

T.C.
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ZOOOTEKNİ ANABİLİM DALI

**KİSMİ EN KÜÇÜK KARELER REGRESYONU VE TEMEL BİLEŞENLER
REGRESYON YÖNTEMLERİNİN İNCELENMESİ VE HAYVANCILIKTA
UYGULAMASI**

YÜKSEK LİSANS TEZİ

HAZIRLAYAN: Seda AKYÜREK
DANIŞMAN: Dr. Öğr. Üyesi Suna AKKOL

VAN-2019

T.C.
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ZOOOTEKNİ ANABİLİM DALI

**KİSMİ EN KÜÇÜK KARELER REGRESYONU VE TEMEL BİLEŞENLER
REGRESYON YÖNTEMLERİNİN İNCELENMESİ VE HAYVANCILIKTA
UYGULAMASI**

YÜKSEK LİSANS TEZİ

HAZIRLAYAN: Seda AKYÜREK

Bu çalışma Van YYÜ Bilimsel Araştırma Projeleri Başkanlığı tarafından
FYL-2018-6792 no'lu proje olarak desteklenmiştir.

VAN-2019

KABUL VE ONAY SAYFASI

Zootekni Anabilim Dalı'nda Dr. Öğr. Üyesi Suna AKKOL danışmanlığında, Seda AKYÜREK tarafından sunulan "Kısmi En Küçük Kareler ve Temel Bileşenler Regresyon Yöntemlerinin İncelenmesi ve Hayvancılıkta Uygulaması" isimli bu çalışma Lisansüstü Eğitim ve Öğretim Yönetmeliği'nin ilgili hükümleri gereğince 11.07.2019 tarihinde aşağıdaki jüri tarafından oy birliği ile başarılı bulunmuş ve Yüksek Lisans tezi olarak kabul edilmiştir.

Başkan: Prof. Dr. Sıddık KESKİN

İmza: 

Üye: Dr. Öğr. Üyesi Nazire MİKAIL

İmza: 

Üye: Dr. Öğr. Üyesi. Suna AKKOL (Danışman)

İmza: 

Fen Bilimleri Enstitüsü Yönetim Kurulu'nun 16.108/2019 tarih ve 16.08.2019 sayılı kararı ile onaylanmıştır.



TEZ BİLDİRİMİ

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

İmza

Seda AKYÜREK

ÖZET

KİSMİ EN KÜÇÜK KARELER REGRESYONU VE TEMEL BİLEŞENLER REGRESYON YÖNTEMLERİNİN İNCELENMESİ VE HAYVANCILIKTA UYGULAMASI

AKYÜREK, Seda
Yüksek Lisans Tezi, Zootekni Anabilim Dalı
Tez Danışmanı: Dr. Öğr. Üyesi Suna AKKOL
Temmuz, 2019 71 sayfa

Bu çalışmanın amacı, kıl keçilerinde bazı vücut ölçüleri ile canlı ağırlık arasındaki ilişkinin Kısmi En Küçük Kareler regresyon yöntemi kullanılarak incelenmesi ve sonuçların Temel Bileşenler regresyonu yöntemi ile karşılaştırılarak çalışma verilerini en iyi açıklayan modelin belirlenmesidir. Bu amaçla, Kısmi En Küçük Kareler ve Temel Bileşenler regresyon yöntemlerine ilişkin teorik bilgiler verilerek Kısmi En Küçük Kareler regresyon yönteminde kullanılan algoritmalar açıklanmıştır.

. Çalışma verisinde yer alan açıklayıcı değişkenler arasında yüksek korelasyon olduğu ve bu veri kümesinin çoklu bağlantı problemine sahip olduğu belirlenmiştir. Çoklu bağlantı sorununun üstesinden gelebilmek amacıyla Kısmi En Küçük Kareler regresyon analizi, SIMPLS ve NIPALS algoritmaları kullanılarak yapılmış ve her iki algoritmanın aynı sonuçlar ürettiği ve iki gizil faktör seçtiği ortaya koyulmuştur. Aynı çalışma verisi için Temel Bileşenler regresyon analizi yapılmış ve iki gizil faktör seçilmiştir. Kısmi En Küçük Kareler ve Temel Bileşenler regresyonu uyum kriterleri bakımından karşılaştırılmıştır. Kısmi En Küçük Kareler regresyon yöntemi için hata kareler ortalaması, hata kareler ortalamasının karekökü ve belirleme katsayısı ve sırasıyla 0.2132, 0.4671 ve % 82.0993 ve temel bileşenler regresyon yöntemi için uyum kriterleri 0.2295, 0.4791 ve % 80.0376 olmuştur. Bu sonuç, Kısmi En Küçük Kareler regresyon yönteminin yine iki gizil faktöre sahip Temel Bileşenler regresyon analiz yönteminden daha iyi olduğunu ortaya koymuştur.

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



ABSTRACT

PARTIAL LEAST SQUARES REGRESSION AND PRINCIPAL COMPONENTS REGRESSION INVESTIGATION METHODS AND APPLICATIONS IN ANIMAL HUSBANDRY

AKYÜREK, Seda

M. Sc. Thesis, Department of Animal Science

Supervisor: Assist. Prof.Dr. Suna AKKOL

July 2019, 71 Pages

The aim of this study is to examine the relationship between some body measurements and live weight in hair goats using partial least squares regression method and determine the model that best describes the study data by comparing the results with the basic components regression method. For this purpose, partial Least Squares and basic components are given theoretical information about regression methods and the algorithms used in the partial least squares regression method are explained.

It was determined that there was a high correlation between the explanatory variables contained in the study data and that this dataset had a multi-link problem. In order to overcome the multiple link problem, partial least squares regression analysis was performed using the SIMPLS and NIPALS algorithms and it was revealed that both algorithms produced the same results and chose two hidden factors. For the same study data, basic components regression analysis was performed and two hidden factors were selected. Partial Least Squares and basic components were compared in terms of regression compliance criteria. Partial least squares mean of error squares for regression method, square root and determination coefficient of error squares average and 0.2132, 0.4671 and 82.0993 % respectively, and compliance criteria for basic components regression method were 0.2295, 0.4791 and 80.0376 %. This result revealed that the partial least squares regression method is again better than the basic components regression analysis method, which has two hidden factors.

Keywords: Multicollinearity, Multiple linear regression, Least square, Partial least squares, Principal component.



ÖN SÖZ

Bu tez çalışmasında, her türlü ilgi, bilgi, hoşgörü ve yardımlarını esirgemeyen danışmanım Sayın Dr. Öğr. Üyesi Suna AKKOL'a, çalışmamın her aşamasında bilgi ve tecrübesinden yoğun olarak faydalandığım çalışma arkadaşım Mehmet Nur ÇİFTSÜREN'e ve ayrıca eğitim hayatım boyunca maddi ve manevi desteklerini esirgemeyen sevgi dolu aileme teşekkür ederim.

2019

Seda AKYÜREK



İÇİNDEKİLER

| | Sayfa |
|---|-------|
| ÖZET | i |
| ABSTRACT | iii |
| ÖN SÖZ..... | v |
| İÇİNDEKİLER..... | vii |
| ÇİZELGELER LİSTESİ | ix |
| ŞEKİLLER LİSTESİ..... | xi |
| SİMGELER VE KISALTMALAR | xiii |
| 1. GİRİŞ..... | 1 |
| 2. KAYNAK BİLDİRİŞLERİ | 5 |
| 3. MATERYAL VE YÖNTEM..... | 11 |
| 3.1. Materyal | 11 |
| 3.2. Yöntem..... | 11 |
| 3.2.1. Çoklu doğrusal regresyon analizi | 11 |
| 3.2.2. Çoklu doğrusal regresyonda en küçük kareler yöntemi | 14 |
| 3.2.2.1. Çoklu doğrusal bağlantı problemi | 17 |
| 3.2.2.2. Çoklu bağlantı sorununu belirlenmesi | 19 |
| 3.2.2.3. Çoklu doğrusal bağlantı probleminin giderilmesi..... | 21 |
| 3.2.3. Temel bileşenler analizi | 22 |
| 3.2.3.1. Temel Bileşen Sayısının Verilmesi | 27 |
| 3.2.4. Kısmi en küçük kareler regresyonu | 29 |
| 3.2.4.1. Kısmi en küçük kareler regresyonu..... | 32 |
| 3.2.4.2. Doğrusal olmayan yinelemeli kısmi en küçük kareler algoritması (NIPALS) | 34 |
| 3.2.4.3. SIMPLS algoritması..... | 40 |
| 3.2.4.4. Standart KEKK algoritması ile SIMPLS algoritmasının karşılaştırılması..... | 43 |
| 4. BULGULAR | 45 |
| 5. TARTIŞMA VE SONUÇ..... | 62 |
| KAYNAKLAR..... | 67 |
| ÖZ GEÇMİŞ..... | 73 |



ÇİZELGELER LİSTESİ

| Çizelge | Sayfa |
|---|--------------|
| Çizelge 3.1. Çoklu doğrusal regresyon analiz tekniğinde verilerin dizaynı | 12 |
| Çizelge 4.1. Çalışmada kullanılan değişkenlere ilişkin tanımlayıcı istatistikler | 45 |
| Çizelge 4.2. Bağımlı ve bağımsız değişkenler arasındaki korelasyonlar ve önem düzeyleri | 46 |
| Çizelge 4.3. Kıl keçilerinde CA için EKK yöntemi kullanılarak elde edilen çoklu doğrusal regresyon analiz sonuçları | 46 |
| Çizelge 4.4. CA için çoklu bağlantı belirleyicilerinden Özdeğer (OD) ve CI değerleri..... | 47 |
| Çizelge 4.5. CA için çoklu bağlantı belirleyicilerinden varyans oranları | 47 |
| Çizelge 4.6. SIMPLS ve NIPALS algoritmaları için gizil faktör sayıları ve model uyum kriterleri..... | 49 |
| Çizelge 4.7. SIMPLS ve NIPALS algoritması kullanılarak CA için elde edilen regresyon katsayıları | 51 |
| Çizelge 4.8. TB regresyon analiz için gizil faktör sayıları ve model uyum kriterleri. ... | 55 |
| Çizelge 4.9. TB regresyon analizi kullanılarak CA için elde edilen regresyon katsayıları | 57 |

ŞEKİLLER LİSTESİ

| Şekil | Sayfa |
|--|-------|
| Şekil 4.1. Çarpaz doğrulama yapılarak elde edilen SIMPLS ve NIPALS algoritmasının gizil faktör seçiminin görseli. | 48 |
| Şekil 4.2. SIMPLS ve NIPALS algoritmaları için korelasyon grafiği | 50 |
| Şekil 4.3. SIMPLS ve NIPALS algoritmaları kullanılarak bağımsız değişkenler için elde edilen VIP istatistikleri..... | 53 |
| Şekil 4.4. TB regresyon analizinde çarpaz doğrulama yapılarak elde edilen gizil faktör seçiminin görseli..... | 54 |
| Şekil 4.5. TB regresyon analizi ile elde edilen korelasyon grafiği..... | 56 |
| Şekil 4.6. TB regresyon analizi kullanılarak elde edilen standardize edilmiş regresyon katsayılarının profili | 57 |



SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış bazı simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

| Simgeler | Açıklama |
|-----------------|--------------------------------|
| β | Regresyon Katsayısı |
| $\hat{\beta}$ | Regresyon Katsayısının Tahmini |
| Cov | Kovaryans |
| $E()$ | Beklenen Değer |
| e_i | Hata (Örneklem) |
| ε_i | Hata (Anakütle) |
| R^2 | Belirleme Katsayısı |
| Var | Varyans |
| X | Bağımsız Değişken |
| Y | Bağımlı Değişken |
| λ | Özdeğer |
| ∂ | Türev |
| k | Parametre Sayısı |
| n | Gözlem Sayısı |
| T ve U | Yük Vektörler |

| Kısaltmalar | Açıklama |
|--------------------|---|
| CN | Koşul İndeksi |
| CV | Çapraz Geçerlilik |
| CY | Cidago Yüksekliği |
| EKK | En Küçük Kareler |
| GC | Göğüs Çevresi |
| GD | Göğüs Derinliği |
| HKO | Hata Kareler Ortalaması |
| KEKKR | Kısmi En Küçük Kareler Regresyonu |
| LASSO | Least Absolute Shrinkage and Selection Operator |
| NIPALS | Doğrusal Olmayan Yinelemeli En Küçük Kareler |
| OLS | Klasik En Küçük Kareler |
| KEKKR1 | Cevap değişkeni tek değişkenli kısmi en küçük kareler algoritması |
| KEKKR2 | Cevap değişkeni çok değişkenli kısmi en küçük kareler algoritması |
| RR | Ridge Regresyonu |
| RSS | Hata Kareler Toplamı |
| SAMPLS | Gözlem-uzaklığı Kısmi En Küçük Kareler |
| SAY | Sağrı Yüksekliği |
| SIY | Sırt Yüksekliği |
| SIMPLS | Örnek-uzaklık kısmi en küçük kareler algoritması |
| TB | Temel Bileşen |
| TV | Tolerans Değeri |
| UNIPALS | Evrensel Kısmi En Küçük Kareler |
| VIF | Varyans Arttırıcı Faktör |
| VK | Varyasyon Katsayısı |

1. GİRİŞ

Hayvancılıkta çok sayıda faktöre bağılı olarak deęişim gösteren verim özellikleri ve ekonomik faktörler arasındaki sebep-sonuç ilişkisini ortaya çıkarabilmek için kullanılan istatistik yöntemlerden biri çoklu regresyon analiz yöntemidir. Çoklu doğrusal regresyon analiz yönteminde bir cevap deęişkeni ve birden çok sayıda açıklayıcı deęişken bulunmaktadır. Cevap deęişkeninin gerçek ölçüm deęerleri (Y) ile regresyon denklemi kullanılarak elde edilen tahmin deęerleri (\hat{Y}) arasındaki uzaklığı en küçük yapan regresyon katsayıları, En Küçük Kareler (EKK) yöntemi ile tahmin edilir. Bu yöntemde, örneklemden elde edilen regresyon denklemiyle, deęişkenler arasında var olan sebep-sonuç ilişkilerini belirlemenin yanında, geleceęe ilişkin tahmin bir dięer ifade ile kestirimin de daha güvenli bir şekilde yapılabilmesi hedeflenmektedir.

Çoklu doğrusal regresyon analizinde EKK yönteminin uygulanabilmesi için varsayımların yerine gelmesi gerekir. Bunlar kısaca hata terimlerinin varyansının sabit olması ve birbirinden bağımsız olmaları, hata terimlerinin beklenen deęerinin "0" olması ve açıklayıcı deęişkenler arasında bir ilişki olmaması şeklinde özetlenebilir. Bu varsayımlardan, açıklayıcı deęişkenlerin arasına bağımsızlığın olması varsayımı, bir dięer ifade ile açıklayıcı deęişkenler arasında çok yüksek bir ilişkinin olmaması varsayımı özellikle biyolojik karakterler söz konusu olduğunda yerine getirilmemektedir. Bu varsayımın sağlanamaması, veri setinde çoklu doğrusal bağlantı probleminin olabileceğini işaret etmektedir. Veri setinin çoklu bağlantı problemine sahip olması durumunda, EKK yöntemi kullanılarak elde edilen tahmin deęerleri güvenilirliğini yitirmektedir. Zira çoklu bağlantı sorununun olduğu veri setinde EKK yönteminin kullanılması varyans deęerlerinin büyük olmasına ve tahminlerin gerçek deęerlerinden uzaklaşmasına neden olmaktadır. Bu sorunun üstesinden gelen yöntemlerden biri Kısmi En Küçük Kareler (KEKK) yönteminin kullanılmasıdır (Bulut ve Alın, 2009).

KEKK yöntemine ilişkin ilk olarak bir algoritma olarak sunulmuştur. İlerleyen dönemlerde, istatistik özellikler ile birleşip uygulanabilir bir regresyon modeli haline almıştır. KEKK yöntemi, kimyasal analizlerde kimyasal veriden saklı bilginin açığa çıkarılmasına imkân tanıyan bir yöntemdir. Bu nedenle fiziksel kimya, analitik kimya,

klinik kimya ve endüstriyel süreç kontrolü gibi kimyanın birçok alanında KEKK yöntemi kullanılmaktadır. Bu yöntemi istatistikçiler tarafından uzun süre bilinmesine karşın, daha çok kemometriciler tarafından kullanıldığı söylenebilir. Günümüzde bu yöntem birçok bilimsel ve teknolojik uygulamalarda sıklıkla kullanılan bir yöntem haline gelmiştir. KEKK yönteminde amaç, açıklayıcı değişkenler X ve cevap değişkeni/değişkenlerinden oluşan iki veri seti için elde edilmiş gizil/gizli (latent) değişkenler kullanılarak, bu iki veri seti arasında ilişki bulmaktır (Polat, 2009). Bu yöntem, çoklu doğrusal regresyon modelini, tahmin bakımından yorumlanması kolay olan daha az bileşenli (gizil faktöre sahip) ve daha kullanışlı bir modele dönüştürmektedir. Genellikle bağımsız değişkenler arasında çoklu bağlantı sorunu olduğunda kullanımı tercih edilen KEKK yönteminin en önemli özelliği, daha az bileşenli sade bir model bulmasıdır (Polat, 2009).

Çalışma verisinin çoklu bağlantı problemine sahip olması durumunda kullanılan bir diğer yöntem Temel Bileşenler (TB) regresyon yöntemidir. Hotelling (1933) ve Rao (1964) tarafından uygulama alanı genişletilmiş olan TB regresyon analiz yöntemi sosyal bilimlerden fen bilimlerine kadar çeşitli çalışma alanında yoğun bir şekilde kullanılmaktadır. Bu regresyon yöntemi, regresyon modelinde yer alan bağımsız değişkenler arasındaki bağımlılık yapısını yok etmek ve/veya boyut indirmek şeklinde başlı başına bir yöntem olarak kullanılmaktadır. Bunun yanı sıra araştırmacılar tarafından veri hazırlama tekniği olarak da sıklıkla kullanılan bir yöntemdir.

Keçiler gerek kolay bakım ve beslenmeleri gerekse sütünün yüksek besin değerine sahip olması nedeniyle farklı toplumlarda giderek artan bir önem kazanmaktadır. Günümüzde Amerika ve Avrupa'nın bazı ülkelerinde inek sütü üretimi yeterli olduğu halde, süt keçisi yetiştirmek için özel çiftlikler kurulmaktadır. Elde edilen keçi sütü peynir, yoğurt, tereyağı gibi değişik süt ürünlerine işlenerek yüksek fiyatla satılmaktadır. Söz konusu ürünlerin pahalı olmasının nedenleri arasında; keçi sütünün insan sağlığına olumlu etkilerinin olduğu söylenilebilir (Koyuncu, 2005b).

TUİK verilerine göre Türkiye'de küçükbaş hayvan sayısının geçen yıl bir önceki yıla göre yüzde 4.1 artarak 46 milyon 117 bine ulaşmış olup koyun sayısının yüzde 4.5'lik artışla 35 milyon 195 bine, keçi sayısının da yüzde 2.7'lik artışla 10 milyon 922 bine yükseldiğini bildirmiştir (TUİK, 2018).

Bu çalışmanın amacı, kıl keçilerinde canlı ağırlık tahmin edilirken veri setinin çoklu bağlantı problemine sahip olması durumunda, KEKK regresyon analiz yöntemini kullanarak elde edilen sonuçları, hayvancılık alanında şimdiye kadar çoklu bağlantı probleminin çözümünde kullanılmış olan TB regresyon analizinden elde edilen sonuçlar ile karşılaştırmak ve çalışma verisini en iyi açıklayan modeli belirlemektir. Bu amaçla, çalışmada öncelikle çoklu doğrusal regresyon analizinde EKK yöntemi, TB analizi ve KEKK analizi yöntemlerine ilişkin bilgiler ve algoritmalar verilmiştir. Çalışma verisinde cevap değişkeni canlı ağırlık, açıklayıcı değişkenler ise göğüs genişliği, cidago yüksekliği, sırt yüksekliği, göğüs derinliği, göğüs çevresi, vücut uzunluğu kullanılmıştır. Çoklu doğrusal regresyon analizi yapılırken veri setine öncelikle EKK yöntemi uygulanarak çoklu bağlantı sorunu olup olmadığı incelenmiştir. Ardından bu sorunun üstesinden gelen ve çalışmanın konusunu oluşturan TB ve KEKK regresyon yöntemleri kullanılarak elde edilen sonuçlar karşılaştırılmıştır.



2. KAYNAK BİLDİRİŞLERİ

KEKK regresyonu başlangıçta bir çeşit kalibrasyon yöntemi olarak Wold (1975) tarafından incelenmiştir. KEKK regresyonu, kendisine kemometri literatüründe önemli bir yer bulmuştur. Bu yöntemin istatistik özellikleri Wold (1984), Naes ve Martens (1986), Manne (1987), Helland (1988) ve Stone ve Brook (1990) gibi çok sayıda istatistikçi ve matematikçinin destekleriyle netlik kazanmaya başlamıştır.

Höskuldsson (1988) tarafından yapılan çalışmada, KEKK'in istatistiksel ve matematiksel yapısı araştırılmış ve KEKK'in diğer regresyon yöntemlerine göre üstünlüğü ve maksimum indirgeme ile minimum değişken sayısını seçmesi ile ilgili detaylı bilgiler verilmiştir. Aynı yıllarda Helland (1988) tarafından yapılan çalışmada, KEKK'in yapısı gösterilmiş ve KEKK regresyonunun TBA'dan daha az bileşen kullanarak daha küçük hata kareler ortalamasına sahip olduğu vurgulanmıştır. Bununla birlikte KEKK'in performansı çeşitli araştırmacılar tarafından simülasyon çalışmaları ile incelenmiştir (Naes ve Martens, 1986; Geladi ve Kowalski, 1986a; Frank, 1987; Frank ve Friedman, 1993).

Naes ve Martens (1986) yaptıkları çalışmada, yakın kızılötesi (NIR) cihazların istatistiksel kalibrasyonuna çoklu doğrusal regresyon, temel bileşen regresyon ve kısmi kareler regresyon yaklaşımı karşılaştırılmıştır. Yanlı regresyon yöntemleri olan son üç yöntemin çoklu doğrusal regresyondan çok daha iyi tahmin sonuçları verdiğini gösterilmiştir.

De Jong (1993), KEKK regresyon yöntemini tamamlayan NIPALS VE SIMPLS algoritmalarını kullanmıştır. KEKK regresyon geometrik yorumunda basit yapı taşlarını kullanarak, NIPALS ve SIMPLS algoritmalarını adım adım açıklamıştır.

Ortabaş (2001) yaptığı yüksek lisans tez çalışmasında çoklu doğrusal bağlantı sorununu ortadan kaldırmaya yönelik olarak kullanılan TB ve Ridge regresyon yöntemlerini ele almıştır. Çoklu bağlantı sorununu ortadan kaldırmaya yönelik kullanılan yöntemlerden yanlı regresyon yöntemleri ile hem çoklu doğrusal bağlantı yapısının açıklanabildiğini, hem de standart hatası daha küçük hata kareler ortalamalı tahminlerin bulunabildiğini ifade etmiştir. Çalışmada TB analizi ve Ridge (RR) regresyon yöntemleri incelenmiş olup, hangi regresyon yönteminin daha iyi sonuç

verdiği araştırılmıştır. Çalışmalar neticesinde temel bileşenler regresyon yönteminin regresyon katsayılarının tahmini bakımından daha iyi sonuç verdiği vurgulanmıştır.

Graham (2003) yapmış olduğu çalışmada TB analiz yöntemini ekolojik verilere uygulamıştır. Çalışmanın sonucunda ekolojik verilerin çoklu doğrusal regresyon modelleri ile analizinde çoklu bağlantı sorununu dikkate alan TB regresyon yönteminin kullanılmasıyla modelin geçerliliğinin artırılacağı ifade edilmiştir

Voncina ve ark., (2005) yapmış oldukları çalışmada Kabak, ayçiçeği, yer fıstığı, zeytin, soya fasulyesi, kolza tohumu, Mısır ve bazı karışık yağlar da dahil olmak üzere, farklı yenilebilir bitkisel yağların 132 örneğinin yağ asidi bileşimini kullanmışlardır. Yağ asit içeriklerinin karakterize edilmesi ile bitkisel yağların sınıflandırılmasında ayırma ve temel bileşenler analizinden yararlanmışlardır. Çalışmada, 7 element için 132 yağ örneğini incelemişler ve ilk 2 temel bileşen ile toplam varyansın % 97.8'inin açıklandığını ve ayırma analizi sonucunda 5 ayırım fonksiyonu ile 7 element arasındaki ayırımı % 96.84 doğruluk oranıyla belirlediklerini vurgulamışlardır.

Albayrak (2005) yaptığı çalışmada beden ağırlığının tahmin edilmesinde RR ve TB analizi EKK yöntemine karşı etkin olup olmadığı araştırmıştır. Aralarında çoklu doğrusal bağlantı bulunan bağımsız değişkenlere RR, TB regresyonu ve EKK yöntemleri uygulanmıştır. RR ve TB regresyon yöntemlerinin EKK yöntemine göre daha düşük standart hatalı, durağan ve kurumsal beklentileri karşılayabilecek tahminler elde edilebileceğini ortaya koymuştur.

