T. C.
ULUDAĞ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK ANABİLİM DALI

RİDGE REGRESYON ANALİZİ ve BİR UYGULAMA

M. Çağatay BÜYÜKUYSAL

(YÜKSEK LİSANS TEZİ)

Bursa-2010

T. C.
ULUDAĞ ÜNİVERSİTESİ
SAĞLIK BİLİMLERİ ENSTİTÜSÜ
BİYOİSTATİSTİK ANABİLİM DALI

RİDGE REGRESYON ANALİZİ VE BİR UYGULAMA

M. Çağatay BÜYÜKUYSAL

(YÜKSEK LİSANS TEZİ)

Danışman: Prof. Dr. İsmet KAN

Bursa-2010

Sağlık Bilimleri Enstitüsü Müdürlüğüne,

Bu tez, jürimiz tarafından ....
tezi olarak kabul edilmiştir.

Adı ve Soyadı İmza

Tez Danışmanı ....

....

Üye ....

Üye ....

Üye ....

Üye ....

Bu tez, Enstitü Yönetim Kurulunun .... tarih, ... sayılı toplantısında alınan .... numaralı kararı ile kabul edilmiştir.

....
Enstitü Müdürü

İÇİNDEKİLER

ÖZET

Bu çalışmada, çoklu doğrusal regresyon analizinde, bağımsız değişkenler arasında çoklu doğrusal bağıntı olması durumunda ortaya çıkan yanlılığı ortadan kaldırmak için en küçük kareler yöntemine alternatif olarak önerilen ridge regresyon yöntemi üzerinde çalışılmıştır. En küçük kareler ve ridge regresyon analizinden elde edilen sonuçlar karşılaştırılmıştır.

İstanbul Üniversitesi İstanbul Tıp Fakültesi Hastanesine şişmanlık şikayeti ile başvuran hastalar arasından rastgele seçilen 20 kişiye ait beden ağırlığı (kg), deri alanı (cm2), uyluk kemiğinin çevresinin uzunluğu (cm) ve belden yukarı ölçülen kasların çevrelerinin uzunluğu (cm) değerleri elde edilmiş ve bu değerlerle hastaların vücut ağırlığıyla olan ilişkileri incelenmiştir.

Yapılan analizlerin sonucunda, ridge regresyon analiziyle elde edilen regresyon katsayıları, kuramsal beklentilere cevap verirken, en küçük kareler yöntemiyle elde edilen regresyon katsayıları, beklenenden uzak sonuçlar vermiştir.

Sonuç olarak çoklu doğrusal bağıntı halinde, ridge regresyon analizinin en küçük kareler yöntemine göre daha doğru sonuçlar verdiği görülmüştür. Bu yüzden çoklu doğrusal bağıntı varlığında, ridge regresyon analizinin kullanılmasını öneriyoruz.

Anahtar kelimeler: En küçük kareler, çoklu doğrusal bağıntı, ridge regresyon.

SUMMARY

At this study, we studied on an alternative method suggested to least squares called "Ridge Regression" which is used to remove multicollinearity between independent variables. Then, results from least squares and ridge regression analysis are compared.

20 patients are randomly selected from a sample from whose had an obesity problem at Istanbul University, Faculty of Medicine. Body mass (kg), skin area (cm2), length of femur circle (cm) and length of muscles' circles which are above the waist (cm) datas are collected and the relationships between these parametres and body mass is examined.

After the analysis, regression coefficients that are found with ridge regression are very close to theoretical expectations, whereas regression coefficients that are found with least squares are so different from expectations.

As a result, when multicollinearity occurs, it is seen that ridge regression analysis gives more appropriate results than least squares. So, at problem of multicollinearity we suggest using ridge regression method.

Key Words: Least Squares, Multicollinearity, Ridge Regression.

GİRİŞ

Sosyal, ekonomik ve tıbbi olayların birçoğu çeşitli faktörlere bağlı olarak meydana gelmektedir. Özellikle ekonomi ve tıp alanındaki olayların ortaya çıkmasında veya gelişmesinde etkili olan faktörler, aynı zamanda birbirleri ile de yakından ilişkilidir. Ekonomik alanda geleceğe yönelik daha gerçekçi çalışmaların yapılabilmesi için o ana kadar yaşanan olayların neden-sonuç ilişkisinin belirlenmesi önemlidir. Benzer şekilde tıp alanında da bir parametreye ilişkin tahminlerde bulunabilmek için o parametreyi etkileyen yan faktörlerin birbirleriyle olan ilişkileri de göz önünde bulundurulmalıdır. Ancak bu şekilde hareket edilerek çalışılan konu ile ilgili olarak yapılan tahminler güvenirlik kazanabilir. Bu çalışmada, alanımızla ilgili olarak tıp alanındaki çalışmalar incelenmiştir.

Sosyal, ekonomik ve psikolojik olaylarda olduğu gibi, çok sayıda etmene bağlı olan tıbbi parametreler de birlikte değişim göstermektedir. Bir tıbbi olayın sebep-sonuç ilişkisini belirleyebilmede regresyon modeli sıkça kullanılan modellerden biridir. Bu modelde, incelenen tıbbi olay ile bu olayın meydana gelmesinde etkili olan diğer parametreler arasında veya olayı etkileyen bağımsız parametreler arasında da bir bağıntının olabileceği öngörülür. Ancak bu bağıntının regresyon analizinde sıkça sorun yarattığı da bilinmektedir. Regresyon sürecinde ortaya çıkan bu sorunun nedeni, bağımsız değişkenlerin bağımsızlık varsayımlarının bozulması ve sonuçta bu değişkenler arasındaki doğrusal bir bağıntının ortaya çıkmasıdır. “Çoklu doğrusal bağıntı” olarak adlandırılan bu soruna önerilen çözüm, yanlı regresyon yöntemlerinin kullanılmasıdır. Söz konusu regresyon yöntemleri, değişken seçimi yaparak veya değişkenlerin hepsini modelde bırakarak en küçük kareler yöntemine göre daha küçük varyansla kestirim yapan yöntemlerdir.

Bu çalışmada, bağımsız değişkenlerin hepsinin modelde bırakılarak parametre kestirimi yapılmasına olanak sağlayan “ridge regresyon yanlı kestirim yönteminin” kullanılması amaçlanmıştır. Bu amaç doğrultusunda çalışmanın ilk bölümünde çoklu regresyon analizi ve en küçük karelere ilişkin varsayımlara yer verilmiş, ikinci bölümde ise çoklu doğrusal bağıntının belirlenmesi, nedenleri ve giderilme yöntemleri anlatılmıştır. Çoklu doğrusal bağıntının var olduğu doğrusal modellerde, çoklu doğrusal bağıntının etkilerini en aza indirgeyerek parametre kestirimi yapılmasını sağlayan ridge regresyon yöntemi de üçüncü bölümde verilmiştir. Bu bölümde ridge regresyon yönteminin matematiksel açıklaması yapılmış, daha sonra da yöntemin kullanılmasına ilişkin ayrıntılara yer verilmiştir. Çalışmanın son bölümünde ise, İstanbul Üniversitesi İstanbul

Tıp Fakültesi Hastanesine şişmanlık şikayeti ile başvuran hastaların, beden ağırlığı (kg), deri alanı (cm^2), uyluk kemiğinin çevresinin uzunluğu (cm) ve belden yukarı ölçülen kasların çevrelerinin uzunluğu (cm) parametrelerinin beden ağırlığı üzerinde etkisini araştıran uygulamaya yer verilmiştir. Uygulamada en küçük kareler yöntemiyle ridge regresyon yönteminden elde edilen sonuçlar karşılaştırılmıştır.

GENEL BİLGİLER

1. Çoklu Doğrusal Regresyon

Sağlık alanındaki bağımlı değişkenlerin iki ya da daha fazla bağımsız değişken tarafından etkilenme olasılığı oldukça yüksektir. Bunun sonucu olarak gözlemlenen bir tıbbi değişkeninin değeri, çok sayıda parametreden etkilenmesi sonucu ortaya çıkmaktadır. Bu parametrelerden bazıları önemli etkileri olan değişkenler (major factors), diğer bazı parametreler ise önemsiz etkiye (minor factor) sahip değişkenlerdir. Bir değişkeni etkileyen iki ve daha fazla bağımsız değişken arasındaki neden-sonuç ilişkilerini doğrusal bir modelle açıklamak ve bu bağımsız değişkenlerin etki düzeylerini belirlemek için yararlanılan yönteme "çoklu doğrusal regresyon analizi" denir (1, 2).

k Sayıda bağımsız değişkeni ve bir bağımlı değişkeni olan model,

$$Y = \beta_0 + \beta_1 X_1 + \beta_1 X_1 + \cdots + \beta_p X_p \qquad [1]$$

şeklinde ifade edilir.

Burada, β_0 sabit (constant) katsayısı, β_i , i=1,2,...,p regresyon katsayılarıdır. β_i katsayılarının her biri önünde bulunan bağımsız değişkenlerin Y'nin değişimi üzerine etkilerini belirtmektedir. ε vektöründeki hata terimleri, ortalaması sıfır ve standart sapması σ olan normal dağılıma sahiptir (1).

Bu fonksiyonel ilişki, matris notasyonuyla gösterilmek istendiğinde,

$$Y = X\beta + \varepsilon \qquad [2]$$

şeklinde ifade edilir. Matris formundaki;

Y: Nx1 boyutlu bağımlı değişken vektörü,

X: Nx(k+1) boyutlu bağımsız değişkenler matrisi,

b: (k+1)x1 boyutlu katsayılar vektörü,

ε: Nx1 boyutlu hata vektörüdür (3, 4)

Çoklu doğrusal regresyon modelinde parametrelerin kestirim değerleri "en küçük kareler" yöntemiyle bulunur.

En küçük kareler yöntemi, basit doğrusal ve çoklu regresyon modellerinin çözümlenmesinde kullanıldığı gibi, çok denklemli ekonometrik modellerin çözümünde de kullanılan tekniklerin temelidir. Örneklem verileri kullanılarak ana kütle parametreleri olan β_0, β_1'in kestirimlerini elde edebilmek için en küçük kareler tekniğinden yararlanılır. Bu parametrelerin kestirim değerleri $\hat{\beta}_0$ve $\hat{\beta}_1$ olarak ele alındığında, regresyon doğrusunun denklemi $\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_i$ şeklinde olur. Bu ilişki matris notasyonuyla gösterilmek istendiğinde,

$$Y = X\hat{\beta} + e \qquad [3]$$

şeklinde yazılır.

Burada, $\hat{\beta}$ (px1) boyutlu parametre katsayılarının kestirim vektörüdür. e ise (nx1) boyutlu artık vektörü olarak tanımlanır.

1.1. Çoklu Doğrusal Regresyon Modelinin Varsayımları

Regresyon analizinde bağımlı ve bağımsız değişkenler arasındaki ilişkiyi gösteren parametrelerin kestirimleri yapılırken bazı varsayımlar göz önünde bulundurulur. Sözkonusu varsayımlar hata vektörü ve bağımsız değişken matrisiyle ilgili varsayımlardır. Kestirimlerin standart hatalarının en küçük ve dolayısıyla parametre kestirimlerinin etkin olmasını sağlayan bu varsayımlar aşağıda belirtilmiştir (5).

1.1.1. Hata Terimlerinin Beklenen Değerinin Sıfır Olması

Çoklu regresyon modelinde birinci varsayım, gözlemlere karşı gelen hata terimlerinin beklenen değerinin sıfır olduğudur. Regresyon modeli;

$$Y_i = \beta_0 + \beta X_i + \varepsilon_i \quad \text{için} \quad i=0,1,2,...,n \qquad [4]$$

Hatalarının beklenen değeri sıfırdır. $E(\varepsilon_i) = 0$

Bu eşitlik matris şeklinde aşağıdaki gibi gösterilebilir.

$$E\begin{bmatrix}\varepsilon_1\\ \varepsilon_2\\ \varepsilon_3\\ \vdots\\ \varepsilon_n\end{bmatrix}=\begin{bmatrix}E(\varepsilon_1)\\ E(\varepsilon_2)\\ E(\varepsilon_3)\\ \vdots\\ E(\varepsilon_n)\end{bmatrix}=\begin{bmatrix}0\\ 0\\ 0\\ \vdots\\ 0\end{bmatrix} \qquad [5]$$

X değişkeni, X_i değerini aldığında, Y değişkeninin beklenen değeri $\beta_0 + \beta X_i$ olur ve $E(Y_i/X_i)$ şeklinde belirtilir. Bütün X_i (i=1,...,n) değerleri için hata terimleri, beklenen değeri sıfır $[E(\varepsilon_i) = 0]$ ve standart sapması σ_ε olan bir normal dağılıma sahiptir (5).

1.1.2. Hata Terimlerinin Varyansının Sabit Olması

Bütün hata terimlerinin varyansı (σ^2) sabittir. Bu varsayım doğrusal regresyon modelinde kestirimlerin standart hatalarının küçük olmasına ve dolayısıyla kestirimlerin güvenilirliğinin artmasına yardımcı olur. Bu varsayıma sabit varyans (homescedasticity) varsayımı denir (5).

1.1.3. Hata Terimlerinin Birbirinden Bağımsız Olmaları

$E(\varepsilon_i\varepsilon_j)=0$ varsayımı, hata terimlerinin birbirlerini etkilemediklerini gösterir. ε_i ve ε_j eğer rasgele değişkenler ise,

$$E(\varepsilon_i\varepsilon_j) = E(\varepsilon_i)\, E(\varepsilon_j) \qquad [6]$$

yazılabilir. Hataların beklenen değeri sıfır olduğundan, $E(\varepsilon_i\varepsilon_j)=0$ olur. Hata terimlerinin birbirinden bağımsız olmaları, değişkenler arasındaki ilişkiyi belirleyen parametrelerin kestirim değerlerinin gerçek değere yakın olmasını sağlar (5).

