ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS

Nurşen YILDIRIM

EN KÜÇÜK KARELER, RİDGE REGRESYON VE ROBUST REGRESYON YÖNTEMLERİNDE ANALİZ SONUÇLARINA AYKIRI DEĞERLERİN ETKİLERİNİN BELİRLENMESİ

ZOOTEKNİ ANABİLİM DALI

ADANA, 2010

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

EN KÜÇÜK KARELER, RİDGE REGRESYON VE ROBUST REGRESYON YÖNTEMLERİNDE ANALİZ SONUÇLARINA AYKIRI DEĞERLERİN ETKİLERİNİN BELİRLENMESİ

Nurşen YILDIRIM

YÜKSEK LİSANS TEZİ

ZOOTEKNİ ANABİLİM DALI

Bu tez **04 /01 /2010** Tarihinde Aşağıdaki Jüri Üyeleri Tarafından **Oybirliği** İle Kabul Edilmiştir.

İmza.....…	İmza.....…....	İmza.....…….
Prof. Dr. G. Tamer KAYAALP	Prof. Dr. Mustafa AKAR	Doç. Dr.Suat ŞAHİNLER
DANIŞMAN	ÜYE	ÜYE

Bu tez Enstitümüz Zootekni Anabilim Dalında hazırlanmıştır.
Kod No:

Prof. Dr. İlhami YEĞİNGİL
Enstitü Müdürü

Bu Çalışma Ç.Ü. Araştırma Fonu Tarafından Desteklenmiştir.
Proje No:ZF2008YL44

Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve Sanat Eserleri Kanunundaki hükümlere tabidir.

ÖZ

YÜKSEK LİSANS TEZİ

EN KÜÇÜK KARELER, RİDGE REGRESYON VE ROBUST REGRESYON YÖNTEMLERİNDE ANALİZ SONUÇLARINA AYKIRI DEĞERLERİN ETKİLERİNİN BELİRLENMESİ

NURŞEN YILDIRIM

ÇUKUROVA ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
ZOOTEKNİ ANABİLİM DALI

Danışman : Prof. Dr. G. Tamer KAYAALP
Yıl: 2009, Sayfa: 69

Jüri : Prof. Dr. G. Tamer KAYAALP
Prof. Dr. Mustafa AKAR
Doç. Dr. Suat ŞAHİNLER

Çoklu doğrusal regresyon analizinin varsayımları sağlanıyorsa En Küçük Kareler Yöntemi (EKK) en iyi tahminleri vermektedir. Ancak veri grubundaki bağımsız değişkenler arasında doğrusal bir ilişki ve gözlem değerleri arasında aykırı değerler varsa varsayımlar sağlanamaz. Bu durumda regresyon katsayıları, "t" değerleri, "F" değerleri, hata kareler ortalamaları ve belirtme katsayıları değişmektedir.

Bu çalışmada çoklu bağlantı problemi ve aykırı değerlerin etkileri araştırılmıştır. Çoklu bağlantı problemi için Ridge Regresyon (RR) yöntemi, aykırı değerlerin belirlenebilmesi içinde bazı hata hesaplama yöntemleri önerilmiştir. EKK, Ridge ve M-tahmin edici kullanılarak parametre tahminleri aykırı değer varlığında ve yokluğunda gerçekleştirilmiştir. Karşılaştırma kriteri olarak belirtme katsayısı (R^2), hata kareler ortalaması, F ve t değerleri kullanılmıştır.

Anahtar Kelimeler: En Küçük Kareler Yöntemi, Ridge Regresyon Yöntemi, M-Tahmin Edici, Aykırı Değer, Çoklu Bağlantı

ABSTRACT

MSc THESIS

DETERMINATION THE EFFECTS OF OUTLIERS AT THE LEAST SQUARES, RIDGE REGRESSION AND ROBUST REGRESSION ANALYSIS RESULTS

NURŞEN YILDIRIM

DEPARTMENT OF ANİMAL SCİENCE
INSTITUTE OF NATURAL AND APPLIED SCIENCES
UNIVERSITY OF ÇUKUROVA

Supervisor : Prof. Dr. G. Tamer KAYAALP
Year: 2009, Pages: 69

Jury: Prof. Dr. G. Tamer KAYAALP
Prof. Dr. Mustafa AKAR
Assoc. Prof. Dr. Suat ŞAHİNLER

Least square method given the best prediction if the assumption of multiple linear regression analysis was provided. Howewer there is any linear relation between independent variables and outliers in the data group, the assumptions can't be provided. In this case regression coefficients "t" values, "F" values, error of mean square and coeeffcient of determination are changed.

Multicollinearity and outliers effect are investigated in this study. Ridge regression (RR) method is suggested for multicollinearity and some outlier calculation method for outliers. Parameter prediction are realized both in outliers existance and nonexistance via Least square, Ridge and M-estimatior. R^2 value, error of mean square, "F" nad "t" values are used as comparision criteria.

KeyWords: Least Square Method, Ridge Regression (RR) Method, M-Estimatior, Outlier, Multicollinearity

TEŞEKKÜR

Bu çalışmanın hazırlanmasında yardımını esirgemeyen danışmanım Prof. Dr. G. Tamer KAYAALP' e yardımlarından dolayı teşekkürü bir borç bilirim. Manevi destekleri ile her zaman yanımda olan değerli hocalarım Prof. Dr. Oya IŞIK, Yrd. Doç. Dr. Gonca KESER ve Yrd. Doç. Dr. Seval SÜZÜLMÜŞ'e teşekkür ediyorum. Ayrıca maddi ve manevi desteğini benden esirgemeyen sevgili annem Cevriye YILDIRIM'a ve babam Mehmet YILDIRIM'a, kardeşlerim Gürcan YILDIRIM ve Ayşen YILDIRIM'a sonsuz teşekkürlerimi sunuyorum. Tez yazım aşamasında yanımda olan ve her türlü manevi desteğini aldığım Mustafa Besim KESER' e şükranlarımı sunuyorum.

İÇİNDEKİLER **SAYFA**

ÖZ…….. .I
ABSTRACT……. ..II
TEŞEKKÜR…… ..III
İÇİNDEKİLER…..…….…….……....…....IV
ÇİZELGE DİZİNİ……...…….…….....VI
ŞEKİLLER DİZİNİ……..VIII
SİMGE VE KISALTMALAR…….…….....IX
1. GİRİŞ……...……1
2. ÖNCEKİ ÇALIŞMALAR…….....…. 4
3. MATERYAL VE YÖNTEM……..10
3.1. Materyal…….. 10
3.2. Yöntem.……10
3.2.1. En Küçük Kareler Yöntemi…….……..... 10
3.2.1.1. Çoklu Doğrusal Regresyon Modelinde
Parametre Tahminlerinin Uygunluğu……...17
3.2.1.2. Çoklu Doğrusal Regresyon Modelinin Varsayımları…….18
3.2.1.3. Çoklu Bağlantı…….…….…21
3.2.1.4. Çoklu Bağlantının Etkileri…….….. 21
3.2.1.5. Çoklu Bağlantıyı Belirleme Yöntemleri...……. 23
3.2.2. Ridge Regresyon Yöntemi…...……..24
3.2.2.1. Ridge Regresyon Yönteminin Kullanım Amaçları……...26
3.2.2.2. Ridge Tahmin Edicisinin Yanlı Olması…….…...26
3.2.2.3. Ridge Tahmin Edicisinin Hata Kareler Ortalaması, Varyansı
ve Hata Kareler Toplamı…...…….…..26
3.2.2.4. Ridge Parametresinin Saptanması …….....28
3.2.3. M-Tahmin Edici…… ……29
3.2.4..Aykırı Değerlerin Belirlenmesinde Kullanılan Başlıca Testler…30
3.2.4.1. Standartlaştırılmış Hatalar……..……..…30
3.2.4.2. Student Türü Hatalar……31

3.2.4.3. R-Student Türü Hatalar. ……..…..…….31
3.2.4.4. Ortalama Değişim (Mean-Shift) Aykırı
Değer Modellemesi……..……..……31
3.2.4.5. Bonferroni Testi……..…….32
4.ARAŞTIRMA BULGULARI..... 33
4.1. En Küçük Kareler Yönteminin Bulguları ……..…… 33
4.2. Çoklu Bağlantının Belirlenmesi……..……34
4.3. Ridge Regresyon Yönteminin Bulguları…….....35
4.3.1. Ridge Parametresini (k^*) Belirleme…….. ……..35
4.4. M-Tahmin Edici Bulguları……..……38
4.5. Aykırı Değerlerin İncelenmesi…….. ……..39
4.6. Aykırı Değerler Veri Setinden Uzaklaştırıldığında Elde Edilen EKK
Bulguları…….…….44
4.7. Aykırı Değerler Uzaklaştırıldığında Elde Edilen Verilerde Çoklu
Bağlantının Belirlenmesi…….....45
4.8. Aykırı Değerler Çıkarıldıktan Sonra Ridge Regresyon Yönteminin
Bulguları……...……....46
4.8.1. Ridge Parametresini (k^*) Belirleme ……...……...46
4.9. Aykırı Değerler Çıkarıldıktan Sonra M-Tahmin Edici
Yönteminin Bulguları……..…….…..…..48
5. TARTIŞMA VE SONUÇ……..……...…….....51
KAYNAKLAR……...……..55
ÖZGEÇMİŞ……...……...59
EKLER……...……...60

ÇİZELGELER DİZİNİ **SAYFA**

Çizelge 4.1. EKK Yöntemi ile Verilere Ait Parametre Tahminleri ve Önem Testleri.....33
Çizelge 4.2. Verilere Ait Varyans Analiz Tablosu..... 34
Çizelge 4.3. R_j^2 Değerleri..... 35
Çizelge 4.4. VBF, l_j ve Koşul Sayısı Değerleri.....35
Çizelge 4.5. k* ve VBF Değerleri.....36
Çizelge 4.6. k*=0.2 Değeri İçin Varyans Analiz Tablosu.....37
Çizelge 4.7. EKK ve RR Yöntemlerine Ait VBF Değerleri.....37
Çizelge 4.8. E.K.K. ve R.R. Yöntemine Ait R^2 ve HKO Değerleri.....38
Çizelge 4.9. M-Tahmin Edici Kullanılarak Elde Edilen Analiz Sonuçları.....38
Çizelge 4.10. Aykırı Değerler İle İlgili İstatistikler……...……...40
Çizelge4.11.Aykırı Değerler Çıkarıldıktan Sonra Elde Edilen EKK Sonuçları.....44
Çizelge4.12.Aykırı Değerler Uzaklaştığında Elde Edilen Varyans Analiz Tablosu..... 44
Çizelge 4.13. R_j^2 Değerleri…….....45
Çizelge4.14. VBF, l_j ve Koşul Sayısı ve Koşul İndeksi Değerleri.....46
Çizelge 4.15. k* ve VBF Değerleri…….....47
Çizelge 4.16. Aykırı Değerler Çıkarıldıktan Sonra k*=0.2 Değeri İçin RR Yöntemi İle Elde Edilen Varyans Analiz Tablosu….....48
Çizelge 4.17. EKK ve RR Yöntemlerine Ait VBF Değerleri….....48
Çizelge 4.18. Aykırı Değerler Veri Setinden Uzaklaştırıldıktan Sonra M- Tahmin Edici Yöntemi İle Elde Edilen Analiz Sonuçları.....49
Çizelge 4.19. Aykırı Değerler Uzaklaştırıldıktan Sonra Elde Edilen HKO ve R^2 Değerleri…….....49

Çizelge 4.20. Aykırı Değer Varlığında EKK ve M-Tahmin Ediciden
Elde Edilen Analiz Sonuçları……50
Çizelge 4.21. Aykırı Değer Yokluğunda EKK ve M-Tahmin Ediciden
Elde Edilen Analiz Sonuçları……50

ŞEKİLLER DİZİNİ **SAYFA**

Şekil 1. NSCC Paket Programında Ridge Regresyon Menüsünün Açılması…….67
Şekil 2. Değişkenlerin Tanımlanması……..…68
Şekil 3. Analiz Sonuçları…….…...68
Şekil 4. Ridge Parametresi ve VIF Değerleri……..…69

SİMGELER VE KISALTMALAR

EKK	: En Küçük Kareler
RR	: Ridge Regresyon
Y	: Bağımlı Değişken
X	: Bağımsız Değişken
k	: Bağımsız Değişken Sayısı
n	: Gözlem Sayısı
k^*	: Ridge Parametresi
b	: EKK Tahmin Edicisi
b_R	: Ridge Regresyon Tahmin Edicisi
b_{GR}	: Genelleştirilmiş Ridge Regresyon Tahmin Edicisi
b_{YR}	: Yönlendirilmiş Ridge Regresyon Tahmin Edicisi
R^2	: Belirtme Katsayısı
HKO	: Hata Kareler Ortalaması
LMS	: En Küçük Medyan Kareler (Least Median of Squares)
LTS	: En Küçük Budanmış Kareler (Least Trimmed of Squares)
WLS	: Ağırlıklandırılmış En Küçük Kareler
adf	:Asit Deterjan Fiber
ndf	:Nötr Deterjan Fiber

1. GİRİŞ

Regresyon analizi, bağımlı değişken adı verilen değişken ile bağımsız değişkenler arasındaki ilişkiye model uydurmak ya da açıklamak için kullanılan bir yöntemdir. Bağımsız değişken sayısı 1 olduğunda basit regresyon, birden fazla olduğunda çoklu (multiple) regresyon olarak isimlendirilir.

Bağımlı değişken genellikle sürekli veri olmalıdır. Fakat bağımsız değişkenler sürekli, kategorik veya farklı veriler olabilmektedir. Regresyon analizi sayesinde;

1. Gelecekteki gözlemlerin tahmini,
2. Bağımlı değişken ile bağımsız değişken arasındaki ilişkinin veya etkinin değerlendirilmesi,
3. Veri yapısı hakkında genel bir tanımlamanın yapılabilmesi mümkün olmaktadır.

Çoklu regresyon analizi yöntemi sayesinde bağımlı değişken ile bağımsız değişkenler arasında ilişki kurularak, parametre tahminleri yapılmaktadır. En küçük kareler yöntemi bu parametre tahminlerini gerçekleştirebilmek için en yaygın kullanılan klasik yöntemlerdendir.

Regresyon analizi ilk olarak 18. yüzyılda ortaya çıkmaya başlamıştır. 1805 yılında Legendre En Küçük Kareler Metodunu geliştirmiştir. 1809 yılında Gauss geliştirdiği metot ile, hatalar normal dağılımlı olduğunda en küçük kareler yönteminin en uygun çözüm olduğunu göstermiştir (Faraway, 2005)

Regresyon analizinde en küçük kareler yöntemi, gözlem değerleri, değişkenler ve hata hakkında birtakım varsayımların sağlandığı durumlarda geçerlilik kazanır. Bu varsayımlar geçerli olmadıkça yapılmış olan hesaplamaların ve elde edilmiş olan regresyon denklemlerinin istatistiki bir değeri olmaz. Çünkü varsayımların bozulmalarının bu değerler üzerine çok önemli etkileri olabilmektedir. Varsayımların tutmaması uydurulan modelin populasyonu iyi temsil etmediğini gösterir. Buna bağlı olarak elde edilen regresyon denkleminden yapılacak tahminlerin hatalı olma ihtimali yüksek olur (Şahinler, 1997).

Regresyon analizinde; bağımsız değişkenler arasında çoklu bağlantı varlığında çoklu bağlantının etkilerini giderebilmek için öncelikle E.K.K. tahmin edicisi yerine

yanlı tahmin yöntemlerinden elde edilen tahmin edicilerin kullanılması önerilmektedir. Ridge Regresyon Yöntemi çoklu bağlantı varlığında çoklu bağlantının parametre tahminleri üzerine olan olumsuz etkilerini giderebilmek için sıklıkla kullanılan yanlı tahmin yöntemlerindendir.

Regresyon analizinde hatalar normal dağılış göstermiyorsa robust regresyon yöntemleri önerilmektedir. Genellikle robust regresyon metotları en küçük kareler yönteminden daha fazla hesaplama gerektirmektedir (Draper ve Smith, 1998).

Robust regresyon, parametrik modelin varsayımlarının gerçekleşmemesi durumunda tahminlerin kararlılığını arttırmak için tasarlanmış istatistiksel yöntemlerin genel bir sınıfıdır. Bir robust regresyon yöntemi, büyük hataların ağırlıklarını azaltarak bu hataların etkisini düşürmektedir. Aykırı değer ve etkili gözlemlerin tespit edilmesi için kullanılan yöntemler robust regresyonun bir parçası olarak ele alınabilir (Şehirli, 2009).

Bir veri setindeki gözlemlerin hatasından farklı olan gözlemler aykırı değer (outlier) olarak bilinmektedir. Bağımlı değişken içerisinde yani y yönünde kuşkulu bir gözlem varsa aykırı değer (outlier), bağımsız değişkenler içerisinde yani x yönünde kuşkulu bir gözlem varsa uç değer (leverage) olarak isimlendirilmektedir. Veri setinden uzaklaştırıldığında regresyon tahminlerinde farklılığa sebep olan gözlemlere etkili gözlem adı verilmektedir. Aykırı değerler ya da uç değerler etkili gözlem olmayabilirler ancak etkili gözlemler genellikle aykırı değer ya da uç değerlerdir (Freund ve ark., 2006).