Pierce ve ark., (2006) yapmış oldukları araştırmada, inceledikleri bitki örneklerine ait kimyasal farklılıkları ortaya koymak amacıyla TB analizinden yararlanmışlardır. Bu amaçla fesleğen, nane ve tatlı ot olmak üzere üç farklı bitki kullanılmıştır. Çalışma sonucunda birinci temel bileşen, tatlı ot bitkisini fesleğen ve nane bitkilerinden ayırarak toplam varyansın % 61.84'ünü açıklamıştır. İkinci temel bileşen fesleğen bitkisini nane kümesinden ayırarak toplam varyansın % 16.78'ini açıklamıştır.

Macciotta ve ark., (2006) süt keçilerinde süt, yağ ve protein verimini tahmin etmek için KEKK regresyon yöntemini kullanılmışlardır. Sarda ırkı süt keçilerinin 1731 laktasyonu içeren veri kümesi, biri model tahmini ve diğeri KEKK regresyonunun tahmin yeteneğinin değerlendirilmesi için olmak üzere iki veri kümesine bölünmüştür. Tahmin edilen ve gözlemlenen değişkenlerin verimleri arasındaki korelasyonlar oldukça

yüksek bulmuştur. Sonuçlar, söz konusu bu çok değişkenli tekniğin büyük bir tahmin doğruluğuna sahip olduğunu ortaya koymuştur.

Shanmugan ve Johnson (2007) melanom kanserli hastalarda sağkalım oranları üzerine yaptıkları çalışmalarında Veri Zarflama Analizi ve TB analizini ayrı ayrı uygulanmışlardır. Her iki analiz yöntemin avantajlı yönleri incelenmiş ve temel bileşenler analizi sonunda birinci bileşenin toplam varyasyonun % 95.53'ünü açıkladığı tespit edilmiştir.

Topal ve ark., (2010) farklı yaşlarda 91 adet sazan balığında çeşitli vücut ölçülerini kullanarak karkas ağırlıklarını tahmin etmeye çalışmışlardır. Vücut ölçüleri arasında ortaya çıkan çoklu bağlantı sorunu nedeniyle çoklu doğrusal regresyon analizinde EKK yöntemi, RR ve TB analiz yöntemlerini kullanmış ve aynı veri seti için bu üç yöntemin sonuçlarını karşılaştırmışlardır. Karşılaştırma kriteri olarak belirleme katsayısı (R^2), hata kareler ortalaması (HKO) ve standart hata (S) ile birlikte modellerin varyasyon katsayısı (VK) kullanılmıştır. Elde edilen sonuçlara göre yanlı tahmin yöntemlerinden olan temel bileşenler ve Ridge regresyon yöntemlerinin EKK ya göre daha güvenilir sonuçlar verdiği ifade edilmiştir.

Mendeş (2011) araştırmasında açıklayıcı değişkenler arasında çoklu bağlantı sorununu gidermek için TB analizini yöntemini kullanmıştır. Regresyon analizi sonucunda, incik genişliği, göğüs çevresi ve canlı ağırlığın kesim sonrası özellikleri tahmin etmedeki doğrusal etkilerinin benzer olduğu bulunmuştur ($P=0.746$). Seçilen ilk 3 temel bileşen toplam varyasyonun % 89.45'lik kısmını açıklamıştır. Sonuç olarak, kesim öncesinde incik genişliği, göğüs çevresi ve canlı ağırlığı fazla olan hayvanların, kesim sonrası özelliklerinin de fazla olmasının beklendiği vurgulanmıştır.

Eker (2011) yaptığı yüksek lisans tez çalışmasında, çoklu regresyon analizinde açıklayıcı değişkenler arasında çoklu bağlantı sorununun var olması durumunda, sıkça kullanılan EKK yönteminin varsayımlarının geçerliliğini kaybetmesi durumunda, EKK yöntemi ile Ridge ve TB regresyon yöntemlerini karşılaştırmalı olarak incelemiştir. EKK, Ridge ve TB regresyon yöntemlerini değerlendirmede hata kareler ortalaması ve belirleme katsayısı (R^2) kullanmıştır. Araştırmanın sonucunda, en düşük hata kareler ortalaması ve en büyük belirleme katsayısının RR modeli ile elde edilmiş olmasına rağmen, parametrelerin önem testlerine göre TB analiz yönteminin daha tutarlı olduğunu ifade edilmiştir.

Korkmazođlu ve Kemalbey (2012) yaptıkları alıřmada Trkiye de endojen byme modeli ekonomi alanında uygulamasını KEKK analizi zerinde alıřmıřlardır. Zaman serisi verilerini ieren birok ekonometrik model oklu dođrusallık problemine sahip olduđundan KEKK regresyonu ile zaman serilerinde daha gvenilir sonular verdiđini sylemiřlerdir. Bu analiz iin iki farklı algoritma olan ekirdek ve NIPALS algoritmaları seilmiřtir. KEKK analizi uygulandıđında, aıklayıcı deđiřkenler, sayıca az olan bileřenlere indirgenmiřtir ve oklu dođrusallık problemi sorunu ortadan kaldırılmıřtır.

Oliveira ve ark., (2013) tarafından suni tohumlama alıřmalarında gebe kalma oranının tahmininde nemli olan sperm deđiřkeninin zellikleri incelenmiřtir. Bu amala alıřmalarında KEKK regresyon yntemini kullanmıřlar ve KEKK regresyonunun gvenilir sonular verdiđini bildirmiřlerdir.

Ferrand-Calmels ve ark., (2014) yaptıkları alıřmada inek, koyun ve kei stndeki yađ asidi bileřimini tahmin etmek iin orta kızıltesi spektrometresi kullanmıřlardır. Bu alıřmada toplamda 349 inek st, 200 koyun st 332 kei st rneđi kullanılmıř ve alıřmada en gvenilir sonuların KEKK regresyon tekniđi ile elde edildiđi grlmřtr.

Godoy ve ark., (2014) TB ve KEKK regresyonu yntemi arasındaki iliřkiyi teorik olarak aıklamıřlardır. Sayısal rnekler, gizli modeller, ıktı tahmin modelleri ve hata algılama endeksleri arasındaki iliřkileri gstermek iin kullanılmıř ve KEKK, TB analizi yntemleri karřılařtırılmıřtır. eřitli zelliklerinin karřılařtırıldıđı bu iki yntem iin ulařtıkları sonulardan biri, KEKK regresyon analizinin tahmini daha gvenilir sonular vermiř olduđudur.

Shaibu ve ark., (2015) alıřmalarında, fenolojik zellikleri kullanarak mısır verimini tahmin etmek iin KEKK regresyon yntemini kullanmıřlardır. Veriler XLSTAT programında analiz edilmiřtir. alıřmada mısır verimini tahmin etmede KEKK ynteminin iyi sonular verdiđi ortaya koyulmuřtur.

Toklu ve Toklu (2015) yaptıkları alıřmada, zeytinyađının isel ve dıřsal iřaretlerine tketicilerin tepkisi zerine bir arařtırma yapmıřlardır. Bu alıřma, zeytinyađı ile iliřkili tanımlanan on iki farklı iřaretin rne ynelik tutuma, rn imajına ve tketicilerin daha fazla deme isteđine etkisini arařtırmıřlardır. Arařtırma modelinin verileri rnekleme yntemiyle seilen zeytinyađı tketicileriyle yapılan

anketle elde edilmiştir. Veriler, KEKK ve yapısal eşitlik modellemesi kullanılarak analiz edilmiştir. Yapısal eşitlik modelinin genel açıklayıcı gücü, bağımsız değişkenler tarafından açıklanan varyansın miktarı ve hipotezlerin her birinin belirli bir yapısal model yolundaki büyüklüğü ve gücü incelenmiştir. Modelin açıklayıcı gücünü ölçmek için belirleme katsayısı ve modeldeki ilişkilerin istatistiki anlamlılığını test etmek için T-İstatistiği kullanılmıştır. Yapıdaki her bir maddenin faktör yüklerinin yüksek bir şekilde anlamlı ($P < 0.0001$) olduğu görülmüştür.

Firinguetti ve ark., (2016) yapmış oldukları çalışmada, çoklu doğrusal bağlantı problemlerine yönelik kullanılan KEKK ve Ridge regresyon yöntemlerini ele almışlardır. Performanslarını karşılaştırmak üzere yapılan simülasyon çalışması sonucunda, hatanın yüksek olduğu durumlarda RR'nin daha iyi performans sergilediği ve modelin daha fazla değişken içerdiği durumlarda KEKK yöntemi ile en iyi sonuçlara ulaştığını ifade etmişlerdir.



3. MATERYAL VE YÖNTEM

3.1. Materyal

Çalışmanın materyalini 2013-2014 yılları arasında Denizli ilinde yürütülmüş olan Kıl keçisi ıslahı projesi kapsamında elde edilen verilerin bir kısmı oluşturmaktadır. Çalışmada kullanılmak üzere düzenlenen veri setinde tamamı dört yaş olmak üzere toplamda 119 baş dişi kıl keçisi bulunmaktadır. Kıl keçilerine ait canlı ağırlık (CA) kg cinsinden, cidago yüksekliği (CY), sağrı yüksekliği (SAY), sırt yüksekliği (SY), vücut uzunluğu (VU), göğüs derinliği (GD), göğüs genişliği (GG), göğüs çevresi (GC) cm cinsinden ölçümü yapılan değişkenlerdir. Çalışmada CA bağımlı (cevap) değişken olarak CY, SAY, SY, VU, GD, GG ve GC bağımsız (açıklayıcı) değişkenler olarak yer almıştır.

Çalışmanın istatistiki analizleri yapılırken SAS ve Stata paket programlarından faydalanılmıştır.

3.2. Yöntem

3.2.1. Çoklu doğrusal regresyon analizi

Regresyon analizi, cevap değişkenindeki değişimi açıklayan açıklayıcı (bağımsız) değişkenlerin belirlenmesi, buradan hareketle ileriye yönelik kestirimler yapılabilmesi, değişkenler arasında karmaşık görünen ilişkinin anlaşılır düzeyde ifade edilerek verinin özetlemesi amaçlarına hizmet eden bir analiz tekniğidir. Çoklu doğrusal regresyon modelinde bir cevap değişkeni olmak üzere iki veya daha fazla açıklayıcı değişken yer almakta ve parametreler modele eklenmeli olarak yazılmaktadır (Alpar, 1997; Özdamar, 2003). Çoklu regresyon modelinde n adet gözlem ve k adet açıklayıcı değişken için veri kümesi Çizelge 3.1’de yazıldığı gibi gösterilmiştir.

Çizelge 3.1. Çoklu doğrusal regresyon analiz tekniğinde verilerin dizaynı

| | Y | X_1 | X_2 | ... | X_k |
|-----|-------|----------|----------|-----|----------|
| 1 | y_1 | x_{11} | x_{12} | ... | x_{1k} |
| 2 | y_2 | x_{21} | x_{22} | ... | x_{2k} |
| 3 | y_3 | x_{31} | x_{32} | ... | x_{3k} |
| · | · | · | · | ... | · |
| · | · | · | · | ... | · |
| · | · | · | · | ... | · |
| n | y_n | x_{n1} | x_{n2} | · | x_{nk} |

Yukarıda verilen k adet açıklayıcı değişkenin olduğu çoklu regresyon eşitliği aşağıdaki gibi yazılır.

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \quad i=1, 2, \dots, n \quad (3.1)$$

Eşitlikte yer alan Y_i 'nci gözleme ait cevap değişkeni,

- ✓ X_{ij} : ($i=1, 2, \dots, n$ ve $j=1, 2, \dots$, olmak üzere) j 'inci değişken ve i 'nci gözlemi ait açıklayıcı değişken
- ✓ β_0 ve β_j , regresyon denkleminin bilinmeyenlerini bir diğer ifade ile katsayıları,
- ✓ ε_i , ortalaması sıfır ve varyansı σ^2 olan, normal dağılımlı şansa bağlı hata terimini ($\varepsilon_i \sim N(0, \sigma^2)$)
- ✓ β_0 , sabit terim (intercept)
- ✓ β_j , ($j=1, 2, \dots, k$) regresyon katsayıları olarak adlandırılır. (Alpar, 1997).

Bir den çok cevap değişkeni olduğunda "Eş.3.1"de verilen doğrusal regresyon modelinin matris gösterimi aşağıdaki gibi ifade edilir.

$$y = X\beta + \varepsilon \quad (3.2)$$

Eş.3.2'de y : $n \times 1$ boyutlu cevap değişkeni için gözlemler vektörü,

X : $n \times p$ boyutlu açıklayıcı değişkenler için gözlemler matrisi (kesim noktasının modele dâhil edilmesi ile $p=I+k$ olur).

β : $p \times 1$ boyutlu bilinmeyen regresyon katsayıları vektörü.

ε : $n \times 1$ boyutlu hata vektörü.

Çoklu doğrusal regresyon analizinde, en küçük kareler (EKK) yöntemi ilk aklı gelen ve yaygın kullanılan tahmin yöntemidir. Ancak bu yöntemin kullanılabilmesi için bazı varsayımların yerine getirilmiş olması gerekir. Bu varsayımlar;

1. Hataların şans değişkeni olduğu,
2. Hataların ortalamasının sıfır diğer bir ifade ile beklenen değerinin sıfır olduğu, $E(\varepsilon_i) = 0$,
3. Hataların normal dağılıma sahip olduğu, $\varepsilon_i \sim N(0, \sigma^2)$,
4. Hataların varyansının sabit olduğu, $Var(\varepsilon_i) = \sigma^2$,
5. Hataların birbirlerinden, $i \neq i'$ için $Cov(\varepsilon_i, \varepsilon_{i'}) = 0$ ve diğer açıklayıcı değişkenlerden ve $Cov(\varepsilon_i, X_j) = 0$, bağımsız olduğu,
6. Model parametrelerinin doğrusal olduğu, bir başka ifade ile cevap ve açıklayıcı değişkenler arasında doğrusal bir ilişki olduğu,
7. Açıklayıcı değişkenler arasında çoklu bağlantı olmadığı,
8. Modelde yer alan parametre sayısının (p) gözlem sayısından (n) küçük olduğu ($p < n$),

Şeklinde sıralanmaktadır (Drapper ve Smith, 1981; Phatak ve De Jong, 1997; Weisberg, 2005). Yukarı da verilen varsayımların sağlanamaması durumunda EKK yöntemi ile elde edilen parametre tahminleri ve güven aralıkları güvenilirliğini yitirmektedir (Albayrak, 2005). Bu nedenle, EKK yöntemi matematiksel işlemler bakımından uygun tahmin yöntemi olmasına rağmen varsayımların bozulması durumunda bu yöntem eleştirilmekte ve yerine alternatif yöntemler olan Ridge, Lasso, Elastik net, temel bileşenler analizi ve kısmi en küçük kareler regresyonu gibi yöntemler önerilmektedir (Neter ve ark., 1996; Wilcox 1997; Ortiz ve ark., 2006; Mohebbi ve ark., 2007).

3.2.2. Çoklu doğrusal regresyonda en küçük kareler yöntemi

Gözlenen (Y_i) ve tahmin edilen (\hat{Y}_i) cevap değişkeni arasındaki fark hata olarak tanımlanmaktadır. EKK regresyon yönteminde hedeflenen amaç hata kareler toplamının minimize edilmesidir. Doğrusal regresyon analizi için yapılan varsayımların sağlanması durumunda EKK yöntemi ile elde edilen tahminler ve güven aralıkları doğru olmaktadır. (Neter vd. 1996, Fox 1997). Buna göre hata, Eş.3.3'de olduğu gibi yazılır.

$$e_i = Y_i - \hat{Y}_i \quad (3.3)$$

n kadar birey için k kadar açıklayıcı değişkene sahip doğrusal regresyon modeline ait Eş. 3.1 hatırlanacak olursa;

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \varepsilon_i \quad i=1,2, \dots, n$$

Eşitliği aşağıdaki formda tekrar yazılır.

$$Y_i = \beta_0 + \sum_{j=1}^k \beta_j X_{ij} + e_i \quad (3.4)$$

Çoklu doğrusal regresyon analizinde β_0 dışında k adet regresyon katsayısı için tahmin eşitliği ve hata için eşitlikler sırası ile aşağıda verildiği gibi yazılır.

$$\hat{Y}_i = \hat{\beta}_0 + \sum_{j=1}^k \hat{\beta}_j X_{ij} \quad (3.5)$$

$$e_i = Y_i - \hat{Y}_i = Y_i - \left(\hat{\beta}_0 + \sum_{j=1}^k \hat{\beta}_j X_{ij} \right) \quad (3.6)$$

Bu katsayılarla ilişkin vektör, $\hat{\beta}_j = \{\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k\}$, EKK yönteminde hata kareler toplamı (residuals sum of square: RSS) en küçük yapan değerleri ifade etmektedir. Buna göre çoklu doğrusal regresyon için EKK tahmini için verilen birbirine denk eşitlikler aşağıda yer almaktadır.

$$RSS\{\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k\} = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n \left(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{i1} - \hat{\beta}_2 X_{i2} - \dots - \hat{\beta}_k X_{ik} \right)^2 \quad (3.7)$$

$$RSS(\hat{\beta}) = \sum_{i=1}^n e_i^2 = \sum_{i=1}^n \left(Y_i - \hat{\beta}_0 - \sum_{j=1}^k \hat{\beta}_j X_{ij} \right)^2 \quad (3.8)$$

Eş 3.7 ve Eş 3.8 de verilen iki denk eşitliği en küçük yapan EKK yöntemiyle elde edilmiş değerler olur. Bu amaçla regresyon eşitliğindeki herhangi regresyonun

($\hat{\beta} = \{\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k\}$) tahmin edilebilmesi için Eş 3.7 ve Eş 3.8 de birinci dereceden kısmi türevi alınır, bu eşitlik, genel olarak Eş.3.9 ve regresyon denkleminde yer alan katsayılar için Eş.3.10 ve Eş.3.12 aşağıda gösterildiği gibidir.

$$\frac{\partial RSS}{\partial \hat{\beta}_j} = 2 \sum_{i=1}^n e_i \frac{\partial e_i}{\partial \hat{\beta}_j} \quad j = 0, 1, \dots, k \quad (3.9)$$

$$\frac{\partial RSS}{\partial \hat{\beta}_0} = 2 \sum_{i=1}^n e_i \frac{\partial e_i}{\partial \hat{\beta}_0} = -2 \sum_{i=1}^n e_i \quad (3.10)$$

$$\frac{\partial RSS}{\partial \hat{\beta}_1} = 2 \sum_{i=1}^n e_i \frac{\partial e_i}{\partial \hat{\beta}_1} = -2 \sum_{i=1}^n X_{i1} e_i \quad (3.11)$$

$$\frac{\partial RSS}{\partial \hat{\beta}_2} = 2 \sum_{i=1}^n e_i \frac{\partial e_i}{\partial \hat{\beta}_2} = -2 \sum_{i=1}^n X_{i2} e_i \quad (3.12)$$

$$\frac{\partial RSS}{\partial \hat{\beta}_k} = 2 \sum_{i=1}^n e_i \frac{\partial e_i}{\partial \hat{\beta}_k} = -2 \sum_{i=1}^n X_{ik} e_i \quad (3.13)$$

Yukarıda verilen Eş.3.10 ve Eş.3.13 numaralı eşitlikleri sıfıra eşitlendikten sonra EKK tahmin denklemleri aşağıdaki gibidir.

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n X_{i1} + \hat{\beta}_2 \sum_{i=1}^n X_{i2} + \dots + \hat{\beta}_k \sum_{i=1}^n X_{ik} = \sum_{i=1}^n Y_i \quad (3.14)$$

$$\hat{\beta}_0 \sum_{i=1}^n X_{i1} + \hat{\beta}_1 \sum_{i=1}^n X_{i1}^2 + \hat{\beta}_2 \sum_{i=1}^n X_{i1} X_{i2} + \dots + \hat{\beta}_k \sum_{i=1}^n X_{i1} X_{ik} = \sum_{i=1}^n X_{i1} Y_i \quad (3.15)$$

$$\hat{\beta}_0 \sum_{i=1}^n X_{i2} + \hat{\beta}_1 \sum_{i=1}^n X_{i1} X_{i2} + \hat{\beta}_2 \sum_{i=1}^n X_{i2}^2 + \dots + \hat{\beta}_k \sum_{i=1}^n X_{i2} X_{ik} = \sum_{i=1}^n X_{i2} Y_i \quad (3.16)$$

$$\hat{\beta}_0 \sum_{i=1}^n X_{ik} + \hat{\beta}_1 \sum_{i=1}^n X_{i1} X_{ik} + \hat{\beta}_2 \sum_{i=1}^n X_{i2} X_{ik} + \dots + \hat{\beta}_k \sum_{i=1}^n X_{ik}^2 = \sum_{i=1}^n X_{ik} Y_i \quad (3.17)$$

Çoklu doğrusal regresyon eşitliklerinde; k adet bilinmeyen katsayı için k adet bilinmeyen eşitliğin çözümü hem uzun süre almakta hem de zor olmaktadır. Bu işlemlerden ancak matris gösterimi kullanılarak daha rahat yapılabilir. Basit veya çoklu doğrusal regresyon eşitlikleri genel bir matris ile gösterilmektedir.

$$Y = X\beta + \varepsilon \quad (3.18)$$

Eş.3.18 de $Y; nx1$ boyutlu cevap değişken vektörünü, $X; nx(k+1)$ boyutlu açıklayıcı değişkenlere ait matrisi, $\hat{\beta}; (k+1).x1$ boyutlu katsayılar vektörü ve $e; nx1$ boyutlu hata vektörünü göstermektedir. Yukarıda verilen Eş.3.18 numaralı eşitlik, n gözlem için aşağıda verildiği gibi yazılmaktadır (Alpar, 1997).

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_n \end{bmatrix}_{nx1} = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdot & \cdot & X_{1k} \\ 1 & X_{12} & X_{22} & \cdot & \cdot & X_{2k} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 1 & X_{n1} & X_{n2} & \cdot & \cdot & X_{nk} \end{bmatrix}_{nx(k+1)} \times \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdot \\ \cdot \\ \beta_k \end{bmatrix}_{(k+1).x1} + \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ e_n \end{bmatrix}_{nx1} \quad (3.19)$$

EKK yönteminde katsayılar, hata kareler toplamının minimize edilmesi esasına dayanmaktadır. Hata kareler toplamı, matris gösteriminde aşağıdaki gibi yazılmaktadır.

$$RSS(\hat{\beta}) = \sum_{i=1}^n e_i^2 = (Y - X\hat{\beta})'(Y - X\hat{\beta}) \quad (3.20)$$

$$RSS(\hat{\beta}) = \sum_{i=1}^n e_i^2 = Y'Y - \hat{\beta}'X'Y - Y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta} \quad (3.21)$$

$$RSS(\hat{\beta}) = Y'Y - 2\hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} \quad (3.22)$$

Yukarıda verilen Eş.3.22 numaralı eşitlik EKK tahmin yöntemini elde etmek için kullanılır. Bunun için eşitliğin $\hat{\beta}$ 'ya göre türevi alınır ve denklem sıfıra eşitlenir. Bu eşitlikler aşağıdaki Eş.3.23 ve Eş.3.25 numaralı eşitliklerde verildiği gibidir.

$$\frac{\partial RSS(\hat{\beta})}{\partial \hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0 \quad (3.23)$$

$$X'X\hat{\beta} = X'Y \quad (3.24)$$

$$\hat{\beta}_{ols} = (X'X)^{-1}X'Y \quad (3.25)$$

Eş.3.25 numaralı eşitlik kullanılarak $\hat{\beta}$ 'nın EKK tahmin edicileri elde edilir. Her bir katsayı için EKK tahmin eşitliğinin matris gösterimi aşağıda gösterilmiştir.

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} n & & & & \\ \sum_{i=1}^n X_{i1} & \sum_{i=1}^n X_{i1}^2 & & & \\ \sum_{i=1}^n X_{i2} & \sum_{i=1}^n X_{i1}X_{i2} & \sum_{i=1}^n X_{i2}^2 & & \\ \vdots & \vdots & \vdots & \ddots & \\ \sum_{i=1}^n X_{ik} & \sum_{i=1}^n X_{i1}X_{ik} & \sum_{i=1}^n X_{i2}X_{ik} & \dots & \sum_{i=1}^n X_{ik}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n X_{i1}Y_i \\ \vdots \\ \sum_{i=1}^n X_{ik}Y_i \end{bmatrix} \quad (3.26)$$

Simetrik

EKK yönteminde varsayımların sağlandığı durumda tahmin edilen model doğrusal, sapmasız/yansız, etkin ve tutarlı bir tahmin edici olur.

3.2.2.1. Çoklu doğrusal bağlantı problemi

Çoklu bağlantı açıklayıcı değişkenlerden bir kısmının veya çoğunluğunun aralarında tam veya tama yakın ilişkisinin olması şeklinde tanımlanmaktadır. Çoklu bağlantı aynı zamanda çoklu doğrusallık olarak da isimlendirilmektedir. Açıklayıcı değişkenler arasında bu şekilde bir durumun olması çoklu bağlantı problemi olarak bilinmektedir.