1.1.4. Hataların Varyans ve Kovaryanslarının Sınırlı Değere Sahip Olması

$Y = X\beta + \varepsilon$ regresyon modeline ait bir diğer varsayım, modeldeki hata terimlerinin beklenen varyanslarının eşit ve kovaryanslarının sıfır kabul edilmesidir.

Hataların varyans-kovaryansı;

$$
\begin{aligned}
E[(\varepsilon\text{-}E\varepsilon)\ (\varepsilon\text{-}E\varepsilon)'] &= E[(\varepsilon\text{-}0)\ (\varepsilon\text{-}0)'] \\
&= E(\varepsilon\varepsilon') \\
&= \sigma^2 I_n \qquad [7]
\end{aligned}
$$

olur.

Eşitlik, matrislerle aşağıdaki gibi gösterilir.

$$
\begin{bmatrix} \sigma^2_{\varepsilon_1} & 0 & \cdots & 0 \\ 0 & \sigma^2_{\varepsilon_2} & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \ldots\ldots & & \sigma^2_{\varepsilon_n} \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & & & \vdots \\ 0 & \ldots\ldots & & 1 \end{bmatrix} \qquad [8]
$$

Elde edilen bu eşitlik, hata terimlerinin varyanslarının σ^2_ε 'ye, kovaryanslarının ise sıfıra eşit olduğunu gösterir (3).

1.1.5. Bağımsız Değişken Matrisi Rankının Gözlem Sayısından Küçük Olması

Bağımsız değişken (X) matrisi nxk boyutludur. İlişkideki parametre kestirimlerinin yapılabilmesi için gözlem sayısının parametre sayısından fazla olması gerekir. Gözlem sayısı parametre sayısına eşit olduğunda (n=k) yalnız bir ilişkinin varlığından söz edilmektedir. Sonsuz sayıda bir ilişki ise n<k olduğu durumda ortaya çıkmaktadır. Bu ise parametre sayısının gözlem sayısından fazla olduğu bir varsayımıdır (5).

1.1.6. Bağımsız Değişkenler Arasında Bir İlişki Olmaması

X matrisinin vektör veya sütunları birbirine bağlı olursa

$$B=(X'X)^{-1}X'Y \qquad [9]$$

formülündeki $|X'X| = 0$ olur. Bu durumda $(X'X)^{-1}$ matrisi olmayacaktır. Regresyon analizinde parametre kestirimlerinin tutarlılığını sağlamak ve varyansları en küçükleyebilmek amacıyla kabul edilen bu varsayımlar, bazı araştırmalarda gerçeği tam olarak yansıtmaz ve bunun sonucunda sapmalar otaya çıkar (3).

1.2. Varsayımlardan Sapmalar

Değişkenler arasındaki ilişkinin çoklu regresyon analizi ile belirlenebilmesi için çoklu doğrusal regresyon modelinin varsayımlarının sağlanması gerekir. Ancak çoklu regresyon modelinin varsayımları, her olayın analizinde geçerli olmamaktadır. Çoğu zaman bu varsayımlardan sapmalar görülmektedir. Bunlar değişen varyans, otokorelasyon ve çoklu doğrusal bağıntıdır.

1.2.1. Değişen Varyans

Çoklu doğrusal regresyon modelinin varsayımlarından birisi olan sabit varyans varsayımlarından sapma, değişen varyans olarak adlandırılmaktadır (heteroscedasticity). Bu durumun ortaya çıkmasının nedenlerinden biri, bağımsız değişkenlerin değerlerinin birbirinden çok farklı olmasıdır. Bu duruma bağlı olarak, hata terimlerinin değerleri de farklı büyüklükte olacaktır (6).

Hata terimlerinin değişen varyans karakterine sahip bulunduğu modellerde hata terimleri (e_j) arasında otokrelasyon olmadığı ve dağılımının normal olduğu kabul edilmektedir. Bir gözlemdeki hata terimi, diğer gözlemdeki hata terimine bağlı değildir ve her gözleme ait hata teriminin varyansları birbirinden farklıdır. Bu durumda varyans-kovaryans matrisi,

$$\begin{bmatrix} \sigma_{11}^2 & 0 & \cdots & 0 \\ 0 & \sigma_{22}^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \sigma_{nn}^2 \end{bmatrix} \qquad [10]$$

olacaktır. Yani hata terimlerinin varyansları $\sigma_{11}^2, \sigma_{22}^2, \ldots, \sigma_{nn}^2$ ve kovaryansı ise sıfırdır.

Değişen varyans durumunda doğrusal regresyon modelinin uygulanması, parametrelerin kestirim değerlerinin sapmasız olmasını etkilemez. Fakat kestirimlerin standart hatalarının büyük olmasına yol açar. Bu durumda değişkenler arasındaki ilişkiyi açıklayan parametre kestirimlerinin etkinliği azalacağından değişen varyansın belirlenmesi gerekir.

Değişen varyansın belirlenmesinin nedeni, parametre kestirimlerinin hatalarını küçültmek ve etkinliğini arttırabilmektir. Değişen varyansın belirlenmesinde, parametrik ve parametrik olmayan, Glejser testi, Goldfield-Quandt testi ve sıra korelasyon testi gibi testler kullanılmaktadır.

Yapılan testlerle varyansların sabit olmadığı belirlendiğinde, söz konusu bu olumsuzluğu gidermek için bazı çalışmalar yapmak gerekir. Bu amaçla;

- Matematiksel model değiştirilir.
- Modele alınmayan değişkenlerden bazıları modele dahil edilir
- Gözlem sayısı arttırılır.

1.2.2. Otokorelasyon

Doğrusal regresyonun modelinin varsayımlarından sapmalarının biri de otokorelasyondur. Otokorelasyon, çoklu regresyon modelindeki hata terimlerinin birbirinden bağımsız olduğu varsayımından olan sapmayı belirtir. Eğer hata terimleri birbirleriyle ilişkili ise, yani otokorelasyon varsa, $E(\varepsilon_i\varepsilon_j) \neq 0$ olacaktır (5, 6).

Değişkenler arasındaki ilişkiyi belirleyen matematiksel modelin yanlış seçilmesi otokorelasyonun nedenlerinden biridir. Kullanılan modele göre regresyon çözümlemesi sonucu hata terimlerinin grafiği çizilir ve hata terimleri grafikte düzenli bir görünüme sahipse otokorelasyon olması beklenir. Eğer grafik düzensiz bir görünümde ise otokorelasyon söz konusu değildir ve bu durumda çoklu regresyon modelinin varsayımı için $E(\varepsilon_i\varepsilon_j)=0$ geçerli olacaktır.

Otokorelasyonun ortaya çıkmasının diğer bir nedeni ise bazı bağımsız değişkenlerin ilişkiye dahil edilmemiş olmasıdır.

Hata terimlerinde otokorelasyon olması, değişkenler arasındaki ilişkiyi belirleyen parametrelerin kestirim değerlerinin gerçek değere yakın olmasını, yani tutarlılığı olumsuz yönde etkilediğinden sebep-sonuç ilişkisini belirlemede hatalara neden olur. β’nın kestirimleri sapmasızdır, fakat varyansları minimum değildir. σ_ε^2 ‘nin kestirimi ise sapmalıdır. Bunun sonucu olarak β’nın varyansları da sapmasız olarak kestirilmiş olur. Hata terimlerinde otokorelasyon olup olmadığını belirlemek için Durbin-Watson, Von-Neumann gibi testler kullanılır (5).

1.2.3. Çoklu Doğrusal Bağıntı

Çoklu regresyonun varsayımlarından biri de, bağımsız değişkenler arasında ilişkinin olmaması durumudur. Bir regresyon modelindeki tahmin ediciler “bağımsız değişkenler” olarak adlandırılırlar. Fakat bu terim, elimizdeki tahmin edici değişkenlerin istatistiki anlamda birbirinden bağımsız olduğu anlamına gelmemektedir. Normalde tahmin ediciler birbirleriyle yüksek ilişkili olabilirler. Bir regresyon modelinin amacı, veri setinden en uygun ve etkin tahminleri seçerek bir model oluşturmaktır. Eğer değişkenler arasında ilişki yoksa bu değişkenlere ortogonal denir. Regresyonda bazı değişkenler arasında yüksek ilişki olduğu zaman ise bu duruma çoklu doğrusal bağıntı sorunu denir. Çoklu doğrusal bağıntı terimi, tahmin ediciler arasında yüksek ilişki olduğu durumlarda kullanılmaktadır (5, 6).

GEREÇ VE YÖNTEM

2. Çoklu Doğrusal Bağıntı

En küçük karelerin uygulanabilmesi için çok önemli bir koşul, açıklayıcı değişkenlerin kendi aralarında tam bir doğrusal bağıntıya sahip olmamalarıdır. Çoklu doğrusal bağıntı, bir regresyon modelinde, bir veya birden fazla bağımsız değişkenin kendi aralarında ilişkili olma durumudur (7). Eğer açıklayıcı değişkenler arasında tam bir korelasyon var ise, yani bu değişkenler için korelasyon katsayısı 1'e eşitse, parametreler belirlenemez hale gelir. Her bir parametre için ayrı ayrı sayısal değerler bulmak olanaksızlaşır ve bu durumda en küçük kareler yöntemi kullanılamaz. Diğer taraftan, eğer açıklayıcı değişkenler arasında bir bağıntı yoksa, yani bu değişkenler için korelasyon katsayısı 0'a eşitse, bunlara ortogonal değişkenler denir ve katsayıların tahmininde bir sorun oluşturmamaktadır (8, 9).

Uygulamada, yukarıda bahsedilen uç durumlarla karşılaşılması, genellikle pek olası değildir. Çoğu zaman, bir çok iktisadi veya tıbbi değişkenin zaman içerisinde birbirlerine bağlı olmaları nedeniyle, açıklayıcı değişkenler arasında bir miktar çoklu doğrusal bağıntı meydana gelir. Bu durumda her açıklayıcı değişken çifti için basit korelasyon katsayısı 0 ile 1 arasında bir değer alacak ve çoklu doğrusal bağıntı sorunları, parametre tahminlerinin doğruluğunu ve kararlılığı belki de sakatlayacaktır. Ancak çoklu doğrusal bağıntının kesin etkileri henüz kuramsal olarak saptanmış değildir.

Çoklu doğrusal bağıntı, iktisadi fonksiyonlarda var ya da yok diyebileceğimiz bir koşul değildir, aksine iktisadi büyüklüklerin doğası gereği çoğu ilişkide içerilmiş bir olgudur. Eğer varsa, parametre tahminlerini ciddi biçimde etkileyebilecek olan çoklu doğrusal bağıntının derecesine ilişkin kesin kanıtlar yoktur. Sezgisel olarak, iki açıklayıcı değişken aşağı yukarı aynı biçimde değişiyorlarsa, her birinin Y üzerindeki tekil etkisini saptamak çok güçleşmektedir. Örneğin, osteoporozlu bir hastaya ait kemik dansitesi, kan kalsiyum iyonu seviyesi, kandaki fosfor seviyesine bağlıdır. Eğer belli bir dönem boyunca bu parametreler aynı oranda değişmişlerse, bunlardan birinin kemik dansitesi üzerinde etkisi yanlışlıkla ötekinin sanılabilir. Aralarında yüksek çoklu bağıntı nedeniyle bu değişkenlerin hastaya ait kemik dansitesi üzerindeki etkileri sağlıklı biçimde araştırılamaz (8).

Çoklu doğrusal bağıntı çeşitli nedenlerden dolayı meydana gelebilir. Söz konusu olası nedenlerin en önemlileri aşağıda sıralanmıştır;

- Sağlık alanında ilgilenilen değişkenlerin zaman içinde birlikte değişme eğilimi göstermesi. Tıbbi değişkenler aynı etmenlerden etkilenebilirler ve bu etmenler bir kez devreye girdi mi bu değişkenler zaman içinde kabaca aynı eğilimi gösterirler. Örneğin, bir hastalık esnasında kişiye ait parametreler, bazıları gecikmeli de olsa, hep beraber artış ya da azalış gösterebilirler.
- Bazı açıklayıcı değişkenlerin gecikmeli değerlerinin ilişkide ayrı birer etmen olarak kullanılması.
- Araştırmacının bilerek ya da bilmeyerek bağımsız değişkenler kümesinden yalnızca bir alt kümeyi örneklem olarak alması.
- Araştırmaya esas olan olayın içinde bulunduğu sosyal ve ekonomik koşullar nedeniyle temelde var olan kısıtlar çoklu doğrusal bağıntının kaynağını oluştururlar.
- Modelin yanlış seçilmesi.

Bağımsız değişken sayısının gözlem sayısından büyük olması durumunda da çoklu doğrusal bağıntı sorunuyla karşılaşılır ve bu durum geniş tanımlı (overdefined) model olarak ifade edilir. Bu modellere genellikle tıp ve davranışla ilgili araştırmalarda rastlanır. Bu tip araştırmalarda örnekleme birimlerinin sayısı az olmasına karşın bağımsız değişken sayısı oldukça fazladır.

Çoklu regresyon modelinde, bağımsız değişkenler arasında yüksek derece ilişki olması varsayımı çoklu doğrusal bağıntı olarak adlandırılır. Çoklu doğrusal bağıntı meydana geldiğinde, aşağıdaki problemlere yol açmaktadır:

- En küçük kareler, tahmin edilen parametrelerin gerçek değerlerinden oldukça farklı olacaktır.
- Regresyon katsayıları belirsiz ve bu katsayıların standart hataları sonsuz olmaktadır.
- Regresyon katsayılarının varyans ve kovaryansları artmaktadır.
- Modelin R^2 değeri yüksek, ancak bağımsız değişkenlerden hiçbiri veya çok azı kısmi t testine göre anlamlı olacaktır.
- İlgili bağımsız değişkenlerin bağımlı değişkenle olan ilişkilerinin yönü kuramsal ve deneysel beklentilerle çelişebilmektedir (10, 11).

