

T.C.
MUĞLA SITKI KOÇMAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İSTATİSTİK ANABİLİM DALI

KİSMİ EN KÜÇÜK KARELER YÖNTEMİNİN
SİMÜLASYON VERİLERİ İLE DİĞER YÖNTEMLERLE
KARŞILAŞTIRILMASI

YÜKSEK LİSANS TEZİ

İSMAİL BAĞCI

AĞUSTOS 2017
MUĞLA

T.C.
MUĞLA SITKI KOÇMAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

İSTATİSTİK ANABİLİM DALI

KİSMİ EN KÜÇÜK KARELER YÖNTEMİNİN
SİMÜLASYON VERİLERİ İLE DİĞER YÖNTEMLERLE
KARŞILAŞTIRILMASI

YÜKSEK LİSANS TEZİ

İSMAİL BAĞCI

AĞUSTOS 2017
MUĞLA

MUGLA SITKI KOÇMAN ÜNİVERSİTESİ

Fen Bilimleri Enstitüsü

TEZ ONAYI

İSMAİL BAĞCI tarafından hazırlanan “**KİSMİ EN KÜÇÜK KARELER YÖNTEMİNİN SİMÜLASYON VERİLERİ İLE DİĞER YÖNTEMLERLE KARŞILAŞTIRILMASI**” başlıklı tezinin, 01/08/2017 tarihinde aşağıdaki jüri tarafından İstatistik Anabilim Dalı’nda yüksek lisans derecesi için gerekli şartları sağladığı oybirliği/oyçokluğu ile kabul edilmiştir.

TEZ SINAV JURİSİ

Yrd. Doç. Dr. Engin YILDIZTEPE (**Jüri Başkanı**)

İstatistik Anabilim Dalı,
Dokuz Eylül Üniversitesi Fen Fakültesi, İzmir



Doç. Dr. Atila GÖKTAŞ (**Üye**)

İstatistik Anabilim Dalı,
Muğla Sıtkı Koçman Üniversitesi, Muğla



Yrd. Doç. Dr. Volkan SEVİNÇ (**Üye**)

İstatistik Anabilim Dalı,
Muğla Sıtkı Koçman Üniversitesi, Muğla



ANA BİLİM DALI BAŞKANLIĞI ONAYI

Doç. Dr. Dursun AYDIN

İstatistik Ana Bilim Dalı Başkanı
Muğla Sıtkı Koçman Üniversitesi, Muğla



Doç. Dr. Atila GÖKTAŞ (**Danışman**)

İstatistik Anabilim Dalı,
Muğla Sıtkı Koçman Üniversitesi, Muğla



Savunma Tarihi: 01/08/2018

Tez çalışmalarım sırasında elde ettiğim ve sunduğum tüm sonuç, doküman, bilgi ve belgelerin tarafımdan bizzat ve bu tez çalışması kapsamında elde edildiğini; akademik ve bilimsel etik kurallarına uygun olduğunu beyan ederim. Ayrıca, akademik ve bilimsel etik kuralları gereği bu tez çalışması sırasında elde edilmemiş başkalarına ait tüm orijinal bilgi ve sonuçlara atıf yapıldığını da beyan ederim.

İsmail BAĞCI

01/08/2017

ÖZET
KİSMİ EN KÜÇÜK KARELER YÖNTEMİNİN SİMÜLASYON VERİLERİ
İLE DİĞER YÖNTEMLERLE KARŞILAŞTIRILMASI

İsmail BAĞCI

Yüksek Lisans Tezi

Fen Bilimleri Enstitüsü

İstatistik Anabilim Dalı

Danışman: Doç.Dr. Atila GÖKTAŞ

Ağustos 2017, 59 sayfa

Çoklu bağlantı mevcut olduğunda tahmin edilen doğrusal regresyon modelinin anlamlı bağımsız değişkenlere ilişkin katsayıların testinde kullanılan t istatistikleri anlamsız olmaktadır. Bu sorunun giderilmesinde kullanılan çok sayıda tahmin yöntemi bulunmaktadır. Çalışmamızda, bağımsız değişkenler arasında çoklu bağlantıya sahip çok değişkenli normal dağılımdan üretilen verilere uygulanan “Kısmi En Küçük Kareler Tahmin Yöntemi (KEKK)”, “Ridge Regresyon (RR)” ve “Temel Bileşenler Regresyonu (TBR)” yöntemleri karşılaştırılmıştır. Simülasyon çalışmasında veriler, farklı ilişki düzeylerinde (0.0, 0.3, 0.5, 0.7, 0.9), örneklem büyüklükleri 30, 50, 100, 200 ve 500 birimlik örneklem için farklı değişken sayılarında çalışma 5000 kere tekrarlanmıştır. Elde edilen bu veriler kullanılarak üç farklı yöntem ile tahmin edilen regresyon modeli için Hata Kareler Ortalamaları (HKO) hesaplanmış, düşük olan HKO değerleri dikkate alınarak, hangi yöntemin hangi durumda daha verimli ve daha iyi sonuçlar verdiği araştırılmıştır.

Simülasyon verileriyle elde edilen bulgulara göre örneklem büyüklüğünün azalması veya artması, tahmin yöntemleri üzerinde önemli bir etki yaratmaktadır. Her örneklem büyüklüğü veya her değişken sayısında anlamlı üstünlük sağlayan bir tahmin yöntemi yoktur. Her tahmin yöntemi örneklem büyüklüğü, bağımsız değişken sayısı veya çoklu bağlantının derecesinden etkilenmektedir. Ancak süper çoklu bağlantı sorununda, bağımsız değişken sayısı ne olursa olsun literatürün aksine ($n \leq 200$ için) TBR yönteminin diğer iki tahmin yöntemine göre daha iyi sonuçlar verdiği gözlenmiştir.

Anahtar Kelimeler: Kısmi En Küçük Kareler, Ridge Regresyon, Temel Bileşenler Regresyon, Çoklu Bağlantı

ABSTRACT
**COMPARISON OF PARTIAL LEAST SQUARES PREDICTION AND
OTHER PREDICTION METHODS WITH SIMULATED DATA**

İsmail BAĞCI

Master of Science (M.Sc.)
Graduate School of Natural and Applied Sciences
Department of Statistical
Supervisor: Assoc. Prof. Dr. Atila GÖKTAŞ
August 2017, 59 pages

When there is multicollinearity, using t test statistics for testing the coefficients related to meaningful independent variable of predicted linear regression is meaningless. To eliminate this problem there are great number of prediction methods used. In our study Partial Least Squares Prediction method (PLS), Ridge Regression (RR) and Principal Components Regression (PCR), which are applied to generated data with multicollinearity among independent variables existing in multiple independent variables from standard normal distribution, are compared. The study was repeated 5000 times in simulation for data in different degree of multicollinearity levels (0.0, 0.3, 0.5, 0.7, 0.9) and for different sample sizes 30, 50, 100, 200 and 500 unit samples. Three prediction regression methods were applied by the help of the gathered data and Error Mean Squares (EMS) of regression parameters were calculated. The lowest EMS was taken into consideration to determine which method was the most fructiferous and had the best results under different circumstances.

According to findings gathered through the Simulated data, increase and decrease in the sample size creates important effect on the predicting methods. There is not a prediction method that has a meaningful superiority to the others in every sample size or every variable number. Each prediction method is affected by the size of the sample, number of independent variables or the degree of the multicollinearity. However in super multi connection problem, whatever the number of dependent variable is, in contrast to literature (for $n \leq 200$), it is observed that PCR method had better results compared to the other two prediction methods.

Key Words: Least Squares, Ridge Regression, Principal Components Regression, Multicollinearity

ÖNSÖZ

Çalışmamda yardım ve katkılarıyla beni yönlendiren ve desteğini esirgemeyen danışman hocam Doç.Dr. Atila GÖKTAŞ' a ve sabırla manevi desteğini esirgemeyen eşim Hatice BAĞCI'ya, kızlarım Naz BAĞCI ve İrem BAĞCI'ya çok teşekkür ederim.



İÇİNDEKİLER

ÖZET.....	iv
ABSTRACT	v
ÇİZELGELER DİZİNİ	ix
ŞEKİLLER DİZİNİ	x
1. GİRİŞ.....	1
2. LİTERATÜR ÇALIŞMALARI	3
3. GENEL BİLGİLER	9
3.1. Çoklu Bağlantının Nedenleri.....	10
3.1.1. Örneklem yöntemi.....	10
3.1.2. Aşırı tanımlanmış modeller.....	10
3.1.3. Model ve anakütle üzerindeki fiziksel kısıtlar	10
3.2. Çoklu Doğrusal Bağlantı Probleminin Sonuçları.....	11
3.3. Çoklu Doğrusal Bağlantı Probleminin Çözümü.....	11
3.4. Çoklu Bağlantıyı Belirleme Yöntemleri.....	12
3.4.1. Bağımsız değişkenler arasındaki belirleme katsayılarının belirlenmesi .	13
3.4.2. Korelasyon matrisi ile belirlenmesi	13
3.4.3. Varyans şişirme faktörünün kullanılması(variance inflation factor-vif).	13
3.4.4. Özdeğer ve özvektörlerin incelenmesi	13
3.4.5. Modele yeni bağımsız değişken eklenmesi.....	14
3.5. Kısmi En Küçük Kareler Yöntemi (KEKK)	14
3.5.1. Kısmi en küçük kareler	15
3.6. Ridge Regression (RR).....	17
3.6.1. Ridge kestiricisinin en küçük kareler kestiricisi ile ilişkisi.....	18
3.6.2. Z ve W matrislerinin özellikleri	19
3.6.3. Ridge kestiricisinin beklenen değeri	20
3.6.4. Ridge tahmin edicinin bazı özellikleri	20
3.6.5. k'nın nokta tahmini yöntemi	22
3.7. Temel Bileşenler Regresyonu (TBR).....	24
3.7.1. Temel bileşen sayısının belirlenmesi	29
4. UYGULAMA	30
4.1. Örnek Uygulama	30

4.2. Simülasyon Sonuçları.....	32
5. SONUÇLAR VE TARTIŞMA	47
KAYNAKLAR	49
Ek A. Simülasyon Çalışması Kullanılan Kodlar	52
ÖZGEÇMİŞ.....	59



ÇİZELGELER DİZİNİ

Çizelge 4. 1. $n=30$, $x=7$ ve $p=0.9$ iken elde edilen simülasyon sonuçları.....	30
Çizelge 4. 2. En Küçük Kareler Yöntemine Göre Regresyon Analiz Sonuçları.....	31
Çizelge 4. 3. EKK Sonucu Elde Edilen Çoklu Regresyon Varyans Analiz Sonuçları	31
Çizelge 4. 4. Değişken sayısı: 4-7-10, Örneklem Büyüklüğü:30-40-100-200-500, İlişki Düzeyi:0, 0.3, 0.5, 0.7, 0.9 iken 5000 denemede elde edilen ortalama HKO değerleri.....	34



ŞEKİLLER DİZİNİ

Şekil 3. 1. $\hat{\beta}(k)$ ve $\hat{\beta}$ SMSE'in k 'ya Bağlı Grafiği(Hoerl ve Kennard,1970)	22
Şekil 4. 1. Örneklem Büyüklüğü:30 iken elde edilen HKO değerleri	36
Şekil 4. 2. Örneklem Büyüklüğü:50 iken elde edilen HKO değerleri	38
Şekil 4. 3. Örneklem Büyüklüğü:100 iken elde edilen HKO değerleri	40
Şekil 4. 4. Örneklem Büyüklüğü:200 iken elde edilen HKO değerleri	42
Şekil 4. 5. Örneklem Büyüklüğü:500 iken elde edilen HKO değerleri	44

KISALTMALAR DİZİNİ

EKK	En Küçük Kareler
HKO	Hata Kareler Ortalaması
KEK	Kısmi En Küçük
KEKK	Kısmi En Küçük Kareler
n	Örneklem Büyüklüğü
RR	Ridge Regresyon
TBR	Temel Bileşenler Regresyonu

1. GİRİŞ

Doğrusal regresyonda, bir veya birkaç bağımsız değişkenin arasında ortaya çıkan anlamlı ilişki, çoklu bağlantı olarak adlandırılmaktadır.

Regresyon analizi, bağımlı bir değişken (Y_i) ile bir veya birden fazla bağımsız değişken (X_i) arasındaki nicel ilişkiyi tahmin etmek amacıyla kullanılan en yaygın istatistiksel uygulamadır. Bağımsız değişken sayısı 1 (bir) tane ise bu duruma basit regresyon, birden daha fazla değişken var ise çoklu regresyon olarak adlandırılır. Regresyon analizinin genel kullanım amaçları:

- a) Önceden ölçülebilen değerlerden yararlanarak, daha sonra ölçülecek ölçüm değerlerini,
- b) Kolay ölçümü yapılabilen veri setinden yararlanarak, ölçümü zor olan bir özelliğin değerlerini
- c) Düşük maliyete sahip ölçüm değerlerinden, yüksek maliyetli ölçüm değerlerini tahmin etmek şeklinde sıralanabilir (Huber ve Dutter, 1974).

Regresyon analizinin yapılmasında gerekli bir takım varsayımları sağlaması koşuluyla en yaygın kullanılan yöntem, En Küçük Kareler (EKK) yöntemidir. Bu yöntem ile denklemin verdiği (teorik) Y değerleri ile ölçümlerden elde edilen (gerçek) X değerleri arasındaki farkların karelerinin toplamını küçültme fikrine dayanmaktadır. Elde edilen bu modelin güvenilirliği EKK yönteminin varsayımlarının sağlanmasına bağlıdır. İncelenen bağımsız değişkenler arasında çoklu bağlantı olması, EKK ile tahmin edilen regresyon parametrelerine ait katsayılar, sonuçların yanlış yorumlanmasına neden olabilir.

Ridge regresyon, regresyon analizinde çoklu bağlantı olması durumunda parametre tahminleri üzerinde oluşacak olan olumsuz etkinin ortadan kaldırılmasını sağlamak amacıyla geliştirilen yanlı bir regresyon yöntemidir.

Temel Bileşenler Regresyon ise aralarında korelasyon bulunan orijinal değişkenlerinden daha az sayıda ve bu değişkenlere ait doğrusal bileşenleri olan yeni değişkenler yardımı ile açıklayan bir regresyon yöntemidir.

Tarım, sosyo-ekonomi, tıp ve biyoloji gibi alanlarda elde edilen sonuçların gerekli varsayımlar dikkate alınmadan EKK yöntemi ile tahmin edilmesi yanlış sonuçlar ortaya çıkarabilmektedir. Bu durumda EKK yöntemi ile elde edilmiş geçerliliklerden şüphe edilmelidir (Alpar, 1997). Bu nedenle yapılan çalışmalarda istatistik alanında önemli bir ölçüt olan çoklu bağlantı varlığı ve çözüm tekniklerine gereken önem verilmemiştir.

Bu çalışmada doğrusal regresyon modelinde yer alan açıklayıcı değişkenler arasında çoklu bağlantı varlığında yanlış tahmin yöntemlerinden “Kısmi En Küçük Kareler” yönteminin diğer yanlış tahmin yöntemleri ile karşılaştırılması amaçlanmaktadır. Karşılaştırma analitik olarak yapılamadığından söz konusu karşılaştırma simülasyon verileri ile gerçekleştirilmiştir. Sonuç olarak bağımsız değişken sayısının çokluğunda yanlış tahmin yöntemleri arasında en iyi tahmin yönteminin “Kısmi En Küçük Kareler Yöntemi” olacağı beklenmektedir.

2. LİTERATÜR ÇALIŞMALARI

En Küçük Kareler Yöntemi, çeşitli alanlarda herhangi bir uygulama ile elde edilmiş verileri modelleyen bir fonksiyon bulmaya çalışır. Genellikle bu kullanılan veri tablosuna tam anlamıyla uyan bir fonksiyon bulmak pek mümkün olmadığı için; veri tablosuna en uygun olabilecek fonksiyon belirlenmeye çalışılır. Bir veri tablosuna en uygun fonksiyonu bulma sürecine regresyon analizi denir.

Doğrusal regresyon analizinde bağımsız değişkenler sabit, bağımlı değişken tesadüfidir. Değişkenler hatasız ölçülmekte, bağımsız ve bağımlı değişkenler arasında var olan ilişki doğrusal olduğu varsayılmaktadır. Ancak bu doğrusallık kriteri, modele ait parametreler için gereklidir. Bağımsız değişkenler arasında güçlü ilişkilerin var olmasına bağlantı (collinearity) veya çoklu doğrusal bağlantı (multicollinearity) olarak ifade edilmektedir. Bu, regresyon analizinde istenmeyen bir durumdur (Orhunbilge, 2000: 240-251).

Regresyon analizlerinde çoklu doğrusal bağlantının olması, çeşitli problemlere neden olmaktadır (Orhunbilge, 2000:240-251). Çoklu doğrusal bağlantı problemi, çoklu regresyon analizindeki her bir gözlem için, bağımsız değişkenlerden biri ya da birkaçı arasında, tam veya tama yakın doğrusal bir ilişki olması durumudur. Ancak uygulamaların hemen hemen tamamında bağımsız değişkenler arasında ilişki olmaması çok sık rastlanan bir durum değildir. Genellikle bağımsız değişkenler arasında oldukça düşük düzeyde bir ilişki vardır. Eğer ilişki doğrusal bir ilişki ise, çoklu doğrusal bağlantı vardır.

Kidwell ve Brown (1982), çoklu bağlantı üzerine bir çalışma yapmıştır. Bu çalışmada RR yöntemine yönelik yapay veriler kullanmıştır. Elde edilen sonuçlar doğrultusunda, tahminleyicilerin ortogonal olmadığı durumda RR modelinin, en küçük kareler yöntemine göre daha farklı sonuçlar verdiği sonucuna varmıştır.

Kurtuluş (2001) yapmış olduğu yüksek lisans tez çalışmasında, bağımsız değişkenler arasında çoklu bağlantı sorunun var olması durumunda, sorunun ortadan kaldırılmasına yönelik kullanılan RR yöntemi üzerinde çalışma yapmıştır. Bu yöntemi ile en küçük kareler yöntemini karşılaştırmıştır. RR metodunda k parametresinin elde

edilmesini saęlayan birkaç yöntemi tanıtmıştır.

Ortabaş (2001) yaptığı yüksek lisans tez çalışmasında çoklu doğrusal bağlantı sorununu ortadan kaldırmaya yönelik olarak kullanılan temel bileşenler ve RR 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 tahminlerinin bulunabildiğini ifade etmiştir.

Topçubaşı ve Billor (2003), yaptıkları çalışmada, açıklayıcı değişkenlerin doğrusal bağımlı olmaları durumunda, en küçük kareler tahmin edicisine alternatif olabilecek çeşitli yanlı tahmin modellerin bulunduğunu ve bu yanlı tahmin modellerinin çoğunu (genelleştirilmiş ridge, ridge, ana bileşenler, ondalıklı rank ve Stein kestiricisi) barındıran bir sınıf tanımlandığını ifade etmişlerdir. Çalışmalarında yanlı kestiricilerin bir kısmını içeren bu sınıf içerisine Liu ve genelleştirilmiş Liu kestiricileri de katmışlardır.

