

**ÇOKLU DOĞRUSAL REGRESYON MODELLERİNDE ÇOKLU
KUŞKULU GÖZLEMLERİN ANALİZİ**

**ANALYSIS OF MULTIPLE INFLUENTIAL OBSERVATION ON
THE MULTIPLE LINEAR REGRESSION MODELS**

ÖMER LÜSER

Hacettepe Üniversitesi

Lisansüstü Eğitim – Öğretim ve Sınav Yönetmeliğinin

İSTATİSTİK Anabilim Dalı İçin Öngördüğü

YÜKSEK LİSANS TEZİ

olarak hazırlanmıştır.

2010

Fen Bilimleri Enstitüsü Müdürlüğü'ne,

Bu çalışma jürimiz tarafından **İSTATİSTİK ANABİLİM DALI 'nda YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Başkan :.....
Prof. Dr. Aydın ERAR

Üye (Danışman) :.....
Prof. Dr. Öniz TOKTAMIŞ

Üye :.....
Doç. Dr. Meral ÇETİN

ONAY

Bu tez/...../..... tarihinde Enstitü Yönetim Kurulunca kabul edilmiştir.

Prof. Dr. Adil DENİZLİ
Fen Bilimleri Enstitüsü Müdürü

Bi'taneme,

ÇOKLU DOĞRUSAL REGRESYON MODELLERİNDE ÇOKLU KUŞKULU GÖZLEMLERİN ANALİZİ

Ömer Lüser

ÖZ

Bu çalışmanın amacı, çoklu doğrusal regresyon modellerinde çoklu kuşkulu gözlemlerin belirlenmesi için kullanılan yöntemlerin tanıtılması ve gerçek bir veri kümesi üzerinde uygulama yapılarak çoklu kuşkulu gözlemlerin etkilerinin incelenmesidir.

Bu çalışmada çoklu kuşkulu gözlem belirlemede kullanılan yöntemlerden, artıklara dayalı yöntem, genelleştirilmiş Cook uzaklığı, genelleştirilmiş Welsch uzaklığı, genelleştirilmiş Welsch-Kuh uzaklığı, genelleştirilmiş düzeltilmiş Cook uzaklığı, genelleştirilmiş Andrews-Pregibon istatistiği, genelleştirilmiş varyans oranı, genelleştirilmiş Cook-Weisberg istatistiği ve rankı etkileyen gözlemlere dayanan bir yöntem incelenmiştir.

Uygulama verisi olarak ülkemizdeki illere ait asayiş suç oranları alınmış ve *Braithwaite*'ın 2005 yılında önerdiği "Birleştirici Utandırma Teorisi" yardımı ile bir toplumda meydana gelen suçu ve suç oranlarını açıklamak için regresyon çözümü yapılmış ve çoklu kuşkulu gözlemlerin etkileri incelenmiştir.

Anahtar Kelimeler: Çoklu kuşkulu gözlemler, rankı etkileyen gözlemler, gizleme ve sürüklenme etkisi, birlikte etki

Danışman: Prof. Dr. Öviz TOKTAMIŞ, Hacettepe Üniversitesi, İstatistik Bölümü, Uygulamalı İstatistik Anabilim Dalı

ANALYSIS OF MULTIPLE INFLUENTIAL OBSERVATION ON THE MULTIPLE LINEAR REGRESSION MODELS

Ömer Lüser

ABSTRACT

The aim of this thesis is to introduce methods using to determine multiple influential observations in multiple linear regression models and examine the effects of multiple influential observations with an application on a real data set.

In this study, a measure based on residuals, the generalized Cook distance, the generalized Welsch distance, the generalized Welsch-Kuh distance, the generalized, modified Cook distance, the generalized Andrews-Pregibon statistic, the generalized variance ratio, the generalized Cook-Weisberg statistic and a measure based on rank-influential points which used to determine the multiple influential observations are examined.

For application data, the ratio of public order crimes of 81 province in our country are taken and the regression analysis were made with help of “Reintegrative Shaming Theory” is proposed by Braithwaite in 2005 for explaining a crime and crime ratio occurring in the community and the multiple influential observations’ influences are investigated.

Keywords: Multiple influential observations, collinearity-influential points, masking and swamping effect, joint effect

Advisor: Prof.Dr. Öñiz TOKTAMIŞ, Hacettepe University, Department of Statistic, Applied Statistic Section

TEŐEKKÜR

Çalıőmamın her aőamasında bilgisi ve manevi desteęi ile her zaman yanımda olan ve katkı ve eleőtirileri ile çalıőmama yön veren deęerli danıőmanım Sayın Prof.Dr. Öviz TOKTAMIŐ'a, çalıőmanın gerçekteőmesi için gerekli ortamı hazırlayan H.Ü. İstatistik Bölümü Baőkanı Sayın Prof.Dr. Süleyman GÜNAY'a en içten dileklerle teşekkür ederim.

Çalıőmanın ortaya çıkmasında fırsat tanıyan ve iş yaşamımda deęerli bilgileri ile yönlendiren Emniyet Genel Müdürlüęü Asayiş Dairesi Baőkanı Sayın Hüseyin ÖZALP'a, Asayiş Dairesi Baőkan Yardımcısı Sayın Esat AKTAŐ'a, Suç Analizi Őube Müdürü Sayın Taőkın ERDİN'e, Büro Amiri Önder ÇELİK'e ve dięer bütün çalıőma arkadaşlarıma teşekkür ederim.

Çalıőmamın her aőamasında manevi desteęi ile bana güç veren ve bilgisini esirgemeyen deęerli arkadaşım Arő. Gör. Hande KONŐUK'a teşekkür ederim.

Tez çalıőmam süresince gösterdikleri sevgi ve sabırla her zaman yanımda olan anneme, babama ve kardeőlerime içtenlikle teşekkür ederim.

İÇİNDEKİLER DİZİNİ

Sayfa

ÖZ.....	i
ABSTRACT	ii
TEŞEKKÜR	iii
İÇİNDEKİLER DİZİNİ.....	iv
ŞEKİLLER DİZİNİ	vii
ÇİZELGELER DİZİNİ.....	viii
1. GİRİŞ.....	1
2. DOĞRUSAL REGRESYONDA KUŞKULU GÖZLEMLERİN ARAŞTIRILMASI... 4	4
2.1. Giriş	4
2.2. Regresyon Çözümlemesi.....	4
2.2.1. Problemin tanımlanması.....	6
2.2.1.1. Değişken seçimi.....	7
2.2.1.2. Modelin matematiksel yapısının belirlenmesi.....	8
2.2.1.3. Parametrelerin kestirim yöntemlerinin seçilmesi	9
2.2.1.4. Varsayımların belirlenmesi.....	10
2.2.2. Modelin oluşturulması	10
2.2.3. Varsayımların geçerliliğinin incelenmesi.....	12
2.2.4. Uyumun değerlendirilmesi	12
2.3. Kuşkulu Gözlemler.....	14
2.3.1. Aykırı gözlemler.....	15
2.3.2. Etkili gözlemler	19
2.4. Tekli Kuşkulu Gözlem Belirleme Yöntemleri	20
2.4.1. Artıklara Dayalı Yöntemler.....	24

2.4.2.	Gözlem uzaklığı matrisine dayalı yöntem.....	25
2.4.3.	Güven elipsoidinin merkezine dayalı yöntemler	28
2.4.3.1.	Cook uzaklığı	28
2.4.3.2.	Welsch uzaklığı.....	29
2.4.3.3.	Welsch-Kuh uzaklığı	29
2.4.3.4.	Düzeltilmiş Cook Uzaklığı	29
2.4.4.	Güven elipsoidinin hacmine dayalı yöntemler	30
2.4.4.1.	Andrews-Pregibon istatistiği.....	30
2.4.4.2.	Varyans oranı	30
2.4.4.3.	Cook-Weisberg istatistiği	31
3.	ÇOKLU KUŞKULU GÖZLEM BELİRLEME YÖNTEMLERİ	32
3.1.	Giriş	32
3.2.	Çoklu Kuşkulu Gözlem Belirlemede Yaşanan Sıkıntılar.....	34
3.3.	Artıklara Dayalı Yöntemler	36
3.4.	Güven Elipsoidinin Merkezindeki Değişime Dayalı Yöntemler.....	37
3.4.1.	Genelleştirilmiş Cook uzaklığı	37
3.4.2.	Genelleştirilmiş Welsch uzaklığı	37
3.4.3.	Genelleştirilmiş Welsch-Kuh uzaklığı	38
3.4.4.	Genelleştirilmiş düzeltilmiş Cook uzaklığı.....	38
3.5.	Güven Elipsoidinin Hacmine Dayalı Yöntemler.....	39
3.5.1.	Genelleştirilmiş Andrews-Pregibon istatistiği.....	39
3.5.2.	Genelleştirilmiş varyans oranı	39
3.5.3.	Genelleştirilmiş Cook-Weisberg istatistiği.....	40
3.6.	Rankı Etkileyen Gözlemlere Dayalı Yöntem	40
4.	UYGULAMA	45

4.1. Giriş	45
4.2. Modelin Oluřturulması	45
4.3. Tekli Kuřkulu Gzlemlerin Belirlenmesi	47
4.4. oklu Kuřkulu Gzlemlerin Belirlenmesi	53
4.4.1. İekli kuřkulu gzlemlerin belirlenmesi ve modele etkileri	53
4.4.2. Geri eleme yntemi ile oklu kuřkulu gzlemlerin belirlenmesi	56
4.4.3. Rankı etkileyen gzlemlere dayalı yntem ile oklu kuřkulu gzlemlerin belirlenmesi	59
5. SONULAR	65
KAYNAKLAR	67

ŞEKİLLER DİZİNİ

	<u>Sayfa</u>
Şekil 2.1 Regresyon Çözümleme Diyagramı	6
Şekil 2.2 Anscombe'nin dört farklı veri kümesinin saçılım grafikleri ve kestirim doğruları (Alpar, 2003).....	14
Şekil 2.3 Bağımlı değişken yönünde aykırı gözlem	16
Şekil 2.4 Açıklayıcı değişken yönünde aykırı gözlem	17
Şekil 2.5 Hem bağımlı hem de açıklayıcı değişken yönünde aykırı gözlemler.....	18
Şekil 2.6 Etkili gözlem.....	19
Şekil 2.7 Press artıklarının klasik artıklar ile grafiği.....	25
Şekil 2.8 Williams Grafiği	26
Şekil 2.9 Pregibon grafiği.....	27
Şekil 3.1 Birlikte etkili gözlemler	32
Şekil 3.2 Ayrı ayrı etkili fakat birlikte etkili olmayan gözlemler	33
Şekil 3.3 Çoklu bağlantıya sebep olan ve gizleyen gözlemler	41
Şekil 4.1 Press Artığı ile Klasik Artık grafiği	51
Şekil 4.2 Williams Grafiği	52
Şekil 4.3 Pregibon Grafiği	52
Şekil 4.4 Rankı Etkileyen Gözlemlere Dayalı Yönteme Ait Kod Penceresi	59

ÇİZELGELER DİZİNİ

Sayfa

Çizelge 2.1 Regresyon analizindeki çeşitli durumlar.....	9
Çizelge 4.1 Bağlanım Katsayıları ve İlgili İstatistikler.....	46
Çizelge 4.2 Tekli Kuşku Gözlem Tespit Yöntemlerine İlişkin Elde Edilen Değerler	47
Çizelge 4.2 (devam) Tekli Kuşku Gözlem Tespit Yöntemlerine İlişkin Değerler.....	48
Çizelge 4.3 Tekli Kuşku Gözlem Belirleme Yöntemlerine Ait Kritik Değerler	49
Çizelge 4.4 Tekli Kuşku Gözlem Belirleme Yöntemlerine Göre Belirlenen Kuşku Gözlemler	50
Çizelge 4.5 En Yüksek Genelleştirilmiş Cook Uzaklığı Alan İkili Gözlemler.....	53
Çizelge 4.6 {6} Nolu Gözlem Çıkarıldığındaki Model Parametrelerine Ait Bilgiler..	55
Çizelge 4.7 {34} Nolu Gözlem Çıkarıldığındaki Model Parametrelerine Ait Bilgiler	55
Çizelge 4.8 {6, 34} Nolu Gözlemler Çıkarıldığındaki Model Parametrelerine Ait Bilgiler.....	56
Çizelge 4.9 Geri Eleme Yöntemi İle Hesaplanan Çoklu Kuşku Gözlemlerin Genelleştirilmiş Cook Uzaklığı Değerleri.....	57
Çizelge 4.10 {6,34,41,59} altkümesi çıkarıldığındaki model parametrelerine ait bilgiler	58
Çizelge 4.11 {6, 34, 41, 35, 16} Nolu Gözlemler Çıkarıldığındaki Model Parametrelerine Ait Bilgiler	60
Çizelge 4.12 Rankı Etkileyen Gözlemlere Dayalı Yöntem İle Elde Edilen Q Matrisi .	61
Çizelge 4.12 (devam) Rankı Etkileyen Gözlemlere Dayalı Yöntem İle Elde Edilen Q Matrisi.....	62
Çizelge 4.13 Rankı Etkileyen Gözlemlere Dayalı Yöntem İle Elde Edilen Q2 Matrisi	63
Çizelge 4.13 (devam) Rankı Etkileyen Gözlemlere Dayalı Yöntem İle Elde Edilen Q2 Matrisi.....	64

1. GİRİŞ

Regresyon çözümlenmesi, değişkenler arasındaki ilişkinin araştırılmasında araştırmacılar tarafından en sık kullanılan yöntemlerden biridir. Analiz sırasında en çok karşılaşılan sorunlardan biri, toplanan veri içinde kuşkulu gözlemlerin olmasıdır. Kuşkulu gözlemler aykırı gözlem ya da etkili gözlem olabilir. Analiz sonucunda yorumların doğru olarak yapılabilmesi için kuşkulu gözlemlerin incelenmesi gerekmektedir. Bu aşamada şu sorular akla gelmektedir:

- Kuşkulu gözlemler nedir, nasıl incelenmelidir?
- Nasıl etkileri vardır?
- Sadece regresyon parametre kestirimlerini, varyanslarını ve sabit değerini mi etkilemektedirler?
- Sadece tekli kuşkulu gözlemlerin tespiti, kuşkulu gözlem incelenmesi için yeterli midir?
- Tekli kuşkulu gözlemlerin tespiti hangi alanlarda yetersiz kalırlar?
- Eğer yetersiz kalıyor ise hangi yollar izlenmelidir?

Doğrusal regresyon modelinde kuşkulu gözlem belirleme yöntemleri literatürde ayrıntılı olarak incelenmiştir. Bu konudaki ilk çalışma 1977 yılında Cook tarafından yapılmıştır. Bu çalışmadan sonra Cook'un kriterinden yararlanılarak Cook uzaklığı, Andrews Pregibon ölçüsü, Hadi ölçüsü, Welsch ve Kuh uzaklığı, Welsch uzaklığı, Cook ve Weisberg olabilirlik uzaklığı (likelihood distance) gibi ölçüler geliştirilmiştir. Hadi 1992 yılında doğrusal regresyonda etkili gözlemleri araştırmak için grafiksel bir düzenleme ve yeni bir etki ölçüsü tanıtmıştır. Meloun ve Militky grafiksel yöntemler ile tekli kuşkulu gözlemlerin etkilerini araştırmak için çalışmalar yapmış, aykırı gözlemleri belirlemek için kullanılan grafikleri, olabilirlik uzaklığına dayanan etki tiplerinin belirlenmesi için elde edilen grafikler ile birleştirmiştir. Bu grafiksel yöntemlerin tümü kuşkulu gözlemlerin hızlı bir şekilde belirlenmesi için uygundur; ancak gizleme ve sürüklenme problemlerini çözmekte genellikle yetersizdirler (Meloun & Militky, 2001). Imon, çoklu gözlem çıkarmaya dayanan DFFITTS'in genelleştirilmiş bir versiyonunu geliştirmiş ve bunu kullanarak çoklu etkili

gözlemleri belirlemek için yeni bir teknik önermiştir (Imon, 2005). Ullah ve Pasha tarafından da, geliştirilen etki ölçüleri tartışılmıştır (Ullah & Pasha, 2009).

Tekli kuşkulu gözlem tespit yöntemleri ile etkileri belirlenemeyen, yani ayrı ayrı etkili olmayan, fakat birlikte etkili olan gözlemlerin olduğu durumlar var olabilir. Böyle durumda bazen bir gözlemin varlığı diğer gözlemin etkisini gizlemektedir. Bu etkiye gizleme etkisi (masking effect) denir. Ayrıca tekli kuşkulu gözlem tespit yöntemleri ile ayrı ayrı etkili oldukları belirlenen, ancak birlikte olduklarında etkili olmayan gözlemlerin bulunduğu durumlar da vardır. Böyle durumlarda bazen bir gözlem diğer gözlemin kuşkulu gözlem olarak belirlenmesine sebep olmaktadır. Bu etkiye sürüklenme etkisi (swamping effect) denir. Böyle durumlarda tekli kuşkulu gözlem tespit yöntemleri yetersiz kalmakta ve çoklu kuşkulu gözlem tespiti yöntemlerine ihtiyaç duyulmaktadır (Aşıkçıl, 2006).

Çoklu kuşkulu gözlemler, genelde tekli kuşkulu gözlemlerin etkisinden daha fazla etkilidir ve uygulayıcılar tarafından sıklıkla gözden kaçırılmaktadır. Çünkü çoklu kuşkulu gözlemlerin belirlenmesi, tekli kuşkulu gözlemlerin belirlenmesinden daha zordur. Çoklu kuşkulu gözlemlerin tespitinin doğasından gelen üç sorunu vardır. Birincisi, birlikte etkili gözlemlerden oluşan altkümenin büyüklüğünün (m) nasıl belirleneceğidir. İkincisi, m büyüklüğü belirlendikten sonra m büyüklüğündeki her alt küme için, başka bir deyişle $n!/[m!(n-m)!]$ kadar olası alt küme için, yapılacak hesaplamalardaki zorluklardır. Üçüncüsü ise, grafiksel yöntemler ile çoklu kuşkulu gözlemlerin gösteriminde yaşanan sıkıntılardır.