2.1. Çoklu Doğrusal Bağıntıyı Tespit Etme Yöntemleri

2.1.1. Frisch'in Kavşak Çözümlemesine Dayalı Bir Yöntem

Frisch'in Kavşak Çözümlemesinde, bağımlı değişkenin, açıklayıcı değişkenlerden her biriyle ayrı ayrı regresyonu yapılır ve başlangıç regresyonları elde edilir. Bu başlangıç regresyonlarından sezgisel veya istatistiksel ölçüler ile inandırıcı olanı seçilir ve sonra öteki değişkenler kademe kademe ilave edilir. Her ilave edilen değişkenden sonra, bu değişkenin, tek tek katsayıları, katsayıların standart hataları ve genel R^2 üzerindeki etkileri incelenir. Burada elde edilen sonuçlarla da;

- Eğer ilave değişken, katsayıları anlamsız yapmıyor ve R^2'yi de yükseltiyorsa faydalı kabul edilir ve modele alınır.
- Eğer ilave değişken, katsayıları etkilemiyor ve R^2'yi yükseltmiyorsa gereksiz sayılır ve modele alınmaz.
- Eğer yeni değişken, katsayıların işaretlerini ya da değerlerini önemli ölçüde etkiliyor ise, zararlı sayılır. Eğer tek tek katsayılar, sezgisel, teorik nedenlerle kabul edilmez hale geliyorsa, bu çoklu doğrusal bağıntının bir işareti olarak da kabul edilir (8, 12).

2.1.2. Yardımcı Regresyon Kriteri

Çoklu doğrusal bağıntı şu iki durumu ifade eder:

- Bağımsız değişkenler arasında tam ilişki olması hali

$$k_1x_1 + k_2x_2 + \cdots + k_kx_k = 0 \qquad [11]$$

x_1 değerini alan yardımcı değişken (sabit terimin değişkeni), k_1, k_2, …, aynı anda sıfıra eşit olmayan sabit sayılar.

- Bağımsız değişkenler arasında tama yakın ilişki olması hali

$$k_1x_1 + k_2x_2 + \cdots + k_kx_k + e_i = 0 \qquad [12]$$

İlk durumda, bir veya birden daha fazla değişken, diğerlerinin tam doğrusal kombinasyonu iken, ikinci durumda yaklaşık doğrusal kombinasyonudur.

İşte, çoklu doğrusal bağıntı, yukarıdaki özelliklerinden faydalanılarak şu yolla da varlığının tespitine çalışılmaktadır. Hangi x değişkeninin, diğer x değişkenleriyle bağlantılı olduğunu bulabilmek için, x_i değişkeni ile kalan x değişkenleri arasındaki doğrusal regresyon araştırılıp çoklu belirlilik katsayısı R^2 hesaplanır. Şöyle ki,

$$y = b_1 + b_2x_2 + b_3x_3 + b_4x_4 + e \qquad [13]$$

Yardımcı regresyonlar,

$$\begin{aligned} x_2 &= a_{12} + a_{32}x_3 + a_{42}x_4 + e_2 \quad ,R^2_{x_2} \\ x_3 &= a_{13} + a_{23}x_3 + a_{43}x_4 + e_3 \quad ,R^2_{x_3} \\ x_4 &= a_{14} + a_{24}x_2 + a_{34}x_3 + e_4 \quad ,R^2_{x_4} \end{aligned} \qquad [14]$$

Yukarıdaki regresyonlardan her birine yardımcı regresyon denilmektedir. Yardımcı regresyonlar, y ile x arasındaki asıl regresyonun yardımcıları olmaktadır. R_i^2'ler hesaplandıktan sonra aşağıdaki F_i oranı ile yardımcı regresyonların anlamlılığı test edilir. f_1=k-2, f_2=n-k+1 serbestlik dereceleriyle belli bir anlamlılık seviyesi α'ya göre bulanacak tablo F_T değeri ile bu F_i oranı karşılaştırılır. Eğer $F_i > F_T$ ise x_i ile diğer x'ler arasında çoklu doğrusal bağlantı varlığına karar verilir. Aksi halde çoklu doğrusal bağlantı olmadığına karar verilir. F_i'nin istatistiki olarak anlamlı olması, yani tablo değerinden büyük olması halinde ilgili x_i değişkeni modelden çıkarılabilir. Bu durumda da önemli bir spesifikasyon hatası yapma durumu ile karşı karşıya geliriz.

$$F_i = \frac{\mathrm{R}^2_{x_i,x_2x_3\ldots x_k} / (k-2)}{(1-\mathrm{R}^2_{x_i,x_2x_3\ldots x_k}) / (n-k+1)} \qquad [15]$$

k: modeldeki değişken sayısı (Y dahil)

$R^2_{x_i,x_2x_3\ldots x_k}$: x_i ile kalan x değişkenleri arasındaki belirlilik katsayısı

2.1.3.Varyans Artış Faktörü (Variance Inflation Factor, VIF)

Çoklu doğrusal bağıntının saptanmasında kullanılan bir başka yaklaşım, varyans artış faktörünün (VIF= Variance Inflation Factor) kullanılmasıdır. İlk olarak Farrar ve Glauber (1967) tarafından çoklu doğrusal bağıntı belirlemek için kullanılmış, fakat Marquardt

(1970) tarafından varyans artış faktörü olarak adlandırılmıştır. Bağımsız değişkenlere ilişkin korelasyon matrisinin tersinin köşegen öğelerine varyans artış faktörü denir. Varyans artış faktörü bir bağımsız değişkenin diğer bağımsız değişkenlerle olan ilişkisinin derecesini belirlemek için hesaplanır (7). Varyans artış faktörlerinin hesaplanmasını göstermek için aşağıdaki gibi üç bağımsız değişkenli bir regresyon modelini inceleyelim.

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + e_i \qquad [16]$$

Üç bağımsız değişkenli bir modelin varyans artış faktörü değerleri aşağıdaki adımlarla hesaplanmaktadır.

- Birinci adımda, X_1 bağımlı, X_2 ile X_3 bağımsız değişken olarak alınıp, modelin R_1^2'si hesaplanır. Böylece X_1 için varyans artış faktörü $VIF(X_1) = 1/(1 - R_1^2)$ olarak hesaplanır.
- İkinci adımda, X_2 bağımlı, X_1 ile X_3 bağımsız değişken olarak alınıp, modelinR_2^2'si hesaplanır. Böylece X_1 için varyans artış faktörü $VIF(X_2) = 1/(1 - R_2^2)$ olarak hesaplanır.
- Üçüncü adımda, benzer şekilde X_3 bağımlı, X_1 ile X_2 bağımsız değişken olarak alınıp, modelin R_3^2'si hesaplanır. Böylece X_3 için varyans artış faktörü $VIF(X_3) = 1/(1 - R_3^2)$ olarak elde edilir.

Bağımlı değişken ile bağımsız değişkenler arasında ilişki yoksa ($R^2 = 0$ olacağından), varyans artış faktörü 1'e $(1 / (1-R^2) = 1 / (1-0) = 1)$ eşittir. Bağımlı ve bağımsız değişkenler arasında tam bir ilişki varsa ($R^2=1$ olacağından) varyans artış faktörü $[VIF=1/(1-R^2)=1/(1-1) = \infty]$ sonsuz olacaktır. R^2 = %90 ise varyans artış faktörü 10 [VIF=1/(1-0,90) = 10] olarak elde edilir. Genelde, yorum için şu genel kural söz konusudur: varyans artış faktörü 10'a eşit veya daha büyük (VIF $\geq$10) ise, çoklu doğrusal bağıntı problemi mevcuttur (10).

2.1.4.Yüksek R^2 Hesaplanmasına Rağmen Regresyon Katsayılarının Anlamsız Olması

Çoklu doğrusal bağıntının varlığının en belirgin belirtisi, bir modelde çok yüksek R^2 değeri hesaplandığı halde regresyon katsayılarının t testi sonucunun anlamsız olmalarıdır.

Bilindiği gibi R^2 yüksek olduğunda (genellikle 0,80'i aştığında) b regresyon katsayılarının topluca F testi genellikle olumlu sonuçlar vermekte, topluca 0'dan farklı olduklarını göstermektedir. Fakat F testi böyle olumlu iken regresyon katsayılarının t testi sonuçları olumsuz sonuçlar veriyorsa, bunu çoklu doğrusal bağıntının bir sonucu olarak yorumlayabiliriz. Çünkü regresyon katsayılarının varyansları çoklu doğrusal bağıntı nedeniyle büyümekte, dolayısıyla t değerleri küçülmektedir. İki bağımsız değişkenli model için b_1 ve b_2 katsayılarının varyansları,

$$V(b_1) = \frac{s^2}{\sum x_1^2(1-r^2)} \;,\; V(b_2) = \frac{s^2}{\sum x_2^2(1-r^2)} \qquad [17]$$

eşit olmaktadır. Burada r^2, x_1 ile x_2 değişkeni arasındaki korelasyon katsayısının karesidir. Yani,

$$r^2 = \frac{(\sum x_1 x_2)^2}{\sum x_1^2 x_2^2} \qquad [18]$$

$V(b_1)$ ve $V(b_2)$ eşitliklerinde görüldüğü üzere r değerindeki artış, yani çoklu doğrusal bağıntıdaki artış, katsayı varyanslarının ve dolayısıyla standart hataların değerinde artışa neden olacaktır.

Varyanslar yalnızca r'ye değil, bunun dışındaki iki faktöre de bağlıdır. Bunlar tahminin varyansı s^2 ile $\sum x_1^2 = \sum(x_1 - \bar{x}_1)^2$ ve $\sum x_2^2 = \sum(x_2 - \bar{x}_2)^2$ dir.

Böylece x_1 ile x_2 arasında yüksek ilişki olma durumunda, eğer s^2 yerine yeterince düşük ve/veya x_1 ve x_2'nin değişkenliği yeterliyse, küçük değerli katsayı varyansları elde etmek mümkündür. Aksi halde x_1 ile x_2 arasında çoklu doğrusal bağlantı bulunmamasına karşın s^2'nin yüksek olması ya da açıklayıcı değişkenlerin yetersiz olma durumunda yüksek değerli katsayı varyansları elde etmek mümkün olabilir. Bağımsız değişkenler arasında çoklu doğrusal bağıntı kararı verirken bu iki durum göz önünde tutulmalıdır.

2.1.5. X'X matrisinin özdeğerinin analizi

Çoklu doğrusal bağıntı problemini ciddi anlamda çalışan ilk araştırmacı Ragnar Frisch (1934), çoklu doğrusal bağıntıyı özdeğerlerle ilişkilendirmiştir. Fakat bilgisayar kullanımının kısıtlı olması nedeniyle (X'X) in özdeğerlerinin nümerik analizi desteklenememiştir. (X'X) matrisi karakteristik yapısı, çoklu doğrusal bağıntının tespitinde yaygın şekilde kullanılan önemli araçlardan biridir. (X'X) korelasyon matrisinin özdeğerleri $\lambda_1, \lambda_2, \ldots, \lambda_p$ ve özvektörleri $v_1, v_2, \ldots, v_p$ olmak üzere, çoklu doğrusal bağıntı durumunda en az bir özdeğer sıfıra çok yakındır ve en az bir özvektör aşağıdaki eşitliği sağlar (13).

$$\sum_{i=p}^{p} v_{ij} X_i = 0 \qquad [19]$$

Burada v_{ij}, j. özvektörün i. elemanı, X_i ise X matrisinin i. sütununu göstermektedir. Özdeğerlerin tek tek incelenmesi pek fazla anlam taşımamaktadır. Bu yüzden birçok istatistikçi özdeğerlerin birbirleriyle kıyaslanması yoluna giderek çoklu doğrusal bağıntının varlığını ve onun derecesini belirlemeye çalışmaktadır. Vinod ve Ullah (1981) en büyük özdeğerin en küçük özdeğere bölümünün karekökü olarak tanımladıkları "şartlı sayı"nın, çoklu doğrusal bağıntı düzeyinin belirlenmesinde iyi bir gösterge olduğunu belirtmişlerdir.

Montgometry ve Peck (1982), şartlı sayı'yı en büyük özdeğerin en küçük özdeğere oranı $k=\lambda_{max} / \lambda_{min}$ olarak tanımlamışlardır. Eğer k, 100'den küçükse çok ciddi boyutta çoklu doğrusal bağıntı problemi olmadığı, 1000'den büyükse veri içerisinde birden fazla çoklu doğrusal bağıntı problemi olduğu kabul edilmektedir. Özdeğerlerin toplamları da çoklu doğrusal bağıntı düzeyini belirlemede kullanılmaktadır. Bu toplam ne kadar büyükse çoklu doğrusal bağıntı düzeyinin o derece yüksek olduğu kabul edilir (8, 14).

2.1.6. Koşul İndeksi

Çoklu doğrusal bağıntının belirlenmesinde bir başka istatistiki yöntemde de koşul indeksinin (condition index, CI) kullanılmasıdır. Koşul indeksi aşağıdaki gibi hesaplanır.

$$CI = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} = \sqrt{K} \qquad [20]$$

Eğer koşul indeksi (CI), 10 ile 30 arasında ise orta dereceden şiddetli derecede çoklu doğrusal bağıntı, eğer 30'un üzerinde ise aşırı derecede çoklu doğrusal bağıntı söz konusudur (15).