Çoklu doğrusal bağlantı, çoklu doğrusal regresyon ile ilgili çalışmalarda sıklıkla ortaya çıkan bir problemdir. Bu problem, çoklu doğrusal regresyon analizi

varsayımlarından bir veya daha fazlasının sağlanamaması durumunda ortaya çıkar. Bu varsayımlardan ilki açıklayıcı değişkenler arasında bağımsızlık varsayımının geçerli olması, bir diğer ifade ile açıklayıcı değişkenler arasındaki miktarının yüksek olması diğeri ise regresyon modelinde yer alan gözlem sayısının parametre sayısından fazla olmasıdır. Bu varsayımlardan bir veya ikisinin sağlanamaması çoklu bağlantı problemi olarak bilinen sorunun ortaya çıkmasına sebep olmaktadır.

Doğrusal bağımlılık tanımı gereğince t_1, t_2, \dots, t_k en az biri sıfırdan farklı olmak üzere, $n \times (k+1)$ boyutlu X matrisinde açıklayıcı değişkenler X_1, X_2, \dots, X_k arasında doğrusal bir ilişki olması durumu;

$$\sum_{j=1}^k X_j t_j = 0 \quad (3.27)$$

Şeklinde tanımlanabiliyorsa X_1, X_2, \dots, X_k vektörleri doğrusal bağımlıdır. Bu durumda $X'X$ matrisinin rankı $(k+1)$ 'den küçük olacak ve $X'X$ matrisinin tersi alınamayacaktır

$((X'X)^{-1})$. Diğer bir ifade ile tam çoklu bağlantı varlığından bahsedilecektir. Aşağıda verilen Eş 3.28 numaralı eşitliğin geçerli olması durumunda ise güçlü çoklu bağlantı söz konusu olur.

$$\sum_{j=1}^k X_j t_j \cong 0 \quad (3.28)$$

Çoklu bağlantının olması durumunda;

1. $(X'X)^{-1}$ Hesaplanabilmesine rağmen EKK yöntemi kullanılarak regresyon analizinden elde edilen, parametre tahminleri gerçek değerlerinden farklı olacak ve katsayıların varyans ve kovaryansları büyük olacaktır.
2. Bu yöntem ile elde edilen tahminler yansız olmasına rağmen, tahminlerin işareti kuramsal beklentilerle çelişecek şekilde farklı olabilir.
3. Modelin belirleme katsayısının yüksek olmasına rağmen açıklayıcı değişkenlerin ya hiçbiri ya da çok az bir kısmı anlamlı olabilir.

4. Tam çoklu bağlantı problemi ise regresyon katsayılarının belirsiz olmasının yanı sıra bu katsayıların standart hatalarının sonsuz olmasına neden olur (Maxwell, 2000; Montgomery ve ark., 2001).

3.2.2.2. Çoklu bağlantı sorununu belirlenmesi

Çoklu bağlantı probleminin belirlenmesinde çeşitli bilgilerden faydalanılabilir. Bu bilgiler aşağıdaki gibi özetlenebilir:

- Regresyon katsayılarının işaretinin ve büyüklüğünün kuramsal beklentilerin tersi olması,
- Veri setindeki herhangi bir açıklayıcı değişkenin modelden çıkarılması veya eklenmesi ve/veya herhangi bir gözlemin modelden çıkarılması veya eklenmesi durumunda katsayılar da belirgin değişimin meydana gelmesi

Bunlarla birlikte çalışmalarda genellikle çoklu bağlantı problemini belirlemede kullanılan yöntemler aşağıda özetlenmiştir. (Alpar 2003, Albayrak, 2006).

1. *Çoklu bağlantının basit korelasyon katsayıları ile belirlenmesi:*
İki açıklayıcı değişken arasındaki basit korelasyon katsayısının 0.80'den büyük veya 1'e çok yakın olması durumunda, çoklu doğrusal bağlantı probleminin irdelenmesi gerektirir. Ancak yalnızca basit korelasyon katsayıları değerlendirilerek çoklu doğrusal bağlantının varlığına karar vermek yeterli olmaz. (Albayrak, 2005).
2. *Çoklu bağlantının F ve t test istatistikleri karşılaştırılarak belirlenmesi:*
Regresyon modelinin anlamlılığının test edilmesi için kullanılan F testi sonucunun anlamlı çıkmasına rağmen, katsayılar için yapılan t testinin tüm açıklayıcı değişkenler için anlamsız çıkması durumunda çoklu doğrusal bağlantının olduğu düşünülür.
3. *Çoklu bağlantının $X'X$ matrisinin öz değerleri ile belirlenmesi:* $X'X$ matrisinin öz değerleri bir veya bire yakın değerlerden meydana geliyor ise çoklu doğrusal bağlantı problemi düşünülmez. Ancak $X'X$ matrisinin öz değerlerinin birkaçının sıfıra çok yakın ya da sıfır değerlerine sahip olması durumunda çoklu bağlantı probleminin varlığından söz edilebilir.
4. *Çoklu bağlantının koşul sayısı veya koşul endeksi ile belirlenmesi:* Koşul Sayısı (Condition Number: CN) ve Koşul Endeksi (Condition Index: CI) $X'X$

matrisinin öz değerleri kullanılarak elde edilen iki ölçüttür. $X'X$ matrisinin en büyük özdeğer λ_{\max} ve en küçük öz değerleri λ_{\min} olmak üzere, CN ve CI aşağıda Eş. 3.29 ve Eş.3.30 numaralı eşitliklerde verildiği gibidir (Montgomery ve Peck, 1982; Gujarati,1995; Albayrak, 2005).

$$CN = \frac{\lambda_{\max}}{\lambda_{\min}} \quad (3.29)$$

$$CI = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}} \quad (3.30)$$

Çoklu doğrusal bağlantının varlığını incelemek amacı ile kullanılan CI; $CI < 10$ olması durumunda çoklu doğrusal bağlantı probleminin olmadığı söylenir. Eğer $10 \leq CI \leq 30$ ise çoklu doğrusal bağlantının varlığın ve derecesinin orta düzeyde olduğu düşünülür Buna ek olarak $CI \geq 30$ ise üzerinde çalışılan veri kümesinde yüksek düzeyde çoklu doğrusal bağlantı probleminin olduğu anlaşılır (Gujarati,1995; Pamukçu, 2010).

5. Çoklu bağlantının varyans artırıcı faktörü ile belirlenmesi:

Varyans artırıcı faktör (Variance Inflation Factors: VIF) $X'X$ matrisinin tersinin kullanılması ile yani $VIF_j = (X'X)^{-1}$ kullanılarak elde edilir. $VIF_j = (X'X)^{-1}$ Eşitliğinde $(X'X)^{-1}$ matrisinin j'ninci köşegen elemanı j'ninci varyans artırıcı faktörü gösterebilir, o halde j'ninci VIF değeri için aşağıdaki eşitlik yazılır (Marquart, 1975).

$$VIF_j = \frac{1}{1 - R_j^2} \quad (3.31)$$

Eş.3.31 de R_j^2 , X_j 'ninci değişken cevap değişkeni kabul edilerek açıklayıcı diğer değişkenler ile arasındaki çoklu korelasyon katsayısının karesini ifade etmektedir. Dolayısıyla açıklayıcı değişkenler arasında herhangi bir ilişki yok ise $VIF=0$ ($R^2 = 0$ olduğu için) olur. Ancak cevap değişkenler arasında tam ilişki var ise $VIF = \infty$ ($R^2 = 1$ olduğu için) olur. Bu konuda yapılan araştırmalar, VIF değerinin 10'a eşit veya daha büyük ($VIF \geq 10$) olması durumunda çoklu doğrusal bağlantı probleminin varlığından bahsedilebileceğini göstermiştir (Webster, 1995).

6. Çoklu bağlantının tolerans değeri ile belirlenmesi:

Tolerans değeri (Tolerance Value: TV) çoklu doğrusal bağlantı problemini belirlemede kullanılan yöntemlerden bir diğeridir. TV, VIF ile yakından ve ters ilişkili olup bu ilişki Eş.3.32' de verilmiştir.

$$TV = \frac{1}{VIF_j} = (1 - R_j^2) \quad (3.32)$$

Dolayısıyla VIF değerinin büyük olması TV'nin küçük olacağı anlamına gelirken, VIF değerinin küçük olması da TV'nin büyük olacağı anlamına gelmektedir. Diğer bir ifade ile TV'nin küçük olması, veri kümesinde çoklu doğrusal bağlantı problemine sahip olduğunun bir göstergesidir (Albayrak, 2005).

3.2.2.3. Çoklu doğrusal bağlantı probleminin giderilmesi

Çoklu doğrusal bağlantı probleminin neden olduğu sorunların giderilmesi için yapılan bazı öneriler vardır. Bunlar; örnek büyüklüğünü artırmak amacıyla veri kümesine yeni gözlemlerin eklenmesi, bağımsız değişkenlerden bir veya birkaçının azaltılması, değişkenlere transformasyon uygulanması ve modelin yeniden tanımlanması şeklinde özetlenebilir (Netter ver ark., 1996; Gujarati, 1995; Orhunbilge, 2002).

Ancak bu öneriler ile her zaman bu problemin giderilmesi mümkün olmamaktadır. Değişkenler ve model üzerinde herhangi bir değişiklik yapmadan çoklu doğrusal bağlantı problemini giderilebilmek amacıyla yapılan bir diğer öneri; Ridge, Lasso, Elastic Net, temel bileşenler regresyonu gibi yanlı tahmin yöntemlerinin kullanılmasıdır (Vinod ve Ullah, 1981; Kalaycı, 2008; Pamukçu, 2010). Zira EKK tahmin edicileri ancak varsayımların yerine getirilmesi durumunda en iyi yansız tahmin edicidir. Çoklu bağlantı probleminin varlığı durumunda bu özellik yerine getirilememektedir. İyi bir tahmin edici özelliklerinden olan yansızlık özelliğinin bir miktar esnetilmesi ile kullanılan yanlı regresyon yöntemleri, çoklu bağlantı problemi olduğunda sıklıkla uygulanan yöntemlerdir. Yansızlık özelliğinin esnetilmesi ile elde edilen tahmin edicinin hata kareler ortalaması (HKO) yansız bir tahmin ediciden daha küçük olabilmektedir (Rawlings ve ark., 1998).

Çalışmanın bu bölümünde temel bileşenler regresyonu ve kısmi en küçük kareler regresyonuna ilişkin bilgiler verilecektir.

3.2.3. Temel bileşenler analizi

İlk kez Karl Pearson tarafından ortaya koyulan Temel Bileşenler Analizinin (TBA), sonraki yıllarda Hotelling (1933) ve Rao (1964) tarafından uygulama alanları geliştirilmiştir (Timm, 2002).

Girshick (1936) temel bileşenlerde bazı alternatif yöntemleri geliştirmiş ve örnek temel bileşenlerinin ve anakütle temel bileşenlerin en yüksek olasılık tahmini olduğunu vurgulamıştır. Gover (1966) bu çalışmada temel bileşenler analizi ile diğer istatistikî yöntemler arasındaki ilişkiyi araştırmış ve önemli geometrik yöntemleri ortaya koymuştur. Jeffers (1967) temel bileşenler analizinin, bileşenleri indirgemesi dışında daha ileri bir seviyeye getirilebileceğini vurgulamıştır (Alkan, 2008).

Sosyal bilimlerden fen bilimlerine kadar geniş alanda yoğun bir şekilde uygulanmakta olan TB regresyon modelinde yer alan açıklayıcı değişkenler arasındaki bağımlılık yapısını yok etmek ve/veya boyut indirgemek amacıyla kullanılmaktadır. TB regresyonu başlı başına bir yöntem olarak uygulanmasının yanı sıra veri hazırlama tekniği olarak da kullanılmaktadır (Arnold, 1981; Ludwig ve Reynolds, 1988; Sharma, 1996.).

Açıklayıcı değişkenlerin bağımsızlığı varsayımının geçerliliğini yitirmesi durumunda bilinen EKK yöntemi ile elde edilen sonuçlar güvenilirliğini yitirir. Açıklayıcı değişkenler arasındaki ilişki çok yüksek ise yani çoklu bağlantı sorunu varsa çözüm belirsiz hale gelebilir (Alpar, 2011

). Bu durumda kullanılan alternatif yöntemlerden biri TB regresyon yöntemidir. Veri kümesinde n adet birey ve p kadar değişkene sahip olduğu düşünülün. Veri kümesindeki değişkenlerin ölçü birimi birbirlerinin aynı ise TB regresyonu yapılırken varyans kovaryans matrisinden faydalanılır. Buna karşın, orijinal veri kümesindeki değişkenlerin ölçü birimleri birbirlerinden farklı ise analiz yapılırken korelasyon matrisinden faydalanılmaktadır (Özdamar, 2004).

TB regresyonu “her doğrusal regresyon modelinin bir dik açıklayıcı değişkenler kümesine dayanarak yeniden açıklanabileceği” gerçeği üzerine oluşturulmuştur. Bu yeni ve dik açıklayıcı değişkenler orijinal açıklayıcı değişkenlerin doğrusal bileşeni olarak elde edilir ve açıklayıcı değişkenlerin temel bileşenleri olarak adlandırılır (Alpar, 2011).

Çok değişkenli istatistik analizinde n tane bireye (nesne) ilişkin p adet değişken (özellik) incelenmektedir. Bu değişkenlerden birçoğunun birbiriyle ilişkili ve değişken sayısının (p) çok büyük olması, çeşitli değerlendirmeler yapılmasını güçleştirmektedir. Yöntemde karşılıklı bağımlılık yapısı gösteren, ölçüm sayısı n olan p adet değişken; doğrusal, ortogonal ve birbirinden bağımsız olma özelliklerini taşıyan k ($k \leq p$) tane yeni değişkene dönüştürülür (Ersungur ark., 2007).

Temel bileşenler analizin 3 temel amacı vardır.

- 1- Tahminleme yapmak
- 2- Boyut indirgemesi yapmak
- 3- Bazı yöntemlerin analiz edebileceği modeli oluşturmak

Bir modeli ifade etmek üzere p değişken varsa bu değişkenlerin belirlediği toplam değişkenliği ifade etmek üzere k sayıda ana değişken bularak daha az sayıda değişkenle çalışmak amaçlanır yani p boyutlu uzay yerine ($k < p$) k boyutlu uzayda çalışılır. Böylece boyut indirgemesi yapılır (Özdamar, 1999).

N adet gözlem ve p adet açıklayıcı değişkenin olduğu orijinal veri setine ait $n \times p$ boyutlu X matrisi aşağıdaki gibi verilmiş olsun.

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1j} & \dots & x_{1p} \\ x_{21} & x_{22} & x_{23} & \dots & x_{2j} & \dots & x_{2p} \\ x_{31} & x_{32} & x_{33} & \dots & x_{3j} & \dots & x_{3p} \\ \cdot & \cdot & \cdot & \dots & \cdot & \dots & \cdot \\ x_{i1} & x_{i2} & x_{i3} & \dots & x_{ij} & \dots & x_{ip} \\ \cdot & \cdot & \cdot & \dots & \cdot & \dots & \cdot \\ x_{n1} & x_{n2} & x_{n3} & \dots & x_{nj} & \dots & x_{np} \end{bmatrix} \quad (3.33)$$

p adet açıklayıcı değişken aşağıda verildiği gibi yazılabilir.

$$X_1 = \begin{bmatrix} x_{11} \\ x_{21} \\ \cdot \\ x_{i1} \\ \cdot \\ x_{n1} \end{bmatrix}, X_2 = \begin{bmatrix} x_{12} \\ x_{22} \\ \cdot \\ x_{i2} \\ \cdot \\ x_{n2} \end{bmatrix}, \dots, X_j = \begin{bmatrix} x_{1j} \\ x_{2j} \\ \cdot \\ x_{ij} \\ \cdot \\ x_{nj} \end{bmatrix}, \dots, X_p = \begin{bmatrix} x_{1p} \\ x_{2p} \\ \cdot \\ x_{ip} \\ \cdot \\ x_{np} \end{bmatrix} \quad (3.34)$$

Burada X_1, X_2, \dots, X_p değişken vektörleridir. Bu doğrusal bileşenler arasında çoklu bağlantı problemi olması, koordinat eksenleri X_1, X_2, \dots, X_p olan orijinal sistemin döndürülmesini gerektirir. Bu döndürme işlemi ile yeni koordinat sistemi

oluşturulurken bu sistemin birbirinden bağımsız hale getirilmesi sağlanmış olur. Bunun bir sonucu olarak değişimin yapısı, başlangıç modelindeki değişken sayısından daha az değişken ile ve daha kolay bir şekilde açıklanabilmektedir.

Açıklayıcı değişkenler matrisinin birinci temel bileşenine ait aşağıdaki eşitlikte verilmiştir.

$$Y_1 = t_1'X = t_1'X' = t_{11}X_1 + t_{21}X_2 + \dots + t_{p1}X_p \quad (3.35)$$

Bu eşitlikte $t_1' = t_{11}, t_{21}, \dots, t_{p1}$ ve $X' = X_1, X_2, \dots, X_p$ olduğu açıklaması yapıldıktan sonra Eş. 3.35'de, matris formunda aşağıdaki gibi yazılır.

$$t_1'X = (t_{11}, t_{21}, \dots, t_{p1}) \begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_p \end{bmatrix} \quad (3.36)$$

Bu çarpma işleminin gerçekleştirilmesi ile Eş. 3.36 ile elde edilmiş olur. Açıklayıcı değişkenler matrisinde yer alan p adet değişkenin doğrusal bileşenlerinin bulunması için bu matrisin varyans-kovaryans matrisi oluşturulur. Daha sonra bu matrisin özdeğer ve öz vektörleri bulunur (Ludwig ve Reynolds1988; Alpar, 1997). Açıklayıcı değişkenlerin varyans-kovaryans matrisi olan S aşağıdaki eşitlikte verilmiştir.

$$S = \begin{bmatrix} s_{11} & s_{12} & s_{13} & \dots & s_{1p} \\ s_{21} & s_{22} & s_{23} & \dots & s_{2p} \\ s_{31} & s_{32} & s_{33} & \dots & s_{3p} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ s_{p1} & s_{p2} & s_{p3} & \dots & s_{pp} \end{bmatrix} \quad (3.37)$$

Burada,

$$s_{jj'} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_{.j})(x_{ij'} - \bar{x}_{.j'})}{n-1} \quad j=1,2,\dots,p \quad j'=1,2,\dots,p \quad (3.38)$$

Eş. 3.35'de verilen birinci temel bileşenin varyansı $t_1' t_1 = 1$ kısıtlaması altında aşağıda verildiği gibi yazılır (Morrison, 1967).

$$Cov(Y_1) = E\left\{ (t_1' X) (t_1' X)' \right\} = E(t_1' X X' t_1) = t_1' S t_1 \quad (3.39)$$

$t_1' t_1 = 1$ Kısıtlaması altında $t_1' t_1 - 1 = 0$ olmaktadır ve Lagranj (Lagrange) fonksiyonunun kullanılması ile, $\varphi(t_1, \lambda_1) = t_1' S t_1 - \lambda_1 (t_1' t_1 - 1) = 0$ elde edilir. Bu eşitliğin t_1 'e göre kısmi türevi alınıp sıfıra eşitlendikten sonra aşağıdaki eşitlikler elde edilir.

$$2S t_1 - 2\lambda_1 t_1 = 0 \quad (3.40)$$

$$(S - \lambda_1 I) t_1 = 0 \quad (3.41)$$

Eş 3.41 de λ_1 değeri, S matrisinin birinci bileşenine ait özdeğeri ve t_1 , birinci özdeğerinden elde edilen birinci özvektörünü ifade eder. $t_1' t_1 = 1$ Kısıtlaması nedeniyle, çözümün elde edilebilmesi için aşağıda verilen eşitliğin sağlanması gerekir.

$$|S - \lambda_1 I| = 0 \quad (3.42)$$

Diğer bir ifade ile Eş. 3.42'de verilen bilgi Eş. 3.41'de birinci terimin determinantının sıfır olmasını gerektirir. Açıklayıcı değişkenler matrisi için p kadar temel bileşen aşağıda verildiği gibi olur (Morrison, 1967; Tatlıdil, 1996; Shrestha ve Kazama, 2007; Widodo ve ark, 2007).

$$\begin{aligned} Y_1 &= t_1' X' = t_{11} X_1 + t_{21} X_2 + \dots + t_{p1} X_p \\ Y_2 &= t_2' X' = t_{12} X_1 + t_{22} X_2 + \dots + t_{p2} X_p \\ Y_p &= t_p' X' = t_{1p} X_1 + t_{2p} X_2 + \dots + t_{pp} X_p \end{aligned} \quad (3.43)$$

Her bir temel bileşenin varyans ve kovaryansı sırasıyla,

$$\text{Var}(Y_i) = t_i' S t_i, i = 1, 2, \dots, p$$

$$\text{Cov}(Y_i, Y_k) = t_i' S t_k, i, k = 1, 2, \dots, p \quad (3.44)$$

Yukarıda verilen Eş.3.42 de'p özdeğer için genelleştirirsek aşağıdaki eşitlik elde edilir.

$$|S - \lambda I| = 0 \quad (3.45)$$

Eş.3.45'in çözümü sonucunda λ 'nın p. dereceden polinomu elde edilmektedir. Bu polinomunda p adet kök bir diğer ifade ile özdeğer bulunur. Bunlar büyüklüklerine göre $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ şeklinde ifade edilir. Bu bileşenlerin varyansı özdeğere eşit olmaktadır. Dolayısıyla özdeğerlerin bu şekilde büyüklüklerine göre sıraya dizilmesi ile varyanslar da büyüklüklerine göre sıralanmış olur. Buna göre ilk bileşenin varyansı en büyük varyans olmaktadır. Dolayısıyla birinci bileşenin varyansı toplam varyasyon içinde en çok varyans miktarına sahip olur (Tatlıldil,1996; Skrbic ve Onjia, 2007).

Birinci bileşenin ardından ikinci bileşen bağımsız doğrusal bileşenler içinde en büyük ikinci varyansa sahip olur ve bu üçüncü, dördüncü vb. şeklinde devam eder. Toplam varyans içindeki en küçük varyansın miktarı son bileşenin varyansına ait olur. Birinci bileşen $Y_1 = t_1' X$ ($t_1' t_1 = 1$ koşulu) ile ikinci bileşen $Y_2 = t_2' X$ 'nin ($t_2' t_2 = 1$ koşulu) birbirinden bağımsızlığı, ($t_1' t_2 = 0$) açıklaması altında, aşağıdaki eşitlik ile ispat gösterilir.

$$E(Y_1 Y_2) = E[(t_1' X)(t_2' X)'] = E(t_1' X X' t_2) = t_1' S t_2 = \lambda_1 t_1' t_2 = 0 \quad (3.46)$$

Bu durumda Y_1 ve Y_2 bileşenleri birbirinden bağımsızdır yani birbirine diktir (Kleinbaum ve ark., 1998; Özdamar, 2004).

Temel bileşenler aynı zamanda standardize edilmiş değişkenlerden de elde edilebilir. Standardize edilmiş değerler genellikle değişkenlerin ölçü birimleri

birbirinden farklı olduğu durumda kullanılır. $z_{ij} = (X_{ij} - \mu)/\sigma_{ij}$ eşitliği kullanılarak her bir değişkene ilişkin standardizasyon yapılır. Standardize edilmiş değişkenler için varyans-kovaryans matrisi olan S ile korelasyon matrisi R birbirinin aynısı olur. Korelasyon matrisi hem standardize edilmiş değişkenler hem de orijinal değişkenler için aynıdır. Bu nedenle orijinal veya standardize edilmiş değişkenler için R matrisi ve standardize edilmiş değişkenler için S veya R matrisi kullanıldığında TB regresyonuna benzer sonuçlar vermektedir. İstatistik paket programları, standardize edilmiş değişkenler için kovaryans matrisi kullanılarak TB regresyon uygulamaları yapmaktadır.

Verilerin standardize edilmesi p adet değişken için Eş.3.47 de verilmiştir.

$$Z = \begin{bmatrix} \frac{X_{11} - \bar{X}_1}{\sqrt{\sigma_{11}}} & \frac{X_{12} - \bar{X}_1}{\sqrt{\sigma_{11}}} & \dots & \frac{X_{1p} - \bar{X}_1}{\sqrt{\sigma_{11}}} \\ \frac{X_{21} - \bar{X}_2}{\sqrt{\sigma_{22}}} & \frac{X_{22} - \bar{X}_2}{\sqrt{\sigma_{22}}} & \dots & \frac{X_{2p} - \bar{X}_2}{\sqrt{\sigma_{122}}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{X_{n1} - \bar{X}_p}{\sqrt{\sigma_{pp}}} & \frac{X_{n2} - \bar{X}_p}{\sqrt{\sigma_{pp}}} & \dots & \frac{X_{np} - \bar{X}_p}{\sqrt{\sigma_{pp}}} \end{bmatrix} \quad (3.47)$$

Eş.3.47 de matris formunda $Z = (S^{1/2})^{-1}(X - \mu)$ olarak yazılır. Korelasyon matrisi R, $R = Cov(Z)$ eşitliği kullanılarak hesaplanır. Bu matrisin özdeğerlerinin bulunması için $|R - \lambda I| = 0$ eşitliği kullanılarak p adet özdeğer $(\lambda_1, \lambda_2, \dots, \lambda_p)$ hesaplanır.