Graham (2003) yapmış olduğu çalışmada ekolojik verilere uyguladığı çoklu regresyon analizlerinde, çoklu bağlantı sorunu ile karşılaşıldığını belirtmiştir. Çalışmasında gerçek ekolojik veriler kullanarak çoklu bağlantı sorununa karşılık farklı istatistiksel tekniklerin kullanımını göstermiştir. Bunun sonucu olarak, ekolojik verilerin çoklu regresyon modellerinde çoklu bağlantılı durumların açıklanmasını ve farklı yöntemlerin kullanılmasıyla modelin geçerliliğinin artırılabilceğini ifade etmiştir.

Ergüneş (2004) yüksek lisans tez çalışmasında en küçük kareler yöntemi ile RR yöntemlerini karşılaştırarak çoklu bağlantı problemini anlatmış ve problemin ortadan kaldırılmasına yönelik Ridge Regresyon (RR) yönteminin kullanımını ele almıştır. Çoklu bağlantı sorunu bulunan bir örneğe En Küçük Kareler ve RR yöntemlerini uygulayarak, daha geçerli parametre tahmini yapmasından dolayı, RR yönteminin EKK yönteminin yerine kullanılmasını önermiştir.

Albayrak (2005), yaptığı çalışmada beden ağırlığının tahmin edilmesinde RR ve Temel Bileşenler Regresyonun, En Küçük kareler yöntemine karşı etkin olup olmadığı araştırılmıştır. Aralarında yüksek çoklu doğrusal bağlantı bulunan bağımsız değişkenlere RR, TBR ve EKK yöntemleri uygulanmıştır. RR ve TBR yöntemlerinin EKK yöntemine göre daha düşük standart hatalı, durağan ve kurumsal beklentileri karşılayabilecek uygun tahminler sağlayacağı gözlenmiştir.

Karadavut ve ark. (2005) yaptıkları çalışmada nohut bitkisinin verimini etkileyen bazı karakterleri, ridge regresyon, en küçük kareler ve robust regresyon yöntemlerinden biri olan M-regresyon yöntemleri ile elde etmiş oldukları parametre tahminlerini karşılaştırmışlardır. Nohut bitkisinde tane ağırlığını etkileyen değişkenlere etki eden regresyon modelinde, ilk önce EKK yöntemini kullanarak parametreleri tahmin edilmiş, bağımsız değişkenler arasındaki çoklu bağlantı tespit edilerek RR yöntemi ile parametre tahminlerini elde etmişlerdir. Yine aynı veri grubuna M-regresyon yöntemini de uygulayarak, elde edilen parametre tahminleri ve analiz sonuçları karşılaştırılmıştır. Sonuç olarak, M regresyon yönteminin diğer yöntemlere göre daha uygun bir tahmin edici olduğu ve tercih edilebileceği ifade edilmiştir.

Yolaçan ve ark. (2005) yaptıkları çalışmalarında çoklu bağlantı probleminin bulunduğu ekonomik verilere RR ve yapay sinir ağıları algoritmalarını uygulamış, sonuçlarını karşılaştırmıştır.

Çankaya ve ark. (2006) yapmış oldukları çalışmada, çoklu lineer regresyon modelinde parametre tahmin yöntemlerini karşılaştırmıştır. Parametre tahminlerinde aykırı değer ortaya çıktığında, en küçük medyan kareler yönteminin diğer yöntemlere göre en yüksek belirtme katsayısına sahip olduğunu ifade etmiştir. Aykırı değerlerin ve yüksek uç değerlerin etkilerinin araştırılması gerektiğini ifade etmişlerdir.

Aktaş (2007) yaptığı çalışmada, çoklu regresyon analizinde karşılaşılan sorunlardan olan çoklu bağlantıyı belirleme tekniklerini ele almış ve enflasyon modeli için Liu kestiricisi ile bir uygulama yapmıştır. Liu kestiricisi ile çoklu bağlantı sorunun kaldırıldığı modele göre, bahsedilen dönemde enflasyon üzerinde en fazla USD ve para arzı değişkenlerinin etkili olduğu belirtilmiştir.

Karakaş (2008) yüksek lisans tez çalışmasında çoklu doğrusal bağlantının nedenleri, teşhis yöntemleri ve istatistiksel sonuçlara varılması süreci üzerindeki etkileri incelemiştir. Özellikle Ridge ve Liu regresyon metodları başta olmak üzere, yanlış regresyon tahmincilerinin çoklu doğrusal bağlantı sorununu düzeltme yöntemi olarak değerlendirilmekte ve istatistiksel özellikleri incelenmektedir. Ridge ve Liu tip regresyon kullanılarak bir istihdam modelinin tahminini elde etmeye yönelik bir uygulama yapılmıştır. İstatistikte, yanlış regresyon tahmincilerine ait standart hataların hesaplanabilmesi için bir formülün bulunmadığını belirtmiştir. İşaretlerin uygunluğu ve parametrelerin anlamlılığı bakımından, Liu tip tahmincinin diğer iki tahminciden daha üstün olduğunu ifade etmiştir.

Çankaya ve arkadaşları (2009) yaptıkları çalışmalarında, Karayaka kuzularına ait bazı vücut ölçülerinden hesaplanmış faktör analiz skorlarını, çoklu regresyon modelinde kullanarak, canlı ağırlığı tahmin etmek ve incelenen vücut ölçüleri arasındaki çoklu bağlantıyı ortadan kaldırmaya çalışmışlardır. Ondokuz Mayıs Üniversitesi Araştırma ve Uygulama Çiftliğinde yetiştirilen sütten kesime döneminde olan 101 adet Karayaka kuzusunun vücut ölçüleri (cidago ve sağrı yüksekliği, vücut uzunluğu, göğüs çevresi ve derinliği, orta ve arka sağrı genişliği, kürekler arası genişlik) ile canlı ağırlık ölçülerinden faydalanmışlardır. Tahmin eşitliğinde kullanılan cidago ve sağrı yükseklik değerlerinde çoklu bağlantı olduğu belirlenmiştir. Faktör analiziyle elde edilen faktör skorlarını kullanıp, çoklu bağlantı problemini ortadan kaldırdıklarını belirtmişlerdir.

Pamukçu (2010) yaptığı yüksek lisans tez çalışmasında, sistolik kan basıncı değerini tahmin etmede RR ile En küçük kareler yöntemleri karşılaştırılmıştır. Buna yönelik olarak, sistolik kan basıncı ile bağımsız değişkenler arasındaki doğrusal ilişkinin tahmininde En küçük kareler regresyonu ve RR tekniklerinin göreceli tahmin geçerlilikleri karşılaştırılmıştır. Yapılan araştırmada, aralarında yüksek çoklu doğrusal bağlantı bulunan bağımsız değişkenler kullanılarak yapılan çalışmada, RR tekniğinin en küçük kareler regresyonuna göre daha uygun tahminler ortaya koyduğu belirlenmiştir.

Topal ve arkadaşları (2010) araştırmalarında, farklı yaşlarda 91 adet sazan balığına ait çeşitli vücut ölçülerini kullanarak karkas ağırlıklarını tahminleyen bir model geliştirmeyi amaçlamışlardır. Elde edilen vücut ölçüleri arasında ortaya çıkan çoklu bağlantı sorununu ortadan kaldırmak için en küçük kareler regresyonu, RR ve temel bileşenler regresyon analiz yöntemleri kullanılmıştır. Elde edilen sonuçlara göre, çoklu doğrusal bağlantı sorununu ortadan kaldırmak için, en küçük kareler metodu yerine Ridge ve temel bileşenler regresyon yöntemlerinin kullanılmasının daha doğru olabileceği belirtilmiştir.

Büyükuysal (2010) yaptığı yüksek lisans tez çalışmasında, çoklu doğrusal regresyon analizinde, bağımsız değişkenler arasındaki çoklu doğrusal bağıntı durumunda ortaya çıkan yanlılığı ortadan kaldırmak için en küçük kareler yöntemine alternatif olarak önerilen RR yöntemi üzerinde çalışılmıştır. En küçük kareler ve RR analizinden elde edilen sonuçları karşılaştırmıştır. İstanbul Tıp Fakültesi Hastanesine şişmanlık şikâyeti ile başvurmuş olan hastalardan rasgele seçilen 20 kişinin beden ağırlığı (kg), deri alanı

(cm²), uyluk kemiğinin çevresinin uzunluğu (cm) ve belden yukarı ölçülen kasların çevrelerinin uzunluğu (cm) değerleri elde edilerek, bu değerler ile hastaların vücut ağırlığıyla olan ilişkileri incelenmiştir. Analizler sonucunda, RR analiziyle elde edilen regresyon katsayılarının, en küçük kareler yöntemiyle elde edilen regresyon katsayılarına göre daha iyi sonuçlar verdiği gözlenmiştir. Elde edilen sonuçlar dikkate alınarak, çoklu doğrusal bağıntı halinde, RR analizinin en küçük kareler yöntemine göre daha doğru sonuçlar verdiği görülmüştür. Çoklu doğrusal bağıntı varlığında, RR analizinin kullanılmasını önermiştir.

Rathert ve arkadaşları (2011) yaptıkları çalışmada, Japon bıldırcınlarında yumurta iç kalite özelliklerinin tahmin edilmesine yönelik olarak, en küçük kareler yöntemi ile temel bileşenler regresyon yöntemini karşılaştırmışlardır. Çalışma kapsamında 20-24 haftalık yaşlarda olan Japon bıldırcınlarından toplanan yumurtaları kullanmışlardır. Sonuç olarak, temel bileşenler regresyon yöntemi ile çoklu problemin ortadan kaldırıldığı ifade edilmiştir.

Eker (2011) yaptığı yüksek lisans tez çalışmasında, çoklu regresyon analizinde bağımsız değişkenler arasında çoklu bağlantı sorununun var olması durumunda, çokça tercih edilen En Küçük Kareler (EKK) metodunun varsayımlarının savunulamaz olmasından dolayı, EKK yöntemi ile Ridge Regresyon (RR) ve Temel Bileşenler Regresyonu (TBR) yöntemlerini karşılaştırarak incelemiştir. Çalışmasında, Ondokuz Mayıs Üniversitesi Araştırma ve Uygulama Çiftliğinde yetiştirilen 85 Karkaya kuzusunun bazı vücut ölçüleri ile canlı ağırlık ölçülerini kullanmıştır. Bu 3 yöntemi değerlendirirken hata kareler ortalaması (HKO) ve parametrelerin önemliliğini (R^2) kullanmıştır. Araştırmanın sonucunda, en düşük HKO değerinin ve en büyük R^2 değerlerinin RR modeli ile elde edilmiş olmasına rağmen, parametrelerin önem testlerine göre Temel Bileşenler Regresyon yönteminin daha tutarlı olduğunu ifade etmiştir.

Mahesh ve arkadaşları (2014) yaptıkları çalışmalarında, farklı yerlerden ve ekin zamanlarında elde edilen Kanada buğdayının protein içerikleri ve sertlik değerleri ile ilgili tahminleri karşılaştırmışlardır. Bu karşılaştırma işlemleri için EKK ve TBR modelleme yöntemleri kullanılmıştır. Çalışmanın sonucunda EKK modelleri, protein içerikleri ve buğday sertliğini öngörmek için kullanılan TBR modellerinden daha iyi tahmin performans sergilediği sonucuna varılmıştır.

Firinguetti ve arkadaşları (2016) yapmış oldukları çalışmada, çok parçalı doğrusallık

problemine yönelik kullanılan Kısmi En Küçük Kareler (KEKK) ile Ridge Regresyon (RR) yöntemlerini ele almışlardır. Performanslarını karşılaştırmak üzere yapılan simülasyon çalışması sonucunda, hata varyansının yüksek olduğu durumlarda Ridge Regresyon (RR) 'un daha iyi performans sergilediği ve modelin daha fazla değişken içerdiği durumlarda Kısmi En Küçük Kareler (KEKK) tahmincisinin en iyi sonuçlara ulaştığını ifade etmişlerdir.



3. GENEL BİLGİLER

Regresyon uygulamalarının pek çoğunda, bağımsız değişkenler arasında ilişki mevcuttur. Bazen bağımsız değişkenler arasındaki doğrusal ilişki çok kuvvetli olmakta ve bu tür durumlarda, regresyon modeli yardımıyla yapılacak yorumlar yanlış olabilmektedir. Aslında çoklu regresyon denklemi yorumlanırken, bağımsız değişkenler arasında kuvvetli bir ilişki olmaması durumuna bağlıdır. Bu durumun bozulması, bir diğer ifade ile bağımsız değişkenler arasında bir ya da daha fazla doğrusal bağıntının var olması çoklu bağlantı (multicollinearity) sorununu ortaya çıkarmaktadır (Alpar 2003).

Çoklu bağlantı, bazı bağımsız değişkenler arasında yüksek oranda korelasyon olması durumudur. Birden çok bağımsız değişken bulunduran regresyon modellerinin pek çoğunda var olan bir durumdur. Çoklu bağlantı var olduğunda, tahminler yansız olsa bile, bağımsız değişkenlerin kuvvetli ilişkisinin değerlendirilmesi ve birlikte etkilerine ait sonuçlara güvenilemez. Temel olarak, eğer bağımsız değişkenler arasındaki korelasyon değeri 0.70'in üstünde ise, bir çoklu bağlantı sorununun olduğu belirtilir.

Tam çoklu bağlantının varlığı, belirlenemeyen katsayıları ve tanımlanamayan standart hata ortaya çıkarmakla beraber, yüksek çoklu bağlantıda daha çok yüksek varyans ve kovaryanslar, geniş güven aralığı ve gerçekte önemsiz olduğu halde, yer aldığı denklemin önemli bulunduğu katsayılara neden olmaktadır.

Regresyon modellerinde de bağımsız değişkenler arasındaki güçlü korelasyondan dolayı, çoklu bağlantı görülür. Çoklu bağlantının var olması, parametrelerin tahminlerinde varyans değerini büyütmektedir. Özellikle küçük ve orta büyüklükteki örnek genişliğine sahip modeller, güçlü bir şekilde istatistiksel olarak önemli bulunurken bağımsız değişkenlerin bireysel olarak önemini azaltır. Çoklu bağlantı, bağımlı ve bağımsız değişkenler arasındaki ilişki hakkında da yanlış sonuçların ortaya çıkmasına neden olabilir.

3.1. Çoklu Bağlantının Nedenleri

Çoklu bağlantı sorunun ortaya çıkması, çok farklı nedenlerden dolayı olabilir. Bu nedeni belirlemek, bize çözüm için bazı ipuçları verebilir. Çoklu bağlantı sorunu, aşağıda verilen bir veya birkaç sebepten dolayı olabilir.

3.1.1. Örnekleme yöntemi

Verilerin elde ediliş şekli; araştırmacı isteyerek ya da istemeyerek bağımsız değişkenler uzayından bir alt uzayı örnekleme alması durumunda çoklu doğrusallığa sebep olur. Elde edilen modelin kendisinde çoklu doğrusal bağlantı var olmasa bile, bağımsız değişkenler içerisinde eksik ya da yetersiz bir alt kümenin seçilmesi nedeniyle bir çoklu bağlantı söz konusu olabilir.

3.1.2. Aşırı tanımlanmış modeller

Eğer bağımsız değişken sayısı, gözlem sayısından büyük ise modellerde çoklu doğrusal bağlantı ortaya çıkmaktadır. Genelde bu tür durumlarda çoklu bağlantının ortadan kalkması için, bazı değişkenleri modelden çıkarmak ya da gözlem sayısının artırılması gerekmektedir.

3.1.3. Model ve anakütle üzerindeki fiziksel kısıtlar

Model ya da anakütle üzerinde kısıtlamalar yapmak, çoklu doğrusal bağlantıya neden olabilir. Örneğin elektrik tüketimi ile ilgili olarak, yaşanılan evin büyüklüğü ve hane gelirinin elektrik tüketimi üzerindeki etkisinin araştırıldığı bir çalışma düşünelim. Bağımsız değişken olan evin büyüklüğü ile hane geliri arasında çoklu doğrusal bağlantı olduğu görülür. Çoklu doğrusal bağlantının nedeni, hane geliri yüksek ailelerin daha geniş evlerde oturuyor olduğu gerçeğidir. Bundan dolayı çoklu doğrusal bağlantının kaynağı, anakütlede var olan gerçek ilişkinin örnekleme de var olmasıdır.

3.2. Çoklu Doğrusal Bağlantı Probleminin Sonuçları

Bağımsız değişkenler arasında var olan güçlü ilişkiler olması bağlantı (collinearity) ya da çoklu doğrusal bağlantı (multicollinearity) olarak adlandırılır. Bu, regresyon analizlerinde istenmeyen bir durumdur (Orhunbilge, 2000:240-251).

Regresyon analizinde çoklu doğrusal bağlantı aşağıda belirtilen problemlere neden olur (Gujarati, 1995:319-399; Orhunbilge, 2000:240-251):

1. Tam çoklu doğrusal bağlantının var olması, hem regresyon katsayılarının belirsiz, hem de bu katsayıların standart hatalarının sonsuz olmasını sağlayabilir.
2. Çoklu doğrusal bağlantı durumunda regresyon katsayılarının varyans ve kovaryansları artış göstermektedir.
3. Modelin R^2 değeri yüksek çıkmasına rağmen, bağımsız değişkenlerden hiçbiri veya çok azı t testine göre anlamlı çıkmaktadır.
4. Bağımsız değişkenler ile bağımlı değişkenin ilişkilerinin yönü kuramsal ve deneysel beklentilerle farklılık gösterebilir.

Bağımsız değişkenler arasında bağıntı var ise, bazılarının modelden çıkarılması gerekebilir. Ama modelden yanlış değişkenin çıkarılması, modelin yanlış tanımlanmasına neden olabilir. Modele bağımsız değişkenleri dâhil edip çıkarmak gibi kullanabileceğimiz kurallar bulunmamaktadır.

3.3. Çoklu Doğrusal Bağlantı Probleminin Çözümü

Çoklu bağlantı sorunun ortadan kaldırılması, çoklu bağlantının nedenine, derecesine ve regresyon modelinin kullanım amacına bağlı olarak değişebilir. Bu sorunun ortadan kaldırılması için önerilen birçok yöntem vardır. Ancak gözlem sayısını arttırmak, modeli yeniden belirlemek ve küçük kareler yönteminden başka yanlı kestirim yöntemlerini kullanmak en genel yaklaşımlardandır. Çoklu bağlantıyı ortadan kaldırmak için izlenen diğer bir yol, değişken seçimidir. Ancak bu yöntemle, bağımlı değişkeni iyi açıklayan değişkenlerin çıkarılması durumunda, değişken seçimi iyi bir yol olmayabilmektedir. Bir diğer yöntem ise, birbiriyle ilişkili olan iki değişkenin yerine, bu iki değişkenin toplamının tek bir değişken olarak modele alınmasıdır.

Çoklu doğrusal bağlantı problemlerinin çözüm yolları aşağıda verilmektedir.

