlerden birinin önceden saptanan bir değeri için diğerinin tahmin edilmesini de sağlar. Değişkenler arasındaki ilişkinin bir eşitlik ile modellenmesi ve açıklanmasını sağlayan regresyon analizi tekniği hemen her alanda uygulamaları bulunmaktadır (Arıcı, 1991; Coşkuntuncel, 2010)

Ancak regresyon modelini tahmin etmeden önce olayı; yani bağımlı değişkenleri açıklayan etkili (bağımsız) değişkenlerin seçilmesi gerekmektedir. Sırf araştırmayı istatistiksel açıdan zengin gösterebilmek amacıyla regresyon model ortaya koymak ve gerekli gereksiz değişkenlerin böyle bir modele konulması birçok problemi beraberinde getirecektir.

Ön kestirici (bağımlı - bağımsız) değişkenler deney düzenleme yoluyla veriler toplansa dahi, matematiksel ve fiziksel kısıtlar nedeniyle birbirleriyle ilişki halinde olabilmektedir. Korelasyon (ilişki) problemini çözerken ileri regresyon teknikleri kullanılsa bile, bütün bu büyük sapmaların varlığı, eksik değerler, birbirleriyle ilişkili hatalar, sabit olmayan değişimler ve diğer problemler anlamsız sonuçlar doğurabilmektedir (Marqurt ve Snee, 1975).

Tahmin edilen modelin yeterli olup olmadığının kontrolü, regresyon analizinin en önemli bölümüdür. Uydurulan modelin doğru modele yeterli derecede yaklaştığını garanti etmek ve en küçük kareler regresyon analizinin tüm varsayımlarını sağlayıp sağlamadığını kontrol etmek gerekir. Eğer varsayımlarda bozulmalar söz konusu olursa; değişen varyanslılık, otokorelasyon ve normal dağılmama gibi sorunlar ortaya çıkar. Bu problemler verilerin artırılması, çeşitli transformasyonlar ile kısmen çözülebilmektedir. Bunun yanında aykırı değer, uç değer ve etkin gözlem gibi gözlemlerin model üzerindeki etkileri incelenir.

Çoklu regresyon analizinde en çok karşılaşılan sorunlardan biri de, bağımsız değişkenlerin birbirleriyle bağlantılı olması, yani bağımsızlık varsayımının bozulması ve bağımsız değişkenler arasında doğrusal bağlantıların mevcut olduğu sorunudur. Çoklu bağlantının EKK kestirimleri üzerinde oldukça olumsuz etkileri bulunduğundan yapılacak yorumların güvenirliliğinden kuşku duyulmalıdır. Bu gibi durumlarda yapılması gereken çoklu bağlantıyı ortadan kaldırmak veya etkisini azaltmaktır. Çoğu araştırmacı bu sorundan kurtulabilmek için çoklu bağlantı içersinde yer alan değişkenlerin bir veya birkaçını modelden çıkartarak çözüme ulaşmaya çalışılır. Ancak değişken seçim yöntemleri çoklu bağlantıdan etkilendiği için bu çözüm yolu da yanlış bulgulara sebep olabilir. (İpek, 2011)

İncelenen değişkenler arasındaki ilşişkilerde varsayımlar sağlanamıyorsa EKK yerine farklı yanlı tahminler veren Ridge Regresyon veya Anabileşenler

Sorumlu yazar: Öztürk, İ., ozirfan23@yahoo.com

Regresyonu analizlerinden yararlanmak gerekir. Çünkü regresyon modeli yeterli uyum sağlamazsa zayıf veya yanıltıcı sonuçlar vermektedir.

MATERYAL ve METOT

Materyal

Çalışmada kullanılan veriler; Harran üniversitesi araştırma uygulama çiftliğinde Turgay ŞENGÜL tarafından yürütülen bir araştırmadan uygun örnekleme yöntemi ile 30 tane broylerin karkas ağırlıklarını etkileyen faktörlere ait veriler ele alınmıştır. Bunlar sırasıyla Bağımlı değişken (Y) olarak karkas ağırlığı (kg), bağımsız değişken (X_i) olarak da X1= protein miktarı (gr gün^{-1}), X2= Tüketilen toplam yem miktarı (kg), X3=Işıklanma süresi (saat gün^{-1}), X4=Cinsiyet, X5= Sıcaklık (C^o) ve X6= Tüketilen Su miktarı (ml gün^{-1}) olarak ele alınmıştır.

Araştırmada broiler'in karkas ağırlığını etkileyen bağımsız değişkenlerin çoklu regresyon modeli tahmin edilmeye çalışılmıştır.

Metot

Çoklu Doğrusal Regresyon

Çok sayıda faktöre bağlı olarak değişim gösteren sosyal, ekonomik, üretim ve verim gibi olayların sebep sonuç ilişkisini ortaya koyabilmek için kullanılan istatistiksel yöntemlerden biri çoklu regresyon analizidir (İpek, 2011).

Çoklu doğrusal regresyon modeli matris formunda $Y = X \quad + \pounds$ şeklinde ifade edilir.

$$Y=\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X=\begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1k} \\ 1 & x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nk} \end{bmatrix} \qquad =\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \quad \pounds=\begin{bmatrix} \pounds_1 \\ \pounds_2 \\ \vdots \\ \pounds_n \end{bmatrix}$$

Burada; **y**, *nx1* boyutunda gözlemlerin vektörü, X matrisi, *n* x *p* boyutunda bağımsız değişken matrisi; , *p* x*1* boyutunda regresyon katsayılarının vektörü ve £ , *n* x1 boyutunda rastgele hataların vektörüdür. *k* bağımsız değişken sayısı olmak üzere *p* = *k* + 1'dir ve 1'ler sabit terim içindir. Ayrıca *k* = 1 alınırsa basit doğrusal regresyon modeli elde edilir. En küçük kareler tahmin edicileri £ = Y – X$_*$ olmak üzere $\sum \pounds_i^2$ hata kareleri toplamını minimize eder. Böylece 'nın en küçük kareler tahmini $\hat{}$,

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

ve varyans-kovaryans matrisi,

$$Var(\hat{}) = \pm^2 (X'X)^{-1}$$

şeklinde elde edilir. En küçük kareler için standartlaştırılmış artık değerler, $e_i = y_i - \hat{y}_i$ olmak üzere,

$$r_i = \frac{e_i}{\pounds}$$

ile verilir. Burada n örnek büyüklüğü p bağımsız değişken sayısı olmak üzere,

$$\hat{\pm}^2 = \frac{1}{n-p} \sum_{i=1}^{n} e_i^2$$

dir ve hatalar bağımsız, sıfır ortalamalı, *S* standart sapmalı, özdeş dağılıma sahip olduğunda

$\pounds^2$, $\pm^2$'nin yansız tahmin edicisidir. (Montgomery, 2013; Çoşkuntuncel, 2010).

Çoklu Regresyon analizinde bağımlı ve bağımsız değişkenler arasındaki ilişkiyi gösteren parametrelerin tahmini yapılırken bazı varsayımlar göz önünde bulundurulur. Bu varsayımlar aşağıdaki gibi özetlenebilir,

- Hata terimlerinin beklenen değeri sıfırdır.
- Hata terimlerinin varyansı sabit olmalı.
- Hata terimleri birbiriyle ilişkisizdir.
- Hata terimleri normal dağılım göstermelidir
- Bağımsız değişkenler arasında herhangi bir ilişki olmamalı. (Alma, 2008; Aktaş, 2007)

Açıklayıcı (bağımsız) değişkenler arasındaki basit doğrusal korelasyon katsayılarının sıfır veya sıfıra çok yakın olması şartı seklinde de açıklanabilen bu varsayım, istatistikte "Çoklu doğrusal bağlantı" bulunmaması olarak ifade edilmektedir *(*Alma, 2008; Albayrak, 2006)

Çoklu Bağlantı Kavramı

İki veya daha fazla bağımsız değişken arasında tam ya da yüksek derecede ilişkinin bulunması durumu çoklu bağlantı (multicollinearity) kavramı ile açıklanmaktadır.