3.2.3.1. Temel Bileşen Sayısının Verilmesi

Temel bileşenler analizi yapılarak bileşenlerin bulunmasının ardından bileşen sayısına karar vermek bu analiz yönteminde önemli bir konudur. Bu amaca hizmet üzere geliştirilmiş birçok yöntem vardır. Bunlardan basit ve kolay olması nedeni ile sıklıkla tercih edilen yöntem, λ değerlerinin toplam varyasyonun 2/3'ünü geçinceye kadar toplanmasını önerir. Bu oranı sağlamak için kullanılan λ değerlerinin sayısı bileşen sayısı olmaktadır. Daha önce bahsedildiği gibi bulunacak bileşen sayısı k olsun. Eş. 3.48 'de verilen temel bileşen sayısının toplam varyasyon içindeki payını açıklamak

mümkün olmaktadır (Rencher, 2002; Özdamar, 2004; Pierce ve ark., 2006; Skrbic ve Onjia, 2007).

$$\frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^p \lambda_i} > \frac{2}{3}, k = 1, 2, 3, \dots, p \quad (3.48)$$

Temel Bileşenler (TB) regresyonu, regresyon katsayılarının ($\hat{\beta}$) tahmin edilmesi için orijinal değişkenler yerine bağımsızlığı sağlayan dik dönüşüm kullanılarak elde edilen yeni değişkenlere EKK yöntemi uygulanması olarak ifade edilebilir. Böylece bağımsız değişkenler arasındaki çoklu bağlantı sorunu giderilmiş olur. TB regresyonu uygulanarak elde edilen hata kareleri ortalamasının EKK yönteminden elde edilenden daha küçük olması beklenir. Çoklu doğrusal regresyon modeline ait eşitliğin matris formunda gösterimini için Eş 3.18 tekrar yazılacak olursa

$$Y = X\beta + \varepsilon$$

EKK yönteminde regresyon katsayılarının tahmini için

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

Eş.3.25'de verilmişti. Bağımsız değişkenler için TB regresyonu ile değişken dönüşümü Eş.3.49 kullanılarak yapılır.

$$X'X = PD'P = W'W \quad (3.49)$$

Eş.3.49 da D; $X'X$ köşegen matrisinin öz değerleri ve P; $X'X$ özvektör matrisini göstermektedir. Temel bileşenlerden elde edilen bu sete uygulanan EKK yöntemi Eş.3.50 ile açıklanır.

$$\gamma = (W'W)^{-1}W'Y \quad (3.50)$$

Eş.3.50 de γ ; TB regresyonu kullanılarak elde edilen katsayısını gösterir. Regresyon katsayıları γ ve $\hat{\beta}$ arasındaki ilişki $\gamma = P'\hat{\beta}$ ve $\hat{\beta} = P\gamma$ şeklinde gösterilir. $\hat{\beta} = P\gamma$ Kullanılarak elde edilen regresyon katsayıları daha sonra orijinal ölçeğine dönüştürülür (Ortabaş, 2001; İşi, 2002; Onay, 2003).

3.2.4. Kısmi en küçük kareler regresyonu

KEKK yönteminde açıklayıcı değişkenler matrisi X ile cevap değişkenler matrisi Y arasındaki ilişkiyi bulunmaya çalışılırken ortogonal gizil (latent) bir diğer ifade ile direk olarak gözlemlenemeyen veya ölçülemeyen değişkenlerden faydalanılır. Bu yöntemde ortogonal gizil değişkenler, Y ile X matrisleri arasındaki kovaryansın maksimum yapılması amacı ile kullanılır (Garthwaite, 1994).

KEKK algoritmasının genel yapısında X birinci kümeyi gösteren N boyutlu bağımsız değişkenler ve Y ikinci kümeyi gösteren M boyutlu bağımlı değişkenler olsun. KEKK bu iki küme arasındaki ilişkiyi modellemek için skor vektörlerinden faydalanır. Her bir değişken kümesinden n adet veri örneği gözlemlendikten sonra, EKK'da sıfır ortalamalı ($n \times N$) boyutlu X matrisini ve sıfır ortalamalı ($n \times M$) boyutlu Y matrisi Eş.4.1 de verilen şekilde ayrıştırılır.

$$X = TP' + E$$

$$Y = UQ' + F \quad (4.1)$$

Eş 4.1 de, T ve U $n \times p$ boyutlu skor matrisleri olmak üzere p elde edilen bileşen veya gizil vektör sayısıdır. Eş.4.1 de yer alan P, $n \times p$ boyutlu ve Q, $m \times p$ boyutlu yük matrislerini, E, $n \times N$ boyutlu ve F, $n \times M$ boyutlu hata matrislerini göstermektedir. Eş.4.1'de verilen bu ayırmanın detayları Martens ve Naes (1989) tarafından verilmiştir (Bulut, 2011). Wold (1975), doğrusal olmayan yinelemeli kısmi en küçük kareler (NIPALS) algoritmasının, KEKK yönteminin bilinen yapısını oluşturduğunu göstermiştir (Rosipal ve Kramer, 2006). KEKK yönteminde yük matrisleri olan w ve c aşağıdaki gibi tanımlanır.

$$[Cov(t, u)]^2 = [cov(Xw, Yc)]^2 = \max_{|r|=|s|=1} [cov(Xr, Ys)]^2 \quad (4.2)$$

Eş.4.2’de verilen skor vektörü olan t ve u arasındaki kovaryansı göstermektedir ve $\text{cov}(t, u) = t'u / n$. NIPALS algoritması, Y uzayındaki herhangi bir u skor vektörü ile işe başlar ve aşağıda verilen adımlar izlenir. Bu işlemler yakınsama gerçekleşinceye kadar devam eder.

$$1- w = (X'u) / (u'u)$$

$$2- \|w\| \rightarrow 1$$

$$3- t = Xw$$

$$4- c = Y't / (t't)$$

$$5- \|c\| \rightarrow 1$$

$$6- u = Yc$$

Yukarıda verilen aşamaların sonunda $u=y$ ise bağımlı değişken sayısının bir olduğu söylenir. Diğer bir ifade ile Y bir boyutlu vektördür ve artık sadece y ile gösterilir. Bu durumda bu algortmada yakınsama ilk iterasyonda gerçekleştirilmiş olur. Eş.4.3’te verilen özdeğer problemi çözüldüğünde ağırlık vektörünün (w) ilk özvektöre eşit olduğu gösterilir.

$$X'YY'Xw = \lambda w \quad (4.3)$$

t ve u için aşağıda verilen tanımlamalar yapılır.

$$t = Xw \quad (4.4)$$

$$u = Yc$$

Eş.4.4’de yer alan iterasyonun dördüncü aşamasında tanımlandığı gibidir ($c = Yt / (t't)$). Bu özdeğer problemlerinden birisinin çözümlenmesi, NIPALS algoritmasında tanımlanan ilişkileri kullanarak skor ve ağırlık vektörlerinin hesaplanmasını sağlamaktadır (Rosipal ve Kramer, 2006).

Kısmi en küçük kareler modeli klasik NIPALS algoritması ya da çekirdek algoritma kullanarak hesaplanabilir. Ancak bunun farklı yolları da vardır. Bunlardan birisi Höskuldsson (1988) tarafından gösterilmiştir. Bu algoritmalar ardışık algoritmalarıdır. Bu algoritmalarda bir gizil vektör bir seferde hesaplanır ve sıradaki gizil vektörün hesaplanması için X ve Y matrisleri indirgenir.

Genellikle KEKK algoritmasında her bir gizil vektör hesaplandıktan sonra X ve Y matrislerinin her ikisi birden indirgenir. Ancak Höskuldsson (1988) Y matrisini indirgemenin isteğe bağlı olduğunu göstermiştir. Lindgren ve arkadaşları bu avantajı kullanarak sadece X matrisini indirgeyerek daha hızlı çekirdek algoritmayı tanımlamışlardır (Dayal ve MacGregor, 1997).

KEKK Formları

İteratif bir sürece sahip KEKK yönteminde t ve u skor vektörlerinin elde edilmesinden sonra X ve Y matrisleri tanımlanır (indirgenir). Farklı şekillerde yapılan indirgeme formları KEKK'nin çeşitli formlarının ortaya çıkmasına sebep olmuştur. Bu çalışmada KEKK1, KEKK2 ve SIMPLS formları açıklanmıştır.

KEKK1, KEKK2 Formu: KEKK regresyon yöntemi olarak sıklıkla kullanılan yöntemler KEKK1 ve KEKK2 yöntemleridir. İki kümeden oluşan veri setinden biri yani cevap değişkeni kümeyi tek bir değişkenden oluşuyor ise KEKK regresyon yöntemi olarak KEKK1 kullanılır. Ancak modelde yer alan her iki küme çok boyutlu ise KEKK2 yaklaşımı kullanılır. Bu yaklaşımlarda, X ve Y arasındaki ilişki simetrik olup iki varsayım vardır. Bunlar aşağıdaki gibi verilir (Rosipal ve Kramer, 2006)

- Skor vektörleri $\{t_i\}_{i=1}^p$ Y'nin iyi bir tahmin edicisidir. Burada p elde edilen skor vektörlerinin sayısıdır.
- t ve u skor vektörleri arasında doğrusal iç ilişki vardır. Bu durumu açıklamak için aşağıdaki eşitlik verilmiştir.

$$U = TD + H \quad (4.5)$$

Eş.4.5'te D, (p x p) boyutlu diagonal matrisi ve H hata matrisini gösterir. Bir diğer durum, X ve Y arasındaki ilişkinin asimetric olmasıdır. Bu durumda öncelikle bir indirgeme yapılarak transformasyon sağlanır (X için ise) ve daha sonra skor vektörlerinin $\{t_i\}_{i=1}^p$ Y'nin iyi bir tahmin edicisi olması sağlanır. Bu skor vektörleri artık Y'nin indirgenmesi için kullanılır. Diğer bir ifade ile t üzerinde Y'nin bir unsuru (component) KEKK'in her bir iterasyonunda Y'den uzaklaştırılmış olur. Bunun için aşağıdaki eşitlikler yazılır.

$$\begin{aligned} X &= X - tp' \\ Y &= Y - tt'Y/(t'T) = Y - tc' \end{aligned}$$

KEKK iterasyonu yapılırken KEKK1'de, y 'nin indirgenmesine ihtiyaç duyulmadığı daha önce de verilmiştir (Höskuldsson, 1998; Rosipal ve Kramer, 2006).

SIMPLS Formu: Jong (1993) tarafından önerilen KEKK formu SIMPLS olarak adlandırılmıştır. Bu algorithmada ağırlık vektörleri $\{\tilde{w}\}_{i=1}^p$ X matrisine uygulanarak doğrudan bulunur. SIMPLS algoritmasının KEKK1 algoritmasına eşit olduğu ortaya koyulmuştur. Ancak cevap değişkenlerin çok boyutlu matris formuna sahip olması durumunda kullanılan KEKK2'den farklı olduğu gösterilmiştir (Jong, 1993).

3.2.4.1. Kısmi en küçük kareler regresyonu

Çoklu doğrusal regresyon analizine ilişkin varsayımların sağlanması durumunda EKK tahmin edicileri, yansız ve en küçük varyansa sahip olur. Çoklu doğrusal bağlantı problemi ile karşılaşıldığında katsayıların varyansı büyümekte ve tahminlerin güvenilirliği ve durağanlığı ortadan kalmaktadır. Alternatif tahmin yöntemleri bu yansızlıktan bir miktar ödün vererek hata karelerini azaltmakta ve bu nedenle yanlı regresyon yöntemleri olarak bilinmektedir. KEKK regresyonu TB, Lasso, Elastic Net, Ridge Regresyon gibi yanlı bir regresyon yöntemidir. Geniş miktarda uygulama alanına sahip olan KEKK regresyon yöntemi ile ilgili gelişmeler aşağıda verildiği gibi sıralanmaktadır (Kondylis, 2006).

1. Wold (1966, 1975): Doğrusal olmayan yinelemeli kısmi en küçük kareler algoritması (NIPALS).
2. Wold et al. (1983): Tek değişkenli KEKK regresyonu (KEKK1).
3. Höskuldsson (1988): KEKK regresyon optimizasyon özellikleri.
4. Helland (1988): KEKK regresyonu ve Krylov uzayları.
5. Stone ve Brooks (1990): Sürekli Regresyon.
6. Phatak (1993): KEKK regresyonu ve Lanczos-Conjugate Gradient yöntemleri.
7. Rosipal ve Trejo (2001): Çoğaltılmış çekirdek Hilbert uzaylarında Çekirdek EKK (Kernel KEKK).
8. Bastien et al. (2005): EKK geliştirilmiş regresyon.

KEKK regresyonunda cevap ve açıklayıcı değişkenlerin çok değişkenli oldukları varsayılmaktadır. Yapılan n adet gözlem için X matrisi $n \times p$ boyutlu ve Y $n \times q$ boyutlu

birer matris olur. KEKK regresyonunda amaç, $p \times q$ boyutlu B regresyon katsayıları vektörü kullanarak, X ve Y matrisleri arasında doğrusal bir ilişki bulmaktır. Bunun için aşağıda verilen regresyon eşitliği düşünülün.

$$Y = XB + E \quad (4.6)$$

KEKK regresyonda yukarıda verilen eşitlikteki X ve Y matrisleri direkt olarak kullanılmamaktadır. Bunun yerine öncelikle X ve Y matrislerinin aşağıda verildiği gibi gizil değişkenler yardımıyla modellenmesi yapılmaktadır.

$$X = TP' + E_X \quad (4.7)$$

$$Y = UQ' + E_Y \quad (4.8)$$

Burada T ve U skor matrislerini, P ve Q yük matrislerini, E_X ve E_Y ise hata matrislerini göstermektedir. Buna göre t_j ve u_j , T ve U skor matrislerinin j 'nci sütunların, p_j ve q_j ise sırasıyla P ve Q yük matrislerinin j 'nci sütunlarıdır ($j=1, 2, \dots, a$). T matrisinde x skorları ve U matrisinde y skorları yer almakta ve bunlar sırasıyla açıklayıcı ve cevap değişkenlerin doğrusal kombinasyonlarını göstermektedir. Aynı zamanda x ve y skorları arasında doğrusal içsel bir ilişki vardır. Bu ilişki aşağıda verildiği gibidir.

$$U_j = d_j t_j + \square_j \quad (4.9)$$

Eş.4.9'da verilen d_j regresyon parametresini ve h_j artıkları (hataları) gösterir. u_1 ile t_1 arasındaki ilişki güçlü ise KEKK için sadece bir bileşen söz konusu olmaktadır. Bu bileşene ait skorlar ise bu durumda y skorları için iyi bir tahmin edici olmaktadır. Genel olarak bu her zaman mümkün olmamakta dolayısıyla birden daha çok KEKK bileşeni elde edilmektedir. Eş.4.9 da verilen eşitlikte verilen x ve y skorları arasındaki ilişki, matris formunda Eş.4.10'da gösterildiği gibi olur.

$$U = TD + H \quad (4.10)$$

Eş.4.10 da D diagonal matrisidir. Bu matrisin elemanları d_1, d_2, \dots, d_n olmaktadır. H matrisi ise hata matrisidir ve sütunları \square_j olarak ifade edilir.

KEKK1'de y skorları olmadığı için Eş.4.11'de ki gibi olur.

$$y = Td + \square \quad (4.11)$$

Literatürde birden fazla KEKK algoritması mevcuttur. Hesaplama hızının artırılması için çeşitli algoritmaların geliştirilmesi bir ihtiyaç olmuştur. Algoritmalar varyans-kovaryans ve ilişki matrislerinin hesaplanmasında farklılıklara sahiptir ve kullanılacak algoritmaya karar verilirken veri matrisi dikkate alınmaktadır (Lindgren ve Rannar, 1998). Wong tarafından önerilen KEKK algoritması NIPALS algoritması ile başlamıştır. Geliştirilen algoritmalar sırasıyla şu şekilde verilmektedir (Kondylis, 2006; Bulut, 2011): Ortogonal skorlar KEKK regresyonu (Wold et al., 1983,1984), Ortogonal yükler KEKK regresyonu (Martens, 1989), Özvektör algoritması (Höskuldsson, 1988), Helland algoritması (Helland, 1988), SIMPLS algoritması (De Jong, 1993), SAMPLS algoritması (Bush ve Nachbar, 1993), Çekirdek algoritması (Rännar, 1994), UNIPALS algoritması (Glen, 1989).

Bu çalışmada NILAPS, SIMPLS algoritmaları kullanılmıştır.

3.2.4.2. Doğrusal olmayan yinelemeli kısmi en küçük kareler algoritması (NIPALS)

Wold tarafından (1966) geliştirilen ve KEKK olarak bilinen bu algoritma yine Wold tarafından (1975) modeldeki cevap değişkenin/değişkenlerin dikkate alınmasını sağlayacak şekilde geliştirilmiştir. NIPALS algoritması olarak bilinen bu algoritma iteratif (yinelemeli) bir yöntemdir. Bu algortmada kovaryans matrisini maksimize eden bileşenlerin bulunması amaçlanmaktadır. Model bileşenlerinin aynı anda elde edilmesi mümkün olmadığı için yinelemeli olarak çalışan bu algortmada, cevap değişken sayısı yinelemeyi etkileyen önemli bir faktör olarak yer almaktadır. Zira tek bir cevap değişkenin yer aldığı bir regresyon modeli için bu algoritma yinelemeli olarak çalışmamaktadır. Yinelemeli olan bu algortmada iki yakınsama kriteri kullanılabilir. Bunlardan biri bileşen sayısının istenilene ulaşması, diğeri ise açıklayıcı değişkenler matrisi X 'in sıfır olmasıdır. Cevap ve açıklayıcı değişkenlere ilişkin X ve Y matrislerinin standartlaştırıldığı varsayımı altında NIPALS algoritması aşağıda verilen aşamalardan oluşur.

1. Cevap deęişkenler matrisinin herhangi bir sütunu başlangıç vektörü olarak alınır. Cevap deęişkenin tek olması durumunda ise cevap deęişken vektörü alınır ($u = y$).
2. X 'in $u=y$ üzerine regresyonundan X ve y arasındaki kovaryansı en çok açıklayan w ağırlık vektörü elde edilir. ($w = X'u/u'u$)
3. Ağırlık vektörü w birim uzunluęuna sahip olacak şekilde ölçeklendirilir. ($w_y = w/\|w\|$)
4. X matrisinin skorları olan t vektörü hesaplanır. ($t = Xw_y$)
5. t vektörü y cevap deęişkeni modellemedeki katkısını açıklayacak c ağırlık vektörü bulunur. ($c = u't/t't$)
6. c ağırlık vektörü birim uzunluęuna sahip olacak şekilde ölçeklendirilir. ($c_y = c/\|c\|$).
7. u vektörü c ağırlık vektörü ile Y 'nin doğrusal bir kombinasyonu olacak şekilde hesaplanır ($u_y = Yc_y/c'_y c_y$).
8. Hesaplanan yeni u_y vektörü ile bir önce hesaplanan u vektörü yakınsama kriteri bakımından deęerlendirilir. Yakınsama kriteri olarak $\|u_y - u\| / \|u_y\| < 10^{-6}$ dikkate alınır. Bu kriter yerine getiriliyor ise algoritmanın sonraki adımlarına geçilir. Yerine getirilemiyor ise 7. aşamada hesaplanan u_y deęeri 2. adım da yerine koyulmak koşulu ile algoritma yeniden çalıştırılır.
9. p yük vektörünün hesaplanması için X matrisi ve t vektörü kullanılır ($p = X't/(t't)$). Burada p yük vektörü, t vektörünün açıklayıcı deęişken üzerindeki etkisini göstermektedir.
10. Q yük vektörünün hesaplanması için Y matrisi ve u vektörü kullanılır ($q = (Y'u_y/(u'_y u_y))$). Burada q yük vektörü, u vektörünün cevap deęişken üzerindeki etkisini göstermektedir.
11. X ve Y matris bileşenleri ayrı ayrı hesaplandıęı için bu bileşenler arasındaki ilişki zayıf olmaktadır. Bu duruma karşı önlem olarak Y 'nin u bileşeninin X 'in t bileşeni üzerine regresyonundan b katsayısı hesaplanır. Burada içsel ilişkiyi ifade eden katsayıdır.

12. Cevap ve açıklayıcı değişkenleri modellemek amacıyla bu algoritmada elde edilen bileşenler ve yükler kullanılır. Bunun için ilgili eşitlikler aşağıda verilmiştir.

$$X = TP'$$

$$Y = BTC' \quad (4.12)$$

Algoritmadaki bu adımı hata matrislerinin hesaplanması takip eder. Bunun için ilgili açıklama aşağıda verilmiştir.

$$X \rightarrow X - tp'$$

$$Y \rightarrow Y - btc' \quad (4.13)$$

Modelde yer alana cevap değişkeninde meydana gelen değişimin önemli ölçüde açıklanması durumunda algoritma sonlandırılır.

Höskuldsson tarafından NIPALS bir özdeğer/özvektör problemi şeklinde dikkate alınan bu algoritmada hesaplanan skor ve ağırlık vektörleri (w, c, t, u) için aşağıdaki eşitlikler verilmiştir.

$$wa_1 = (X'YY'X)w \quad (4.14)$$

$$ca_2 = (Y'XX'Y)c \quad (4.15)$$

$$ta_3 = (XX'YY')t \quad (4.16)$$

$$ua_4 = (YY'XX')u \quad (4.17)$$

Eş.4.14 –Eş.4.17’de a_1, a_2, a_3 ve a_4 öz değerleri gösterir. Yine aynı eşitliklerde yer alan w, c, t ve u matrislerinin normları bire eşittir. Eş.4.14’ de yer alan matrisin boyutu p x p Eş.4.17’de yer alan matrisin boyutu ise m x m’dir. Diğer bir ifade ile bu matrislerin boyutu cevap ve açıklayıcı değişkenlerin sayıları ile açıklanır. Eş.4.16 ve Eş.4.17 ‘de verilen matrislerin boyutları ise verideki gözlem sayısı ile ifade edilir. Buna

göre, gözlem veya değişken sayısının çok fazla olmasına rağmen bu algoritmanın çalışması sırasında küçük boyutlu matrislerle tanımlanabilmektedir (Lindgren ve Rännar, 1998).