Veri seti içerisinde aykırı değer varlığında bu değerler belirlenerek analizden uzaklaştırılmamalıdır. Bunun nedenleri dikkatli bir şekilde araştırılmalıdır (Şahinler, 1997).

Bu çalışmada öncelikle EKK yöntemi çoklu doğrusal regresyon modeline uygulanarak elde edilen parametre tahminleri ile regresyon analizine ait varyans analiz tablosu oluşturulmuştur. Ardından bağımsız değişkenler arasında çoklu bağlantı varlığı araştırılmış, çoklu bağlantının nedenleri, etkileri, çoklu bağlantının belirlenmesi ve etkisinin giderilmesi üzerinde çalışılmıştır. Çoklu bağlantı varlığında, çoklu bağlantı probleminin regresyon analizinde parametre tahminleri üzerine olan etkilerini azaltmak için önerilen yanlı tahmin edicilerden Ridge Regresyon

Yönteminin kullanım amaçları, Ridge Regresyonda parametre tahminlerinin yapılabilmesi için gerekli olan k* değerini belirleme yöntemleri verilerek Ridge Regresyon Yöntemine ait parametre tahminleri elde edilmiştir. Daha sonra Robust Regresyon Yöntemlerinden olan M-Tahmin Edici Yönteminin özellikleri ve çalışma prensibi anlatılarak bu yönteme ait parametre tahminleri yapılmıştır.

Diğer taraftan veri setinde aykırı değer varlığı Standartlaştırılmış Artıklar, Student Türü Artıklar, R-Student Türü Artıklar, Ortalama Değişim Aykırı Değer Modellemesi ve Bonferroni Testleri kullanılarak kontrol edilmiştir. Yapılan analizler sonucunda bazı gözlemler aykırı değer olarak tespit edilmiştir. Aykırı değerler veri setinden çıkarıldıktan sonra yukarıda anlatılan tüm yöntemler yeniden uygulanmış ve elde edilen sonuçların karşılaştırılmasına yer verilmiştir. Karşılaştırma kriteri olarak belirtme katsayısı, hata kareler ortalaması ve parametre tahminleri kullanılmıştır.

Bu tezin amacı, regresyon analiz sonuçları üzerine etkili olan aykırı değerlerin belirlenerek, aykırı değer varlığında ve veri setinden aykırı değerler çıkarıldıktan sonra E.K.K., Ridge Regresyon ve Robust Regresyon Yöntemlerinden M-Tahmin ediciden elde edilen analiz sonuçlarının karşılaştırmalı olarak incelenmesidir. Bu sayede kullanılan veri seti için regresyon modeli oluşturmada en uygun yöntem belirlenecektir. Diğer yandan aykırı değerlerin regresyon analiz sonuçları üzerinde hangi değerlere etki ettiği, hangileri üzerinde etkili olmadığı saptanacaktır. Ayrıca aykırı değerlerin veri setinden çıkarılması ve/veya çıkarılmaması için bazı önerilerde bulunulacaktır.

2. ÖNCEKİ ÇALIŞMALAR

Lawrence ve Marsh (1984), tarafından yapılan bir çalışmada U.S kömür madenciliği endüstrisinde ölümlerin tahmininde alternatif ridge ve robust tekniklerini kullanmışlardır. Veri setinde aykırı değerler ve çoklu bağlantı varlığında robust ridge yaklaşımına ihtiyaç olduğunu bildirmişlerdir.

Chattergee ve Hadi (1986), yaptıkları çalışmada regresyon analizinde aykırı değer ve etkili gözlemlerin incelenmesinin gerekliliğini vurgulamışlardır. Bir gözlem tahmin değerlerine, *b'* nın varyansına ya da uyum iyiliği istatistiklerine etki edebildiğini saptamışlardır. Bununla beraber bir regresyon denkleminde etkili olan unsurların değişkenler ve modelde ileri sürülen varsayımlar olduğunu bildirmişlerdir. Aynı çalışmada bir aykırı değerin etkili gözlem olmak zorunda olmadığını benzer şekilde etkili gözlemlerinde aykırı değer olmayabileceğini bildirmişlerdir.

Bek ve ark.(1996), tarafından yapılan çalışmada doğrusal regresyon modellerinde en küçük kareler yönteminin tahminlerinde veri noktalarının ne kadar etkisi olduğunu belirlemek için çeşitli istatistikler kullanılmıştır. Kullanılan bu istatistiklerin (Cook İstatistiği ve Welcsh-Kuh İstatistiği) parametre tahminleri ve varyanslar üzerine olan etkileri kullanılan veri setinde birlikte değişim varlığında ve yokluğunda araştırılmıştır. Birlikte değişim problemi olduğunda etkili olan gözlemlerin bu sorun çözüldükten sonra olan etkilerinin önceki kadar olmadığını bildirmişlerdir. Diğer yandan daha önce etkili olmayan bazı gözlemlerin aslında etkili olduğunu ve birlikte değişim probleminin bu etkiyi gizlediğini belirtmişlerdir. Sonuç olarak kullanılan bu yöntemlerin birbirleri ile karşılaştırılması yerine birlikte kullanılarak daha sağlıklı bir sonuca ulaşılabileceğini vurgulamışlardır.

Şahinler (1997), aykırı değerleri ve etkili gözlemleri belirlemek amacıyla yaptığı çalışmasında birlikte değişim problemi olan veri grubunu incelemiş ve bu problemin gerçekte aykırı olan gözlemleri gizleyebildiği ve daha başka gözlemleri aykırı değer gibi gösterip araştırıcıyı yanıltabildiğini belirtmiştir.

Ahn ve James (1999), Güney Florida'da atmosferde biriken fosfor miktarının ölçümünde çeşitli sebeplerden dolayı bir sorun olduğunu belirtmişlerdir. Bu sebeplerin neden olduğu hatalı ölçümleri belirlemek ve veri setinden uzaklaştırmak

için yaptıkları çalışmada Hadi ve Simonoff (1993)'un önerdiği aykırı değer belirleme yöntemini kullanmışlardır. Bu yönteme göre öncelikle n gözlemli veri setinden oluşturulan regresyon modeline göre hata değerleri hesaplanarak mutlak değerleri oluşturulur. Bu değerlerden en küçük değer sahip olan a tanesine sahip olan gözlemler ile oluşturulan alt küme oluşturulur. Burada ifade edilen a değeri bağımsız değişken sayısının bir fazlasına eşittir. Ardından oluşturulan regresyon modellerinden çeşitli test istatistikleri hesaplanır. Bu test istatiği değerleri göz önüne alınarak koşulları sağlayan tüm gözlemler aykırı değer olarak belirlenmiştir.

Şahinler (2000), regresyon modeli uydurulduktan sonra modelin yeterli olup olmadığının kontrolünün önemi üzerine yaptığı çalışmada modelin yeterliliğini belirlemek için kullanılan varyans analizi ve belirtme katsayısına ilave olarak çeşitli testleri önermiştir. Regresyon modeline girecek olan değişken seçimi, düzeltilmiş belirtme katsayısının değeri, hataların incelenmesi, varsayımların kontrolü, bağımsız değişkenler arasında birlikte değişim probleminin olup olmadığının kontrolü ve etkili gözlemlerin incelenmesinin regresyon modelinin oluşturulmasında önemli bir yere sahip olduğunu belirtmiştir.

Türkay (2004), tarafından yapılan çalışmada Türkiye ekonomisinde enflasyonun faiz, döviz kuru ve para arzı ile olan ilişkisini incelediği ekonometrik bir modelde parametre tahminlerini E.K.K ve M-tahmin edici yöntemini kullanarak elde etmiştir. M-tahmin edici yöntemini uygulayarak elde edilen parametre tahminleri ile aykırı değerlerin belirlenerek veri setinden çıkarıldıktan sonra E.K.K. yönteminden elde edilen tahmin değerlerinin yakın sonuçlar verdiğini saptamıştır. Sonuç olarak hiçbir gözlem değerinin veri setinden uzaklaştırılmadan parametre tahmininin yapıldığı M-tahmin edicinin E.K.K.' ye göre iyi bir alternatif olduğunu bildirmiştir.

Albayrak (2005), tarafından yapılan çalışmada En Küçük Kareler, Ridge ve Temel Bileşenler Regresyon analizlerini kullanarak elde edilen analiz sonuçları karşılaştırılmıştır. Kullanılan veri setinde çoklu bağlantı olması durumunda standart hataların yüksek olabileceğini ve yanlı tahmin tekniklerinin en küçük kareler tekniğine göre daha tutarlı, geçerli ve uygun tahminler sağladığını bildirmiştir.

Gündoğan (2005), aykırı değerlerin varlığında M tahmin yöntemlerinin kendi içinde tutarlı olduğunu bildirmiştir. Bunun yanında En Küçük Kareler yönteminden

elde edilen sonuçların aykırı değer varlığında farklılaştığını belirtmiştir. Aynı çalışmada tüm aykırı değerler veri setinden uzaklaştırıldıktan sonra EKK ve M tahmin yöntemleri arasında çok fazla bir farklılık gözlemlenmemiştir. Aykırı değerler eklendiğinde M tahminlerinin EKK sonuçlarına göre daha sağlam olduğunu saptamıştır. Ayrıca çoklu regresyonda aykırı değerlerin birbirini gizlemesi, örnek büyüklüğü yeteri kadar olmadığında veri çıkarma işleminin başka sorunlara sebep olabileceği ihtimali, veri çıkarma işlemi sonucunda elde edilecek yeni dağılım hakkında yeterli bilgiye sahip olunamaması gibi sorunlarla karşılaşılabileceği için veri çıkarma işleminin çok fazla önerilen bir yol olmadığını, bunun yerine aykırı değerlere karşı dirençli tahmin kriterleri kullanmanın gerekliliğini belirtmiştir.

Karadavut ve ark., (2005), Nohut (*Cicer arietinum* L.) bitkisinde verime etki eden bazı karakterleri En Küçük Kareler, Ridge Regresyon ve Robust regresyon yöntemlerinden olan M-Regresyon yöntemleri ile elde ettikleri parametre tahminlerini karşılaştırmışlardır. Nohut bitkisinde tane ağırlığına etki eden değişkenlere ilişkin regresyon modelinde öncelikle En Küçük Kareler yöntemini kullanarak parametreleri tahmin etmişlerdir. Bağımsız değişkenler arasında çoklu bağlantı tespit edildikten sonra Ridge Regresyon yöntemini kullanarak parametre tahminlerini gerçekleştirmişlerdir. Aynı verilere M-regresyon yöntemini de uygulamışlardır. Elde edilen parametre tahminleri ve analiz sonuçlarını karşılaştırmışlar ve M-regresyon yönteminin ilgili verilere ait parametre tahminlerinde diğer yöntemlere göre daha uygun bir tahmin edici olarak tercih edilebileceğini belirlemişlerdir.

Aşıkgil (2006), yapmış olduğu çalışmada EKK yöntemi ile parametre tahminlerini gerçekleştirmiştir. Ardından tekli ve çoklu aykırı değer, uç değer ve etkili gözlemleri (tekli ve çoklu kuşkulu gözlemler) belirleyen teknikleri çalışmıştır. Aynı veri setine robust regresyon yöntemlerini de uygulayarak elde ettiği sonuçları karşılaştırmıştır. Sonuçta bütün veri seti için tek kuşkulu ve çoklu kuşkulu gözlemlerin incelenmesi gerektiğini, tek kuşkulu gözlemlerin saptanmasında kullanılan bazı istatistikler için belirlenen kritik değerlerin kuşkulu gözlemleri saptamada uygun olmaması durumunda gözlemlerin kendi aralarında

karşılaştırılabileceğini, tekli ve çoklu kuşkulu gözlemlerin birbirini maskeleyebilme durumundan dolayı kesinlikle robust yöntemlerin kullanılması gerektiğini belirtmiştir.

Çankaya ve ark., (2006), yaptıkları çalışmada çoklu lineer regresyon modelinde parametre tahmin yöntemlerini karşılaştırmışlardır. En küçük kareler, nonparametrik yöntem ve robust regresyon yöntemlerini kullanarak yaptıkları parametre tahminlerinde aykırı değer varlığında en küçük medyan kareler yönteminin diğer yöntemlere göre en yüksek belirtme katsayısına sahip olduğunu belirtmişlerdir. Aykırı değerlerin ve yüksek uç değerlerin veri setinden uzaklaştırılmamasını etkilerinin araştırılmasını tavsiye etmişlerdir.

Ergül (2006), yapmış olduğu çalışmada çeşitli robust regresyon yöntemleri ile en küçük kareler yöntemlerini kullanarak parametre tahminlerini elde etmiştir. Robust regresyon yöntemlerinden LMS, LTS, M-Regresyon ve WLS tekniklerini çalışmıştır. Yapılan analizler sonucunda EKK yöntemine ait belirtme katsayısı değerinin diğer yöntemlerden daha düşük olduğunu belirlemiştir. Bu nedenle çoklu doğrusal regresyon modeli elde etmek için robust regresyon yöntemleri ve çok değişkenli robust regresyon tekniklerinde güvenilir bir şekilde çalışılabileceğini bildirmiştir.

Kontrimas ve Verikas (2006), aykırı değer belirleme yöntemlerini 4 grup altında toplamıştır. Buna göre birinci grupta veri merkezinden uzaklık üzerine çalışan yöntemler (temel bileşenler analizine dayalı teknikler), ikinci grupta tahmin ve gerçek değerler arası uzaklık üzerine çalışan yöntemler (hataların grafiksel analizi ve etkilerin ölçümü), üçüncü grupta robust regresyon yöntemleri, dördüncü grupta da sınıflama yöntemlerine dayalı teknikler yer almaktadır.

Martin ve Roberts (2006), yaptıkları çalışmada en küçük kareler student türü artıklarda aykırı değer belirlemede kullanılmak üzere kritik noktanın oluşumuna bootstrap (yeniden örnekleme) yaklaşımını önermişlerdir ve bu yaklaşımın belirlenen hataların normal dağılım göstermemesi gibi bir durumda kullanılmasını önermişlerdir.

Ortiz ve ark., (2006), kimyasal analiz verilerinde aykırı değerlerin belirlenmesi için robust regresyon tekniklerini önermişlerdir. Bu çalışmada Huber'in M tahmin edicisi, GM-tahmin edici, en küçük medyan kareler (LMS) yöntemlerini kullanmışlardır. Sonuç olarak En Küçük Medyan Kareler regresyon yönteminin

matematiksel özelliklerinin aykırı değer belirlemede daha etkili olabileceğini belirtmişlerdir.

Koç (2007), çalışmasında LMS (Least Median Square), LTS (Least Trimmed Square), S, M, GM ve EKK yöntemlerini kullanarak regresyon denkleminde parametre tahminleri yapmıştır. En iyi sonucun S tahmin edicisinde alındığını bildirmiştir. Regresyon analizi yapılmadan önce sapan değerleri bulmak için robust tahmin edicilerinden herhangi birinin kullanılmasını önermiştir.

Alma ve Vupa (2008), yaptıkları çalışmada en küçük kareler ve en küçük medyan kareler yöntemlerini karşılaştırmışlardır. Veri setinde birden fazla sapan değer bulunduğunda bu değerlerin birbirini maskeleyebildiği ve güvenilir verilerin sapan değer olarak görülmesine sebep olabildiğini belirtmişlerdir. Bu nedenle hata terimlerinin normal dağılmadığı veya bağımlı değişkenin sapan değer içermesi durumlarında küçük örneklemeler için regresyon modelinde, en küçük medyan kareler yönteminin en küçük kareler yöntemine göre daha az etkilendiğini bildirmişlerdir. En küçük medyan kareler parametre tahmin değerlerinin regresyon modelini daha iyi açıkladığını tespit etmişlerdir.

Kim ve ark. (2008), tarafından yapılan çalışmada mean-shift (ortalama değişim) aykırı değer modellemesi çalışılmıştır. Öncelikle veri setinde aykırı değerler belirlenmiştir. Ortalama değişim aykırı değer modeli yöntemi kullanılarak aykırı değer olduğu düşünülen gözlemlerle birlikte değişken seçimi analizlerini gerçekleştirmişlerdir. Bu amaçla aykırı değerler ve bağımsız değişkenlerden oluşan mümkün olan bütün alt kümelere regresyon analizi uygulanmıştır. Hiçbir gözlem ve değişken veri setinden uzaklaştırılmadan regresyon denklemine ait varyans analiz tablosu elde edilmiş ve belirtme katsayısı değerleri karşılaştırma kriteri olarak kullanılmıştır. Ardından stepwise yöntemi ile değişken eleme işlemi gerçekleştirilmiştir. En yüksek belirtme katsayısı ve en düşük hata kareler ortalamasının hangi kombinasyonda elde edildiği saptanmıştır. Böylelikle aykırı değer olan gözlemlerinde dahil edildiği ve en iyi model belirlenmiştir.

Billor ve Kıral (2008), yapmış oldukları çalışmada çok sayıdaki aykırı değer belirleme yöntemlerinden hangilerinin pratikte kullanılabilmesinin daha uygun olduğu konusunda ortak bir görüş olmadığını bildirmişlerdir. Bu amaçla aykırı değer

belirleme literatürleri içerisinde yer alan çoklu aykırı değer belirleme yöntemlerinin performansını belirlemek için karşılaştırmalı Monte Carlo Simülasyon çalışmasını gerçekleştirmişlerdir. Hangi durumda hangi yöntemin daha üstün olduğuna dair önerilerde bulunmuşlardır. Yaptıkları karşılaştırmada Hadi ve Sinonoff (1993) tarafından önerilen aykırı değer belirleme yöntemi ile yeniden ağırlıklandırılmış en küçük kareler yönteminin aykırı değerlerin belirlenmesinde hata riskinin düşük olduğunu buna karşın kümeleme algoritma tekniğinin aykırı değerlerin belirlenmesinde hata riskinin yüksek olduğunu saptamışlardır.