Tekli kuşkulu gözlem belirlemede kullanılan istatistikler geliştirilerek, çoklu kuşkulu gözlemleri belirlemek için ölçütler elde edilebilir. Bu çalışmanın ikinci bölümünde doğrusal regresyon modelinin kurulum aşamaları ayrıntılı olarak verilecek ve tekli kuşkulu gözlem belirleme yöntemleri tanıtılacaktır. Üçüncü bölümde çoklu kuşkulu gözlem belirlemede yaşanan sıkıntılar ayrıntılı olarak anlatılacak ve çoklu kuşkulu gözlem belirlemede kullanılan yöntemlerden, artıklara dayalı yöntem, geliştirilmiş Cook uzaklığı, geliştirilmiş Welsch uzaklığı, geliştirilmiş Welsch-Kuh uzaklığı, geliştirilmiş düzeltilmiş Cook uzaklığı, geliştirilmiş Andrews-Pregibon istatistiği, geliştirilmiş varyans oranı, geliştirilmiş Cook-Weisberg istatistiği ve rankı etkileyen gözlemlere dayanan

bir yöntem tanıtılacaktır. Dördüncü bölümde uygulama verisi olarak ülkemizdeki illere ait asayiş suç oranları alınarak ve *Braithwaite*'ın 2005 yılında önerdiği “Birleştirici Utandırma Teorisi” yardımı ile bir toplumda meydana gelen suçu ve suç oranlarını açıklamak için regresyon çözümlenmesi yapılacak ve çoklu kuşkulu gözlemlerin etkileri incelenecektir. Beşinci bölümde ise, çalışma kapsamında elde edilen sonuçlar tartışılacaktır.

2. DOĞRUSAL REGRESYONDA KUŞKULU GÖZLEMLERİN ARAŞTIRILMASI

2.1. Giriş

Regresyon çözümlenmesi, değişkenler arasındaki ilişkinin araştırılmasında araştırmacılar tarafından en sık kullanılan yöntemlerden biridir. Bu analiz yönteminin uygulanması sürecinde izlenen yol ve yöntemler literatürde ayrıntılı olarak verilmiştir. Analiz sırasında en çok karşılaşılan sorunlardan biri, toplanan veri içinde aykırı gözlemlerin olmasıdır. Analiz sonucunda yorumların doğru olarak yapılabilmesi için verinin yapısı doğru olarak incelenmeli ve varsayım bozulmaları, aykırı değerler gibi sorunlar çözümlenmelidir.

Bu bölümde regresyon çözümlenmesinde kuşkulu gözlemlerin incelenmesine geçmeden önce, Alt Bölüm 2.2’de regresyon çözümlenmesi, amacı ve regresyon çözümlenmesinin aşamaları, Alt Bölüm 2.3’te kuşkulu gözlemlerin incelenmesi adımı ayrıntılı olarak anlatılacak olup, Alt Bölüm 2.4’te tekli kuşkulu gözlemlerin belirlenmesinde kullanılan yöntemler incelenecektir.

2.2. Regresyon Çözümlenmesi

Regresyon çözümlenmesi, değişkenler arasındaki karşılıklı ilişkiyi anlamaya yardım etmekte kullanılan veri çözümlenme yöntemleri kümesi olarak tanımlanabilir (Chatterjee & Hadi, 2006). Bu ilişki, bağımlı değişken (dependent variable) ile bir ya da daha fazla açıklayıcı değişkeni (explanatory or predictor variables) birbirine bağlayan bir model ya da bir denklem biçiminde ifade edilir.

Açıklayıcı değişkenler (explanatory variables), değişik kaynaklarda “bağımsız değişkenler” (independent variables) ya da “tahmin edici değişkenler” (predictor variables) olarak da adlandırılmaktadır. “Bağımsız değişkenler” ifadesi, yaygın kullanılmasına rağmen, tercih edilmemelidir. Çünkü uygulamada bu değişkenler nadiren kendi aralarında bağımsızdırlar (Chatterjee & Hadi, 2006).

Açıklanan ya da bağımlı değişken Y ile, açıklayıcı değişkenler X_1, X_2, \dots, X_k ile gösterilir. k , açıklayıcı değişkenlerin sayısını ifade etmektedir. Y ile X_1, X_2, \dots, X_k

arasındaki gerçek ilişki aşağıdaki regresyon modeli ile yaklaşık olarak elde edilebilir.

$$Y = f(X_1, X_2, \dots, X_k) + \varepsilon \quad (2.1)$$

Burada ε 'un yaklaşımdaki farklılığı temsil eden rastgele hata olduğu varsayılır. ε , veriye tam olarak uydurulan modeldeki başarısızlığı açıklar.

$f(X_1, X_2, \dots, X_k)$ fonksiyonu, Y ile X_1, X_2, \dots, X_k arasındaki ilişkiyi açıklamaktadır.

Doğrusal regresyon modeli,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon \quad (2.2)$$

biçiminde ifade edilir. $\beta_0, \beta_1, \dots, \beta_k$, regresyon parametreleri ya da katsayılarını göstermekte olup, veriden kestirilen bilinmeyen sabit değerlerdir.

Doğrusal regresyon modeli matris formunda,

$$Y = X\beta + \varepsilon \quad (2.3)$$

biçiminde yazılır. Burada, $p = k + 1$ olmak üzere,

Y : $(n \times 1)$ boyutlu bağımlı değişken vektörünü,

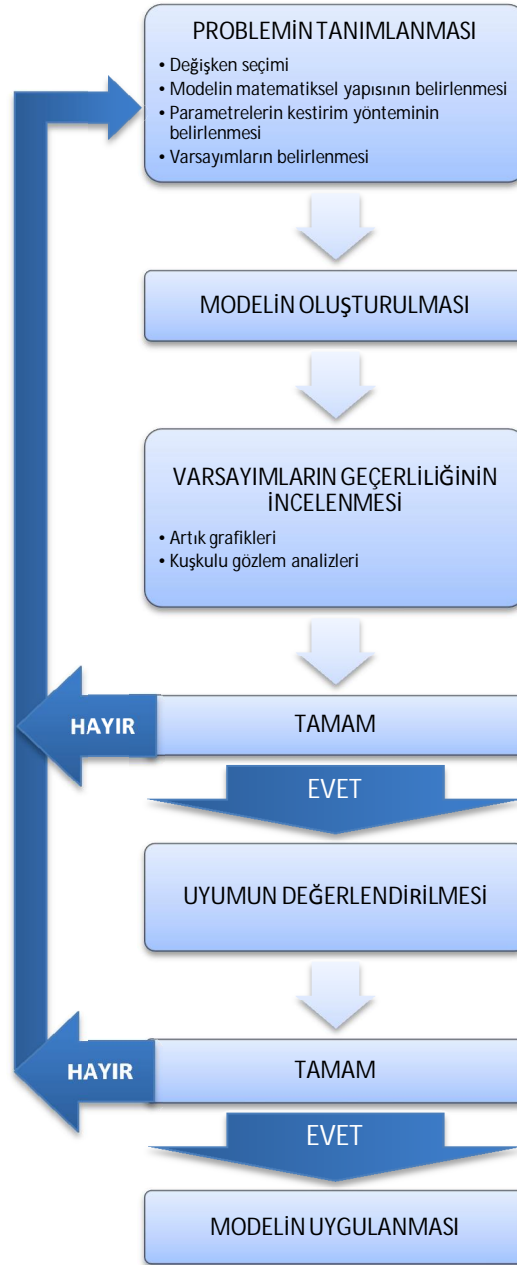
X : $(n \times p)$ boyutlu tasarım matrisini,

β : $(p \times 1)$ boyutlu parametreler vektörünü,

ε : $(n \times 1)$ boyutlu hata vektörünü,

temsil etmektedir.

Regresyon çözümlemesinde model oluşturmanın belirli aşamaları vardır. Bu aşamalar Şekil 2.1'de verilmiştir.



Şekil 2.1 Regresyon Çözümleme Diyagramı

Bu aşamalar aşağıdaki alt bölümlerde açıklanmaktadır.

2.2.1. Problemin tanımlanması

Regresyon çözümü, genellikle problemin açık ve kesin olarak ifade edilmesi ile başlar. Problemin tanımlanması regresyon analizinin ilk ve en önemli aşamasıdır. Çünkü yanlış tanımlanan bir problem ya da yanlış ifade edilen bir

soru, konu dışı deęişkenlerin ya da yanlış istatistiksel analiz yönteminin seçimine neden olup, çabaların boşa gitmesine yol açabilir. Örnek olarak; bir şirkette çalışan elemanlar ile ilgili araştırma yapılıyor olsun. Şirketin kayıtlarında, elemanların aldığı maaş, vasıf ve cinsiyet bilgileri var. Bu şirketin elemanları ile ilgili iki durum araştırılsın:

- a) Aynı vasıflı bay ve bayan elemanın aynı maaşı alıp almadığı, bayan elemanın aldığı maaşın bay elemanların aldığı maaştan daha az olup olmadığı.
- b) Aynı maaşı alan bay ve bayan elemanın aynı vasıflara sahip olup olmadıkları, bayan elemanın vasıflarının bay elemanın vasıflarından daha fazla olup olmadığı.

Eğer “Aynı vasıftaki bay ve bayan elemanlardan, ortalama olarak bayan elemanlara daha az maaş mı ödeniyor?” sorusu araştırılırsa; maaş, bağımlı deęişken, vasıf ve cinsiyet açıklayıcı deęişken olarak ele alınır.

Eğer “Aynı maaşı alan bay ve bayan elemanlardan, ortalama olarak bayan elemanlar daha fazla vasıflı mıdır?” sorusu araştırılırsa da; vasıf, bağımlı deęişken, maaş ve cinsiyet açıklayıcı deęişken olarak ifade edilir. Yani, araştırılan konuya ve öğrenilmek istenen soruya göre deęişkenlerin rolleri yer deęiştirmektedir.

Bu aşamada uygulanacak adımlar aşağıda verilmektedir

2.2.1.1. Deęişken seçimi

Problem tanımlandıktan sonra, bağımlı deęişkeni açıklayan deęişkenler seçilir. Modele sadece bağımlı deęişken ile ilgili açıklayıcı deęişkenler dahil edilmelidir. Bağımlı deęişkendeki deęişimi açıklamaya önemli derecede katkısı olan açıklayıcı deęişkenler uygun yöntemler kullanılarak seçilmelidir. Regresyon modelinin mümkün olan en az açıklayıcı deęişken ile en iyi tahmini yapabilecek şekilde oluşturulması esastır.

Örneğin, bir bölgedeki satılık ev fiyatları ile ilgili bir araştırmada bağımlı değişken ev fiyatları ve açıklayıcı değişkenler evin bulunduğu arsanın alanı, evin alanı, evin yaşı, evdeki oda sayısı, evdeki banyo sayısı, evin tipi, alınan vergi miktarı v.s. olabilir.

Değişkenler nicel (quantitative) ve nitel (qualitative) olarak ikiye ayrılırlar. Nicel değişkenler sayılabilir ve ölçülebilir değişkenlerdir. Örneğin ev fiyatı, evin oda sayısı, yaş, ödenen vergi nicel değişkenlerdir. Nitel değişkenler ise gruplanabilir ve sıralanabilen değişkenlerdir. Nitel değişkenlere de evin tipi, cinsiyet gibi gruplanabilen ve eğitim durumu gibi sıralanabilen değişkenler örnek verilebilir.

Regresyon çözümlemesinde bağımlı değişken nicel değişken olup, açıklayıcı değişkenler, nicel ya da nitel olabilir. Eğer bütün açıklayıcı değişkenler nitel değişken ise varyans analizi yöntemleri, eğer açıklayıcı değişkenlerden bazıları nicel, diğerleri nitel değişken ise, bu durumda kovaryans analizi yöntemleri kullanılır. Eğer bağımlı değişken nitel değişken olarak seçilmiş ise lojistik regresyon yöntemi kullanılır.

2.2.1.2. Modelin matematiksel yapısının belirlenmesi

Modelin şekli, başlangıçta araştırma konusunda uzman olan kişiler yardımıyla belirlenebilir. Model belirlendikten sonra toplanan verinin analizi sonucunda önerilen model, ya kabul edilecek ya da reddedilecektir. Hipotezin test edilmesi için modelin matematiksel yapısının belirlenmesine, bir başka deyişle, $f(X_1, X_2, \dots, X_k)$ fonksiyonunun matematiksel yapısının belirlenmesine ihtiyaç vardır.

Bu çalışmada Eşitlik (2.3)'te verilen doğrusal modeller ele alınacaktır. Bu doğrusal regresyon denklemi sadece bir tane açıklayıcı değişken içerirse buna basit doğrusal regresyon denklemi, eğer birden fazla açıklayıcı değişken var ise buna da çoklu doğrusal regresyon denklemi adı verilir.

Bir bağımlı değişkenin olduğu regresyon çözümlemesine tek değişkenli (univariate) regresyon, birden fazla bağımlı değişkenin olduğu regresyona ise çok

değişkenli (multivariate) regresyon denir. Regresyon analizindeki çeşitli durumlar Çizelge 2.1’de verilmektedir.

Çizelge 2.1 Regresyon analizindeki çeşitli durumlar

Durumlar	Koşulları
Tek değişkenli (Univariate)	Sadece bir nicel bağımlı değişken var.
Çok değişkenli (Multivariate)	Birden fazla nicel bağımlı değişken var.
Basit (Simple)	Sadece bir açıklayıcı değişken var.
Çoklu (Multiple)	Birden fazla açıklayıcı değişken var.
Doğrusal (Linear)	Denklem tüm parametrelere göre doğrusaldır.
Doğrusal Olmayan (Nonlinear)	Bağımlı değişken ile bazı açıklayıcı değişkenler arasındaki ilişki doğrusal değildir ya da parametrelere göre doğrusal olmayan şekilde yer alır.
Varyans Analizi	Tüm açıklayıcı değişkenler nitel değişkenlerdir.
Kovaryans Analizi	Bazı açıklayıcı değişkenler nitel, bazıları nicel değişkenlerdir.
Lojistik Regresyon	Bağımlı değişken niteldir.

2.2.1.3. Parametrelerin kestirim yöntemlerinin seçilmesi

Problemin tanımlanması, değişkenlerin belirlenmesi, modelin matematiksel yapısının belirlenmesi aşamalarından sonra bu aşamada toplanan veri yardımı ile modelin parametreleri, belirlenen kestirim yöntemine göre elde edilir. Parametre kestiriminde en yaygın kullanılan yöntem, en küçük kareler kestirim yöntemidir.

Belirli varsayımlar altında en küçük kareler kestirim yöntemi, istenen özelliklere sahip kestirimler oluşturur. Ancak bazı durumlarda diğer kestirim yöntemleri en küçük kareler kestirim yöntemine tercih edilebilir. Diğer kestirim yöntemlerinden bir kaç şunlardır; En Çok Olabilirlik Yöntemi, Ridge Yöntemi, Temel Bileşenler Yöntemi, Robust Yöntemi.

2.2.1.4. Varsayımların belirlenmesi

Regresyon analizi gibi bir istatistiksel yöntemin geçerliliği belirli varsayımlara dayanmaktadır. Varsayımlar genellikle hatalar, açıklayıcı değişkenler, veri ve modelin yapısı hakkındadır. Analizlerin ve analizlerden çıkarılan sonuçların doğruluğu bu varsayımların geçerliliğine bağlıdır. Regresyon analizi sonucunda oluşturulan model kullanılmadan önce, varsayımların geçerliliği kontrol edilmelidir.

Çoklu doğrusal regresyonda şu varsayımların sağlanması gerekir (Draper & Smith, 1966) :

- $E(\varepsilon_i) = 0, i = 1, 2, \dots, n,$
- $V(\varepsilon_1) = V(\varepsilon_2) = \dots = V(\varepsilon_n) = \sigma^2$ (Değişmez varyanslılık varsayımı),
- $i \neq j$ için $Cov(\varepsilon_i, \varepsilon_j) = 0$ (Hataların bağımsızlığı varsayımı),
- $\varepsilon \sim N(0, \sigma^2)$ (Normallik varsayımı),
- Eğer X_j bir rastlantı değişkeni ise $E(\varepsilon_i, x_{ij}) = 0,$
- X matrisinin tam ranklı olması, yani X 'in sütunları arasında tam ya da yaklaşık doğrusal bağımlılık (çoklu bağlantı) olmaması.

2.2.2. Modelin oluşturulması

Bu adımda, Alt Bölüm 2.2.1.3'te belirlenen kestirim yöntemi kullanılarak toplanan veri için model uydurulur. Başka bir deyişle, regresyon parametreleri tahmin edilir ve model denkleminde yerlerine konulur.

β 'nin kestiriminde en sık kullanılan kestirim yöntemlerinden birisi en küçük kareler kestirim yöntemidir. Bu yöntemde hata kareler toplamını veren ,

$$S(\beta) = (Y - X\beta)'(Y - X\beta) \quad (2.4)$$

fonksiyonunu en küçük yapan β değerleri,

$$\hat{\beta} = (XX)^{-1} X'Y \quad (2.5)$$

biçiminde elde edilir. Burada $\hat{\beta}$, $(p \times 1)$ boyutlu parametre kestirim vektörüdür. Bu değerler β 'nin en küçük kareler kestiricisi adını alır.

Kestirilen regresyon denklemini ise,

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \dots + \hat{\beta}_k X_k \quad (2.6)$$

biçimindedir ve

$$\begin{aligned} \hat{Y} &= X \hat{\beta} \\ &= X(XX)^{-1} X'Y \end{aligned} \quad (2.7)$$

biçiminde de ifade edilebilir.

$x_{i1}, x_{i2}, \dots, x_{ik}$, i . gözlem için açıklayıcı değişkenlerin aldıkları değerler olmak üzere Eşitlik (2.6) kullanılarak, bağımlı değişkenin kestirim değeri (estimated value) \hat{y}_i ile gösterilirse,

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_k x_{ik}, \quad i = 1, 2, \dots, n \quad (2.8)$$

biçiminde elde edilir. Açıklayıcı değişkenin veri setinde bulunmayan herhangi bir değeri için bağımlı değişken kestirilebilir. Veri setinde bulunmayan herhangi bir \tilde{x} değerine karşılık gelen kestirim değerine önkestirim değeri (prediction value) denir ve \tilde{y} ile gösterilir.