Bunun yanı sıra çok kesin olmamakla birlikte çoklu doğrusal bağıntıyı ortaya çıkaran diğer bazı bulgulardan da söz edilebilir. (X'X) matrisinin determinantı çoklu doğrusal bağıntı göstergesi olarak kullanılabilir. (X'X) matrisi korelasyon biçimindeyken determinantının alacağı değer aralığı $0 \leq |X'X| \leq 1$ dir. Eğer $|X'X| = 1$ ise açıklayıcı değişkenler ortogonaldir. $|X'X| = 0$ ise açıklayıcı değişkenler arasında tam çoklu doğrusal bağıntı mevcuttur. Bu gösterge kolayca uygulanabilir olmasına rağmen çoklu doğrusal bağıntının kaynağı konusunda herhangi bir bilgi vermez.

Çoklu doğrusal bağıntının tespitinde kullanılan bütün bu farklı yaklaşımlar, sorunun daha açık ve kesin olarak ortaya çıkarılabilmesi için birlikte değerlendirilmelidir. Özdeğerlerin büyüklükleri, bağımlılığın derecesini belirlerken, varyans artış faktörü verideki olası bağımlılıkların her bir regresyon katsayısı üzerindeki olumsuz etkilerini açıklayabilir. Ancak buna rağmen araştırmacılar, çoklu doğrusal bağıntı derecesini ölçmede sadece şartlı sayı kriterini kullanmayı tercih etmektedirler.

2.1.7. Farrar-Glauber Çoklu Doğrusal Bağıntı Sınaması

Farrar ve Glauber tarafından çoklu doğrusal bağıntının tespit edilmesi için geliştirilmiş olan bu yöntem, çoklu doğrusal bağıntının olup olmadığını ortaya koyacak olan χ^2 sınamasıdır.

Farrar ve Galuber, bir örneklemdeki çoklu doğrusal bağıntıyı, gözlemlenen X'lerin ortogonallikten bir sapması olarak ele almaktadır. Ortogonallikten sapma arttıkça, yani determinantın değeri sıfıra yaklaştıkça, çoklu doğrusal bağıntı güçlenir. Bu olgudan yola çıkan Farrar ve Glauber, bütün bağımsız değişkenler kümesi için çoklu doğrusal bağıntı gücünü saptamak için aşağıdaki χ^2 sınamasını önermektedirler.

H_0 : X'ler ortogonaldir.

H_1 : X'ler ortogonal değildir.

χ^2 test istatistiği ise,

$$\chi_h^2 = -\left[n - 1 - \left(\frac{1}{6}\right) * (2k + 5)\right] + \ln|X'X| \qquad [21]$$

şeklinde hesaplanır ve serbestlik derecesi;

s.d.=(1/2) * k(k-1)

olarak belirlenmektedir. $\chi^2_h > \chi^2_{\alpha; sd}$ ise, sıfır önsavı reddedilir ve çoklu bağıntı olduğuna karar verilir. χ^2 değeri arttıkça, çoklu bağıntının derecesi de artar (16).

2.2. Çoklu Doğrusal Bağıntıyı Giderme Yöntemleri

Bağımlı ve bağımsız değişkenler arasındaki ilişkiyi belirleyen katsayı kestirimlerinin standart hatasını küçültmek ve daha tutarlı kestirim yapabilmek için çoklu doğrusal bağıntının giderilmesi zorunludur.

2.2.1. Daha fazla bilgi toplama

Çoklu doğrusal bağıntı sorununa bir çözüm olarak Farrar ve Glauber daha fazla veri toplamayı önermişlerdir. Bu yöntemde gözlem sayısının arttırılmasının yanında bu verilerin bilgi kapsamlarının arttırılması da söz konusudur. Başlangıçta yıllık verilerin kullanılmış ve çoklu doğrusal bağıntı ortaya çıkmış ise, 3 aylık veya aylık veriler denenmelidir. Gözlem sayısının arttırılması, parametre kestirimlerine ait varyansların küçülmesini sağlayabilir (3). Benzer şekilde C.F.Christ örneğin büyütülmesiyle, bir denklemdeki çoklu doğrusallıktan doğan, tahmin edilmiş parametreler arasındaki, yüksek ortak varyansların düşürülebileceğini, çünkü bu ortak varyansların örneklem büyüklüğüyle ters orantılı olduğunu söylemektedir. Sonuç olarak, eğer mümkün ise, daha fazla veri toplayarak parametre kestirimlerinin yeniden yapılması yerinde olur (8).

2.2.2.Bağımsız Değişkenlerden Bir veya Birkaçının Modelden Çıkartılması

Çoklu regresyon modelinde iki veya daha fazla bağımsız değişken arasında önemli ölçüde bir ilişki olduğunda, bu değişkenlere ait regresyon katsayılarının en küçük kareler kestirimleri sapmasız, fakat bu kestirimlerin varyansları büyük olur. Varyansı küçültmek için bağımlı değişken üzerinde daha az etkili olan bağımsız değişkenler regresyon modelinden çıkartılır. Bunun sonucunda regresyon modelinde kestirimlerin varyansları daha küçük, dolayısıyla kestirimlerin güvenirliği yüksek olur. Bu yöntem uygulandığında, regresyon modelindeki parametrelerin kestirimleri sapmalı olur. Modelden çıkartılacak değişkenin değişimini açıklamada “çok önemli” olmadığı durumlarda bu yolun uygulanması önerilir (5).

2.2.3. Modelin Yeniden Belirlenmesi

Çoklu doğrusal bağıntının nedeni, model seçiminden kaynaklanabilir. Örneğin, ilişkili açıklayıcı değişkenlerin kullanılması gibi. Bu durumda ya açıklayıcı değişkenler yeniden tanımlanır, ya da ilişkili açıklayıcı değişkenlerin biri çıkartılır. Ancak açıklayıcı değişkenlerin çıkartılması modelin etkinliğini azaltabilir. Çünkü Lipovetsky ve Conklin’in (2001) belirttiği gibi çoklu doğrusal bağıntı olsa bile değişkenler birbirini tam olarak temsil etmeyebilirler. Açıklayıcı değişkenlerin her biri uyumda ve bağımlı değişkenin yapısını açıklamada özel bir role sahip olabilirler (11).

2.2.4. Oranların veya Birinci Farkların Kullanılması

Çoklu doğrusal bağıntının varlığında bazı zaman serisi analizlerinde esas veriler yerine birincil farkları kullanılabilmektedir. Bu durumda birbiriyle bağlantısı olan bağımsız değişkenlerin birinci farkları o denli birbiriyle ilişkili olmayabilmektedir. Fakat çoklu doğrusal bağıntının bu şekilde giderilmesi hata terimlerine etki ederek otokorelasyona neden olduğundan, kestirimlerin etkinliğini azaltır.

Kullanılabilecek diğer bir yöntem ise, esas değişken yerine bir oranın kullanılmasıdır. Modeldeki bütün değişkenler bağımsız değişkenlerden birisine oranlanır. Bu uygulama çoklu doğrusal bağıntı sorununa biraz olsun çözüm getirebilmesine rağmen, hata terimlerinin değişen varyanslı olmasına neden olur. Hata terimlerinin değişen varyanslı olması ise parametre kestirimlerinin standart hatalarının büyük olmasına ve Y’lerin kestirim etkinliğinin azalmasına yol açar.

Çoklu doğrusal bağıntının giderilmesi için oranların veya birinci farkların kullanılması, regresyon modelinde kabul edilen varsayımları geçersiz kıldığından, sağlıklı bir çözüm yolu değildir (5).

2.2.5.Değişkenleri Dönüştürme Yöntemi

Daha çok zaman serisi verileri için kullanılan, çoklu doğrusal bağıntıyı giderme yöntemi, verilerin orjinal değerleri yerine dönüşümlü değerlerini alarak model tahminidir. Değerleri aynı zamanda yıldan yıla birlikte artan veya azalan değişkenler arasındaki ilişkinin işareti pozitif, aynı yöndedir. Söz konusu ilişki, bağımsız değişkenler arasında kuvvetli ise çoklu doğrusal bağıntı söz konusudur. Böyle bir zaman serisi modelinin aşağıdaki gibi olur.

$$y_t = \beta_0 + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \qquad [22]$$

Burada X'ler arasındaki bağımlılığın derecesini azaltabilmek için yukarıdaki model bir dönem gecikmeli olarak alınır;

$$y_{t-1} = \beta_0 + \beta_1 X_{1t-1} + \beta_2 X_{2t-1} + \beta_3 X_{3t-1} + u_{t-1} \qquad [23]$$

Daha sonra, ikinci denklemden birinci denklem çıkartılır;

$$y_t - y_{t-1} = \beta_1 (X_{1t} - X_{1,t-1}) + \beta_2 (X_{2t} - X_{2,t-1}) + ... + v_t \qquad [24]$$

$$v_t = u_t - u_{t-1}$$

elde edilir. Son elde edilen eşitliğe "birinci farklar biçimi" denilir. Burada Orjinal y ve X verileri yerine değişkenlerin birbirini takip eden değerleri arasındaki farklar alınarak regresyon tahmini yapılmaktadır. İlk modelde X'ler arasında kuvvetli ilişki olduğu halde, son dönüşümlü modelde çoklu doğrusal bağıntı önemli derecede azaltılmış olmaktadır. Zira X_1 ve X_2 arasında kuvvetli ilişki varken, bir önceki yıl değerlerine göre farkları arasında da önemli bir ilişki olacağına dair bir sebep yoktur.

Birinci farklar dönüşümünün sakıncası, orijinal modelin hata terimi u_t'nin değerleri arasında otokorelasyon olmadığı halde, dönüşümlü model hata terimi v_t otokorelasyonlu olabilmektedir (12).

2.2.6.Yanlı Kestirim Yöntemleri

2.2.6.1.Shrunken Regresyon Yöntemi

Shrunken regresyon yöntemi, yanlı kestirim yöntemlerinden birisidir. Bu yöntemle parametre kestirimi yapılabilmesi için bağımsız değişken sayısı ikiden büyük (k>2) ve (X'X) matrisi birim matrise eşit (X'X=1) olması gerekir.

Shrunken kestiricisi $\hat{\beta}_s$, a ve b sabit olmak üzere aşağıdaki gibidir:

$$\hat{\beta}_s = \left\{1 - \frac{b(1-R^2)}{a(1-R^2)+R^2}\right\} \qquad [25]$$

Shrunken kestiricisi $\hat{\beta}_s$, EKK kestiricisinden daha küçük hata kareler ortalamasını vermektedir.

Bu yöntemle bulunan parametre kestirimlerinin en küçük hatayla bulunabilmesi,

$$a = 0 \; ve \; b = \frac{k-2}{n-k+2} \qquad [26]$$

olmasına bağlıdır.

Çoklu doğrusal bağıntı varlığında X'X matrisi birim matrise eşit değildir ve bu nedenle Shrunken regresyon yöntemiyle değişken seçimi yapılırken en küçük özdeğere karşılık gelen değişkenler modelden çıkarılır. X, (X'X) ' in özvektörlerinin matrisi olmak üzere Shrunken kestiricisi;

$$\hat{\beta}_s(r) = X \begin{bmatrix} \mathrm{I}_r \; 0 \\ 0 \; \mathrm{dI}_{k-r} \end{bmatrix} X \hat{\beta}_{\mathrm{EK}} \qquad [27]$$

formülüyle bulunur. Burada;

$\hat{\beta}_s(r)$: En küçük özdeğere karşılık gelen bağımsız değişkenler çıkartıldıktan sonra elde edilen (β) parametre tahmincisi.

r: Çıkarılan bağımsız değişken sayısıdır.

Modelde bırakılan bağımsız değişkenlerle bağımlı değişkenler arasındaki ilişki katsayısı olan $\hat{\beta}_s(r)$'nin hesaplanmasında d katsayısının seçimi bir problem olarak görülmektedir.

V_r: Çıkarılan değişkenlere ait özvektörlerin matrisi, Λ_r ise çıkarılan değişkenlere ait (X'X) matrisinin özdeğerlerinden köşegen matrisi olmak üzere d'nin aşağıdaki şekilde hesaplanması önerilmektedir.

$$d = enb\left\{0,\ 1\text{-c}\ (1\text{-R}^2)\, w^{-1}\right\}$$

$$w = \hat{\beta}' V_r \Lambda_r V_r{}' \hat{\beta}$$

$$0 \le c \le 2(r-2)/(n-k+2) \qquad [28]$$

Gunst ve Mason isimli istatistikçilere göre d katsayısı 0,823 ile 0,860 arasında değer alabilmektedir (5, 12).

2.2.6.2.Temel Bileşenler Regresyon Yöntemi

Çoklu doğrusal bağıntıyı gidermek için ileri sürülen yanlı tahmin edicilerden birisi de temel bileşenler tahmin edicisidir. Temel bileşenler analizini ilk olarak Hotelling (1933), Sperman'ın yaptığı çalışmalara dayanarak ortaya koymuştur. Çok değişkenli istatistiksel yöntemler içersinde, kovaryans yapısının incelenmesinde kullanılan etkin bir yöntem olan temel bileşenler analizi, başlangıç değişkenlerinin en büyük varyansa sahip standartlaştırılmış doğrusal bileşimlerinin araştırılması esasına dayanır. Verilerin çok boyutlu olması halinde bu yöntem başarı ile uygulanmaktadır. Analizde amaç, p değişkeni ile ilgili veri toplanmış ise, bu değişkenlerin belirlediği toplam değişkenliği ifade etmek üzere k sayıda ana bileşen bulmak, böylece daha az sayıda değişken ile çalışarak p boyutlu uzay yerine k boyutlu (k<p) bir uzayda çalışmak, böylece boyut indirgemek amaçlanır (17).

p tane değişken üzerine n gözlem yapıldığında, p boyutlu R_p uzayından n tane noktadan oluşan bir nokta bulutu elde edilmiş olur. Temel bileşenler analizinde çok boyutlu uzaydaki bu nokta bulutunu daha az boyutlu bir alt uzayda incelemektir. Bunu yaparken doğal olarak veri matrislerindeki bazı bilgiler kaybolacaktır. Ancak, yöntem bilgi kaybını en az yapacak şekilde en optimum çözümü bulma esasına dayanır.