Bağımsız değişkenler arasındaki bu yüksek korelasyon; özellikle en küçük kareler (EKK) yöntemi ile elde edilen regresyon katsayılarına ait tahmincilerin varyans-kovaryans değerlerinin büyük olmasına sebep olmaktadır. Bu durum regresyon modeline bağlı yapılacak yorumların hatalı olmasına neden olabilmektedir (Montgomery ve Peck, 1992).

Yukarıda bahsi geçen varsayımlar bazı araştırmacılar tarafından ya bilinçsizlikten dolayı göz ardı edilmekte veya özellikle çoklu bağlantı problemini gidermek için yüksek korelasyonlu değişkenlerin elemine edilmesi yoluna gidilmektedir. Bazen tahmin edilmek istenen regresyon modelinde, bağımlı değişkeni etkileyen olmazsa olmaz bağımsız değişkenler bulunmaktadır. Örneğin aralarında yüksek korelasyon bulunan ve canlı ağırlığı etkileyen yem tüketimi, yemdeki enerji ve protein miktarı ile cevre sıcaklığı gibi faktörlerin elemine edilmesi bilgi kaybına sebep olacaktır.

Söz konusu bağımsız değişkenlerin modelde bulundurulması belki R^2 belirtme katsayısının yüksek çıkmasını sağlayacaktır; ancak bu durum regresyon katsayılarına ait EKK tahmin edicilerinin varyans ve kovaryans değerlerinin büyük olmasına sebep olmaktadır. Bunun sonucunda regresyon modeline bağlı yapılacak yorum ve tahminlerin hatalı olmasına sebep olabilmektedir (Montgomery, 2013)

Ayrıca çoklu bağlantılı veriler de örnekten örneğe veya yıldan yıla tahmin edilecek doğrular arasında önemli derecede farklılıklar gözlemleneceğinden, anlamlı ve durağan regresyon doğrularının elde edilmesi güçleşmektedir (Maxwell, 2000) . Bundan dolayı bu tip regresyon doğrularının geçerliliğinin sınanmasında çok büyük örneklere ihtiyaç vardır. Ancak araştırmacı her zaman bu amaç doğrultusunda yeteri büyüklükte materyal temin edemeyebilir. Özellikle spesifik konularda çalışma yapılması durumunda (deve kuşu gibi) veya büyük başlarla ilgili araştırma uygulama çiftliğinde istenilen nitelikte materyal bulunamayabilir. Bu durumda da EKK yöntemi yerine yanlı tahmin yöntemlerinden faydalanmakta fayda vardır.

Bazı durumlarda EKK yönteminde çoklu bağlantının varlığı problemi bir veya daha fazla regresyon katsayısının yanlış işarete sahip tahminler verdiğini görürüz. Bu duruma regresyon katsayısı pozitif olması gerekirken kestirimin negatif olması örnek olarak verilebilir. Bu problem araştırmacı için oldukça sıkıntı verici bir durumdur. Araştırmacı katsayının pozitif olduğuna inanırken, kullanıcıya modeldeki parametrenin kestiriminin negatif olduğunu açıklamaya kalkması zor ve çelişkili bir durumdur (Montgomery, 2013).

Regresyon katsayılarının neden yanlış işaretli olabileceğini Mullet [1976] aşağıdaki şekilde maddeler halinde açıklamıştır:

1. Bazı hesaplama hatalarının yapılması
2. Bağımsız değişkenler arasında çoklu bağlantının varlığı
3. Tahmin edilen modelin, önemli bağımsız değişkenleri içermemesi
4. Bazı bağımsız değişkenlerin değer aralığının çok küçük olması

Bağımsız değişkenler arasında çoklu bağlantı varsa, EKK yöntemiyle çözüme gitmek uygun olmayacağından; öncelikle aşağıdaki yöntemlerle değişkenler arasında çoklu bağlantının olup olmadığının belirlenmesi gerekmektedir.

a. Bağımsız değişkenler arasında korelasyon katsayılarından bazılarının 1'e yakın olması,

b. $X'X$ matrisinin özdeğerlerinin bir yada birden fazlasının sıfır veya sıfıra yakın çıkması,

c. $X'X$ matrisinin rankı bağımsız değişken sayısından küçük olması,

d. Standartlaştırılmış $X'X$ matrisinin determinantı sıfır veya sıfıra çok yakın olması,

e. Koşul sayısı (Condition Number $CN = \frac{a_{max}}{a_j}$) 100 ile 1000 arasında ise orta derecede çoklu bağlantı; 1000 in üzerinde ise güçlü çoklu bağlantı vardır.

f. Koşul endeksi (Condition Index $CI = \sqrt{\frac{a_{max}}{a_j}}$) hesaplanan en büyük özdeğerin diğer özdeğerlere bölümünün karekökü 10 ile 30 arasında olması orta dercede; 30 üzerinde olması güçlü çoklu bağlantı olduğu anlamına gelmektedir (Gujarati, 1995)

g. Özdeğerlerin terslerinin toplamı bağımsız değişken sayısından çok büyük olması,

h. VIF_j = 1/(1-R_j^2) (Varyans Büyütme Faktörü) j = 1,2,.......k değerinin en büyüğü

$\mathrm{VIF} \geq 10$ ise anlamlı çoklu bağlantı söz konusu olduğunu Webster (1995) beyan etmektedir.

İ. j'nci bağımsız değişkenin Çoklu Belirtme Katsayısı R_j^2 değeri 1 'e yakın çıkması,

j. Hoerl ve Kennard'ın (1970) önerdiği Ridge İzinde, katsayıların grafiğinde dalgalanmalar olursa, çoklu bağlantı sorunu ortaya çıkmaktadır.

Böyle bir durumda en uygun regresyon modelinin çözümü ise yanlı Ridge regresyon yöntemi ile tahmin edilebilmesidir. (İpek, 2011)

Çoklu bağlantı probleminin saptanmasında kullanılan yukarıdaki yaklaşımlardan her biri tek başına yeterli olmamakla birlikte, bazı dezavantajlara da sahiptir. Ayrıca hangi durumda hangi yaklaşımın kullanılacağı konusunda da herhangi bir öneride bulunmamaktadır (Gujarati, 1995). Bu nedenle sadece bir yaklaşıma bağlı kalmak yerine, bunlardan birkaç tanesinin birlikte kullanılması daha güvenilir sonuçlar verecektir.

Yanlı ve Yansız Tahminler

Daha önceden bahsedilen EKK yöntemi ile regresyon modellerinin tahmin edilmesi için gerekli olan doğrusallık, sabit varyanslılık, bağımsızlık gibi varsayımlar yanlı tahminler için de gereklidir.

Doğrusal regresyon analizinde EKK tahminleri yansızdır. Yansızlık, örnekler üzerinden hesaplanan tahmin değerinin populasyon değerine eşit olması durumunu ifade eder. Herhangi bir örnekten elde edilecek yansız bir parametre tahmininin, büyük varyansa sahip olabileceğinden, her zaman populasyon parametre değerine eşit veya yakın olacağı söylenemez. Bu nedenle bazen, daha küçük varyanslı yanlı tahminler yansız tahminlere tercih edilebilir (Freund ve Minton, 1979).