Kestirim ve önkestirim değerlerinin birbirinden farkı, kestirim değeri açıklayıcı değişkenin veri kümesinde bulunan n gözlemden herhangi birinin tahmini iken, önkestirim değeri, açıklayıcı değişkenin veri kümesinde olmayan bir değerinin tahmini olmasıdır. Veri kümesinden çok uzaktaki bir değer için önkestirimi genellikle önerilmemektedir. Bu durumdaki önkestirim değerine öngörü değeri (forecasted value) denilmektedir.

2.2.3. Varsayımların geçerliliğinin incelenmesi

Çoklu doğrusal regresyon varsayımları Alt Bölüm 2.2.1.4'te belirtilmişti. Analizden herhangi bir sonuç çıkarmadan önce varsayımların sağlandığının gösterilmesi gerektiğini vurgulamak gerekir. Regresyon analizi yinelemeli (iterative) bir süreçtir. Bu süreç, sorunun tanımlanması, modelin oluşturulması, varsayımların geçerliliğinin incelenmesi, uyumun değerlendirilmesi ve modelin uygulanması aşamalarının yer aldığı bir süreçtir. Bu süreç memnun edici bir sonuç alınana kadar tekrarlanır. Memnun edici bir sonuç ise, varsayımları sağlayan ve veri kümesine oldukça iyi uyan bir kestirim modelidir. Bu yinelemeli süreç Şekil 2.1'den de görülebilmektedir.

Model varsayımlarının kontrolü artıkların grafiklerinin incelenmesi ile yapılabilmektedir. Artıkların grafiklerinde varsayım bozulmaları kolaylıkla tespit edilebilmektedir. Model varsayımlarının bir başka kontrolü ise modele uyum sağlamayan gözlemlerin incelenmesidir. Çünkü bir gözlem yeri geldiğinde tüm modeli etkileyebildiği ve varsayımları bozabildiğinden bu kontrol çok önemlidir. Kuşkulu gözlemlerin çeşitleri, modele etkileri ve tekli kuşkulu gözlem tespit yöntemleri daha ayrıntılı olarak Alt Bölüm 2.3 ve 2.4'te incelenecektir.

2.2.4. Uyumun değerlendirilmesi

Model oluşturulduktan sonra, sadece doğrusal bir ilişkinin varlığı ile değil, aynı zamanda modelin uyum kalitesinin ölçülmesi ile de ilgilenilmelidir. Uyumun kalitesini değerlendirmek için önem kontrol testleri ve belirtme katsayısı R^2 kullanılmaktadır.

Regresyon katsayılarının ve modelin önem kontrol testleri yapıldığında, test değerinin büyüklüğü, X ve Y arasındaki ilişkinin gücü hakkında bilgi verir. Test değerinin mutlak değeri ne kadar büyük ise X ve Y arasındaki ilişki o kadar güçlüdür.

R^2 , modelde kullanılan açıklayıcı değişkenlerin bağımlı değişkendeki toplam değişimi açıklayabilme oranı denir ve belirtme katsayısı olarak adlandırılır. R^2 değeri 1'e yaklaştıkça modelin iyi bir uyuma sahip olduğu söylenebilir. Fakat şu

husus gözden kaçırılmamalıdır ki, R^2 'nin her zaman büyük çıkması modelin iyi olduğunu göstermez. Çünkü modele, konu ile ilgili olsa da olmasa da eklenen her açıklayıcı değişken, R^2 'nin artmasına sebep olmaktadır. Eğer değişken seçimi doğru yapılmış ve Y ile X_j ler arasındaki ilişki gerçekten doğrusal ise iyi bir ölçü olabilir. Belirtme katsayısı, R^2 ,

$$R^2 = \text{RKT} / \text{GKT} = 1 - (\text{HKT} / \text{GKT}) \quad (2.9)$$

dir. Eğer modele anlamlı bir katkısı olmayan değişkenlerin dahil edildiği kuşkusu var ise düzeltilmiş belirtme katsayısına (R^2_d) bakılmalıdır. R^2_d , belirtme katsayısının serbestlik derecesine göre düzeltilmesi ile elde edilmektedir. Düzeltilmiş belirtme katsayısı, R^2_d ,

$$\begin{aligned} R^2_d &= 1 - \left\{ \left[\text{HKT} / (n-p) \right] / \left[\text{GKT} / (n-1) \right] \right\} \\ &= 1 - \left[(n-1) / (n-p) \right] (1-R^2) \end{aligned} \quad (2.10)$$

dir. Eğer R^2 ile R^2_d değerleri çok farklı değilse; kullanılan gözlem sayısının yeterli olduğu, aksi durumda ise anlamlı katkıları olmayan değişkenlerin modele dahil edildiği anlamını taşır.

Yukarıda verilen kriterler kullanılarak, sadece Y ile X arasında doğrusal bir ilişkinin olup olmadığı hakkında değil, ilişkinin kuvveti hakkında da bilgi elde edilir. Mutlak t değeri büyüdükçe Y ile X arasındaki ilişki kuvvetlenir. Bu testlerin güvenilir ve objektif sonuçlar vermesi için Alt Bölüm 2.2.1.4'te belirlenen bütün varsayımların sağlanması gerekmektedir.

Regresyon çözümleme aşamaları sonucunda herhangi bir problemle karşılaşılmamış ve modelin uyum değerlendirilmesi tatmin edici bir sonuç vermişse, model uygulanabilir bir modeldir. Bu aşamalarda yaşanacak herhangi bir sorun olduğunda Şekil 2.1'de gösterilen yol izlenmelidir.

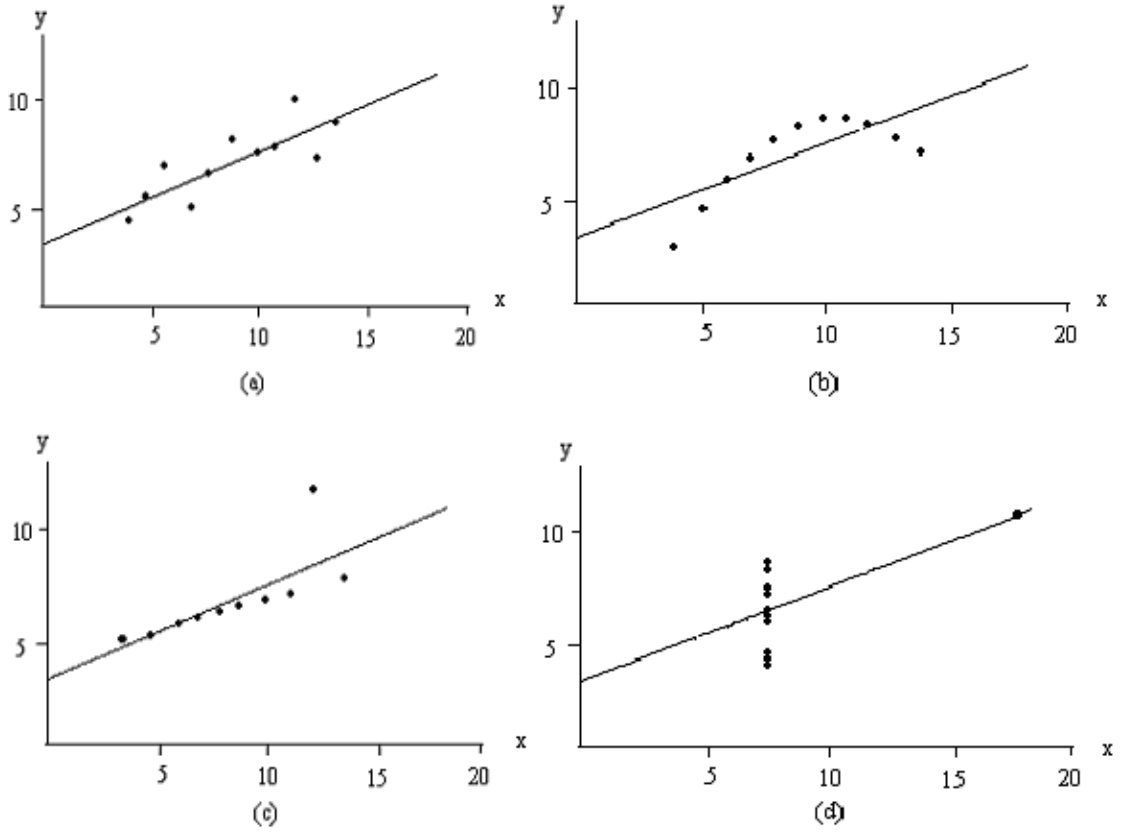
Regresyon çözümlemesi aşamalarından varsayım geçerliliğinin incelenmesi aşamasında belirtilen, kuşkulu gözlemlerin incelenmesi aşaması Alt Bölüm 2.3'te

daha ayrıntılı anlatılacaktır. Öncelikle kuşkulu gözlemin ne olduğunu açıklamak gerekir.

2.3. Kuşkulu Gözlemler

Veri kümesinin saçılım grafiğine bakıldığında verilerin çoğunluğunun dışında bulunan ve sonuçlar üzerinde yanıltıcı bir etkisi olacağından şüphelenen gözlemlere kuşkulu gözlemler denir.

Kuşkulu gözlemler, toplanan veriye uydurulan modeli birçok açıdan etkileyebilmektedir. Anscombe 1973 yılında aynı istatistik sonuçları veren dört ayrı veri kümesi için, aynı kestirim denklemini elde etmiştir. Şekil 2.2’de bu dört örnekleme ait grafikler verilmiştir. Bu grafiklerde kuşkulu gözlemlerin model üzerindeki etkileri açıkça görülebilmektedir.



Şekil 2.2 Anscombe'nin dört farklı veri kümesinin saçılım grafikleri ve kestirim doğruları (Alpar, 2003)

Şekil 2.2 (a)'daki saçılım grafiğinde Y ile X arasında doğrusal bir ilişki olduğu görülmektedir. Bu grafikte herhangi bir gözlem, diğer gözlemlerin vermek istediği bilgiye engel olmadığından kuşkulu gözlem yoktur. Şekil 2.2 (b)'deki saçılım grafiğinde, açıklayıcı değişken ile bağımlı değişken arasında doğrusal olmayan bir ilişkinin olduğu görülmektedir. Şekil 2.2 (c)'deki saçılım grafiğinde kuşkulu bir gözlemin bulunduğu ve veri kümesi dışında kalan bu gözlemin doğruyu nasıl yukarı çektiği görülmektedir. Şekil 2.2 (d)'de yer alan saçılım grafiğinde veri kümesinden uzakta bulunan tek bir gözlemin bile model denklemini ne denli büyük bir ölçüde değiştirebildiği görülmektedir.

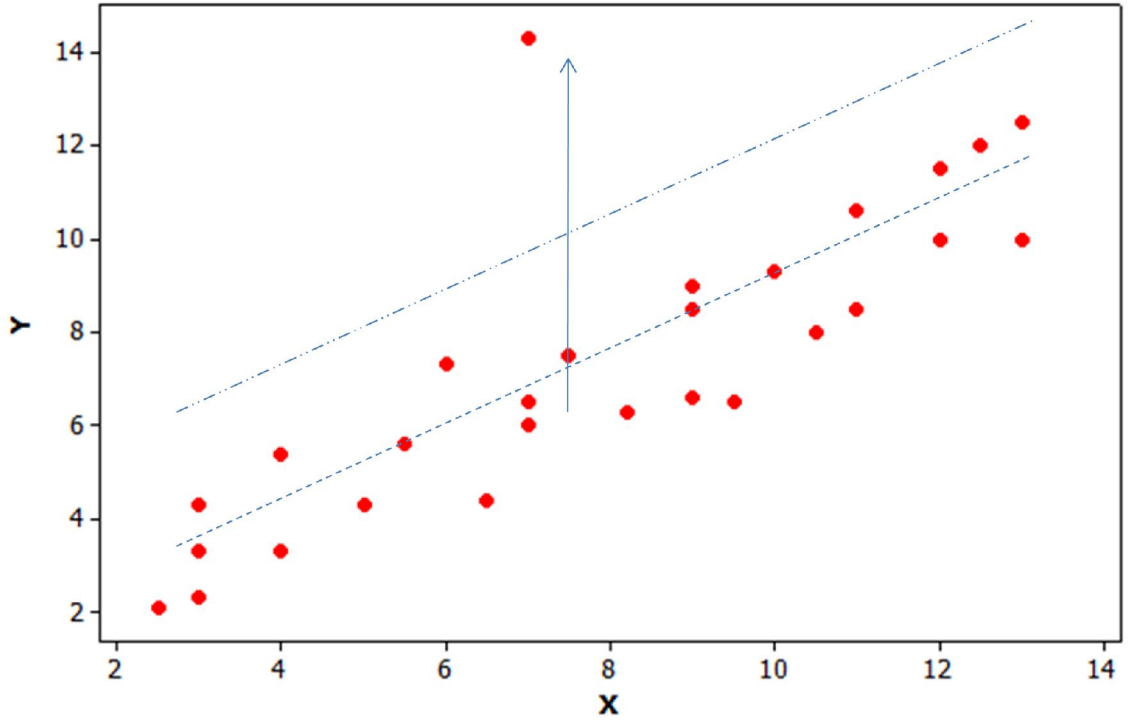
Verilere ait saçılım grafiğindeki gözlemlerin kestirilen regresyon doğrusu etrafında toplanması beklenir. Fakat Şekil 2.2 (c) ve Şekil 2.2 (d)'de olduğu gibi bir ya da birden fazla gözlemin diğer gözlemlerden uzakta yer alması mümkündür. Genel olarak veri kümesinin çoğunluğunun dışında kalan bu tür gözlemlere *kuşkulu gözlem*, *aykırı değer*, *sıra dışı gözlem* ya da *sapan değer* adları verilmektedir. Tez kapsamında bu tür gözlemler için kuşkulu gözlem ifadesi kullanılacaktır. Bu tip gözlemler, ölçüm hatası, verilerin girişindeki yazım hatası ya da ölçüm aracının yanlış ölçmesi vb. nedenlerden ortaya çıkabilir. Model denkleminin kestiriminde bu gözlemlerin varlığı, modeldeki parametre kestirimlerini, hata varyansının kestirimini, parametre kestirimlerinin varyanslarını ve buna bağlı olarak test istatistiklerini önemli ölçüde değiştirebilir. Kısaca bu gözlemlerin varlığında kestirim denklemine güvenilmez. Bu nedenle, mümkünse bu değerler düzeltilmeli, düzeltilemiyorsa bu gözlemler veri kümesinden çıkarılmalıdır. Kuşkulu gözlemler, aykırı gözlemler ve etkili gözlemler olarak iki başlık altında incelenebilir.

2.3.1. Aykırı gözlemler

Aykırı gözlemler; bağımlı değişken yönünde aykırı, açıklayıcı değişken yönünde aykırı ya da hem bağımlı hem de açıklayıcı değişken yönünde aykırı olabilir.

Bağımlı değişken yönünde aykırı gözlemler Y eksenini yönünde verilerin çoğunluğundan aykırılık gösteren gözlemlerdir. Bundan dolayı bu gözlemlere *Y eksenini yönünde aykırı gözlemler* ya da kısaca *aykırı değer* denir. Tez çalışması içinde aykırı değer ifadesi bu gözlemler için kullanılacaktır.

Şekil 2.3'de aykırı bir gözlem görülmektedir. Bu gözlem, bağımlı değişken yönünde bir sapma göstermiştir. Ayrıca bu noktanın apsisi, \bar{x} değerine yakın olduğu için regresyon doğrusunu sadece Y eksenini yönünde sürükleyecek, doğrunun eğimi üzerinde önemli bir etkisi olmayacaktır. Başka bir deyişle, bu durumda yalnız sabit değerin kestirim değeri $\hat{\beta}_0$ etkilenecektir. Ayrıca Y eksenini yönünde aykırı gözlemlerin varlığında parametre kestirimlerinin varyansları büyüyecektir.