3. MATERYAL VE YÖNTEM

3.1. Materyal

Bu çalışmada materyal olarak kullanılan veriler, Ç.Ü. Ziraat Fakültesi Zootekni Bölümü Yemler ve Hayvan Besleme Anabilim Dalı'nda yapılan tercihli yemleme çalışmasından elde edilmiştir. Ham yağ (X_1) (g/gün), ham protein (X_2) (g/gün), asit deterjan fiber (X_3) (g/gün) ve nötr deterjan fiber (X_4) (g/gün) tüketim miktarları bağımsız değişken, yem tüketim miktarı (kg/gün) ise bağımlı değişken (Y) şeklinde alınarak

$$Y_i = b_0 + b_1 X_1 + b_2 X_2 + b_3 X_3 + b_4 X_4 + e_i$$

Regresyon modeli kullanılmıştır. Yorumlamada kolaylık sağlamak amacı ile veriler standardize edildikten sonra regresyon analizleri uygulanmıştır. Bağımlı ve bağımsız değişken olarak alınan verilerde gözlem sayısı 114 'tür. Çalışmada MINITAB, SPSS, R ve NCSS paket programları yardımı ile istatistiki analizler gerçekleştirilmiştir. Materyal olarak kullanılan ham verilerin bir kısmı ve bu verilerin standardize edilmiş halleri Ek-1 ve Ek-2'de verilmiştir.

3.2. Yöntem

3.2.1. En Küçük Kareler Yöntemi

Basit doğrusal regresyon modeli

$$y_i = b_0 + b_1 x_i + e_i \quad ; \quad (i=1,\ldots,k) \qquad (3.1)$$

şeklinde ifade edilir.

eşitlikte,

y_i : bağımlı değişkenin i. gözlem değerini,

x_i : bağımsız değişkenin i. gözlem değerini,

β_0 : regresyon doğrusunun Y eksenini kestiği noktanın orijine olan uzaklığını,

β_1 :regresyon katsayısı olup bağımsız değişkendeki bir birimlik değişime karşılık bağımlı değişkende kendi birimi cinsinden meydana gelen ortalama değişim miktarını,

ε_i: i. hata terimi olup $e_i \sim (0, s^2)$ şeklinde bir dağılış göstermektedir.

Buna göre gözlem noktalarını temsil edebilen öyle bir doğru çizmeliyiz ki, gözlem noktalarının doğruya olan uzaklıklarının kareler toplamı minimum olsun. Başka bir ifade ile modeldeki b_0 ve b_1 parametrelerini öyle bir tahmin etmeliyiz ki gözlem noktalarının tahmin edilen doğruya olan uzaklıklarının kareler toplamı en küçük olsun. En küçük kareler yöntemi bu durumu sağlayan yöntemlerden biridir. Böylece b_0 ve b_1 parametrelerini tahmin etmek için,

$$KT = \sum_{i=1}^{n} e_i^2, \text{ minimum} \qquad (3.2)$$

$$e_i = y_i - b_0 - b_1 x_i \qquad (3.3)$$

$$\sum e_i^2 = \sum (y_i - b_0 - b_1 x_i)^2 \qquad (3.4)$$

ifadesi minimum yapılır. olarak yazılabilir. Buna göre bu ifadeyi minimum yapan b_0 ve b_1 değerleri bu parametrelerin en küçük kareler tahminini verirler. İlgili eşitliği minimum yapan b_0 ve b_1 değerlerini bulmak için bu ifadenin b_0 ve b_1'e göre kısmi türevlerini alarak sıfıra eşitlemek gerekmektedir. Böylece;

$$\frac{\partial KT}{\partial b_0} = 0 \Rightarrow -2\sum (y_i - b_0 - b_1 x_i) = 0 \qquad (3.5)$$

$$\frac{\partial S}{\partial b_1} = 0 \Rightarrow -2\sum x_i (Y_i - b_0 - b_1 x_i) = 0 \qquad (3.6)$$

Böylece;

$$n_0 \hat{b}_0 + \hat{b}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \qquad (3.7)$$

$$\hat{b}_0 \sum_{i=1}^{n} x_i + \hat{b}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \qquad (3.8)$$

olarak gösterilmektedir. Bu denklem sisteminin çözümü ile;

$$\hat{b}_1 = \frac{\sum XY - \frac{(\sum X)(\sum Y)}{n}}{\sum X^2 - \frac{(\sum X)^2}{n}} = \frac{S_{XY}}{S_{XX}} \qquad (3.9)$$

$$\hat{b}_0 = \frac{\sum_{i=1}^{n} Y_i + \hat{b}_1 \sum_{i=1}^{n} X_i}{n} = \overline{Y} - \hat{b}_1 \overline{X} \quad \text{olur.} \qquad (3.10)$$

Böylece, hesaplanan $\hat{b}_0$ ve $\hat{b}_1$ değerleri eşitlik (3.1)' de yerine konulursa,

$$\hat{Y}_i = \hat{b}_0 + \hat{b}_1 X_i \tag{3.11}$$

elde edilmektedir. Eşitlik (3.1)' de yer alan e_i daha önce belirtildiği üzere, hata terimi olup ortalaması sıfır varyansı s^2 olan normal dağılış gösterdiği varsayılırsa, bu varyansı;

$$S_{yx}^2 = \frac{\sum (y_i - \hat{y})^2}{n-2} \tag{3.12}$$

formülü yardımı ile tahmin edilebileceğini göstermek mümkündür. Burada $\hat{y}$ 'nin değeri yerine yazılırsa;

$$S_{yx}^2 = \frac{\sum (y_i - \hat{b}_0 - \hat{b}_1 x_i)^2}{n-2} \tag{3.13}$$

eşitliği elde edilmektedir. Bu eşitlikte $\hat{b}_0 = \bar{y} - \hat{b}_1 \bar{x}$ değeri yerine konursa,

$$S_{yx}^2 = \frac{\sum (y_i - \bar{y} + \hat{b}_1 \bar{x} - \hat{b}_1 x_i)^2}{n-2} \tag{3.14}$$

$$= \frac{\sum \left\{(y_i - \bar{y}) - \hat{b}_1 (x_i - \bar{x})\right\}^2}{n-2} \tag{3.15}$$

$$= \frac{\sum \left\{(y_i - \bar{y})^2 - 2\hat{b}_1 (y_i - \bar{y})(x_i - \bar{x}) + \hat{b}_1^{\,2} (x_i - \bar{x})^2\right\}}{n-2} \tag{3.16}$$

$$S_{yx}^2 = \frac{\sum (y_i - \bar{y})^2 - 2\hat{b}_1 \sum (y_i - \bar{y})(x_i - \bar{x}) + \hat{b}_1^{\,2} \sum (x_i - \bar{x})^2}{n-2} \tag{3.17}$$

olur. Eşitlik (3.17)' de

$$\hat{b}_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \qquad (3.18)$$

değeri yerine konursa;

$$S_{yx}^2 = \frac{\sum (y_i - \bar{y})^2 - \hat{b}_1 \sum (y_i - \bar{y})(x_i - \bar{x})}{n-2} \qquad (3.19)$$

$$S_{yx}^2 = \frac{S_{yy} - \dfrac{(S_{xy})^2}{S_{xx}}}{n-2} \qquad (3.20)$$

olarak yazılabilir. Buna göre hesaplanan varyans değerinin karekökü alınırsa tahmin denkleminin standart hatası hesaplanmış olur.

Çoklu regresyon analizi yönteminde, bir bağımlı değişkeni etkileyen birden çok bağımsız değişken vardır. Böylece çoklu regresyon modeli aşağıdaki gibi gösterilmektedir.

$$Y_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + b_k X_{ik} + e_i \quad ; \quad (i = 1,\ldots,n) \quad (j = 1,\ldots,k) \qquad (3.21)$$

Eşitlikte,

Y_i : bağımlı değişkenin i. gözlem değerini,

$X_{1i},\ldots,X_{ki}$: bağımsız değişkenlerin i. gözlem değerini,

β_0 : regresyon doğrusunun Y eksenini kestiği noktanın orijine olan uzaklığını

$\beta_1,\ldots,\beta_k$: regresyon katsayısı olup bağımsız değişkendeki bir birimlik değişime karşılık bağımlı değişkende kendi birimi cinsinden meydana gelen ortalama değişim miktarlarını,

ε_i: i. hata terimi olup $e_i \sim (0, s^2)$

k : bağımsız değişken sayısını,

n : gözlem sayısını ifade etmektedir.

Bu modelde 2 bağımsız değişken olduğunda en küçük kareler yöntemi uygulanarak her bir değişken için n tane gözlem değerinin bulunduğunu varsayarak aşağıdaki denklemleri elde etmek mümkün olmaktadır.

$$\hat{b}_0 n + \hat{b}_1 \sum X_1 + \hat{b}_2 \sum X_2 = \sum Y \qquad (3.22)$$

$$\hat{b}_0 \sum X_1 + \hat{b}_1 \sum X_1^2 + \hat{b}_2 \sum X_1 X_2 = \sum X_1 Y \qquad (3.23)$$

$$\hat{b}_0 \sum X_2 + \hat{b}_1 \sum X_1 X_2 + \hat{b}_2 \sum X_2^2 = \sum X_2 Y \qquad (3.24)$$

Bu denklemleri matris notasyonuna göre yazarsak,

$$\begin{pmatrix} n & \sum X_1 & \sum X_2 \\ \sum X_1 & \sum X_1^2 & \sum X_1 X_2 \\ \sum X_2 & \sum X_1 X_2 & \sum X_2^2 \end{pmatrix} \begin{pmatrix} \hat{b}_0 \\ \hat{b}_1 \\ \hat{b}_2 \end{pmatrix} = \begin{pmatrix} \sum Y \\ \sum X_1 Y \\ \sum X_2 Y \end{pmatrix}$$

A B Y

şeklinde gösterilmektedir. Buradaki A matrisine katsayılar matrisi, B matrisine tahmin matrisi ve Y matrisine ise çarpımlar toplamı matrisi adı verilmektedir. A matrisinin tersi alınarak eşitiğin her iki tarafı bu matris ile çarpılırsa ,

$$B=A^{-1} Y$$

olmaktadır.

Matris notasyonu parametre tahminleri eşitlik (3.25) yardımı ile de hesaplanabilmektedir. Matris ve vektör gösterimleri aşağıda belirtilmiştir.

$$Y=\begin{pmatrix} y_1 \\ y_2 \\ . \\ . \\ . \\ y_n \end{pmatrix} \quad X=\begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ 1 & x_{21} & \cdots & x_{2k} \\ . & . & \cdots & \\ . & . & \cdots & \\ . & . & \cdots & \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix} \quad \beta=\begin{pmatrix} b_0 \\ b_1 \\ . \\ . \\ . \\ b_k \end{pmatrix} \quad e=\begin{pmatrix} e_1 \\ e_2 \\ . \\ . \\ . \\ e_n \end{pmatrix}$$

$$\hat{b} = (X'X)^{-1}X'Y \qquad (3.25)$$

Çoklu regresyon modelinde parametrelerin tahmini için kullanılan X matrisi ve Y vektörünün genel görünümü çizelge (3.1)'de gösterilmiştir.

Çizelge 3.1. Çoklu Regresyon Modelinde Verilerin Gösterimi

	Bir bağımlı değişken ve bağımsız değişkenler						Hata Terimi
Gözlem No	Y	1	X_1	X_2	...	X_k	ε
1	Y_1	1	X_{11}	X_{12}	...	X_{1k}	ε_1
2	Y_2	1	X_{21}	X_{22}	...	X_{2k}	ε_2
3	Y_3	1	X_{31}	X_{32}	...	X_{3k}	ε_3
.	.	.	.	.	…	.	.
.	.	.	.	.	…	.	.
.	.	.	.	.	…	.	.
n	Y_n	1	X_{n1}	X_{n2}	.....	X_{nk}	ε_n

Bu çizelge matris notasyonuna dönüştürüldüğünde regresyon modeli eşitlik (3.11)'te gösterildiği gibi ifade edilebilir.

$$Y = Xb + e \tag{3.26}$$

Burada;

Y: nx1 boyutlu bağımlı değişken vektörünü,

X: nx(k+1) boyutlu bağımsız değişken matrisini,

b : (k+1)x1 boyutlu parametre vektörünü ifade etmektedir.

e : hata vektörü olup $e_i \sim (0, s^2)$ 'dir.

3.2.1.1. Çoklu Doğrusal Regresyon Modelinde Parametre Tahminlerinin Uygunluğu

Regresyon analizinde parametre tahminleri yardımı ile regresyon denklemleri elde edilmektedir. Regresyon denklemlerinde değişkenler arasındaki ilişkiyi gösteren parametre tahminlerinin istatistiki olarak önemli olup olmadığına karar vermek gerekmektedir. Bu tahminlerde bu duruma karar verebilmek için F ve t testi gibi testler yapılmaktadır. Bu testler haricinde belirtme katsayısı da kriter olarak kullanılmaktadır.

F testinde bütün bağımsız değişkenlerin bağımlı değişken üzerindeki etkisini test edebilmek için oluşturulan alternatif hipotezde parametrelerin bütünüyle ele alındığında formüle edilen ilişkinin yani elde edilen regresyon denkleminin istatistiki olarak önemli olduğu şeklinde açıklanabilir. Buna göre F değeri eşitlik (3.27) yardımıyla hesaplanmaktadır.

$$F = \frac{R^2/(k+1)}{(1-R^2)/(k+1)} \tag{3.27}$$

Eşitlikte;

R^2: belirtme katsayısını,

k : bağımsız değişken sayısını,

n : gözlem sayısını ifade etmektedir.

t testi ise incelenen olaydaki bağımlı değişken ile bağımsız değişkenler arasındaki ilişkiyi gösteren $\hat{b}$ parametrelerinin test edilmesini sağlamaktadır. Tek bir parametreyi test etmek için gerekli t istatistiği (3.28) eşitliğindeki gibi hesaplanmaktadır.

$$t = \frac{\hat{b}_k - b_k}{s_{\hat{b}}} \qquad (3.28)$$

Eşitlikte;

$\hat{b}_k$: test edilen parametrenin değerini,

$s_{\hat{b}}$: ilgili parametrenin standart hatasını ifade etmektedir.

n-k-1 serbestlik derecesine göre t tablosundan bulunan değer, t istatistiğinden küçükse ilgili katsayının istatistiki olarak anlamlı olduğu sonucuna ulaşılmaktadır (İmir, 1986).

3.2.1.2. Çoklu Doğrusal Regresyon Modelinin Varsayımları

1. Hata Teriminin Normal Dağılış Göstermesi

Normal dağılış varsayımı parametrik testlerin yapılabilmesi için oldukça önemlidir. Kolmogorov-Smirnov, Shapiro-Wilks istatistiği gibi çeşitli testler yardımı ile hataların normal dağılışa sahip olup olmadıkları belirlenebilir. Hataların normal dağılışa sahip olmadığı belirlenirse en küçük kareler yöntemi uygulanamaz (Şahinler, 2000).

$Y = X\beta + e$ eşitliğinden

$e = Y - E(Y)$ olduğundan

$$E(e) = E(Y) - E(Y) = 0 \tag{3.29}$$

elde edilir.

$$E(Y) = Y - e = X\beta \tag{3.30}$$

$$Var(e) = E(x)^2 - [E(x)]^2$$

$$Var(e) = E[e - E(e)]\,E[e - E(e)]' = E(ee') = \sigma^2 I$$

Böylece hataların varyans-kovaryans matrisi eşitlik (3.31)'da görüldüğü gibidir.

$$Var(e) = \sigma^2 I \tag{3.31}$$

2. Hata Terimlerinin Birbirinden Bağımsız Olması

ε'ler arasında ilişki (otokorelasyon) olmaması, $Kov(e_i, e_j) = 0,\ i \neq j$ şeklinde ifade edilmektedir.

3. Hata Terimlerinin Varyanslarının Sabit Olması

Bütün ε'lerin varyansı sabittir. Bu varsayım $(Var(e) = \sigma^2 I)$ şeklinde ifade edilmektedir.

4. Bağımsız Değişkenler Arasında Bir İlişki Olmaması

Bağımsız değişkenler arasında bir ilişkinin olması birlikte değişim problemini ortaya çıkarmaktadır. ε'nin ortalamasının 0 olması, otokorelasyon olmaması ve eşit varyanslı olması varsayımları sağlandığında, $\hat{b}$ ve HKO istenilen bazı özelliklere sahip olurlar. Bu özellikler '' Gauss- Markov Teoremi'' olarak bilinmektedir. Bu teoreme göre:

$\hat{b}$, b 'nın yansız bir tahmin edicisidir.

B vektörünün EKK tahmin edicisi $\hat{b} = (X'X)^{-1}X'Y$ idi. X sabit kabul edildiğinden $\hat{b}$,Y'nin doğrusal tahmin edicisidir.

Burada ;

$$\hat{b} = (X'X)^{-1}X'(Xb + e)$$

$$= (X'X)^{-1}X'Xb + (X'X)^{-1}X'e$$

$$= b + (X'X)^{-1}X'e \qquad (3.32)$$

ilişkisi bulunur.