Daha kararlı tahminler tercih edildiğinden dolayı yansız ve büyük varyanslı tahminler yerine, yanlı ve küçük varyanslı tahminler veren yöntemler geliştirilmiştir. Bu yöntemlerden biri de ridge regresyon yöntemidir.

Yanlı tahmin yöntemi olan ridge regresyon yöntemi ile çoklu bağlantılı verilerde daha durağan tahmin doğrularının elde edilmesi mümkün olmakla birlikte, kısıtlı çalışmalarda daha küçük örnekler kullanılması durumunda EKK yöntemine göre daha etkili ve durağan sonuçlar sağlamaktadır (Vinod, 1995).

Canlı ağırlık, kesim ağırlığı, karkas ağırlığı gibi beden ağırlığı ile ilgili çalışmalarda bağımsız değişkenler arasında genellikle bir örtüşme yani ilişki olacağından EKK yerine ridge regresyon yönteminin kullanılması daha güvenilir sonuçlar vermektedir (Maxwell, 2000)

Bir tahminin güvenirliğinin ölçüsü olarak tahmine ilişkin hata kareler ortalaması yanlılık ve örnek değişkenliğinin birleşik etkisiyle ifade edilebilmektedir (Netter, Wasserman ve Kutner, 1990).

$$HKO(\hat{\ }_r) = E(\hat{\ }_r - \ )^2 = V(\hat{\ }_r) + [E(\hat{\ }_r) - \ ]^2$$

$$= V(\hat{\ }_r) + \hat{\ }_r$$

deki sapma kare daha a çık bir ifadeyle, tahminin “hata kareler ortalaması” tahminin varyansı ile sapmanın karesi toplamına eşittir. Eğer tahmin yansız ise hata kareler ortalaması (HKO), tahminin varyansına eşittir. Küçük bir sapmaya karşılık varyansda meydana gelen azalma büyük ise, yanlı tahmin olan ridge regresyon, EKK yansız tahmininin sağladığından daha küçük bir hata kareler ortalamasına sahip olacaktır. Böylece ridge regresyonun esas amaçlarından biri gerçekleştirilmiş olunur.

Ridge Regresyon

Araştırma sonuçlarına ait sebep sonuç ilişkilerini uygun bir modelle ortaya koymak amacıyla regresyon analizinden yararlanılmaktadır. Elde edilen modele ilişkin regresyon katsayılarını tahmin etmek için yaygın olarak kullanılan yöntem En Küçük Kareler (EKK) yöntemidir. EKK tahmin yönteminde amaç hatayı en küçüklemektir. Ancak bağımsız değişkenler arasında yüksek derecede bir ilişki bulunuyorsa, bu tür veriler hata değerlerinde, dolayısıyla varyansta yanıltıcı bir büyümeye sebep olur. Bu büyüme parametre tahminlerine ve kestirim sonuçlarına olumsuz şekilde yansır. Diğer bir ifadeyle, bağımsız değişkenler arasında şiddetli bir ilişki bulunduğu durumlarda, varyanstaki yanıltıcı büyümeden dolayı, EKK tahmin yönteminin kullanılması yanlış model bulgularına ve kullanımına neden olur. Bu olumsuz etkiyi yok etmek için yanlı tahmin yöntemlerine başvurulur. Bu yöntemlerden birisi de ridge regresyon olup küçük bir yanlılık değerine karşılık varyans alanını dolayısıyla hatayı küçültür. Yanlı yöntemlere ilişkin bu tahmin ediciler, EKK tahmin edicilerine göre yanlı, ancak çok daha küçük varyanslı tahminler verirler. Kısacası, yanlı tahmin yöntemlerinin kullanılmasındaki genel amaç, EKK tahmin yönteminde büyük olan varyans alanını küçük bir yanlılık karşılığında daraltmaktır. Böylece EKK yöntemine göre daha doğru sonuçlar elde edilir (Ebegil, 2009).

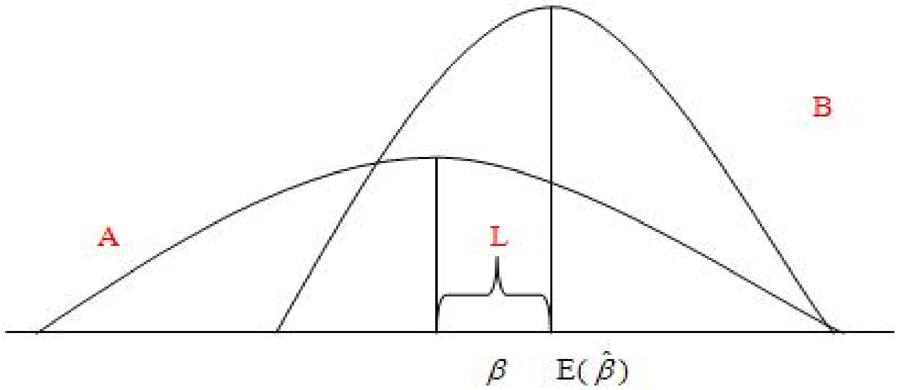
Çoklu regresyon modelinin tahmininde olayla ilgili bağımsız değişkenler deney düzenlemesi yardımı olmadan verilerin toplanması veya sadece deney düzenleme yoluyla veriler toplansa dahi bu kez de olayın yapısındaki fiziksel ve matematiksel kısıtlar nedeniyle birbirleriyle ilişki halinde olabilmektedir. Regresyonda çoklu bağlantının ortaya çıkması ile problemdeki sebep-sonuç ilişkisini ortaya koyan parametrelerin tahmin edilmesiyle duyarlı sonuçlar elde edilememektedir.

EKK yöntemi ile Ridge regresyon tekniğinin işleyiş mantığı temelde aynıdır. Aralarında sadece varyans-kovaryans matrisinin köşegen elemanlarına küçük bir (k) sabiti eklenerek, k-oranında yanlı tahminler elde etmemizi sağlamaktadır (Albayrak, 2006).

Ridge tahminleyicisinin EKK tahminleyicisine göre avantajı Şekil 1’de verilmiştir. Görüldüğü gibi merkezi olan normal eğri (A dağılımı) ‘nın sapmasız tahminleyicisinin EKK’e ait olasılık dağılımını göstermektedir. Bu eğrinin yayılışı sapmasız tahminleyicinin varyansı hakkında bilgi vermektedir. Merkezi $E(\hat{\ })$ olan normal eğri (B dağılımı) ise sapmalı ridge tahminleyicisinin olasılık dağılımını göstermektedir. Burada sapma $E(\hat{\ })$ ile arasındaki farktır. B dağılımına ait yayılışın dar olması daha küçük bir varyansın varlığını göstermektedir. Sonuç olarak bir miktar sapmaya müsaade etmek, sapmasız tahminleyicinin varyans ve HKO’sından daha küçük varyans ve HKO’sına sahip bir tahminleyici elde etmemize olanak sağlamaktadır (Şenyay L. 1993).

Burada yanlılık katsayısı olarak tanımlanan k sabiti 1 ile 0 arasında değer almaktadır. k değeri arttıkça varyans azalmaktadır. Asıl önemli olan optimum bir k sabitinin belirlenmesidir. Ancak henüz optimum bir çözümü garanti edecek bir yaklaşım bulunamamıştır (Albayrak, 2006).

Pratikte ise genellikle yanlı regresyon grafiği (Ridge Trace) olarak adlandırılan, yanlı standartlaştırılmış regresyon katsayıları ile k sabiti arasında hesaplanan grafiklerden yararlanılmaktadır (Hoerl, 1970).