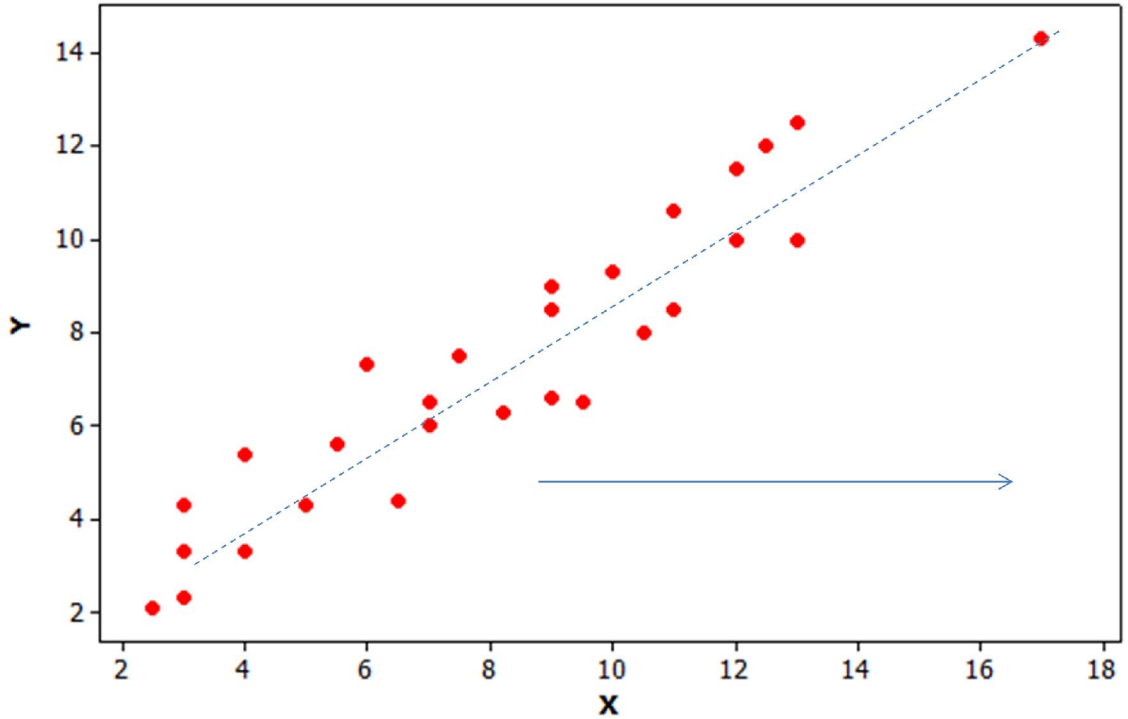
Şekil 2.3 Bağımlı değişken yönünde aykırı gözlem

Açıklayıcı değişken yönünde aykırı gözlemler, apsileri açıklayıcı değişkenin tanım aralığı dışında yer alan, başka bir deyişle X eksenini yönünde, uzaklaşmış gözlemlerdir. Bu gözlemler aynı zamanda, model denkleminin eğimini etkileyerek regresyon doğrusunu aşağı ya da yukarı çekebilmektedirler. X eksenini yönünde aykırı değerler, kestirim doğrusunun uzantısında yer alıyor ise doğruyu aşağı ya da yukarı çekmemekte, ancak doğrunun etkinliğini artırmaktadırlar. Tez çalışması boyunca bu tür gözlemler için X eksenini yönünde aykırı değer ifadesi

kullanılacaktır. X ekseninde aykırı değerler iki gruba ayrılabilir (Rousseeuw & Zomeren, 1990);

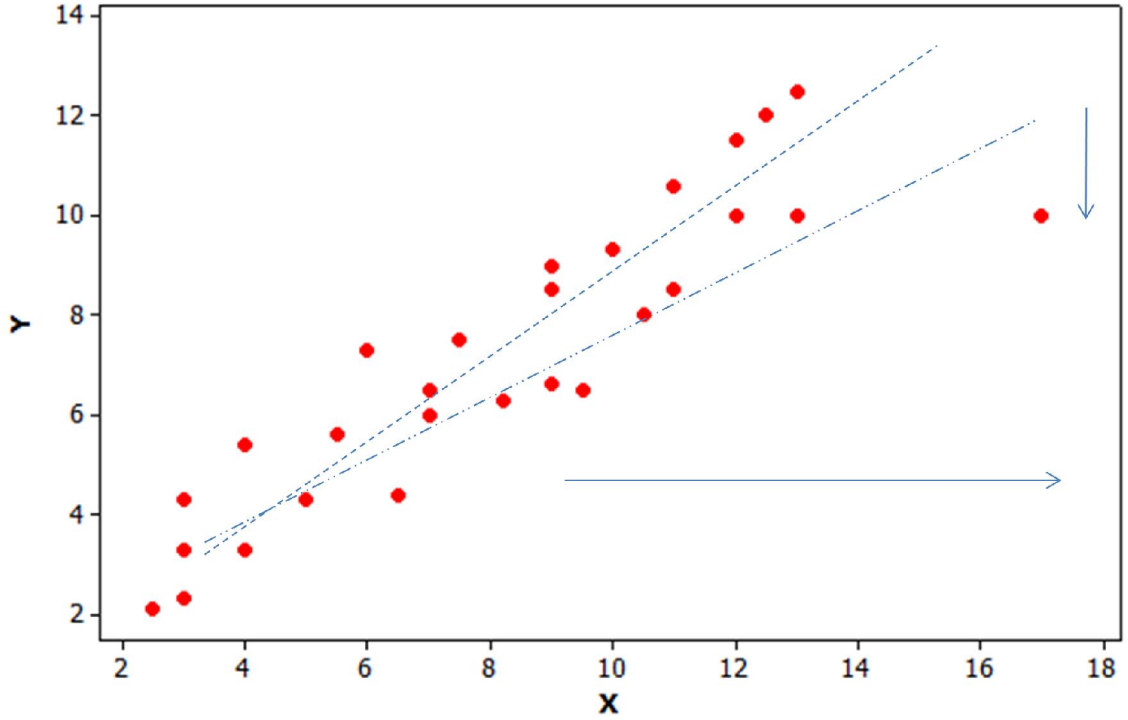
- İyi değerler, regresyon doğrusunun uzantısında olanlardır. Bir başka deyişle, regresyon katsayılarının doğruluğunu (precision) arttıran değerler, iyi değerlerdir.
- Kötü değerler ise, regresyon doğrusuna uzak olan değerlerdir. Yani, regresyon katsayılarının doğruluğunu azaltan değerler, kötü değerlerdir.

Şekil 2.4'te görülen X yönündeki aykırı değer, veri kümesinin oluşturacağı doğrunun uzantısında yer alır ve veri kümesinin diğer noktaları ile tutarlıdır. Bir başka deyişle, bu gözlemin standart hatası düşüktür. Bu durumda parametre kestirimlerinin varyansları küçülür. Başka bir deyişle, regresyon modelinin kalitesi arttığından iyi değerdir (Scanlon, 1994).



Şekil 2.4 Açıklayıcı değişken yönünde aykırı gözlem

Aykırı gözlemler, hem Y yönünde hem de X yönünde aykırı olabilmektedirler. Bu gözlemler, hem parametre kestirimlerinin değişmesine, hem de varyanslarının büyümesine neden olmaktadır.



Şekil 2.5 Hem bağımlı hem de açıklayıcı değişken yönünde aykırı gözlemler

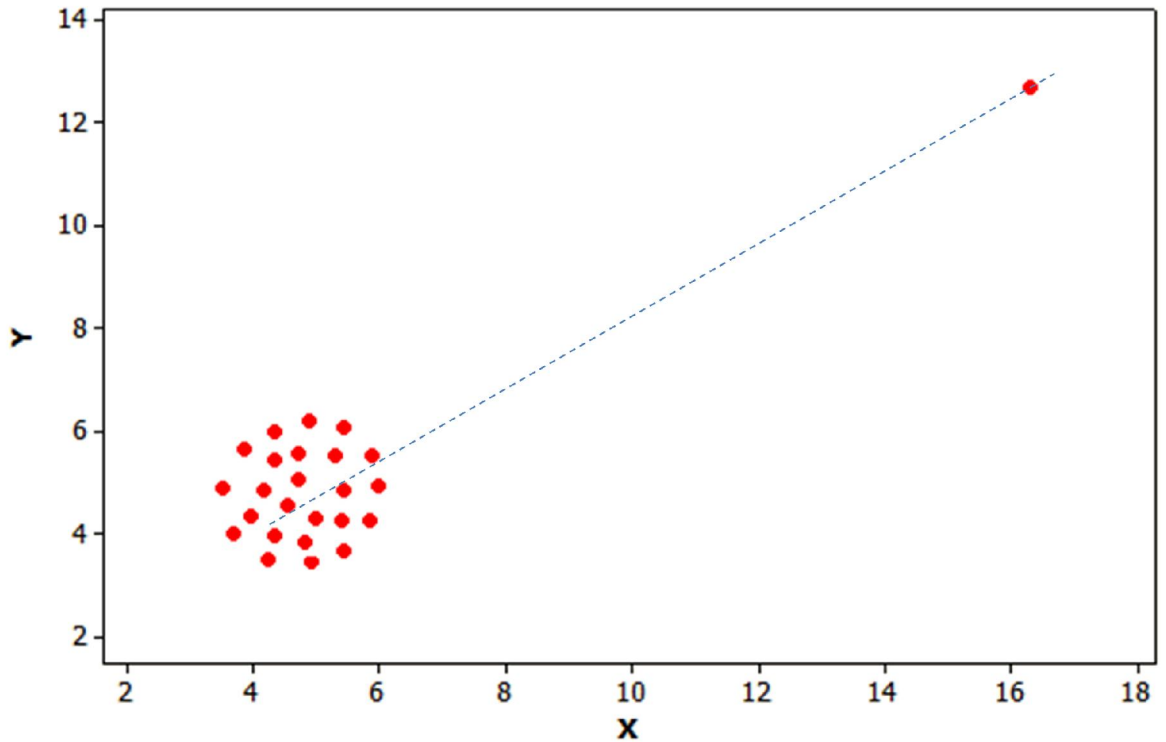
Şekil 2.5'deki aykırı gözlem, hem Y yönünde hem de X yönünde aykırı bir gözlemdir. Bu gözlem, veri kümesinin oluşturduğu regresyon doğrusundan Y yönünde uzaklaştığı için Y yönünde, \bar{x} değerinden X yönünde uzaklaştığı için X yönünde aykırı bir gözlemdir. Ayrıca bu gözlem, Şekil 2.3'te gösterilen aykırı gözlem ile karşılaştırıldığında farklı etkilere sahip olduğu görülür. Şekil 2.3'teki aykırı gözlem Y yönünde aykırı bir gözlemdir ve olması gereken yerden yaklaşık 8 birim uzaktadır. Şekil 2.5'teki aykırı gözlem ise hem Y hem de X yönünde aykırı bir gözlemdir ve olması gereken yerden yaklaşık 4 birim uzaktadır. Buna rağmen bu gözleminin parametre kestirimleri üzerindeki etkisi, Şekil 2.3'teki aykırı gözlemin parametre kestirimleri üzerindeki etkisinden büyüktür. Daha önceden de bahsedildiği gibi Şekil 2.3'teki aykırı gözlemi sadece sabit terimin kestirimi üzerinde etkilidir. Şekil 2.5'teki aykırı gözlemin hem sabit terim hem de diğer

parametre kestirimleri üzerindeki etkisi Şekil 2.3'teki aykırı gözleminkinden daha fazladır.

2.3.2. Etkili gözlemler

Veri kümesinden çıkarıldığı zaman, regresyon parametrelerinin kestirimleri üzerinde fark edilir bir değişime sebep olan gözlemlere *etkili gözlemler (influential observations)* denir (Shih & Weisberg, 1986). Sonuçlar üzerinde yanıltıcı bir etki yaratan gözlemlerdir (Rousseeuw & Zomeren, 1990).

Şekil 2.6'da veri kümesinden uzakta bulunan tek bir gözlem, açıkça model denklemini belirlemektedir. Bu gibi gözlemlere etkili gözlem denir. Model denklemini bu gözlem üzerinden geçeceği için bu gözlem aykırı değer olmayacaktır ve \bar{x} değerinden X yönünde çok uzaklaştığı için de X eksenini yönünde aykırı değer olacaktır. Bu gözlem, model parametreleri üzerinde aşırı etkili olduğu için, veri kümesinden çıkarıldığı zaman parametre kestirimleri büyük ölçüde değişecektir.



Şekil 2.6 Etkili gözlem

Y yönünde aykırı gözlemler, X yönünde aykırı gözlemler ve etkili gözlemler arasındaki ilişkiler aşağıdaki gibidir (Myers, 1990):

- Y yönünde aykırı gözlemler her zaman etkili gözlem olmak zorunda değildir.
- X yönünde aykırı gözlemler her zaman etkili gözlem değildir.
- Etkili gözlem her zaman Y yönünde aykırı gözlem ya da X ekseninde aykırı gözlem olmayabilir, ancak genellikle Y yönünde ya da X yönünde aykırı gözlemlerdir.

Regresyon modelinde tek açıklayıcı değişken olması durumunda yukarıdaki grafiklerde görüldüğü gibi herhangi bir kuşkulu gözlemi kolay bir şekilde tespit etmek mümkündür. Ancak iki ve ikiden fazla açıklayıcı değişken olması durumunda kuşkulu gözlemleri grafikler ile belirleme zor olacaktır. Bu durumda kuşkulu gözlemleri tespit etmek için bazı istatistiklerden yararlanılır. Aşağıdaki alt bölümde bu gözlemlerin nasıl ortaya çıkarılabileceği konusunda geliştirilen çeşitli yöntemler incelenecektir.

2.4. Tekli Kuşkulu Gözlem Belirleme Yöntemleri

Regresyon artıklarının analizi, modelin yeterliliğini belirlemede ya da modelde sorun yaratabilecek gözlemlerin belirlenmesinde oldukça yararlıdır. Kuşkulu gözlem tespit ederken artıklara dayalı istatistiklerinin yanı sıra artıkların grafikleri de kullanılmaktadır. Artıklara dayalı yöntemlere geçmeden önce, artıkların tanımını yapmak gerekir.

a) Klasik artıklar

Gözlemleri en iyi temsil edecek olan regresyon modelini belirleme sürecinde artıklar, gözlemler ile model arasındaki uygunluk hakkında açıklayıcı bilgi vermektedir.

Bir anlamda e_i artıkları, ε_i rastgele hatalarını ölçer. Artıklar, bağımlı değişkenin gözlenen değeri ile kestirilen değeri arasındaki farktır.

$$e_i = y_i - \hat{y}_i, \quad i = 1, 2, \dots, n \quad (2.11)$$

Burada şöyle bir problemle karşılaşmaktadır; ε_i 'ler (hata) ne gözlenebilir, ne de tahmin edilebilir. Ancak artıklar gözlenebilir. Eşitlik (2.7)'den yararlanarak

$$\begin{aligned} e &= Y - \hat{Y} = Y - X(X'X)^{-1}X'Y = (1 - H).Y \\ &= (1 - H).(X\beta + \varepsilon) \\ &= (1 - H).\varepsilon \end{aligned} \quad (2.12)$$

yazılabilir. Eşitlik (2.12)'ye bakıldığında, H matrisinin köşegen dışındaki öğeleri yeterli derecede küçük olduğunda e 'nin ε 'nin mantıklı bir tahmin edicisi olduğu görülür. Ayrıca ε 'nin elemanları bağımsız ve aynı varyanslı olmalarına karşın, Eşitlik (2.12)'den artıkların bağımsız ve aynı varyanslı olmadıkları söylenebilir. Artıkların dağılımı;

- Hataların dağılımına,
- H matrisinin elemanlarına,
- Örneklemin büyüklüğüne (n),

bağılıdır. Büyük örneklerde, $\frac{1}{n} \approx 0$ olduğu için $e_i \approx \varepsilon_i$ olur ve artık dağılımının analizi hataların dağılımı hakkında doğrudan bilgiler verir. Klasik artıklar sabit varyansa sahip değildir. Küçük ve orta büyüklükteki örnekler için klasik artıklar, etkili noktaların belirlenmesi için iyi sonuç vermeyebilirler. Bu nedenle kuşkulu gözlem araştırmalarında, klasik artıkların (e) dönüştürülmüş biçimleri kullanılmaktadır (Aydın, 2006).

b) Standartlaştırılmış artıklar

i. artığın standart sapması σ_i olmak üzere, standartlaştırılmış artık,

$$f(e_i, \sigma_i) = \frac{e_i}{\sigma_i} \quad (2.13)$$

biçiminde tanımlanır.

Eşitlik (2.13)'de σ_i yerine $(e'e)^{1/2}$ yerleştirilerek, $(e'e)^{1/2}$ 'ye göre standartlaştırılmış (normalized residual) i. artık, a_i ,

$$a_i \equiv e_{N,i} = f(e_i, \sqrt{e'e}) = \frac{e_i}{\sqrt{e'e}}, \quad i=1,2,\dots,n, \quad (2.14)$$

biçiminde elde edilir.

Eşitlik (2.13)'de σ_i yerine

$$\sqrt{\frac{e'e}{n-p}} = \hat{\sigma} \quad (2.15)$$

yerleştirilerek, $\sqrt{\frac{e'e}{n-p}}$ 'ye göre standartlaştırılmış (standartized residual) i. artık, b_i ,

$$b_i \equiv e_{S,i} = f(e_i, \hat{\sigma}) = \frac{e_i}{\sqrt{\frac{e'e}{n-p}}}, \quad i=1,2,\dots,n, \quad (2.16)$$

biçiminde elde edilir. Genelde b_i birim normal sapma olarak adlandırılır.

Eşitlik (2.13)'de σ_i yerine $\hat{\sigma}\sqrt{1-h_{ii}}$ yerleştirilerek, içsel student türü (internally studentized residual) i. artık, r_i ,

$$r_i \equiv f(e_i, \hat{\sigma}\sqrt{1-h_{ii}}) = \frac{e_i}{\hat{\sigma}\sqrt{1-h_{ii}}}, \quad i=1,2,\dots,n, \quad (2.17)$$

biçiminde elde edilir. Genelde bu artık Student tipi artık olarak bilinir.

Ayrıca, i. gözlem çıkarıldığında, artık kareler ortalamasının kestirimi $\hat{\sigma}_{(i)}^2$ ile gösterilirse,

$$\hat{\sigma}_{(i)}^2 = \frac{Y_{(i)}^T (I - H_{(i)}) Y_{(i)}}{n-k-1}, \quad i=1,2,\dots,n, \quad (2.18)$$

ve i. gözlem çıkarıldığında H matrisi $H_{(i)}$ ile gösterilirse, $H_{(i)}$,

$$H_{(i)} = X_{(i)}^T (X_{(i)}^T X_{(i)})^{-1} X_{(i)}, \quad i = 1, 2, \dots, n, \quad (2.19)$$

biçiminde elde edilir. Eşitlik (2.13)'de σ_i yerine $\hat{\sigma}_{(i)}\sqrt{1-h_{ii}}$ yerleştirilerek, dışsal student türü (externally studentized residual) i. artık, r_i^* ,

$$r_i^* \equiv e_{J,i} = f(e_i, \hat{\sigma}_{(i)}\sqrt{1-h_{ii}}) = \frac{e_i}{\hat{\sigma}_{(i)}\sqrt{1-h_{ii}}}, \quad i = 1, 2, \dots, n, \quad (2.20)$$

biçiminde elde edilir. Burada h_{ii} , H matrisinin i. köşegen elemanıdır. Bu artıklar, jackknife artıklar olarak da adlandırılmaktadır.

$\hat{\sigma}_{(i)}$ yerine $\hat{\sigma}_{(i)} = \hat{\sigma} \left(\frac{n-p-e_{S,i}}{n-p-1} \right)$ ifadesi yazılırsa $e_{J,i}$ nin $e_{S,i}$ cinsinden ifadesi,

$$e_{J,i} = e_{S,i} \left(\frac{n-p-1}{n-p-e_{S,i}} \right) \quad (2.21)$$

biçiminde olur.