Ortogonalitenin (bağımsızlığın) önemli olduğu bu algotirmada t_i vektörü ve E_i hata matrisleri için aşağıda verilen eşitliklerin sağlanması gereklidir.

$$\begin{aligned} t_i &= E_{i-1} w_i E_i \\ E_i &= X - \sum_{j=1}^i t_j p \\ (4.18) \end{aligned}$$

$$E_0 = X$$

Çok değişkenli ve tek değişkenli KEKK algoritmasının ilk adımı t_i vektörlerinin doğrusal kombinasyonlarının hesaplanmasında kullanılan w_i ağırlık vektörlerinin hesaplanması ve ikinci adım olarak p_i yüklerinin hesaplanmasıdır. Hesaplanan m tane faktör için aşağıda eşitlikler kullanılmaktadır.

$$T_m = X R_m \quad (4.19)$$

$$P_m = X' T_m (T_m' T_m)^{-1} \quad (4.20)$$

$$R_m = W_m (P_m' W_m)^{-1} \quad (4.21)$$

X matrisindeki varyansın önemli bir kısmı ilk m faktör tarafından açıklamaktadır. Doğrusal dönüşümler kullanılarak Eş.4.20'de ağırlık vektörlerinin iki seti arasında bağlantı kurmaktadır Eş.4.18 ve Eş.4.19'da $P_m' R_m = I_m$ 'ye eşit olduğu Eş.4.22'de gösterilmiştir.

$$R_m' P_m = R_m' X' T_m (T_m' T_m)^{-1} = T_m' T_m (T_m' T_m)^{-1} = I_m \quad (4.22)$$

m kadar faktör için KEKK kullanılarak hesaplanan T_m değerler vektörünü gösterebilir. Buna göre,

$$\hat{y}_{pls}^m = T_m (T_m' T_m)^{-1} T_m' y \quad (4.23)$$

Eşitliği elde edilir. Tek değişkenli durum için yazılan Eş.4.23'de çok değişkenli bir model için \hat{y}_{pls}^m yerine \hat{Y}_{pls}^m yazılarak elde edilmektedir (Bulut, 2011). Eş.4.23'de verilen $T_m = XR_m$ ve $y = X\hat{\beta}_{ols}$ yazıldığında, eşitlik aşağıdaki gibi olur (Phatak ve Jong, 1997).

$$\hat{y}_{pls}^m = XR_m (R_m' X' XR_m)^{-1} R_m' X' X\hat{\beta}_{ols} \quad (4.24)$$

Eş.4.24 aynı zamanda aşağıdaki eşitliğe denktir.

$$\hat{\beta}_{pls}^m = R_m (R_m' X' XR_m)^{-1} R_m' X' X\hat{\beta}_{ols} \quad (4.25)$$

Eş.4.19, Eş.4.20'de yerine koyulduğunda elde edilen eşitlik aşağıdaki gibi yazılabilir.

$$P_m = X' XR_m (R_m' X' XR_m)^{-1} \quad (4.26)$$

P_m için yazılan eşitliğin Eş.4.26'da yerine koyulduğunda $\hat{\beta}_{pls}^m$ için aşağıdaki eşitlik yazılır.

$$\hat{\beta}_{pls}^m = R_m (R_m' X' XR_m)^{-1} R_m' X' X\hat{\beta}_{ols}$$

$$\hat{\beta}_{pls}^m = R_m P_m' \hat{\beta}_{ols} \quad (4.27)$$

$$\hat{\beta}_{pls}^m = W_m (P_m' W_m)^{-1} P_m' \hat{\beta}_{ols}$$

Çok değişkenli regresyonda KEKK tahmini, \hat{B}_{pls}^m yukarıdaki eşitliğe benzerdir. Aradaki tek fark, $\hat{\beta}_{pls}^m$ yerine \hat{B}_{pls}^m gelmesi ile açıklanabilir. Çok değişkenli regresyon için KEKK tahmini Eş.4.28'de verildiği gibidir (Phatak ve Jong, 1997).

$$\hat{B}_{pls}^m = W_m (P_m' W_m)^{-1} P_m' \hat{B}_{ols} \quad (4.28)$$

NIPALS algoritmasında ikinci bir yaklaşım, KEKK faktörlerinin orijinal değişken olarak kullanılmasıdır. Bu durumu açıklamak için aşağıdaki tanımlamalar yapılmaktadır.

$$t_a = X_{a-1}w_a, (a=1,2,\dots,A) \quad (4.29)$$

Eş.4.29'da w_a ($a=2,3,\dots,A$) ağırlık matrisleridir ve her bir matrisler faktör skorlarını tanımlamak için kullanılmaktadır. Aynı zamanda X_a indirgenmiş matristir ve bu matriste indirgeme işlemi devam ettiği sürece bilgi kaybı meydana gelmektedir. Bunun bir sonucu olarak faktörlerin yorumlanması daha zor hale gelmektedir. Zira ilk faktörde bazı X değişkenleri kullanılıyorken sonradan gelenlerde ise başka faktörler kullanılmaktadır. Ancak p_a yük vektörleri, faktörler ile değişkenler arasındaki ilişkiyi iyi bir şekilde açıklayabilir. Bu nedenle orijinal merkezleştirilmiş veri matrisi X_0 yerine t_a yük vektörlerinin ifade edilmesi durumunda aşağıdaki eşitlik verilir.

$$t_a = X_0r_a, (a=1,\dots,A) \quad (4.30)$$

Eğer ağırlık vektörlerinin matrisi $R \equiv [r_1, r_2, \dots, r_A]$ ifade edilirse eşitlik aşağıdaki gibi yazılır.

$$T = X_0R \quad (4.31)$$

İndirgenmiş X matrisleri yardımıyla hesaplanan T faktör skorları, standartlaştırılmış X matrislerinin doğrusal kombinasyonu olarak ifade edilebilir. İndirgenmiş X_a matrislerinin tamamı ve t_a faktör skorları X_0 'ın sütun uzayında yer almaktadır. Böylece R matrisi aşağıdaki gibi verilir.

$$R = X_0^+T = (X_0'X_0)^- X_0'T = (X_0'X_0)^- P(T'T)^{-1} \quad (4.32)$$

Eş.4.32 de $P = (p_1, p_2, \dots, p_a)$ faktör yüklerine ait matrisi, (-) genelleştirilmiş tersi ve (+) ise tek Moore-Penrose tersini göstermektedir. R matrisi P' matrisinin genelleştirilmiş inversi olması nedeni ile R matrisi aşağıdaki gibi yazılabilir.

$$R = W(P'W)^{-1}$$

(4.33)

Höskuldsson KEKK-NIPALS algoritması için aşağıdaki tekrarlayan ilişkiyi sunmuştur (Höskuldsson, 1988).

$$r_a = G_a w_a \quad a = 1, 2, \dots, A \quad (4.34)$$

$$G_{a+1} = G_a - r_a p_a' \quad a = 1, 2, \dots, A-1 \quad (4.35)$$

Burada $G_1 = I_p$ olarak işleme başlanmaktadır. Bununla birlikte r_a vektörü standardize edilmemiştir. Dolayısıyla Eş.4.30 daki t_a ve r_a vektörleri de normleştirilmemiştir (De Jong, 1993).

3.2.4.3. SIMPLS algoritması

De Jong (1993) tarafından önerilen ve kısaca SIMPLS olarak isimlendirilen bu algorithmada, skorları indirgenmiş X matrisleri yerine orijinal X matrislerinin kombinasyonları kullanılarak hesaplama yapılmaktadır. Bu nedenle bu algoritma ile elde edilen sonuçlar, klasik algoritma kullanılarak elde edilen sonuçlardan farklı olmaktadır. Ancak, bahsedilen bu farklılık, çok küçük miktarda olup genel olarak dikkate alınmamaktadır. SIMPLS algoritmasında önceki bileşenlerden elde edilen yük vektörleri üzerine orijinal $X'Y$ matrisi yansıtılarak özvektör analizi yapılmaktadır. İndirgeme işlemindeki ortaya çıkan bu farklılık SIMPLS algoritmasının diğer algoritmalarından farklı olmasına ve dolayısıyla çok küçük de olsa sonuçların da farklı olmasına neden olmaktadır (Lindgren ve Rännar, 1998).

SIMPLS yaklaşımı çok değişkenli istatistikî yöntemlerde kullanılmakta olan klasik bir yaklaşımdır. Bu algorithmada öncelikle amaç optimizasyon kriteri belirlenir ve son olarak algoritma oluşturulur.

X 'in ardışık ortogonal faktörlerini hesaplamak için yük vektörleri $t_a = X_0 r_a$ olsun. Bu yük vektörleri Y 'nin $u_a = Y_0 q_a$ şeklindeki faktörleri ile arasındaki kovaryans maksimum olacak şekilde hesaplanır. Önerilen bu düzeltmede, R ağırlık matrisi direk olarak hesaplanmaktadır. Böylece hesaplama işlemlerinde kolaylık sağlanmaktadır.

Ağırlık vektörleri r_a ve q_a merkezileştirilmiş veri kullanılarak doğrudan aşağıdaki gibi hesaplanmaktadır (De Jong, 1993).

$$t_a = X_0 r_a, \dots a = 1, 2, \dots, A \quad (4.36)$$

$$u_a = Y_0 q_a, a = 1, 2, \dots, A \quad (4.37)$$

Ağırlık vektörlerinin belirlenmesi bazı varsayımların sağlanmasını gerektirir. Aşağıda verilmiş olan koşulların sağlanması önemlidir. Özellikle aşağıda verilen 4 koşul çözümü kontrol etmektedir:

- 1) Kovaryansın (t_a ve u_a arasındaki) maksimize edilebilmesi gerekir,

$$u_a' t_a = q_a' (Y_0' X_0) r_a = \max$$

- 2) r_a Ağırlık vektörünün normalleştirilmesi gerekir, $r_a' r_a = 1$

- 3) q_a Ağırlık vektörünün normalleştirilmesi gerekir, $q_a' q_a = 1$

- 4) t skorlarının ortogonalize edilmiş olması gerekir, $t_b' t_a = 0$ $a > b$ koşulu ile

$a > b$ için aşağıda verilen eşitlik kullanılmaktadır.

$$t_b' t_a = t_b' X_0 r_a = (t_b' t_a) P_b' r_a = 0 \quad (4.38)$$

Eş.4.38'de p_b orijinal X değişkenleri ile b'inci KEKK faktörü arasındaki ilişkiyi açıklayan yük vektörüdür. Eş.4.38'de herhangi yeni r_a ağırlık vektörü önceki bütün yük vektörlerine dik olmak zorundadır.

KEKK çözümleri için temel adımlar basit olarak aşağıda verildiği gibidir (De Jong, 1993):

- Çapraz çarpımı hesaplanır: $S = X_0' Y_0$ $a=1, 2, \dots, A$ için;
- $a=1$ için S matrisinin tekil değer ayrışımını (SVD) hesaplanır.
- $a>1$ için $S - P(P'P)^{-1}P'S$ matrisinin tekil değer ayrışımını (SVD) hesaplanır.
- Ağırlıklar bulunur: r =ilk sol singüler vektör.
- Skorlar hesaplanır. $t = X_0 r$.

- Yükle hesaplanır. $p = X_0' t / (t' t)$.
- r , t ve p 'yi sırası ile R , T , P içine depolanır.
- Sonlandırılır.
- Regresyon katsayıları hesaplanır: $B_{pls} = RT^{-1} Y_0$

SIMPLS algoritması ile standart KEKK algoritması arasındaki temel farklılık indirgeme işleminin yapılış şeklidir. SIMPLS yönteminde indirgeme işlemi daha büyük boyutlu X_0 ve Y_0 matrislerine değil çapraz çarpımlar matrisi olan S_0 'a uygulanmaktadır.

SIMPLS algoritmasının standart PLS algoritmasından daha basit bir şekilde uyguladığı diğer bir durum X ve Y matrislerinin merkezleştirilmesidir. Standart indirgeme işlemindeki ilk adım olan merkezleştirme işlemini burada uygulamak önemli değildir.

Daha iyi bir yorum yapılabilmesi amacıyla standart KEKK algoritması üzerinde bazı değişiklikler yapılmıştır. Geliştirilen yeni algoritmaya SIMPLS ismi verilmiştir (De Jong, 1993).

SIMPLS algoritması için aşamalar aşağıda sıralandığı gibidir.

- 1) $Y_0 = Y - ort(y)$
- 2) $S = X' Y_0$ $a = 1, 2, \dots, A$
- 3) $q = S' S$ Matrisinin baskın vektörü
- 4) $r = S q$
- 5) $t = X r$
- 6) $t = t - ort(t)$
- 7) $normt = \sqrt{t' t}$
- 8) $t = t / normt$
- 9) $r = r / normt$
- 10) $p = X' t$
- 11) $q = Y_0' t$
- 12) $u = Y_0 q$
- 13) $v = p$ Eğer $a > 1$ ise

- 14) $v = v - V(V'p)$
- 15) $u = u - T(T'u)$ Sonlanır.
- 16) $v = v / \sqrt{v'v}$
- 17) $S = S - v(v'S)$
- 18) r, t, p ve p'yi R, T, P Q U ve P içine kaydet ve sonlandır.
- 19) $B = RQ'$
- 20) $h = \text{diag}(TT') + 1/n$
- 21) $\text{var}(X) = \text{diag}(PP')/n - 1$
- 22) $\text{var}(X) = \text{diag}(PP')/n - 1$

3.2.4.4. Standart KEKK algoritması ile SIMPLS algoritmasının karşılaştırılması

Hem SIMPLS ve hem de standart KEKK algoritmalarında kovaryans maksimize edilmektedir. İlk bakışta benzer gibi görünen bu algoritmalarından NIPALS-KEKK kullanıldığında, veri matrisi $X_A = (I_n - TT')X_0$ matrisine ve SIMPLS algoritması kullanıldığında dolaylı olarak $X_0(I_P - VV')$ matrislerine indirgenmektedir. Bununla birlikte, SIMPLS algoritmasında faktörler direkt olarak orijinal veri matrisi kullanılarak hesaplanmaktadır. SIMPLS algoritmasında R ağırlıkları ters matrislerin hesaplanmasını gerektirmez ve bu ağırlıkların yorumu KEKK algoritmasında hesaplanan w ağırlıklarının yorumundan daha basittir. Aynı zamanda SIMPLS algoritması standart KEKK algoritmasına göre daha hızlıdır. Bunun nedeni, SIMPLS algoritmasında X matrisine herhangi bir indirgenme işlemi yapılmamasıdır (De Jong, 1993).

Cevap değişken sayısının bir ($m=1$) olması bir diğer ifade ile cevap değişken matrisi tek boyutlu olması durumunda, her iki algoritmanın benzer sonuçlar üreteceği bildirilmiştir. Ancak cevap değişken matrisinin çok olması durumunda bu durum geçerli olmamakta, sonuçlar birbirlerinden farklı olmaktadır (De Jong, 1993).



4. BULGULAR

Bu çalışmada kıl keçilerine ait canlı ağırlıklar (CA) bağımlı değişken, cidago yüksekliği (CY), sağrı yüksekliği (SAY), sırt yüksekliği (SIY), vücut uzunluğu (VU), göğüs derinliği (GD), göğüs genişliği (GG), göğüs çevresi (GC) ise modelde bağımsız değişkenler olarak yer alan değişkenlerdir. Çalışmada kullanılan veri setine ait tanımlayıcı istatistikler Çizelge 4.1’de verilmiştir.

Çizelge 4.1. Çalışmada kullanılan değişkenlere ilişkin tanımlayıcı istatistikler

| | n | \bar{X} | S | $S_{\bar{X}}$ | CV | Minmum | Maksimum |
|-----|-----|-----------|------|---------------|-------|--------|----------|
| CA | 119 | 57.34 | 9.85 | 0.90 | 17.17 | 38.00 | 86.45 |
| GG | 119 | 20.08 | 1.96 | 0.18 | 9.77 | 15.00 | 25.00 |
| SAY | 119 | 78.82 | 4.68 | 0.43 | 5.94 | 64.00 | 89.00 |
| CY | 119 | 81.28 | 5.00 | 0.46 | 6.16 | 67.00 | 95.00 |
| SIY | 119 | 76.71 | 4.70 | 0.43 | 6.13 | 62.00 | 88.00 |
| GD | 119 | 33.84 | 2.13 | 0.20 | 6.29 | 28.00 | 39.00 |
| GC | 119 | 89.57 | 5.75 | 0.53 | 6.42 | 75.30 | 106.00 |
| VU | 119 | 75.23 | 4.07 | 0.37 | 5.41 | 65.00 | 86.00 |

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

Çizelge 4.1 sırasıyla gözlem sayısı, ortalama, standart sapma, standart hata, varyasyon katsayısı en küçük ve en büyük değerleri içermektedir. Ortalama ve standart hatalar CA, GG, SAY, CY, SIY, GD, GC ve VU için sırasıyla 57.34 ± 0.90 , 20.08 ± 0.18 , 78.82 ± 0.43 , 81.28 ± 0.46 , 76.71 ± 0.43 , 33.84 ± 0.20 , 89.57 ± 0.53 ve 75.23 ± 0.37 şeklinde elde edilmiştir. Kullanılan değişkenler için varyasyon katsayısı %5.41 ila %17.17 arasında değerler almıştır. Veri setinde kullanılan değişkenler arasındaki korelasyonlar Çizelge 4.2’de verilmiştir.

Çizelge 4.2. Bağımlı ve bağımsız değişkenler arasındaki korelasyonlar ve önem düzeyleri

| | CA | GG | SAY | CY | SIY | GD | GC | VU |
|-----|--------------|--------------|---------------------|---------------------|--------------|---------------------|--------------|----|
| CA | 1 | | | | | | | |
| GG | 0.543 *** | 1 | | | | | | |
| SAY | 0.455 *** | 0.318 *** | 1 | | | | | |
| CY | 0.474 *** | 0.309 *** | 0.955 *** | 1 | | | | |
| SIY | 0.428 *** | 0.311 *** | 0.980 *** | 0.949 *** | 1 | | | |
| GD | 0.815 *** | 0.458 *** | 0.466 *** | 0.475 *** | 0.443 *** | 1 | | |
| GC | 0.883 *** | 0.591 *** | 0.484 *** | 0.495 *** | 0.474 *** | 0.854 *** | 1 | |
| VU | 0.722 *** | 0.328 *** | 0.497 *** | 0.546 *** | 0.454 *** | 0.638 *** | 0.617 *** | 1 |

Çizelge 4.2 incelendiğinde, özellikle CY ile SAY, SIY ile SAY ve CY, CG ve GD değişkenleri arasındaki korelasyonların çok yüksek olduğu dikkat çekmektedir. Bu durum, çalışma verisinde çoklu bağlantı probleminin varlığının araştırılması gerektiğini işaret etmektedir. Veri setinde bu problemin varlığını araştırmak amacıyla, EKK yöntemi kullanılarak çoklu doğrusal regresyon analizi yapılmıştır. Regresyon katsayıları, standart hata, t ve p, TV ve VIF değerleri ile birlikte model uyum kriterleri Çizelge 4.3'te verilmiştir. Çoklu bağlantı belirleyicilerinden özdeğerler ve CI değerleri Çizelge 4.4 ve varyans oranları Çizelge 4.5'te sunulmuştur.

Çizelge 4.3. Kıl keçilerinde CA için EKK yöntemi kullanılarak elde edilen çoklu doğrusal regresyon analiz sonuçları

| Bağımsız Değişkenler | Regresyon Katsayıları | Standart Hata | t | p | TV | VIF |
|----------------------|-----------------------|---------------|--------|--------|--------|---------|
| Sabit | -101.8811 | 7.9491 | -12.82 | <.0001 | . | 0.0000 |
| GG | 0.2849 | 0.2396 | 1.19 | 0.2369 | 0.6407 | 1.5608 |
| SAY | 0.2969 | 0.4430 | 0.67 | 0.5041 | 0.0329 | 30.4096 |
| CY | -0.0400 | 0.2764 | -0.14 | 0.8853 | 0.0741 | 13.5018 |
| SIY | -0.3579 | 0.4308 | -0.83 | 0.4079 | 0.0346 | 28.9436 |
| GD | 0.5644 | 0.3555 | 1.59 | 0.1152 | 0.2476 | 4.0382 |
| GC | 1.0232 | 0.1423 | 7.19 | <.0001 | 0.2113 | 4.7328 |
| VU | 0.6654 | 0.1361 | 4.89 | <.0001 | 0.4611 | 2.1689 |
| Uyum Kriteri | | | | | | |
| R^2 | 0.8379 | | | | | |
| R_d^2 | 0.8277 | | | | | |
| HKO | 16.71093 | | | | | |
| \sqrt{HKO} | 4.08790 | | | | | |

TV: Tolerans değeri, VIF: Varyans şişirme faktörü, R^2 : Belirleme katsayısı, R_d^2 : Düzeltilmiş belirleme katsayısı, HKO: Hata Kareler Ortalaması.

Çizelge 4.3'de sabit dışında modelde yer alan yedi açıklayıcı değişkenden sadece GC ve VU anlamlı ($P < 0.001$) bulunduğu görülmektedir. Buna rağmen aynı çizelgede verilen R^2 ve R_d^2 değerlerinin yüksek olması çoklu bağlantı varlığını destekleyen bulgular olmuştur. Çizelgede çoklu bağlantı belirleyicilerden olan VIF bakımından incelendiğinde, 10 değerinin üzerinde VIF'e sahip değişkenler olduğu görülmektedir. Bunlar SAY, CY ve SIY değişkenleri olup, sırasıyla SAY için 30.4096, CY için 13.5018 ve SIY için 28.9436 olarak bulunmuştur. Bu değişkenlere karşılık gelen TV değerlerinin sırasıyla 0.0329, 0.0741, 0.0346 olduğu Çizelge 4.3'te görülmektedir. $VIF \geq 10$ ve TV'nin sıfıra yakın değerler alması, üzerinde çalışılan veri setinin çoklu bağlantı sorununa sahip olduğunu işaret eden diğer iki kriterdir.

Çizelge 4.4. CA için çoklu bağlantı belirleyicilerinden Özdeğer (OD) ve CI değerleri

| No | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|----|---------|---------|----------|---------|---------|---------|---------|---------|
| OD | 7.98458 | 0.0073 | 0.00389 | 0.00223 | 0.00121 | 0.00053 | 0.00019 | 6.7E-05 |
| CI | 1 | 33.0691 | 45.28165 | 59.8668 | 81.3417 | 122.244 | 206.034 | 344.018 |

OD: Öz değerler, CI: Koşul İndeksi

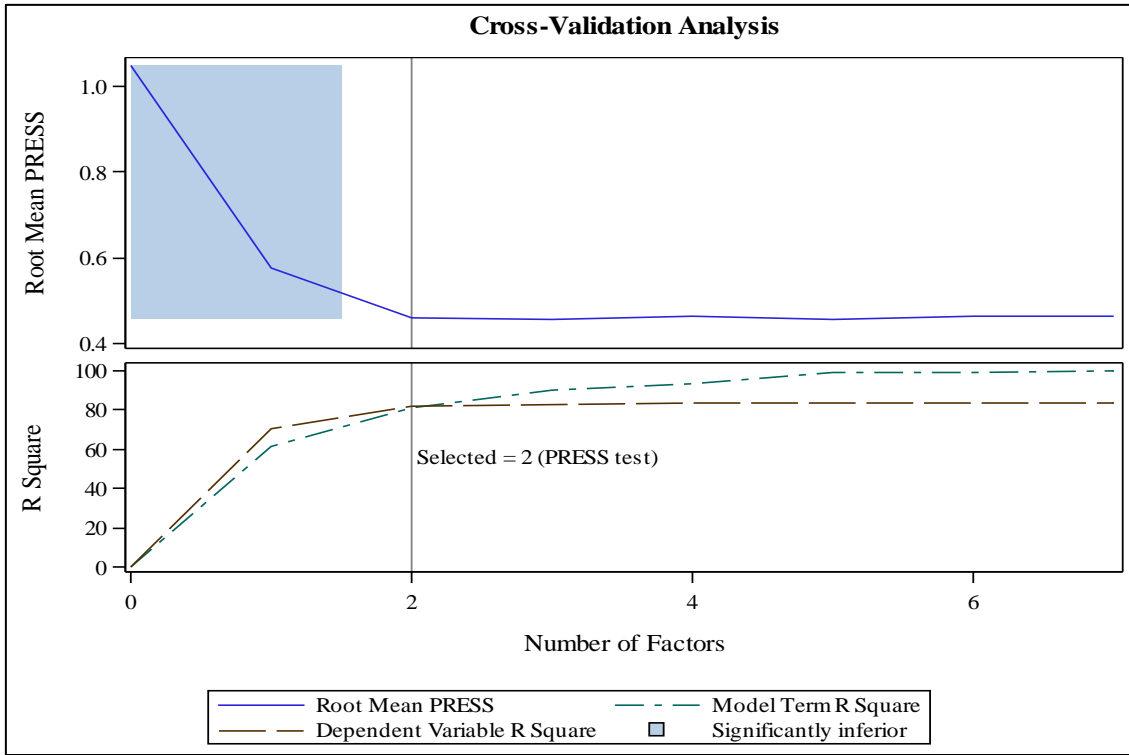
Çizelge 4.5. CA için çoklu bağlantı belirleyicilerinden varyans oranları

| No | Varyans Oranları | | | | | | | |
|----|------------------|---------|----------|---------|---------|---------|---------|---------|
| | Sabit | GG | SAY | CY | SIY | GD | GC | VU |
| 1 | 3.479E-05 | 9.4E-05 | 1.8E-06 | 4.3E-06 | 2E-06 | 1.5E-05 | 1.3E-05 | 2.1E-05 |
| 2 | 0.0044 | 0.54854 | 0.00136 | 0.00345 | 0.00162 | 0.00126 | 0.00402 | 0.00109 |
| 3 | 0.03656 | 0.26761 | 0.003 | 0.00644 | 0.00423 | 0.0507 | 0.02375 | 0.04467 |
| 4 | 0.65008 | 0.05419 | 0.000412 | 0.00196 | 0.00063 | 0.0616 | 0.0531 | 0.00844 |
| 5 | 0.25793 | 0.04037 | 0.00016 | 0.00209 | 0.00129 | 0.04986 | 0.04441 | 0.77972 |
| 6 | 0.00293 | 0.08662 | 0.000433 | 2.4E-06 | 6.2E-05 | 0.82456 | 0.85676 | 0.00354 |
| 7 | 0.04776 | 0.00125 | 0.08625 | 0.98532 | 0.11519 | 0.00173 | 0.00049 | 0.11829 |
| 8 | 0.0003013 | 0.0013 | 0.90838 | 0.00073 | 0.87698 | 0.01028 | 0.01745 | 0.04422 |

Çoklu bağlantının varlığını belirlemede kullanılan diğer iki belirleyici, özdeğerler ve CI değerlidir. Çizelge 4.4 incelendiğinde özdeğerlerin çok küçük olduğu ve CI değerlerinin 30'un üzerinde olduğu görülmektedir. Bu iki kriter dikkate alındığında veri setinin çoklu bağlantı problemine sahip olduğu görülmektedir. Çizelge 4.5'te verilen varyans oranları dikkate alındığında SAY ve SIY değişkenlerine ait varyans oranlarının sırasıyla 0.90838, 0.87698 olduğu görülmektedir. Bu değerlerin

yüksek olması, veri kümesinde çoklu bağlantının var olduğunu gösteren bir diğer belirleyici olarak değerlendirilmiştir.