1. Bir yada daha fazla bağımsız değişken modelden çıkartılabilir. Ancak bu yöntem, modeli yanlış tanımlattırabilir.
2. Gözlemler arası farklar alınarak değişkenler dönüştürülebilir. Böyle bir yöntem sadece zaman serilerine uygulanabilirken, kesit verilerine uygulanamamaktadır.
3. Bazı durumlarda yeni gözlem değerlerinin elde edilmesi, çoklu doğrusal bağlantı probleminin ortadan kalkmasına neden olabilir. Ancak, örneği büyütmek her zaman mümkün olmayabilir.
4. Birbirleri ile ilişkili olan iki değişken yerine, bu iki değişkenin doğrusal birleşimi alınarak tek bir değişken alınabilir.
5. Birbirinden bağımsız şekilde bileşenler türeten “Temel Bileşenler Regresyon Analizi” ya da “En Küçük Kareler” tekniğinin düzeltilmiş şekli olan ve yanlış standartlaştırılmış regresyon katsayılarını tahmin eden “Ridge Regresyon” teknikleri kullanılabilir.
6. Temel bileşenler analizindeki gibi çoklu bağlantı ilişkisindeki bağımsız değişkenler yerine yeni bileşenler elde ederek kullanılmasına olanak veren KEKK yöntemi de kullanılmaktadır. TBR’den farkı bağımlı değişkeni de dikkate alarak bu bileşenler elde edilmektedir.

3.4. Çoklu Bağlantıyı Belirleme Yöntemleri

Daha önce yapılmış olan uygulamalarda, çoklu bağlantının ortaya çıkması yönünde bilgiler veren bazı durumlar saptanabilir. Modele bir bağımsız değişken eklenmesi veya çıkartılması, bir gözlemin değiştirilmesi durumlarında, regresyon katsayıları tahminlerinde önemli değişimler bulunuyorsa çoklu doğrusal bağlantıdan kuşku duyulur. İşaretine ve büyüklüğe göre regresyon katsayılarının beklenenden farklı olması, regresyon katsayılarının istatistiksel olarak önemsiz çıkması, regresyon katsayıları tahminlerine ait güven aralıklarının geniş olması, çoklu bağlantı sorununun belirtilerindendir. Çoklu bağlantıyı belirlemeye yönelik birçok yöntem kullanılmaktadır. Bu yöntemler şunlardır:

3.4.1. Bağımsız değişkenler arasındaki belirleme katsayılarının belirlenmesi

Çoklu bağlantı, bağımsız değişkenler arasındaki ilişkiden dolayı ortaya çıkmaktadır. İlişkinin derecesi, bağlantının derecesi hakkında karar verilebilir. Her bağımsız değişkenin, diğer bağımsız değişkenlerle arasındaki korelasyon katsayılarının karesi yani belirleme katsayısı değeri 1'e yakın çıktıkça, yüksek derecede çoklu bağlantı olduğu ifade edilir.

3.4.2. Korelasyon matrisi ile belirlenmesi

Bu yöntem, çoklu bağlantının ortaya çıkarılmasında kullanılan en yaygın ve basit yöntemdir. İki bağımsız değişken arasındaki korelasyon katsayısının mutlak değerinin 1'e yakın olması, iki bağımsız değişkenin yakın doğrusal bağlantılı olduğunu ifade eder. Eğer ikiden fazla bağımsız değişken arasında bağımlılık varsa, bu durum korelasyon katsayılarına yansıyamayabilir.

3.4.3. Varyans şişirme faktörünün kullanılması(variance inflation factor-vif)

Bağımsız değişkenlere ait korelasyon matrisinin tersinin köşegen değerlerine Varyans Şişirme Değerleri denmektedir.

Çoklu doğrusal bağlantının ortaya çıkarılmasında kullanılan en yaygın ve kullanışlı yöntemlerden birisi de varyans şişirme faktörü (VIF) değerlerinin yorumlanmasıdır.

$$VIF = \frac{1}{1-R_j^2} \quad (3.1)$$

Denklemden R_j^2 değerinin sıfıra eşit olması durumunda VIF_j değeri de 1 olacaktır. R_j^2 değeri 1'e yaklaştığında ise VIF_j değeri sonsuza yaklaşır. Bu durum, tam çoklu bağlantının göstergesidir. VIF değeri 10'dan büyük olursa, güçlü çoklu bağlantı olduğu ifade edilir.

3.4.4. Özdeğer ve özvektörlerin incelenmesi

Öz vektör, değişkenlerin doğrusal bileşenleri olarak tanımlanır. p adet değişken için p adet öz vektör vardır. Öz değer kavramı ise öz vektörlerce açıklanan varyans olarak

bilinmektedir. $(X'X)$ matrisinin öz değeri göz önüne alınarak çoklu bağlantının derecesi konusunda bilgi sahibi olunur.

Sıfırdan farklı öz değeri sayısı $(X'X)$ matrisinin rankına eşit olmaktadır. $(X'X)$ matrisinin özdeğerlerinden bir tanesi sıfıra eşit olduğunda, matris tekil matris olarak ifade edilir. $(X'X)$ matrisinin özdeğerlerden birinin sıfıra eşit ya da çok küçük ($\lambda_j < 0.001$) özdeğerlere sahip olması, veride doğrusal bağımlılık söz konusu olduğu anlamı taşır.

3.4.5. Modele yeni bağımsız değişken eklenmesi

Modele yeni bağımsız değişken eklendikten sonra, R^2 değeri önemli bir gelişme sağlanmazsa, çoklu doğrusal bağlantı probleminin ortaya çıktığı söylenebilir.

3.5. Kısmi En Küçük Kareler Yöntemi (KEKK)

Kısmi En Küçük Kareler yöntemi, 1960'lı yıllarda Helman Wold tarafından geliştirilen ve özellikle bağımsız değişken sayısının fazla ve örnek sayısının az olduğu durumlarda oldukça kullanışlı olan bir yöntemdir. KEKK regresyonu ile bağımsız değişkenler ile bağımlı değişkenler arasındaki doğrusal ilişki tanımlanmaktadır.

KEKK yöntemiyle veri kümesinde yer alan bağımlı ve açıklayıcı değişkenler arasındaki kovaryansı en çoklayan bileşen sayısını bulmak amaçlanmaktadır. Bu yöntemde, analiz işlemleri yapılmadan önce bütün değişkenlerin birbirleri ile aynı önem sırasında olmaları için merkezileştirme, sonrasında ise ölçeklendirme işlemi yapılmaktadır. Merkezileştirme işlemi için, bağımlı ve açıklayıcı tüm değişkenlerden ortalama değerlerinin çıkartılması işlemi yapılır. Açıklayıcı değişkenlerin boyutunun indirgenmesi ile elde edilen gizli değişkenler, açıklayıcı değişkenlerin doğrusal birleşimidirler. Gizli değişkenler sayesinde bağımlı değişkenleri açıklayıcı değişkenler ile döndürme mümkün olmaktadır. Bunun dışında gizli değişkenler yeni açıklayıcı değişkenler olarak regresyon analizi için de kullanılmaktadırlar.

3.5.1. Kısmi en küçük kareler

KEKK varsayımı, Ridge ve TBR varsayımlarındaki gibi değişkenlerin ölçeklendirilmesi yöntemi gibidir. Yorumlamanın daha kolay olabilmesi ve sayısal kararlılığa varabilmek için, analiz yapmadan önce verinin merkezileştirilmesi önerilir. Ölçeklendirmeden önce ya da sonra, X ve Y'deki tüm değişkenlerden ortalamalarının farkı alınır. Bu sayede her bir değişkene aynı ağırlık, analizde aynı öncelikli önem verilmiş olunur.

KEKK yöntemi ile Y değişkenlerini $X_1, \dots, X_p$ bağımsız değişkenleriyle döndürmek için, X değişkenleriyle benzer role sahip olan ve genellikle gizli değişkenler olarak tanımlanan yeni bileşenler elde edilir. Bağımsız değişkenler matrisinin bir küme indirgenmiş doğrusal birleşimleriyle yer değiştirmesi yönünden ele alındığında KEKK ile TBR yöntemleri benzerlik taşırlar. KEKK ve TBR yöntemlerinde, Y değişkenleriyle döndürülebilecek bazı bileşenler bulmak adına, X değişkenlerinin sayısından daha az sayıda bileşen kullanılarak, regresyon probleminin boyutluluğu azaltılmaktadır. Her bir bileşen, $X_1, \dots, X_p$ 'nın doğrusal bir birleşimi demektir. KEKK ile TBR arasındaki temel fark, KEKK' de temel bileşenler belirlenirken gözlenen bağımlı değişkenler önemli bir role sahip iken, TBR yönteminde ise temel bileşenler, bağımlı değişkenleri referans olarak kullanmaz.

KEKK yöntemi, Temel Bileşenler Analizine benzer bir yol izleyerek X'in doğrusal ayrışımını verir (3.2). Burada X_{rj} şeklinde yazılan t_j 'ler, X'in doğrusal birleşimidir. $p \times 1$ boyutlu olan p_j 'ler, yükler olarak tanımlanırlar. KEKK tahmin edicilerini bulabilmek için NIPALS ve SIMPLS adında algoritmalar kullanılır. Geleneksel olan NIPALS algoritmasında t_j 'ler E_j artık matrislerinin doğrusal birleşimleri elde edilir (3.3).

$$X = t_1 p_1' + t_2 p_2' + \dots + t_p p_p' = \sum_{j=1}^p t_j p_j' = TP' \quad (3.2)$$

$$t_j = E_{j-1} w_j, \quad E_j = X - \sum_{i=1}^j t_i p_i', \quad E_0 = X \quad (3.3)$$

Buradaki w_j 'ler birimdir. w_j ve r_j , $j = 1, \dots, h$ ağırlık kümeleri aynı uzayda bulunmaktadır. Tek değişkenli ve aynı zamanda çok değişkenli KEKK için birçok agoritmada ilk önce t_j 'nin doğrusal birleşimini hesaplayabilmek için w_j veya r_j 'yi, $i=1, \dots, h$ elde etmektir. Sonra ise X matrisini t_j üzerinde döndürerek, p_j bulunur. h adet

boyut elde edilip aşağıdaki ilişkiler yazılabilmektedir (Phatak ve De Jong, 1997; Yeniay ve Göktaş, 2002).

$$T_h = XR_h \quad (3.4)$$

$$P_h = X' T_h (T_h' T_h)^{-1} \quad (3.5)$$

$$R_h = W_h (P_h' W_h)^{-1} \quad (3.6)$$

Buradaki h alt kümesi, matrisin ilgili vektörlerin ilk h dizisinden oluşmaktadır. İki ağırlık vektörü kümesinin bir doğrusal dönüşüme bağlandığı 3.6. da gösterilmiştir. 3.4 ve 3.5 eşitsizliklerine göre $P_h' R_h = I_h$ olduğu görülmektedir.

Yine, $R_h' P_h$ ifadesinin I_h 'a eşit olduğu görülmektedir:

$$R_h' P_h = R_h' X' T_h (T_h' T_h)^{-1} = T_h' T_h (T_h' T_h)^{-1} = I_h \quad (3.7)$$

h adet boyut elde edildikten sonra KEKK ile elde edilen uygun değerler vektörü, 3.7 eşitsizliğinde gösterildiği gibi gözlenen bağımlı değişkenlerin ilk h KEKK doğrusal birleşimleri olan T_h üzerindeki yansımasıdır.

$$\hat{y}_{KEKK}^h = T_h (T_h' T_h)^{-1} T_h' y \quad (3.8)$$

Eşitsizliğinde T_h yerine XR_h , y yerine $X\hat{\beta}_{KEK}$ konulduğunda;

$$\hat{y}_{KEKK}^h = XR_h (R_h' X' XR_h)^{-1} R_h' X' X \hat{\beta}_{KEK} \quad (3.9)$$

elde edilir. 3.9 'dan yola çıkarak;

$$\hat{\beta}_{KEKK}^h = R_h (R_h' X' XR_h)^{-1} R_h' X' X \hat{\beta}_{KEK} \quad (3.10)$$

elde edilir.

Çok değişkenlide $\hat{y}_{KEKK}^h$ vektörü yerine $\hat{Y}_{KEKK}^h$ matrisi kullanılır. 3.9 ve 3.10 daki eşitsizliklerinde bulunan R_h , R_h 'ın tekil olmayan bir dönüşümü sayesinde, sonucu değiştirmeden değişebilir. Mesela, W_h , R_h yerine kullanılabilir.

$\hat{\beta}_{KEKK}^h$ 'ın daha basit haline ulaşmak için şu yol izlenebilir: 3.5 teki T_h ifadeleri yerine 3.4 teki karşılığı yazıldığında;

$$P_h = X'XR_h(R_h'X'XR_h)^{-1} \quad (3.11)$$

elde edilir. Buradan 3.10 daki eşitsizlik ile;

$$\hat{\beta}_{KEKK}^h = R_h P_h' \hat{\beta}_{KEK} = W_h (P_h' W_h)^{-1} P_h' \hat{\beta}_{KEK} \quad (3.12)$$

elde edilir. Çok değişkenli durum için ise;

$$\hat{B}_{KEKK}^h = R_h P_h' \hat{B}_{KEK} = W_h (P_h' W_h)^{-1} P_h' \hat{B}_{KEK} \quad (3.13)$$

şeklindedir.

$W_h (P_h' W_h)^{-1}$ matrisi bir izdüşüm matrisidir. Fakat bu matris simetrik olmadığından, eğik projektör olarak ifade edilir. 3.12 de gösterilen h bileşeni için Kısmi En Küçük Kareler Regresyonu tahmin edicisi, P_h 'a dik olan P_h uzayı boyunca $\hat{\beta}_{KEK}$ 'nin W_h üzerine eğik izdüşümüdür. Kısmi En Küçük Kareler Regresyonunda yanın derecesi, $\hat{\beta}_{KEK}$ dik izdüşümünün alındığı uzayın boyutu ile kontrol edilir. Boyut ne kadar küçük olursa, yan da o kadar büyük olur.

3.6. Ridge Regression (RR)

RR yöntemi, 1970 yılında Hoerl ve Kennard tarafından çoklu doğrusal bağlantı sorununu ortadan kaldırmak için geliştirilmiş bir yöntemdir.

Ridge regresyon, $X'X$ korelasyon matrisinin köşegen elemanlarına pozitif ve küçük k değerleri ekleyerek koşul sayısını küçültmeyi hedeflemektedir. RR yöntemi kullanılarak elde edilen tahminler, EKK yöntemine göre daha güvenilirlerdir.

RR yöntemi; aşağıda belirtilen durumlar için önerilmektedir;

- Çoklu doğrusal regresyon modellerinde bağımsız değişkenlerin birbirleriyle ilişkili olması durumlarda EKK tahminine göre daha küçük varyanslı tahminlerin bulunmasında,
- Kuvvetli bir çoklu bağlantının olması durumunda, katsayılarda oluşan kararsızlıkların grafiksel olarak gösterilmesinde,
- Regresyondaki yanlılık karesi ile varyansı değiştirerek Hata Kareler Ortalamasını (HKO) azaltmak;

d) Bağımsız değişkenlerde oluşan çoklu bağlantıyı ortadan kaldırmaktır.

Ridge regresyonun çözüm tekniği, basit en küçük kareler yöntemine benzer bir şekildedir. RR yönteminde, regresyon katsayı tahminlerini hesaplamadan önce standart formda yer alan değişkenlerden meydana gelen $(X'X)$ matrisinin köşegen elemanlarına küçük ve pozitif bir sabitin eklenmesiyle gerçekleştirilir. Buna göre RR çözümü,

$$\hat{\beta} = (X'X + kI)^{-1} X'Y \quad (3.14)$$

şeklindeir. Burada k değeri, küçük pozitif sayısal bir değerdir. Bu değerin seçimi konusunda çok fazla çalışma yapılmaktadır. Optimum k değeri üzerinde henüz kesin bir çözüm yoktur.

3.6.1. Ridge kestiricisinin en küçük kareler kestiricisi ile ilişkisi

En küçük kareler kestiricisi,

$$\hat{\beta} = (X'X)^{-1} X'Y \quad (3.15)$$

olarak belirlenmiştir. Buradan

$$X'X \hat{\beta} = X'Y \quad (3.16)$$

olarak yazılabilir. Ridge kestiricisi,

$$\hat{\beta}^* = (X'X + kI)^{-1} X'Y \quad (3.17)$$

olarak verilmişti. $X'Y$ yerine eşiti yazıldığında buradan

$$\hat{\beta}^* = (X'X + kI)^{-1} X'X \hat{\beta} \quad (3.18)$$

elde edilir.

$X'X$ matrisinin tersinin tersi kendi olduğundan,

$$\hat{\beta}^* = (X'X + kI)^{-1} [(X'X)^{-1}]^{-1} \hat{\beta} \quad (3.19)$$

yazılabilir. Her iki matris tekil olmadıklarından,

$$\hat{\beta}^* = [(X'X)^{-1} (X'X + kI)^{-1}]^{-1} \hat{\beta} \quad (3.20)$$

yazılabilir. Buradan da,

$$\hat{\beta}^* = [(X'X)^{-1} X'X + k (X'X)^{-1}]^{-1} \hat{\beta} \quad (3.21)$$

yazılabilir. Gerekli işlemlerden sonra,

$$\hat{\beta}^* = [I + k (X'X)^{-1}]^{-1} \hat{\beta} \quad (3.22)$$

olur.

$$Z = [I + k (X'X)^{-1}]^{-1} \quad (3.23)$$

olarak tanımlarsak,

$$\hat{\beta}^* = Z \hat{\beta} \quad (3.24)$$

olarak yazılır. Bu eşitlik ridge kestiricisinin en küçük kareler kestiricisinin bir dönüşümü olduğunu gösterir.

Aynı zamanda $\hat{\beta}^*$ boyu $k \neq 0$ için $\hat{\beta}$ 'dan daha kısadır ve aşağıdaki şekilde gösterilir:

$$\hat{\beta}^* \hat{\beta}^* < \hat{\beta} \hat{\beta} \quad (3.25)$$

3.6.2. Z ve W matrislerinin özellikleri

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

$$W = (X'X + kI)^{-1} \quad (3.26)$$

şeklinde tanımlanınca,

$$\hat{\beta}^* = W X'Y \quad (3.27)$$

olur. Z matrisini,

$$Z = [I + k(X'X)^{-1}]^{-1} \quad (3.28)$$

şeklinde tanımlanmıştı. Buradan,

W matrisinin özdeğerleri,

$$\tau_i(w) = \frac{1}{\lambda_i + k} \quad (3.29)$$

dır.

Z matrisinin özdeğerleri ise şöyledir:

$$\tau_i(z) = \frac{\lambda_i}{\lambda_i + k} \quad (3.30)$$

Z matrisi aynı zamanda şu eşitliği de verir:

$$Z = I - k(X'X + kI)^{-1} = I - kW \quad (3.31)$$

3.6.3. Ridge kestiricisinin beklenen değeri

$$E(\hat{\beta}^*) = E[(Z\hat{\beta})] \quad (3.32)$$

$$= E[(I - kW)\hat{\beta}] \quad (3.33)$$

$$= E[(I - k(X'X + kI)^{-1})\hat{\beta}] \quad (3.34)$$

$$= E[\hat{\beta} - k(X'X + kI)^{-1}\hat{\beta}] \quad (3.35)$$

u_i matrisin özdeğerine karşılık gelen özvektörler olmak üzere $\hat{\beta}^*$ beklenen değeri;

$$E(\hat{\beta}^*) = \beta - k \sum_{i=1}^p (\lambda_i + k)^{-1} u_i u_i' \beta \quad (3.36)$$

olarak ifade edilir.