Klasik artıklar ve standartlaştırılmış artıklar veri kümesindeki bütün gözlemleri hesaba katarak elde edilen model üzerinden hatayı tahmin etmektedirler. Hatayı tahmin etmenin başka bir yolu da, i. gözlem çıkarıldıktan sonra geriye kalan gözlemler üzerinden model kurarak kestirim yapmaktır. Bu artıklara Press artığı denir.

i. gözlem çıkarıldıktan sonra geriye kalan $n-1$ gözlem üzerine regresyon modeli kurulup bu model yardımı ile i. gözlemin önkestirim değeri $\hat{y}_{i(i)}$ elde edilerek, i. Press artığı,

$$e_{(i)} = e_{P,i} = y_i - \hat{y}_{i(i)} \quad (2.22)$$

biçiminde hesaplanır. Press artıkları kolayca,

$$e_{(i)} = \frac{e_i}{1 - h_{ii}} \quad (2.23)$$

yardımı ile hesaplanabilir.

Kuşkulu gözlemlerin tespit edilmesi ve analizler üzerindeki etkilerinin ölçülmesi için çok sayıda ilişkili yöntemler bulunmaktadır. Bu yöntemler, her birinin ilgilendiği özel durumlara bağlı olarak beş grupta toplanabilir.

2.4.1. Artıklara Dayalı Yöntemler

Regresyon çalışmalarında artıklar önemli rol oynamaktadır. Hiçbir analiz artıklarını ayrıntılı olarak incelemeyen tamamlanamaz. Alt Bölüm 2.4'te verilen standartlaştırılmış artıklara göre kuşkulu gözlemler belirlenebilmektedir. Artıklara dayalı bu yöntemler aşağıda verilmektedir.

a) $\sqrt{(e^t \cdot e) / (n - p)}$ 'ye göre standartlaştırılmış artıklara dayalı yöntem

Eşitlik (2.16) ile ifade edilen artıklar birim varyansa sahip olur ve $[-2, +2]$ aralığında bulunur. t rastgele değişkenine benzer davranış gösterir. Bu yüzden çoğu tanısal grafikte klasik artıklar yerine bu artıklar kullanılmaktadır.

b) İçsel student türü artıklara dayalı yöntem

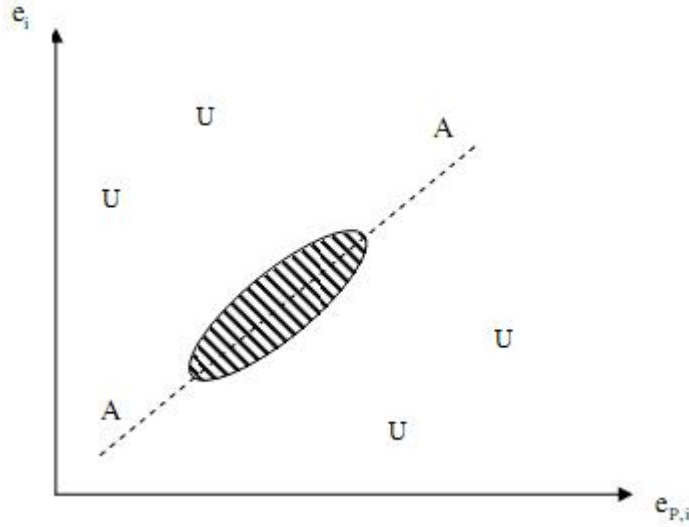
Büyük artışa ve büyük h_{ii} değerine sahip bir gözlemin en küçük kareler kestirimi üzerinde etkili olacağı düşünüldüğünden içsel student türü artıkların kullanılması önerilir. r_i değerleri $[-3, +3]$ aralığında yer alır. (Montgomery & Peck, 1992)

c) Dışsal student türü artıklara dayalı yöntem

Eşitlik (2.20) ile ifade edilen artıklar, aykırı değerlerin tespiti için sık sık kullanılır. $e_{j,i} \leq F_{(\alpha;1,n-p-1)}$ şartı geçerli olduğu zaman ilgili gözlemin etkili gözlem olmadığını belirtir. Bu artıklar, X ekseninde aykırı değerler hakkında hiçbir bilgi vermez. (Aydın, 2006)

Kuşkulu gözlem belirlemek için verilen bu yöntemlerin grafikleri oluşturularak istenen gözlemler daha kolay tespit edilebilmektedir. Bu grafikler, X eksenine sıralı olarak gözlem numaraları, Y ekseninde de artıkların çeşitleri olacak şekilde çizdirilir. Bir ekseninde gözlemlerin sıralı halinin bulunduğu bu grafiklere indeks grafikleri denir. Verilen kriteri geçen gözlemler kuşkulu gözlem olarak belirlenir.

Williams'ın 1973 yılında önerdiği press artıkları grafiği aykırı gözlem belirlemede oldukça yararlıdır. Şekil 2.7'de verilen bu grafikte X ekseninde press artıkları, Y ekseninde de klasik artıklar yer alır. Grafikteki $y = x$ doğrusu dışında kalan gözlemler X eksenine yönünde aykırı değer, doğru uzantısı üzerinde yer alan fakat verilerin çoğundan uzakta kalan gözlemler aykırı değer olarak belirlenirler (Aydın, 2006).



Şekil 2.7 Press artıklarının klasik artıklar ile grafiği

2.4.2. Gözlem uzaklığı matrisine dayalı yöntem

Eşitlik (2.7) 'de verilen $\hat{Y} = HY$ ifadesindeki

$$H = X(XX)^{-1}X' \quad (2.24)$$

matrisine, $(n \times n)$ boyutlu gözlem uzaklıkları matrisi denir. Ayrıca H matrisi, gözlenen değerlerden oluşan Y vektörünü onun en küçük kareler kestiricisi olan

\hat{Y} vektörüne dönüştürdüğü için şapka matrisi ya da dönüşüm matrisi olarak da tanımlanmaktadır (Rousseeuw & Leroy, 2003).

H matrisi, simetrik ($H' = H$) ve eşgüçlü (idempotent) ($H.H = H$) bir matristir.

$x_i' = (1 \ x_{i1} \ x_{i2} \ \dots \ x_{ik})$ gözlem vektörü iken, h_{ii} değeri Eşitlik (2.24)'den,

$$h_{ii} = x_i'(XX)^{-1}x_i \quad (2.25)$$

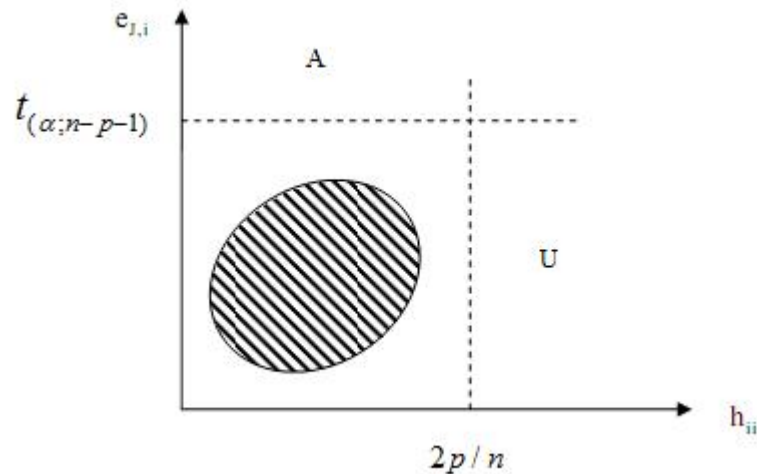
ile hesaplanır.

Gözlem uzaklığı matrisinin köşegen ögesi olan h_{ii} değeri, X eksen yönünde X 'in tanım aralığının merkezinden ayrılışın bir ölçüsüdür. (Draper & John, 1981)

Gözlemler, \bar{x} değerinden ne kadar uzaklaşırsa, h_{ii} değeri o kadar büyümektedir.

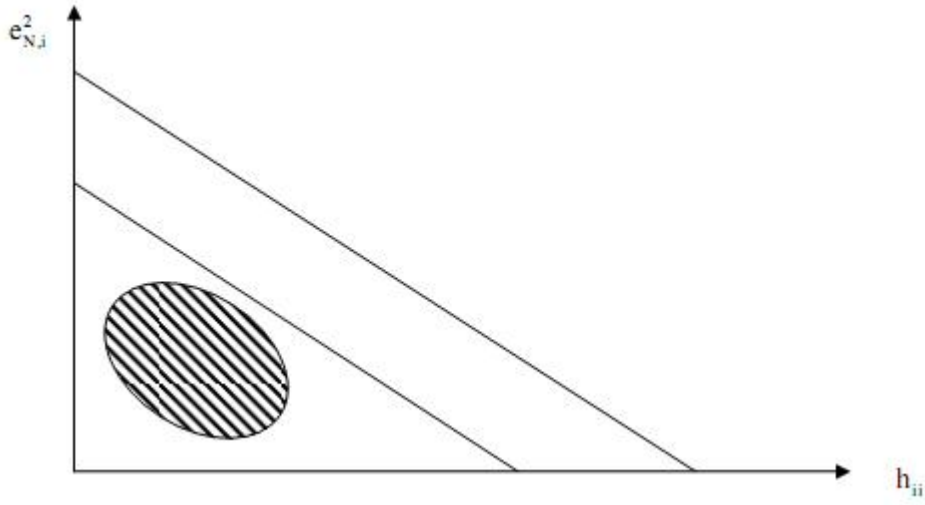
Hoaglin ve Welsch tarafından $h_{ii} > 2p/n$ olan gözlemler, X eksen yönünde aykırı değer olarak belirtilmiştir. Fakat $1/n \leq h_{ii} \leq 1$ olduğunda, $n \leq 2p$ ise, bu kriter geçerli olamaz. (Hoaglin & Welsch, 1978)

Kuşkulu gözlem belirlemede, gözlem uzaklık matrisinin köşegen değeri olan h_{ii} değerinin kullanıldığı çeşitli grafikler önerilmiştir. Bu grafikler Williams grafiği ve Pregibon grafiğidir.



Şekil 2.8 Williams Grafiği

Williams'ın 1973'te önerdiği ve Şekil 2.8'de verilen Williams grafiğinin X ekseninde h_{ii} değerleri, Y ekseninde de jacknife artıkları yer alır. İki adet sınırlayıcı çizgi çizilir. İlki aykırı değer belirlemek için olan $y = t_{(\alpha; n-p-1)}$ ve ikincisi X eksenini yönünde aykırı değer belirlemek için olan $x = 2p/n$ dir. Burada ifade edilen $t_{(\alpha; n-p-1)}$; $n-p-1$ serbestlik dereceli t dağılımının α için tablo değeridir. Grafikte X eksenini yönünde sınırı geçen noktalara ait gözlemler X eksenini yönünde aykırı değer, Y eksenini yönünde sınırı geçen noktalara ait gözlemler aykırı değer olarak belirlenir (Meloun & Militky, 2001).



Şekil 2.9 Pregibon grafiği

Pregibon grafiğinde ise, X ekseninde h_{ii} değerleri, Y ekseninde de $e_{N,i}^2$ değerleri bulunmaktadır. Şekil 2.9'da verilen bu grafikte sınırlar

$$E(h_{ii} + e_{N,i}^2) = (p+1)/n \quad (2.26)$$

ifadesinden elde edilen $y = -x + 2(p+1)/n$ ve $y = -x + 3(p+1)/n$ doğrularıdır.

Bu grafik yardımı ile etkili gözlemler belirlenir. Eğer bir nokta en üstteki doğrunun dışında ise etkili bir gözlemdir. Bu gözlem, doğrular arasında kalan bir noktaya karşılık gelen gözlemden daha güçlü bir etkiye sahiptir (Meloun & Militky, 2001).

2.4.3. Güven elipsoidinin merkezine dayalı yöntemler

Normallik varsayımı altında, β için %100.(1- α) güvenilirlik derecesindeki birleşik güven bölgesi,

$$\frac{(\beta - \hat{\beta})'(X'X)(\beta - \hat{\beta})}{p\hat{\sigma}^2} \leq F_{(\alpha;p,n-p)} \quad (2.27)$$

biçiminde ifade edilmektedir. $F_{(\alpha;p,n-p)}$, p ve $n-p$ serbestlik dereceli F dağılımının üst α değeridir.

Eşitsizlik (2.27), merkezi $\hat{\beta}$ olan ve elipsoid biçimindeki bir bölgeyi ifade etmektedir. i . gözlemin etkisi, i . gözlem çıkartıldığı zaman bu eşitsizlik ile ifade edilen elipsoidin merkezinde meydana gelen değişim tarafından ölçülebilir. Bu değişim göz önüne alınarak çeşitli yöntemler önerilmiştir. Bu yöntemler aşağıdaki alt bölümlerde tanıtılacaktır.

2.4.3.1. Cook uzaklığı

Cook 1977'de i . gözlemin Eşitsizlik (2.27) ile ifade edilen elipsoidin merkezine ve kestirilen parametrelere etkisini belirlemek için aşağıdaki ölçütü önermiştir:

$$C_i = \frac{(\hat{\beta} - \hat{\beta}_{(i)})'(X'X)(\hat{\beta} - \hat{\beta}_{(i)})}{p\hat{\sigma}^2}, \quad i = 1, 2, \dots, n \quad (2.28)$$

Burada $\hat{\beta}_{(i)}$, i . gözlem çıkartıldığı durumda elipsoidin merkezidir. Bu ölçüte Cook uzaklığı denir ve $\hat{\beta}$ ile $\hat{\beta}_{(i)}$ arasındaki uzaklığın bir ölçüsü olarak düşünülebilir.

Cook uzaklığı, aynı zamanda artıklar ve gözlem uzaklığı değerleri ile

$$C_i = \frac{r_i^2}{p} \frac{h_{ii}}{1-h_{ii}} \quad (2.29)$$

biçiminde ifade edilir. $F_{(\alpha;p,n-p)}$ değerinden büyük Cook uzaklığına sahip gözlemler kuşkulu gözlemlerdir. (Cook, 1977)

2.4.3.2. Welsch uzaklığı

Welsch 1982 yılında i . gözlemin Eşitsizlik (2.27) ile ifade edilen elipsoidin merkezine ve kestirilen parametrelere etkisini ölçmek için aşağıdaki yöntemi önermiştir.

$$W_i = \sqrt{(n-1) \cdot r_i^{*2} \frac{h_{ii}}{(1-h_{ii})^2}} \quad (2.30)$$

Burada r_i^* , i . gözlemin jacknife artığını ve h_{ii} , i . gözlemin gözlem uzaklığını ifade etmektedir. Eğer n , p 'ye göre çok büyük ise, W_i değeri $3\sqrt{p}$ ile karşılaştırılır. W_i değeri $3\sqrt{p}$ 'den büyük olan gözlemler kuşku gözlem olarak belirlenir (Welsch, 1982).

2.4.3.3. Welsch-Kuh uzaklığı

Welsch ve Kuh tarafından 1977 yılında i . gözlemin, Eşitsizlik (2.27) ile ifade edilen elipsoidin merkezine ve kestirilen parametrelere etkisini belirlemek için aşağıdaki ölçütü önermiştir. Bu ölçüt,

$$WK_i = |r_i^*| \sqrt{\frac{h_{ii}}{1-h_{ii}}} \quad (2.31)$$

biçiminde ifade edilmektedir. Aynı zamanda DFITTS istatistiği de denir. Welsch uzaklığı ile arasındaki ilişki de

$$WK_i = W_i \sqrt{\frac{1-h_{ii}}{n-1}} \quad (2.32)$$

ile ifade edilir. WK_i değeri $2\sqrt{p/n}$ 'den büyük olan gözlemler kuşku gözlem olarak belirlenir (Welsch & Kuh, 1977).

2.4.3.4. Düzeltilmiş Cook Uzaklığı

Welsch ve Kuh tarafından 1977 yılında önerilen bu ölçüt, Eşitlik (2.28)'in uyarlanmış biçimidir ve

$$\begin{aligned}
C_i^* &= |r_i^*| \sqrt{\frac{h_{ii}}{1-h_{ii}} \frac{n-p}{p}} \\
&= WK_i \sqrt{\frac{n-p}{p}}
\end{aligned}
\tag{2.33}$$

biçiminde ifade edilir. $2\sqrt{(n-p)/n}$ değerinden büyük C_i^* değerine sahip gözlemler kuşkulu gözlemlerdir (Welsch & Kuh, 1977).

2.4.4. Güven elipsoidinin hacmine dayalı yöntemler

Bu bölümde i . gözlemin etkisi, Alt Bölüm 2.4.3'te ifade edilen merkezi $\hat{\beta}$ olan elipsoidin hacmine bağlı olarak ölçülmüştür. i . gözlem çıkarıldığı zaman bu elipsoidin merkezinde meydana gelen değişimi ölçerek tekli kuşkulu gözlem belirlemesi yapılabilmektedir. Bu değişimin ölçüsüne dayalı olan yöntemler aşağıdaki alt bölümlerde tanıtılacaktır.

2.4.4.1. Andrews-Pregibon istatistiği

Andrews ve Pregibon tarafından 1978 yılında önerilen ölçüt,

$$AP_i = h_{ii} + (1-h_{ii}) \frac{r_i^2}{n-p}
\tag{2.34}$$

biçiminde ifade edilmektedir. (Andrews & Pregibon, 1978) $AP_i > 2(p+1)/n$ koşulunu sağlayan gözlemler kuşkulu gözlemdir (Chatterjee & Hadi, 1986).

2.4.4.2. Varyans oranı

Belsley, Kuh ve Welsch tarafından 1980 yılında önerilen bu yöntem, COVRATIO istatistiği de denir. Belsley, Kuh ve Welsch, i . gözlemin etkisini, $\hat{\beta}$ ve $\hat{\beta}_{(i)}$ kestiricilerinin varyanslarının karşılaştırılması ile elde etmektedirler. Varyans oranı

$$VR_i = \frac{\det\{V[\hat{\beta}_{(i)}]\}}{\det\{V[\hat{\beta}]\}} \quad (2.35)$$

$$= \left(\frac{n-p-r_i^2}{n-p-1}\right)^p \frac{1}{1-h_i}$$

ile ifade edilir. İdeal olarak, tüm gözlemlerin kovaryans matrisi üzerindeki etkisi aynı ise VR_i yaklaşık olarak 1'e eşittir. 1'den sapmalar i . gözlemin muhtemelen kuşkulu gözlem olduğunu göstermektedir. $|VR_i - 1| \geq 3p/n$ ise i . gözlem kuşkulu gözlemdir (Belsley et al., 1980).