β vektörü ve $(X'X)^{-1}X'$ matrisindeki elemanlar sabit değer olduğundan $\hat{b}$ vektörü e vektörünün doğrusal bir fonksiyonudur. $\hat{b}$ 'nın beklenen değeri:

$$E(\hat{b}) = E(b) + E\left[(X'X)^{-1}X'e\right]$$

$$= b + (X'X)^{-1}X'E(e)$$

$$= b + (X'X)^{-1}X'0$$

$$= b \qquad (3.33)$$

Böylece $\hat{b}$, b 'nın yansız bir tahmin edicisi olur.

Hata terimlerinin varyanslarının sabit olmaması tahminlerin standart hatalarının büyük olmasına yol açmaktadır. Böyle bir durumda kullanılan matematiksel model değiştirilir veya modele alınmayan değişkenlerden bazıları modele dahil edilir veya gözlem sayısı arttırılır (Ergüneş, 2004).

3.2.1.3. Çoklu Bağlantı

Çoklu regresyon modeli ile ilgili varsayımlardan sapmaların bir tanesi de bağımsız değişkenler arasındaki çoklu bağlantıdır. Çoklu bağıntının varlığında değişkenler arasındaki ilişkiyi belirleyen parametre tahminlerinin standart hataları büyük olur, bu da gerçek ilişki katsayısının yönü ve değeri açısından önemli derecede farklılığa yol açar. Neden olduğu sorunların azaltılarak olabildiğince sağlıklı karar verebilmek için çoklu bağıntının giderilmesi gerekmektedir.

3.2.1.4. Çoklu Bağlantının Etkileri

Çoklu bağlantıdan en belirgin şekilde etkilenen parametre tahminlerinin varyanslarıdır. E.K.K. tahmin edicisi $\hat{b}$ 'nın kovaryans matrisi aşağıda belirtilmiştir.

$$Kov(\hat{b}) = s^2 (X'X)^{-1} \qquad (3.34)$$

$(X'X)^{-1}$ j. köşegen elemanı c_{jj} olmak üzere,

$$Var(\hat{b}_j) = s^2 c_{jj}$$

$$= s^2 (1 - R_j^2)^{-1} \qquad (3.35)$$

Eşitlikte;

R_j^2 : Bağımsız değişkenler arasında her bir bağımsız değişken bağımlı değişken olarak alındığında elde edilen regresyon denklemine ait belirtme katsayısını,

s^2 : Parametre tahminlerinin varyansını ifade etmektedir.

şeklinde ifade edilmektedir. Bağımsız değişkenler arasındaki belirtme katsayısı olan $R_j^2 = 0$ ise $Var(\hat{b}_j) = s^2$ olacaktır. R_j^2' nin sıfırdan farklı olduğu durumlarda ise $Var(\hat{b}_j) > s^2$' dir.

Yüksek derecede olan çoklu bağlantı parametreler üzerinde kurulan hipotez testlerini de olumsuz yönde etkileyebilmektedir.

$$t = \frac{\hat{b}_j}{\sqrt{s^2 c_{jj}}} = \hat{b}_j \left[\sqrt{\frac{1 - R_j^2}{s^2}} \right] \qquad (3.36)$$

olmak üzere, R_j^2 bire yaklaştığında t_j sıfıra yaklaşmakta ve b_j'nin anlamlılığını belirleme olanağı azalmaktadır. Test istatistiği değerinin sıfıra yaklaşması, modele alınan bağımsız değişkenler gerçekte bağımlı değişkeni etkilemesine karşın, sıfır hipotezinin kabul edilmesine yol açarak b_j nin anlamlılığı konusunda olumsuz karar verilmesine neden olmaktadır.

Çoklu bağlantı $\hat{Y}$'ların tahmininin tutarlılığını azaltmaktadır. Yönü ve büyüklüğü bakımından regresyon katsayılarının gerçek katsayılardan çok farklı olması $\hat{Y}$'ları da etkilediğinden, $\hat{Y}$ tahminlerinin standart hataları büyük olmaktadır.

3.2.1.5 Çoklu Bağlantıyı Belirleme Yöntemleri

1. Bağımsız Değişkenler Arasındaki Korelasyon Katsayılarının Belirlenmesi

Çoklu bağlantının bağımsız değişkenler arasındaki ilişkiden dolayı ortaya çıktığı bilinmektedir. Bu ilişkinin derecesine bakılarak bağlantının derecesi hakkında karar verilebilir. Herbir bağımsız değişkenin diğer bağımsız değişkenlerle arasındaki korelasyon katsayılarının karesi olan değer 1'e yakınsa, yüksek derecede çoklu bağlantı olduğu ortaya çıkar (İmir, 1986).

2. Korelasyon Matrisi İle Belirlenmesi

Albayrak (2005)' in bildirdiğine göre Neter ve ark.,. (1996); Gujarati (1995) yüzeysel olarak, iki bağımsız değişken arasındaki basit korelasyon katsayısı oldukça anlamlı (r>%75) ise, bu durum çoklu doğrusal bağlantı problemine yol açabildiğini saptamıştır. Buna rağmen, istatistik açıdan anlamlı korelasyonlar her zaman çoklu doğrusal bağlantı problemine yol açmamaktadır. Benzer şekilde korelasyon katsayısının mutlak değeri 0.8'den büyük ise çoklu doğrusal bağlantı olduğu çeşitli çalışmalarda bildirilmiştir.

3. Çoklu Bağıntının Varyans Büyütme Faktörü İle Belirlenmesi

Korelasyon matrisinin matrisinin j. köşegen elemanı c_{jj} j. bağımsız değişkene ait varyans büyütme faktörünü verir. Bu ölçü Hoerl ve Kennard' a göre ikiden fazla ilişkinin belirlenmesinde en iyi ölçüdür . Bazı araştırıcılara göre bu değer 5'in üzerinde olduğunda çoklu bağlantıdan söz edilebilir (İmir, 1986).

4. Özdeğer Ve Özvektörlerin İncelenmesi

$(X'X)$'in özdeğerleri $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \ldots \geq \lambda_j > 0$ ve karşılık gelen birim dik özvektörleri $V_1, V_2, \ldots, V_j$ olmak üzere

$$\lambda_j = V'_j X\, XV_j = (XV_j)'(XV_j), \qquad j = 1, 2, \ldots, k \qquad (3.37)$$

dır. Küçük özdeğerler ve karşılık gelen özvektörler iç ilişkileri belirlemektedir. ($X'X$)'in son r tane özdeğerleri yeterince küçük ise,

$$0 \approx (XV_j)'(XV_j) \Rightarrow XV_j \approx 0 \qquad (3.38)$$

olacaktır. Çoklu bağlantının belirlenmesi amacıyla ($X'X$) korelasyon matrisinin özdeğerleri incelendiğinde,

$$\frac{mak \mid l_j \mid}{\min \mid l_j \mid} \qquad (3.39)$$

oranı 10'dan küçükse bağımsız değişkenler arasında çok az bir ilişki vardır. Bu oranın 30'dan büyük olması ise kuvvetlı bir ilişkinin varlığını belirler (İmir,1986).

3.2.2. Ridge Regresyon Yöntemi

Ridge regresyon tahmin edicisi hakkında Hoerl ve Kennard başta olmak üzere 1970'den bugüne kadar yüzlerce çalışma yapılmıştır. Ridge regresyon yönteminde en küçük kareler yönteminde izlenen aşamalar birden fazla tekrarlanmaktadır. Ridge yönteminin en küçük karelerden farklılığı k^* ridge parametresinin varlığıdır. 0 ile 1 arasında değer olan her k^* için hesaplanan parametre kestirimleri arasında, aranan kriterlere sahip olanları belirlenir. Albayrak (2005), Ridge regresyon analizinde korelasyon matrisinin köşegen değerlerine küçük bir yanlılık sabiti eklenerek, yanlı standartlaştırılmış regresyon katsayıları aşağıdaki gibi hesaplandığını bildirmektedir.

$$\hat{b}_R = (X'X + kI)^{-1} X'Y \qquad (3.40)$$

$W = (X'X + kI)^{-1}$ olmak üzere

$$\hat{b}_R = WX'Y \tag{3.41}$$

E.K.K. tahmin edicisi $\hat{b} = (X'X)^{-1}X'Y$ olarak ifade edilmiştir. Burada X'Y=(X'X)$\hat{b}$ olarak yazıldığında $\hat{b}$ ile $\hat{b}_R$ arasındaki ilişki:

$$\hat{b}_R = \left[I_k + k*(X'X)^{-1}\right]^{-1}\hat{b} \tag{3.42}$$

$$= Z\hat{b} \tag{3.43}$$

$$Z = (I_k + k*(X'X)^{-1})^{-1} \tag{3.44}$$

olmaktadır.

Ergüneş (2004)' in bildirdiğine göre Hoerl ve Kennard (1970), $\hat{b}_R$, Z, W'nin bazı özelliklerini şu şekilde sıralamışlardır.

* $x_j(W)$ ve $x_j(Z)$, W, Z'nin özdeğerleri, l_j ise (X'X)'in özdeğerleri olmak üzere

$$x_j(W) = (1/\, l_j + k^*) \tag{3.45}$$

$$x_j(Z) = l_j(l_j + k^*) \tag{3.46}$$

$$Z = I - k*(X'X + k*I)^{-1} = I - k*W \tag{3.47}$$

Her iki tarafı W^{-1} ile soldan çarpıldığında $Z = (X'X + k*I)^{-1}X'X = WX'X$ olur.

$k^* \neq 0$ için $\hat{b}_R$, $\hat{b}$ 'dan kısadır. $\hat{b}_R = Z\hat{b}$ tanımı ile

$$(\hat{b}_R)(\hat{b}_R) < (\hat{b})(\hat{b}) \text{ dır.} \qquad (3.48)$$

3.2.2.1. Ridge Regresyon Yönteminin Kullanım Amaçları

Ridge regresyon yönteminin kullanım yerleri şöyle sıralanabilir.

* Çoklu regresyon modelinde bağımsız değişkenler birbirleri ile bağlantılı olduklarında E.K.K. β tahmin edicisinden daha küçük varyanslı β tahmin edicilerinin elde edilmesinde,

* Güçlü çoklu bağlantı etkisi ile regresyon katsayılarında oluşan kararsızlıkların grafik üzerinde gösterilmesinde,

* Modeldeki gereksiz değişkenlerin çıkarılmasında (İmir, 1986).

3.2.2.2. Ridge Tahmin Edicisinin Yanlı Olması

$\hat{b}(k^*)$ tahmin edicisi yanlıdır. k*=0 veya Zk*=I olduğunda,

$$E[\hat{b}(k^*)] = b \qquad (3.49)$$

olur ki o zaman en küçük kareler tahmin edicisi olan b 'yı verir.

3.2.2.3 Ridge Tahmin Edicisinin Hata Kareler Ortalaması, Varyansı ve Hata Kareler Toplamı

HKO bakımından $\hat{\beta}_R$'nin özelliklerini açıklayabilmek için $E[L_1^2(k^*)]$' ya bakılması gerekir.

$$E[L_1^2(k^*)] = \mathrm{E}\left[(\hat{\beta}_R - \beta)'(\hat{\beta}_R - \beta) \right] \qquad (3.50)$$

$$= E\left[\left(\hat{\beta}_R - \beta\right)' Z' Z \left(\hat{\beta}_R - \beta\right) \right] + (Z\beta - \beta)'(Z\beta - \beta)$$

$$= \sigma^2 \text{İz}(X'X)^{-1} Z'Z + \beta'(Z - I)'(Z - I)\beta$$

$$= \sigma^2 [\text{İz}(X'X + k^* I)^{-1} - k^* \text{İz}(X'X + k^* I)^{-2}] + k^{*2} \beta' (X'X + k^* I)^{-2} \beta$$

$$= \sigma^2 \sum_{j=1}^{k} \lambda_j / (\lambda_j + k^*)^2 + k^{*2} \beta' (X'X + k^* I)^{-2} \beta$$

$$= \gamma_1(k^*) + \gamma_2(k^*)$$

$\gamma_1(k^*)$, Zb'dan b'ya olan uzaklığın karesidir. Bu ifade k^*= 0 olduğu zaman sıfırdır. Çünkü $Z=I$ olur. Böylece, $\gamma_2(k^*)\beta$ yerine $\hat{\beta}_R$ kullanıldığında yanlılığın karesi olarak ele alınabilir. Birinci terim $\gamma_1(k^*)$ parametre tahminlerinin varyanslarının toplamı, yani toplam varyansı göstermektedir. Şans değişkeni Y'nin terimlerinde

$$\hat{\beta}_R = Z\hat{\beta} = Z(X'X)^{-1}X'Y \qquad (3.51)$$

O halde;

$$Var(\hat{\beta}_R) = Z(X'X)^{-1}X'Var[Y]X(X'X)^{-1}Z'$$

$$= \sigma^2 Z(X'X)^{-1}Z' \qquad (3.52)$$

dir. Bütün $\beta_{R_{(j)}}$'lerin varyanslarının toplamı formül (3.41) 'in köşegen elemanları toplamıdır.

$\hat{\beta}_R$'ler için HKT ise aşağıdaki formülle bulunur.

$$\phi(\hat{\beta}_R) = (Y - X\hat{\beta}_R)'(Y - \hat{\beta}_R)$$

$$= Y'Y - (\hat{\beta}_R)'X'Y - k^*(\hat{\beta}_R)'(\hat{\beta}) \qquad (3.53)$$

3.2.2.4 Ridge Parametresinin (k*) Saptanması

İstatistikçilerden bazıları k^*'ın tek bir değer olmadığını, ancak $\hat{b}_{EK}$'dan daha iyi olan $\hat{b}_{RR}$'nin her zaman bulunabileceğini belirtmişlerdir. Ridge parametresi k^*'ın saptanması, çoklu bağıntı varlığındatahminlerin küçük varyansa sahip olması ve yanlılığı bakımından önemlidir. σ^2'nin küçük olduğu çoklu bağlantılı verilerde k^*'nın 0.198-0.272 arasında değişebileceği ileri sürülmektedir.

[0-1] aralığında değer alan k^*'ın belirlenmesine ilişkin verilen önerilerden bazıları şunlardır:

i) k^*'ın [0-1] aralığındaki değerlerine karşı bu değerlerden bulunan β_j'lerin tek tek çiziminden oluşan ridge izinden yararlanılır. Her β_j için çizilen eğrilerin yatay eksene paralel olmaya başladıkları k^* değeri, ilgili olaya ait ridge regresyon modeli için ridge parametresi olarak belirlenir.

ii) k^*'nın belirlenmesinde kullanılan diğer bir faktör varyans büyütme faktörüdür. Varyans büyütme faktörünün 1 ile 10 arasındaki değerlerine karşılık gelen k^*'ın belirlenmesi önerilmektedir. Fakat genellikle uygulamada 7 ve civarındaki değerlere karşılık gelen k^* değerleri kullanılmaktadır.

iii) Çoklu bağıntı varlığında kullanılan ridge regresyon yönteminde k^*'ın aşağıdaki formül yardımı ile belirlenebilmesi de mümkündür

$$k^* = \frac{s^2}{\hat{b}'\hat{b}} \qquad (3.54)$$

iv) Yukarıda gösterilen k^* parametresine eşdeğer olan

$$k^* = \frac{ks^2}{\sum_{j=1}^{k} s_j^2} \tag{3.55}$$

şeklindeki k^*değeri, ridge parametresi olarak alınabilir (İmir,1986).

3.2.3. M-Tahmin Edici

Robust istatistiğin asıl amacı: hata terimlerinin normal dağılışa sahip olmaması ve/veya aykırı değerlerin bulunması durumunda güvenilebilir sonuçlar verebilmektir. Normallik varsayımı hipotez testleri ve güven aralıklarının oluşturulması için gerekli olup β katsayılarının tahmini için gerekli değildir. Ancak hatalar normal olmayan dağılımlardan geliyorlarsa en küçük kareler tahmin edicilerinden elde edilen sonuçlar güvenilir olmayacaktır. Regresyon analizinde en çok kullanılan yöntem olan en küçük kareler tahmin edicileri $e_i = Y_i - X\beta$ olmak üzere hata kareler toplamını minimum yapma düşüncesi ile çalışır. Verilerde aykırı değerler varsa bu durumda e_i artacak ve kareler toplamını minimum yapmaya çalışan en küçük kareler yöntemi çeşitli varsayımlar altındaki güvenilirliğinden uzaklaşacaktır (Coşkuntuncel,2005).