Şekil 1. En küçük kareler tahminleyicisinin varyansı (A dağılımı) ile Ridge tahminleyicisinin varyansına (B dağılımı) ilişkin dağılım

Ridge regresyon katsayılarına ($\hat{\beta}_r$) karşılık k değerlerine (0< k <1) çizilen ridge iz (Ridge Trace) grafiğinde k arttıkça ridge kestiricileri ciddi anlamda değişkenlik göstermektedir. k'nın belirli bir değerinde ise ridge kestiricileri ($\hat{\beta}_r$) kararlı hale gelmektedir. Diğer bir ifade ile her bir k değerine karşılık elde edilen ridge kestiricileri $\hat{\beta}_r$ (regresyon katsayıları), k değeri arttıkça katsayılar daha kararlı hale gelmektedir. Ayrıca k değeri arttıkça residuallere (artıklara) ait KO (kareler ortalaması) arttığını, buna karşılık R^2 değerinin de azaldığını görmekteyiz. Gözlemlenen bu durumlar EKK yönteminin kararsızlığını göstermektedir.

k yanlılık parametresinin seçimi için bir çok araştırmacı farklı yöntemler ortaya koymuş olmalarına karşılık, bütün bu yöntemlerin ridge izinin doğrudan analizinden daha üstün olduğunun garantisi yoktur (Montgomery, 2013).

Burada k sabiti; yanlı standartlaştırılmış regresyon katsayılarının durağanlaştığı bölgeden seçilmektedir. Genellikle standartlaştırılmış regresyon katsayıları önce küçük k-değerleri ile anormal bir şekilde değişmekte, daha sonra ise durağan hale gelmektedir. Regresyon katsayılarının durağanlaştığı bu bölgeden en küçük optimum k değeri seçilmektedir.

Pratikte ridge izinin incelenmesi sonucu belirlenen k değerine bağlı olarak yapılan regresyon tahminleri tercih edilmektedir. Çünkü araştırmacıların bu yöntemi pratikte uygulayabilmeleri, hem de yorumlayabilmeleri daha kolaydır.

Ridge ve diğer yanlı kestiriciler üzerinde bugüne kadar onlarca çalışma yapılmıştır. Buna rağmen yöntemlerden herhangi biri en iyi yöntem olarak doğrudan ortaya konulamazken, çoklu bağlantının olduğu durumlarda yanlı kestirimin en küçük karelerden üstün olduğuna ilişkin önemli kanıtlar ortaya konulmuştur (Montgomery, 2013).

$X'X$ korelasyon matrisinin birim matrise yakın olması durumunda EKK yöntemi güvenilir sonuçlar vermektedir. Ancak $X'X$ korelasyon matrisinin birim matris olmaktan uzaklaşması, EKK tahminlerinin VIF (varyans büyütme faktörü) değerlerinde büyümesine neden olmakta ve dolayısıyla parametre tahminlerinin hatalarını arttırmaktadır(İpek, 2011).

Bağımsız değişkenler arasındaki ilişkiler, EKK katsayı tahminlerinin varyanslarını azaltmaktadır. Modeldeki her bir değişken için VIF değerleri, regresyon katsayılarının varyansları üzerindeki basit korelasyonların toplam etkisini göstermektedir. Çoklu bağlantı durumunda korelasyon matrisinin tersinin köşegen elemanları VIF değerleri; her bir tahminin diğer tahminlerle olan çoklu korelasyonunda sonsuz hale geldiği belirtilmiştir. Bu durumda EKK tahminleri yansız tahmin ediciler sınıfında en küçük varyanslı tahminler olma özelliklerini kaybetmektedirler. Çünkü çoklu bağlantı $\hat{\beta}$ ile gerçek β değerleri arasında sapmaya neden olmaktadır.

L_1, $\hat{\beta}$'dan β'ya olan uzaklık olmak üzere;

$$L_1^2 = (\hat{\beta} - \beta)'(\hat{\beta} - \beta)$$

Şeklinde yanlılığın karesi yazılabilir. L_1^2'nin beklenen değeri ise;

$$E(L_1^2) = \sigma^2 \, İz\,(X'X)^{-1}$$

Şeklindedir. Burada iz, bir kare matrisinin esas köşegenleri üzerindeki elemanlarının toplamıdır.

$X'X$ matrisinin öz değerleri $\lambda_1 > \lambda_2 > \ldots\ldots\ldots \lambda_k$ ile gösterildiğinde, $\hat{\beta}$'dan β'ya uzaklığının karesinin ortalama değeri şöyle verilmektedir.

$$E(L_1^2) = \sigma^2 \sum_{j=1}^{k} \lambda_j^{-1}$$

Çoklu bağlantının olması durumunda EKK yöntemiyle çözüm yapıldığında parametre tahminlerinin hata kareler toplamı, öz değerlerden yararlanılarak hesaplanırsa, bir yada daha fazla öz değerin sıfır veya sıfıra yakın olması, $\hat{\beta}$'nin β'dan sapmalarının beklenen değeri büyük olabilmektedir.

Bu nedenle regresyon katsayılarının tahmini için, bağımsız değişkenlerin birbirleri üzerindeki etkilerini minimum yapmak ve kararlı katsayı tahminleri elde edilebilmek için yanlı regresyon yöntemlerinden birisi olan ridge regresyonun kullanılması gerekmektedir (Albayrak, *2005).*

Ridge Regresyonun yanlı regresyon yöntemi olmasına karşın EKK yöntemine göre iki etkisi vardır.

1. Bağımsız değişkenlerde çoklu bağlantıyı gidermek.

2. Regresyonda yanlılık karesiyle varyansı değiştirerek hata kareler ortalamasını azaltmaktadır.

Hoerl ve Kennard (1970) ridge regresyonun 3 amaçla önermişlerdir;

a. Güçlü çoklu bağlantı etkisiyle katsayılarda oluşan kararsızlıkların grafik üzerine yansıtılması,

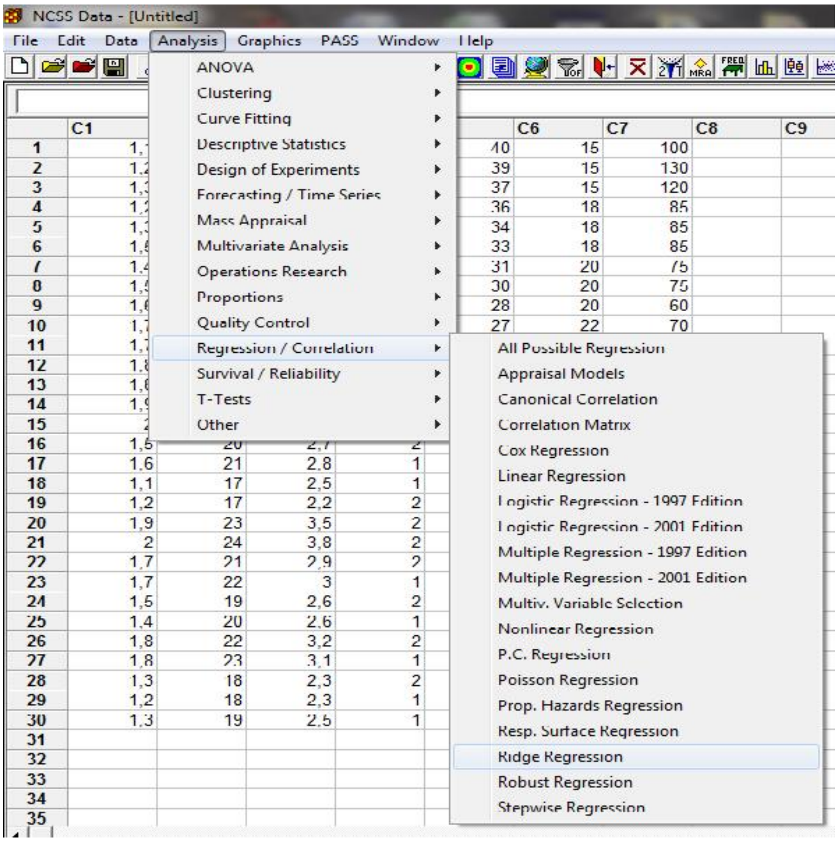
b. Değişkenlerin dik olmadıkları durum için EKK kestiricisinden daha küçük varyanslı kestiriciler elde edilmesi,

c. Modeldeki gereksiz değişkenlerin ayıklanması (İpek, 2011).