2.4.4.3. Cook-Weisberg istatistiği

Cook ve Weisberg tarafından 1982 yılında önerilen bu istatistik

$$CW_i = -\frac{1}{2} \log(VR_i) + \frac{p}{2} \log\left(\frac{F_{(\alpha;p,n-p)}}{F_{(\alpha;p,n-p-1)}}\right) \quad (2.36)$$

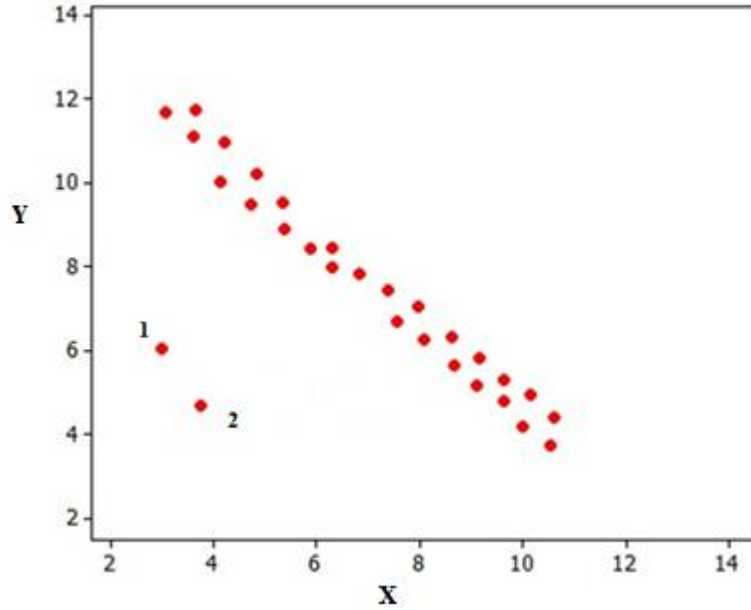
biçiminde ifade edilir. Eşitlik (2.36)'da ki ikinci terim i . gözleme bağlı olmadığından CW_i ve VR_i eşdeğerdir ve genellikle aynı sonuçları vermektedir (Chatterjee & Hadi, 1988).

3. ÇOKLU KUŞKULU GÖZLEM BELİRLEME YÖNTEMLERİ

3.1. Giriş

Alt Bölüm 2.4'te tekli kuşkulu gözlemlerin belirlenmesi için çeşitli yöntemlere değinildi. Bu bölümde tekli kuşkulu gözlemlerin tespit yöntemlerinin daha genel bir durumu olan çoklu kuşkulu gözlemlerin belirlenmesi durumu anlatılacaktır. Amaç, çeşitli regresyon sonuçları üzerinde gözlem gruplarının birlikte etkilerini belirlemektir.

Bu aşamada doğal olarak "Neden çoklu kuşkulu gözlem tespit yöntemlerine ihtiyaç var?" sorusu akla gelebilir. Gözlemlerin, tek tek kuşkulu gözlem olmadığı ancak birlikte iken kuşkulu gözlem oldukları ya da tek tek kuşkulu gözlem oldukları ancak birlikte olduklarında kuşkulu gözlem olmadıkları durumlar olabilir. Örneğin,

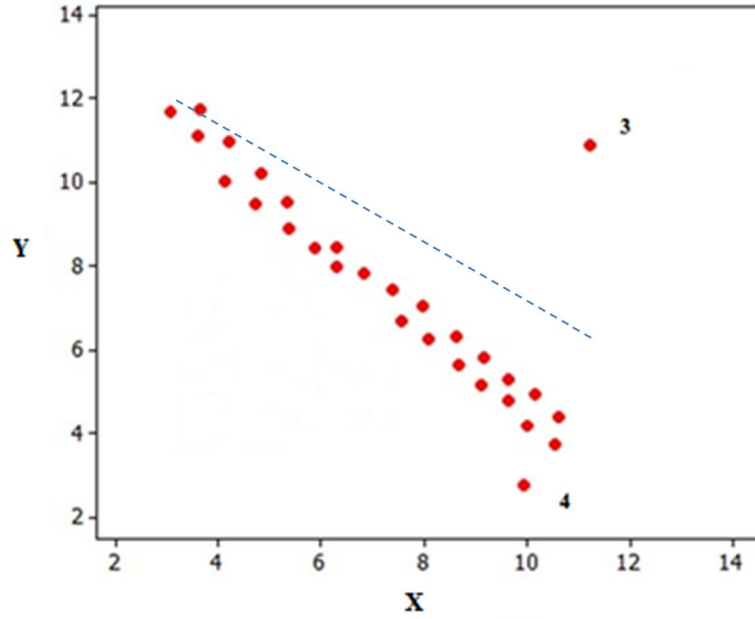


Şekil 3.1 Birlikte etkili gözlemler

Şekil 3.1'de verilen grafikte 1 ve 2 ile gösterilen gözlemler ayrı ayrı kuşkulu gözlem değildirlir. Fakat birlikte büyük bir etkiye sahiptirler. Bu durumda bir kuşkulu gözlem, başka bir gözlemin varlığı ile gizlenmektedir. Başka bir deyişle, iki kuşkulu gözlem olması durumunda, gözlemlerden biri veri kümesinden çıkarıldığında diğeri

kuşkulu gözlem olarak belirlenmektedir. Bu duruma gizleme (masking) etkisi denir (Chatterjee & Hadi, 1988). Çünkü bir kuşkulu gözlem, başka bir gözlemin varlığı ile maskelenmiştir (Seber & Lee, 2003).

Diğer yandan tekli kuşkulu gözlem tespit yöntemleri ile etkileri belirlenebilen, yani tek tek kuşkulu gözlem olan, ancak birlikte olduklarında kuşkulu gözlem oldukları belirlenemeyen durumlar da olabilir. Örneğin, Şekil 3.2 'de verilen grafikte 3 ve 4



Şekil 3.2 Ayrı ayrı etkili fakat birlikte etkili olmayan gözlemler

ile gösterilen gözlemler tek tek incelendiğinde iki gözlem de kuşkulu gözlemdir, ancak birlikte olduklarında kuşkulu gözlem olarak belirlenemezler. Kuşkulu gözlemlerin regresyon doğrusunu kendilerine doğru çekmesiyle diğer gözlemlerin modelden uzaklaşması sonucu kuşkulu gözlem olarak ele alınmaları durumu gerçekleşmektedir. Bu duruma sürüklenme etkisi (swamping effect) denir (Meloun & Militky, 2001).

Birlikte etki problemi kuramsal açıdan olduğu kadar uygulama yönünden de önemlidir. Birlikte etki, genelde tek gözlem etkisinden daha fazla etkilidir ve uygulayıcılar tarafından sıklıkla gözden kaçırılmaktadır. Çünkü çoklu kuşkulu gözlemlerin belirlenmesi, tekli kuşkulu gözlemlerin belirlenmesinden daha zordur.

Birlikte etkili gözlemlerin çoğu Alt Bölüm 2.4'te anlatılan yöntemlerin genelleştirilmiş biçimleri ile tespit edilebilmektedir.

3.2. Çoklu Kuşkulu Gözlem Belirlemede Yaşanan Sıkıntılar

Çoklu kuşkulu gözlemlerin tespitinin doğasından gelen üç sorunu vardır. Birincisi, birlikte etkili olan gözlemlerin oluşturduğu alt kümenin büyüklüğünün belirlenememesidir. Birlikte olduklarında kuşkulu oldukları düşünülen gözlemlerin, büyüklüğü $m = 2, 3, \dots$ olan tüm alt kümelerinin belirlenmesiyle ilgilenildiği varsayalım. m nasıl belirlenebilir? m 'nin belirlenmesi zor olduğundan ardışık bir yöntemin kullanılması kolaylık sağlayabilir. $m = 2$, sonra $m = 3$, vb. ile başlanır. Ancak bu durumda da nerede durulacağı sorusu karşımıza çıkar.

İkinci problem ise hesaplamadır. Tekli kuşkulu gözlem tespitindeki her bir yöntemde veri setindeki her bir gözlem için hesaplama yapılmaktadır, başka bir deyişle aynı işlem n sayıda tekrarlanır. Oysa çoklu kuşkulu gözlem durumunda her bir tespit yönteminde m büyüklüğündeki her alt küme için, başka bir deyişle $n!/[m!(n-m)!]$ kadar olası alt küme için, hesaplama yapılır. Buna ek olarak, tekli kuşkulu gözlem tespit yöntemleri, modelin yapısındaki basit temel fonksiyonları (örneğin artıklar e , artık kareler ortalaması $\hat{\sigma}^2$, H matrisinin i . köşegen elemanı h_{ii} , vs.) kullanıyor iken, çoklu kuşkulu gözlem tespit yöntemlerinde genellikle daha fazla hesaplama gerektiren fonksiyonlar (örneğin H matrisinin köşegen dışı elemanları h_{ij} , H ve $(I - H)$ matrislerinin tersleri, vs.) kullanılmaktadır. Bugünün hızlı bilgisayarları bile, n ve m çok büyük olduğunda çok fazla hesaplama yapmaktadırlar.

Üçüncü problem ise grafiksel gösterimle ilgilidir. Çoklu kuşkulu gözlemleri grafik üzerinde göstermek, etkilerini incelemek ve tespit etmek, özellikle n ve m büyükse, kolayca mümkün olmamaktadır. Gerek değişken sayısının artmasından gerekse grafik ölçeğinin büyümesinden, çok sayıda gözlem arasından m gözlemin grafiksel gösterimi sorun olmaktadır.

Aşağıdaki alt bölümlerde çoklu kuşkulu gözlem tespit yöntemleri tanıtılmaktadır.

Çoklu kuşkulu gözlem tespit yöntemlerindeki hesaplama güçlüğü en aza indirmek için kullanılan yöntemlerden birisi geri eleme (backward elimination) yöntemidir. m boyuttaki çoklu kuşkulu gözlemlerin tespit edilmesinde olası tüm altkümelerin hesaplanması gerekliliğini ortadan kaldırmaktadır. N gözlemlili bir veri kümesinde, çoklu kuşkulu gözlem tespit yöntemlerinden genelleştirilmiş Cook uzaklığını kullanarak adimsal regresyondaki geriye doğru çıkarmaya benzer bir yaklaşım ile en çok etkili m gözlemlili alt kümeyi araştırılır. İşlemlere başlamak için veri kümesinden bir başlangıç alt küme seçilir. Başlangıç alt küme boyutu s ile gösterilirse, $m < s < n-p$ olmalıdır. Veri setinden $n-p$ gözlemlili rasgele bir örneklem seçerek başlangıç alt kümesi oluşturulur. Daha sonra, başlangıç alt kümesindeki her gözlem teker teker çıkarılarak geriye kalan $s-1$ gözlemlili altkümelerin C_1 değeri hesaplanır. En büyük C_1 değerine sahip $s-1$ gözlemlili altküme alınır. $s-1$ gözlemlili altkümedeki her gözlem teker teker çıkarılarak geriye kalan $s-2$ gözlemlili altkümelerin C_1 değeri hesaplanır. En büyük C_1 değerine sahip $s-2$ gözlemlili altküme alınarak aynı işlem bu altküme üzerine uygulanır. En büyük C_1 değerine sahip m gözlemlili altküme elde edene kadar devam edilir.

Tek bir başlangıç alt kümesi en etkili m gözlemi içermeyebileceğinden, bu süreci yeni başlangıç alt kümeleri ile birçok kez tekrarlamak gereklidir. m gözlemlili en etkili altküme başarılı olarak bulunmanın olasılığına ulaşmak için gerekli tekrar sayısı basit olasılık hesaplamaları kullanılarak yaklaşık olarak belirlenebilir. Belirlenen sayıdaki aranılan gözlemler bir ya da birden çok başlangıç alt kümelerinde yer aldığı varsayımı altında, yapılan tekrarların en az birindeki gözlem çıkarmalarından kurtulacağı varsayımına dayanır ve bu gözlemler, en iyi etkili alt küme kabul edilir. Yineleme sayısını iki sebepten dolayı yaklaşık bir sayıdır. Birincisi, başlangıç alt küme büyüklüğü rastgele seçme işleminden dolayı en iyi altküme oluşturulan gözlemlerden bir ya da daha fazlasını içermeyebilir. İkincisi, en iyi altkümedeki gözlemlerden biri, başlangıç alt kümesindeki diğer gözlemlerden dolayı ilk aşamalarda çıkarılabilir. Uygulamada, tespiti önemli olan belli sayıda etkili bir altkümenin olduğunu varsaymak için önsel bir temel bulunmamaktadır. Burada m değerini her yinelemedeki gözlem çıkarma işlemi için bir durma noktası olarak görebiliriz. Sonuç olarak bu yöntem, çoklu kuşkulu gözlem tespitinde

kullanılan her yöntem için geçerli olan olası tüm altkümelerin hesaplanması yükünü aza indirgeyerek bir çözüm oluşturmuştur. (Delozier & Orlich, 2005)

Çoklu kuşkulu gözlem tespit yöntemleri Alt Bölüm 2.4'teki yöntemlerin genelleştirilmiş durumları anlatılacak şekilde sınırlandırılmıştır. Bu yöntemlere ek olarak alt küme büyüklüğünün (m), belirtilmesine gerek duymayan bir yöntem sunulmaktadır. Bu yöntemler aşağıdaki alt bölümlerde verilmektedir.

3.3. Artıklara Dayalı Yöntemler

Veri kümesinden çıkarılan m sayıda gözlemin bulunduğu alt küme

$$I = \{i_1, i_2, \dots, i_m\}, \quad m < (n - p) \quad (3.1)$$

biçiminde gösterilirse, Y ve X matrisleri,

$$Y = \begin{pmatrix} Y_{(I)} \\ Y_I \end{pmatrix} \begin{matrix} (n-m) \times 1 \\ m \times 1 \end{matrix} \quad (3.2)$$

$$X = \begin{pmatrix} X_{(I)} \\ X_I \end{pmatrix} \begin{matrix} (n-m) \times p \\ m \times p \end{matrix} \quad (3.3)$$

biçiminde parçalanabilir. I altkümesindeki gözlemlerin gözlem uzaklıkları matrisi,

$$H_I = X_I'(X'X)X_I \quad (3.4)$$

ve klasik artıkları,

$$e_I = Y_I - X_I'\hat{\beta} \quad (3.5)$$

biçiminde ifade edilirse, Alt Bölüm 2.4.1'de tanımlanan içsel Student türü artıkların genelleştirilmiş hali olan r_I ,

$$r_I^2 = \frac{e_I'(I - H_I)^{-1}e_I}{\hat{\sigma}^2} \quad (3.6)$$

biçiminde ifade edilir. Burada $\hat{\sigma}^2$, tam model için artık kareler toplamı, (n-p-m) serbestlik derecesi ile bölünerek elde edilir. En yüksek değeri alan altküme çoklu kuşkulu gözlemlerin bulunduğu alt kümedir.

3.4. Güven Elipsoidinin Merkezindeki Değişime Dayalı Yöntemler

Alt Bölüm 2.4.3'te verilen Eşitsizlik (2.27), merkezi $\hat{\beta}$ olan ve elipsoid biçimindeki bir bölgeyi ifade etmektedir. Çoklu kuşkulu gözlemlerin etkisi, I altkümesinde yer alan gözlemler çıkartıldığı zaman bu elipsoidin merkezinde meydana gelen değişim tarafından ölçülebilir. Bu değişim göz önüne alınarak çeşitli yöntemler önerilmiştir. Bu yöntemler aşağıdaki alt bölümlerde tanıtılacaktır.

3.4.1. Genelleştirilmiş Cook uzaklığı

Alt Bölüm 2.4.3.1'de verilen tekli kuşkulu gözlem tespit yöntemlerinden Cook uzaklığının genelleştirilmesi ile çoklu kuşkulu gözlemler tespit edilebilir. I alt kümesindeki gözlemlerin etkisi, I alt kümesindeki gözlemler veri setinden çıkarıldığında elde edilen Eşitlik (2.27)'de verilen ve merkezi $\hat{\beta}$ olan elipsoid biçimindeki bir bölgenin merkezindeki değişim ölçülerek belirlenebilir. Bu değişimin ölçüsü olarak Cook, 1977'de aşağıdaki ölçütü önermiştir:

$$C_I = \frac{(\hat{\beta} - \hat{\beta}_{(I)})'(X'X)(\hat{\beta} - \hat{\beta}_{(I)})}{p \cdot \hat{\sigma}^2} \quad (3.7)$$

$$= \frac{e_I'(I - H_I)^{-1} H_I (I - H_I)^{-1} e_I}{p \cdot \hat{\sigma}^2}$$

Bu istatistik Eşitlik (2.28)'in genelleştirilmiş halidir. En yüksek C_I değerini alan I alt kümesindeki gözlemler, çoklu kuşkulu gözlemler olarak belirlenir (Chatterjee & Hadi, 1988).

3.4.2. Genelleştirilmiş Welsch uzaklığı

Eşitlik (2.27)'de verilen ve merkezi $\hat{\beta}$ olan elipsoid biçimindeki bir bölgenin merkezindeki değişimin ölçümüne dayanan yöntemlerden birisi olan bu yöntem,

Alt Bölüm 2.4.3.2'de verilen Welsch uzaklığının genelleştirilmiş biçimidir. Welsch'in 1982'de önerdiği bu ölçüt,

$$W_I^2 = \left(\frac{n-m}{\hat{\sigma}_{(I)}^2} \right) e_I' (I - H_I)^{-2} H_I (I - H_I)^{-1} e_I \quad (3.8)$$

Eşitlik (2.30)'ün genelleştirilmiş biçimidir. En yüksek W_I değerini alan I alt kümesi çoklu kuşkulu gözlemlerin bulunduğu alt küme olarak belirlenir (Imon, 2005).