En küçük mutlak sapma (Least Absolute Deviation) (LAD) tahmin edicisi hataların mutlak değerleri toplamını en küçük yapma düşüncesi ile çalışmaktadır. M tahmin edicide bu fikir genelleştirilmiş ve parametrelerin tahmini $\sum p(\hat{e}_i)$ ifadesi en küçük olacak şekilde yapılmaktadır. Burada p(e) e'nin bir fonksiyonudur. Huber'in M tahmin edicisi hataların kareleri ile mutlak değerleri arasında uygunluk sağlayan bir fonksiyon kullanmaktadır. LAD tahmin edicilerinin En küçük kareler tahmin edicilerine göre avantajı aykırı değerlere karşı çok hassas olmamasıdır. Huber bu iki yönteminde avantajlarını birleştiren "e" sıfıra yakınsa e^2 ve sıfırdan uzaksa |e| olmak üzere bir fonksiyon tanımlamıştır.

$$p(e)=\begin{cases} e^2, -k \le e \le k \\ 2k|e|-k^2, e<-k, k<e \end{cases} \quad (3.56)$$

şeklinde p(e) fonksiyonunu tanımlamıştır. Huber k değerini $k=1.5\hat{s}$ olarak önermiştir. $\hat{s}$ ifadesi populasyonun standart sapmasının tahminidir. 2k|e|-k^2 ifadesi |e| yerine kullanılmaktadır. $\hat{s}$ değeri $\hat{s}=1.483MAD$ ve MAD ise $|\hat{e}_i|$ mutlak sapmalarının medyanı şeklinde tanımlanmaktadır. Parametrelerin tahmin edilmesi için $\sum p(\hat{e}_i)$ ifadesinin minimize edilmesi gerekmektedir. M tahmin edicisinde kullanılan algoritma, parametre tahminlerinin en küçük kareler yöntemi kullanılarak elde edilmesi ile başlamaktadır. Elde edilen regresyon denkleminden hesaplanan hataların mutlak değeri alınarak mutlak sapmalarının medyanı belirlenmektedir. Bu sonuçlar populasyonun standart sapmasının tahmininde ($\hat{s}^0$) kullanılmaktadır. Hesaplanan hata değerleri (e_i^0), $-1.5\hat{s}^0$ ve $1.5\hat{s}^0$ arasında ise $e_i^*=e_i^0$, eğer $-1.5\hat{s}^0$'dan küçükse $e_i^*=-1.5\hat{s}^0$ ve eğer $1.5\hat{s}^0$'dan büyükse $e_i^*=1.5\hat{s}^0$ olacak şekilde tekrar yazılır. Yeni oluşturulan hata değerleri e^* şeklinde gösterilmektedir. Elde edilen e* değerleri kullanılarak düzeltilmiş Y değerleri oluşturulmaktadır. Daha sonra b parametre vektörünün M tahminleri Y yerine Y* kullanılarak EKK yöntemi ile elde edilmektedir . Bu algoritmada bir sonraki tahminlerle önceki tahminler aynı oluncaya kadar iterasyona devam edilmektedir (Birkes ve Dodge, 1993; Karadavut ve ark., 2005).

3.2.4. Aykırı Değerlerin Belirlenmesinde Kullanılan Başlıca Testler

3.2.4.1. Standartlaştırılmış Hatalar

Aykırı değerlerin belirlenmesinde kullanılan bu hatalar aşağıdaki eşitlik (3.56) yardımı ile hesaplanmaktadır ve [-2,+2] aralığında bulunmaktadır.

$$d_i=\frac{e_i}{\sqrt{HKO}} \quad (3.57)$$

3.2.4.2. Student Türü Hatalar

İçsel hata olarak da adlandırılan bu hatalar eşitlik 3.57 yardımıyla hesaplanmaktadır.

$$r_i = \frac{e_i}{\sqrt{HKO(1-h_{ii})}} \tag{3.58}$$

$h_{ii} = X_i^{'}(X'X)^{-1}X_i$ şeklinde hesaplanır.

3.2.4.3. R-Student Türü Hatalar

t dağılışı gösteren bu hatalar eşitlik 3.58 yardımı ile hesaplanmaktadır.

$$t_i = \frac{e_i}{\sqrt{s_{(i)}^2(1-h_{ii})}} \tag{3.59}$$

$$s_{(i)}^2 = \frac{(n-k-1)HKO - (e_i^2/(1-h_{ii}))}{n-k-2} \tag{3.60}$$

şeklinde hesaplanmaktadır (Aşıkgil,2006).

3.2.4.4. Ortalama Değişim (Mean-Shift) Aykırı Değer Modellemesi

v 'inci gözlemin aykırı değer olup olmadığının araştırılmasında U ile gösterilen yeni bir bağımsız değişken tanımlanmıştır. U 'nun i' inci elemanı $i \neq v$ iken $u_1 = 0$ ve v' inci elemanı $u_v = 1$ olarak belirtilir. X 'in v' inci satırı çıkartıldığında,

$$Y_i = bX_i + e_{i,} \qquad i \neq v$$

denkleminin tahmini ve tüm gözlemler için X ve U üzerinden

$$Y_v = bX_v + d + e_{v,} \qquad i = v$$

denkleminin tahmini elde edilebilir. Burada d , U' nun katsayısı olup ortalama değişim olarak adlandırılmaktadır. v' inci gözlemin aykırı değer olup olmadığına karar verebilmek için

$H_0 : d = 0$ sıfır hipotezinin

$H_1 : d \neq 0$ alternatif hipotezine karşı test edilmesi gerekir.

Eğer hata terimleri normal dağılıma sahipse, test istatistiği $n-k-1$ serbestlik derecesi ile $t-$dağılımıdır (Weisberg, 2005). Buna göre aykırı değer olduğu düşünülen gözlemler için bağımsız değişkenler oluşturulmaktadır. Örneğin, 100 gözlemden oluşan bir veri grubunda 52 nolu gözlemin aykırı değer olduğu düşünülürse, 52 nolu gözlemin denk geldiği hücreye bir, diğer gözlemlerin karşışına sıfır yazılarak yeni bir bağımsız değişken elde edilerek tekrar regresyon analizi uygulanır. Analiz sonuçlarına göre yeni oluşturulan bağımsız değişken istatistiki olarak önemli ise ilgili gözlemin aykırı değer olduğu ifade edilebilmektedir.

3.2.4.5. Bonferroni Testi

Aykırı değerleri belirlemede r-student türü hataların mutlak değerleri göz önünde bulundurulur. Hangi gözlemin $|t_i|$ değerinin büyük olduğu bilinmediğinden Bonferroni testi uygulanarak t_i değerleri hesaplanmaktadır. Hesaplanan bu değerlerden Bonferroni kritik değeri $t_{n-k-2;(a/2n)}$ den büyük olan gözlemler aykırı değer olarak tahmin edilmektedir. t_i değeri aşağıdaki eşitlik (3.61)' den hesaplanmaktadır.

$$t_i = \frac{\hat{e}_i}{\hat{s}_{(i)}\sqrt{1-h_{ii}}} \qquad (3.61)$$

Eşitlikte;

$\hat{e}_i$: Hata teriminin tahmin edicisini,

$\hat{s}_{(i)}$: Parametre tahminlerinin varyansını (varyans – kovaryans matrisindeki köşegen elemanları) ifade etmektedir.

4. ARAŞTIRMA BULGULARI

4.1. En Küçük Kareler Yönteminin Bulguları

Ek-1' de verilen verilerin EKK yöntemine ait $\hat{b}^*$ katsayıları, önem testleri ve regresyon eşitliği aşağıdaki gibi bulunmuştur.

Çizelge 4.1. EKK Yöntemi ile Verilere Ait Parametre Tahminleri ve Önem Testleri

Değişkenler	Katsayılar	Standart Hata	t	P
X_1 (ham yağ)	0.406	0.254	1.599	0.113
X_2 (ham protein)	-0.625	0.188	-3.320	0.001
X_3 (adf)	2.816	0.373	7.549	0.000
X_4 (ndf)	-3.085	0.502	-6.144	0.000
R^2 = %62		R^2 (düzeltilmiş) = %60		

Yukarıdaki tabloya göre regresyon denklemi eşitlik (4.1)' de ifade edilmektedir.

$$\hat{Y} = 0.00000034 + 0.406X_1 - 0.625X_2 + 2.816X_3 - 3.085X_4 \qquad (4.1)$$

Buna göre diğer değişkenler sabit tutulursa, ham yağ miktarı bir birim arttığında yem tüketiminin 0.406 birim artması, ham protein miktarı bir birim arttığında yem tüketiminin 0.625 birim azalması, adf miktarı bir birim arttığında yem tüketiminin 2.816 birim artması ve ndf tüketimi bir birim arttığında yem tüketiminin 3.085 birim azalması beklenmektedir. Çizelge 4.1 incelendiğinde ham protein, adf ve ndf miktarının elde edilen regresyon modeline katkısı istatistiki olarak önemli ($p<0.01$) bulunmuş olmasında rağmen ham yağ miktarı önemsiz ($p>0.05$) olarak belirlenmiştir.

Çizelge 4.2' de F değerine bakıldığında, H_0 hipotezinin reddedildiği, yani regresyon modelinin önemli olduğu görülmektedir.

Çizelge 4.2. Verilere Ait Varyans Analiz Tablosu

V.K.	S.D.	K.T.	K.O.	F	P
Regresyon	4	69.853	17.463	44.117	0.000
Hata	109	43.147	0.396		
Genel	113	113			

4.2. Çoklu Bağlantının Belirlenmesi

* Bağımsız değişkenler arasındaki korelasyon matrisi incelendiğinde bağımsız değişkenler arasındaki korelasyon katsayılarının yüksek ilişkili olduğu görülmektedir. Ancak bu durum çoklu bağlantının olup olmadığının belirlenmesinde yeterli değildir. Bağımsız değişkenlere ait korelasyon matrisi aşağıdaki gibi bulunmuştur.

	X_1	X_2	X_3	X_4
X_1	1.0000	0.937	0.859	0.929
X_2	0.937	1.0000	0.881	0.918
X_3	0.859	0.881	1.0000	0.978
X_4	0.929	0.918	0.978	1.0000

* Çizelge 4.3 incelendiğinde R_j^2 değerlerinin tümü belirtme katsayısından büyük ve 1'e yakın olduğu görülmektedir. Bu da çoklu bağlantının var olduğu konusunda bilgi vermektedir.

Çizelge 4.3. R_j^2 Değerleri

Bağımsız Değişkenler	R_j^2
X_1	0.95
X_2	0.90
X_3	0.97
X_4	0.97

* Çizelge 4.4'de görüldüğü gibi korelasyon matrisinin VBF değerleri 10'dan büyüktür. Ayrıca en büyük özdeğerin en küçük özdeğere bölümü ile elde edilen koşul sayısı değeri 30 ' dan büyüktür. Bu da çoklu bağlantının olduğunu belirtmektedir.

Çizelge 4.4. VBF, l_j ve Koşul Sayısı Değerleri

	VBF	l_j	Koşul Sayısı
X_1	18.428	3.751343	1.00
X_2	10.118	0.173691	21.6
X_3	39.720	0.066576	56.35
X_4	71.961	0.008390	447.1

4.3. Ridge Regresyon Yönteminin Bulguları

4.3.1. Ridge Parametresini (k^*) Belirleme

NCSS (2004) istatistik paket programının kullanımı ile elde edilen k^*'ya karşılık gelen VBF değerleri çizelge 4.5'te verilmiştir.

Çizelge 4.5. k* ve VBF Değerleri

k*	VBF_1	VBF_2	VBF_3	VBF_4
0.001	16.0116	9.7397	32.2273	57.7092
0.002	14.2156	9.4053	26.7834	47.3631
0.004	11.7402	8.8257	19.5610	33.6588
0.006	10.1192	8.3275	15.1124	25.2423
0.008	8.9715	7.8864	12.1720	19.7001
0.01	8.1100	7.4891	10.1217	15.8540
0.02	5.6939	5.9428	5.400	7.1488
0.04	3.6859	4.0695	2.8886	2.8652
0.06	2.6990	2.9918	2.0434	1.6538
0.08	2.0964	2.3077	1.5989	1.1241
0.1	1.6909	1.8435	1.3142	0.8366
0.2	0.7831	0.8190	0.6677	0.3444
0.3	0.4708	0.4804	0.4240	0.2111
0.4	0.3233	0.3250	0.3008	0.1524
0.5	0.2409	0.2400	0.2290	0.1203
0.6	0.1897	0.1879	0.1829	0.1002
0.7	0.1555	0.1534	0.1514	0.0866
0.8	0.1312	0.1292	0.1287	0.0766
0.9	0.1133	0.1114	0.1117	0.0691
1	0.0995	0.0978	0.0986	0.0631

Çizelge 4.5 incelendiğinde k* değerlerine karşılık gelen VBF değerlerinin k*=0.2 iken 1'e yakın olduğu görülmektedir. k*=0.2 alındığında regresyon denklemi eşitlik (eşitlik 4.2)'deki gibi bulunmuştur.

$$\hat{Y} = 0.000001 + 0.19838X_1 - 0.57613X_2 + 2.40976X_3 - 2.53496X_4 \qquad (4.2)$$

Buna göre diğer değişkenler sabit tutulursa, ham yağ miktarı bir birim arttığında yem tüketiminin 0.19838 birim artması, ham protein miktarı bir birim arttığında yem tüketiminin 0.57613 birim azalması, adf miktarı bir birim arttığında yem tüketiminin 2.40976 birim artması ve ndf tüketimi bir birim arttığında yem tüketiminin 2.53496 birim azalması beklenmektedir.

Çizelge 4.6. k*=0.2 Değeri İçin Varyans Analiz Tablosu

V.K.	S.D.	K.T.	K.O.	F
Regresyon	4	66.844	16.71	39.51
Hata	109	46.156	0.423	
Genel	113	113		

$$R^2 = \frac{\text{Re}\, gKT}{GKT} = \frac{66.844}{113} = 0.59 \qquad (4.3)$$

k*=0.2 değeri için F değerine bakıldığında H_0 hipotezinin reddedildiği ve regresyon modelinin önemli olduğu görülmektedir.

Çizelge 4.7'de EKK ve RR yöntemine ait (k*=0.2 için) VBF değerleri ifade edilmiştir.

Çizelge 4.7. EKK ve RR Yöntemlerine Ait VBF Değerleri

VBF Değerleri	EKK	k* =0.2 için VBF Değerleri
VBF_1	18.428	0.7831
VBF_2	10.118	0.8190
VBF_3	39.720	0.6677
VBF_4	71.961	0.3444

Çizelge 4.8' de E.K.K. ve Ridge regresyon yöntemlerine ait R^2 ve HKO değerleri karşılaştırmalı olarak verilmiştir.

Çizelge 4.8. E.K.K. ve R.R. Yöntemine Ait R^2 ve HKO Değerleri

	R^2	HKO
E.K.K.	%62	0.396
R.R.	%59	0.423

4.4. M-Tahmin Edici Bulguları

M-tahmin edici bulguları R istatistik paket programda MASS kütüphanesi kullanılarak hesaplanmıştır. İlgili kütüphane içerisinde iken V5 bağımlı değişken, V1(ham yağ), V2(ham protein), V3(adf) ve V4(ndf) bağımsız değişkenler olmak üzere aşağıdaki komutlar yardımı ile parametre tahminleri elde edilmiştir. İlgili program parametre tahminlerinde Huber fonksiyonunu kullanmaktadır.

```
> library(MASS)
> gr=rlm(V5~V1+V2+V3+V4,a)
> summary(gr) (Faraway, 2005)
```

Çizelge 4.9. M-Tahmin Edici Kullanılarak Elde Edilen Analiz Sonuçları

Değişkenler	$\hat{b}$	Standart Hata	t	
X_1	0.4479	0.2346	1.9091	
X_2	-0.5721	0.1738	-3.2907	
X_3	2.8635	0.3445	8.3132	
X_4	-3.2292	0.4636	-6.9651	R^2=0.61

Elde edilen regresyon denklemi eşitlik 4.4' da verilmiştir.

$$\hat{Y} = -0.0373 + 0.4479X_1 - 0.5721X_2 + 2.8635X_3 - 3.2292X_4 \qquad (4.4)$$

Denkleme göre diğer değişkenler sabit tutulmak şartıyla, ham yağ miktarı bir birim arttığında yem tüketimi 0.4479 birim artmaktadır. Benzer şekilde ham protein miktarı bir birim arttığında yem tüketimi 0.5721 birim azalmaktadır. Diğer taraftan adf miktarı bir birim arttığında yem tüketimi 2.8635 birim artmakta, ndf miktarı bir birim arttığında yem tüketimi 3.2292 birim azalmaktadır. Çizelge 4.9'daki t değerleri $t_{0.025,109} = 1.982$ cetvel değeri ile karşılaştırıldığında ham yağ bağımsız değişkeni hariç diğer bağımsız değişkenlerin modele olan katkısının istatistiki olarak önemli olduğu tespit edilmiştir.

4.5. Aykırı Değerlerin İncelenmesi

Çizelge 4.10'da ifade edilen değerler incelendiğinde 19, 33, 51, 57 ve 95 nolu gözlemlerin t_i, d_i ve r_i değerleri 2'den büyük olduğu için aykırı değer olduğu tahmin edilmektedir. Ortalama değişim aykırı modellemesi yardımıyla bu gözlemler tekrar incelenmiş olup 0.05 önem seviyesinde 19, 33, 51, 57 ve 95 nolu gözlemlerin aykırı değer oldukları tespit edilmiştir. Diğer yandan Bonferroni testine göre $t_{108,0.000219}$=3.391 kritik değeri ile $|t_i|$ değerleri karşılaştırıldığında sadece 57 nolu gözlemin aykırı değer olduğu görülmektedir.