3.6.4. Ridge tahmin edicinin bazı özellikleri

1. $\hat{\beta}(k)$, yarıçapı $\hat{\beta}(k)$ 'in uzunluğu olan orjin merkezli küre üzerindeki rezidü kareler toplamını minimum yapar.
2. Rezidü kareler toplamı k 'ın artan bir fonksiyonudur.
3. $\hat{\beta}(k)' \hat{\beta}(k) < \hat{\beta}' \hat{\beta}$ ve $k \rightarrow \infty$ $\hat{\beta}(k)' \hat{\beta}(k) \rightarrow 0$ dır.

4. $\lambda_1 \geq \dots \geq \lambda_p$, $X'X$ özdeğerleri olmak üzere $X'X + kI$ en büyük özdeğerinin en küçük özdeğere oranı $(\lambda_1 + k)/(\lambda_n + I)$ olup k azalan bir fonksiyonudur. Bu oranın kareköküne X koşul sayısı denmektedir (Judge ve ark,1985).

5. Ridge tahmin edici $k=0$ için EKK tahmin edicisini verir. Ayrıca Ridge tahmin edici EKK tahmin edicinin bir lineer dönüşümü olarak da yazılabilir:

$$\hat{\beta}(k) = (X'X + kI)^{-1}X'y.$$

6. Hoerl ve Kennard (1970) Ridge tahmin edicinin toplam varyansının k 'in sürekli monoton azalan bir fonksiyonu ve yanlılığın karesinin k 'in sürekli monoton artan bir fonksiyonu olduğunu göstermiştir. Bu nedenle varyanstaki azalma, yanlılığın karesindeki artıştan fazla olduğu müddetçe Ridge tahmin edicinin iyi bir teknik olduğu söylenir. $S_k = (X'X + kI)$ olsun. Bu durumda:

$$E(\hat{\beta}(k)) = S_k^{-1}X'X\beta \quad (3.37)$$

$$BIAS(\hat{\beta}(k)\beta) = -kS_k^{-1}\beta \quad (3.38)$$

$$Var(\hat{\beta}(k)) = \sigma^2 S_k^{-1}X'X S_k^{-1} \quad (3.39)$$

olduğundan Ridge tahmin edicinin matris hata kareler ortalaması

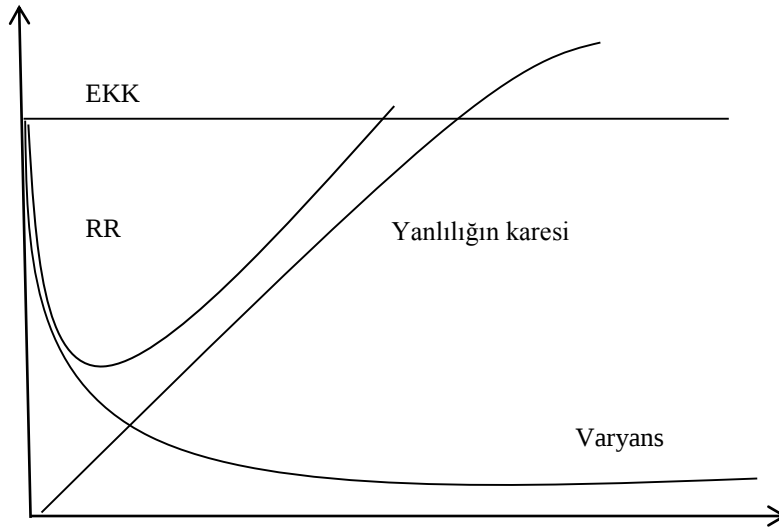
$$MMSE(\hat{\beta}(k)\beta) = S_k^{-1}(\sigma^2 X'X + k^2 \beta\beta')S_k^{-1} \quad (3.40)$$

olur. $\lambda_1 \geq \dots \geq \lambda_p$, $X'X$ 'in özdeğerleri olmak üzere skaler hata kareler ortalaması

$$SMSE(\hat{\beta}(k)\beta) = trMMSE(\hat{\beta}(k)\beta) \quad (3.41)$$

$$= \sum_{i=1}^p \frac{\sigma^2 \lambda_i}{(\lambda_i + k)^2} + k^2 \beta'(X'X + kI)^{-2}\beta \quad (3.42)$$

elde edilir.



Şekil 3. 1. $\hat{\beta}(k)$ ve $\hat{\beta}$ SMSE'in k'ya Bağlı Grafiği(Hoerl ve Kennard,1970)

7. Sabit k için $\lim_{\beta \rightarrow \infty} MMSE(\hat{\beta}(k) \beta) = \infty$ olduğundan $\hat{\beta}(k)$ sınırsız riske sahip olmamaktadır(Özkale,2007).

3.6.5. k'nın nokta tahmini yöntemi

RR modeline ait olan k parametresinin belirlenmesi özdeğerlere dayalıdır. RR işleminin hangi noktada durağanlaştığını ya da özdeğer 1'e en yakın olduğu noktayı tespit etmek için, Ridge iz grafiği dikkate alınarak ya da k parametresinin değerinin bulunması ile tespit edilir. Çoğu araştırmacı k değerinin bulunması için farklı formüller önermiştir. Bunlar içerisinde özdeğere bağlı olarak k sabitinin belirlenmesinde koşul indeksinden faydalanarak,

$$k \leq \frac{\lambda_{max} - 100\lambda_{min}}{99}, k \neq 0 \quad (3.43)$$

denklemini elde edilmiştir. Bu eşitlikten k parametresinin VBF değerini 1'e yakın yaptığı nokta belirlenebilir. Albayrak (2005)'in Anderson (1998)'dan bildirdiğine göre, optimum k değerinin elde edilmesine yönelik kullanılan diğer kriterler arasında katsayıların kuramsal beklentilere uygunluğunu, durağanlığını, kabul edilebilir büyüklüğünü, makul hata kareleri toplamını ve minimum VBF 'leri sağlayan k sabiti yaklaşımları sayılabilir.

Ridge regresyon (RR) modelinde $X'X$ matrisi korelasyon formunda olmak üzere $X'X$ matrisinin köşegen değerlerine küçük k değerlerinin ($0 \leq k \leq 1$) eklenmesiyle RR elde edilmektedir (Hoerl ve Kennard 1970).

$$\hat{\beta}(k) = (X'X + kI)^{-1}X'y, \quad k \geq 0 \quad (3.44)$$

Ridge tahmin edici, Hoerl ve Kennard (1970) tarafından aşağıda gibi elde edilmiştir. $\tilde{\beta}$ rastgele bir tahmin edici iken, $\tilde{\beta}$ 'nın kareler toplamı, $\tilde{\beta}$ 'nın optimallik özelliğine göre

$$\begin{aligned} S(\tilde{\beta}) &= (y - X\tilde{\beta})'(y - X\tilde{\beta}) \\ &= (y - X\hat{\beta})'(y - X\hat{\beta}) + (\hat{\beta} - \tilde{\beta})'X'X(\hat{\beta} - \tilde{\beta}) \\ &= S(\hat{\beta}) + \Phi(\tilde{\beta}) \end{aligned} \quad (3.45)$$

olarak yazılabilir.

Burada, $(\hat{\beta} - \tilde{\beta})'X'X(\hat{\beta} - \tilde{\beta})$, $\hat{\beta}$ yerine $\tilde{\beta}$ 'nın olmasından kaynaklanan yanlılığın karesidir. $X'X$ kötü koşullu iken, $\hat{\beta}$ ile β arasındaki mesafe artmaktadır. Bu yüzden $\tilde{\beta}'\tilde{\beta}$ tahmin edicisinin mesafesinin toplamının verilen belli bir seviyesi için uzaklığın karesini minimum yapmamız gerekir.

$\Phi > 0$ hata kareler toplamı için verilen bir sabit iken;

$S(\hat{\beta}) + \Phi(\tilde{\beta})$ koşulunu sağlayan tahmin edicilerin bir kümesi vardır $\{\tilde{\beta}\}$.

Bu kümede her zaman en küçük uzunluklu $\tilde{\beta}$ tahmini elde edilmek istenir. Diğer bir deyişle $1/k$ Lagrange çarpanı olmak üzere;

$$\underset{\tilde{\beta}}{Min} \left\{ \tilde{\beta}'\tilde{\beta} + \frac{1}{k} [(\hat{\beta} - \tilde{\beta})'X'X(\hat{\beta} - \tilde{\beta}) - \Phi_0] \right\} \quad (3.46)$$

ifadesinin $\tilde{\beta}$ ve $1/k$ ya göre türevleri

$$\tilde{\beta} + \frac{1}{k}(X'X)(\tilde{\beta} - \hat{\beta}) = 0 \quad (3.47)$$

ve

$$\Phi_0 = (\hat{\beta} - \tilde{\beta})'X'X(\hat{\beta} - \tilde{\beta}) \quad (3.48)$$

normal denklemlerini verir.

3.7. Temel Bileşenler Regresyonu (TBR)

Çoklu doğrusal bağlantıyı gidermeye yönelik olarak kullanılan yanlı tahmin edicilerden birisi de Temel Bileşenler tahmin edicisidir. Temel Bileşenler Analizi, ilk kez Hotelling (1933) tarafından kullanılmıştır.

Temel Bileşenler Regresyonunda; aralarında yüksek korelasyon bulunan açıklayıcı değişkenlere Temel Bileşenler Analizi uygulanır. Bu analiz, bir boyut indirgeme yöntemi olmakla beraber, bu analiz sonucunda orijinal değişkenler arası varyansı en yüksek derecede verebilen ve birbirinden bağımsız daha az sayıda yeni değişkenler elde edilir. Bu yeni değişkenler, bileşen olarak adlandırılır. Bu bileşenlere ait skor değerleri, sonuç değişkenini açıklamak üzere kurulan regresyon modelinde kullanılır. Boyut indirgeme işlemi için bazen Temel Bileşenler Analizi yerine Faktör Analizi de yapılmaktadır. Faktör Analizi sonucunda elde edilen faktör skorları regresyon analizinde kullanılır.

Temel Bileşenler Analizinde, n (gözlem sayısı) ve p (değişken sayısı) dan meydana gelen veri matrisi X'in p boyutlu uzaydaki durumuna göre, veri matrisi çok sayıda noktadan meydana gelen bir topluluk olarak ifade edilebilir. Bu matriste ham veri kullanılıyorsa, varyans-kovaryans matrisinden yararlanılır. Fakat standartlaştırılmış veri kullanılıyorsa korelasyon matrisinden yararlanılır. Farklı sonuçlar doğuran bu 2 yoldan hangisinin seçileceği ise verilerin ölçü birimlerine bağlıdır. Eğer değişkenlerin ölçü birimleri aynı ise varyans-kovaryans matrisi, aynı değilse korelasyon matrisi kullanılmalıdır.

Temel Bileşenler Analizi, bir değişkenler setinin varyans-kovaryans yapısını, yine bu değişkenlere ait doğrusal birleşimleri ile açıklayıp, boyut indirmesini ve yorumlanmasını sağlayan çok değişkenli bir istatistik yöntemidir.

$$X = \begin{pmatrix} X_{11} & X_{12} & X_{13} & \cdots & X_{1J} & \cdots & X_{1p} \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ X_{n1} & X_{n2} & X_{n3} & \cdots & X_{nJ} & \cdots & X_{np} \end{pmatrix}$$

p adet tesadüfi değişkeni;

$$X_1 = \begin{pmatrix} X_{11} \\ X_{21} \\ \vdots \\ X_{i1} \\ \vdots \\ X_{n1} \end{pmatrix}, X_2 = \begin{pmatrix} X_{12} \\ X_{22} \\ \vdots \\ X_{i2} \\ \vdots \\ X_{n2} \end{pmatrix}, \dots, X_j = \begin{pmatrix} X_{1j} \\ X_{2j} \\ \vdots \\ X_{ij} \\ \vdots \\ X_{nj} \end{pmatrix}, \dots, X_p = \begin{pmatrix} X_{1p} \\ X_{2p} \\ \vdots \\ X_{ip} \\ \vdots \\ X_{np} \end{pmatrix}$$

şeklinde gösterilmektedir. Matematiksel anlamda temel bileşenler $X_1, X_2, \dots, X_p$ değişkenlerinin doğrusal kombinasyonlarıdır. Geometrik olarak bu doğrusal bileşenler birbiriyle ilişkili koordinat eksenleri $X_1, X_2, \dots, X_p$ olan orijinal sistemleri döndürerek birbirinden bağımsız yeni koordinat sisteminin oluşturulmasını amaçlamaktadır. Elde edilen yeni eksenler, maksimum değişkenliği içeren yönleri göstermesiyle beraber, aynı zamanda değişim yapısının daha basit ve daha az sayıda değişken ile açıklanmasına imkân vermektedir. X gözlem matrisinin birinci temel bileşenine ait olan denklem şu şekildedir;

$$Y_1 = t_{11} X_1 + t_{21} X_2 + \dots + t_{p1} X_p \quad (3.49)$$

$$Y_1 = t'_1 \text{ olur.}$$

Buradan;

$$t'_1 = (t_{11}, t_{21}, \dots, t_{p1}) \text{ ve } X' = (X_1, X_2, \dots, X_p) \quad (3.50)$$

$$t'_1 X = (t_{11}, t_{21}, \dots, t_{p1}) \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ \vdots \\ X_p \end{pmatrix} \quad (3.51)$$

$$= t_{11} X_1 + t_{21} X_2 + \dots + t_{p1} X_p \quad (3.52)$$

şeklinde yazılabilir. X matrisinde bulunan p değişkenine ait doğrusal bileşenlerini bulmak için varyans-kovaryans matrisinin özdeğerleri ve özvektörleri kullanılır (Alpar, 1997; Ludwig ve Reynolds, 1988). Bahsedilen Varyans- kovaryans matrisini oluşturmaya yönelik olarak gerekli olan ortalamalar, varyans ve kovaryanslar aşağıda verilen denklemler yardımıyla hesaplanmaktadır (Shanmugan ve Johonson, 2007).

$$\bar{x}_j = \frac{\sum_{i=1}^n x_{ij}}{n}, \quad j = 1, 2, \dots, p \quad (3.53)$$

$$s_{jj'} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{ij'} - \bar{x}_{j'})}{n-1}, \quad j = 1, 2, \dots, p \quad i = 1, 2, \dots, p \quad (3.54)$$

Yukarıdaki şekliyle hesaplanmış değerler ile oluşturulan varyans-kovaryans matrisi $p \times p$ boyutunda olmakta ve şu şekilde verilmiştir.

$$S = \begin{pmatrix} s_{11} & s_{12} & s_{13} \dots & s_{1p} \\ s_{21} & s_{22} & s_{23} \dots & s_{2p} \\ s_{31} & s_{32} & s_{33} \dots & s_{3p} \end{pmatrix}$$

Böylelikle temel bileşenlerin doğrusal denklemlerini oluşturmak için $Y_1 = t_{11} X_1 + t_{21} X_2 + \dots + t_{p1} X_p$ eşitliğinde verilen birinci temel bileşenin varyansı verildiği gibi hesaplanmaktadır.

$$E\left\{ (t_1' X) (t_1' X) \right\} = E(t_1' X X t_1) = t_1' S t_1 \quad (3.55)$$

Yine $t_1' t_1 = 1$ kısıtlaması altında $(t_1' t_1 - 1) = 0$ olmaktadır. Lagrange fonksiyonu,

$$\phi(t_1, \lambda_1) = t_1' S t_1 - \lambda_1 (t_1' t_1 - 1), \quad (3.56)$$

bulunur.

Buradan;

$$(S - \lambda_1 I) t_1 = 0 \quad (3.57)$$

elde edilmektedir. Buradaki λ_1 değeri, S matrisinin birinci bileşenine ait özdeğer; t_1 ise S matrisinin birinci özdeğerinden elde edilmiş özvektördür. Söz konusu eşitliğin

solundaki ikinci terim için $t_1' t_1 = 1$ kısıtlama olması nedeniyle $t = 0$ başka çözüm olması, sadece birinci terimin determinantının sıfır olması ile mümkün olabilmektedir.

$$(S - \lambda_1 I) = 0 \quad (3.58)$$

eşitliğini p özdeğeri için genelleştirir ve λ_1 yerine λ koyulduğunda,

$$(S - \lambda I) = 0 \quad (3.59)$$

elde edilir. Bu eşitlik çözümlendiğinde λ 'ın p dereceden bir polinomu bulunur. Yine bu polinomun $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$ olacak şekilde toplam p adet kökü, diğer bir ifadeyle özdeğeri vardır. Bileşenlerin varyans değeri özdeğerlere eşit olduğundan ve özdeğerleri de büyükten küçüğe doğru sıralanması durumunda, birinci bileşenin varyansı en yüksek olmaktadır. Bu yüzden birinci bileşenin varyans değeri, toplam varyasyondaki en yüksek paya sahiptir. Hesaplanan p adet özdeğer için, yine p adet özvektör hesaplanmaktadır. Özvektörleri hesaplamadan önce, $\Phi_0 = (\hat{\beta} - \tilde{\beta})' X' X (\hat{\beta} - \tilde{\beta})$ eşitliğinden hesaplanan her özdeğeri tek tek aynı eşitliğe koyarak matrisin eşleniği bulunursa;

$$Eş(S - \lambda I)$$

Eşlenik matris kofaktör matrisine eşit olmaktadır.

$$K = Eş(S - \lambda I) \quad (3.60)$$

K , kofaktör matrisin boyutu $p \times p$ boyutta olup, sütunların birbirinin aynısı veya farklı olmasının yanında en önemli özelliği birbirine orantılı olmasıdır. Bu özellikten dolayı, her sütunun standartlaştırılmış özvektörlerin değerleri birbirine eşit olmaktadır. Bunun sonucu olarak, özvektörlerin standartlaştırılması işlemi yapılır. Birinci özvektörün standartlaştırılması ile ilgili denklem aşağıda şekildedir.

$$t_{i1} = \frac{k_{i1}}{\sqrt{\sum_{i=0}^p (k_{i1})^2}}, \quad i = 1, 2, 3, \dots, p \quad (3.61)$$

Eşitlikteki k_{i1} ifadesi K matrisinin i . satır elemanıdır. Elde edilen p adet standartlaştırılmış özvektör eşitlikte koyulursa;

$$Y_1 = t'_1 X = t_{11} X_1 + t_{21} X_2 + \dots + t_{p1} X_p$$

$$Y_2 = t'_2 X = t_{12} X_1 + t_{22} X_2 + \dots + t_{p2} X_p$$

.....

$$Y_p = t'_p X = t_{1p} X_1 + t_{2p} X_2 + \dots + t_{pp} X_p$$

bileşenleri elde edilir. Standardize hale getirilmiş özvektörlerin kareleri toplamı 1'dir.

$$\sum_{i=1}^p (k_{ij})^2 = 1, \quad i=1,2,\dots,p \quad (3.62)$$

Bu şekilde, toplam varyasyondaki en büyük katkırı sağlayan bileşen için bulunan standardize hale getirilmiş özvektör değerinin,

$$t_{ij}^2, \quad i=1,2,\dots,p$$

şeklinde karesini alarak, birinci bileşende yeralan her değişkenin toplam varyasyona olan katkısı bulunur ve değişken sayısı,

$$\sum_{i=1}^p t_{i1}^2 < 0.95, \quad i=1,2,\dots,p$$

eşitliği ile bulunur. Aynı işlem diğer bileşenler yani özdeğerler için de yapılır. Y_1 sonra bağımsız doğrusal bileşenler arasında en büyük varyansa ikinci bileşen sahiptir.