3.4.3. Genelleştirilmiş Welsch-Kuh uzaklığı

Bu yöntem, Alt Bölüm 2.4.3.3'de verilen ve merkezi $\hat{\beta}$ olan elipsoidin merkezindeki değişimin ölçümüne dayanan yöntemlerden birisi olan Welsch-Kuh uzaklığı yönteminin çoklu kuşkulu gözlemleri tespit etmek için genelleştirilmiş biçimidir. I ile belirtilen altkümedeki m tane gözlemin birlikte etkisi, 1977'de Welsch ve Kuh tarafından önerilen aşağıdaki istatistik ile ölçülebilir.

$$\begin{aligned} WK_I^2 &= \frac{(\hat{\beta} - \hat{\beta}_{(I)})' X_I H_I^{-1} X_I' (\hat{\beta} - \hat{\beta}_{(I)})}{\hat{\sigma}_{(I)}^2} \\ &= \frac{e_I' (I - H_I)^{-1} H_I (I - H_I)^{-1} e_I}{\hat{\sigma}_{(I)}^2} \end{aligned} \quad (3.9)$$

Bu istatistik Eşitlik (2.31)'in genelleştirilmiş biçimidir. En yüksek WK_I değerini alan altküme, çoklu kuşkulu gözlemlerin bulunduğu altküme olarak tespit edilir (Imon, 2005).

3.4.4. Genelleştirilmiş düzeltilmiş Cook uzaklığı

Bu yöntem, Alt Bölüm 2.4.3.4'te verilen ve merkezi $\hat{\beta}$ olan elipsoid biçimindeki bir bölgenin merkezindeki değişimin ölçümüne dayanan bir başka yöntem olan düzeltilmiş Cook uzaklığının çoklu kuşkulu gözlemleri tespit etmek amacı ile genelleştirilmiş biçimidir. Welsch ve Kuh tarafından 1977 yılında önerilen bu ölçüt,

$$C_I^* = \left(\frac{(\hat{\beta} - \hat{\beta}_{(I)})'(X'X)(\hat{\beta} - \hat{\beta}_{(I)})}{(p/(p-k))\hat{\sigma}_{(I)}^2} \right)^{1/2} = WK_I \sqrt{\frac{n-p}{p}} \quad (3.10)$$

Eşitlik (2.33)'ün genelleştirilmiş halidir. En yüksek C_I^* değerini alan altküme, çoklu kuşku gözlemlerin bulunduğu alt kümedir (Imon, 2005).

3.5. Güven Elipsoidinin Hacmine Dayalı Yöntemler

Bu bölümde çoklu kuşku gözlemlerin etkisi, Alt Bölüm 2.4.4'te ifade edildiği gibi, merkezi $\hat{\beta}$ olan elipsoidin hacmine bağlı olarak ölçülmüştür. I alt kümesinde yer alan gözlemler çıkarıldığı zaman, bu elipsoidin merkezinde meydana gelen değişim ölçülerek çoklu kuşku gözlemler belirlenmektedir. Bu değişim ölçüsüne dayalı olan yöntemler aşağıdaki alt bölümlerde tanıtılacaktır.

3.5.1. Genelleştirilmiş Andrews-Pregibon istatistiği

Bu ölçüt, Alt Bölüm 2.4.4.1'de verilen ve 1978 yılında Andrews ve Pregibon tarafından önerilen Andrews-Pregibon istatistiğinin çoklu kuşku gözlemleri belirlemek için elde edilen genelleştirilmiş biçimidir. Eşitlik (2.34)'ün genelleştirilmiş hali,

$$AP_I = 1 - \det(I - H_I) \left(1 - \frac{r_I^2}{n-p} \right) \quad (3.11)$$

biçiminde ifade edilir. En yüksek AP_I değerini alan altküme, çoklu kuşku gözlemlerin bulunduğu alt kümedir (Chatterjee & Hadi, 1988).

3.5.2. Genelleştirilmiş varyans oranı

Alt Bölüm 2.4.4.2'de verilen ve tekli kuşku gözlemlerin etkisini ölçen varyans oranı yönteminin çoklu kuşku gözlemlerin etkisini ölçmek için genelleştirilmiş hali,

$$VR_i = \frac{\det\{V[\hat{\beta}_{(i)}]\}}{\det\{V[\hat{\beta}]\}} \quad (3.12)$$

$$= \left(\frac{n-p-r_i^2}{n-p-m}\right)^p \frac{1}{\det(1-H_i)}$$

ile ifade edilir. Bu ölçüt, Eşitlik (2.35)'te ifade edilen ölçütün genelleştirilmiş halidir. En yüksek VR_i değerini alan altküme, çoklu kuşkulu gözlemlerin bulunduğu alt kümedir (Imon, 2005).

3.5.3. Genelleştirilmiş Cook-Weisberg istatistiği

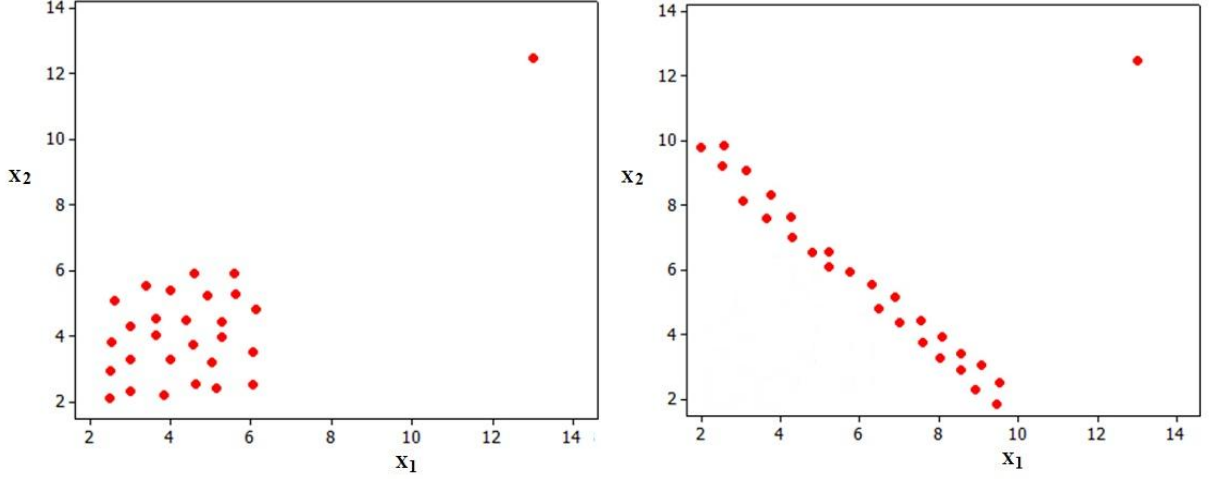
Alt Bölüm 2.4.4.3'te verilen Cook-Weisberg istatistiğinin çoklu kuşkulu gözlemlerin etkisini ölçmek için genelleştirilmiş hali,

$$CW_i = -\frac{1}{2} \log(VR_i) \times \frac{p}{2} \log\left(\frac{F_{(\alpha;p,n-p)}}{F_{(\alpha;p,n-p-m)}}\right) \quad (3.13)$$

biçimindedir. En yüksek CW_i değerini alan altküme, çoklu kuşkulu gözlemlerin bulunduğu alt kümedir.

3.6. Rankı Etkileyen Gözlemlere Dayalı Yöntem

Toplanan veriden bir gözlemin çıkarılması ile veri kümesinin özyapısı (eigenstructure) da önemli derecede değişebilmektedir. X eksenini yönünde aykırı değerler özyapıyı etkilemeye eğilimlidir. Bu noktalar çoklu bağlantıya sebep olabilir ya da çoklu bağlantıyı gizleyebilir (Kempthorne, 1989). Şekil 3.3'te verilen sol taraftaki grafikte çoklu bağlantıya sebep olan bir gözlem, sağ taraftaki grafikte çoklu bağlantıyı gizleyen bir gözlem görülmektedir.



Şekil 3.3 Çoklu bağlantıya sebep olan ve gizleyen gözlemler

Verinin özyapısını etkileyen bu gözlemlere çoklu bağlantılı-etkili (collinearity-influential) gözlemler denir. Bu gözlemlerin tespiti için 1989'da Kempthorne bir yöntem önermiştir. Bu önerilen yöntem ile, çoklu kuşkulu gözlemler belirlenebilmektedir. Bu yöntem, m alt küme büyüklüğünün belirlenmesine ihtiyaç duymamaktadır (Chatterjee & Hadi, 1988).

Veri kümesinden çıkarılan m gözlem I ile, H matrisindeki m gözlem ile ilişkili kısım H_I ile ifade ediliyor olsun. Eğer H_I matrisinin özdeğerleri $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m$ ve $\lambda_m = 1$ ise, $rank(X_{(I)}) < p$ olur. Yani I alt kümesinde bulunan m gözlem, eğer oluşturdukları H_I matrisinin en büyük özdeğeri 1'e eşitse, normalde p ranklı olan $X_{(I)}$ matrisinin rankının p 'den küçük olmasına sebep olacaklardır (Chatterjee & Hadi, 1988). Yani bu gözlemler rankı etkileyeceklerdir. Bu nedenle, I altkümesindeki gözlemlere rankı etkileyen gözlemler de denir.

H matrisinin büyük özdeğerlere sahip altkümelerini aramak, özellikle n ve m büyük olduğunda, çok fazla işlem yapılmasını gerektirir. Bu gözlemleri belirlemenin alternatif bir yolu aşağıdaki teorem ile verilmiştir.

Teorem : H_I matrisinin özdeğerleri $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_m$ ve X matrisinin kolon uzayı \mathfrak{R}_X olsun. Eğer $V = (v_1, v_2, \dots, v_n)' \in \mathfrak{R}_X$ sıfırdan farklı bir vektör ve $\forall i \notin I$ için $v_i = 0$ ise, $\lambda_m = 1$ olur (Kempthorne, 1989) .

Başka bir deyişle, eğer çoklu kuşku gözlemlerin bulunduğu bir veri kümesi var ise, \mathfrak{R}_X , birkaç elemanı büyük değerde ve diğer elemanları sıfır ya da sıfıra yakın değerleri olan vektörler içerir. Bu nedenle çoklu kuşku gözlemler, \mathfrak{R}_X 'teki bu tip vektörler araştırılarak belirlenebilir. Bu durumda \mathfrak{R}_X için bir taban oluşturan herhangi bir vektörler kümesi, \mathfrak{R}_X için birim dikey tabanlı (orthonormal bases) eşdeğer bir kümeye dönüştürülebilir. Bu nedenle çoklu kuşku gözlem araştırmasını sadece \mathfrak{R}_X 'in birim dikey tabanlarını incelemekle sınırlandırabiliriz. \mathfrak{R}_X için fazla sayıda birim dikey taban bulunabilir. Eğer B , \mathfrak{R}_X 'in herhangi bir birim dikey tabanı ve A herhangi bir $p \times p$ boyutlu birim dikey matris ise, $Q = B.A$ matrisi de \mathfrak{R}_X için bir birim dikey tabandır. Sonuç olarak araştırma, \mathfrak{R}_X için herhangi bir birim dikey taban bularak başlar. Örneğin B , \mathfrak{R}_X için birim dikey bir taban olsun. Daha sonra ise B 'yi, kolonları mümkün olduğunca çok sayıda sıfır ya da sıfıra yakın değerler ile az sayıda mutlak değerce büyük değerler içeren ve bir birim dikey taban olan Q 'ya dönüştürerek araştırma devam eder.

Birim dikey taban olan B matrisi, X matrisinin Tekil Değer Ayrıştırma (Singular Value Decomposition, SVD) yöntemi ile ayrıştırılmasından bulunan U matrisi olarak elde edilir. (Chatterjee & Hadi, 1988)

SVD yöntemine göre, eğer X tam ranklı ise,

$$X_{(n \times p)} = U_{(n \times p)} D_{(p \times p)} V'_{(p \times p)} \quad (3.14)$$

biçiminde yazılabilir ve $U'U = V'V = VV' = I$ dir. D matrisi, köşegen öğeleri $X'X$ (ya da XX') matrisinin sıfırdan farklı özdeğerlerinin karekök değerlerinden oluşan köşegen bir matristir. V matrisi, $X'X$ matrisinin standartlaştırılmış

özvektörlerinden elde edilir. U matrisi ise, XX' matrisinin standartlaştırılmış özvektörleri ile ilişkilidir.

Eşitlik (3.14)'dan U matrisi

$$U = X(V')^{-1}D^{-1} \quad (3.15)$$

ile elde edilir. Böylece B matrisi bulunmuş olur.

B matrisini istenilen Q matrisine dik olarak (orthogonally) dönüştürmenin çeşitli yolları vardır. Kempthorne, dönüştürme yollarından varimax-quartimax yöntemini önermiştir. Amaç $p \times p$ boyutlu dik bir matris olan A ile $Q = B.A$ matrisini elde etmektir. A matrisi,

$$\psi = \sum_{j=1}^k \psi_j \quad (3.16)$$

değerini en büyük yapan matristir.. Varimax yöntemine göre ψ_j , Q 'nun j . kolonunun elemanlarının karelerinin varyansdır ve

$$\psi_j = \frac{1}{n} \sum_{i=1}^n (q_{ij}^2)^2 - \left(\frac{1}{n} \sum_{i=1}^n q_{ij}^2 \right)^2 \quad (3.17)$$

biçiminde ifade edilir. Burada q_{ij} , Q matrisinin (i, j) 'nci elemanıdır. Quartimax yöntemine göre ise,

$$\psi_j = \frac{1}{n} \sum_{i=1}^n q_{ij}^4 \quad (3.18)$$

biçimindedir (Chatterjee & Hadi, 1988).

Q matrisi, kolonları mümkün olduğunca çok sayıda sıfır ya da sıfıra yakın değerler, az sayıda mutlak değerce büyük değerler içerdiği için, her kolonu incelenerek q_{ij} değerleri büyük olan gözlemlerin oluşturduğu altküme belirlenir. I

altkümesindeki gözlemlerin ilişkili olduğu Q matrisindeki elemanların Kempthorne tarafından verilen teoremi (Kempthorne, 1989) sağlaması gerekmektedir.

$$\lambda_m \geq \max_{j=1,2,\dots,p} \left\{ \sum_{i \in I} q_{ij}^2 \right\} \quad (3.19)$$

Eşitlik (3.19)'in sağ yanı ile belirlenen altküme çoklu kuşkulu gözlemlerin bulunduğu altkümedir.

Sonuç olarak özetle şu adımlar izlenir:

1. X matrisinin kolon uzayı olan \mathfrak{R}_X 'in bir birim dikey taban olan B matrisi elde edilir.
2. B matrisi varimax/quartimax dönüştürme yöntemleri uygulanarak Q matrisine dönüştürülür.
3. Q matrisinin her kolonu için; $(q_{ij})^2 \approx 0$ ise $i \notin I$ ve $(q_{ij})^2 > 0$ ise $i \in I$ koşulunu sağlayan altküme belirlenir.
4. I altkümesi için Eşitlik (3.19)'in sağ tarafı bire eşitse, I altkümesi çoklu kuşkulu gözlemlerin yer aldığı altkümedir.

Bu yöntem genelde olası bütün çoklu kuşkulu gözlem altkümelerini belirleyememektedir. Ayrıca 2. adımda ifade edilen varimax/quartimax dönüşümü sonucunda, kolonları mümkün olduğunca çok sayıda sıfır ya da sıfıra yakın değerler, az sayıda mutlak değerce büyük değerler içeren bir birim dikey taban olan matris elde edilemez ise bu yöntem çoklu kuşkulu gözlemleri belirlemede yetersiz kalmaktadır.

4. UYGULAMA

4.1. Giriş

Bir toplumda meydana gelen suçu açıklamak için literatürde birçok yöntem mevcuttur. Suç kavramı çok fazla etkene bağlı bir olgudur ve kaynağının ne olduğu, hangi durumlardan etkilendiği bir tartışma konusu olmuştur (Dolu, 2010). Bu çalışmada, suç teorilerinden biri olan *Braithwaite*'in 2005 yılında önerdiği "Birleştirici Utandırma Teorisi" yardımı ile bir toplumda meydana gelen suçu ve suç oranlarını açıklamak için regresyon çözümlemesi yapılacaktır.

Tez kapsamında ifade edilen doğrusal regresyon modeli kurularak, kuşkulu gözlemlerin incelemesine geçilecektir. Tekli ve çoklu kuşkulu gözlemlerin tespiti için önerilen yöntemler uygulanarak, gözlemlerin modele olan etkileri ayrıntılı olarak incelenecektir. Tez kapsamında tekli ve çoklu kuşkulu gözlemler ile ilgili ifade edilen yorumların gerçek bir uygulamada geçerli olup olmadığı gözlenecektir.

4.2. Modelin Oluşturulması

Çoklu doğrusal regresyon modelinde, çoklu kuşkulu gözlemlerin belirlenmesi ve model üzerindeki etkileri, bir uygulama ile daha açık bir şekilde görülebilir. Uygulama için *Braithwaite*'in 2005 yılında bir toplumda meydana gelen suçu ve suç oranlarını açıklamakta önerdiği "Birleştirici Utandırma Teorisi" ülkemizde meydana gelen asayiş suçlarını açıklamak için kullanılacaktır (Braithwaite, 2005).

Braithwaite, bir takım özelliklere sahip bireylerden oluşan toplumlarda karşılıklı bağlılığın arttığını, bu özelliklere sahip bireylerin az olması durumunda bireylerin arasındaki karşılıklı bağlılığın azalacağını ve bireylerin birbirinden kopacağını söylemiştir. Bu özellikler; 15 yaşından küçük ve 25 yaşından büyük olmak, bayan olmak, evli olmak, bir iş-güç sahibi olmak, eğitimde yüksek hedeflere sahip olmak. Bu özellikler toplumlara oluşturan bireyleri birbirine bağlayıcı bir işlev görür. Bu özelliklerin olmaması durumunda bireyler arasındaki karşılıklı bağlılık azalır ve bireyler birbirinden kopar. Bu özelliklerin yanı sıra şehirleşme ve nüfus hareketliliğinin artması da toplumculuğun azalmasına, yani bireylerin toplumun çıkarları yerine kendi çıkarlarını daha ön planda tutmalarına sebep olduğunu ifade

etmiştir. *Braithwaite*, bu nedenlerle bireyler arasındaki karşılıklı bağlılığın ve toplumsalluğun azaldığı yerlerde suç oranlarının artacağını öne sürmüştür (Dolu, 2010).