BULGULAR ve TARTIŞMA

Ridge regresyona ait istatistik analizler SPSS, SAS gibi istatistik programlarında makro komutları yazılarak yapılabildiği gibi; NCSS istatistik paket programında da herhangi bir makro komutu yazmaya gerek kalmadan pratik bir şekilde analizler yapıla bilmektedir.

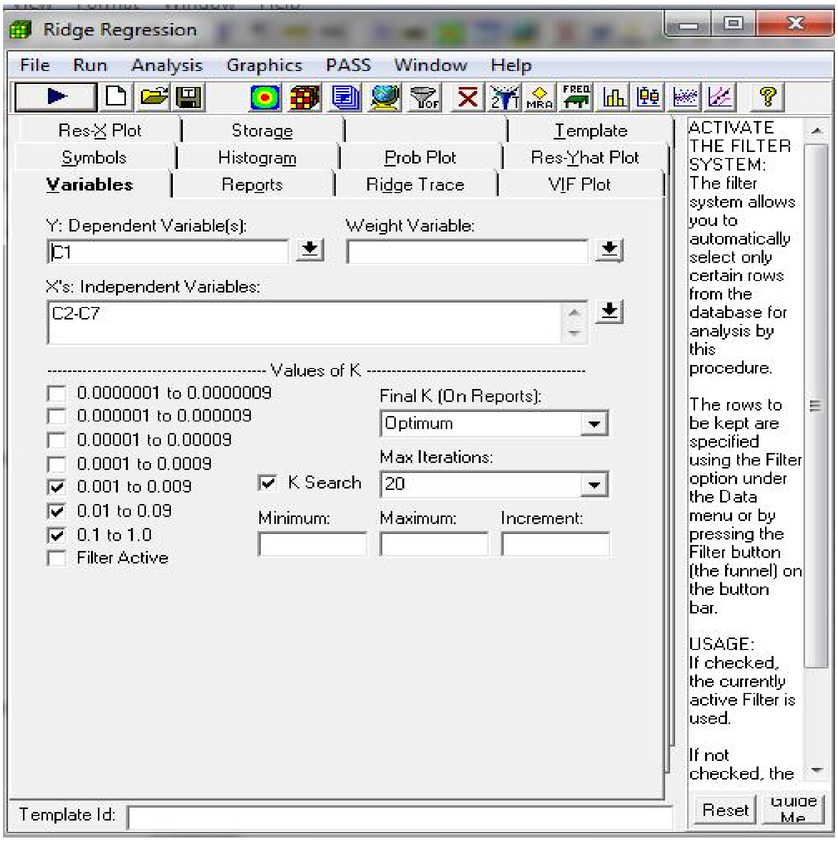
Burada hem NCSS istatistik programının nasıl kullanılacağı, hem de analiz sonuçlarının nasıl değerlendirilebileceği ele alınmaktadır. Şekil 2'i inceleyecek olursak, NCSS data sayfasında her bir değişkene ait veriler girildikten sonra Analysis menüsünden Regression/ Correlation sekmesi tıklanarak Ridge Regression analizi seçilir.



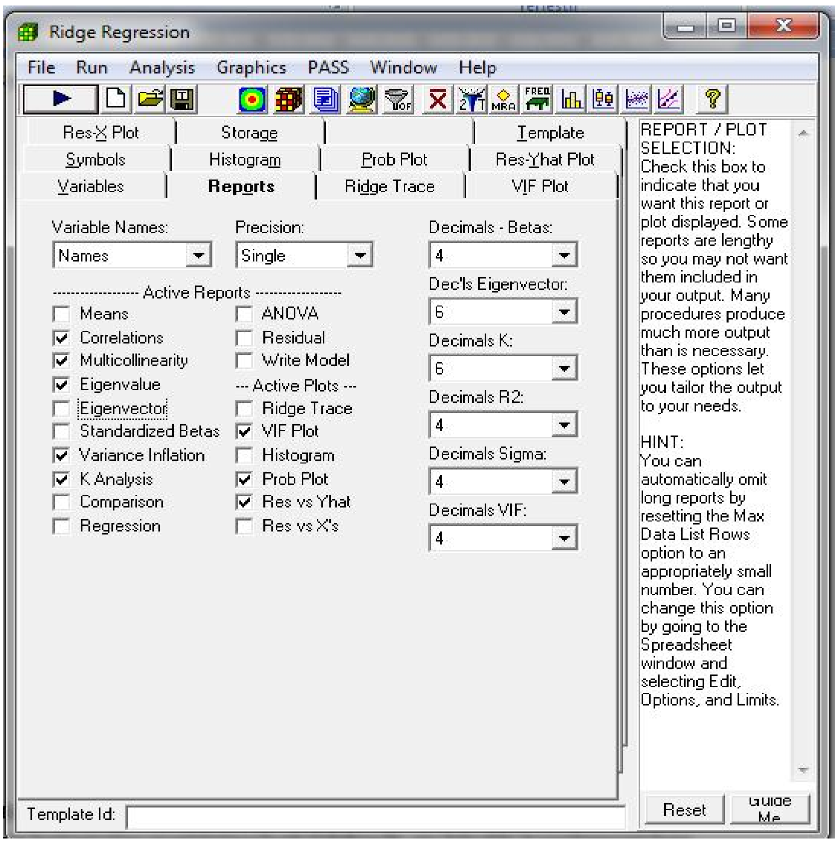
Şekil 2. NCSS Programına ait data sayfası ve ridge regression analiz seçeneklerine ait menüler

Görüldüğü gibi Şekil 3'de açılan pencerede Y bağımlı değişkenler Dependent Variable sekmesine, X bağımsız değişkenleri ise Independent Variables sekmesine atanır. Açılan penceredeki alt sekmelerde ise yine k-yanlılık değerine ait seçenekler bulunmaktadır. k- sabiti ridge iz grafiğinden yararlanarak belirlenebileceği gibi menüdeki optimum seçeneği de kullanılabilir.

Montgomery (2013), k yanlılık parametresinin seçimi için birçok araştırmacının farklı yöntemler ortaya koymuş olduklarını; ancak bu yöntemlerin ridge izinin doğrudan analizinden daha üstün olduğunun bir garantisi olmadığını belirtmiştir. Ayrıca program her menünün üzerine farenin imleci ile gezinince pencerenin sağ tarafında o menü ile ilgili özet açıklamalarda bulunmaktadır.



Şekil 3. NCSS Programına ait Ridge regresyon ile ilgili analiz seçenekleri



Şekil 4. Reports sekmesi, Ridge regresiyon analiz sonuçları ile ilgili detaylı istatistik analiz seçenekleri

Bununla birlikte Şekil 4'de görüldüğü gibi Ridge trace (iz), Vif Plot (Varyans büyütme faktörleri) Prob Plot (artıkların normallik dağılımı), Res-Yhat Plot (artık yani Residual değerleri ile y değişkeni arasında ilgiyi gösteren grafik) seçeneklerinden ilgili grafikler şekillendirip düzenlenebilir. Yukarıda bahsedilen grafiklerin çizimi ve bunlara ait diğer ridge regresyonla ilgili istatistik analizlerinin dökümünü Reports seçeneğini tıklayarak ayrı ayrı seçtikten sonra elde etmemiz mümkündür.