Çizelge 4.10. Aykırı Değerler İle İlgili İstatistikler

Gözlem No	e_i	t_i	r_i	d_i	h_{ii}
1	-1.11285	-1.81195	-1.79326	-1.76879	0.027111
2	-.32394	-.52022	-.52197	-.51488	0.026979
3	-.25321	-.40644	-.40800	-.40245	0.027014
4	-.42210	-.67847	-.68016	-.67090	0.027030
5	-.54826	-.88254	-.88343	-.87141	0.027041
6	.55731	.89725	.89805	.88580	0.027096
7	-.05479	-.08789	-.08829	-.08708	0.027194
8	.57359	.92383	.92445	.91167	0.027451
9	.47898	.77112	.77256	.76130	0.028936
10	-.17649	-.28465	-.28586	-.28052	0.036987
11	-.24749	-.40094	-.40249	-.39336	0.044880
12	-.12341	-.19761	-.19849	-.19615	0.023456
13	1.14436	1.86887	1.84786	1.81886	0.031139
14	-.35458	-.57243	-.57420	-.56357	0.036694
15	-.89573	-1.45949	-1.45198	-1.42369	0.038588
16	-.41183	-.66472	-.66643	-.65458	0.035261
17	-.03491	-.05619	-.05645	-.05549	0.033828
18	.01754	.02815	.02828	.02787	0.028395
19	1.60878	2.65045	2.58010	2.55702	0.017808
20	.51728	.83105	.83223	.82218	0.024007
21	-.75241	-1.21389	-1.21126	-1.19589	0.025225
22	-.73027	-1.17606	-1.17399	-1.16071	0.022502
23	-.83089	-1.33959	-1.33474	-1.32063	0.021035
24	-.26722	-.43237	-.43400	-.42472	0.042286
25	-.48735	-.78467	-.78606	-.77460	0.028950
26	-.23480	-.37982	-.38132	-.37319	0.042173
27	-.46090	-.74734	-.74885	-.73256	0.043038
28	-.54019	-.86721	-.86820	-.85859	0.022009
29	-.50446	-.82373	-.82495	-.80180	0.055330

30	-.43657	-.73593	-.73748	-.69389	0.114730
31	.65321	1.07922	1.07841	1.03823	0.073128
32	-.49309	-.79669	-.79803	-.78373	0.035520
33	-1.53852	-2.60554	-2.53899	-2.44535	0.072401
34	-.32951	-.55058	-.55235	-.52374	0.100902
35	.13974	.22892	.22992	.22211	0.066762
36	.22035	.41459	.41617	.35023	0.291808
37	-.94195	-1.54016	-1.53056	-1.49716	0.043163
38	-.66960	-1.11007	-1.10889	-1.06427	0.078863
39	-.81358	-1.31563	-1.31124	-1.29311	0.027457
40	-.26300	-.42332	-.42492	-.41802	0.032208
41	-.35675	-.57354	-.57532	-.56702	0.028636
42	-.14231	-.22902	-.23002	-.22619	0.033010
43	-.27377	-.44108	-.44272	-.43514	0.033957
44	.71782	1.16243	1.16057	1.14092	0.033575
45	.43551	.70459	.70622	.69221	0.039286
46	.94467	1.53805	1.52850	1.50148	0.035045
47	.29103	.46729	.46897	.46257	0.027121
48	.43234	.69426	.69592	.68717	0.024988
49	.21236	.34080	.34219	.33752	0.027085
50	.80455	1.30456	1.30038	1.27876	0.032971
51	1.33839	2.18989	2.15273	2.12727	0.023518
52	.09242	.14876	.14943	.14690	0.033562
53	-.32379	-.52342	-.52517	-.51464	0.039703
54	.25792	.41681	.41840	.40994	0.040003
55	.34388	.54930	.55106	.54657	0.016250
56	.41670	.66789	.66959	.66231	0.021636
57	2.31193	3.95014	3.70946	3.67463	0.018688
58	1.01016	1.63679	1.62433	1.60557	0.022962
59	-.52100	-.83643	-.83758	-.82809	0.022527
60	.00548	.00878	.00882	.00872	0.024508

61	-.34563	-.55388	-.55565	-.54935	0.022553
62	.39721	.64364	.64537	.63133	0.043065
63	.31705	.51106	.51280	.50393	0.034323
64	.51646	.84084	.84197	.82086	0.049508
65	.58714	.94971	.95013	.93321	0.035308
66	.29740	.47828	.47998	.47269	0.030167
67	.33957	.56470	.56647	.53971	0.092240
68	.40099	.66940	.67110	.63734	0.098084
69	-.88062	-1.51736	-1.50837	-1.39968	0.138922
70	.18295	.30111	.30237	.29079	0.075141
71	-.52272	-.85604	-.85709	-.83081	0.060373
72	.47823	.79325	.79461	.76011	0.084946
73	.41807	.68898	.69065	.66448	0.074342
74	-.19298	-.31228	-.31358	-.30672	0.043240
75	-.26844	-.43375	-.43537	-.42667	0.039578
76	-.37799	-.64127	-.64301	-.60078	0.127038
77	-.96429	-1.56406	-1.55379	-1.53266	0.027006
78	-.29347	-.47149	-.47318	-.46644	0.028267
79	-.30495	-.48973	-.49145	-.48470	0.027281
80	-.28115	-.45178	-.45343	-.44686	0.028780
81	-.41103	-.66129	-.66301	-.65330	0.029077
82	.63650	1.02697	1.02672	1.01167	0.029096
83	.19142	.30752	.30880	.30425	0.029297
84	.75915	1.22803	1.22518	1.20660	0.030088
85	.38500	.61867	.62043	.61192	0.027241
86	.12790	.20545	.20636	.20329	0.029464
87	-.01864	-.03001	-.03014	-.02962	0.034244
88	.34054	.54607	.54784	.54126	0.023874
89	1.24244	2.02774	1.99940	1.97476	0.024498
90	-.13109	-.21099	-.21192	-.20836	0.033303
91	-.61081	-.99004	-.99013	-.97084	0.038582

92	-.07694	-.12407	-.12463	-.12230	0.037150
93	.15549	.24897	.25005	.24714	0.023162
94	.21712	.34753	.34894	.34509	0.021934
95	1.96034	3.28240	3.14444	3.11581	0.018133
96	.76371	1.23126	1.22836	1.21386	0.023470
97	-.63776	-1.02620	-1.02595	-1.01367	0.023801
98	-.36237	-.58106	-.58284	-.57596	0.023464
99	-.58931	-.94657	-.94702	-.93666	0.021758
100	.06390	.10333	.10381	.10157	0.042673
101	-.08516	-.13692	-.13754	-.13536	0.031505
102	.13977	.22642	.22741	.22215	0.045722
103	.06313	.10189	.10236	.10034	0.038998
104	-.12248	-.19634	-.19721	-.19467	0.025559
105	-.08247	-.13487	-.13549	-.13107	0.064048
106	-.01669	-.02780	-.02792	-.02653	0.097120
107	-.11267	-.18309	-.18391	-.17909	0.051719
108	-.15613	-.25264	-.25373	-.24815	0.043477
109	-1.03062	-1.70980	-1.69491	-1.63808	0.065927
110	.07435	.12334	.12390	.11818	0.090227
111	.27893	.45479	.45645	.44334	0.056647
112	.01474	.02509	.02520	.02343	0.135416
113	-.60413	-.97971	-.97989	-.96021	0.039750
114	-.52375	-.86406	-.86507	-.83245	0.073994

4.6. Aykırı Değerler Veri Setinden Uzaklaştırıldığında Elde Edilen EKK Bulguları

19, 33, 51, 57 ve 95 nolu gözlemler veri setinden uzaklaştırıldıktan sonra elde edilen EKK bulguları aşağıdaki çizelgede belirtilmiştir.

Çizelge 4.11. Aykırı Değerler Çıkarıldıktan Sonra Elde Edilen EKK Sonuçları

Değişkenler	Katsayılar	Standart Hata	t	P
X_1	0.442	0.212	2.082	0.04
X_2	-0.529	0.155	-3.411	0.001
X_3	2.847	0.302	9.419	0.000
X_4	-3.255	0.409	-7.962	0.000
R^2 = %72		R^2 (düzeltilmiş) = %71		

Yukarıdaki tabloya göre regresyon denklemi eşitlik (4.5)' da verilmiştir.

$$\hat{Y} = -0.0532 + 0.442X_1 - 0.529X_2 - 2.847X_3 - 3.255X_4 \quad (4.5)$$

. İlgili verilere ait varyans analiz tablosu Çizelge 4.12'de belirtilmiştir.

Çizelge 4.12. Aykırı Değerler Uzaklaştığında Elde Edilen Varyans Analiz Tablosu

V.K.	S.D.	K.T.	K.O.	F
Regresyon	4	69.74	17.435	68.04
Hata	104	26.649	0.256	
Genel	108	96.389		

4.7. Aykırı Değerler Uzaklaştırıldığında Elde Edilen Verilerde Çoklu Bağlantının Belirlenmesi

İlgili verilerde bağımsız değişkenlere ait korelasyon matrisi aşağıdaki gibi bulunmuştur.

	X_1	X_2	X_3	X_4
X_1	1.0000	0.940	0.863	0.932
X_2	0.940	1.0000	0.881	0.918
X_3	0.863	0.881	1.0000	0.978
X_4	0.932	0.918	0.978	1.0000

* Çizelge 4.13 incelendiğinde R_j^2 değerlerinin belirtme katsayısından büyük ve 1'e yakın olduğu görülmektedir. Bu da çoklu bağlantının var olduğu konusunda bilgi vermektedir.

Çizelge 4.13. R_j^2 Değerleri

Bağımsız Değişkenler	R_j^2
X_1	0.945
X_2	0.905
X_3	0.97
X_4	0.98

* Çizelge 4.14'de görüldüğü gibi korelasyon matrisinin VBF değerleri 10'dan büyüktür. Ayrıca en büyük özdeğerin en küçük özdeğere bölümü ile elde edilen koşul sayısı değeri 30 ' dan büyüktür. Bu da çoklu bağlantının olduğunu belirtmektedir.

Çizelge 4.14. VBF, l_j ve Koşul Sayısı ve Koşul İndeksi Değerleri

	VBF	l_j	Koşul Sayısı
X_1	19.6585	3.756166	1.00
X_2	10.5125	0.171901	21.85
X_3	39.7659	0.063678	58.99
X_4	72.9871	0.008255	455.01

Çizelge 4.15 incelendiğinde aykırı değerler veri setinden çıkarıldıktan sonra k* değerlerine karşılık gelen VBF değerlerinin k*=0.2 olduğunda 1'e yakın olduğu görülmektedir.

4.8. Aykırı Değerler Çıkarıldıktan Sonra Ridge Regresyon Yönteminin Bulguları

4.8.1. Ridge Parametresini (k^*) Belirleme

NCSS (2004) istatistik paket programının kullanımı ile elde edilen k^* ve VBF değerleri çizelge 4.15'te verilmiştir.

Çizelge 4.15. k* ve VBF Değerleri

k*	VBF_1	VBF_2	VBF_3	VBF_4
0.001	17.0030	10.0884	32.1800	58.3406
0.002	15.0380	9.7177	26.6936	47.7579
0.004	12.3431	9.0830	19.4511	33.8120
0.006	10.5874	8.5439	15.0127	25.2928
0.008	9.3489	8.0706	12.0892	19.7044
0.01	8.4220	7.6471	10.0556	15.8374
0.02	5.8388	6.0193	5.3855	7.1222
0.04	3.7223	4.0831	2.8992	2.8535
0.06	2.6998	2.9859	2.0561	1.6475
0.08	2.0833	2.2957	1.6099	1.1199
0.1	1.6723	1.8301	1.3232	0.8334
0.2	0.7654	0.8102	0.6707	0.3428
0.3	0.4583	0.4751	0.4251	0.2100
0.4	0.3143	0.3216	0.3012	0.1516
0.5	0.2342	0.2376	0.2290	0.1197
1	0.0973	0.0971	0.0984	0.0628

Buna göre k*=0.2 alındığında regresyon denklemi aşağıdaki gibi bulunmuştur (eşitlik 4.6).

$$\hat{Y} = -0.05360 + 0.20445X_1 - 0.46385X_2 + 2.41349X_3 - 2.66505X_4 \qquad (4.6)$$

İlgili denkleme göre diğer değişkenler sabit tutulmak şartıyla, ham yağ tüketimi bir birim arttığında yem tüketiminin 0.20445 birim artması, ham protein tüketimi bir birim arttığında yem tüketiminin 0.46385 birim azalması, adf tüketimi bir birim arttığında yem tüketiminin 2.41349 birim artması ve ndf tüketimi bir birim arttığında yem tüketiminin 2.66505 birim azalması beklenmektedir.

Çizelge 4.16. Aykırı Değerler Çıkarıldıktan Sonra k*=0.2 Değeri İçin RR Yöntemi İle Elde Edilen Varyans Analiz Tablosu

V.K.	S.D.	K.T.	K.O.	F
Regresyon	4	66.478	16.619	57.91
Hata	104	29.911	0.287	
Genel	108	96.389		

$$R^2 = \frac{\text{Re}\, gKT}{GKT} = \frac{66.478}{96.389} = 0.69 \qquad (4.7)$$

Çizelge 4.16 incelendiğinde regresyon denkleminin istatistiki olarak önemli olduğu gözlenmektedir.

Çizelge 4.17'de Aykırı değerler çıkarıldıktan sonra EKK ve RR yöntemine ait (k*=0.2 için) VBF değerleri ifade edilmiştir.

Çizelge 4.17. EKK ve RR Yöntemlerine Ait VBF Değerleri

VBF Değerleri	EKK	k*=0.2 için VBF Değerleri
VBF_1	19.6585	0.7654
VBF_2	10.5125	0.8102
VBF_3	39.7659	0.6707
VBF_4	72.9871	0.3428

4.9. Aykırı Değerler Çıkarıldıktan Sonra M-Tahmin Edici Yönteminin Bulguları

R istatistik paket program yardımıyla yapılan M-Tahmin Edici' ye ait analiz sonuçları Çizelge 4.18' da verilmiştir.

Çizelge 4.18. Aykırı Değerler Veri Setinden Uzaklaştırıldıktan Sonra M-Tahmin Edici Yöntemi İle Elde Edilen Analiz Sonuçları

Değişkenler	$\hat{b}$	Standart Hata	t	
X_1	0.4424	0.2190	2.0204	
X_2	-0.5130	0.1603	-3.2009	
X_3	2.8711	0.3121	9.1992	
X_4	-3.2916	0.4222	-7.7966	R^2=0.68

Elde edilen regresyon denklemi eşitlik (4.8)' da verilmiştir.

$$\hat{Y} = -0.06515 + 0.4424X_1 - 0.5130X_2 + 2.8711X_3 - 3.2916X_4 \quad (4.8)$$

Buna göre ham yağ miktarı bir birim arttığında yem tüketiminin 0.4424 birim artması, ham protein miktarı bir birim arttığında yem tüketiminin 0.5130 birim azalması, adf miktarı bir birim arttığında yem tüketiminin 2.8711 birim artması ve ndf miktarı bir birim arttığında yem tüketimini 3.2916 birim azalması beklenmektedir.

Çizelge 4.19, 4.20 ve 4.21'de aykırı değer varlığında ve veri setinden aykırı değerler çıkarıldıktan sonra elde edilen bazı analiz sonuçları karşılaştırmalı olarak verilmiştir.

Çizelge 4.19. Aykırı Değerler Uzaklaştırıldıktan Sonra Elde Edilen HKO ve R^2 Değerleri

	R^2	HKO
E.K.K.	%72	0.256
R.R.	%69	0.287

Çizelge 4.20. Aykırı Değer Varlığında EKK ve M-Tahmin Ediciden Elde Edilen Analiz Sonuçları

Değişkenler	Standart Hata (EKK)	Standart Hata (M-Tahmin)	t (EKK)	t (M-Tahmin)
X_1	0.254	0.2346	1.599	1.9091
X_2	0.188	0.1738	-3.320	-3.2907
X_3	0.373	0.3445	7.549	8.3132
X_4	0.502	0.4636	-6.144	-6.9651

Aykırı değer varlığında EKK ve M-tahmin ediciye ait t değerlerinde her iki yöntemde de birinci bağımsız değişkenin modele olan katkısının istatistiki olarak önemli olmadığı görülmektedir.

Çizelge 4.21. Aykırı Değer Yokluğunda EKK ve M-Tahmin Ediciden Elde Edilen Analiz Sonuçları

Değişkenler	Standart Hata (EKK)	Standart Hata (M-Tahmin)	t (EKK)	t (M-Tahmin)
X_1	0.212	0.2190	2.082	2.0204
X_2	0.155	0.1603	-3.411	-3.2009
X_3	0.302	0.3121	9.419	9.1992
X_4	0.409	0.4222	-7.962	-7.7966

Aykırı değerler veri setinden uzaklaştırıldıktan sonra EKK ve M-tahmin edici yöntemlerinden elde edilen regresyon denkleminde tüm bağımsız değişkenlerin istatistiki olarak önemli olduğu belirlenmiştir.

5. TARTIŞMA VE SONUÇ

En küçük kareler yöntemi kullanılarak yapılan parametre tahminleri için gerekli varsayımlardan bağımsız değişkenler arasında bir ilişki olmaması varsayımı sağlanmadığında çoklu bağlantı problemi ortaya çıkmaktadır. Bu nedenle EKK yöntemi kullanılarak elde edilen analiz sonuçları yanlış sonuçlara ve yanlış modellerin oluşumuna neden olabilmektedir (Ergüneş, 2004).

Çalışmada öncelikle En Küçük Kareler yöntemi ile parametre tahminleri elde edilmiştir. Veriler incelendiğinde bağımsız değişkenler arasında korelasyon katsayılarının ilişkili olması, R_j^2 değerleri ilk oluşturulan modelin belirtme katsayısından büyük ve bire yakın olması, VBF değerlerinin 10'dan büyük olması, özdeğerlerin birbirine oranının 30'un üzerinde olması sebebiyle bağımsız değişkenler arasında çoklu bağlantı olduğu söylenebilir. EKK yöntemi ile elde edilen VBF değerleri, Ridge Regresyon yöntemi kullanılarak elde edilen VBF değerlerinden daha yüksek bulunmuştur. Ridge Regresyon yönteminde hesaplanmış olan belirtme katsayısı EKK yöntemi ile elde edilen belirtme katsayısı değerinden farklı olduğu belirlenmiştir.