Bir veri setinde yer alan p değişken arasında önemli kovaryans ya da korelasyonlar saptanıyorsa bu veri setini analizlerde doğrudan kullanmak yerine, bu setin doğrusal bileşenlerinden oluşan daha az sayıdaki doğrusal bileşenden oluşan bir veri setini kullanmak yararlıdır.

Temel bileşenler analizi, daha önceden ortaya çıkarılmamış ilişkileri ortaya çıkarma ve sıradan sonuçlar diye nitelendirilemeyecek tahminler yapmaya izin veren bir yöntemdir (12).

Temel bileşenler yönteminde, X_i değişkenler, (i=1,2,…,k) kümesinde X_i'lerin doğrusal bileşenleri olan ve ana bileşenler olarak adlandırılan yeni değişkenler (P_i) oluşturulur.

$$\begin{aligned} P_1 &= a_{11}X_1 + a_{12}X_2 + \ldots + a_{1k}X_k \\ P_2 &= a_{21}X_1 + a_{22}X_2 + \ldots + a_{2k}X_k \\ &\vdots \\ P_k &= a_{k1}X_1 + a_{k2}X_2 + \ldots + a_{kk}X_k \end{aligned} \qquad [29]$$

Ana bileşenler, P_1, P_2,...,P_k birbirleri ile ilişkili değildiler ve onlar varyanslarının büyüklüklerine göre sıralanırlar. Her bir ana bileşen özdeğerlere karşılık gelir (17).

Ana bileşenler yönteminin, en küçük kareler yöntemine göre, örnekteki bilginin daha azını kullandığı gerekçesiyle, çoklu doğrusal bağıntının etkilerinden kurtulmak için uygulanması bazı şüpheler taşımaktadır. Ekonomistler, örnekte yeterli bağımsız değişkenin olmaması nedeniyle çoklu doğrusal bağıntının, tahminleri bozduğu amacında birleşirler. Örnekteki bilgi, modelin bütün katsayılarının güvenli tahmin edilmesi için yetersizdir. Çoklu doğrusal bağıntının derecesi arttıkça, en küçük kareler Y'deki değişimi tek tek açıklayıcı değişkenler arasında, gittikçe daha gelişigüzel ve güven vermez biçimde dağıtır (8).

Daha önce önerilen düzeltici çözümler, model dışı tahminlerden, büyütülmüş örneklerden, bazı katsayılara ya da açıklayıcı değişlenlerin arasındaki ilişkilere konan kişisel sınırlamalardan elde edilen daha çok bilginin kullanılmasını içerirler. Temel bileşenler yönteminde, (X'lerin doğrusal bileşimlerinden) bazı yapay değişkenler çıkarılır. Böylece dikey değişkenlerdeki çoklu doğrusal X'leri dönüştürmüş oluruz. Eğer, bu yapma değişkenlere belli iktisadi anlamlar verilebiliyorsa, bunlar kendi başlarına değişkenler olarak kullanılabilirler ve bu dönüştürme, çoklu doğrusal bağıntı sorununa savunulabilecek bir çözüm getirir. Çünkü bu durumda, ilk modele ait parametrelerinin sayısında anlamlı bir azalma sağlanmış olur. Ancak çoğu durumda, yapma değişkenler doğrudan iktisadi değişkenler olarak yorumlanamaz ve kullanılmaları, örnekteki bütün bilginin yalnızca bir bölümünden (temel bileşenlere giren bölümünden) yararlanıldığı, yani örnekteki bilgide aranan çoğaltma yerine azaltma yapıldığı anlamına gelir (8).

2.2.6.3.Özdeğerler Regresyonu Yöntemi

Özdeğerler regresyonu yöntemi, temel bileşenler regresyonu yöntemine benzemektedir. Temel bileşenler regresyonundaki gibi bu yöntem de, bağımsız değişkenler arasında ilişki var olduğunda en küçük kareler tahmincilerinden elde edilen sonuçları düzenler. Burada; A(nxn), boyutlu bir kare matris olmak üzere f(x) = AX = αX eşitliğini sağlayan X vektörüne özvektör, α değerine ise özdeğer denir.

Özdeğer regresyonunun temel bileşenler regresyonundan başlıca iki farkı vardır;

- Bu analizde özdeğerler ve özvektörler farklı matrislerden elde edilir.
- Çoklu doğrusal bağıntıyı gösteren özvektörlerin çıkarılmasına karar vermek için kullanılan teknikler farklıdır.

Y vektörü;

$$Y_i^* = \frac{Y_i - \bar{Y}}{\sqrt{\sum (Y_i - \bar{Y})^2}} \qquad [30]$$

olarak standartlaştırılıp W bağımsız değişkenler matrisi, Y* ile genişletilerek,

Z = (Y* ; W)

matrisinin özvektörlerine dayalı bir yöntem geliştirilmiştir. $I_0 \geq I_1 \geq \ldots \geq I_k$, (Z'Z)'nin özdeğerleri, a_j, (Z'Z)'nin j. özvektörü olmak üzere A özvektörler matrisi olarak gösterilmektedir.

(Z'Z)'nin j. özdeğeri

$$I_j = \sum_{i=1}^{n} (Y_i^* a_{oj} + \sum_{r=1}^{k} W_{ir} a_{rj})^2 \qquad [31]$$

olarak elde edilir.

1) $I_j \approx 0$ ise, Z'nin sütunları doğrusal bağımlıdır, yani değişkenler arasında kuvvetli bir çoklu doğrusal bağıntı vardır. $a_{0j} \neq 0$ ise çoklu doğrusal bağıntı olan değişkenlerin önkestirim yapabilme özelliği vardır.

2) $I_i \approx 0$ ve $a_{0j} \approx 0$ ise, yukarıdaki eşitlikte yer alan

$$\sum_{r=1}^{k} W_{ir} a_{rj}$$

değeri sıfıra eşit olur. I_j'nin ve a_{0j}'nin sıfıra eşit veya sıfıra yakın bir değer olması, Z'nin sütunları arasında tam bir doğrusal bağıntının varlığını gösterir. Bu durumda β parametre kestirimi yapmak oldukça zordur. j <0,05 içi $I_j \approx 0$ ve $| a_{0j}| < 0{,}10$ için de $a_{0j} \approx 0$ varsayılmıştır.

$a_{0j} = (a_{1j} a_{2j} ... a_{kj})$ yani, a_{oj}; a_j 'nin ilk elemanından başka tüm elemanlarını içeren vektörse, özdeğer regresyonu için en küçük kareler kestiriminin;

$$\hat{\beta}_{ÖR} = \sum_{i=1}^{n} (Y_i - \overline{Y})^2 (\sum_{j=0}^{k} a_{oj}^2 / I_j)^{-1} \sum_{j=0}^{k} a_{oj} I_{j-1} a_{oj} \qquad [32]$$

olduğu gösterilmiştir. Bu kestirim öngörüldüğü gibi tüm özvektörlerin doğrusal bileşimidir. Çoklu doğrusal bağıntının birden çok olduğu durumlar için yukarıdaki eşitlikteki toplam j=1,...,t olarak alınmakta ve k+1-t özdeğere karşılık gelen özvektörler çıkarılmaktadır. Dolayısıyla bu özvektörlerin ilgili oldukları bağımsız değişkenler modelden çıkarılmış olacaklardır (5).

2.2.6.4.Ridge Regresyon Yöntemi

Ridge regresyon yönteminin temel amacı, çoklu doğrusal bağıntı varlığında en küçük varyansla parametre kestirimi olan yanlı bir kestirim yöntemidir. Bu yöntem detaylı olarak bir sonraki bölümde incelenmiştir.

3. Ridge Regresyon

Ridge regresyon yöntemi ilk defa 1970 yılında Technometrics dergisinde yayınladıkları iki makale ile Hoerl ve Kennard tarafından geliştirilmiştir. Hoerl ve Kennard bu dergide yayınlanan ilk makalelerinde "Ridge Regresyon: Ortogonal Olmayan Problemler İçin Sapmalı Tahmin" başlığı altında, tam ranklı genel doğrusal hipotez modeline uyan çoklu regresyonda sapmasız tahminleme probleminin detaylı bir tartışmasını ortaya koymuşlardır (12).

Ridge regresyonun çözüm tekniği, basit en küçük kareler yöntemine benzer bir yöntemdir. Ridge regresyon yönteminde, regresyon katsayı tahminlerini hesaplamadan önce standart formdaki değişkenlerin oluşturdukları (X'X) matrisinin köşegen elemanlarına küçük ve pozitif bir sabitin eklenmesiyle gerçekleştirilir. Buna göre ridge regresyon çözümü,

$$\hat{\beta}_{RR} = (X'X) + kI)^{-1}X'Y \qquad [33]$$

şeklindedir. Açıklayıcı değişkenler korelasyon matrisinin köşegen elemanlarına pozitif sabitlerin eklenmesindeki amaç, matris şartlı sayısının önemli ölçüde küçültülmesidir. k=0 için ridge çözümü en küçük kareler çözümüne eşdeğer olduğundan ridge tahmini, en küçük kareler tahmininin bir doğrusal dönüşümü olarak da ifade edilebilir (18).

3.1.Ridge Kestiricisinin En Küçük Kareler Kestiricisi ile İlişkisi

En küçük kareler kestiricisi,

$$\hat{\beta} = (X'X + kI)^{-1}X'Y \qquad [34]$$

olarak verilmişti. Buradan

$$X'X\hat{\beta} = X'Y \qquad [35]$$

olarak yazılabilir. Ridge kestricisi,

$$\hat{\beta}^{*} = (X'X + kI)^{-1}X'Y \qquad [36]$$

olarak verilmişti. X'Y yerine eşiti yazıldığında buradan

$$\hat{\beta}^* = (X'X + kI)^{-1}X'X\hat{\beta} \quad [37]$$

elde edilir. X'X matrisini tersi kendisi olduğundan,

$$\hat{\beta}^* = (X'X + kI)^{-1}[(X'X)^{-1}]^{-1}\hat{\beta} \quad [38]$$

yazılabilir. Her iki matris tekil olmadıklarından,

$$\hat{\beta}^* = [(X'X)^{-1}(X'X + kI)]^{-1}\hat{\beta} \quad [39]$$

yazılabilir. Buradan da,

$$\hat{\beta}^* = [(X'X)^{-1}X'X + k(X'X)^{-1}]^{-1}\hat{\beta} \quad [40]$$

elde edilir. Gerekli işlemlerden sonra,

$$\hat{\beta}^* = [I + k(X'X)^{-1}]^{-1}\hat{\beta} \quad [41]$$

olur. Z=$[I + k(X'X)^{-1}]^{-1}$ olarak tanımlanırsa

$$\hat{\beta}^* = Z\hat{\beta} \quad [42]$$

olarak yazılır. Bu eşitlik ridge kestiricisinin en küçük kareler kestiricisinin bir dönüşümü olduğunu göstermektedir (19).

3.2. Z ve W matrislerinin özellikleri

Ridge kestricisi $\hat{\beta}^* = (X'X + kI)^{-1}X'Y$ olarak verilmişti. W matrisini,

$$W = (X'X + kI)^{-1} \quad [43]$$

şeklinde tanımladığımızda

$$\hat{\beta}^* = WX'Y \quad [44]$$

olur. Z matrisini,

Z=$[I + k(X'X)^{-1}]^{-1}$ şeklinde tanımlamıştık. Buradan,

W matrisinin özdeğerleri

$$\tau_i = \frac{\lambda_i}{\lambda_i + k} \quad [45]$$

Z matrisi aynı zamanda aşağıdaki eşitliklerle de verilebilir:

$$Z = I - k(X'X + kI)^{-1} = I - kW \quad [46]$$

3.3. Ridge Kestiricisinin Beklenen Değeri

$$\begin{aligned} E(\hat{\beta}^*) &= E[(Z\hat{\beta})] \\ &= E[(I - kW\hat{\beta})] \\ &= E[(I - k(X'X + kI)^{-1})\hat{\beta}] \\ &= E[\hat{\beta} - k(X'X + kI)^{-1}\hat{\beta}] \end{aligned} \quad [47]$$

υ_i ‘ler matrisin özdeğerlerine karşılık gelen özvektörler olmak üzere $\hat{\beta}^*$ ‘ın beklenen değeri,

$$E(\hat{\beta}^*) = \beta - k \sum_{i=1}^{p} (\lambda_i + k)^{-1} \upsilon_i \upsilon_i' \beta \quad [48]$$

olarak yazılır.

3.4. Ridge Kestiricisinin Varyansı

$$V(\hat{\beta}^*) = V[(X'X + kI)^{-1}X'Y]$$

$$= \sigma^2 (X'X + kI)^{-1} X'X (X'X + kI)^{-1} \quad [49]$$

Özdeğerler ve özvektörler cinsinden,

$$= \sigma^2 \sum_{i=1}^{p} (\lambda_i + k)^{-2} \upsilon_i \upsilon_i' \quad [50]$$

olarak yazılır (19).