Veri setinde çoklu bağlantı sorunu tespit edildikten sonra çoklu doğrusal regresyon analizi, KEKK yöntemi kullanılarak yapılmıştır. KEKK regresyon analizi yapılırken öncelikle SIMPLS algoritması ve daha sonra NIPALS algoritması kullanılmıştır. Her iki algoritma çalıştırıldığında 10 katmanlı çapraz doğrulama ile elde edilen sonuçlar birbirinin aynısı olmuştur. Sekiz faktör kadar seçimin yapılabileceği hem SIMPLS hem de NIPALS algoritmaları kullanıldığında optimum gizil faktör sayısının seçimine ilişkin görsel Şekil 4.1’de sunulmuştur.



Şekil 4.1. Çarpaz doğrulama yapılarak elde edilen SIMPLS ve NIPALS algoritmasının gizil faktör seçiminin görseli.

Şekil 4.1 incelendiğinde yedi açıklayıcı değişkenin yer aldığı model için çarpaz doğrulama sonunda iki gizil faktör seçilmiştir. Buna karar verilirken HKO’nun karekökü, cevap değişkeni ve açıklayıcı değişkenlerdeki değişimi ayrı ayrı ifade eden belirleme katsayıları dikkate alınmıştır. Bu üç kriterin optimum koşullarda bulunduğu noktada KEKK regresyon analizinde seçilecek gizil faktör sayısına karar verilmiştir. Buna göre hem SIMPLS hem de NIPALS algoritmaları kullanıldığında gizil faktör

sayısı iki olarak belirlenmiştir. Modele ait HKO, HKO'nun karekökü, cevap ve açıklayıcı değişkenlerdeki değişimi açıklamak üzere kullanılan belirleme katsayıları ve gizil faktör sayısına göre belirleme katsayılarındaki değişim Çizelge 4.6'da verilmiştir.

Çizelge 4.6. SIMPLS ve NIPALS algoritmaları için gizil faktör sayıları ve model uyum kriterleri

| Gizil faktör sayısı | HKO | \sqrt{HKO} | SIMPLS | | | |
|---------------------|--------|--------------|---------------------------|---------|---------------------------|---------|
| | | | Model Etkisi | | Bağımlı Değişken Etkisi | |
| | | | Gizil Faktör için geçerli | Toplam | Gizil Faktör için geçerli | Toplam |
| 1 | 0.3317 | 0.5759 | 61.2385 | 61.2385 | 70.7149 | 70.7149 |
| 2 | 0.2132 | 0.4617 | 20.2772 | 81.5158 | 11.3844 | 82.0993 |

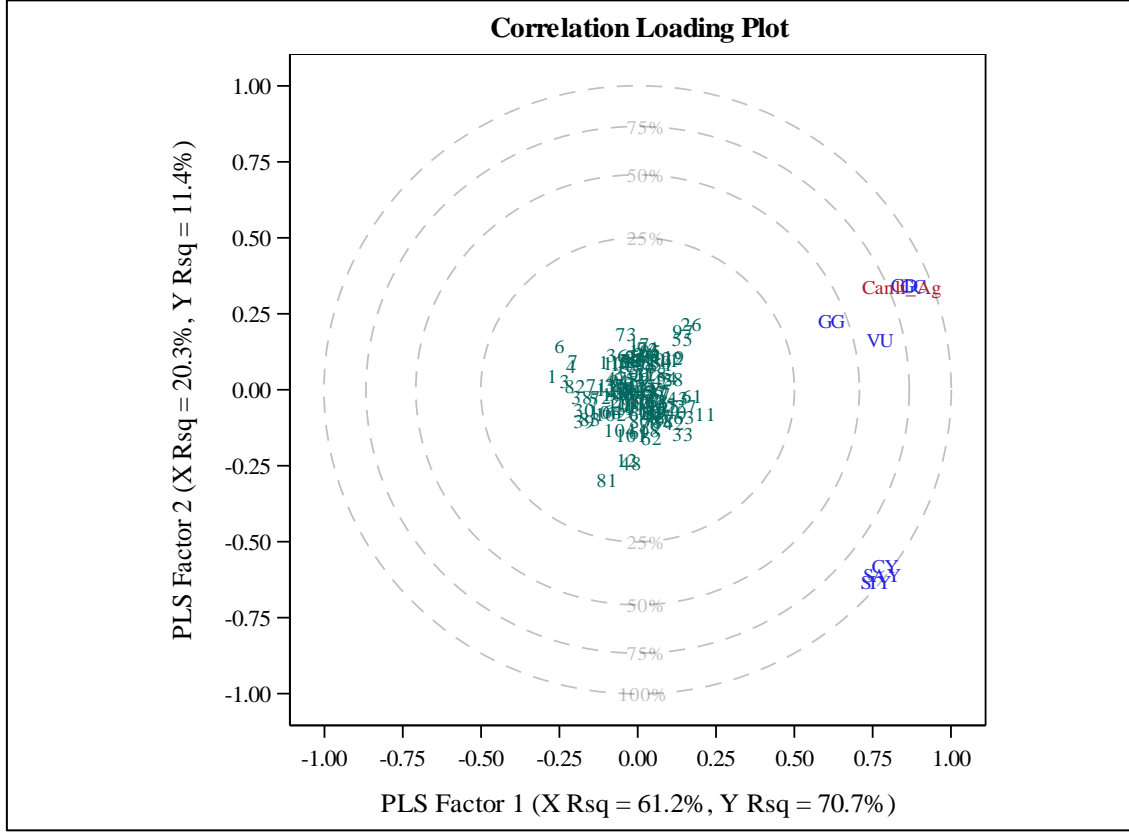
| Gizil Faktör Sayısı | HKO | \sqrt{HKO} | NIPALS | | | |
|---------------------|--------|--------------|---------------------------|---------|---------------------------|---------|
| | | | Model Etkisi | | Bağımlı Değişken Etkisi | |
| | | | Gizil Faktör için geçerli | Toplam | Gizil Faktör için geçerli | Toplam |
| 1 | 0.3317 | 0.5759 | 61.2385 | 61.2385 | 70.7149 | 70.7149 |
| 2 | 0.2132 | 0.4617 | 20.2772 | 81.5158 | 11.3844 | 82.0993 |

HKO: Hata kareler ortalaması, \sqrt{HKO} : Hata kareler ortalamasının karekökü

Çizelge 4.6 incelendiğinde, gizil faktör sayısı bir iken HKO'nun 0.3317 olduğu, gizil faktör sayısı iki olduğunda ise 0.2132'ye düştüğü görülmektedir. Modelde yer alan açıklayıcı değişkenlerdeki açıklama miktarı bir gizil faktörün olduğu model için % 61.2385 iken iki gizil faktörün olduğu model için bu miktar % 20.2772'lik artışla toplamda % 81.5185'e ulaşmıştır. Bağımlı değişkendeki açıklama miktarı da gizil faktör sayısındaki artışla birlikte % 70.7149'dan % 82.0993'e çıkmıştır. İkinci gizil faktör ile meydana gelen açıklama miktarı % 11.3844 olmuştur.

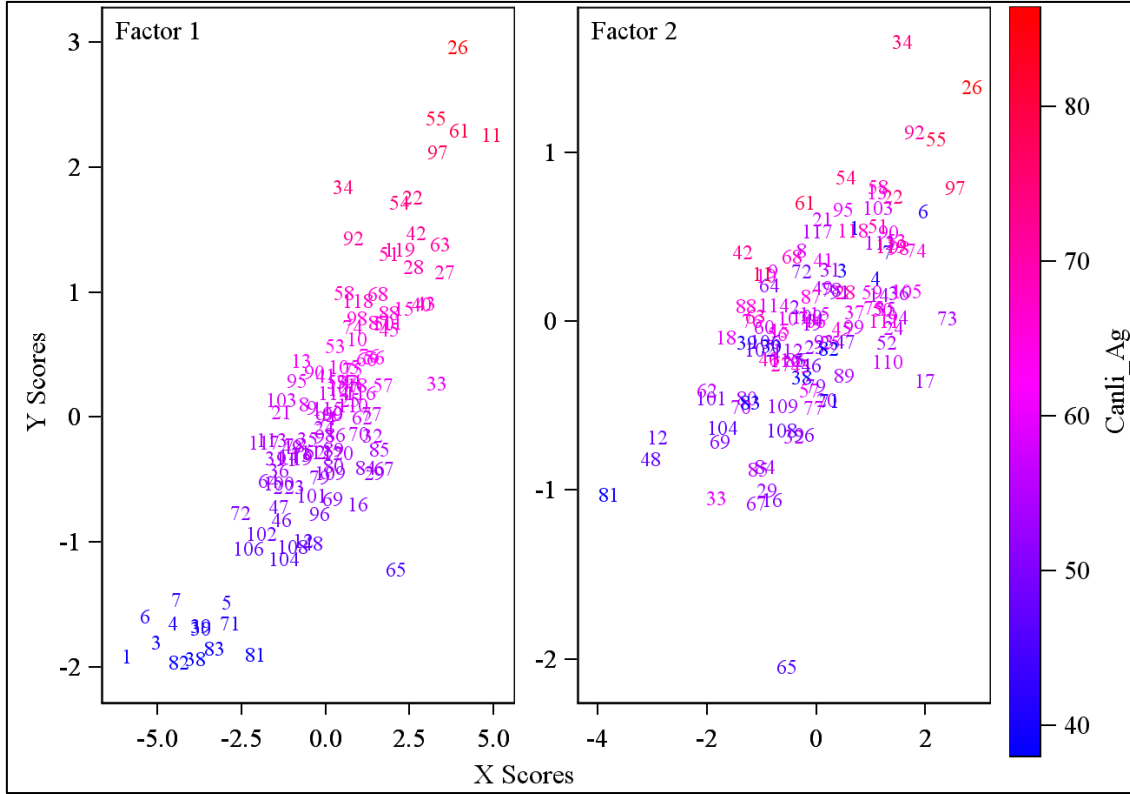
Optimum gizil faktör sayısı iki olan KEKK regresyon analizinin çıktılarından olan korelasyon yük grafiği Şekil 4.2'de verilmiştir. Bu grafikte her bir X puanı numaralandırılarak, bağımlı ve bağımsız tüm değişkenler hem gizil faktörler ile açıklanma miktarlarına göre hem de birbirleriyle aralarında olan ilişkiye göre işaretlenerek görselde yer almıştır. Buna göre numaralarda herhangi bir gruplama olmaması, modele bir sınıflandırma etkisi dâhil etmenin gerekli olmadığını ortaya koymuştur. Aynı zamanda grafikte CA'nın birinci gizil faktör için tüm değişkenler ile pozitif yönlü korelasyona sahip olduğu görülmektedir. Çizelge 4.2'de verilen değişkenler arasındaki korelasyonlar incelendiğinde bu bilgilerin örtüştüğü görülebilir.

Bununla birlikte, ilk iki gizil faktör tarafından değişkenlerdeki değişimin açıklanma miktarları yine aynı grafik üzerinde gösterilmiştir.



Şekil 4.2. SIMPLS ve NIPALS algoritmaları için korelasyon grafiği

SIMPLS ve NIPALS algoritmasına göre seçilen iki gizil faktör için bazı açıklayıcı bilgiler Şekil 4.3'te verilmiştir. Şekilde birinci gizil faktör için açıklayıcı değişken skorları ile cevap değişkeni skorları arasındaki korelasyonun en yüksek olduğu görülmektedir. Bu korelasyon miktarı ikinci gizil faktör için daha düşük olmuştur. Gizil faktör sayısının çoğalması ile bu korelasyon miktarı giderek azalacağı açıktır.

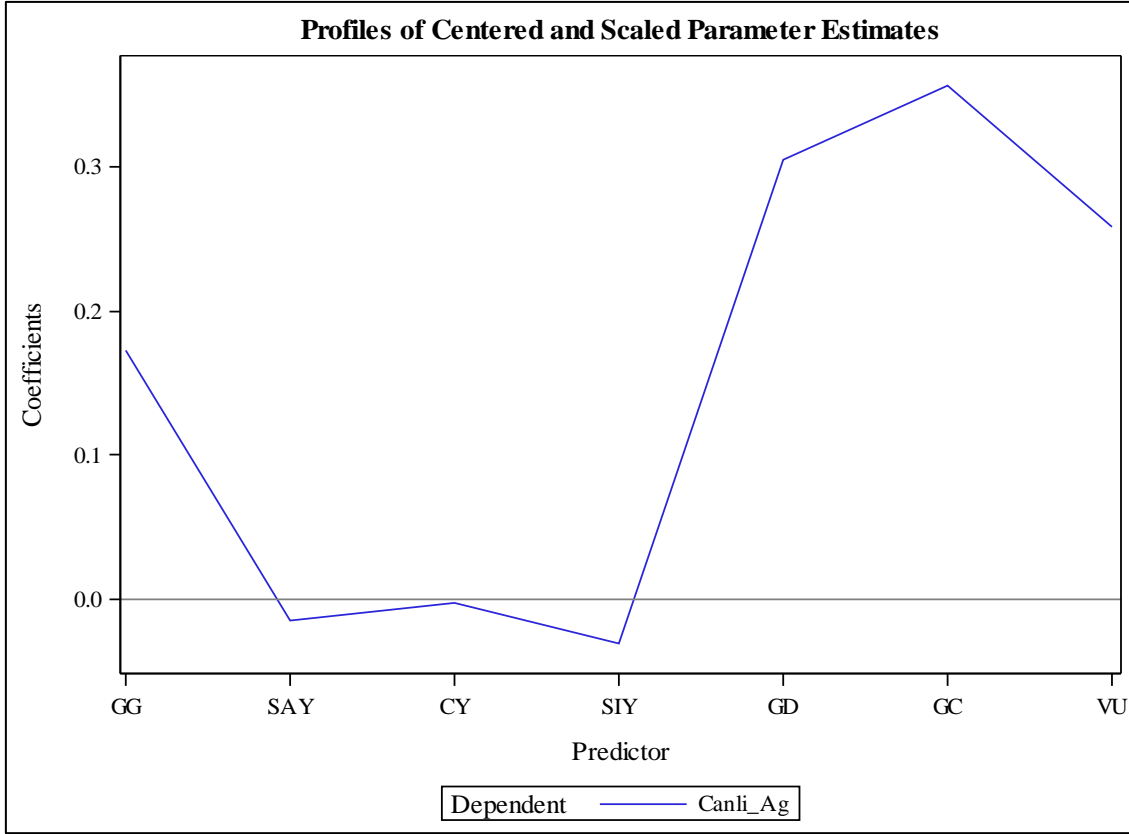


Şekil 4.3. KEKK regresyon analizinde (hem SIMPLS hem de NIPALS algoritmaları) iki gizil faktör için açıklayıcı değişkenlere karşı cevap değişkeninin grafiği

KEKK regresyon analizinde SIMPLS ve NIPALS algoritması kullanılarak tahmin edilen standardize edilmiş ve edilmemiş katsayılar ve standardize edilmiş katsayılara ilişkin profil grafiği sırasıyla Çizelge 4.7 ve Şekil 4.3'te verilmiştir.

Çizelge 4.7. SIMPLS ve NIPALS algoritması kullanılarak CA için elde edilen regresyon katsayıları

| Katsayılar | Sabit | GG | SAY | CY | SIY | GD | GC | VU |
|---------------|-----------|--------|---------|---------|---------|--------|--------|--------|
| Std edilmiş | 0.0000 | 0.1727 | -0.0148 | -0.0025 | -0.0308 | 0.3042 | 0.3561 | 0.2579 |
| Std edilmemiş | -101.4411 | 0.8665 | -0.0311 | -0.0049 | -0.0645 | 1.4083 | 0.6096 | 0.6237 |

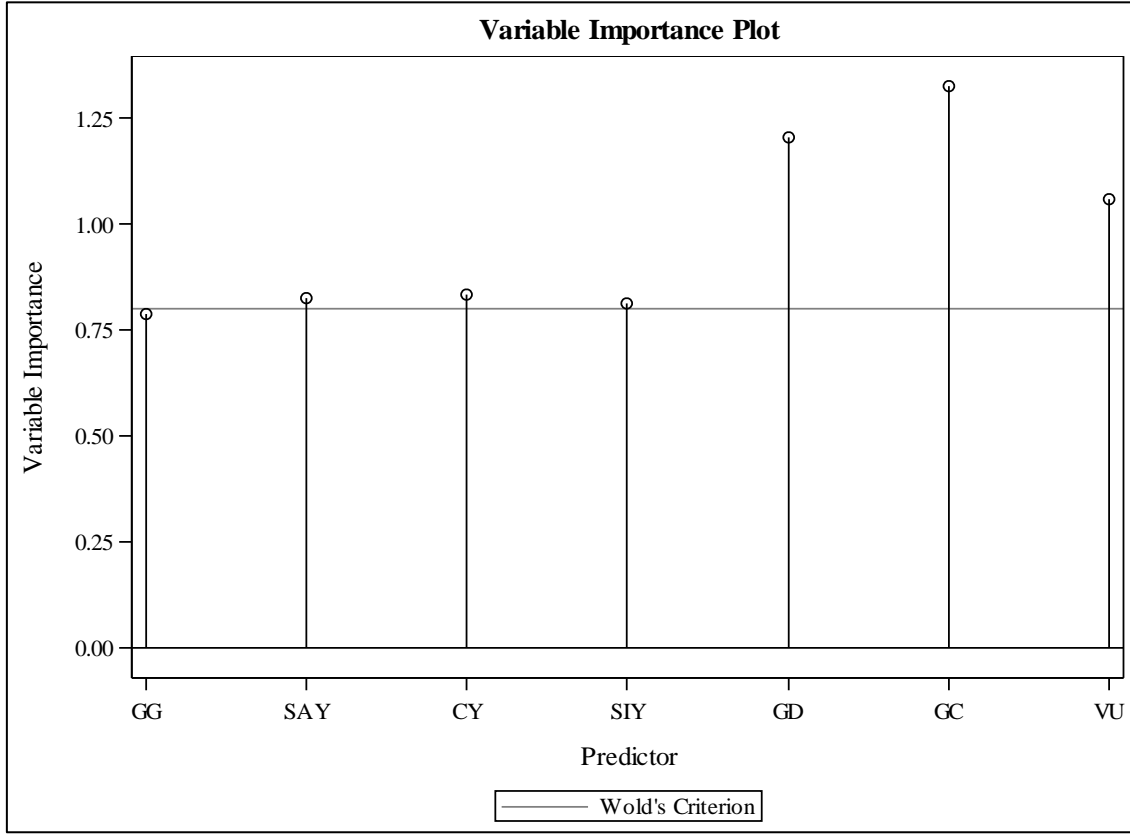


Şekil 4.3. SIMPLS ve NIPALS algoritmaları kullanılarak elde edilen standardize edilmiş regresyon katsayılarının profili

Çizelge 4.7 ve Şekil 4.3 birlikte değerlendirildiğinde CA'yı açıklamada etkisi en az olan değişkenler sırasıyla SAY, CY ve SIY olmuştur. SAY için standardize edilmiş katsayı -0.0148 CY için -0.0025 ve SIY için -0.0308 şeklinde tahmin edilmiştir. En düşük etki miktarı CY değişkenine ait olduğu söylenebilir. CY'de meydana gelen birim değişimin CA'da -0.0049 (standardize edilmemiş değer) kg'lık bir azlamaya neden olduğu söylenebilir. Etki miktarları büyüklüklerine göre sıralanacak olursa GG, VU, GD ve GC değişkenlerine ait olduğu görülmektedir. Bu etkiler GG, VU, GD ve GC için sırasıyla 0.1727, 0.2579, 0.3042 ve 0.3561 tahmin edilmiştir. Buna göre, CA'yı açıklamadaki etkisi en yüksek olan değişken GC olduğu ve GC'deki birim artışın, CA'da 0.3561 kg artış meydana getirdiği söylenebilir.

Etkisi düşük olan değişkenlerin değerlendirilmesi için KEKK regresyon analizinde bir prosedür uygulanmaktadır. Bu prosedür, bazı bağımsız değişkenlerin modelden çıkarılmasına olanak sağlamaktadır. Bunun için açıklayıcı değişkenler için Wold'un (1994) VIP (değişken önem grafiği) istatistiği kullanılmaktadır. Etkisi küçük

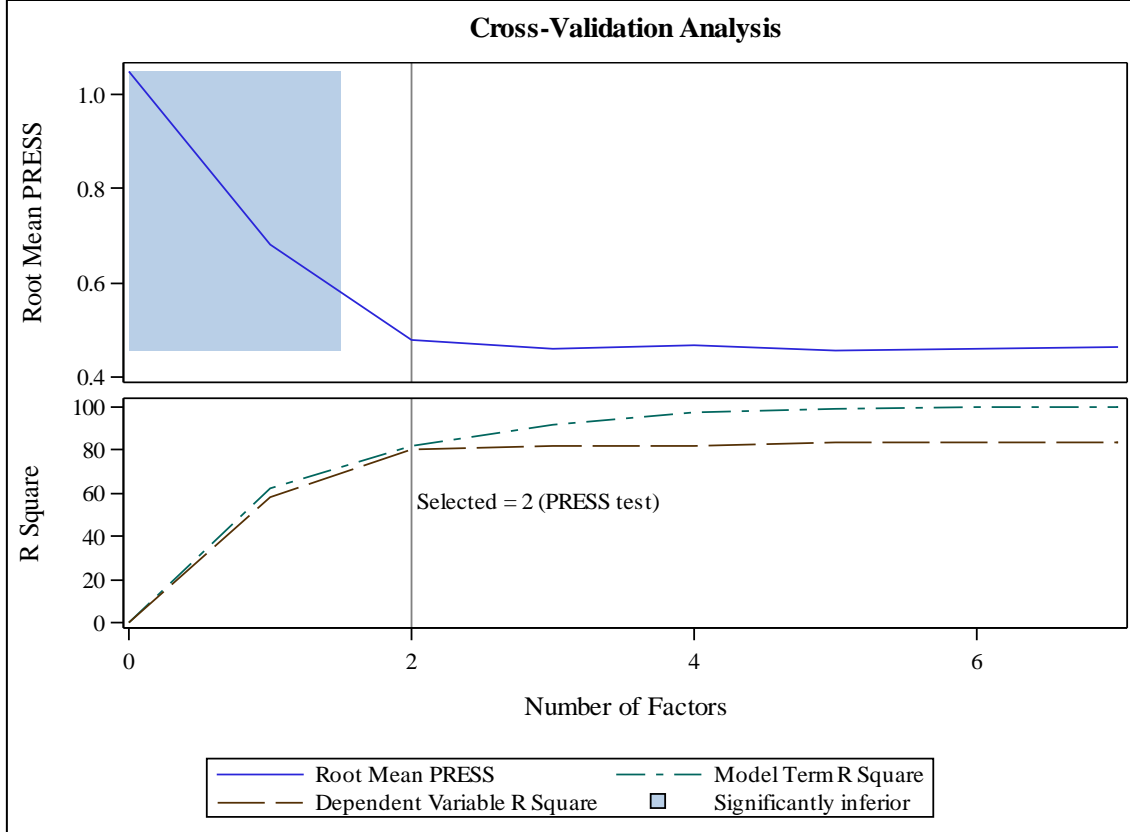
olan deęişkenlerin deęerlendirilmesi için VIP istatistięinden faydalanmak üzere, VIP istatistięini gsteren grsel Őekil 4.4'te verilmiřtir.



Őekil 4.3. SIMPLS ve NIPALS algoritmaları kullanılarak baęımsız deęişkenler için elde edilen VIP istatistikleri.

Őekil 4.4 incelendięinde 0.8 deęerinin altında olan tek bir deęişken dikkat çekmektedir. Bu deęişken GG'dir. GG dıřındaki deęişkenlerin tamamının bu sınırın üzerinde olduęu Őekil 4.4'de grlmektedir. Dolayısıyla GG, VIP deęeri ile standardize edilmiř regresyon katsayısı (0.1727) birlikte deęerlendirmeye alınmıřtır. GG'den daha kçük standardize edilmiř regresyon katsayısına sahip deęişkenler olduęu, Őizelge 4.7 ve Őekil 4.3'de grlmektedir. Dolayısıyla her ne kadar VIP deęeri kritik deęerin çok az altında olsa da GG deęişkeninden daha kçük standardize edilmiř deęerlere sahip deęişkenler olması nedeni ile GG'nin modelden ıkarılmayacaęına, bir dięer ifade modelde kalmasına karar verilmiřtir. Zira yedi aıklayıcı deęişkene sahip bu model için standardize edilmiř regresyon katsayısına bakıldıęında, GG'nin modeli aıklamada drdnc sırada nemli bir etkiye sahip olduęu grlmektedir (Őizelge 4.7).

Çalışmanın bir sonraki aşamasında veriler, TB regresyonu kullanılarak analiz edilmiştir. KEKK regresyona benzer şekilde, TB regresyon analizinde de 10 katmanlı çapraz doğrulama yapılmıştır. Bu yöntem ile optimum gizil faktör sayısı iki olarak belirlenmiştir. Yöntemin çalışmasını ve karar aşamasını gösteren görsel, Şekil 4.5'te verilmiştir.



Şekil 4.4. TB regresyon analizinde çapraz doğrulama yapılarak elde edilen gizil faktör seçiminin görseli.