Regresyon modelini tahmin etmeden önce bağımlı ve bağımsız değişkenler arasındaki ilişkinin yönü ve derecesinin korelasyon analizi ile ortaya konulması gerekmektedir. Çizelge 1'de verilen bağımsız değişkenler arasındaki korelasyon katsayılarının 1'e yakın olması değişkenler arasında çoklu bağlantının olabileceğini göstermektedir. Korelasyon katsayılar matrisinde bağımlı değişkenlerden protein ile yem tüketimi, protein ile sıcaklık, protein ile ışıklanma süresi, yem tüketimi ile ışıklanma süresi arasında %90'ların üzerinde önemli derecede ilişki bulunmaktadır. Bu durum değişkenler arasında çoklu bağlantı probleminin olabileceğini göstermektedir.

Çizelge1. Bağımlı ve bağımsız değişkenler arasındaki korelasyon katsayıları

	Y1 Can.ağ.	**X2** protein	**X3** yem t.	**X4** cinsiyet	**X5** sıcak	**X6** ışık.s.	**X7** su t.
protein	0.973 0.000						
yem t.	0.940 0.000	0.938 0.000					
cinsiyet	0.313 0.092	0.115 0.546	0.242 0.197				
sıcak	-0.976 0.000	-0.993 0.000	-0.921 0.000	-0.148 0.437			
ışıklanma	0.933 0.000	0.970 0.000	0.890 0.000	0.068 0.719	-0.980 0.000		
su tük.	-0.385	-0.499	-0.290	0.536	0.506	-0.575	1
P-	0.036	0.005	0.121	0.002	0.004	0.001	0.000

*: Pearson korelasyon ve P önem düzeyi

Çoklu bağlantı olup olmadığını tespit etmek için kullanılabilecek bir diğer yol varyans büyütme faktörü (VIF) değerlerinin kullanılmasıdır. Varyans büyütme değerleri Çizelge 2'de verilmiştir.

Çizelge 2. Bağımsız değişkenlere ait varyans şişirme (büyütme) VIF değerleri

Bağımsız değ.	VIF	R^2 ve diğer X'ler	Tolerans
X2	145.9388	0.9931	0.0069
X3	14.7223	0.9321	0.0679
X4	2.7293	0.6336	0.3664
X5	178.8926	0.9944	0.0056
X6	36.0038	0.9722	0.0278
X7	3.9823	0.7489	0.2511

*Bazı varyans büyütme değerleri (VIF's) 10'dan büyük olduğundan çoklu bağlantı problemi vardır

Protein tüketimi; yem tüketimi; sıcaklık; ışıklanma süresi ile su tüketimine ait VIF değerleri 10'dan çok büyük olduğu için değişkenler arasında ciddi çoklu bağlantının olduğu söylenebilir.

Değişkenler arasında çoklu bağlantının olup olmadığının tespit edilmesinde kullanılan bir diğer yöntem bağımsız değişkenlere ait korelasyon matrisinin özdeğerlerinin kullanılmasıdır. Bağımsız değişkenlere ait korelasyon matrisinin özdeğerleri sırasıyla çizelge 3 de ki gibi belirlenmiş olup buna göre son dört özdeğer (karakteristik kök) sıfıra yakın olduğu için değişkenler arasında çoklu bağlantının varlığından söz edilebilir. Ayrıca en büyük özdeğer en küçük özdeğere bölündüğünde,

$$\frac{a_{max}}{a_{min}} = \frac{4.133308}{0.003117} = 1326.11 \text{ elde edilir.}$$

Bu değer 1000 den büyük olduğu için yine ciddi çoklu bağlantı probleminin varlığına işaret etmektedir.

Çizelge 3. Korelasyonların özdeğerleri

No.	Özdeğer	Göreceli yüzde	Birikimli yüzde	Koşul sayısı(CN)
1	4.133308	68.89	68.89	1.00
2	1.504020	25.07	93.96	2.75
3	0.274591	4.58	98.53	15.05
4	0.060120	1.00	99.53	68.75
5	0.024843	0.41	99.95	166.37
6	0.003117	0.05	100.00	1326.11

*Bazı koşul sayıları 1000'den büyük olup, ciddi çoklu bağlantı problemi söz konusudur.

Çoklu bağlantı probleminin olup olmadığının tespit edilmesinde kullanılan bir diğer yöntem özdeğerlerin terslerinin toplamı, bağımsız değişken sayısından fazla ise bağımsız değişkenler arasında çoklu bağlantının varlığını göstermektedir.

$$L = \Sigma \frac{1}{a_i} = \frac{1}{4.13} + \frac{1}{1.50} + \frac{1}{0.27} + \frac{1}{0.06} + \frac{1}{0.02} + \frac{1}{0.003} \cong 396.3$$

olup, değişken sayısı $P = 6$'dan oldukça çok büyüktür. Bu durum yine çoklu bağlantının varlığına işaret etmektedir.

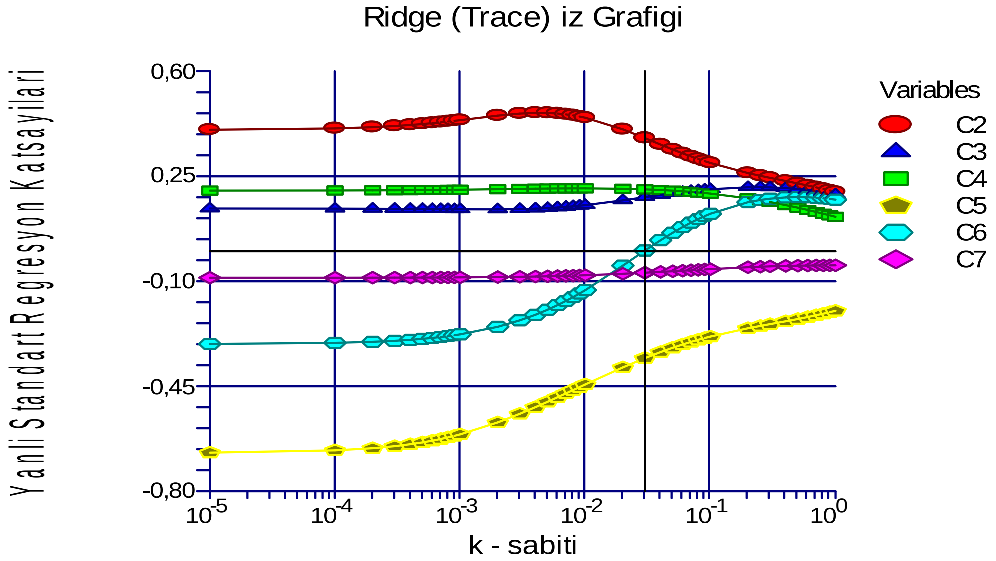
Ayrıca EKK yöntemine göre diğer bir ifade ile k yanlılık sabiti 0 alınarak regresyon analizi yapıldığında; varyans analizi sonucunda F= 493.6 gibi P= 0.000 önemlilik düzeyi modeli oldukça güvenilir olduğunu göstermektedir. Hâlbuki elde edilen regresyon modelindeki bazı katsayıların önemsiz olduğu görülmektedir (Çizelge 4). Bu çelişkili durum yine çoklu bağlantının varlığına ve değişkenlerin durağan olmadığı anlamına geldiği belirtilmiştir (Vupa ve Alma, 2008).