Diğer taraftan EKK yöntemi ile elde edilmiş olan hata kareler ortalaması değeri RR yöntemi ile elde edilmiş olan hata kareler ortalaması değerinden düşük iken belirtme katsayısı değeri büyüktür. Elde edilen bu sonuç Ergüneş (2004) 'in bildirişi ile benzer özellik göstermektedir. EKK regresyon yöntemi ile elde edilen denklemde her bir bağımsız değişkenin bir birim artması ile bağımlı değişken kendi birimi cinsinden -0.487 birim azalması beklenirken, aynı değer RR yöntemi uygulandığında -0.5029 olarak tespit edilmiştir. Çoklu bağlantı durumunun EKK yöntemi ile yapılan parametre tahminleri üzerine etkisi sözkonusudur. EKK yöntemine ait varyans analiz tablosu incelendiğinde regresyon kareler toplamının RR yönteminde düştüğü buna karşılık hata kareler toplamının RR yönteminde arttığı gözlemlenmiştir. Burada da çoklu bağlantının etkileri görülmektedir.

Veri setinde çoklu bağlantı problemini giderebilmek için Ridge Regresyona alternatif olarak önerilen Stein ve Liu tahmin edicileri de kullanılabilir. Ayrıca bağımsız değişkenlerin bir veya birkaçının modelden çıkarılması (modelden

çıkarılacak değişkenin bağımlı değişkeni açıklamada çok önemli olmadığı durumlarda), veri setine yeni gözlem eklenmesi, değişkenler dönüştürme işlemlerinin yapılması, bağımsız değişkenlerin kümeleştirilmesi yoluyla çoklu bağlantılı değişkenlerin birleştirilerek yeni bir değişken oluşturmak gibi farklı tekniklerde önerilmektedir (İmir,1986).

EKK yöntemi ve M-Tahmin edici ile elde edilmiş olan standart hata değerleri karşılaştırıldığında M-Tahmin ediciye ait parametrelerin standart hatalarının daha düşük olduğu gözlemlenmiştir. Ayrıca EKK yöntemi ile elde edilen parametrelere ait "t" değerleri M-Tahmin ediciye oranla daha düşük bulunmuştur.

Aykırı değerleri belirlemek için kullanılan artık değerlere ait tablo incelendiğinde bazı değerlerin aykırı değer olabileceği düşünülmüştür. Ortalama değişim aykırı değer modellemesi ve bonferroni testi uygulandıktan sonra aykırı değerler olduğu belirlenen gözlemler veri setinden uzaklaştırılmıştır. Aykırı değerler veri setinden çıkarıldıktan sonra EKK yöntemi ile elde edilen analiz sonuçları karşılaştırıldığında belirtme katsayısının aykırı değer varken elde edilen belirtme katsayısından yüksek olduğu belirlenmiştir. Hata kareler ortalaması değeri aykırı değerler veri setinden çıkarıldıktan sonra düşmüştür. Benzer durum p değerlerinde de ortaya çıkmaktadır. Aykırı değer varlığında EKK yöntemi sonuçlarına göre ham yağ bağımsız değişkenine ait p değeri yüksek ve önemsiz bulunurken, aykırı değerler veri setinden uzaklaştırıldıktan sonra elde edilen sonuçlarda aynı bağımsız değişkene ait p değeri düşmüş ve istatistiki olarak önemli bulunmuştur. Bu sonuca göre veri setinde aykırı değerlerin varlığı p değerinin sapmasına ve hata kareler ortalaması değerinin yükselmesine neden olduğu belirtilebilir. Benzer bir durum parametre tahminlerine ait standart hata değerlerinde de görülmektedir. Aykırı değer varlığında EKK yöntemi ile elde edilen parametrelerin standart hata değerleri aykırı değerler çıkarıldıktan sonra hesaplanan standart hata değerlerinden daha yüksek bulunmuştur. Aykırı değerler çıkarıldıktan sonra EKK yöntemi ile elde edilmiş olan bağımsız değişkenlere ait "t" istatistiği değerleri aykırı değer varlığındaki "t" değerlerinden daha yüksek bulunmuştur.

Aykırı değer varlığında EKK yöntemi ile elde edilen denkleme ait F değeri aykırı değerler uzaklaştırıldıktan sonra elde edilmiş olan F değerinden daha düşüktür. Benzer durum RR sonuçlarında da görülmektedir.

Ridge regresyon yöntemine ait sonuçlar aykırı değer varlığında ve aykırı değerler veri setinden çıkarıldıktan sonra karşılaştırıldığında EKK yöntemi ile elde edilmiş olan sonuçlarla benzerlik gösterdiği belirlenmiştir. Elde edilen bu sonuçlar High (2004)' ün elde ettiği sonuçlarla uyumludur.

Diğer taraftan aykırı değer varlığında ve aykırı değerler veri setinden uzaklaştırıldıktan sonra çoklu bağlantı durumu incelenirse; aykırı değerler veri setinde iken elde edilmiş olan bağımsız değişkenler arasındaki korelasyon katsayılarının aykırı değerler çıkarıldıktan sonra elde edilen bağımsız değişkenlere ait korelasyon katsayılarından daha düşük olduğu görülmektedir. Aykırı değer varlığı bağımsız değişkenler arasında var olan yüksek bir ilişkiyi daha düşük göstermektedir. Benzer bir durum VBF değerlerinde de ortaya çıkmaktadır. Aykırı değerler veri setinden uzaklaştırıldıktan sonra elde edilen VBF değerleri aykırı değer varlığında elde edilen VBF değerlerinden az bir farkla yüksek bulunmuştur. Bağımsız değişkenlere ait koşul sayıları incelendiğinde özellikle 3. ve 4. bağımsız değişkenin koşul sayısı aykırı değerler veri setinden uzaklaştırıldıktan sonra artış göstermiştir. Bu sonuçlar göz önünde bulundurularak, veri setinde aykırı değer varlığının, bağımsız değişkenler arasındaki iç ilişki durumunu yani çoklu bağlantı durumunu maskeleyebildiği ifade edilebilir. Yani kullanılan veri setinde aykırı değer bulunuyorsa bağımsız değişkenler arasında ilişki olmasına rağmen çoklu bağlantı olmadığı gibi yanlış bir sonuca yönlendirebilir. Elde edilmiş olan bu sonuçlar Bek ve ark. (1996) ve Şahinler (1997)' nin çalışmalarında elde ettiği sonuçlarla benzerlik göstermektedir.

Aykırı değerler veri setinden uzaklaştırıldıktan sonra EKK yöntemi ve M-Tahmin edici ile elde edilmiş olan standart hata değerleri karşılaştırıldığında M-Tahmin ediciye ait parametrelerin standart hata değerlerinin daha yüksek olduğu gözlemlenmiştir. Ayrıca EKK yöntemi ile elde edilen parametrelere ait "t" değerleri M-Tahmin ediciye oranla daha yüksek bulunmuştur.

Aykırı değerler varken ve veri setinden uzaklaştırıldıktan sonra M-Tahmin ediciden elde edilmiş olan belirtme katsayısı değerleri ile en küçük kareler yöntemi ile elde edilmiş olan belirtme katsayısı değerleri karşılaştırıldığında M-Tahmin ediciye ait belirtme katsayısı değerlerinin daha az etkilendiği görülmektedir. M-tahmin edicide aykırı değer varlığında bu değer %61 iken aykırı değerler uzaklaştırıldıktan sonra %68 olmaktadır.

Sonuç olarak;

- Aykırı değer varlığında belirtme katsayısı düşmekte, hata kareler ortalaması yükselmektedir.
- Regresyon denklemine ait belirtme katsayısı ve hata kareler ortalaması değerleri aykırı değer varlığından etkilenmektedir.
- Aykırı değer veri setinde iken ve veri setinden çıkarıldıktan sonra parametre tahminlerinden elde edilen t, F değerleri farklılık göstermektedir.
- Aykırı değerler çoklu bağlantı durumunu maskeleyebilmektedir.

.

KAYNAKLAR

AHN, H., JAMES, R.T., 1999. Outlier Detection İn Phosphorus Dry Deposition Rates Measured İn South Florida. Atmospheric Environment ,33: 5123-5131.

ALBAYRAK, A.S., 2005. Çoklu Doğrusal Bağlantı Halinde Enküçük Kareler Tekniğinin Alternatifi Yanlı Tahmin Teknikleri Ve Bir Uygulama. ZKÜ Sosyal Bilimler Dergisi Cilt 1, Sayı 1.

ALMA, Ö.G., ve VUPA, Ö., 2008. Regresyon Analizinde Kullanılan En Küçük Kareler ve En Küçük Medyan Kareler Yöntemlerinin Karşılaştırılması. Sdü Fen Edebiyat Fakültesi Fen Dergisi (E-Dergi). 3(2) 219-229.

AŞIKGİL, B., 2006. Çoklu Doğrusal Regresyonda Aykırı, Etkili Değerlerin Araştırılması ve Bir Uygulama. Mimar Sinan Güzel Sanatlar Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı Yüksek Lisans Tezi.

BEK, Y., ŞAHİNLER, S., KAYAALP, G.T., 1996. Etkili Gözlemlerin Belirlenmesinde Cook Ve Welsch-Kuh İstatistiğinin Karşılaştırmalı Olarak İncelenmesi. MKÜ Ziraat Fakültesi Dergisi, 1(1): 85-100.

BILLOR, N., and KIRAL, G., 2008. A Comparison of Multiple Outlier Detection Methods for Regression Data, Communications in Statistics - Simulation and Computation,37:3,521 -545.

BIRKES, D., and DODGE, Y., 1993. Alternative Methods of Regression, A Wiley İnterscience Publication.

CHATTERJEE, S and HADİ, A.S., 1986. Influential Observations, High Leverage Points, and Outliers in Linear Regression. Statistical Science, Vol. 1, No. 3, pp. 379-393.

COOK, R. D. and WEISBERG, S., 1982. Residuals and Influence in Regression, First Edition, Chapman and Hall.

COŞKUNTUNCEL, O., 2005. Karma Denemelerde ve Modellerde Robust İstatistiksel Analizler. Çukurova Üniversitesi Fen Bilimleri Enstitüsü Matematik Anabilim Dalı Doktora Tezi.

ÇANKAYA, S., KAYAALP, G.T., SANGUN, L., TAHTALİ, Y., AKAR, M., 2006. A Comparative Study Of Estimation Methods For Parameters İn Multiple Linear Regression Model. J. Appl. Anim. Res., 29 : 43-47.

DRAPER, N. R., and SMİTH, H. 1998. Applied Regression Analysis , Wiley İnterscience Publication, p: 567.

ERGÜL, B., 2006. Robust Regresyon ve Uygulamaları. Eskişehir Osmangazi Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı Yüksek Lisans Tezi.

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.

FARAWAY, J. J., 2005. Linear Models with R, Chapman&Hall / CRC, USA, pg:229.

FREUND, J. R., WİLSON, W. J. ve SA, P. 2006. Regression Analysis Statistical Modelling of a Response Variable (Second Edition), Elseiver Inc. page: 120.

GÜNDOĞAN, Y., 2005. Sağlam Regresyonda Kısmi Artık Grafiği. Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı, Yüksek Lisans Tezi.

HADİ, A. S. and SİMONOFF, J. S., 1993. Procedures for the Identification of Multiple Outliers in Linear Models. Journal of the American Statistical Association. Vol. 88, No. 424, 1264-1272.

HİGH, R., 2004. http://darkwing.uoregon.edu/~robinh/outl.txt. Erişim tarihi 21.07.2008.

İMİR, E., 1986. Çoklu Bağıntılı Doğrusal Modellerde Ridge Regresyon Yöntemiyle Parametre Kestirimi. T.C Anadolu Üniversitesi Yayınları, No. : 212.

KARADAVUT, U., GENÇ, A., TOZLUCA, A., KINACI, İ., 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.

KIM, S.S., PARK, S. H. and KRZANOWSKİ, W. J., 2008. Simultaneous Variable Selection And Outlier İdentification İn Linear Regression Using The Mean-Shift Outlier Model. Journal Of Applied Statistics,35:3,283 - 291.

KOÇ, Y.S., 2007. Robust Tahmin Edicileri ve Özellikleri. Çukurova Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalı Yüksek Lisans Tezi.

KONTRIMAS, V., and VERIKAS, A., 2006. Trackıng Of Doubtful Real Estate Transactıons By Outlıer Detectıon Methods: A Comparatıve Study, Informatıon Technology And Control, Vol. 35, No. 2.

LAWRENCE, K. D., and MARSH, L.C., 1984. Robust Ridge Estimation Methods For Predicting U. S. Coal Mining Fatalities. Commun. Statist.-Theor. Meth., 13(2), 139-149.

MARTİN, M. A. and ROBERTS, S., 2006. An Evaluation Of Bootstrap Methods For Outlier Detection İn Least Squares Regression',Journal Of Applied Statistics,33(7),703- 720.

NETER, J., KUTNER, M. H., NACHTSHEİM, C. J., WASSERMAN, W., 1996. Applied Linear Statistical Models.

ORTIZ, M.C., SARABIA, L.A., and HERRERO, A., 2006. Robust Regression Techniques A Useful Alternative For The Detection Of Outlier Data İn Chemical Analysis. Talanta 70:499–512.

SMYTH G. K. and HAWKİNS D. M., 2000. Robust Frequency Estimation Using Elemental Sets, Journal of Computational and Graphical Statistics, Vol. 9, 196-214.

ŞAHİNLER, S.,1997. Regresyon Analizinde Etkili Gözlemlerin (Influential Observations) Belirlenmesinde Kullanılan İstatistiklerin Karşılaştırmalı Olarak İncelenmesi, Ç.Ü. Fen Bilimleri Enstitüsü Zootekni Anabilim Dalı Doktora Tezi .

ŞAHİNLER, S., 2000. En Küçük Kareler Yöntemi ile Dogrusal Regresyon Modeli Olusturmanin Temel Prensipleri. MKÜ Ziraat Fakültesi Dergisi 5 (1-2): 57-73.

ŞEHİRLİ,2009.http://www.deu.edu.tr/userweb/kemal.sehirli/dosyalar/regresyon1-2.pdf., erişim tarihi 02.07.2009.

TÜRKAY, H., 2004. Doğrusal Regresyon Analizinde M Tahminciler ve Ekonometrik Bir Uygulama. Doğu Anadolu Bölgesi Araştırmaları, http://web.firat.edu.tr/daum/docs/31/19%20DO%C4%9ERUSAL%20REGRASYON%20ANAL%C4%B0Z%C4%B0NDE---HAKAN%20T%C3%9CRKAY%20(MUZAFFER%20DEM%C4%B0RBA%C5%9E-05334222697----10-%20SYF-106-115--%20%C3%96DEN.doc, Erişim Tarihi 02.12.2009.

WEISBERG, S., 2005. Applied Linear Regression, Third Edition, John Wiley&Sons, Inc. Page:197.

ÖZGEÇMİŞ

1982 yılında Adana'da doğdu. İlk ve Orta Öğrenimimi Adana'da tamamladı. 1999 yılında Çukurova Üniversitesi Fen Edebiyat Fakültesi Biyoloji Bölümünü kazandı. 2003 yılında mezun olduktan sonra aynı yıl Çukurova Üniversitesi Fen Bilimleri Enstitüsü Ortaöğretim Alan Öğretmenliği programını ve hazırlık eğitimini tamamladı. 2006 yılında Ç.Ü. Ziraat Fakültesi Zootekni Bölümü Biyometri ve Genetik Anabilim Dalı'nda Araştırma Görevlisi olma hakkını kazandı. Aynı bölümde başladığım yüksek lisans eğitimime devam etmekte.