Ayrıca, Y_1 ile Y_2 'in bağımsızlığı $(t'_1 t_2) = 0$,

$$E(Y_1 Y_2) = E [(t'_1 X) (t'_2 X)']$$

$$= E (t'_1 X X' t_2)$$

$$= t'_1 S t_2$$

$$= \lambda_1 t'_1 t_2$$

$$= 0$$

şeklinde ispat edilir. Bu Y_1 ve Y_2 istatistiksel anlamda hem bağımsız hem de dik olarak görünür.

Temel bileşenler regresyonunda kullanılacak olan bileşenler, standardize hale

getirilmiş değişkenlerden (Z_{ij}) ve temel bileşenlerde değişkenlerin ağırlıklarından (t_{ij}) yararlanılarak bulunur.

$$S_{kj} = t_{1k}Z_{1j} + t_{2k}Z_{2j} + \dots + t_{pk}Z_{pj}$$

Eşitlikte yer alan S_{kj} ifadesi, k-ıncı temel bileşendeki j-inci gözlem değeri için skor değerlerini ifade etmektedir. Bu skorların kullanılması ile temel bileşenler regresyon yöntemine dayalı model tahmini şu şekildedir:

$$\hat{Y} = a + b_1s_1 + b_2s_2$$

Buradaki a ifadesi, regresyon sabitini, b_i ($i=1,2$) ifadesi ise i-inci temel bileşen skorlarının regresyon katsayısını ifade etmektedir.

3.7.1. Temel bileşen sayısının belirlenmesi

Temel bileşenler analizinde bileşenler elde edildikten sonra, bileşen sayısının kaç olacağına karar vermek önemlidir. Buna yönelik olarak çok fazla yöntem geliştirilmiştir ve fakat en basit ve en çok kullanılan yöntem, toplam varyasyonun 2/3'ünü geçinceye kadar λ değerleri toplanarak bileşen sayısına karar verme yöntemidir. Rencher, (2002) ise açıklanan toplam varyasyon oranını % 80 olarak ifade etmiştir. Eğer bulunacak olan yeni bileşen sayısı k olarak ifade edilirse, bunun sayısı ve toplam varyasyondaki payı aşağıdaki denklem yardımıyla bulunur.

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

Başka bir yöntem ise, çoklu lineer regresyon analizi için temel bileşenler analizi sonucunda seçilecek olan temel bileşen sayısı 1'den büyük özdeğerlerin sayısıdır. Bir diğer yöntem; temel bileşenlere ait çizilen grafiğe bakarak, grafiğin monotonlaşmaya başladığı noktayı temel bileşenlerin sayısı olarak belirlemektir.

4. UYGULAMA

4.1. Örnek Uygulama

Örneklem büyüklüğü $n=30$, değişken sayısı $x=7$ ve ilişki düzeyi $\rho = 0.9$ olarak belirlenmiş şekilde türetilmiş deneme sonucunda elde edilen veriler aşağıda verilmiştir.

Çizelge 4. 1. $n=30$, $x=7$ ve $\rho=0.9$ iken elde edilen simülasyon sonuçları

	Y	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
1	-10.6599	-1.7004	-1.36334	-1.25266	-1.55361	-0.95063	-1.98978	-1.84709
2	6.8852	0.53564	1.26376	1.00403	0.57264	0.45586	0.89183	1.10429
3	1.3776	0.31484	-0.43723	-0.71063	0.08681	-0.27339	-0.02124	-0.11721
4	-9.9437	-1.30236	-1.02924	-1.59775	-0.93113	-1.1617	-0.54371	-2.30794
5	-0.3138	-0.74986	-0.92202	-0.2169	-0.39981	0.29124	0.18182	-0.1533
6	-5.0700	-0.65366	-0.28755	-0.79842	-0.98458	-0.42364	-0.3635	-0.75852
7	9.2272	1.05997	1.44330	1.20458	0.88500	0.89783	0.75137	1.32055
8	3.9213	0.11613	-0.5156	-0.39962	0.51550	-0.28823	0.95717	-0.00126
9	0.5928	0.06176	0.19200	-0.28336	-0.74934	0.81040	0.70119	0.00096
10	4.3151	0.12243	0.41320	0.55354	0.38153	0.63010	0.08886	0.81078
11	3.2708	0.29706	0.83681	0.11983	-0.06607	0.13988	0.78917	0.16158
12	0.9139	0.39620	-0.18565	0.08462	-0.38977	-0.1779	0.28823	-0.02114
13	12.2480	1.17196	1.50687	1.55514	1.30308	0.73669	1.88902	2.56991
14	9.9170	1.02835	1.53527	1.09699	1.42754	0.25912	1.22168	1.34462
15	-2.5537	-0.24405	-1.22008	-0.3577	-1.60832	0.72385	-0.5496	-0.87961
16	-5.6884	-0.65175	-1.33859	-0.73308	-0.31816	-0.84728	-0.5682	-1.05105
17	3.0740	-0.00466	0.22568	0.55109	0.14358	-0.03187	0.41968	0.54932
18	8.0769	0.91446	1.7282	1.34058	1.42237	0.80692	1.28485	1.49444
19	-6.1496	-0.75357	-1.16452	-1.04021	-0.90178	-0.77447	-0.53934	-0.70147
20	22.7757	2.69416	2.60693	3.58822	3.25159	2.76456	2.81879	4.48173
21	4.2781	0.51768	0.31862	0.64645	0.12951	-0.78084	1.24012	0.88277
22	-1.2111	-0.3663	-0.68935	-0.32175	-0.67339	-0.37253	0.03112	-0.77221
23	4.5969	0.86943	-0.13051	0.04046	1.02638	0.40562	0.26393	0.56452
24	2.5851	0.14742	-0.67631	0.32616	0.67150	0.16087	0.28521	0.44968
25	-7.0684	-1.01972	-1.31506	-0.68085	-0.68039	-1.85439	-0.31205	-2.09113
26	7.8680	0.78790	1.23622	1.37808	0.7683	0.63056	1.66394	1.67637
27	-1.2904	-0.42021	-0.64057	-0.41749	-0.71642	-0.03515	0.29127	-0.22439
28	-2.6621	-1.24602	-0.55181	-0.16821	0.30807	-1.60857	-1.45214	-1.28372
29	16.7458	1.37414	2.33942	1.15150	2.18070	2.5958	2.15190	2.72120
30	10.8699	1.02324	1.52117	1.12785	1.08413	1.47739	0.85442	1.57515

Tahmin edilen regresyon denklemi

$$\hat{Y}_{EKK} = 1.119 + 1.533 X_1 + 0.567 X_2 + 0.736 X_3 + 1.738 X_4 + 0.924 X_5 + 0.643 X_6 + 1.080 X_7$$

$$HKO_{EKK} = \frac{(1-1.533)^2 + (1-0.567)^2 + \dots + (1-1.080)^2}{7} = 0.175078$$

Çizelge 4. 2. En Küçük Kareler Yöntemine Göre Regresyon Analiz Sonuçları

Değişkenler	Katsayılar	Standart Hata	T-Değeri	P-Değeri	VIF
Sabit	1.119	0.276	4.06	0.001	
X1	1.533	0.832	1.84	0.079	12.13
X2	0.567	0.503	1.13	0.272	6.89
X3	0.736	0.708	1.04	0.31	11.2
X4	1.738	0.504	3.45	0.002	5.99
X5	0.924	0.479	1.93	0.067	4.85
X6	0.643	0.574	1.12	0.275	6.78
X7	1.08	0.807	1.34	0.194	28.3

Çizelge 4. 3. EKK Sonucu Elde Edilen Çoklu Regresyon Varyans Analiz Sonuçları

Kaynak	Serbestlik Derecesi	Kareler Toplamı	Kareler Ortalaması	F-Değeri	P-Değeri
Regresyon	7	1653.13	236.162	160.10	0.000
X1	1	5.00	5.004	3.39	0.079
X2	1	1.87	1.869	1.27	0.272
X3	1	1.59	1.593	1.08	0.310
X4	1	17.55	17.552	11.90	0.002
X5	1	5.49	5.494	3.72	0.067
X6	1	1.85	1.849	1.25	0.275
X7	1	2.64	2.642	1.79	0.194
Hata	22	32.45	1.475		
Toplam	29	1685.58		Düzeltilmiş R Kare = 0.9746	

Çizelge 4.1. ‘deki veriler kullanıldığında KEKK yöntemi ile elde edilen model ve HKO ‘sı aşağıdaki gibi elde edilmiştir.

$$\hat{Y}_{KEKK} = 1.08972 + 1.37404X_1 + 0.709557X_2 + 1.05466X_3 + 1.71159X_4 + 0.826148X_5 + 0.755764X_6 + 0.855327X_7 + e$$

$$HKO_{KEKK} = \frac{(1-1.37404)^2 + (1-0.709557)^2 + \dots + (1-0.855327)^2}{7} = 0.120631$$

Çizelge 4.1. ‘deki veriler kullanıldığında TBR yöntemi ile elde edilen model ve HKO ‘sı aşağıdaki gibi elde edilmiştir.

$$\hat{Y}_{TBR} = 1.03187 + 0.839032X_1 + 1.05727X_2 + 1.01401X_3 + 1.07609X_4 + 0.639942X_5 + 0.950688X_6 + 1.37355X_7 + e$$

$$HKO_{TBR} = \frac{(1-0.839032)^2 + (1-1.05727)^2 + \dots + (1-1.37355)^2}{7} = 0.043827$$

Çizelge 4.1. ‘deki veriler kullanıldığında RR yöntemi ile elde edilen model ve HKO ‘sı aşağıdaki gibi elde edilmiştir.

$$\hat{Y}_{RR} = 1.06385 + 1.30472X_1 + 0.612636X_2 + 0.770036X_3 + 1.63457X_4 + 0.892892X_5 + 0.727120X_6 + 1.18392X_7 + e$$

$$HKO_{RR} = \frac{(1-1.30472)^2 + (1-0.612636)^2 + \dots + (1-1.18392)^2}{7} = 0.11689$$

4.2. Simülasyon Sonuçları

Bu çalışmada Minitab 16.0 programı ile ayrı ayrı 4, 7 ve 10 değişkenli, normal dağılıma uygun, farklı ilişki düzeylerinde ve örneklem büyüklüklerinde doğrusal regresyon modelleri oluşturulmuştur. Oluşturulan bu modellerin her birinin parametreye ilişkin HKO (Hata Kareler Ortalamaları) hesaplanmış ve ortalamaları alınmıştır. Elde edilen bu hata kareler ortalama değerleri dikkate alınarak Kısmi En küçük Kareler Yöntemi, Ridge Regresyon ve Temel Bileşenler Regresyonu yöntemlerinden hangisinin daha tercih edilebilir olduğu saptanmaya çalışılmıştır. Türetilen verilerle elden edilen HKO değerleri Çizelge 4.2’de sunulmuştur.

Açıklayıcı değişkenler (x) ;

$$x_{ij}=(1-\rho^2)^{1/2}u_{ij}+\rho u_{ip}$$

ile üretilmiştir.



Çizelge 4. 4. Değişken sayısı: 4-7-10, Örneklem Büyüklüğü:30-40-100-200-500, İlişki Düzeyi:0, 0.3, 0.5, 0.7, 0.9 iken 5000 denemede elde edilen ortalama HKO değerleri

	n=30				n=50				n=100				n=200				n=500			
	ρ	KEKK	TBR	RR	ρ	KEKK	TBR	RR	ρ	KEKK	TBR	RR	ρ	KEKK	TBR	RR	ρ	KEKK	TBR	RR
x=4	0	0.0423	0.4303	0.0349	0	0.0222	0.4283	0.0227	0	0.0121	0.3952	0.0109	0	0.00516	0.39876	0.00508	0	0.00178	0.39566	0.00184
	0.3	0.0380	0.1824	0.0378	0.3	0.0211	0.1386	0.0216	0.3	0.0108	0.1328	0.0104	0.3	0.00464	0.12587	0.00483	0.3	0.00232	0.1278	0.00202
	0.5	0.0440	0.127	0.047	0.5	0.0238	0.1022	0.0264	0.5	0.0116	0.0967	0.0129	0.5	0.00492	0.08982	0.0054	0.5	0.00219	0.08896	0.00238
	0.7	0.0538	0.0748	0.0591	0.7	0.0365	0.0746	0.0385	0.7	0.0206	0.0622	0.021	0.7	0.01167	0.05867	0.00942	0.7	0.00598	0.05653	0.00359
	0.9	0.1745	0.0662	0.1695	0.9	0.0786	0.0432	0.0843	0.9	0.0500	0.034	0.0491	0.9	0.02614	0.02884	0.0237	0.9	0.0136	0.02409	0.01109
x=7	0	0.0673	0.6716	0.0477	0	0.0326	0.5923	0.0256	0	0.0125	0.5995	0.0113	0	0.00564	0.61701	0.00503	0	0.00209	0.58347	0.00199
	0.3	0.0431	0.2655	0.0421	0.3	0.0276	0.2055	0.0261	0.3	0.0137	0.1840	0.0107	0.3	0.00928	0.17067	0.00541	0.3	0.00415	0.15777	0.00207
	0.5	0.0407	0.1454	0.0556	0.5	0.0224	0.1225	0.0301	0.5	0.0127	0.1128	0.0157	0.5	0.0059	0.10184	0.00619	0.5	0.00284	0.10083	0.00285
	0.7	0.0534	0.0806	0.0835	0.7	0.0297	0.0693	0.0460	0.7	0.0187	0.0610	0.0233	0.7	0.00907	0.05818	0.01012	0.7	0.00417	0.0558	0.00406
	0.9	0.1237	0.0404	0.1767	0.9	0.0791	0.0339	0.1166	0.9	0.0426	0.0263	0.0543	0.9	0.02343	0.02195	0.0288	0.9	0.01193	0.01997	0.01156
x=10	0	0.0957	0.7329	0.0543	0	0.0432	0.7429	0.0247	0	0.0169	0.7265	0.012	0	0.00627	0.71927	0.00527	0	0.00214	0.72321	0.00196
	0.3	0.0567	0.2868	0.0538	0.3	0.0342	0.2396	0.0250	0.3	0.0195	0.1974	0.0123	0.3	0.01203	0.17592	0.00601	0.3	0.00689	0.1668	0.00222
	0.5	0.0394	0.1451	0.0659	0.5	0.0254	0.1252	0.0347	0.5	0.0129	0.1093	0.0148	0.5	0.0074	0.10191	0.00727	0.5	0.00371	0.09617	0.00267
	0.7	0.043	0.076	0.0991	0.7	0.03	0.0651	0.0504	0.7	0.0153	0.0561	0.0223	0.7	0.00894	0.05304	0.01105	0.7	0.00379	0.05143	0.0044
	0.9	0.1089	0.0304	0.2185	0.9	0.0783	0.0267	0.1132	0.9	0.0409	0.0204	0.0602	0.9	0.02385	0.01825	0.03536	0.9	0.01091	0.01656	0.01306

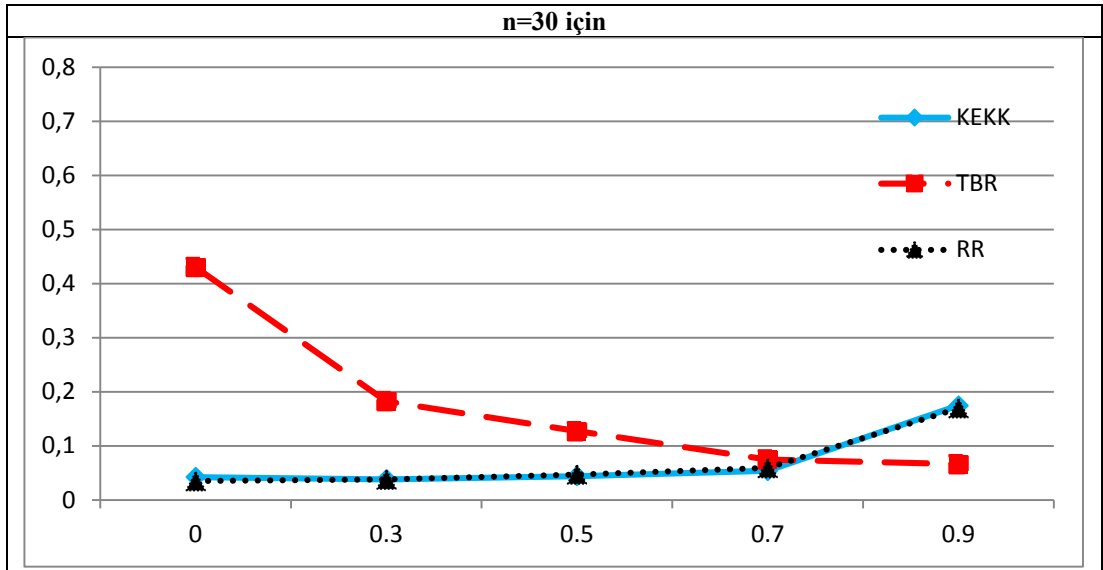
Çizelgedeki sonuçlar incelendiğinde örneklem büyüklüğü ne olursa olsun ilişki düzeyi artıkça TBR ile elde edilen HKO değeri hep düşmektedir.

n değişken sayısı artıkça kullanılan yöntem ne olursa olsun, elde edilen modellerin HKO değerleri düşmektedir.

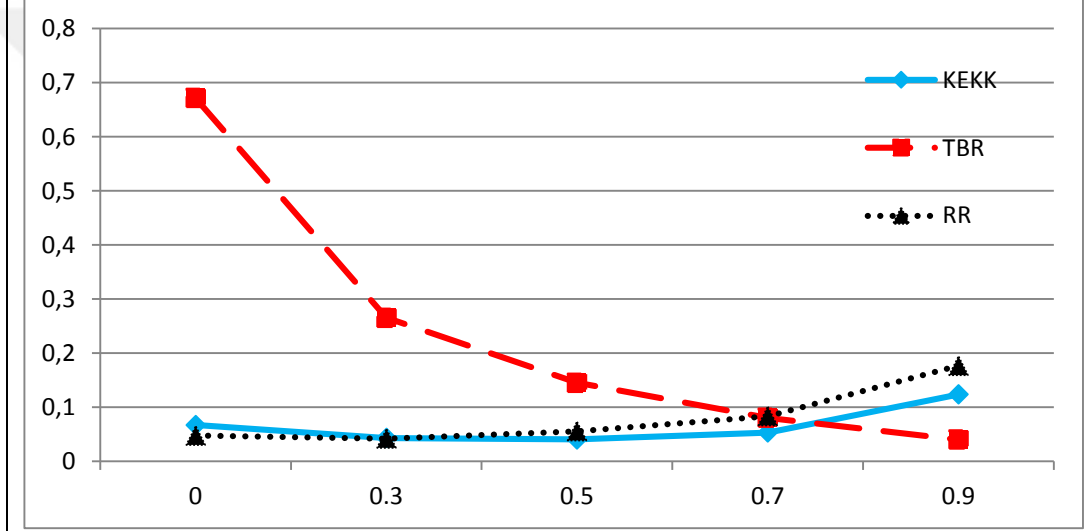
Değişken sayısı artıkça örneklem büyüklüğü ne olursa olsun yüksek ilişki düzeyi hariç tüm HKO değerlerinde artış gözlemlenmiştir.