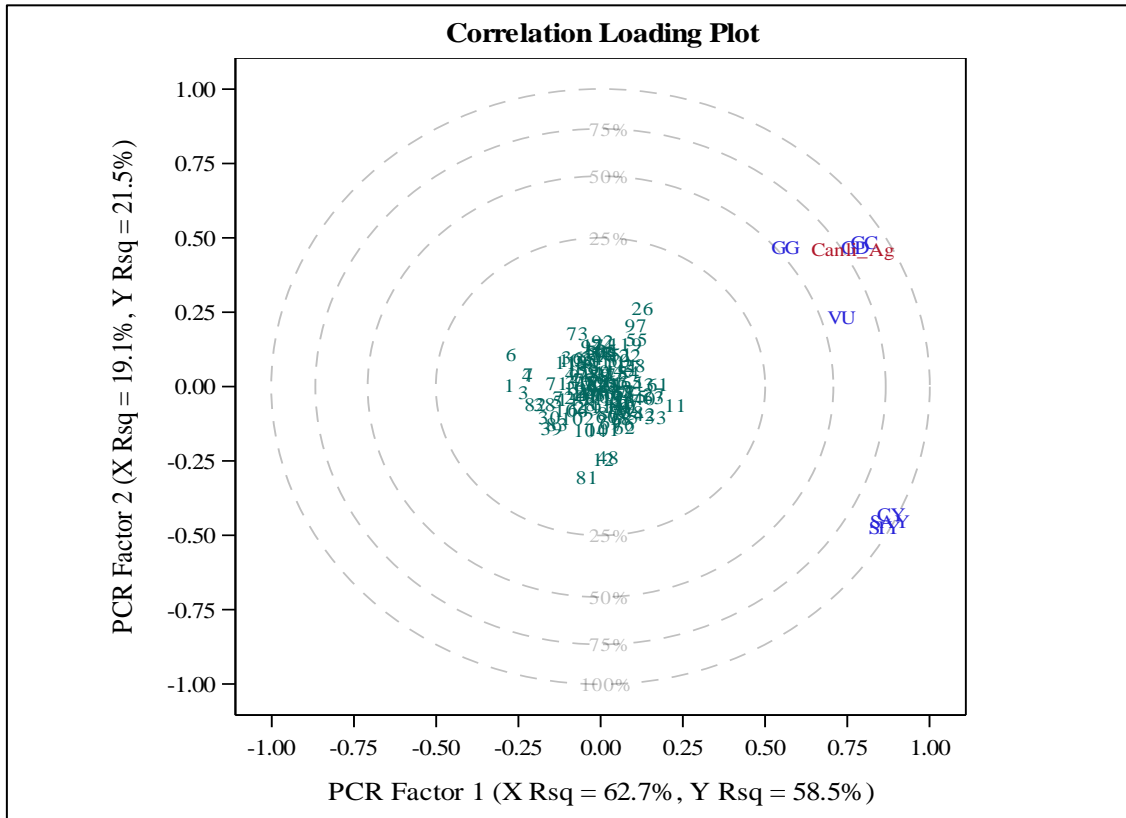
Şekil 4.5 incelendiğinde yedi açıklayıcı değişkenin yer aldığı modelde çapraz doğrulama sonunda iki gizil faktör seçildiği görülmektedir. Daha önce de ifade edildiği gibi optimum gizil faktör sayısının belirlenmesinde HKO'nun karekökü, bağımlı ve bağımsız değişkenlerdeki değişimi ayrı ayrı ifade eden belirleme katsayıları dikkate alınmıştır. Bu üç kriterin optimum koşullarda bulunduğu noktada TB regresyon analizinde seçilecek gizil faktör sayısına karar verilmiştir. Gizil faktör sayılarına göre HKO, HKO'nun karekökü, bağımlı ve bağımsız değişkenlerdeki değişimi açıklamak üzere kullanılan belirleme katsayılarındaki değişim Çizelge 4.8'de verilmiştir.

Çizelge 4.8. TB regresyon analiz için gizil faktör sayıları ve model uyum kriterleri.

| Gizil faktör sayısı | HKO | \sqrt{HKO} | Model Etkisi | | Bağımlı Değişken Etkisi | |
|---------------------|--------|--------------|---------------------------|---------|---------------------------|---------|
| | | | Gizil Faktör için geçerli | Toplam | Gizil Faktör için geçerli | Toplam |
| 1 | 0.4629 | 0.6804 | 62.7213 | 62.7213 | 58.5396 | 58.5396 |
| 2 | 0.2295 | 0.4791 | 19.1407 | 81.8620 | 21.4980 | 80.0376 |

Çizelge 4.8’de gizil faktör sayısı bir iken HKO 0.4629 ve HKO’nun karekökü 0.6840, gizil faktör sayısı iki olduğunda HKO 0.2295 ve HKO’nun karekökü 0.4791 olmuştur. Modelde yer alan açıklayıcı değişkenlerdeki ve cevap değişkenindeki açıklama miktarları ise gizil faktör sayısı arttıkça artmıştır. Bağımlı değişkenler için bu miktar gizil faktör sayısı bir olduğunda % 62.7213 ile başlayıp, ikinci gizil faktör için % 19.1407 ile toplamda % 81.8620 olmuştur. Cevap değişkeni olan CA’daki değişimin % 58.5396 birinci gizil faktör tarafından açıklanmıştır. İkinci gizil faktör ile açıklanan miktar % 21.4980 olup toplamda bu miktar % 80.0376 olarak elde edilmiştir (Çizelge 4.8).

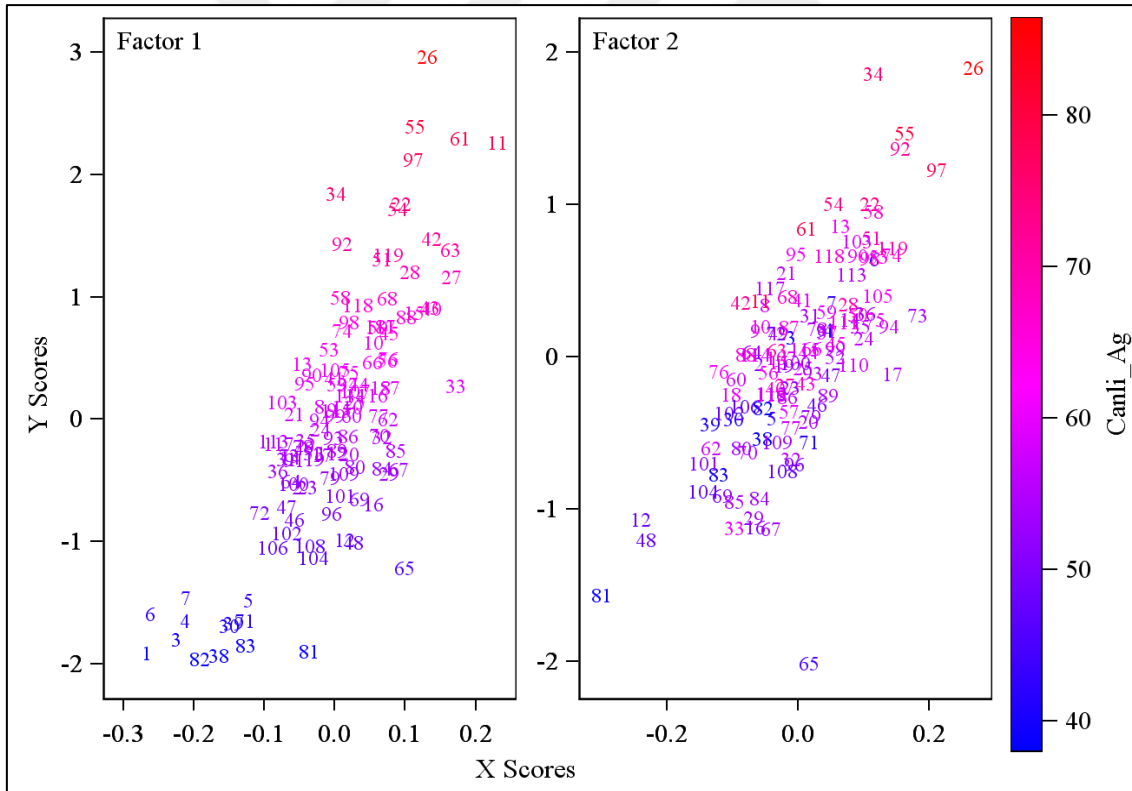
TB regresyon analizinin bir diğer çıktısı olan korelasyon yük grafiği birinci ve ikinci gizil faktörler ile birlikte Şekil 4.6’da verilmiştir.



Şekil 4.5. TB regresyon analizi ile elde edilen korelasyon grafiği.

Şekil 4.6'nın Şekil 4.2'den tek farklı yönü TB regresyonunun birinci ve ikinci gizil faktörler tarafından yapılan açıklanma miktarlarıdır. Çizelge 4.8'de verilenlerle birlikte değerlendirilecek olursa, birinci ve ikinci gizil faktörlerin açıklama miktarları açıklayıcı değişkenler için sırasıyla % 62.7213 ve % 19.1407 olurken, cevap değişkeni için sırasıyla % 58.5396 ve % 21.4980 olduğu görülmektedir. Bu bilgiler, Şekil 4.6'da birinci gizil faktör için x eksenini ve ikinci gizil faktör için y eksenini üzerinde açık bir şekilde gösterilmiştir.

TB regresyon analizinde seçilen iki gizil faktör için açıklayıcı değişkenlere karşı cevap değişkenlerinin yer aldığı grafiği Şekil 4.7'de verilmiştir. Şekilde, en yüksek korelasyon birinci gizil faktörde gözlenirken ikinci gizil faktörde bu korelasyonun azaldığı görülmektedir.

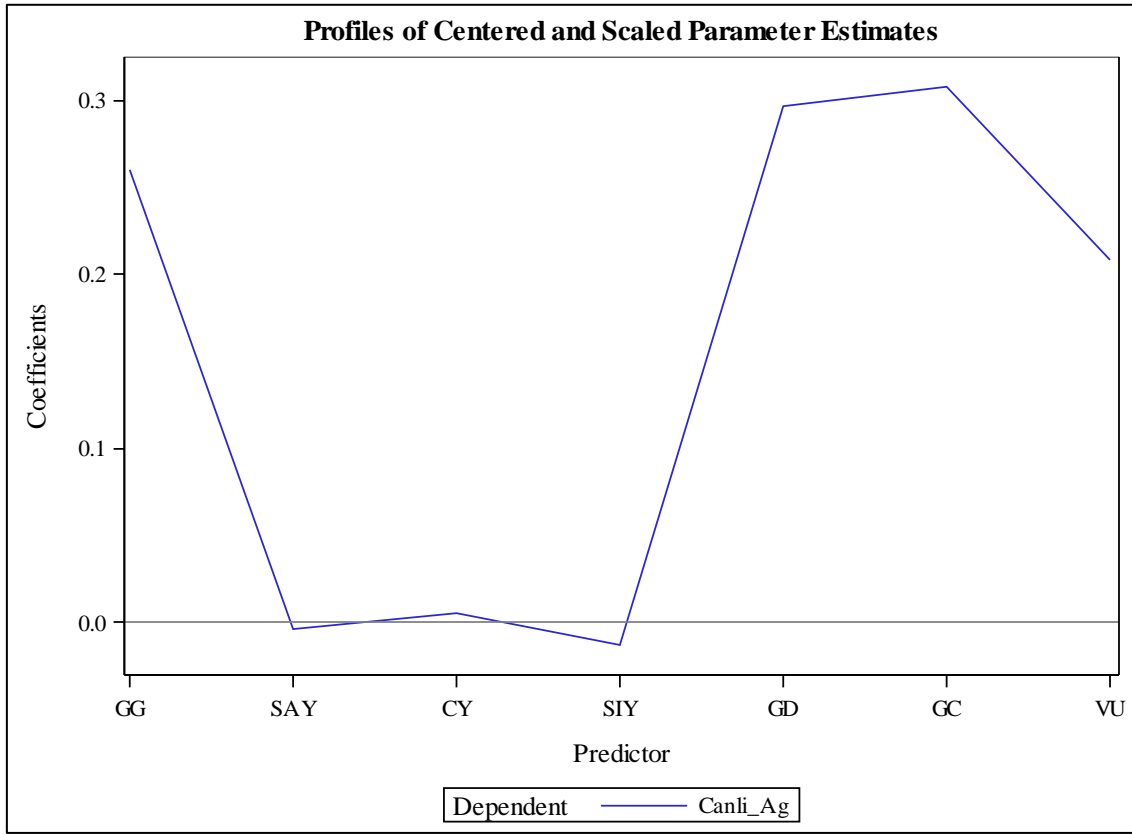


Şekil.4.7. TB regresyon analizinde iki gizil faktör için açıklayıcı değişkenlere karşı cevap değişkeninin grafiği.

TB regresyon analiz yöntemi kullanılarak tahmin edilen standardize edilmiş ve edilmemiş regresyon katsayılar Çizelge 4.9 ve standardize edilmiş katsayılara ilişkin profili grafiği Şekil 4.7’de verilmiştir.

Çizelge 4.9. TB regresyon analizi kullanılarak CA için elde edilen regresyon katsayıları

| Katsayılar | Sabit | GG | SAY | CY | SIY | GD | GC | VU |
|---------------|----------|--------|---------|--------|---------|--------|--------|--------|
| Std edilmiş | 0.0000 | 0.0905 | -0.0123 | 0.0026 | -0.0161 | 0.3533 | 0.4085 | 0.2100 |
| Std edilmemiş | -98.2144 | 0.6007 | -0.0278 | 0.0051 | -0.0364 | 1.3787 | 0.7004 | 0.5071 |



Şekil 4.6. TB regresyon analizi kullanılarak elde edilen standardize edilmiş regresyon katsayılarının profili

Standardize edilmiş regresyon katsayıları için oluşturulmuş Çizelge 4.9’deki görsel incelendiğinde, SAY, CY ve SIY değişkenlerinin modeli açıklamada sifıra yakın önemleri olduğu görülebilir. Bu önem miktarları SAY, CY ve SIY için sırasıyla-0.0123, 0.0026, -0.0161 tahmin edilmiştir. En çok açıklamaya sahip olan değişkenler ise büyüklük sırasına göre GC, GD, VU ve GG olmuştur. CA’daki değişimi açıklamadaki katkılarının yüksek olduğu bu değişkenlerin standardize edilmiş katsayıları sırasıyla 0.4085, 0.3533, 0.2100, 0.0905 şeklinde elde tahmin edilmiştir. Değişkenler için bu

büyüklik sıralaması KEKK regresyon analizi ile elde edilen sıralamanın aynı olmuştur.

KEKK regresyonu için SIMPLS, NIPALS algoritmaları ve TB regresyon için elde edilen model etki yükleri ve model etki ağırlıkları sırasıyla Çizelge 4.10 ve 4.11’de verilmiştir.

Çizelge 4.10. SIMPLS, NIPALS algoritması kullanan KEKK ve TB regresyon analizlerinden elde edilen elde edilen model etki yükleri

| SIMPLS | | | | | | | |
|----------------------|-----------|------------|-----------|------------|-----------|-----------|-----------|
| GFS | GG | SAY | CY | SIY | GD | GC | VU |
| 1 | 0.297759 | 0.376710 | 0.381705 | 0.367384 | 0.411217 | 0.424315 | 0.373653 |
| 2 | 0.190453 | -0.512237 | -0.486068 | -0.529983 | 0.288221 | 0.286810 | 0.137349 |
| NIPALS | | | | | | | |
| 1 | 0.297759 | 0.376710 | 0.381705 | 0.367384 | 0.411217 | 0.424315 | 0.373653 |
| 2 | 0.190453 | -0.512237 | -0.486068 | -0.529983 | 0.288221 | 0.286810 | 0.137349 |
| TB regresyonu | | | | | | | |
| 1 | 0.267077 | 0.419510 | 0.421747 | 0.412260 | 0.369491 | 0.382298 | 0.349538 |
| 2 | 0.404713 | -0.390703 | -0.369879 | -0.408270 | 0.404195 | 0.419853 | 0.201031 |

GFS: Gizil faktör sayısı

Çizelge 4.11. SIMPLS, NIPALS algoritması kullanan KEKK ve TB regresyon analizlerinden elde edilen model etki ağırlıkları

| SIMPLS | | | | | | | |
|----------------------|-----------|------------|-----------|------------|-----------|-----------|-----------|
| GFS | GG | SAY | CY | SIY | GD | GC | VU |
| 1 | 0.328817 | 0.275940 | 0.287109 | 0.259291 | 0.493619 | 0.534667 | 0.437411 |
| 2 | 0.138062 | -0.447950 | -0.420508 | -0.480500 | 0.366299 | 0.490544 | 0.283425 |
| NIPALS | | | | | | | |
| 1 | 0.328817 | 0.275940 | 0.287109 | 0.259291 | 0.493619 | 0.534667 | 0.437411 |
| 2 | 0.062144 | -0.511660 | -0.486797 | -0.540366 | 0.252331 | 0.367098 | 0.182434 |
| TB regresyonu | | | | | | | |
| 1 | 0.267077 | 0.419510 | 0.421747 | 0.412260 | 0.369491 | 0.382298 | 0.349538 |
| 2 | 0.404713 | -0.390703 | -0.369879 | -0.408270 | 0.404195 | 0.419853 | 0.201031 |

GFS: Gizil faktör sayısı

Gözlenen CA, SIMPLS, NIPALS algoritmaları ve TB regresyonu kullanılarak tahmin edilen canlı ağırlıklar (sırasıyla CA, CS_S, CA_N ve CS_TB) için tanımlayıcı

istatistikler ve gözlenen ve tahmin edilen değerler arasındaki korelasyonlar Çizelge 4.12’te verilmiştir.

Çizelge 4.12. Gözlenen ve tahmin edilen CA değişkenine ilişkin temel istatistikler, gözlenen ve tahmin edilen CA’ya ait korelasyonlar

| | N | Ortalama | Std sapma | Minimum | Maksimum | CA | p |
|-------|-----|----------|-----------|----------|----------|---------|-----|
| CA | 119 | 57.34454 | 9.84726 | 38.0000 | 86.45000 | 1 | |
| CA_S | 119 | 57.34454 | 8.92247 | 34.73990 | 80.59811 | 0.90609 | *** |
| CA_N | 119 | 57.34454 | 8.92247 | 34.73990 | 80.59811 | 0.90609 | *** |
| CA_TB | 119 | 57.34454 | 8.80973 | 35.22933 | 80.88948 | 0.89464 | *** |

***: P<0.001

Çizelge 4.12 incelendiğinde CA ile CA_S ve CA_N arasındaki korelasyon değeri aynı olup % 90.609 olarak bulunmuştur. Her iki algoritma ile elde edilen bu değer yaklaşık % 91 olup anlamlı bulunmuştur (P<0.001). TB regresyonu kullanılarak gözlenen ve tahmin edilen CA değerleri arasındaki ilişki % 89.464 olarak elde edilmiştir. Bu değer yaklaşık %90 olup anlamlı bulunmuştur (P<0.001).

Yapılan bu çalışmada, verilere KEKK ve TB regresyon analiz yöntemleri uygulanarak elde edilen standardize edilmiş katsayılar Çizelge 4.13’de verilmiştir. Bu çizelgede katsayıların birbirinden farklı olduğu görülmektedir. Ancak modele katkısı düşünüldüğünde, her iki yöntemde de GC en yüksek etkiye sahip olup bunu büyüklük sırasına göre GD, VU, GG takip etmektedir. Çizelge 4.13 incelendiğinde, her iki yöntem için CA’daki değişimi açıklamada en az katkısı olan değişkeninin, CY olduğu söylenebilir.

Çizelge 4.13. KEKK ve TB regresyon analiz yöntemleri uygulanarak elde edilen standardize edilmiş katsayılar.

| Regresyon Yöntemi | Standardize edilmiş katsayılar | | | | | | | |
|-------------------|--------------------------------|--------|---------|---------|---------|--------|--------|--------|
| | Sabit | GG | SAY | CY | SIY | GD | GC | VU |
| KEKK | 0.0000 | 0.1727 | -0.0148 | -0.0025 | -0.0308 | 0.3042 | 0.3561 | 0.2579 |
| TB | 0.0000 | 0.0905 | -0.0123 | 0.0026 | -0.0161 | 0.3533 | 0.4085 | 0.2100 |

GG: Göğüs genişliği, SAY: Sağrı yüksekliği, CY: Cidago yüksekliği, SIY: sırt yüksekliği, GD: Göğüs derinliği, GC: Göğüs çevresi, VU: vücut uzunluğu.





5. TARTIŞMA VE SONUÇ

Çoklu doğrusal regresyon analizinde sıklıkla kullanılan yöntem EKK yöntemidir. Bunun nedeni EKK yönteminin uygulamadaki ve yorumlamadaki kolaylığı ve anlaşılır olmasıdır. Çoklu doğrusal regresyon analizinde EKK yönteminin uygulanabilmesi için bazı varsayımların geçerli olması gereklidir. Bu varsayımlar kısaca şu şekilde özetlenebilir: hatalar şans değişkeni olup ortalaması $E(\varepsilon_i) = 0$ ve varyansı $Var(\varepsilon_i) = \sigma^2$, normal dağılıma sahip olmalıdır. Hatalar birbirlerinden ($Cov(\varepsilon_i, \varepsilon_j) = 0$) ve açıklayıcı değişkenlerden $Cov(\varepsilon_i, X_j) = 0$ bağımsızdırlar. Açıklayıcı değişkenler arasında önemli büyüklükte ilişki olmamasının yanı sıra ve cevap değişken ile doğrusal bir ilişkiye sahip olmalıdır. Veri kümesinin sahip olduğu gözlem sayısı parametre sayısından büyük ($n > k$) olmalıdır. Varsayımların bir veya birden fazlasının sağlanamaması EKK yöntemi ile elde edilen tahminlerin güvenilirliğinin azalmasına neden olmaktadır. Regresyon modelinde yer alan açıklayıcı değişkenlerin bağımsızlığının sağlanamaması çoklu bağlantı sorununun ortaya çıkmasına neden olmaktadır. EKK'ya alternatif olarak geliştirilen bazı yöntemler kullanılarak bu sorunun üstesinden gelinebilmektedir. Bu amaçla kullanılan tahmin yöntemlerinden birkaçı Ridge regresyon, TB regresyonu, Elastic_net, Lasso regresyon ve KEKK regresyon yöntemleri şeklinde özetlenebilir.

KEKK regresyon yöntemi çeşitli çalışma alanlarında kullanılmasına (Albayarak, 2005; Eker, 2011; Topal ve Ark, 2010) rağmen hayvancılık alanındaki veriler için yaygın bir şekilde kullanılmamıştır. Bu yöntem çoklu bağlantı problemi ile karşılaşıldığında bu sorunun üstesinden gelmesinin yanı sıra modelden bağımsız değişken veya değişkenleri çıkarma prosedürüne de sahiptir. Bu çalışmada, K1 keçilerinde bazı vücut ölçüleri kullanılarak canlı ağırlıklar tahmin edilmeye çalışılmıştır. Bu amaçla, KEKK ve TB regresyon yöntemleri kullanılmış ve yöntemlerin teorik bilgiler verildikten sonra sonuçlar tahmin doğruluğu bakımından değerlendirilmiştir.

Çalışmaya tanımlayıcı istatistiklerin elde edilmesi ile başlanmış ve sonuçlar Çizelge 4.1'de verilmiştir. Özellikle varyasyon katsayılarının % 30'un üzerinde olmaması veri kümesinin güvenilir olduğunu ortaya koymuştur. Zira literatürde

varyasyon katsayısının % 30'un üzerinde olması verinin güvenilirliği ortadan kaldırmaktadır (Gujarati, 1995, Alpar, 2011). Çalışmada cevap değişkeni olarak modelde yer alan CA ve açıklayıcı değişkenler olarak yer alan GG, SAY, CY, SIY, GD, GC ve VU arasındaki ilişkileri gösteren Çizelge 4.2, açıklayıcı değişkenler arasında çoklu bağlantı sorunu olabileceğini akla getirmiştir. Zira, SAY ile CY arasındaki korelasyon % 95.5, SAY ile SIY arasındaki korelasyon % 98.0, CY ile SIY arasındaki korelasyon % 94.9 ve GD ile GC arasındaki korelasyon ise %85.4 olup veri kümesinin çoklu doğrusal bağlantı problemi açısından incelenmesi gerektiğini ortaya koymuştur (Akyürek ve Akkol, 2018; Çiftsüren ve Akkol, 2018; Pamukçu,2010). Çoklu bağlantı sorununun tespiti için çoklu bağlantı belirleyicileri incelenmiştir. (Çizelge 5.3, 5.4 ve 5.5). Bu kriterlerden $VIF \geq 10$, $TV < 0.1$, özdeğerlerin sıfıra çok yakın, $CI \geq 30$ olması ve varyans oranlarının çok büyük olması veri setinin çoklu bağlantı sorununa sahip olduğunu ortaya koymuştur (Alpar, 2011; Bulut ve Alın,2009). Çalışmada VIF ve TV değerleri sırasıyla SAY için 30.4096 ve 0.00329, CY için 13.5018 ve 0.0741, SIY için 38.9436 ve 0.0346 şeklinde elde edilmiştir. Özdeğerler sıfıra çok yakın ve CI, 30 değerinin çok üzerinde (Çizelge 4.4), SAY ve SIY değişkenleri için varyans oranlarının çok büyük (Çizelge 4.5, sırasıyla 0.90838 ve 0.87698) olduğu tespit edilmiştir. Elde edilen bu sonuçlar, literatürle uyumlu olarak çalışmada kullanılan veri setinin çoklu bağlantı problemine sahip olduğunu ortaya koyulmuştur (Albayrak,2005; Alpar,2011; Belsley 1991; Gujarati, 1995; Marquardt ve Snee 1975).