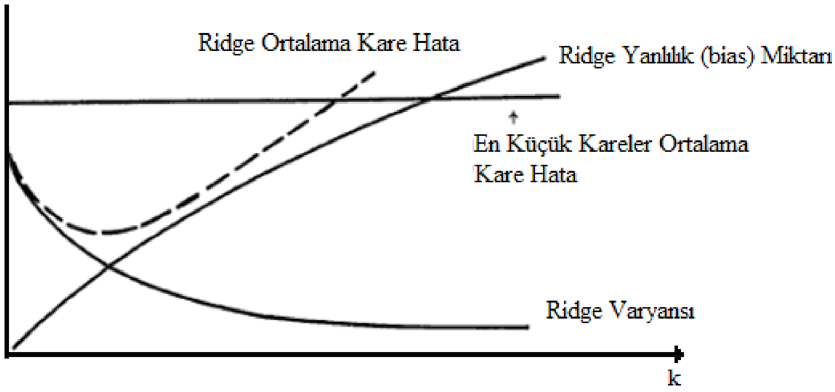
Yanlı olmasına karşın varyansı küçülttüğünden dolayı tercih edilen bir yöntem olan ridge regresyon yönteminin başlıca kullanım amaçları şunlardır:

- Çoklu regresyon modelinde bağımsız değişkenler birbirleriyle bağıntılı olduklarında, en küçük kareler β kestiricisinden daha küçük varyanslı β kestiricilerinin elde edilmesinde,
- Güçlü çoklu doğrusal bağıntı etkisiyle regresyon katsayılarında oluşan kararsızlıkların grafik üzerinde gösterilmesinde,
- Modeldeki gereksiz değişkenlerin çıkarılmasında kullanılmaktadır.

Yukarıdaki amaçlar için kullanılan ridge regresyon yönteminde en küçük kareler yönteminde izlenen aşamalar birden fazla tekrarlanmaktadır. Ridge regresyon yönteminin en küçük karelerden farklılığı, k ridge parametresinin varlığıdır. 0 ile 1 arasında değer alan her k için hesaplanan parametre kestirimleri arasından, aranan kritere sahip olanlar belirlenir. Aşağıda ayrı başlıklar altında ridge kestiricisi β, ridge parametresi k ve ridge yöntemiyle ilgili diğer özellikler açıklanmıştır (5).

3.5.Ridge Varyansı, Sapma ve k Arasındaki İlişki

Varyans, sapma kare ve k arasındaki ilişki aşağıdaki şekilde gösterilmeye çalışılmıştır. Şekil- 1'deki en küçük kareler varyansı hata kareler ortalamasına eşittir ve k'dan etkilenmediği için yatay doğru ile temsil edilmiştir. Şekilde görüleceği üzere, k artarken ridge'in varyansı azalmakta, sapma kare artmaktadır. Şekil üzerinde kesikli çizgi ile temsil edilen ridge tahmin yönteminin hata kareler ortalaması, optimal bir k'da minimum değeri almakta ve sonra tekrar beklenildiği gibi en küçük kareler hata kareler ortalamasını geçmektedir. Böylece, küçük bir sapmaya göz yumularak, tahminin artık kareler ortalamasında meydana gelen az bir artışa karşılık varyansta kayda değer bir azalma sağlayabilecek k değerinin seçilmesi mümkün olabilmektedir.



Şekil- 1: Ridge varyansı, sapma ve k arasındaki ilişki

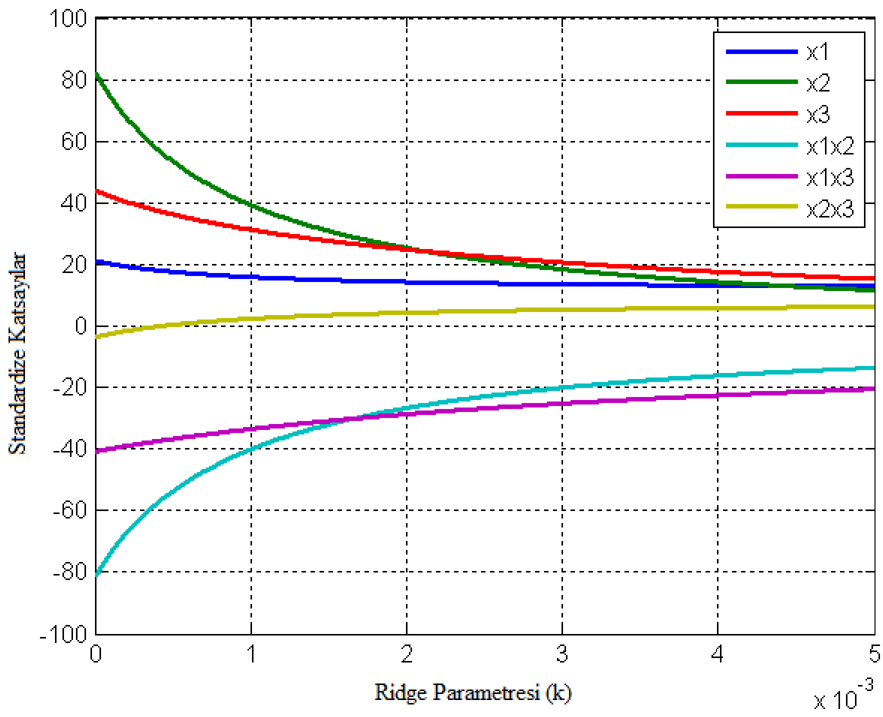
3.6. En Uygun k Parametresini Seçme Yöntemleri

3.6.1.Ridge İzi

Ridge regresyon konusunda yapılan çalışmalarda özellikle k değerini bulmak için onlarca çalışma yapılmış ve yayınlanmıştır. Bu makalelerde özellikle k değerinin 0'a çok yakın bir değer olarak belirlenmesi için çalışılmış ve k ile ilgili bir takım formüller geliştirilmiştir. Bu formüller simülasyon ve Monte-Carlo incelemeleriyle sınanarak en iyi k değeri tespit edilmeye çalışılmıştır. Optimum k'nın seçilmesi konusunda Lee ve Campbell'in 1985 yılında yayınlanan makalelerinde istatistik literatüründeki k elemesi için yapılan bir takım mekanik kurallar tarif edilerek, bu kurallar optimal k'nın seçimi için denenmişlerdir. Bu çalışmalar sonucunda optimal k'nın varlığı ispat edilmiş ve genel olarak optimal k'nın bir cebirsel kapalı forma sahip olamamasından dolayı iteratif sayısal yöntem, optimal k hesaplamalarında kullanılmıştır. Bu stokastik optimal k'nın örnek dağılımının matematiksel özelliklerini sınamak için bir Monte-Carlo incelemesi yürütülmüş ve simülasyon sonuçları verilmiştir (12).

Çoklu doğrusal regresyon problemlerinde çoklu doğrusal bağıntı olduğunda katsayı tahminleri duyarlıdır. Yani veri kümesine birkaç gözlemin ilave edilmesiyle bu tahmin edicilerde değişikliklerin olduğu görülür. Böyle durumlarda $\hat{\beta}$ regresyon katsayıları genellikle kararsız katsayılar olarak bulunur. Bu kararsızlıkları izleyebilmek ve çoklu doğrusal bağıntının etkisini açıkça görebilmek için grafiksel anlatım olan ridge izinden yararlanılır. Ridge regresyonun grafiksel bir gösterimi olan ridge izi, regresyon katsayıları $\hat{\beta}_R$ 'lar düşey eksende, k değerleri yatay eksende olacak şekilde iki boyutlu bir uzayda grafik elde edilir ($0 \leq k \leq 1$). Ridge izi araştırmacıya hangi katsayıların verilere duyarlı olduğu konusunda yardımcı olur.

Ridge izi k'ya karşılık gelen her bir katsayı değerinin grafiğidir. Her bir katsayı için bir eğri ya da iz oluşur. Ridge izinde amaç, k değerinin en küçük karelerden daha küçük hata kareler ortalaması veren k değerini bulmak ve kararlı katsayılar kümesini oluşturmaktır. Doğal olarak k arttıkça, hata kareler ortalaması da artacaktır. Bu çok fazla dikkate alınması gereken bir problem değildir, çünkü önemli olan en yakın tahmin değerini bulmak değil, gelecek gözlemlerin tahmininde işe yarayacak sabit katsayı dizileri geliştirmektir. Burada sabit ifadesinde kastedilen, tahmini verilerdeki küçük değişmeler katsayıların pek duyarlı olmamasıdır. Eğer açıklayıcı değişkenler arasında güçlü bir ilişki varsa, katsayılar k'nın küçük değerleri için hızla değişecek, büyük değerler için ise yavaş yavaş dengelenecektir. Katsayıların sabitleştiği k değeri istenen katsayıları verir. Eğer açıklayıcı değişkenler ortogonal ise, katsayılar çok az değişir ve bu da en küçük kareler sonuçlarının iyi bir katsayı kümesi olduğunu gösterir.



Şekil- 2: Ridge İzi

Ridge izinin kullanımını kolaylaştıran özellikler şunlardır;

- k’nın belirli bir değerinde sistem durağanlaştıracak ve ortogonal sistemin, genel karakteristiklerine sahip olacaktır.
- Katsayılar, değişme oranlarını temsil ettikleri faktörlere göre teorik olarak uygun olmayan mutlak değerlere sahip olmayacaklardır.
- k=0 ‘da görünüşte hatalı işaretlere sahip olan katsayılar, k arttıkça uygun işarete sahip olacak şekilde değişeceklerdir (13).

Ridge izi yöntemiyle seçilen k, teknik bir anlatımla, rastgele bir değişkendir. Regresyon tahminleri seçiminde bu o kadar çok önemli olmasa bile, sapma tanımına bağlı olarak, hipotez testleri ve aralık sınırları teorilerini zorlaştırır. Sapma sebebiyle ortalama karesel hataları bilinmeyen gerçek katsayı vektörü β ile bağımlıdır.

3.6.2.Genelleştirilmiş Ridge Tahmini

Hoerl ve Kennard (1970), optimum bir k değeri kullanarak ridge izi için alternatif bir yöntem geliştirmişlerdir. Genelleştirilmiş ridge tahmin yöntemi olarak adlandırılan bu yöntemde katsayılar aşağıdaki gibi tahmin edilir.

$$\hat{\alpha}_r = [(X^*)'(X^*) + K]^{-1}(X^*)Y \qquad [51]$$

Burada k, köşegen elemanları $k_j \geq 0$ (j=1,2,...,p) olan pxp boyutlu köşegen matrisidir. Tüm k_j değerleri birbirine eşit iken eşitlik 33’de verilen ridge tahmin edici elde edilir. α_i kanonik formda regresyon katsayılarının tahmini olmak üzere $k_i = \sigma^2 / \alpha_i^2$ (i=1,2,...,p) genelleştirilmiş tahmin edicinin hata kareler ortalaması değerini mininum yapar. Hoerl ve Kennard (1970), k_i değerlerinin ardışık seçim yöntemini vermiştir. k_i ‘nin ilk tahmini için en küçük kareler ile başlanır.

$$k_i^0 = \frac{\hat{\sigma}^2}{\hat{\alpha}_i^2} \quad , \ i=1,2,...,p \qquad [52]$$

k_i^0 kullanılarak K=diag(k_1^0,...k_p^0) olmak üzere başlangıç genelleştirilmiş ridge tahmin edici ($\hat{\alpha}_i^0$) hesaplanır ve bunun yardımıyla bir sonraki k_i değeri bulunur.

$$k_i^1 = \frac{\hat{\sigma}^2}{(\hat{\alpha}_i^0)^2} \quad , \ i=1,2,...,p \qquad [53]$$

Bu şekilde durağan parametre tahminleri elde edilene kadar iterasyona devam edilir. Durağanlığın bir ölçüsü olarak regresyon katsayılarının tahminlerinin uzunluklarının karesi kullanılır. Eğer i-1 'den i adımına geçerken parametre tahminlerinin boylarının karesinde herhangi bir değişiklik olmuyorsa iterasyon sonlandırılır. Aksi halde devam edilir.

Genelleştirilmiş ridge tahmin edicide farklı k_i değerleri olduğundan k_i 'lerin seçimi için ridge izi kullanılamaz (13).

BULGULAR

4. Ridge Regresyon Analizi Üzerine Bir Uygulama

Çalışmanın uygulama kısmında kullanılan veriler, Dr.Ali Sait Albayrak'ın İstanbul Üniversitesi İstanbul Tıp Fakültesi Hastanesine şişmanlık şikayeti ile başvuran hastalar arasından basit rasgele örnekleme yöntemi ile elde ettiği verilerdir. Araştırmada 4 değişken kullanılmaktadır: Beden ağırlığı (kg), deri alanı (cm^2), uyluk kemiğinin çevresinin uzunluğu (cm) ve belden yukarı ölçülen kasların çevrelerinin uzunluğu (cm). Araştırmada 20 kişiden elde edilen ölçüm sonuçları aşağıdaki Tablo-1'de verilmiştir (10).

Tablo- 1: Basit tesadüfi örnekleme yöntemiyle seçilen hastalardan elde edilen ölçüm sonuçları

Hasta No	Beden Ağrılığı	Deri Alanı	Uyluk Kemiği Çevre Uzunluğu	Ölçülen Kasların Çevre Uzunluğu
1	75,57	1,81	109,47	73,91
2	144,78	2,30	126,49	71,63
3	118,74	2,86	131,83	93,98
4	127,64	2,77	137,92	78,99
5	81,91	1,78	107,19	78,49
6	137,79	2,38	136,91	60,20
7	172,09	2,92	148,59	70,10
8	161,29	2,59	132,33	77,72
9	135,26	2,06	126,75	58,93
10	122,56	2,37	135,89	62,99
11	161,29	2,89	143,76	76,20
12	172,72	2,83	144,02	71,88
13	74,29	1,74	118,11	58,42
14	113,03	1,83	112,27	72,64
15	81,28	1,36	108,46	54,10
16	151,76	2,74	138,18	76,45
17	143,51	2,58	140,46	65,28
18	161,29	2,81	148,84	62,48
19	93,98	2,11	122,43	68,83
20	133,99	2,34	129,54	69,85

Elde edilen veriler, en küçük kareler yöntemiyle (EKK) yanlı regresyon tekniklerinden ridge regresyon (RR) yöntemiyle karşılaştırılmış, elde edilen sonuçlara göre uygun yöntem belirlenmeye çalışılmıştır. Aşağıdaki tablolarda sonuçları verilen analizler için NCSS 2007 (Versiyon 07.1.5) paket programı kullanıldı.