Literatürde KEKK yönteminin TBR yöntemine göre daha iyi olduğu ifade edilmektedir. Ancak yapılan çalışmada ve elde ettiğimiz bulgulara göre, her örneklem ve her ilişki düzeyi için bunun doğru olmadığı belirlenmiştir.

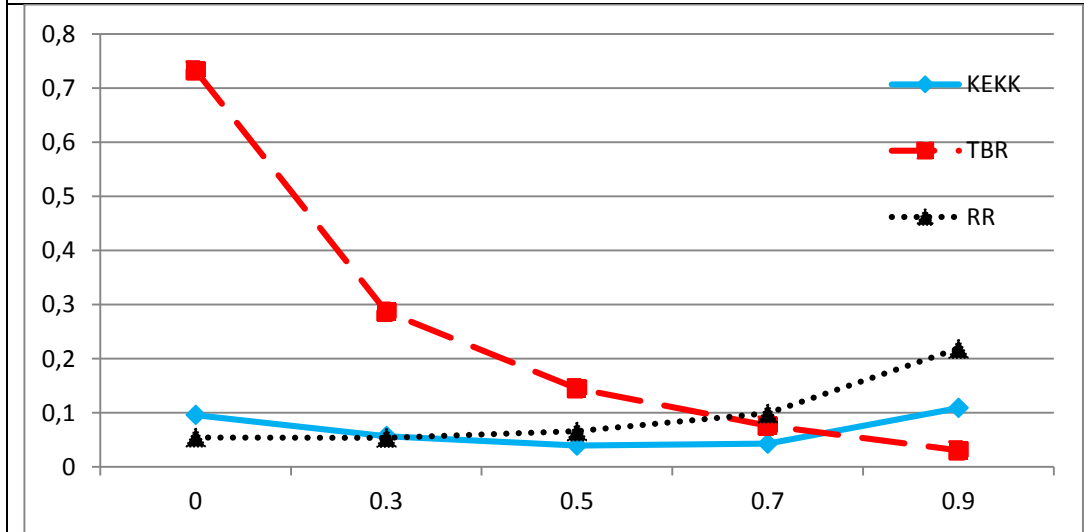




Şekil 4.1.A – Değişken Sayısı=4



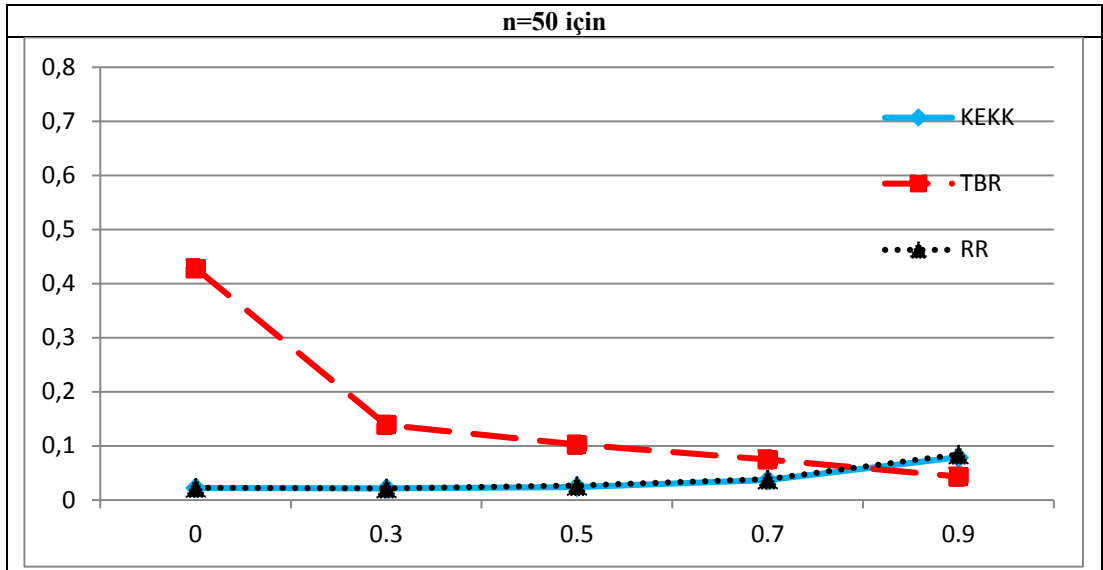
Şekil 4.1.B – Değişken Sayısı=7



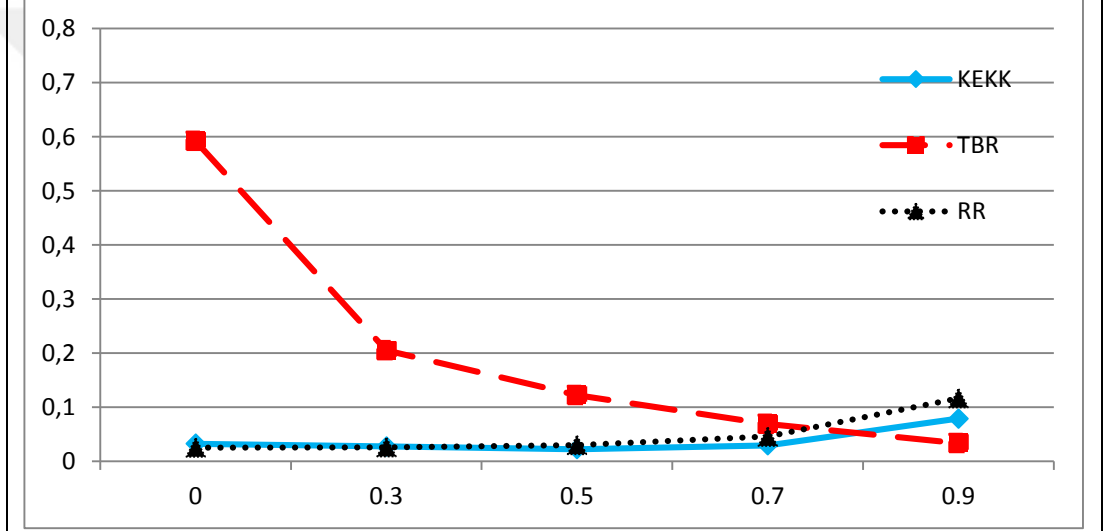
Şekil 4.1.C – Değişken Sayısı=10

Şekil 4. 1. Örneklem Büyüklüğü:30 iken elde edilen HKO değerleri

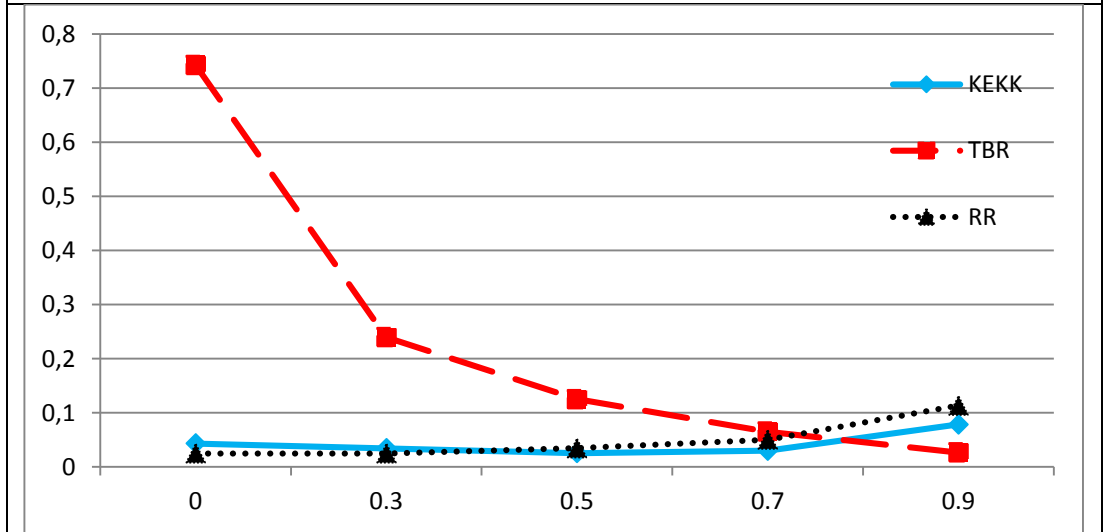
Örneklem büyüklüğü $n=30$ ve 4 değişkenli regresyon modelinde (Şekil 4.1.A) bağımsız değişkenler arasında ilişki yokken, RR diğerlerine göre daha iyi sonuçlar vermiştir. Ancak KEKK yöntemi de RR modelinden çokta kötü sonuçlar vermemiştir. İlişki düzeyi 0.3 olduğunda, TBR ‘nin HKO değeri düşmeye başlamıştır. Fakat yine de TBR yöntemi diğerlerine göre daha kötü sonuç vermiştir. İlişki düzeyi biraz daha artırıldığında (0.5), TBR nin HKO değeri anlamlı bir şekilde düşmesine rağmen, yine de diğerlerine göre daha kötü sonuçlar vermektedir. İlişki düzeyinin artmasıyla KEKK ile RR arasında bir denge olmaktadır. İlişki değeri 0.7 gibi bir ilişkide TBR diğerlerine yakın sonuçlar vermektedir. Fakat bağımsız değişkenler arasında çok yüksek bir ilişkinin olması durumunda TBR diğerlerine göre HKO bakımından daha anlamlı modeller tahmin etmektedir. Yine benzer şekilde örneklem büyüklüğü $n=30$ için hem 7 değişkenli regresyon modelinde (Şekil 4.1.B), hem de 10 değişkenli regresyon modelinde (Şekil 4.1.C) aynı durumlar söz konusudur.



Şekil 4.2.A – Değişken Sayısı=4



Şekil 4.2..B – Değişken Sayısı=7



Şekil 4.2.C – Değişken Sayısı=10

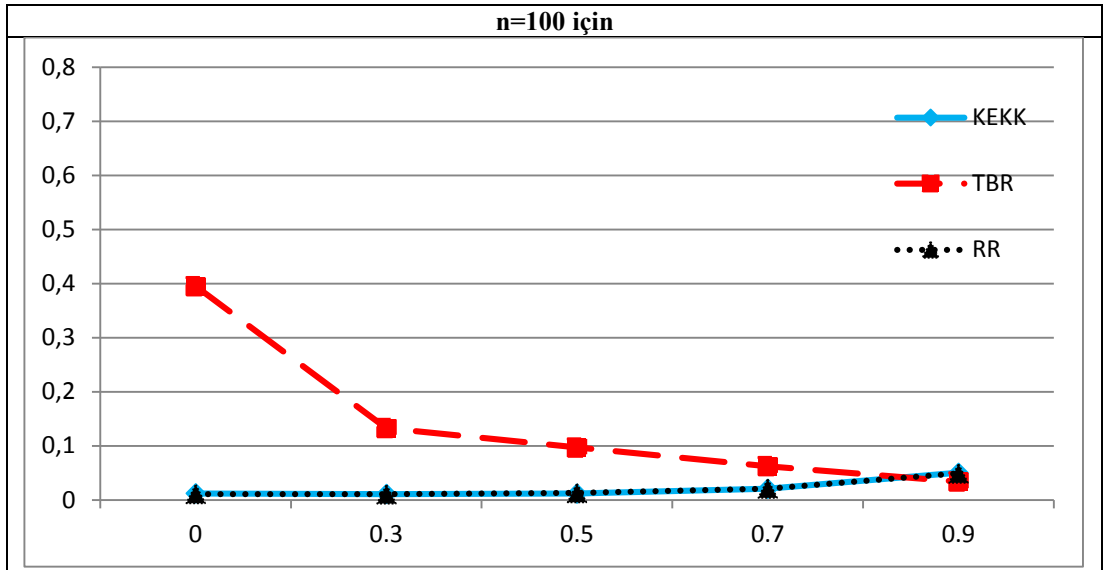
Şekil 4. 2. Örneklem Büyüklüğü:50 iken elde edilen HKO değerleri

Örneklem büyüklüğü $n=50$ iken farklı değişken sayılarına ait regresyon modellerine ait elde edilen HKO değerleri Şekil 4.2. 'de verilmiştir.

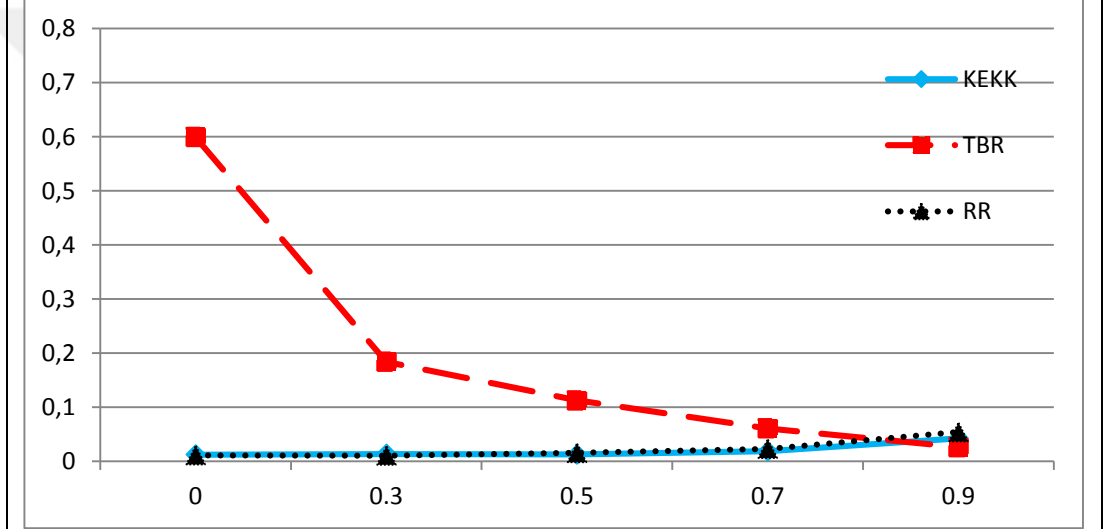
Örneklem büyüklüğü $n=50$ ve 4 değişkenli regresyon modelinde (Şekil 4.2.A) bağımsız değişkenler arasında ilişki yokken, KEKK diğerlerine göre daha iyi sonuçlar vermiştir. Ancak RR yöntemi de KEKK modelinden çokta kötü sonuçlar vermemiştir. İlişki düzeyi 0.3 olduğunda, TBR 'nin HKO değeri düşmeye başlamıştır. Fakat yine de TBR yöntemi diğerlerine göre daha kötü sonuç vermiştir. İlişki düzeyi biraz daha artırıldığında (0.5), TBR 'nin HKO değeri anlamlı bir şekilde düşmesine rağmen, yine de diğerlerine göre daha kötü sonuçlar vermektedir. İlişki değeri 0.7 gibi bir ilişkide TBR diğer yöntemlere yakın sonuçlar vermektedir. Fakat bağımsız değişkenler arasında çok yüksek bir ilişkinin olması durumunda TBR diğerlerine göre HKO bakımından daha anlamlı modeller tahmin etmektedir.

Örneklem büyüklüğü $n=50$ ve 7 değişkenli regresyon modelinde (Şekil 4.2.B) bağımsız değişkenler arasında ilişki yokken, RR diğerlerine göre daha iyi sonuçlar vermiştir. İlişki düzeyi 0.3 iken RR yine daha sonuçlar verirken, KEKK yöntemi de yakın sonuçlar vermiştir. İlişki düzeyi 0.5 ve 0.7 olduğunda ise KEKK yöntemi daha iyi sonuçlar vermektedir. İlişki düzeyi arttıkça TBR ile elde edilen HKO değerleri anlamlı bir şekilde düşmektedir. Fakat ilişki düzeyi 0.9 olduğunda TBR yöntemi diğer 2 yöneme göre daha iyi sonuçlar vermektedir.

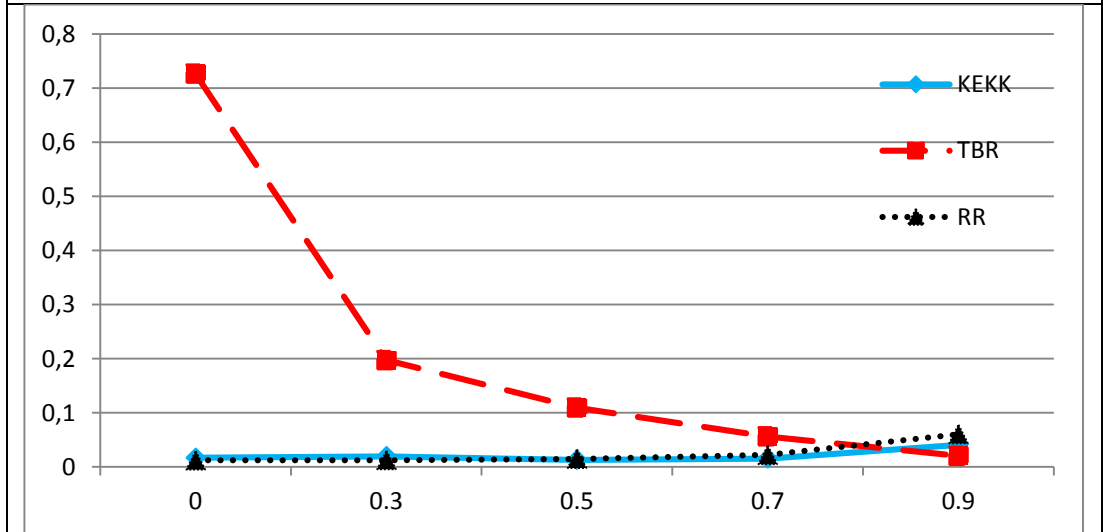
Örneklem büyüklüğü $n=50$ ve 10 değişkenli regresyon modelinde (Şekil 4.2.C) bağımsız değişkenler arasında ilişki yokken, RR diğerlerine göre daha iyi sonuçlar vermiştir. İlişki düzeyi 0.3 iken RR yine daha sonuçlar verirken, KEKK yöntemi de yakın sonuçlar vermiştir. İlişki düzeyi 0.5 ve 0.7 olduğunda ise KEKK yöntemi daha iyi sonuçlar vermektedir. İlişki düzeyi arttıkça TBR ile elde edilen HKO değerleri anlamlı bir şekilde düşmektedir. Fakat ilişki düzeyi 0.9 olduğunda TBR yöntemi diğer 2 yöneme göre daha iyi sonuçlar vermiştir.