Çizelge 4. EKK're göre elde edilen regresyon modelinin katsayıları ve önem testi

Değişken	Katsayılar	Sta.Hata	T	P....................
Sabit	1.584	0.903	1.754	0.093
X_2-Protein	0.050	0.027	1.831	0.08
X_3-Yem tük.	0.087	0.043	2.028	0.054
X_4-Cinsiyet	0.109	0.016	6.680	0.00
X_5-Sıcak	-0.029	0.011	-2.742	0.012
X_6-Işıklanma	-0.027	0.009	-2.814	0.01
X_7-Su tük.	-0.001	0.000	-2.418	0.024

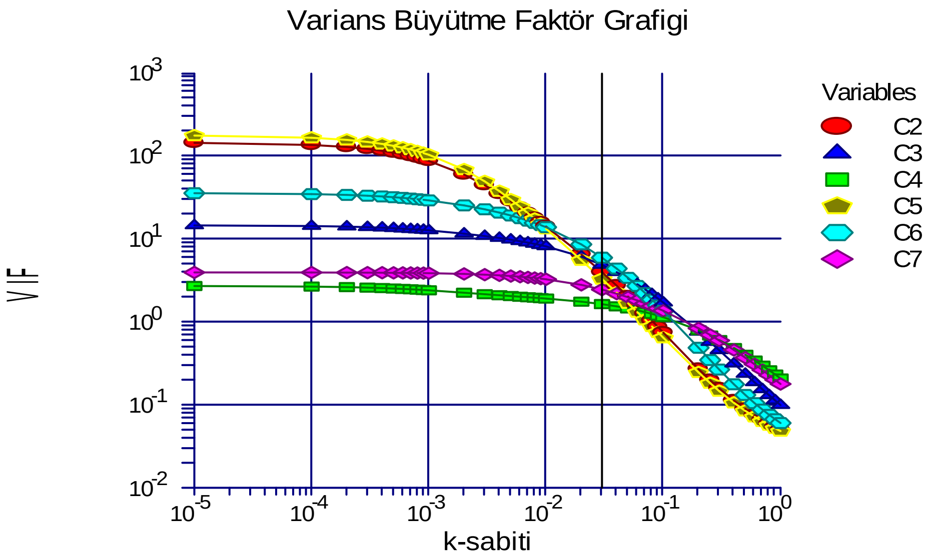
Modelin özeti:
S = 0.0271 R^2 = 99.23% R^2 (adj) = 99.03%

Ayrıca değişkenlere ait korelasyon matrisi incelendiğinde sıcaklık ile su tüketiminin canlı ağırlığı olumsuz etkilediği diğer bağımsız değişkenlerin ise pozitif etkilediği görülmektedir. Hâlbuki ışıklanma süresi ile canlı ağırlık arasında doğrusal pozitif ilişki varken, regresyon katsayısının negatif olduğunu Çizelge 4'de görmekteyiz. Bu durum açıklanamayacak bir çelişki doğurmaktadır. Bu nedenle EKK'le tahmin edilen model pek güvenilir sonuçlar vermeyecektir. Bunun yerine küçük bir yanlılık sabiti belirlenerek ridge regresyon analizinden faydalanmak gerekmektedir. Bunun için öncelikle k sabitinin belirlenmesi gerekir. Daha öncede bahsettiğimiz gibi k sabitini belirlemek için bir kısım yöntemler önerilmiştir. Ancak ridge iz grafiğinden daha üstün oldukları söylenemediğinden, burada pratik olan ridge iz grafiğinden yararlanarak optimum k yanlılık sabiti Şekil 5'den belirlenebilir.



Şekil 5. Ridge (Trace) iz grafiği ile k –sabitinin kestirimi

Şekil 5'den ridge iz grafiğini incelediğimizde k-sabitinin askari 0.02 den sonraki değer alınması gerekmektedir. Çünkü değişkenler 10^{-5} yaklaşık 0 noktasından 1'e doğru yaklaştıkça regresyon katsayılarının durağan hale geldiği görülmektedir. Örneğin C6 yani X6-ışıklanma süresine ait regresyon katsayısı k= 0.02 den sonra işaret değiştirmekte ve daha durağan hale gelmektedir. Bu nedenle k yanlılık değerini asgari 0.03 almamız gerekmektedir. Benzer olayı Şekil 6'da verilen varyans büyütme faktör grafiğinden de izleyebilmekteyiz.



Şekil 6. Varyans büyütme faktör (VIF) grafiği

Ayrıca yine Çizelge 5'de belirli bir orandaki k-sabitine karşılık bazı değişkenlere ait katsayılarda meydana gelen değişimleri görebilmekteyiz. Örneğin k=0.03 yanlılık değerinden sonra X6–ışıklanma süresi işaret değiştirmiştir. Bu nedenle k sabitini 0.03 olarak almayı uygun gördük. Burada k= 0.00 iken EKK için elde edilen katsayı tahminleri elde dilmektedir. R^2 belirtme katsayısı her ne kadar burada %99,2 gözükmekte ise de max VIF değerleri 10 'un çok üzerinde olup (178.893) ciddi çoklu bağlantı nın varlığını göstermektedir. Ancak k=0.03 yanlılık sabitinde R^2=%97.8 oranına düşse de max VIF değerleri 10'un altında olduğundan daha stabil ve güvenilir sonuçlar elde etmek mümkündür.

Çizelge 5. k-değerlerine karşılık elde edilen standardize edilmiş Ridge regresyon katsayıları

k	X2	X3	X4	X5	X6	X7	R2	Sigma	Max VIF
0.000	0.405	0.142	0.202	-0.671	-0.309	-0.088	0.992	0.027	178.89
0.001	0.424	0.141	0.204	-0.637	-0.292	-0.088	0.992	0.028	134.28
0.002	0.452	0.139	0.207	-0.571	-0.252	-0.086	0.991	0.03	69.62
0.004	0.461	0.142	0.208	-0.52	-0.212	-0.084	0.99	0.031	38.07
0.006	0.459	0.146	0.209	-0.487	-0.18	-0.083	0.989	0.033	24.63
0.008	0.453	0.15	0.209	-0.463	-0.153	-0.081	0.987	0.035	18.03
0.01	0.445	0.154	0.209	-0.444	-0.13	-0.08	0.986	0.036	14.26
0.02	0.406	0.17	0.208	-0.387	-0.048	-0.075	0.982	0.041	8.675
0.03	0.377	0.18	0.206	-0.356	0.002	-0.072	0.978	0.045	6.001
0.04	0.355	0.188	0.204	-0.336	0.037	-0.069	0.975	0.049	4.432
0.05	0.339	0.193	0.202	-0.322	0.062	-0.067	0.972	0.052	3.423
0.1	0.294	0.207	0.192	-0.284	0.125	-0.06	0.958	0.063	1.719
0.4	0.234	0.21	0.154	-0.234	0.179	-0.049	0.888	0.103	0.487
0.8	0.207	0.196	0.125	-0.208	0.176	-0.047	0.811	0.134	0.262
1	0.198	0.189	0.115	-0.199	0.172	-0.047	0.778	0.145	0.209

Değişkenlere ilişkin VIF değerlerinin küçüldüğünü Çizelge 6 da görebildiğimiz gibi ve k=0.03 yanlılık sabiti için elde edilen regresyon modeli de verilmiştir. Modele ilişkin belirtme katsayısının R^2=%97.84, varyasyon katsayısının ise %2.96 oranında olup oldukça güvenilir olduğunu söyleyebiliriz.