4.1. EKK ve RR Regresyon Teknikleriyle Elde Edilen Sonuçlar

EKK ve Çoklu Doğrusal Bağıntı

Tablo- 2: Model Özeti (EKK)

R	R^2	Düzeltilmiş- R^2	Tahminin standart hatası	D-Watson (d)
0,895	0,801	0,764	15,756	2,202

Tablo- 2'ye bakıldığında en küçük kareler yöntemine göre beden ağırlığı ile açıklayıcı değişkenler arasındaki doğrusal ilişkinin %89,5 olduğu ve beden ağırlığında meydana gelen değişimlerin yaklaşık %80,1'i bağımsız değişkenler tarafından açıklandığı anlaşılmaktadır.

Tablo- 3: ANOVA Tablosu (EKK)

Kaynak	Kareler Toplamı	Serbestlik Derecesi	Kareler Ortalaması	F	p
Regresyon	16003,875	3	5334,625	21,488	<0,001
Hata	3972,212	16	248,263		
Toplam	19976,087	19			

Tablo- 3'teki ANOVA tablosu incelendiğinde α=0,05 anlamlılık düzeyinde modelin anlamlı olduğu (F =21,488 ve $p < 0,001$) görülmektedir.

Tablo- 4: Tahmin Edilen Regresyon Katsayıları ve Güven Aralıkları (EKK)

						$\hat{b}$ **için %95 güven aralığı**	
	$\hat{b}$	**SH**	$\hat{\beta}$	**t**	**p**	**Alt**	**Üst**
Sabit	768,420	653,842		1,175	,257	-617,663	2154,503
X1	304,366	212,716	4,383	1,431	,172	-146,571	755,203
X2	-7,404	6,667	-3,036	-1,111	,283	-21,538	6,729
X3	-5,619	4,113	-1,605	-1,366	,191	-14,338	3,101

Modelin yüksek belirlilik (R^2 =%80,1) ve korelasyon katsayısının (R =%89,5) aksine Tablo- 4'deki katsayılar tablosu incelendiğinde hiçbir değişkenin kısmi t testine göre anlamlı olmadığı görülmektedir. Bu durum, çalışmada belirtildiği üzere, çoklu doğrusal bağıntı probleminin bir göstergesidir.

Tablo- 5: En Küçük Kareler ve Çoklu Doğrusal Bağıntının Saptanması

	Korelasyon Matrisi				
	X_1	X_2	X_3	**VIF**	R^2 **ve diğer X'ler**
X_1	1			754,923	$R^2_{1.23} = 0,999$
X_2	,924	1		601,325	$R^2_{2.13} = 0,999$
X_3	,475	,085	1	111,104	$R^2_{3.12} = 0,991$

Tablo- 5 incelendiğinde, Tablo- 4'te elde edilen çoklu doğrusal bağıntı sonucuna ulaşılabildiği görülmektedir. Burada bağımsız değişkenler için hesaplanan korelasyon matrisi, varyans artış faktörü (VIF) ve $R_{1;23}$, $R_{2;13}$, $R_{3;\,12}$ değerleri verilmiştir. Bağımsız değişkenlerden her biri bağımlı değişken olarak alınıp geriye kalan diğer bağımsız değişkenler arasındaki ilişkiler incelendiğinde bu ilişkilerin 0,90'dan büyük olduğu görülmektedir. Korelasyon matrisi incelendiğinde ise X_1 ve X_2 değişkenleri arasında çok yüksek bir ilişki (%92,4) olduğu açıktır. X_1, X_2 ve X_3 bağımsız değişkenleri için hesaplanan VIF'ler sırasıyla yaklaşık olarak 755, 601 ve 111 olarak elde edilmiş ve bu değerler kritik varyans artış faktörü (VIF =10) değerinin oldukça üzerindedir. Varyans artış faktörü değerleri 10'dan büyük olduğundan verilerde çoklu doğrusal bağıntı söz konusudur.

Tablo- 6: Korelasyonların Özdeğerleri

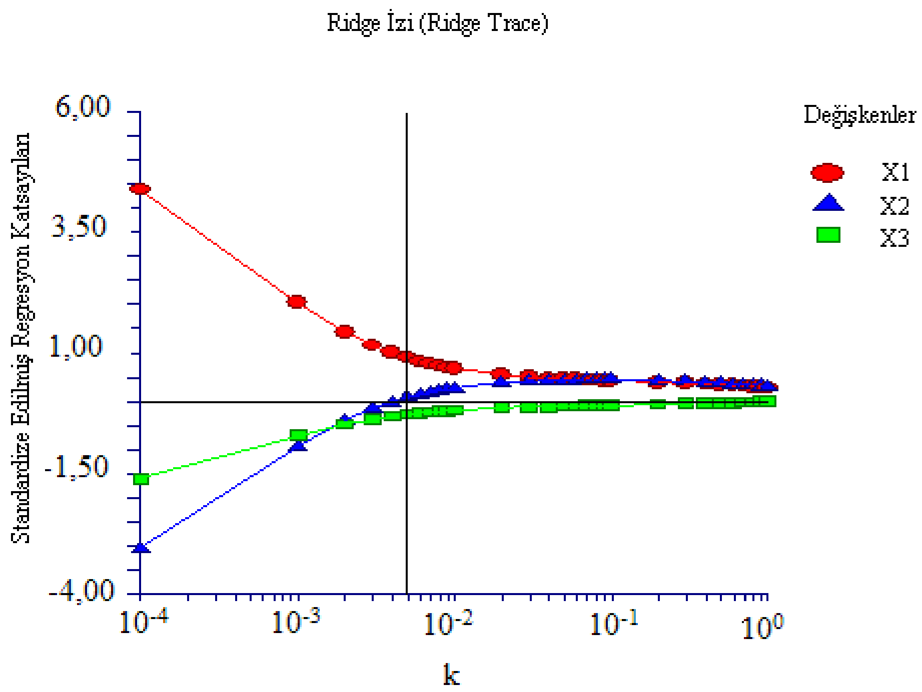
No	Özdeğer	Göreli Yüzde	Birikimli Yüzde	Koşul Sayısı	Koşul İndeksi
1	2,066	68,88	68,88	1,00	1,00
2	0,933	31,10	99,98	2,22	1,490
3	0,001	0,02	100,00	3028,97	55,036

Hesaplanan korelasyonlara ait öz değerlerden yararlanarak elde edilen koşul sayısı (CN = 3028,97) ve koşul indeksi (CI = 55,036) değerleri kritik değerlerden (koşul sayısı için 1000 ve koşul endeksi için 30) büyüktür (Tablo- 6). Elde edilen bu istatistiklerin tamamı verilerde çok güçlü çoklu doğrusal bağlantı problemi olduğunun bir kanıtıdır.

Tablo -7. k, R^2, Standartlaştırılmış Ridge Regresyon Katsayılar ve VIF'ler

k	R^2	Standartlaştırılmış Ridge Regresyon Katsayıları X_1	X_2	X_3	Varyans Artış Faktörü (VIF) X_1	X_2	X_3
0,000	0,801	4,3828	-3,0360	-1,6051	754,9231	601,3251	111,1043
0,001	0,788	2,0243	-0,9320	-0,7038	124,3590	99,2328	19,0402
0,004	0,782	0,8909	0,0773	-0,2702	16,2573	13,1545	3,2518
0,005	0,781	0,8909	0,0773	-0,2702	11,1139	9,0587	2,4991
0,006	0,779	0,8194	0,1406	-0,2408	8,1013	6,6594	2,0574
0,007	0,779	0,7665	0,1874	-0,2224	6,1864	5,1342	1,7760
0,008	0,779	0,7257	0,2232	-0,2067	4,8941	4,1047	1,5855
0,009	0,778	0,6933	0,2516	-0,1942	3,9811	3,3773	1,4503
0,010	0,777	0,6669	0,2747	-0,1840	3,3121	2,8442	1,3509
0,020	**0,772**	**0,5423**	**0,3808**	**-0,1351**	**1,0528**	**1,0410**	**1,0032**
0,030	0,768	0,4976	0,4157	-0,1169	0,6026	0,6783	0,9202
0,040	0,764	0,4739	0,4319	-0,1066	0,4395	0,5445	0,8796
0,050	0,759	0,4588	0,4406	-0,0997	0,3618	0,4788	0,8520
0,080	0,747	0,4330	0,4494	-0,0867	0,2727	0,3969	0,7931
0,090	0,743	0,4274	0,4498	-0,0836	0,2594	0,3828	0,7767
0,100	0,739	0,4225	0,4496	-0,0808	0,2493	0,3715	0,7612
0,200	0,703	0,3910	0,4350	-0,0610	0,2047	0,3072	0,6342
0,300	0,670	0,3700	0,4155	-0,0477	0,1835	0,2682	0,5386
0,400	0,640	0,3527	0,3967	-0,0375	0,1674	0,2381	0,4635
0,500	0,613	0,3376	0,3791	-0,0294	0,1540	0,2135	0,4034
0,800	0,545	0,3000	0,3344	-0,0128	0,1227	0,1603	0,2802
1,000	0,507	0,2797	0,3101	-0,0059	0,1070	0,1357	0,2274

Tablo- 7'de ridge regresyon analizi sonuçlarının birinci kısmında belirli yanlılık katsayılarına göre R^2, standartlaştırılmış ridge regresyon katsayıları ile bu katsayılara ait VIF değerlerindeki değişimler verilmektedir. Bu verilerden yararlanarak optimum k sabiti, standartlaştırılmış ridge regresyon katsayılarının durağanlaşmaya başladığı ve bu katsayılara ait VIF değerlerinin birlikte 1'e yaklaştığı noktada (k=0,02) seçilir. Yanlılık sabiti k ile modelin R^2 değerleri arasındaki ilişki incelendiğinde k, 0 ile 1 aralığında bir değişim gösterirken; R^2, %50,7 ile %80,1 aralığında değişmektedir.



Şekil- 3. Ridge İzi

Yanlı k sabitinin hesaplanmasında ridge izinden yararlanıldığında, k değeri 0,02'den sonra standardize edilmiş katsayıların durağanlaşma eğilimi gösterdiği görülmektedir.

Tablo -8. k=0,02 Yanlılık Sabitiyle Ridge Regresyon Raporu

R	R^2	Adj- R^2	Tahminin standart hatası
0,879	0,772	0,741	16,857

Tablo -9. k=0,02 için ANOVA Tablosu (RR)

Kaynak	SS	SD	Kareler Ort.	F	p
Regresyon	15612,213	3	5204,071	19,081	<0,001
Hata	4363,871	16	272,472		
Toplam	19976,084	19			

Tablo- 9'u incelediğimizde k=0,02 için elde edilen ANOVA tablosunda %5 anlamlılık düzeyinde modelin anlamlı olduğu ($F = 19{,}081$ ve $p < 0{,}001$) görülmektedir.

Tablo- 10. Tahmin Edilen Ridge Regresyon Katsayıları ve VIF değerleri

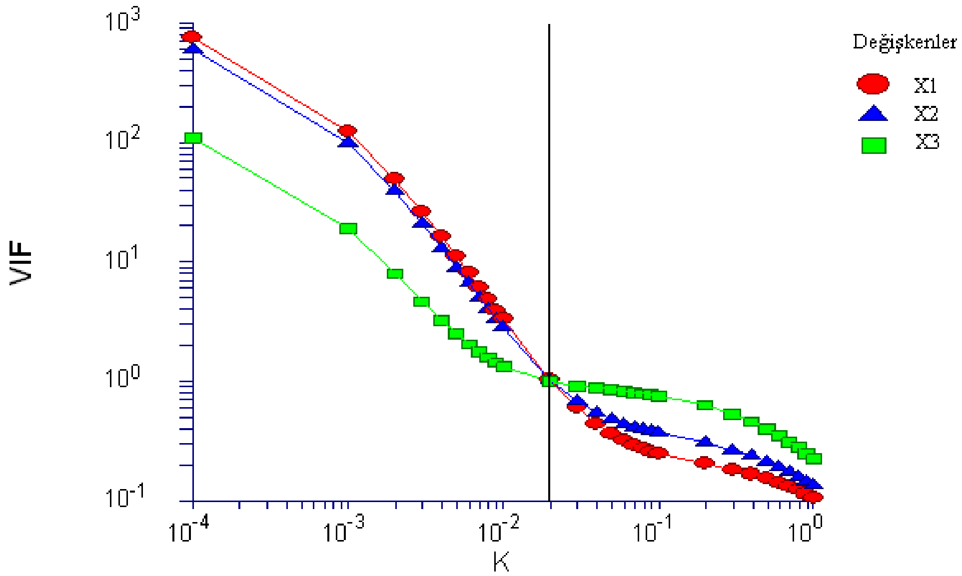
	$\hat{b}$	**SH**	$\hat{\beta}$	**t**	**VIF**
Sabit	-47,910	40,329		-1,188	
X1	37,661	8,499	,542	4,523	1,053
X2	,929	,297	,381	3,193	1,041
X3	-,473	,418	-,135	-1,155	1,003

Tablo- 10'u incelediğimizde k yanlılık parametresi 0,02 alınmış ridge regresyon modelinde, X_1, X_2 ve X_3 bağımsız değişkenleri için hesaplanan varyans artış faktörleri sırasıyla 1.053, 1.041 ve 1.003 olarak elde edilmiştir. Bu değerler kritik varyans artış faktörü (VIF =10) değerinin oldukça altında değerlerdir. Bu durumda yeni modelde çoklu doğrusal bağıntının ortadan kalktığını söyleyebiliriz.