Şekil 4.3.A – Değişken Sayısı=4



Şekil 4.3.B – Değişken Sayısı=7



Şekil 4.3..C – Değişken Sayısı=10

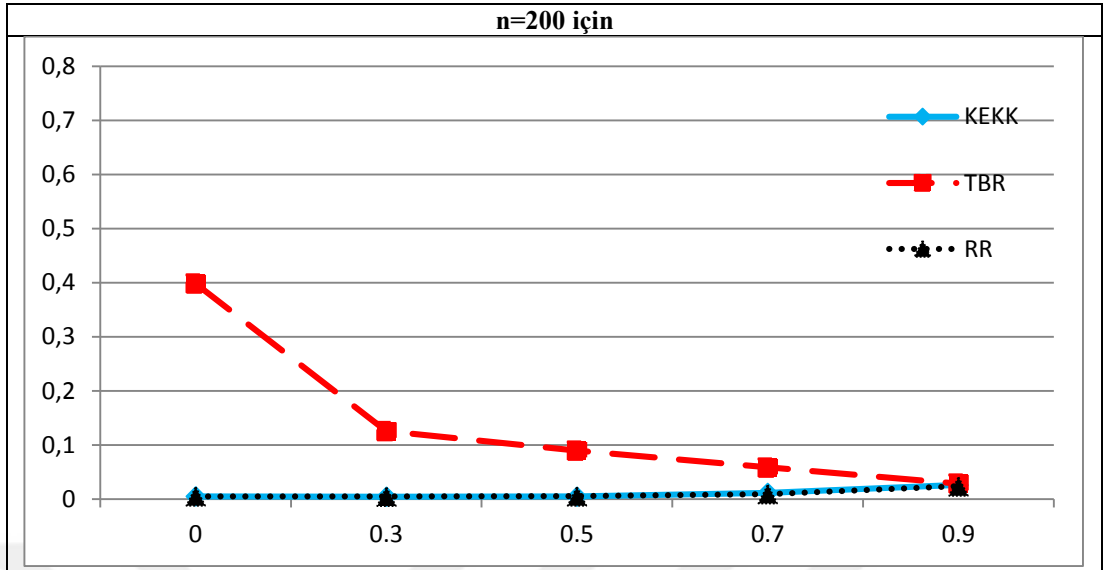
Şekil 4. 3. Örneklem Büyüklüğü:100 iken elde edilen HKO değerleri

Örneklem büyüklüğü $n=100$ iken farklı değişken sayılarına ait regresyon modellerine ait elde edilen HKO değerleri Şekil 4.3. 'te verilmiştir.

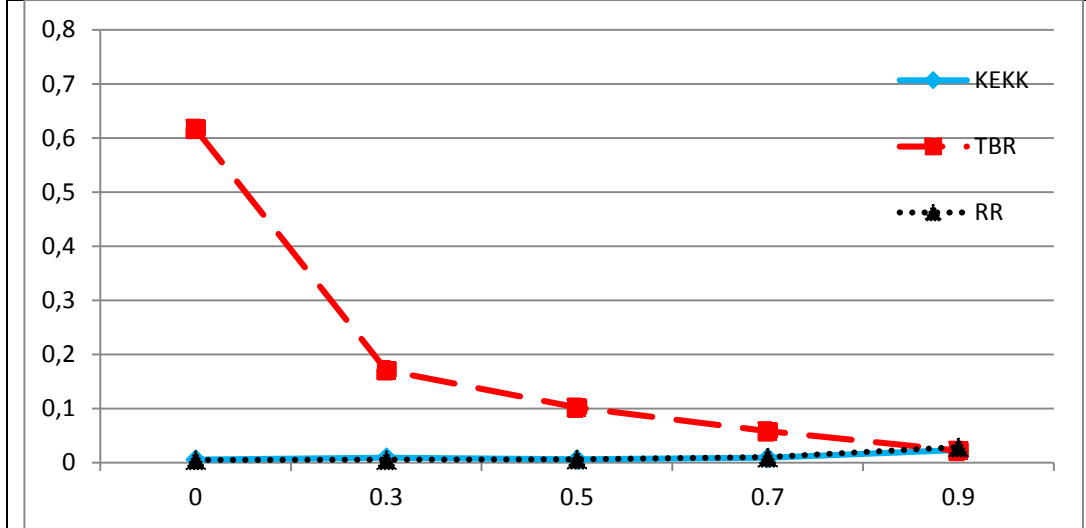
Örneklem büyüklüğü $n=100$ ve 4 değişkenli regresyon modelinde (Şekil 4.3.A) bağımsız değişkenler arasında ilişki yokken, RR diğerlerine göre daha iyi sonuçlar vermiştir. Ancak KEKK yöntemi de iyi sonuçlar vermiştir. İlişki düzeyi 0,3 olduğunda, TBR 'nin HKO değeri düşmeye başlamıştır. Fakat yine de TBR yöntemi diğerlerine göre daha kötü sonuç vermiştir. İlişki düzeyi biraz daha artırıldığında (0.5), TBR nin HKO değeri anlamlı bir şekilde düşmesine rağmen, yine de diğerlerine göre daha kötü sonuçlar vermektedir. Bu ilişki düzeyinde KEKK yöntemi daha iyi sonuçlar vermiştir. İlişki değeri 0.7 olduğunda KEKK ve RR yöntemleri daha iyi sonuçlar vermiştir. Fakat bağımsız değişkenler arasında çok yüksek bir ilişkinin olması durumunda TBR diğerlerine göre HKO bakımından daha anlamlı modeller tahmin etmektedir.

Örneklem büyüklüğü $n=100$ ve 7 değişkenli regresyon modeline (Şekil 4.3.B) ait bağımsız değişkenlere ait ilişki düzeylerine ait verilere bakıldığında, elde edilen sonuçlar 4 değişkenliye ait sonuçlarla benzerlik göstermektedir. İlişki düzeyi 0 ve 0.3 düzeyinde iken RR yöntemi, 0.5 ve 0.7 düzeyinde iken KEKK, 0.9 düzeyinde ise TBR yönteminin daha iyi olduğu görülmektedir.

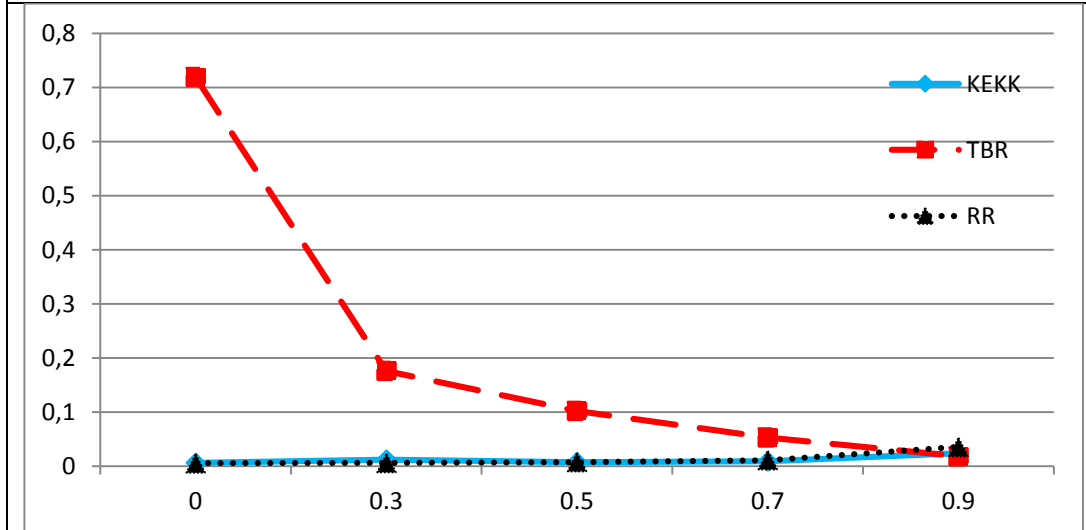
Örneklem büyüklüğü $n=100$ ve 10 değişkenli regresyon modeline (Şekil 4.3.C) ait bağımsız değişkenlere ait ilişki düzeylerine ait verilere bakıldığında, elde edilen sonuçlar 4 değişkenliye ait sonuçlarla benzerlik göstermektedir. İlişki düzeyi 0 ve 0.3 düzeyinde iken RR yöntemi, 0.5 ve 0.7 düzeyinde iken KEKK, 0.9 düzeyinde ise TBR yönteminin daha iyi olduğu görülmektedir.



Şekil 4.4.A – Değişken Sayısı=4



Şekil 4.4.B – Değişken Sayısı=7



Şekil 4.4.C – Değişken Sayısı=10

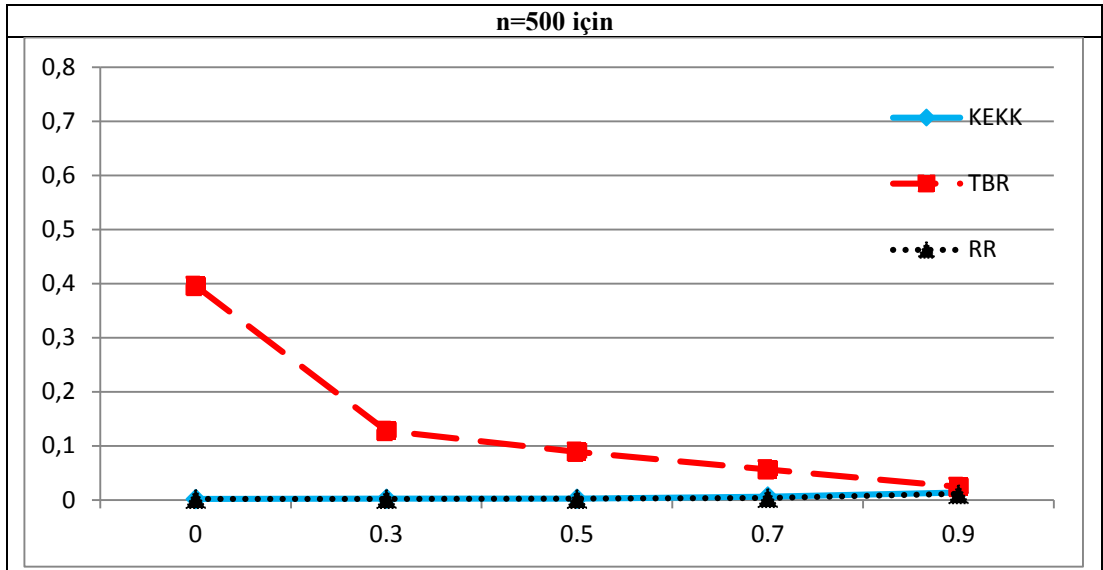
Şekil 4. 4. Örneklem Büyüklüğü:200 iken elde edilen HKO değerleri

Örneklem büyüklüğü $n=200$ iken farklı değişken sayılarına ait regresyon modellerine ait elde edilen HKO değerleri Şekil 4.4. 'te verilmiştir.

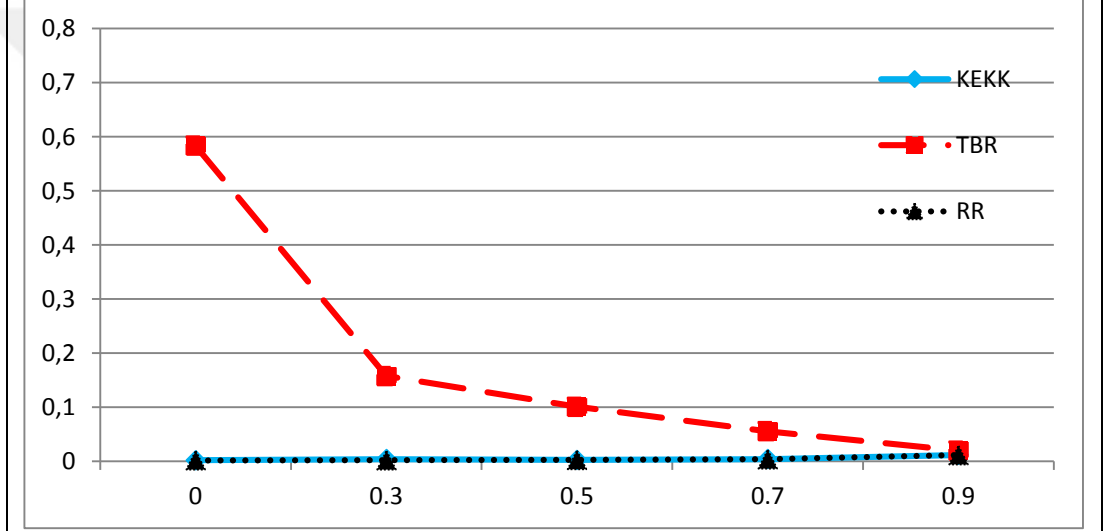
Örneklem büyüklüğü $n=200$ ve 4 değişkenli regresyon modelinde (Şekil 4.4.A) bağımsız değişkenler arasında ilişki yokken, RR diğerlerine göre daha iyi sonuçlar vermiştir. Ancak KEKK yöntemi de iyi sonuçlar vermiştir. İlişki düzeyi 0.3 olduğunda, KEKK yöntemi daha tercih edilebilir olmuş, TBR yönteminin de HKO değerinin düşmeye başladığı gözlenmiştir. Fakat yine de TBR yöntemi diğerlerine göre daha kötü sonuç vermiştir. İlişki düzeyi biraz daha artırıldığında (0.5), KEKK yöntemi diğerlerine göre daha iyi sonuçlar vermiştir. İlişki değeri hem 0.7 iken, hem de 0.9 iken RR yöntemi diğerlerine göre HKO bakımından daha anlamlı modeller tahmin etmektedir.

Örneklem büyüklüğü $n=200$ ve 7 değişkenli regresyon modelinde (Şekil 4.4.B) bağımsız değişkenler arasında ilişki düzeyi 0.0 ve 0.3 iken RR yöntemi daha iyi sonuçlar vermiştir. İlişki düzeyi 0.5 ve 0.7 iken KEKK daha iyi sonuçlar vermiştir. Fakat yüksek düzeyli ilişki durumunda (0.9) ise TBR daha iyi bir yöntem olarak görülmüştür.

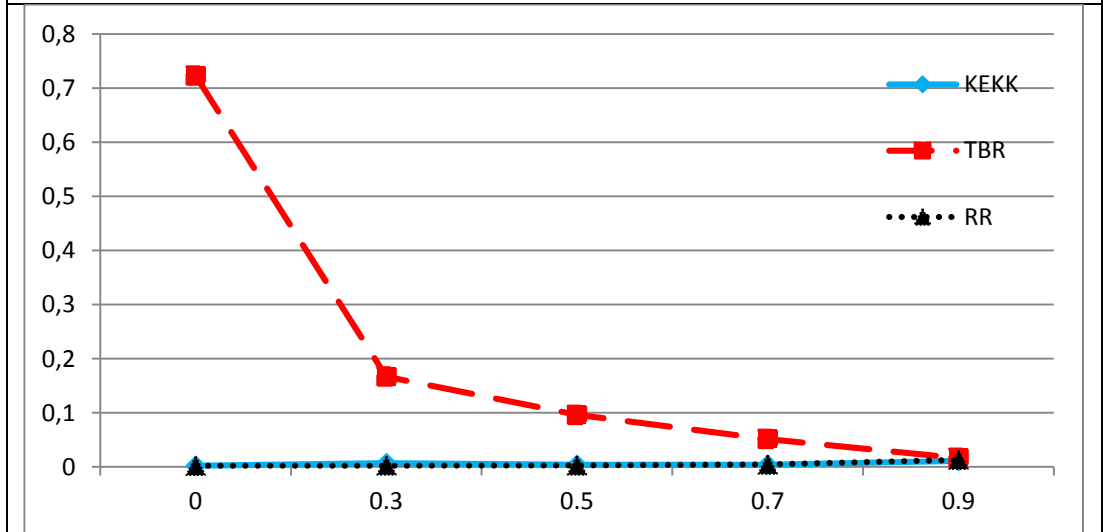
Örneklem büyüklüğü $n=200$ ve 10 değişkenli regresyon modelinde (Şekil 4.4.C) bağımsız değişkenler arasında ilişki düzeyi 0.0 ve 0.3 iken RR yöntemi daha iyi sonuçlar vermiştir. İlişki düzeyi 0.5 ve 0.7 iken KEKK daha iyi sonuçlar vermiştir. Fakat yüksek düzeyli ilişki durumunda (0.9) ise TBR daha iyi bir yöntem olarak görülmüştür.



Şekil 4.5.A – Değişken Sayısı=4



Şekil 4.5.B – Değişken Sayısı=7



Şekil 4.5.C– Değişken Sayısı=10

Şekil 4. 5. Örneklem Büyüklüğü:500 iken elde edilen HKO değerleri

Örneklem büyüklüğü $n=500$ iken farklı değişken sayılarına ait regresyon modellerine ait elde edilen HKO değerleri Şekil 4.5. 'te verilmiştir.

Örneklem büyüklüğü $n=500$ ve 4 değişkenli regresyon modelinde (Şekil 4.5.A) bağımsız değişkenler arasında ilişki yokken, KEKK yöntemi diğerlerine göre daha iyi sonuçlar vermiştir. Ancak RR yöntemi de iyi sonuçlar vermiştir. İlişki düzeyi 0,3 olduğunda, TBR 'nin HKO değeri düşmeye başlamıştır. Fakat yine de TBR yöntemi diğerlerine göre daha kötü sonuç vermiştir. 0,3 ilişki düzeyinde en iyi sonuçlar RR ile elde edilmiştir. İlişki düzeyi biraz daha artırıldığında (0,5), KEKK yöntemi diğerlerine göre daha iyi sonuçlar vermiştir. İlişki değeri hem 0,7 iken hem KEKK hem de RR metotları ile edilen HKO değerleri 0,5 ilişki düzeyine göre artmıştır. Yine de 0,7 düzeyde en iyi sonuçlar RR ile elde edilmiştir. Bağımsız değişkenler arasında çok yüksek düzeyde ilişki olması durumunda yine RR yöntemi diğerlerine göre HKO bakımından daha anlamlı modeller tahmin etmektedir.

Örneklem büyüklüğü $n=500$ ve 7 değişkenli regresyon modelinde (Şekil 4.5.B) tüm ilişki düzeylerinde RR yöntemi diğer 2 yöntemeye göre çok daha iyi sonuçlar vermiştir. Fakat KEKK yöntemi de oldukça yakın sonuçlar vermiştir.

Örneklem büyüklüğü $n=500$ ve 10 değişkenli regresyon modelinde (Şekil 4.5.C) ilişki yokken ve ilişki düzeyleri düşük iken (0.3 ve 0.5) RR yöntemi daha iyi sonuçlar vermiştir. İlişki düzeyleri yüksek iken (0.7 ve 0.9) KEKK yöntemiyle daha iyi sonuçlar bulunmuştur.

4 değişkene sahip bir modelde örneklem büyüklüğü $n \leq 200$ olduğu durumlarda ilişki düzeyi çok az iken KEKK ve RR metodları TBR ye göre daha iyi sonuçlar vermektedir. Sadece ilişki düzeyi çok yüksek (0.9) olduğunda TBR yöntemi diğer 2 yöntemeye göre daha iyi sonuçlar vermiştir. Örneklem büyüklüğü $n > 200$ olduğunda yine düşük düzeyli ilişkilerde KEKK ve RR yöntemi, TBR ye göre çok iyi daha sonuçlar verirken, çok yüksek ilişki düzeyindeki durumlarda RR daha iyi sonuçlar verdiği görülmüştür.

7 değişkene sahip regresyon modellerinde tüm örneklem büyüklüklerinde ilişki düzeyleri düşük iken (0.0 ve 0.3) RR yöntemi daha iyi sonuçlar vermiştir. İlişki düzeyi 0.5 ve 0.7 iken $n \leq 200$ için KEKK yöntemi, $n=500$ iken RR yöntemi tercih edilebilir görülmüştür. İlişki düzeyi 0.9 için ise $n \leq 200$ için TBR, $n=500$ için ise RR daha verimli sonuçlar vermiştir.