Ek-1. Materyal Olarak Kullanılan Ham Verilerin Bir Kısmı

X_1	X_2	X_3	X_4	Y
203.60	2187.93	2954.78	4045.74	19.07
210.32	2260.18	3052.36	4179.35	22.68
207.01	2224.63	3004.34	4113.59	23.06
215.08	2311.36	3121.46	4273.97	22.15
215.58	2316.78	3128.78	4283.99	21.55
217.31	2335.31	3153.82	4318.27	26.72
201.60	2166.52	2925.86	4006.14	24.07
224.01	2407.30	3251.04	4451.38	26.70
183.40	1970.94	2661.74	3644.50	26.84
157.37	1509.64	2049.98	3150.84	21.96
192.22	2366.24	2351.52	3580.26	19.54
200.23	1935.85	2565.82	3969.89	20.81
172.66	1672.68	2050.66	3213.33	27.52
189.02	1561.01	2355.06	3693.88	20.60
186.09	1499.54	2375.91	3669.17	18.61
195.17	1608.18	2423.74	3784.14	20.34
162.15	1456.94	2257.83	3304.87	23.91
174.59	1710.58	2090.26	3148.28	23.12
283.61	3144.17	4404.77	6418.30	27.12
337.27	3739.09	5238.21	7632.72	20.54

Ek-2. Standardize Edilmiş Verilerin Bir Kısmı

X_1	X_2	X_3	X_4	Y
-1.01664	-0.85979	-0.75118	-0.97783	-0.08741
-0.93063	-0.7823	-0.6707	-0.90206	0.6809
-0.973	-0.82042	-0.7103	-0.93935	0.76177
-0.8697	-0.7274	-0.61371	-0.84839	0.5681
-0.86331	-0.72159	-0.60767	-0.84271	0.4404
-0.84116	-0.70171	-0.58702	-0.82327	1.54072
-1.04224	-0.88275	-0.77503	-1.00029	0.97673
-0.7554	-0.6245	-0.50684	-0.74777	1.53646
-1.2752	-1.09252	-0.99287	-1.2054	1.56626
-1.60837	-1.5873	-1.49742	-1.48539	0.52766
-1.1623	-0.66854	-1.24872	-1.24184	0.01262
-1.05978	-1.13016	-1.07198	-1.02085	0.28291
-1.41266	-1.41243	-1.49686	-1.44995	1.71098
-1.20326	-1.5322	-1.2458	-1.1774	0.23822
-1.24076	-1.59813	-1.2286	-1.19141	-0.18531
-1.12454	-1.48161	-1.18916	-1.1262	0.18288
-1.54719	-1.64382	-1.32599	-1.39803	0.94267
-1.38796	-1.37178	-1.4642	-1.48684	0.77454
0.00745	0.16584	0.44471	0.36779	1.62585
0.69428	0.80393	1.13209	1.05657	0.22545

Ek-3 Standardize Edilmiş Verilerin Bir Kısmına Ait "X" Matrisi

1	-1.01664	-0.85979	-0.75118	-0.97783
1	-0.93063	-0.7823	-0.6707	-0.90206
1	-0.973	-0.82042	-0.7103	-0.93935
1	-0.8697	-0.7274	-0.61371	-0.84839
1	-0.86331	-0.72159	-0.60767	-0.84271
1	-0.84116	-0.70171	-0.58702	-0.82327
1	-1.04224	-0.88275	-0.77503	-1.00029
1	-0.7554	-0.6245	-0.50684	-0.74777
1	-1.2752	-1.09252	-0.99287	-1.2054
1	-1.60837	-1.5873	-1.49742	-1.48539
1	-1.1623	-0.66854	-1.24872	-1.24184
1	-1.05978	-1.13016	-1.07198	-1.02085
1	-1.41266	-1.41243	-1.49686	-1.44995
1	-1.20326	-1.5322	-1.2458	-1.1774
1	-1.24076	-1.59813	-1.2286	-1.19141
1	-1.12454	-1.48161	-1.18916	-1.1262
1	-1.54719	-1.64382	-1.32599	-1.39803
1	-1.38796	-1.37178	-1.4642	-1.48684
1	0.00745	0.16584	0.44471	0.36779
1	0.69428	0.80393	1.13209	1.05657

Ek-4 Ek-3'te Belirtilen Matrisin Transpozu (X')

1.00000 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000
-1.01664 -0.93063 -0.97300 -0.86970 -0.86331 -0.84116 -1.04224 -0.75540
-0.85979 -0.78230 -0.82042 -0.72740 -0.72159 -0.70171 -0.88275 -0.62450
-0.75118 -0.67070 -0.71030 -0.61371 -0.60767 -0.58702 -0.77503 -0.50684
-0.97783 -0.90206 -0.93935 -0.84839 -0.84271 -0.82327 -1.00029 -0.74777

1.00000 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000 1.00000
-1.27520 -1.60837 -1.16230 -1.05978 -1.41266 -1.20326 -1.24076 -1.12454
-1.09252 -1.58730 -0.66854 -1.13016 -1.41243 -1.53220 -1.59813 -1.48161
-0.99287 -1.49742 -1.24872 -1.07198 -1.49686 -1.24580 -1.22860 -1.18916
-1.20540 -1.48539 -1.24184 -1.02085 -1.44995 -1.17740 -1.19141 -1.12620

1.00000 1.00000 1.00000 1.00000
-1.54719 -1.38796 0.00745 0.69428
-1.64382 -1.37178 0.16584 0.80393
-1.32599 -1.46420 0.44471 1.13209

Ek-5 Standardize Edilmiş Verilerin Bir Kısmına Ait X'X Matrisi

20.0000	-19.6124	-18.6692	-16.4072	-18.4406
-19.6124	24.4494	24.0535	22.4320	24.1370
-18.6692	24.0535	24.5226	22.4792	23.8013
-16.4072	22.4320	22.4792	21.5009	22.5738
-18.4406	24.1370	23.8013	22.5738	24.1566

Ek-6 Standardize Edilmiş Verilerin Bir Kısmına Ait Y Vektörü

-0.08741
0.6809
0.76177
0.5681
0.4404
1.54072
0.97673
1.53646
1.56626
0.52766
0.01262
0.28291
1.71098
0.23822
-0.18531
0.18288
0.94267
0.77454
1.62585
0.22545

Ek-7 Standardize Edilmiş Verilerin Bir Kısmına Ait (X'Y) Matrisi

14.3224
-13.8396
-12.5109
-10.9063
-12.8478

Ek-8 Standardize Edilmiş Verilerin Bir Kısmına Ait $(X'X)^{-1}(X'Y)$ Matrisi (En Küçük Kareler Yöntemine Göre Elde Edilmiş Olan Parametre Tahminleri)

0.07849
-2.62591
0.51975
1.44630
0.28819

Ek-9 Standardize Edilmiş Verilerin Bir Kısmına Ait k*=0,2 Değeri İçin Elde Edilmiş Olan $(X'X+k^*I)^{-1}(X'Y)$ Matrisi

0.504467
-0.856318
0.314264
0.740612
-0.290457

Ek-10 Standardize Edilmiş Verilerin Bir Kısmına Ait EKK Yöntemi İle Elde Edilen Regresyon Denklemi

$$\hat{Y} = 0.078 - 2.63X_1 + 0.52X_2 + 1.45X_3 + 0.29X_4$$

Ek-11 M-Tahmin Edici Yönteminde Standardize Edilmiş Verilerin Bir Kısmına Ait Hata Değerleri

e_i
-0,02154
0,030546
-0,03134
-0,05975
-0,14421
-0,14626
-0,20375
-0,28305
0,313804
-0,35294
-0,39257
-0,40737
0,491822
-0,5685
-0,60381
0,705286
0,729718
0,748269
-1,0193
1,24308

Ek-12 Ek-11' de Verilen Hata Değerlerinin Mutlak Değerleri Alındıktan Sonra Elde Edilmiş Olan Medyan ve $\hat{s}$ Değeri

Medyan=0.372755

$\hat{s} = (1.483)(0.372755) = 0.55279$

$1.5\hat{s} = 0.829$

Ek-13 -0.829 ve 0.829 Değerleri İle Hata Değerleri Karşılaştırıldıktan Sonra Elde Edilmiş Olan Yeni Hata Değerleri

e_i^*

-0,02154
0,030546
-0,03134
-0,05975
-0,14421
-0,14626
-0,20375
-0,28305
0,313804
-0,35294
-0,39257
-0,40737
0,491822
-0,5685
-0,60381
0,705286
0,729718
0,748269
-0,829
0,829

Ek-14 Standardize Edilmiş Verilerin Bir Kısmına Ait $\hat{Y}$ Değerleri

$\hat{Y}$
0,9642091
0,9461836
0,214223
0,2979738
0,4271207
0,9080251
0,8846485
0,8511504
0,4607356
0,880595
0,618023
0,8477711
1,0744381
0,3831923
0,6164306
0,8354343
0,8961319
0,7881907
0,9318907
0,4678997

Ek-15 $\hat{Y}$ ile Yeni Hata Değerleri Toplamından Elde Edilen Y_i^* Değerleri

Y_i^*
0,942669
0,97673
0,182883
0,238224
0,282911
0,761765
0,680899
0,5681
0,77454
0,527655
0,225453
0,440401
1,56626
-0,18531
0,012621
1,54072
1,62585
1,53646
0,102891
1,2969

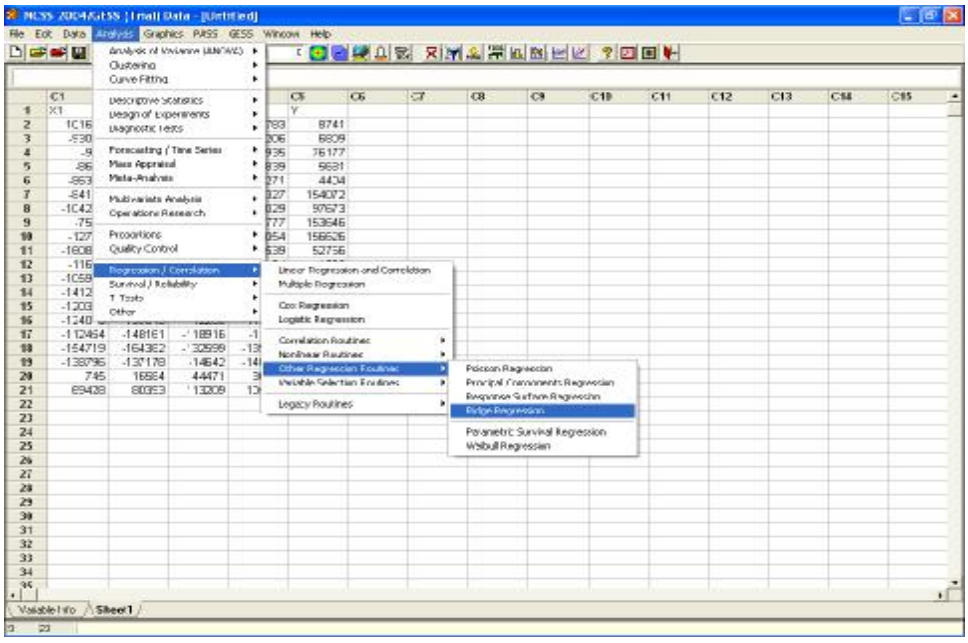
Ek-16 Y_i^* Değerleri Bağımlı Değişken Olarak Alındığında EKK Yöntemi İle Elde Edilmiş Olan Regresyon Denklemi

$$\hat{Y} = 0.640 - 0.055X_1 - 0.135X_2 - 0.204X_3 - 0.307X_4$$

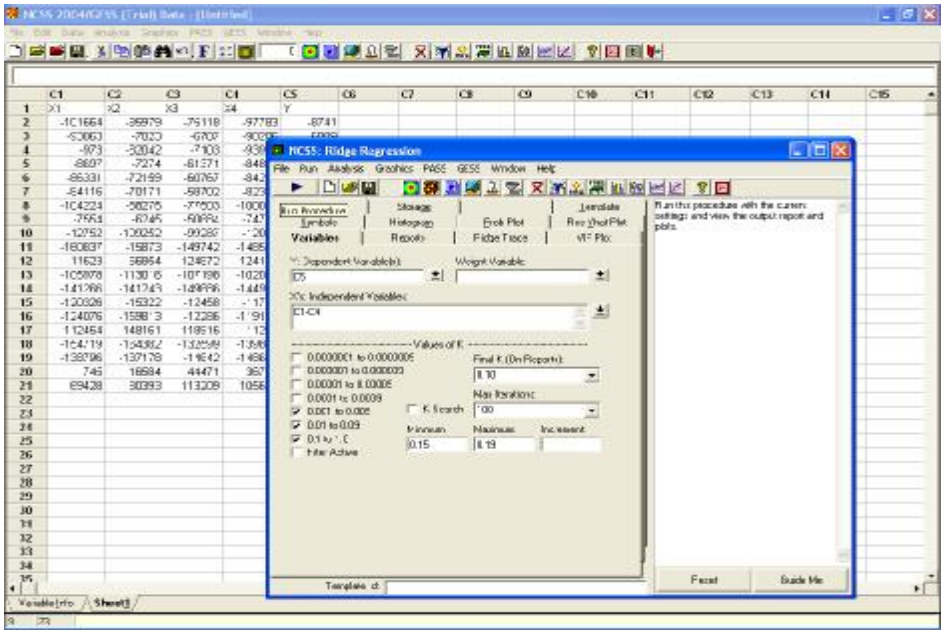
Ek-17 Aykırı Değer Olduğu Düşünülen Gözlemlerin Ortalama Değişim Aykırı Modellemesi Yöntemi Kullanılarak Elde Edilmiş Olan Önem Seviyeleri

Değişkenler	P
U19	0.009
U33	0.10
U51	0.31
U57	0.000
U95	0.001

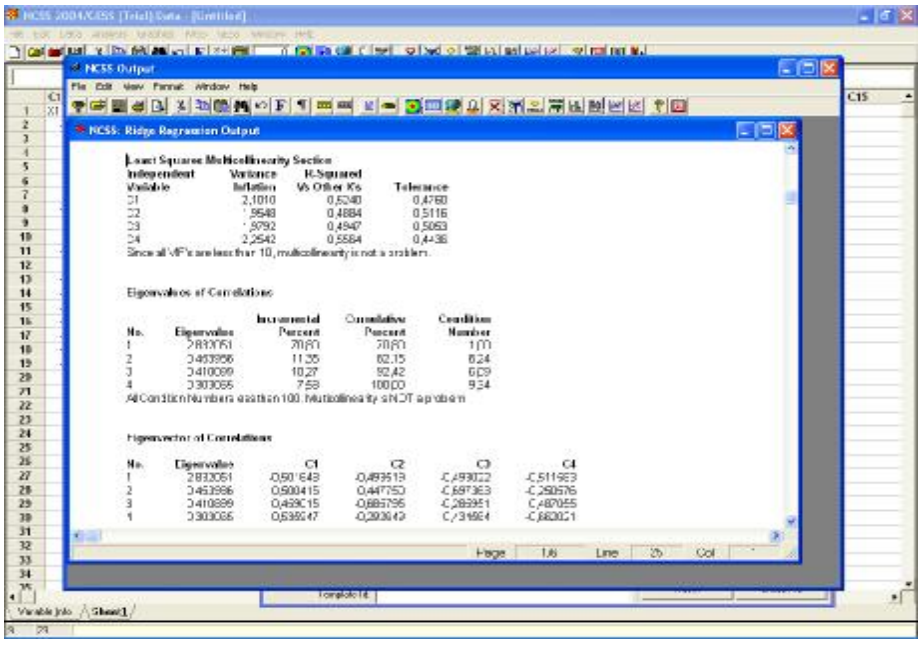
Ek-18 NCSS Programında Ridge Parametresi (k*) ‘ nin Belirlenmesi



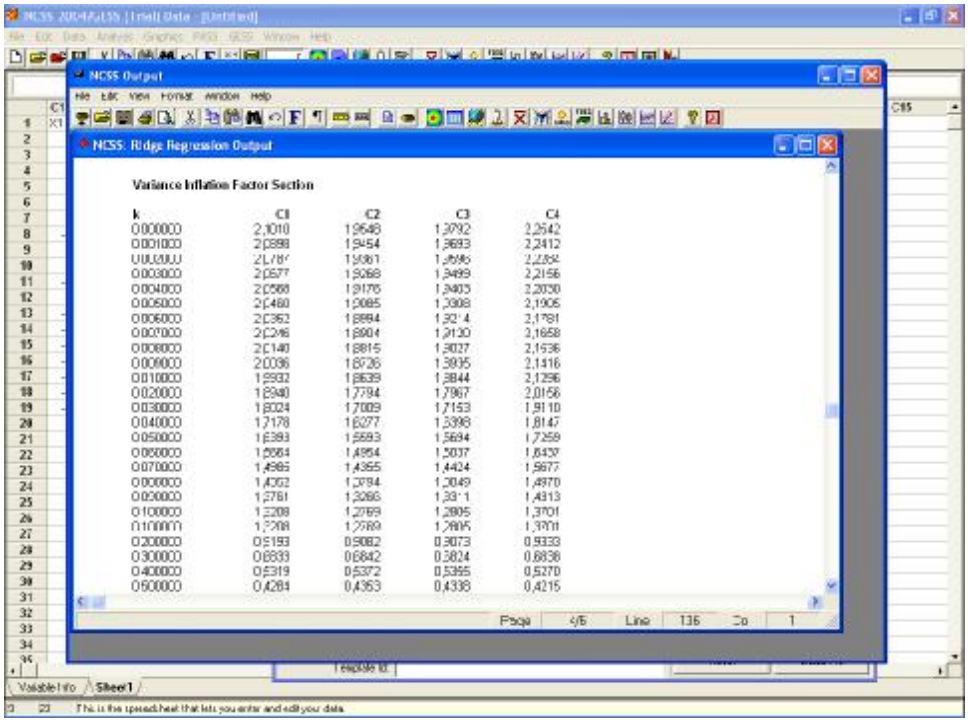
Şekil 1. NSCC Paket Programında Ridge Regresyon Menüsünün Açılması



Şekil 2. Değişkenlerin Tanımlanması



Şekil 3. Analiz Sonuçları



Şekil 4. Ridge Parametresi ve VIF Değerleri

Ek-19 Ridge Parametresi k* Belirlendikten Sonra MINITAB Paket Programında Ridge Regresyon Yöntemine Göre Parametre Tahminlerinin Elde Edilmesinde Kullanılan Syntax Komutları

```
MTB > read 20 5 m1
MTB> read 20 1 m2
MTB > transpose m1 m3
MTB > multiply m3 m1 m4
MTB > read 5 5 m6
MTB > add m6 m4 m7
MTB > invert m7 m8
MTB > multiply m3 m2 m9
MTB > multiply m8 m9 m10
MTB > print m10
```