Çizelge 6. k = 0.03 için Ridge Regresyon katsayıları

Bağımsız Değişkenler	Regresyon katsayıları	Standard Hata	Stand. Regressiyon katsayıları	VIF
sabit	0.632			
X2	0.047	0.0076	0.3769	4.0069
X3	0.110	0.0415	0.1803	4.8982
X4	0.112	0.0213	0.2061	1.6492
X5	-0.015	0.0024	-0.3562	3.3150
X6	0.0002	0.0065	0.0024	6.0011
X7	-0.0009	0.0006	-0.0717	2.4711

Model: Y=0.632+ 0.047*X2+0.110*X3+ 0.112*X4 - 0.015*X5 + 0.0002*X6 - 0.0009*X7

Çizelge 7. k = 0,03 Yanlılık Sabiti İçin Varyans Analizi

Source	SD	KT	K.O.	F-Ratio	P
Sabit	1	70.53	70.533		
Model	6	2.139	0.3566	173.27	0.000
Hata	23	0.047	0.0021		
Toplam(Düzeltilmiş)	29	2.187	0.0754		
R^2			0.9784		
Standart Hata			0.0454		
Varyasyon katsayısı			0.0296		

SONUÇ ve ÖNERİLER

Olaylar arasındaki ilişkileri uygun bir model ile açıklayabilmek için, regresyon katsayıları yaygın olarak EKK yöntemi ile tahmin edilmektedir. EKK tahmin yönteminde amaç hatayı en küçüklemektir. Ancak örnek büyüklüğünün küçük olması ve bununla birlikte bağımsız değişkenler arasında kuvvetli bir ilişki bulunması durumunda, EKK yöntemi bu tür verilerde varyansta yanıltıcı bir büyümeye sebep olmaktadır. Bu yanıltıcı büyüme parametre tahminlerine ve kestirim sonuçlarına olumsuz şekilde yansımaktadır. Söz konusu olumsuz etkiyi yok etmek için yanlı tahmin yöntemlerinden biri olan ridge regresyon yöntemine başvurulur. Bu yöntem Şekil 1'de de görüldüğü gibi küçük bir yanlılık sabitine karşılık, varyans alanını dolayısıyla hatayı küçültmektedir.

Ridge regresyon yöntemi EKK tahmin edicilerine göre yanlı tahminler vermektedir, ancak bunun neticesinde çok daha küçük varyanslı tahminler elde edilmektedir. Burada amaç, EKK tahmin yönteminde büyük olan varyans alanını küçük bir yan karşılığında daraltmaktır. Böylece EKK yöntemine göre daha doğru sonuçlar elde edilmektedir.

Araştırmamızda çoklu doğrusal bağlantı probleminin bir sonucu olarak EKK yöntemine bağlı olarak tahmincilerin standart hatalarının yüksek ve X6-ışıklanma süresine ait katsayının işareti kuramsal beklentilerle çeliştiği görülmektedir. Söz konusu çelişki ve olumsuzlukları düzeltmek için uygulanan yanlı tahmin yöntemlerinden ridge regresyon tutarlı ve kuramsal beklentilere uygun sonuçlar verdiği görülmektedir. Optimum yanlılık sabitini araştırmak amacıyla VIF ve ridge grafiklerinden yararlanarak, yanlı regresyon katsayılarının durağanlaştığı ve bu katsayılara ait VIF değerlerinin birlikte 1'e yaklaştığı bölgede yaklaşık bir k değeri seçilerek iterasyonlar sonucunda ridge regresyon için seçilen optimum k= 0,03 yanlılık sabiti belirlenir. Elde edilen sonuçlarda X6-ışıklanma süresine ait katsayının işareti değişmekte ve tahmincilerin standart hataları küçülmektedir.

Sonuç olarak, karkas ağırlığını açıklayan değişkenler arasında çoklu doğrusal bağlantı olduğundan, yanlı tahmin teknikleri EKK yöntemine göre daha tutarlı, güvenilir, durağan ve kuramsal beklentilere uygun tahminler sağladığı görülmektedir.

KAYNAKLAR

Aktaş, C., 2007. Çoklu Bağıntı ve Liu Kestiricisiyle Enflasyon Modeli için bir Uygulama. ZKÜ sosyal Bilimler Dergisi, cilt 3, sayı 6, 67-79 .

Albayrak, A.S., *2005.* Çoklu doğrusal bağlantı halinde en küçük kareler tekniğinin alternatifi yanlı tahmin teknikleri ve bir uygulama *ZKÜ Sosyal Bilimler Dergisi Cilt 1, Sayı 1, 105-126.*

Albayrak, A.S. 2006. Uygulamalı Çok Değişkenli İstatistik Teknikleri. Asil yayın dağıtım, Ankara, 265-284.

Alma, G.Ö., Vupa, Ö. 2008. Regresyon Analizinde Kullanılan En Küçük Kareler Ve En Küçük Medyan Kareler Yöntemlerinin Karşılaştırılması. SDÜ Fen Edebiyat Fakültesi Fen Dergisi (E-Dergi). 3(2) 219-229.

Arıcı, H. 1991. İstatistik Yöntemler ve Uygulamalar. Ankara: Meteksan.281 s.

Coşkuntuncel, O. 2010. Sosyal Bilimlerde Yanlı Regresyon Tahmin Edicilerin Kullanılması. Eğitim ve Psikolojide Ölçme ve Değerlendirme Dergisi, 1(2), 100-108.

Ebegil M. 2009. Ridge tahminine dayalı yanlı tahmin edici için bir test istatistiği **,** SAÜ Fen Edebiyat Dergisi, II, 1-14.

Freund, R.J. and Minton, P.D. 1979. Regression methods. Marcel Dekker, New York 261 s.

Gujarati, D.N. 1995. Basic Econometrics. McGraw-Hill, New York, 319-399 s.

Hoerl, A.E. and Kennard, R.W. 1970 Ridge Regression: Biased Estimation for Nonorthogonal Problems. Technometrics, Cilt:12, No: 1, 55-67.

İpek, O. 2011. Ridge Regresyon Üzerine Bir Çalışma. idari.cu.edu.tr/sempozyum/bil28.htm Erişim Tarihi: 06.06.2014.

Marqurt, D.W.,Snee, R.D. 1975. Ridge Regression in Pratice. The American Statistician, ,Vol. 29 No. 1,4.

Maxwell, Scott E. 2000. Sample Size in Multiple Regression Analysis. Psychological Methods, Cilt: 5, No: 4, 434-458.

Montgomery, D. C., Peck, E. A., & Vining, G. G. 2013. Doğrusal Regresyon analizine Giriş (Introduction to linear regression analysis New York: John Wiley and Sons.) 5.Basımdan çeviri. Yayın No: 717, 142 Nobel Akademik Yayıncılık, Ankara, 645 S.

Mullet, G, M. 1976. Wye Regression Coefficients have the wrong sign. Journal of Quality Technology, 8, 121-126.

NCSS Statistical System 2001. *User's Guide*, Kaysville, NCSS Inc.

Netter, J., Wasserman W. and Kunter M. 1990. Applied Linear Statistical Models. Irwin: Homewood, IL.

Şahinler, S. 2000. En Küçük Kareler Yöntemi ile Doğrusal Regresyon Modeli Oluşturmanın Temel Prensipleri. MKÜ. Ziraat Fakültesi Dergisi 5 (1-2). Hatay, 57–73.

Şenyay, L. ve Özler, C. 1993. "Ridge Tahminleyicisinin Özellikleri. Dokuz Eylül Üniversitesi", I. Ulusal Ekonometri ve İstatistik Sempozyumu, İzmir, 217-236.

Vinod, H.D. 1995. Double Bootstrap for Shrinkage Estimators, Journal of Econometrics, 68, 287-302.

Vupa Ö ve Alma, Ö.G. 2008. Doğrusal Regresyon çözümlemesinde Çoklu bağlantı probleminin sapan değer içeren küçük örneklemlerde incelenmesi. SÜ Fen Ed. F. Derg. Sayı 31, Konya, 97-107.

Webster, A.1995. Applide Statistics for Business and Economics 683-684.