10 deęiřkene sahip regresyon modellerinde iliřki d zeyleri 0.0 ve 0.3 iken t m  rneklem b y kl r  i in RR y ntemi daha iyi sonu lar vermiřtir. 0.5 iliřki d zeyine g re, $n \leq 200$ iken KEKK, $n=500$ i in ise RR metodu daha saęlıklı sonu lar vermiřtir. 0.7 iliřki d zeyinde t m  rneklem b y kl kl r  i in KEKK y ntemi daha iyi sonu lar vermiřtir.  ok y ksek iliřki d zeyinde (0.9) ise, $n \leq 200$ i in TBR, $n=500$ i in ise KEKK daha verimli sonu lar vermiřtir.



5. SONUÇLAR VE TARTIŞMA

Yapılan çalışmada, çok değişkenli doğrusal regresyon (çoklu regresyon) modellerinde ortaya çıkan çoklu bağlantı sorununu çözmek için kullanılan yöntemlerden temel düzeyde KEKK, RR ve TBR yöntemleri karşılaştırılmıştır. Simülasyon ile elde edilen veri seti için, değişken sayıları, örneklem büyüklükleri ve ilişki düzeylerine göre bu yöntemlerin kendi aralarında hangisinin etkin sonuçlar verdiği farklılıklar gösterilmiştir. Özellikle, ilişki düzeylerinin belli derecelerinde, hangi yöntemin daha tercih edilebilir sorusuna yanıt verilebildiği gözlenmiştir. İlişki düzeyinin çok düşük olduğu (≤ 0.3) durumlarda RR yöntemi daha olumlu sonuçlar elde edilirken, orta düzeyde ilişkiye sahip durumlarda KEKK yöntemi daha iyi sonuçlar vermiştir. Sadece çok yüksek düzeydeki ilişki düzeylerde (0.9) genel olarak TBR yöntemi diğer iki yöntemle göre daha iyi sonuçlar vermiştir.

Yine yapılan bu tez çalışmasında elde edilen bir başka bulguda şu şekildedir: Örneklem büyüklüğü yüksek olduğunda ($n=500$) RR regresyon daha iyi sonuçlar vermiştir.

Literatürde yapılan çalışmaların pek çoğunda RR yönteminin diğer iki yöntemle daha uygun tahminler yaptığı görülmüştür. Ergüneş (2004) yaptığı çalışmada RR ile KEKK yöntemlerini karşılaştırarak, daha etkin parametre tahmini yapması nedeniyle RR yönteminin KEKK yerine kullanılması gerektiği belirtmiştir. Albayrak (2005) çalışmasında RR, TBR ve KEKK yöntemlerini karşılaştırmış, RR ve TBR yöntemlerinin KEKK 'ye göre daha uygun tahminler sağladığını ifade etmiştir. Pamukçu (2010), aralarında yüksek çoklu bağlantı bulunan bağımsız değişkenler kullanarak yaptığı çalışma sonucunda RR 'nin KEKK yöntemine göre daha uygun tahminler ortaya koyduğu belirtmiştir. Bizim çalışmamızda ise yüksek düzeydeki ilişki durumlarında ($n=500$ hariç) TBR yönteminin daha sağlıklı sonuçlar verdiği gözlemlenmiştir. Eker (2011) KEKK, TBR ve RR yöntemlerini karşılaştırmış, en düşük HKO ve en büyük R^2 değerlerini RR yöntemiyle elde etmiştir. Fakat parametrelerin önem testlerine göre TBR 'nin daha tutarlı olduğu sonucunu elde etmiştir.

Arařtırma sonucunda her üç yntemin, hangi durumlarda daha saėlıklı ve seilebilir olduėu ile ilgili net kararlara varılabilmektedir. rnekleme byklėu, iliřki dzeyi ve deėiřken sayıları dikkate alınarak en uygun yntem seilebilir.



KAYNAKLAR

- Aktaş, C. (2007) Çoklu Bağntı ve Liu Kestiricisiyle Enflasyon Modeli İçin Bir Uygulama, *ZKÜ Sosyal Bilimler Dergisi*. 3(6): 67–79.
- Albayrak, S, A. (2005) Çoklu Doğrusal Bağlantı Halinde En Küçük Kareler Tekniğinin Alternatifi Yanlı Tahmin Teknikleri Ve Bir Uygulama, *ZKÜ Sosyal Bilimler Dergisi*. 1(1), 105-126.
- Alpar, R. (2003) *Uygulamalı çok değişkenli istatistiksel yöntemlere giriş*, 1. Nobel Yayın, Yayın No:452, Ankara.
- Büyükuysal , M.Ç. (2010) *Ridge Regresyon Analizi Ve Bir Uygulama*, Uludağ Üniversitesi, Sağlık Bilimleri Enstitüsü, Biyoistatistik Anabilim Dalı, Yüksek Lisans Tezi, 42-43s.
- Cankaya, S., Kayaalp, G.T., Sangun, L. Tahtali Y. and Akar, M. (2006) *A Comparative Study of Estimation Methods for Parameters in Multiple Linear Regression Model*, J. Appl. Anim. Res., 29(1), 43-47.
- Çankaya, S., Altop, A., Kul, E., Erener, G. (2009) Faktör Analiz Skorları Kullanılarak Karayaka Kuzularında Canlı ağırlık Tahmini, *Anadolu Tarım Bilim. Derg.*, 24(2):98-102.
- Eker, S. (2011) *Regresyon Analizinde Çoklu Bağlantı Durumunda En Küçük Kareler, Ridge Regresyon Ve Temel Bileşenler Yaklaşımlarının Karşılaştırılması*, Ondokuz Mayıs Üniversitesi, Fen Bilimleri Enstitüsü, Zootekni Anabilim Dalı, Yüksek Lisans Tezi, 11-36s.
- Ergüneş, E. (2004) *En Küçük Kareler Yöntemi İle Ridge Regresyon Yönteminin Karşılaştırılmalı Olarak İncelenmesi*, Çukurova Üniversitesi, Fen Bilimleri Enstitüsü, Zootekni Anabilim Dalı, Yüksek Lisans Tezi, 78s.
- Firinguetti, L., Kibria G., ve Araya R. (2016) *Study of Partial Least Squares and Ridge Regression Methods*, Communications in Statistics - Simulation and Computation, DOI: 10.1080/03610918.2016.1210168
- Graham, M. H. (2003) Confronting Multicollinearity in Ecological Multiple

Regression, *Ecology*, 84:2809–2815. [doi:http://dx.doi.org/10.1890/02-3114].

Karadavut, U., Genç, A., Tozluca, A., Kınacı, D., Aksoyak, Ş., Palta, Ç., Pekgör, A. (2005) Nohut (*Cicer arietinum* L.) Bitkisinde Verime Etki Eden Bazı Karakterlerin Alternatif Regresyon Yöntemleriyle Karşılaştırılması, *Tarım Bilimleri Dergisi*, 2005, 11 (3) 328-333.

Karakaş, S. (2008) *Çoklu doğrusal bağlantı problemi ve yanlı regresyon tahmincileri*, İstanbul Üniversitesi, Sosyal Bilimler Enstitüsü, Ekonometri Anabilim Dalı, Yüksek Lisans Tezi, 159s.

Kidwell, J. S. ve Brown, L. H. (1982) *Ridge Regression as a Technique for Analyzing Models with Multicollinearity*, J. Marriage and Family, Vol. 44(2): 287-299.

Kurtuluş, M. (2001) *Ridge regresyon Üzerine Bir Çalışma*, Gazi Üniversitesi, Fen Bilimleri Enstitü. Yüksek Lisans Tezi, 35.

Mahesh, S., Jayas, D. S., Paliwal, J., ve White, N. D. G. (2014) *Comparison of Partial Least Squares Regression (PLSR) and Principal Components Regression (PCR) Methods for Protein and Hardness Predictions using the Near-Infrared (NIR) Hyperspectral Images of Bulk Samples of Canadian Wheat*, Food and Bioprocess Technology, 8(1), 31–40.https://doi.org/10.1007/s11947-014-1381-z

Orhunbilge, N. (2000) *Uygulamalı Regresyon ve Korelasyon Analizi*, Avcioğlu Basım Yayın, İstanbul.

Ortabaş, N. (2001) *Principal components in the problem of multicollinearity*, Dokuz Eylül Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, 45s.

Pamukçu, E. (2010) *Sistolik Kan Basıncını Etkileyebilecek Faktörlerin Ridge Regresyon Analizi İle İncelenmesi Ve Çoklu Bağlantı Problemi*, Fırat Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Anabilim Dalı, Yüksek Lisans Tezi, 84s.

Phatak, A., De Jong, S. (1997) The geometry of partial least squares, *Journal of Chemometrics*, 11, 311–338.

Polat, E. (2009) *Kısmi En Küçük Kareler Regresyonu*, Hacettepe Üniversitesi, Fen Bilimleri Enstitüsü, İstatistik Ana Bilim Dalı, Yüksek Lisans Tezi, 60-63.

Rathert, T. Ç., Üçkardeş, F., Narinç, D. ve Aksoy, T. (2011) *Comparision of Principal*

Component Regression with the Least Square Method in Prediction of Internal Egg Quality Characteristics in Japanese Quails, Kafkas Üniversitesi, Veterinerlik Fakültesi Dergisi. 17 (5): 687-692.

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.

Topçubaşı, A.S. ve Billor, N. (2006) *Yanlı Kestiricilerin Bir Sınıfı Ve Bu Sınıf İçin Tanılama Ölçüleri*, Erişim adresi: <http://idari.cu.edu.tr/sempozyum/bil26.htm>, Erişim Tarihi: 10.10.2011.

Yeniay, Ö. ve Göktaş, A. (2002) A comparison of partial Least Squares regression with other prediction methods, *Hacettepe Journal of Mathematics and Statistics*, 31, 99-111.

Yolacan, S., Yazıcı, B. ve Agaoglu, E. (2005) *Artificial Neural Network Approach as an Alternative to Ridge Regression on an Economic Problem*, 4th WSEAS Int. Conf. on Non-linear Analysis, Non-Linear Systems and Chaos, Sofia, Bulgaria, October 27-29, 2005, pp:120-124.

Ek A. Simülasyon Çalışması Kullanılan Kodlar

Bu çalışmada, simülasyon sonucu elde edilen tüm veriler, Minitab 16.0 programı için hazırlanmış kodlar ile elde edilmiştir. Daha sonra bu alanda benzeri çalışmalar yapmak isteyen araştırmacılar için kullanılmak üzere bu kodlar aşağıda verilmiştir.

Değişken sayısı: 4 ve ilişki düzeyi: 0.5 için yazılan Minitab kodları:

Random k400 c7-c10;

Normal 0 1.

Random k400 c11;

Normal 0 1.

*let c3=((1-(k300*k300))**0,5)*c7+k300*c10 # ro değerinin k300 olarak belirlenmesi*

*let c4=((1-(k300*k300))**0,5)*c8+k300*c10*

*let c5=((1-(k300*k300))**0,5)*c9+k300*c10*

*let c6=((1-(k300*k300))**0,5)*c10+k300*c10*

let c1=1+c3+c4+c5+c6+c11

set c2

l(1:1/1)k400 # k400 gözlem sayısını vermektedir.

End.

#PLS başlangıç

PLS c1 = c3-c6;

NComp 2;

Coding 1;

coefficients c19.

*let c22(k1)=mean((1-c19)**2)*

#PLS bitiş.

#PCR

copy c3-c6 m1

trans m1 m2

multi m2 m1 m3

eigen m3 c26 m4.

copy m4 c50-c53

multi m1 m4 m5

```

copy m5 c28-c31
regress c1 2 c28 c29;
coefficients c49;
constant;
Brief 0.
let c55(1)=c49(1)
let c55(2)=c49(2)*c50(1)+c49(3)*c51(1)
let c55(3)=c49(2)*c50(2)+c49(3)*c51(2)
let c55(4)=c49(2)*c50(3)+c49(3)*c51(3)
let c55(5)=c49(2)*c50(4)+c49(3)*c51(4)
let c23(k1)=mean((1-c55)**2)
#PCR bitiş.

```

#Ridge başlangıç

```

copy c2-c6 m1
trans m1 m2
multi m2 m1 m3
set c47
1( 1:1/1 )5

```

End.

Diagonal C47 m5.

```

regress c1 4 c3-c6;
coefficients c44;
MSE k3;

```

Constant;

Brief 0.

```

let k4=k3/(maximum(C44))
multi k4 m5 m5
add m5 m3 m9
inver m9 m6
multi m2 c1 m4
multi m6 m4 m7
copy m7 c60
let c24(k1)=mean((1-c60)**2)
#Ridge bitiş.

```

let k1=k1+1

Değişken sayısı: 7 ve ilişki düzeyi: 0.5 için yazılan Minitab kodları:

Random k400 c10-c16;

*Normal 0 1. # standart sapma değerinin k155 olarak
belirlenmesi*

Random k400 c17;

Normal 0 1.

*let c3=((1-(k300*k300))**0,5)*c10+k300*c16 # ro değerinin k300 olarak
belirlenmesi*

*let c4=((1-(k300*k300))**0,5)*c11+k300*c16*

*let c5=((1-(k300*k300))**0,5)*c12+k300*c16*

*let c6=((1-(k300*k300))**0,5)*c13+k300*c16*

*let c7=((1-(k300*k300))**0,5)*c14+k300*c16*

*let c8=((1-(k300*k300))**0,5)*c15+k300*c16*

*let c9=((1-(k300*k300))**0,5)*c16+k300*c16*

let c1=1+c3+c4+c5+c6+c7+c8+c9+c17

set c2

1(1:1/1)k400 # k400 gözlem sayısını vermektedir.

End.

#PLS başlangıç

PLS c1 = c3-c9;

NComp 2;

Coding 1;

coefficients c19.

*let c22(k1)=mean((1-c19)**2)*

#PLS bitiş.

#PCR

copy c3-c9 m1

trans m1 m2

multi m2 m1 m3

eigen m3 c26 m4.

copy m4 c60-c66

multi m1 m4 m5

copy m5 c28-c34

```

regress c1 2 c28 c29;
coefficients c59;
constant;
Brief 0.
let c68(1)=c59(1)
let c68(2)=c59(2)*c60(1)+c59(3)*c61(1)
let c68(3)=c59(2)*c60(2)+c59(3)*c61(2)
let c68(4)=c59(2)*c60(3)+c59(3)*c61(3)
let c68(5)=c59(2)*c60(4)+c59(3)*c61(4)
let c68(6)=c59(2)*c60(5)+c59(3)*c61(5)
let c68(7)=c59(2)*c60(6)+c59(3)*c61(6)
let c68(8)=c59(2)*c60(7)+c59(3)*c61(7)
let c23(k1)=mean((1-c68)**2)
#PCR bitiş.

```

#Ridge başlangıç

```

copy c2-c9 m1
trans m1 m2
multi m2 m1 m3

```

```

set c47

```

```

1( 1:1/1 )8

```

End.

Diagonal C47 m5.

```

regress c1 7 c3-c9;

```

```

coefficients c44;

```

```

MSE k3;

```

```

Constant;

```

Brief 0.

```

let k4=k3/(maximum(C44))

```

```

multi k4 m5 m5

```

```

add m5 m3 m9

```

```

inver m9 m6

```

```

multi m2 c1 m4

```

```

multi m6 m4 m7

```

```

copy m7 c70

```

```
let c24(k1)=mean((1-c70)**2)
```

```
#Ridge bitiş.
```

```
let k1=k1+1
```

Değişken sayısı: 10 ve ilişki düzeyi: 0.5 için yazılan Minitab kodları:

```
Random k400 c13-c22;
```

```
Normal 0 1. # standart sapma değerinin k155 olarak belirlenmesi
```

```
Random k400 c23;
```

```
Normal 0 1.
```

```
let c3=((1-(k300*k300))**0,5)*c13+k300*c22 # ro değerinin k300 olarak belirlenmesi
```

```
let c4=((1-(k300*k300))**0,5)*c14+k300*c22
```

```
let c5=((1-(k300*k300))**0,5)*c15+k300*c22
```

```
let c6=((1-(k300*k300))**0,5)*c16+k300*c22
```

```
let c7=((1-(k300*k300))**0,5)*c17+k300*c22
```

```
let c8=((1-(k300*k300))**0,5)*c18+k300*c22
```

```
let c9=((1-(k300*k300))**0,5)*c19+k300*c22
```

```
let c10=((1-(k300*k300))**0,5)*c20+k300*c22
```

```
let c11=((1-(k300*k300))**0,5)*c21+k300*c22
```

```
let c12=((1-(k300*k300))**0,5)*c22+k300*c22
```

```
let c1=1+c3+c4+c5+c6+c7+c8+c9+c10+c11+c12+c23
```

```
set c2
```

```
1( 1:1/1 )k400 # k400 gözlem sayısını vermektedir.
```

```
End.
```

```
#PLS başlangıç
```

```
PLS c1 = c3-c12;
```

```
NComp 2;
```

```
Coding 1;
```

```
coefficients c25.
```

```
let c26(k1)=mean((1-c25)**2)
```

```
#PLS bitiş.
```

```
#PCR
```

```
copy c3-c12 m1
```

```

trans m1 m2
multi m2 m1 m3
eigen m3 c32 m4.
copy m4 c70-c79
multi m1 m4 m5
copy m5 c34-c43
regress c1 2 c34 c35;
coefficients c69;
constant;
Brief 0.
let c81(1)=c69(1)
let c81(2)=c69(2)*c70(1)+c69(3)*c71(1)
let c81(3)=c69(2)*c70(2)+c69(3)*c71(2)
let c81(4)=c69(2)*c70(3)+c69(3)*c71(3)
let c81(5)=c69(2)*c70(4)+c69(3)*c71(4)
let c81(6)=c69(2)*c70(5)+c69(3)*c71(5)
let c81(7)=c69(2)*c70(6)+c69(3)*c71(6)
let c81(8)=c69(2)*c70(7)+c69(3)*c71(7)
let c81(9)=c69(2)*c70(8)+c69(3)*c71(8)
let c81(10)=c69(2)*c70(9)+c69(3)*c71(9)
let c81(11)=c69(2)*c70(10)+c69(3)*c71(10)
let c27(k1)=mean((1-c81)**2)
#PCR bitiş.

```

```

#Ridge başlangıç
copy c2-c12 m1
trans m1 m2
multi m2 m1 m3
set c56
1( 1:1/1 )11
End.
Diagonal C56 m5.
regress c1 10 c3-c12;
coefficients c53;
MSE k3;

```

Constant;

Brief 0.

let k4=K3/(maximum(C53))

multi k4 m5 m5

add m5 m3 m9

inver m9 m6

multi m2 c1 m4

multi m6 m4 m7

copy m7 c83

*let c28(k1)=mean((1-c83)**2)*

#Ridge bitiş.

let k1=k1+1

ÖZGEÇMİŞ

Adı Soyadı : İsmail BAĞCI
Doğum Yeri : Çelikhan / ADIYAMAN
Doğum Yılı : 1977
Medeni Hali : Evli

Eğitim ve Akademik Durumu

Lise : (1989-1992) Eski Malatya Lisesi
Lisans : (1994-1999) Muğla Üniversitesi / İstatistik ve Bilgisayar Bilimleri
Yabancı Dil : İngilizce

İş Tecrübesi

2000-2002 Güney Ege Bilgisayar – Satış Sorumlusu
2005- Muğla Sıtkı Koçman Üniversitesi Enformatik Bölümü -
Okutman