Çalışma verisindeki çoklu bağlantı probleminin üstesinden gelebilmek amacıyla KEKK regresyon yöntemi kullanılmıştır. Bu yöntemde araştırmacılar tarafından sıklıkla kullanılan SIMPLS ve NIPALS algoritmaları veriye uygulanmış ve elde edilen sonuçlar birbirinin aynı olmuştur. Çalışmada varılan bu sonuç Jong (1993)'ün çalışması ile uyumlu bulunmuştur. Jong 1993'de yaptığı çalışmasında cevap değişkeninin sayısının bir olması durumunda SIMPLS ve NIPALS algoritmalarının aynı sonuçları ürettiğini bildirmiştir (Jong,1993).

Veriye KEKK regresyon yöntemi uygulanırken VIP istatistiği kullanılarak modelden açıklayıcı değişken çıkarılıp çıkarılmayacağı test edilmiştir. Modelde yer alan değişkenlerden GG, 0.8 değerinin altında kalmasına rağmen bu değişkenin standardize edilmiş katsayısı çok küçük olmadığı için modelde kalmasına karar verilmiştir (Wold, 1966). Hem SIMPLS hem de NIPALS algoritması çalıştırıldığında

iki gizil faktör ile bağımlı değişkendeki değişimim % 82.0993'lik kısmı açıklamıştır (Çizelge 4.6). Bu miktarın büyük kısmı birinci gizil faktör tarafından açıklanmıştır (% 70.7149).

Çalışmada kullanılan bir diğer yöntem TB regresyon analiz yöntemidir. Bu yöntem, boyut indirgeme amaçlı olduğu kadar çoklu bağlantı sorununun üstesinden gelme amaçlı olarak da kullanıldığı için araştırmacılar tarafından sıklıkla tercih edilen yöntemlerden biri olmuştur (Albayrak, 2005; Eker, 2011; Graham, 2003; Mendeş, 2011; Ortabaş, 2001). TB regresyon analizinde iki gizil faktör seçilmiştir. Bu regresyon yöntemi ile cevap değişkeni olan CA'daki değişimim % 80.0376'lık kısmı açıklanmıştır. Hem KEKK hem de TB regresyon yönteminde iki gizil faktör seçilmiş ancak CA'daki değişim KEKK yönteminde TB regresyon yönteminden daha iyi olmuştur. Bir diğer ifade ile CA'daki değişim miktarı KEKK regresyon yöntemi ile TB analiz yönteminden daha iyi (% 82.0993-% 80.0376=% 2.0617) açıklanmıştır. Bu miktar % 2.0617 olmuştur. Boyut indirgeme bakımından her iki regresyon yönteminin iki gizil faktöre sahip olması, Maitra ve Yan (2008)'nin çalışmasının sonuçlarından farklı olmuştur. Zira Maitra ve Yan (2008), yaptıkları çalışmada KEKK yönteminin TB regresyon analiz yönteminden boyut indirgeme konusunda daha iyi olduğunu ortaya koymuşlardır.

Yöntemleri karşılaştırması amacıyla gözlenen CA ile kullanılan yöntemlere göre beklenen CA değerleri arasındaki korelasyonlar hesaplanmış ve en yüksek korelasyonun KEKK yöntemi (hem SIMPLS hem de NIPALS) ile elde edildiği belirlenmiştir (Çizelge 4.12). Elde edilen bu sonuç, Macciotta ve ark., (2006), El-Fallah ve El-Salam (2014) ve Akkol ve Akyürek (2018)'un çalışmaları ile uyumlu olmuştur. Bu çalışmaların tamamında KEKK regresyon yöntemi ile elde edilen sonuçların TB regresyon analizinden daha yüksek doğrulukta sonuçlar ürettiği ortaya koyulmuştur.

Karmaşıklığı en az olan sade tahmin denklemleri yorumlamada kolay ve anlaşılır olması nedeni ile, araştırmacılar tarafından her zaman tercih edilmektedir. Bu tez çalışmasında kullanılan regresyon yöntemlerinin her ikisi de aynı gizil faktör sayısına sahip olması nedeni ile aynı sade tahmin denklemine sahiptir. Ancak HKO, CA'daki açıklama miktarları, gerçek ve tahmin edilen CA değerleri arasındaki korelasyonlar dikkate alındığında, KEKK regresyon analiz yönteminin TB regresyon analiz yönteminden daha doğru ve güvenilir sonuçlar verdiği ortaya koyulmuştur.

Yapılan bu çalışmada, hayvancılık alanında karşılaşılan çoklu bağlantı problemine sahip veri kümelerinin analizinde KEKK regresyon yönteminin kullanılmasının TB regresyon analiz yöntemine iyi bir alternatif olacağı sonucuna ulaşılmıştır.





KAYNAKLAR

- Alpar, R., 1997. *Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş*. Bağırğan Yayın Evi, Ankara.
- Alpar, R., 2003. *Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş 1*, Nobel Yayın Dağıtım, Ankara.
- Alpar, R., 2011. *Uygulamalı Çok Değişkenli İstatistiksel Yöntemler*, Detay Yayıncılık, Ankara.
- Albayrak, A., S., 2005. Çoklu doğrusal bağlantı halinde en küçük kareler tekniğinin alternatifleri yanlı tahmin teknikleri ve bir uygulama. *Sosyal Bilimler Dergisi*, 1(1):106-126.
- Albayrak, S.A., 2005. Çoklu bağlantı halinde en küçük kareler teknikleri ve bir uygulama. *Sosyal Bilimler Dergisi*. 1(1): 105-126.
- Albayrak, A. S., 2006. *Uygulamalı Çok Değişkenli İstatistik Teknikleri*. Asil Yayın Dağıtım, ISBN: 9759091984, Ankara.
- Alkan, Ö., 2008. *Temel Bileşenler Analizi ve Bir Uygulama Örneği* (yüksek lisans tezi) Sosyal Bilimler Enstitüsü, Erzurum.
- Akyürek, S., Akkol, S., 2018. Yumurta iç kalite özelliklerinin kısmi en küçük kareler regresyonu kullanılarak tahmin edilmesi *YYÜ Tarım Bil*, 28(4): 473-481.
- Arnold, SF., 1981. *The Theory of Linear Models and Multivariate Analysis*, John Wiley and Sons, Inc., USA.
- Bastien, P., Vinzi, V.E., Tenenhaus, M., 2005. PLS Generalised linear regression. *Computational Statistics and Data Analysis* 48(1):17-46
- Bush, B. L., Nachbar, J. B., 1993. Sample-distance partial least squares: PLS optimized for many variables, with application to CoMFA. *J. Comput. Aided Mol. Des*, 7(5): 587-619.
- Bulut, E., Alın, A., 2009. Kısmi en küçük kareler regresyon yöntemi algoritmalarından nıplars ve pls-kernel algoritmalarının karşılaştırılması ve bir uygulama. *Dokuz Eylül Üniversitesi İktisadi ve İdari Bilimleri Fakültesi Dergisi*.24(2): 127-138.
- Bulut, Y.M., 2011. *Çoklu İç İlişki Durumunda Kısmi En Küçük Kareler Regresyonu ve Alternatif Yöntemlerle Karşılaştırılması*. Eskişehir Osmangazi Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek lisans tezi. Eskişehir.
- Belsley, D.A., 1991. *Conditioning Diagnostics, Collinearity and Weak Data in Regression*. John Wiley and Sons, New York, NY, USA.
- Çiftsüren M.N., Akkol, S., 2018. Prediction of internal egg quality characteristic and variable selection using regularization methods: ridge, LASSO and elastic net. *Archives Animal Breeding*, 61(3): 279-284.
- Dayal, B. S., MacGregor, J. F., 1997. Improved PLS Algorithms, *Journal of Chemometrics*, 11(1): 73-85.
- De Jong, S.,1993. Simpls: an alternative approach to partial least squares regression. *Chemometrics and Intelligent Laboratory Systems*, 18(3): 251-263.
- Dielman, T.E., 1991. *Applied Regression Analysis for Business and Economics*. (First Edition) Boston: PWS-KENT Publishing Company, 2, 62.
- Draper, N.R.; H. Smith 1981. *Applied Regression Analysis*. 2.ed., New York: John Wiley&Sons.

- El-Fallah M., El-Salam A., (2014). A note on partial least squares regression for multicollinearity (A comparative study). *International Journal of Applied Science and Technology*. **4**(1):163-169
- Eker, S., 2011. *Regresyon Analizinde Çoklu Bağlantı Durumunda En Küçük Kareler, Ridge Regresyon ve Temel Bileşenler Yaklaşımlarının Karşılaştırılması* (yüksek lisans tezi) Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, Zootekni Anabilim Dalı, 11-36
- Ersungur, Ş., Kızıltan, M., Özgür, A.P., 2007. Türkiye’de bölgelerin sosyo-ekonomik gelişmişlik sıralaması: temel bileşenler analizi. *Atatürk Üniversitesi İktisadi ve İdari Bilimler Dergisi*. **21**(2):55-66.
- Ferrand-Calmels, M., Palhière, I., Brochard, M., Leray, O., Astruc, J. M., Aurel, M. R., Douguet, M., 2014. Prediction of fattyacid profiles in cow, ewe, and goat milk by mid-infrareds pectrometry. *Journal of Dairy Science*. **97**(1): 17-35.
- Firinguetti, L., Kibria G., Araya R., 2016. Study of partial least squares and ridge regression methods. *Journal Communications in Statistics Simulation and Computation*, **46** (8): 6631-6644.
- Frank, I. E., 1987. Intermediate least squares regression methods. *Chemometrics and Intelligent Laboratory Systems*, **1**(3): 233-242.
- Frank, I. E., Friedman, J. H., 1993. A statistical view of some chemometrics regression tools. *Technometrics*, **35**(2): 109-147.
- Fox J, 1997. *Applied Regression Analysis: Linear Models and Related Methods*. Sage Publication, USA :123–240.
- Garthwaite, P.H., 1994. An Interpretation of partial least squares, *Journal of the American Statistical Association*, **89**(425):122-127.
- Geladi, P., Kowalski, B. R., 1986. An example of 2-block predictive partial least squares regression with simulated data. *Anal. Chim. Acta*, **185**:19-32.
- Godoy, J. L., Vega, J. R., Marchetti, J. L., 2014. Relation ships between PCA and PLS-regression. *Chemometrics and Intelligent Laboratory Systems*, **130**: 182-191.
- Gujarati, D.N., 1995. *Basic Econometrics*. 3rd Ed., McGraw-Hill, New York. 1002p
- Glen, W.G., Dunn, III, W.J., Sarker, M. and Scott, D.R., 1989. UNIPALS: Software for principal components analysis and partial least squares regression. *Tetrahedron Computer Methodology* **2**(6): 377–396
- Graham, M. H., 2003. Confronting multicollinearity in ecological multiple regression, *Ecology*, **84**(11):2809–2815.
- İşi, A., 2002. *Yanlı Tahmin ediciler ve Kombinasyonları*, (yüksek lisans tezi), Gazi Üniversitesi, Fen Bilimleri Enstitüsü.
- Helland, I.S., 1988, On the structure of partial least squares regression. *Journal Commun. Statist. – Simula*, **17**(2): 581-607.
- Höskuldsson, A., 1988. PLS regression methods, *Journal of Chemometrics*, **2**(3): 211-228.
- Kalaycı, Ş., 2008. *SPSS Uygulamalı Çok Değişkenli İstatistik Teknikleri*, Asil Yayın Dağıtım, Ankara.
- Kemalbay, G., ve Korkmazoğlu, O.B., 2012. Effect of multicollinearity on electricity consumption forecasting using partial least squares regression. *Procedia- Social and Behavioral Sciences*, **62**:(11),50-1154.
- Koyuncu, M., 2005b. *Keçi Yetiştiriciliğinin Dünya ve Türkiye Stratejileri, Süt Keçiciliği Ulusal Kongresi Bildirisi*, 26-27 Mayıs 2005, İzmir.59-65.

- Kondylis, A., 2006, *PLS Methods in Regression: Model Assessment and Inference*, Ph.D. Thesis, University of Neuchâtel, Switzerland.
- Kleimbaum, D.G., Kupper, L.L., Muller, K.E., Nizam, A., 1998. *Applied Regression Analysis And Other Multivariable Methods*, R.R. Donnelley and Sons Duxbury Press, ISBN 0-534-20910-6, 798p USA.
- Lindgren, F., Rannar, S., 1998. Alternative partial least squares algorithms. *Perspective in Drug and Discovery and Design*, 3:105-113
- Ludwig, J.A., Reynolds J.F., 1988. *Statistical Ecology*, A Wiley-Interscience publication, ISBN 0-471-83235-9, 337p. USA.
- Maitra, S., Yan, J., 2008. Principal component analysis and partial least squares: two dimension reduction techniques for regression. *Casualty Actuarial Society Discussion Paper Program* June 15-18, Canada 79-90.
- Marquardt, D.W., Snee, R.D., (1975). Ridge Regression in Practice. *The American Statistician*, 29(1): 3-20.
- Macciotta, N. P. P., Dimauro, C., Bacciu, N., Fresi, P., Cappio-Borlino, A., 2006. Use of a partial least-squares regression model to predict test day of milk, fat and protein yields in dairy goats. *Animal Science*, 82(04): 463-468.
- Martens, H., Naes, T., 1989. *Multivariate Calibration*. John Wiley & Sons, Chichester, New York, Brisbane, Toronto, Singapore.
- Manne, R., 1987, Analysis of two partial-least-squares algorithms for multivariate calibration, *Chemometrics and Intelligent Laboratory Systems*, 2(1-3): 187-197.
- Maxwell S.E., 2000. Sample size in multiple regression analysis. *Psychological Methods*, 5(4): 434-458.
- Mendes, M., 2011. Multivariate multiple regression analysis based on principal component scores to study relationships between some pre- and post-slaughter traits of broilers. *Journal of Agricultural Sciences*, 17(1): 77-83
- Montgomery, D., Peck, E., 1982. *Introduction To Linear Regression Analysis*, New York.
- Morrison, D.F., 1967. *Multivariate Statistical Methods*, McGraw-Hill Book Company, 338p., New York.
- Mohebbi, M., Nourijelyani, K., Zeraati, H., 2007. A simulation study on robust alternatives of least squares regression. *Journal of Applied Sciences*, 7(22): 3469-3476.
- Montgomery, D.C., Peck, E.A., Vining, G.G., 2001. *Introduction to Linear Regression Analysis*, 3rd Edition, John Wiley & Sons, New York.
- Naes, T., Irgens, C., Martens, H., 1986. Comparison of Linear Statistical Methods for Calibration of NIR Instruments. *Journal of the Royal Statistical Society* 35(2): 195-206.
- Neter, J., Kutner, M., Nachtsheim, C., Wasserman, W., 1996. *Applied Linear Regression Models*, Irwin, USA.
- Oliveira, L. Z., Arruda, R.P., Andrade, A.F.C., Celeghini, E.C.C., Reeb, P.D., Martins, J.P.N., Santos, R., F.G.P., Beletti, M.E., Peres, G.F.R., Monteiro, F.M., Lima, V.F.M.H., 2013. Assessment of in vitro sperm characteristics and their importance in the prediction of conception rate in a bovine timed-AI program. *Animal Reproduction Science* 137 (3-4):145-155.
- Onay, İ., 2003. *Ortalama Artelyel Kan Basıncını Etkileyen Faktörlerin Temel Bileşenler Regresyonu ile Belirlenmesi*, (yüksek lisans tezi). Gazi Üniversitesi, Fen Bilimleri Enstitüsü. Ankara.

- Ortabaş, N. 2001. *Principal components in the problem of multicollinearity*, (yüksek lisans tezi). Dokuz Eylül Üniversitesi, Fen Bilimleri Enstitüsü, 45s.
- Orhunbilge, A., N., 2002. *Uygulamalı Regresyon ve Korelasyon Analizi*, İ.Ü Basım ve Yayınevi, İstanbul.
- Ortiz, M., Sarabia, L., Herrero, A., 2006. Robust Regression Techniques: A Useful Alternative for the Detection Data in Chemical Analysis. *Talanta*, **70**(3): 499–512.
- Özdamar, K., (1999). *Paket Program ile İstatistiksel Veri Analizi*. Kaan Kitabevi, Eskişehir.
- Özdamar, K., 2003. *SPSS ile Biyoistatistik*, Eskişehir.
- Özdamar K., 2004. *Paket Programlar ile İstatistiksel Veri Analizi* (Çok değişkenli analizler), Kaan Kitabevi, 502, Eskişehir.
- Phatak, A., 1993. *Evaluation of some Multivariate Methods and their Applications in Chemical Engineering*. PhD thesis, University of Waterloo, Ontario, Canada (unpublished).
- Phatak, A., De Jong, S., 1997, The geometry of partial least squares. *Journal of Chemometrics*, **11**(4) 311–338.
- Pamukçu, E., 2010. *Sistolik Kan Basıncını Etkileyebilecek Faktörlerin Ridge Regresyon Analizi ile İncelenmesi ve Çoklu Bağlantı Problemi*, (yüksek lisans tezi). Fırat Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, 84s.
- Pierce, KM., Hope, JL., Hoggard, JC., Synovec, RE., 2006. A principal component analysis based method to discover chemical differences in comprehensive two-dimensional gas chromatography with time-of-flight mass spectrometry (GCxGC-TOFMS) Separation of Metabolites in Plantsamples. *J. Talanta*, **70**(4): 797-804.
- Polat, E., 2009. *Kısmi En Küçük Kareler Regresyonu*, (yüksek lisans tezi). Hacettepe Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, Ankara.
- Rawlings, JO., Pantula, SG., Dickey, DA., 1998. *Applied Regression Analysis: A Research Tool, Second Edition*, springer texts in statistics, USA., P. 658.
- Rännar, S., Fredrik, L., Geladi, P., Wold, S., 1994. A PLS kernel algorithm for data sets with many variables and fewer objects. Part 1: theory and algorithm, *Journal of Chemometrics*, **8**(2): 111-125.
- Rencher, A. C., (2002), *Methods of Multivariate Analysis*, John Wiley and Sons, New York
- Rosipal, R., Trejo, L., 2001. Kernel PLS in reproducing kernel hilbert spaces. *Journal of Machine Learning Research*, **2**(Dec): 97-123.
- Rosipal, R., Kramer, N., 2006, *Overview and Recent Advances in Partial Least Squares. In: Saunders C, Grobelnik M, Gunn S, Shawe-Taylor J (Eds.)*, subspace, latent structure and feature selection techniques springer: 34-51.
- Skırbic, B., Onjic, A., 2007. Multivariate analysis of microelement content in wheat cultivate in serbia. *Food Control*. **18**: 338-345p.
- Sharma, S., 1996. *Applied Multivariate Techniques*, John Wiley and Sons, Inc., ISBN 0.471.31064.6, 493p. USA
- Shrestha, S., Kazama, F., 2007. Assessment of surface water quality using multivariate statistical techniques: a case study of the fuji river basin, japan. *Environmental Modelling and Software*. **22**(4): 464-475p.
- Shaibu, A. S., Adnan, A. A., 2015. Predicting grainyield of maize using drought tolerance traits. *African Journal of Agricultural Research*, **10**(33): 3332-3337.

- Shanmugam, R., Johnson, C., 2007. At a crossroad of data envelopment and principal component analyses, *The International Journal of management Science*, **35**(4): 351-364.
- Stone, M., Brooks, R. J., 1990. Continuum Regression: Cross-Validated sequentially constructed prediction embracing ordinary least squares, partial least squares and principal component regression. *J. R. Statist. Soc.* **52** (2): 237–269.
- Tatlıdil, H., 1996. *Uygulamalı Çok Değişkenli İstatiksel Analiz*, Akademi Matbaası Ankara.
- Timm, NH., 2002. *Applied Multivariate Analysis*. Springer- verlag, USA.
- Toklu, İ.T. ve Toklu, A.T. 2015. “Zeytinyağının İçsel ve Dışsal İşaretlerine Tüketicilerin Tepkisi: Nicel Bir Araştırma”, *International Review of Economics and Management*, **3**(1), 61-84.
- Topal, M., Eyduran, E., Yağanoğlu, A.M., Sönmez, A.Y. ve Keskin, S. 2010 Çoklu Doğrusal Bağlantı Durumunda Ridge ve Temel Bileşenler Regresyon Analiz Yöntemlerinin Kullanımı, *Atatürk Üniversitesi Ziraat Fakültesi Dergisi*, **41** (1):53-57.
- Tuik, www.tuik.gov.tr. Erişim 10 Haziran 2018.
- Voncina DB., Kodba ZC., Novic, M., 2005. Multivariate data analysis in classification of vegetable oil characterized by the content of fattyacids. *Chemometrics and Intelligent laboratory Systems*, **75**(1): 31-43.
- Vinod, H. D., Ullah, A., 1981. *Recent Advanced in Regression Methods*. Marcel Dekker, New York.
- Webster A., 1995. *Applied Statistfor Business and Economics*, 683-684
- Weisberg, S., 2005. *Applied Linear Regression*. John Wiley & Sons, Inc. Hoboken, New Jersey.
- Wilcox, RR., 1997. *Introduction to Robustestimation and Hypothesis Testing*. Academic Press. San Diego.
- Widodo, A., Yang, B.S., Han, T., 2007. Combination of independent analysis and support vector machines for intelligent faults diagnosis of induction motors. *Expert Systems With Applications*, **32**(2): 299-312.
- Wold, S., Martens, H., & Wold, H., 1983. *The Multivariate Calibration Problem in Chemistry solved by the PLS Method*. Lecture notes in mathematics. Proc. Conf. Matrix Pencils, In A. Ruhe, and B. Kgstrm (ed.).
- Wold, S., Ruhe, A., Wold, H., Dunn III, W.J., 1984. The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses, *SIAM J. Sci. Stat. Comput*, **5**(3), 735–743.
- Wold, S., 1994. “*PLS for Multivariate Linear Modeling*,” *QSAR: Chemometric Methods in Molecular Design*. Methods and principles in medicinal chemistry.
- Wold, H., 1966. “*Estimation of Principal Components and Related Models by Iterative Least Squares*,” in P. R. Krishnaiah, ed., *Multivariate Analysis*, New York: Academic Press.



ÖZ GEÇMİŞ

1989 yılında Van'da doğdu. İlk orta ve lise eğitimini tamamladıktan sonra 2008 yılında Yüzüncü Yıl Üniversitesi Fen-Edebiyat Fakültesi Matematik Bölümü'ne yerleşti. 2012 yılında lisans eğitimini tamamladı. 2015 yılında Van Yüzüncü Yıl Üniversitesi, Fen Bilimleri Enstitüsü, Zootekni Anabilim Dalı'nda yüksek lisans programına başladı. Şu an Gençlik spor bakanlığı gençlik lideri olarak görev yapıyor.



T.C.
VAN YÜZÜNCÜ YIL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
LİSANSÜSTÜ TEZ ORJİNALLİK RAPORU

Tarih: 10/07/2019

Tez Başlığı / Konusu: **KİSMİ EN KÜÇÜK KARELER REGRESYONU VE TEMEL
BİLEŞENLER REGRESYON YÖNTEMLERİNİN İNCELENMESİ VE
HAYVANCILIKTA UYGULAMASI**

Yukarıda başlığı/konusu belirlenen tez çalışmamın Kapak sayfası, Giriş, Ana bölümler ve Sonuç bölümlerinden oluşan toplam 75 sayfalık kısmına ilişkin, 10/07/2019 tarihinde şahsım/tez danışmamın tarafından turnitin intihal tespit programından aşağıda belirtilen filtreleme uygulanarak alınmış olan orijinallik raporuna göre, tezimin benzerlik oranı % 3 (üç) tür.

Uygulanan filtreler aşağıda verilmiştir:

- Kabul ve onay sayfası hariç,
- Teşekkür hariç,
- İçindekiler hariç,
- Simge ve kısaltmalar hariç,
- Gereç ve yöntemler hariç,
- Kaynakça hariç,
- Alıntılar hariç,
- Tezden çıkan yayınlar hariç,
- 7 kelimedenden daha az örtüşme içeren metin kısımları hariç (Limit inatch size to 7 words)

Van Yüzüncü Yıl Üniversitesi Lisansüstü Tez Orijinallik Raporu Alınması ve Kullanılmasına İlişkin Yönergeyi inceledim ve bu yönergede belirtilen azami benzerlik oranlarına göre tez çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini bilgilerinize arz ederim.

16.05.2019

Tarih ve İmza



Adı Soyadı: Seda AKYÜREK

Öğrenci No:

Anabilim Dalı: ZOOTEKNI

Programı:

Statüsü: Y. Lisans

Doktora

DANIŞMAN ONAYI

UYGUNDUR

Dr. Öğr. Üyesi Suna AKYOL