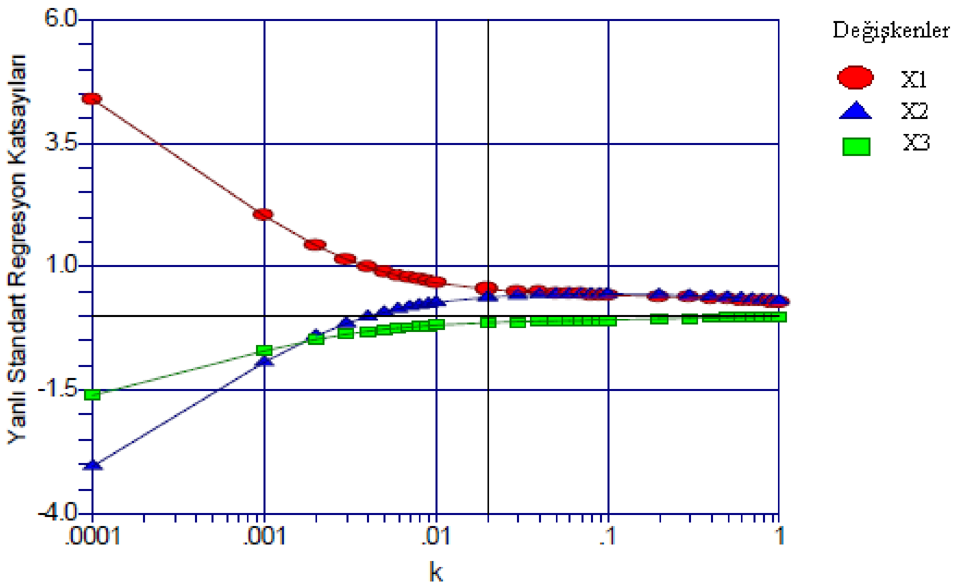
En küçük kareler ve ridge regresyon yöntemleriyle elde edilen sonuçların karşılaştırmaları Tablo- 11'de verilmiştir.

Tablo- 11. k=0,02 İçin Ridge Regresyon ve En Küçük Kareler Karşılaştırması

	RR katsayıları	**EKK katsayıları**	**Standardize RR Katsayıları**	**Standardize EKK katsayıları**	**RR Standart Hata**	**EKK Standart Hata**
Sabit	-47,909	768,42				
X_1	37,661	304,366	0,5423	4,3828	8,498757	212,7157
X_2	0,928	-7,404	0,3808	-3,0360	0,2967767	6,667062
X_3	-0,47	-5,619	-0,1351	-1,6051	0,418141	4,113124
R^2	0,7724	0,8012				



Şekil- 4. Varyans Artış Faktörü (VIF) Grafiği



Şekil- 5. Ridge Grafiği

Logaritmik ölçekli grafikte standartlaştırılmış ridge regresyon katsayılarının nasıl durağanlaştığı çok açık bir biçimde görülmektedir (Şekil- 4). Diğer taraftan, değişkenlerin VIF değerlerinin hangi yanlılık sabiti ($k \cong 0,02$) ile 1'e birlikte yaklaştıkları Şekil- 5'te görülmektedir.

Yukarıdaki bölümlerde ridge regresyon ve en küçük kareler yöntemiyle elde edilen sonuçlara yer verilmiştir. Aşağıdaki tabloda yine NCSS paket programıyla aynı verilere ana bileşenler regresyon yöntemi de uygulanmış ve bu üç yöntem tek bir grafikte gösterilmiştir.

Tablo- 12. En küçük kareler (EKK) ve Ridge Regresyon (RR) Teknikleriyle Elde Edilen Sonuçların Karşılaştırılması

	$\hat{b}_i$		$\hat{\beta}_i$		**Standart Hata**		**VIF**	
	EKK	RR	EKK	RR	EKK	RR	EKK	RR
Sabit	768,42	-47,91						
X_1	304,37	37,66	4,38	0,54	212,72	8,50	754,92	1,05
X_2	-7,40	0,93	-3,04	0,38	6,67	0,30	601,33	1,04
X_3	-5,62	-0,47	-1,61	-0,14	4,11	0,42	111,10	1,00

$$EKK = \; y_i' = 768{,}42 + 304{,}37X_1 - 7{,}40X_2 - 5{,}62X_3$$

$$RR = y_i' = \; -47{,}91 + 37{,}661X_1 + 0{,}93X_2 - 0{,}47X_3$$

Çalışmanın geneline baktığımızda en küçük kareler tekniğinde olduğu gibi ridge regresyon sonuçlarının da belirlenen anlamlılık düzeylerinde anlamlı olduğu görülmektedir. Diğer bir anlatımla beden ağırlığı (Y), açıklayıcı değişkenler tarafından anlamlı bir şekilde açıklanmaktadır.

Ridge regresyon yöntemi tercih edildiğinde, en küçük kareler yöntemine göre tahmincilerin standart hataları da küçüldiği Tablo 12'den açıkça anlaşılaktadır. En küçük kareler tekniği ile yanlı tahmin tekniği ridge regresyon sonuçları arasındaki en önemli çelişki ise, X_2 (Uyluk Kemiği Çevre Uzunluğu) değişkenine ait regresyon katsayısının işaretinde görülmektedir. En küçük kareler tekniğinde kuramsal beklentilerin aksine beden ağırlığı (Y) ile uyluk kemiğinin çevresinin uzunluğu (X_2) değişkeni arasında negatif yönlü bir ilişki söz konusu iken, yanlı tahmin teknikleriyle elde edilen sonuçlarda kuramsal beklentilere uygun bir şekilde pozitif yönlü bir ilişki elde edilmiştir.

TARTIŞMA VE SONUÇ

Genellikle sebep sonuç ilişkisinin belirlenmesi amacıyla kullanılan çoklu regresyon modelinde bağımsız değişkenler matrisindeki en az bir vektör diğer bir vektörle doğrusal olarak bağımlı ise ilişki matrisinin tersi olmayacaktır. İki veya daha fazla bağımsız değişkenin kendi aralarında sıkı bir şekilde bağımlı oldukları durumlarda, istatistikte de çok önemli bir problem olan çoklu bağıntı konusu ortaya çıkmaktadır.

Açıklayıcı değişkenler arasında çoklu doğrusal bağıntı olması durumunda yanlı regresyon teknikleri ile, en küçük kareler tekniğine göre daha durağan ve kuramsal beklentilere uygun sonuçlar elde edilmektedir. Her ne kadar en küçük kareler ve yanlı tahmin tekniklerinden herhangi birisinin seçimi, yanlı veya yansız tahmincilerden birisinin seçimi anlamına gelse de, gerçekte durum böyle değildir. Bilindiği gibi, pratik anlamda, en küçük kareler tahmin edicileri sadece modelin hatasız tanımlanması durumunda yansızdırlar. Bu nedenle pratikte en küçük kareler tahmin edicilerinin genelde yanlı olacağı kabul edilmektedir. Kısaca yanlı tahmin teknikleriyle, çoklu doğrusal bağlantı sorununu azaltmak amacıyla, birbirleriyle anlamlı ilişki içinde olan açıklayıcı değişkenler birlikte analiz edilebilmektedir.

Çalışmanın son bölümünde yapılan uygulamanın sonuçları incelendiğinde, çalışmanın amacına uygun sonuçlar elde edildiği görülecektir. Buna göre, en küçük kareler tekniğinde olduğu gibi ridge regresyon sonuçlarının da belirlenen anlamlılık düzeylerinde anlamlı bulunmuştur. Tablo- 12'ye bakıldığında ridge regresyon yöntemi ile elde edilen varyans artış faktörlerinin, en küçük kareler yönteminde elde edilen değerlerden çok düşük olduğu görülecektir. En küçük kareler yönteminde 3 değişken için sırasıyla 754.92, 601.33 ve 111.10 olarak elde edilirken, ridge regresyon analizi sonucunda varyans artış faktörleri sırasıyla 1.05, 1.04 ve 1 olarak bulunmuş ve bu değerler kritik (VIF=10) değerin altında olduğundan ilk durumda söz konusu olan çoklu doğrusal bağıntının ortadan kalktığı görülmüştür. Yine Tablo- 12'yi incelediğimizde ridge regresyon yöntemiyle, en küçük kareler yöntemindeki standart hatanın oldukça azaldığı görülür. En küçük kareler tekniğiyle ridge regresyon sonuçları arasındaki en önemli çelişki, X_2 (uyluk kemiği çevre uzunluğu) değişkenine ait regresyon katsayısının işaretinde ortaya çıkmıştır. En küçük kareler yönteminde X_2 değişkeninin işareti negatif elde edilmiştir. Bunun anlamı, hastaların uyluk kemiği çevre uzunluğu arttıkça vücut ağırlıklarının azaldığı şeklindedir. Bu da kuramsal beklentilerin tam tersi bir sonuçtur. Alternatif yöntem olarak ridge regresyon uygulandığında ise X_2 değişkenin işareti pozitif elde edilir ve bu da kuramsal beklentilerle örtüşmektedir. Yani kişinin uyluk kemiğinin çevre uzunluğundaki artış, vücut ağırlığındaki artış ile paralellik göstermektedir.

Sonuç olarak, çoklu regresyon analizinde eğer çoklu doğrusal bağıntı söz konusu ise, en küçük kareler yöntemiyle parametre tahmininde bulunmak yanlış sonuçlar almamıza ve yorumlamamıza neden olabilir. Bu yanlışlığın önüne geçebilmek için çoklu doğrusal bağıntı varlığında, en küçük kareler yöntemi yerine yanlı regresyon yöntemleri tercih edilmelidir.

KAYNAKLAR

1. OZDAMAR K. Spss ile biyoistatistik, 5.Baskı, Kaan Kitabevi, Eskişehir, 2003.

2. PAULSON DS. Handbook of regression and modeling, Taylor & Francis Group, page 153-155, New York, 2007.

3. AKTAŞ C. Çoklu bağıntı ve liu kestiricisiyle enflasyon modeli için bir uygulama. Zonguldak Karaelmas Üniversitesi Sosyal Bilimler Dergisi, 3: 67-69, 2007.

4. VAGO E, KEMENY S. Logistic ridge regression for clinical data analysis. Applied Ecology and Environmental Research, 4(2): 171-179, 2006.

5. İMİR E. Çoklu bağıntılı doğrusal modellerde ridge regresyon yöntemiyle parametre kestirimi, Anadolu Üniversitesi Basımevi, Eskişehir, 1986.

6. AYHAN S. Sıralı lojistik regresyon analiziyle Türkiye'deki hemşirelerin iş bırakma niyetini etkileyen faktörlerin belirlenmesi. Osmangazi Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Eskişehir, 2006.

7. SALKIND NJ, RASMUSSEN K. Encyclopedia of measurements and statistics. Volume 1, SAGE Publications, California, 2007.

8. KOUTSOYİANNİS A. Ekonometri kuramı. Çeviren: ŞENESEN Ü, 1. Baskı, Verso Yayıncılık, Ankara, 1989.

9. RAWLINGS OJ, PANTULA SG, DİCKEY DA. Applied regression analysis: A research tool, 2nd Edition, Springer, New York, 1998.

10. ALBAYRAK AS. Çoklu doğrusal bağlantı halinde en küçük kareler tekniğinin alternatifi yanlı tahmin teknikleri ve bir uygulama. Zonguldak Karaelmas Üniversitesi Sosyal Bilimler Dergisi, 1(1), 2005.

11. AKDENİZ F, ÇABUK A. Ridge regresyon teorisinde 1970-2001 arasındaki gelişmeler. V. Ulusal Ekonometri ve İstatistik Sempozyumu Bildiri Kitabı, Adana, 2001.

12. ÖZKALE MR. Çoklu iç ilişki ile ilgili problemler. Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, Adana, 2007.

13. USLU VR. Ridge regresyon ve öğrenci başarısı üzerine bir uygulama. Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Samsun, 1991.

14. ORTABAŞ N. Principal components in the problem of multicollinearity. Dokuz Eylül Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İzmir, 2001.

15. AKTAŞ C. Lojistik regresyon analizinde ridge kestiricisi ve bir uygulama. Eskişehir Osmangazi Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Doktora Tezi, Eskişehir, 1995.

16. BALDEMİR E. Ridge regresyon yöntemi ile parametre tahmini ve bir uygulaması, Dokuz Eylül Üniversitesi, Sosyal Bilimler Enstitüsü, Doktora Tezi, İzmir, 1995.

17. OZDAMAR K. Paket programlar ile istatistiksel veri analizi-2, 5. Baskı, Kaan Kitabevi, Eskişehir, 2004.

18. SAKALLIOĞLU S, KAÇIRANLAR S. A new biased estimator based on ridge estimation. Statistical Papers, 49: 669-689, 2008.

19. KURTULUŞ M. Ridge regresyon üzerine bir çalışma. Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, Ankara, 2001.

TEŞEKKÜR

- Yüksek lisans tezimi hazırlamamda bana yardımcı olan değerli danışmanım, hocam Prof.Dr. İsmet KAN'a,

- Yüksek lisans tezimi hazırlamamdaki desteklerinden dolayı değerli hocalarım, Doç.Dr. İlker ERCAN'a ve Yard.Doç.Dr.Bülent EDİZ'e,

- Katkılarından dolayı değerli meslektaşlarım Arş. Gör. Gökhan OCAKOĞLU'na ve Arş. Gör. Güven ÖZKAYA'ya

- Yardımlarından dolayı Sağlık Bilimleri Enstitüsü personeline,

- Manevi desteklerinden dolayı babam R.Levent BÜYÜKUYSAL'a, annem Sevtap BÜYÜKUYSAL'a ve kardeşim Yağmur BÜYÜKUYSAL'a

sonsuz teşekkür ederim.

ÖZGEÇMİŞ

14 Ağustos 1985 tarihinde Bursa'da doğdum. İlkokulu İnal Ertekin İlkokulunda bitirdim. 1996 yılında Turhan Tayan Anadolu Lisesi'ni kazandım. Ortaokul ve lise öğrenimimi yine bu okulda tamamladım. 2007 yılında Eskişehir Osmangazi Üniversitesi Fen-Edebiyat Fakültesi İstatistik bölümününden mezun oldum.