81 ildeki aynı yıla ait asayiş suç oranlarını bu teori yardımı ile açıklamak için, illerdeki asayiş suçu oranının (ASA) bağımlı değişken; erkek bireylerin oranının (ERK), 15-25 yaş nüfusun oranının (15-25YAS), bekarların oranının (BKR), işsizlik oranının (ISZ), lise ve dengi okullardan mezun olanlar ile daha düşük düzeyde eğitilmiş bireylerin oranının (EGT), şehirleşme endeksinin (S.END) ve toplam göç miktarının nüfusa oranının (GOC) açıklayıcı değişkenler olduğu ve daha önce suç teorileri alanında yapılan çalışmalar göz önüne alınarak asayiş suç oranını açıklayan modelin doğrusal bir model olduğu varsayılmıştır. Uygulamada açıklayıcı değişkenlerle ilgili bilgiler Türkiye İstatistik Kurumu'ndan ve bağımlı değişkene ait bilgiler de Emniyet Genel Müdürlüğü'nden alınmıştır.

Modelin yapısı belirlendikten sonra, varsayımların geçerliliğinin kontrolleri yapıldığında varsayımların geçerli olduğu, ancak çoklu bağlantı sorunu olduğu ortaya çıkmıştır. 15-25 YAS ile BKR açıklayıcı değişkenlerinin arasında çoklu bağlantı olduğundan dolayı BKR değişkeni modelden çıkarılarak çoklu bağlantı sorunu ortadan kaldırılmış ve model tekrar kurulmuştur.

$$ASA = \beta_0 + \beta_1 \times ERK + \beta_2 \times 15-25YAS + \beta_3 \times ISZ + \beta_4 \times EGT + \beta_5 \times S.END + \beta_6 \times GOC + \varepsilon \quad (3.20)$$

Bu model için belirtme katsayısı R^2 , % 45,3 ve modelin standart hatası 0,003321 olarak hesaplanmıştır. Çizelge 4.1'deki bağlantı katsayılarına ait tabloya bakıldığında sadece EGT değişkeninin anlamlı olduğu görülmektedir.

Çizelge 4.1 Bağlanım Katsayıları ve İlgili İstatistikler

	Sabit	ERK	15-25YAS	ISZ	EGT	S.END	GOC
Katsayı	0.158	-0.061	-0.037	0.000	-0.116	0.000	0.005
Std. Hata	0.043	0.046	0.030	0.000	0.039	0.001	0.019
t	3.669	-1.316	-1.223	1.324	-3.023	-0.087	0.278
Sig.	0.000	0.192	0.225	0.190	0.003	0.931	0.782

4.3. Tekli Kuşku Gözlemlerin Belirlenmesi

Model kurulumunun ardından kuşku gözlemlerin analizine geçilebilir. Öncelikle tekli kuşku gözlem tespit yöntemleri ile kuşku gözlem araştırmasına başlanır. Alt Bölüm 2.4'te verilen tekli kuşku gözlem tespit yöntemleri ile ilgili elde edilen değerler her bir gözlem için, Çizelge 4.2'de verilmiştir. Çizelge 4.2'de verilen tabloda tekli kuşku gözlem tespit yöntemlerine göre kuşku olan değerler gölgeli olarak verilmiştir.

Çizelge 4.2 Tekli Kuşku Gözlem Tespit Yöntemlerine İlişkin Elde Edilen Değerler

Göz. No	b_i	r_i	r_i^*	C_i	h_{ii}	WK_i	W_i	C_i^*	AP_i	$ 1-VR_i $
34	-1.0779	-1.6811	-1.7025	0.5783	0.5889	2.0376	28.0061	6.6250	0.5927	1.0371
6	-1.7705	-2.1432	-2.1980	0.3054	0.3176	1.4994	16.0888	4.8750	0.3483	0.0292
71	3.6008	3.7494	4.1377	0.1692	0.0777	1.2010	11.1116	3.9050	0.2429	0.7271
62	0.3668	0.5506	0.5480	0.0543	0.5563	0.6135	8.1256	1.9948	0.5458	1.4086
61	-1.6955	-1.7910	-1.8187	0.0531	0.1038	0.6190	5.8083	2.0125	0.1308	0.1000
24	1.8850	1.9593	1.9986	0.0441	0.0744	0.5665	5.2318	1.8419	0.1107	0.1816
31	-1.8252	-1.9011	-1.9361	0.0439	0.0783	0.5643	5.2226	1.8348	0.1116	0.1595
7	2.1209	2.1774	2.2355	0.0366	0.0512	0.5194	4.7385	1.6887	0.1004	0.2707
57	-1.9039	-1.9552	-1.9942	0.0299	0.0519	0.4664	4.2567	1.5165	0.0891	0.1997
29	-0.8966	-0.9797	-0.9794	0.0266	0.1625	0.4314	4.1852	1.4025	0.1612	0.1986
30	-0.9195	-0.9979	-0.9978	0.0253	0.1509	0.4207	4.0539	1.3678	0.1502	0.1782
26	0.9403	1.0137	1.0139	0.0239	0.1397	0.4086	3.9126	1.3286	0.1395	0.1593
27	1.1735	1.2347	1.2392	0.0233	0.0968	0.4057	3.7922	1.3190	0.1033	0.0527
45	1.7623	1.8051	1.8337	0.0229	0.0469	0.4066	3.7013	1.3220	0.0770	0.1580
35	1.0827	1.1423	1.1446	0.0211	0.1016	0.3850	3.6081	1.2517	0.1053	0.0810
76	1.1995	1.2485	1.2533	0.0186	0.0770	0.3621	3.3491	1.1774	0.0844	0.0267
73	-0.7791	-0.8404	-0.8387	0.0165	0.1405	0.3391	3.2478	1.1024	0.1365	0.1966
9	1.3555	1.3898	1.3988	0.0141	0.0487	0.3165	2.8841	1.0292	0.0615	0.0392
33	1.7889	1.8133	1.8424	0.0129	0.0268	0.3055	2.7520	0.9932	0.0582	0.1778
47	-0.8322	-0.8778	-0.8764	0.0124	0.1012	0.2941	2.7559	0.9563	0.0984	0.1373
22	-1.0433	-1.0806	-1.0818	0.0121	0.0678	0.2918	2.6851	0.9486	0.0704	0.0556
77	-0.7434	-0.7855	-0.7835	0.0103	0.1043	0.2673	2.5093	0.8692	0.0995	0.1581
58	1.4039	1.4268	1.4371	0.0096	0.0318	0.2606	2.3535	0.8471	0.0465	0.0655
4	0.5519	0.5951	0.5925	0.0082	0.1399	0.2390	2.2881	0.7769	0.1317	0.2366
11	-0.7228	-0.7561	-0.7539	0.0077	0.0862	0.2316	2.1526	0.7531	0.0811	0.1401
42	1.0696	1.0928	1.0942	0.0075	0.0419	0.2290	2.0789	0.7444	0.0453	0.0245
40	-0.7663	-0.7960	-0.7940	0.0072	0.0733	0.2233	2.0605	0.7259	0.0690	0.1175
43	-1.0065	-1.0292	-1.0296	0.0069	0.0437	0.2202	2.0012	0.7160	0.0453	0.0398
81	-1.0511	-1.0722	-1.0733	0.0067	0.0390	0.2162	1.9598	0.7028	0.0418	0.0257
39	-0.7199	-0.7489	-0.7466	0.0066	0.0758	0.2139	1.9766	0.6953	0.0706	0.1284
67	-1.0290	-1.0496	-1.0504	0.0064	0.0389	0.2112	1.9149	0.6868	0.0410	0.0303
14	-1.0568	-1.0765	-1.0777	0.0062	0.0363	0.2091	1.8931	0.6799	0.0392	0.0219
8	-0.9215	-0.9436	-0.9428	0.0062	0.0463	0.2077	1.8898	0.6752	0.0456	0.0596
17	-0.6403	-0.6692	-0.6667	0.0059	0.0844	0.2024	1.8789	0.6579	0.0776	0.1513

Çizelge 4.2 (devam) Tekli Kuşuklu Gözlem Tespit Yöntemlerine İlişkin Değerler

G. No	b_i	r_i	r_i^*	C_i	h_{ii}	WK_i	W_i	C_i^*	AP_i	$ I-VR_i $
72	0.8048	0.8283	0.8265	0.0058	0.0559	0.2011	1.8391	0.6538	0.0524	0.0915
52	-0.7149	-0.7367	-0.7344	0.0048	0.0581	0.1824	1.6700	0.5930	0.0528	0.1091
55	0.9925	1.0083	1.0084	0.0047	0.0311	0.1807	1.6319	0.5876	0.0323	0.0305
59	-0.5647	-0.5904	-0.5878	0.0047	0.0854	0.1796	1.6685	0.5839	0.0774	0.1635
36	-0.4393	-0.4696	-0.4671	0.0045	0.1248	0.1764	1.6747	0.5735	0.1151	0.2308
41	0.4920	0.5165	0.5140	0.0039	0.0929	0.1645	1.5341	0.5348	0.0839	0.1823
54	-0.8695	-0.8842	-0.8829	0.0038	0.0332	0.1636	1.4787	0.5319	0.0312	0.0562
79	0.6644	0.6830	0.6805	0.0038	0.0536	0.1620	1.4795	0.5266	0.0473	0.1119
49	-0.5950	-0.6151	-0.6125	0.0037	0.0643	0.1606	1.4752	0.5221	0.0568	0.1341
60	-0.6647	-0.6826	-0.6801	0.0037	0.0519	0.1592	1.4527	0.5175	0.0456	0.1100
25	-0.4550	-0.4788	-0.4763	0.0035	0.0968	0.1559	1.4570	0.5068	0.0872	0.1916
37	-0.5827	-0.6000	-0.5974	0.0031	0.1123	0.1465	1.3403	0.4763	0.0490	0.1269
75	0.3894	0.4133	0.4110	0.0031	0.0567	0.1462	1.3783	0.4753	0.1020	0.2192
16	0.5164	0.5341	0.5315	0.0028	0.0650	0.1402	1.2881	0.4557	0.0563	0.1451
69	0.4149	0.4350	0.4326	0.0027	0.0900	0.1360	1.2668	0.4423	0.0800	0.1872
70	0.7986	0.8089	0.8070	0.0024	0.0400	0.1302	1.1719	0.4232	0.0217	0.0605
53	0.6262	0.6391	0.6365	0.0024	0.0254	0.1300	1.1792	0.4227	0.0331	0.1023
65	-0.4585	-0.4745	-0.4720	0.0023	0.0662	0.1256	1.1553	0.4085	0.0567	0.1530
2	-0.3973	-0.4132	-0.4109	0.0020	0.0753	0.1173	1.0834	0.3812	0.0651	0.1704
46	0.4968	0.5097	0.5071	0.0020	0.0500	0.1163	1.0607	0.3783	0.0410	0.1297
63	-0.3512	-0.3674	-0.3652	0.0018	0.0863	0.1122	1.0431	0.3649	0.0756	0.1885
78	0.6365	0.6461	0.6435	0.0018	0.0295	0.1121	1.0116	0.3646	0.0227	0.0893
64	-0.5023	-0.5131	-0.5105	0.0016	0.0415	0.1063	0.9647	0.3455	0.0326	0.1193
80	0.3748	0.3884	0.3862	0.0016	0.0689	0.1050	0.9670	0.3415	0.0584	0.1645
5	-0.5854	-0.5940	-0.5914	0.0015	0.0288	0.1018	0.9177	0.3309	0.0211	0.0952
20	0.4300	0.4408	0.4384	0.0014	0.0486	0.0991	0.9029	0.3222	0.0388	0.1350
48	0.2697	0.2853	0.2836	0.0014	0.1063	0.0978	0.9190	0.3180	0.0950	0.2213
28	-0.4972	-0.5058	-0.5032	0.0013	0.0337	0.0939	0.8492	0.3054	0.0247	0.1110
21	0.3322	0.3434	0.3413	0.0012	0.0640	0.0893	0.8200	0.2903	0.0532	0.1621
51	0.5378	0.5442	0.5416	0.0010	0.1127	0.0838	0.7532	0.2723	0.0150	0.0951
32	-0.2218	-0.2355	-0.2340	0.0010	0.0234	0.0834	0.7861	0.2710	0.1010	0.2331
66	-0.3625	-0.3713	-0.3692	0.0010	0.0470	0.0820	0.7461	0.2665	0.0364	0.1392
18	0.0893	0.1100	0.1092	0.0009	0.3411	0.0786	0.8580	0.2555	0.3288	0.6674
19	0.3443	0.3509	0.3488	0.0007	0.0373	0.0687	0.6223	0.2234	0.0266	0.1294
1	0.1539	0.1641	0.1630	0.0005	0.1199	0.0602	0.5695	0.1956	0.1079	0.2466
10	0.2443	0.2510	0.2494	0.0005	0.0524	0.0587	0.5355	0.1907	0.0409	0.1539
12	-0.2079	-0.2149	-0.2135	0.0005	0.0644	0.0560	0.5146	0.1821	0.0527	0.1705
13	-0.2112	-0.2172	-0.2158	0.0004	0.0553	0.0522	0.4776	0.1699	0.0436	0.1592
44	0.2633	0.2680	0.2664	0.0004	0.0733	0.0507	0.4582	0.1647	0.0235	0.1320
56	-0.1747	-0.1815	-0.1803	0.0004	0.0349	0.0507	0.4680	0.1649	0.0614	0.1832
23	0.1453	0.1514	0.1503	0.0003	0.0787	0.0440	0.4069	0.1429	0.0667	0.1914
74	-0.1445	-0.1484	-0.1474	0.0002	0.0520	0.0345	0.3153	0.1123	0.0400	0.1579
50	0.1523	0.1550	0.1540	0.0001	0.0343	0.0290	0.2625	0.0944	0.0223	0.1364
3	-0.1178	-0.1197	-0.1189	0.0001	0.0327	0.0214	0.1933	0.0696	0.0192	0.1340
38	0.1190	0.1210	0.1202	0.0001	0.0314	0.0221	0.1997	0.0719	0.0205	0.1355
68	-0.0590	-0.0599	-0.0594	0.0000	0.0292	0.0103	0.0930	0.0335	0.0169	0.1326
15	-0.0122	-0.0124	-0.0123	0.0000	0.0413	0.0026	0.0232	0.0083	0.0290	0.1473

Kuşkulu gözlem tespit yöntemleri için kritik değerler hesaplanarak Çizelge 4.3'te verilmiştir.

Çizelge 4.3 Tekli Kuşkulu Gözlem Belirleme Yöntemlerine Ait Kritik Değerler

Yöntem	Kritik Değer
ri^*	3.98
hi	0.1728
Ci	2.14
$WK i$	0.59
Wi	0.794
Ci^*	1.91
$AP i$	0.198
$ I-VR i $	0.2592

Çizelge 4.2'de elde edilen değerler, Çizelge 4.3'te verilen kritik değerler ile karşılaştırılmıştır. Çizelge 4.2'de, Alt Bölüm 2.4.1'de verilen artıklara dayalı yöntemlere göre, b_i değeri $[+2,-2]$ aralığı dışında olan 71 ve 7 nolu gözlemler, r_i değeri $[-3, +3]$ aralığı dışında kalan 71 nolu gözlem ve r_i^* değeri $F_{(\alpha;1,n-p-1)} = F_{(0,05;1,73)} \cong 3,98$ kritik noktasından büyük olan 71 nolu gözlem kuşkulu gözlem olarak belirlenmiştir.

Alt Bölüm 2.4.2'de verilen gözlem uzaklığı matrisine dayalı yöntemine göre, gözlem uzaklık değeri $2p/n = 14/81 = 0,1728$ kritik değerinden büyük olan 34, 62, 18 ve 6 nolu gözlemler kuşkulu gözlem olarak tespit edilmiştir.

Alt Bölüm 2.4.3.1'de verilen güven elipsoidinin merkezine dayalı yöntemlerden biri olan Cook uzaklığı yöntemine göre, C_i değeri $F_{(\alpha;p,n-p)} = F_{(0,05;7,74)} \cong 2.14$ kritik değerinden büyük hiçbir gözlem olmadığı için hiçbir gözlem kuşkulu olarak tespit edilememiştir.

Alt Bölüm 2.4.3.2'de verilen Welsch uzaklığı yöntemine göre, W_i değeri $3\sqrt{p} \cong 0,794$ kritik değerinden büyük olan 34, 6, 71 ve 62 nolu gözlemler kuşkulu gözlem olarak belirlenmişlerdir.

Alt Bölüm 2.4.3.3.'te verilen Welsch-Kuh uzaklığı yöntemine göre, WK_i değeri $2\sqrt{p/n} \cong 0,59$ kritik değerinden büyük olan 34, 6 71, 61 ve 62 nolu gözlemler kuşkulu gözlemlerdir.

Alt Bölüm 2.4.3.4'te verilen düzeltilmiş Cook uzaklığı yöntemine göre, C_i^* değeri $2\sqrt{(n-p)/n} \cong 1,91$ kritik değerinden büyük olan 34, 6 71, 61 ve 62 nolu gözlemler kuşkulu gözlemlerdir.

Alt Bölüm 2.4.4'te verilen güven elipsoidinin hacmine dayalı yöntemlere göre; AP_i değeri $2(p+1)/n \cong 0,198$ kritik değerinden büyük olan 34, 62, 18, 6 ve 71 nolu gözlemler ve $|VR_i - 1|$ değeri $3p/n \cong 0,2592$ kritik değerinden büyük olan 62, 34, 71, 18 ve 7 nolu gözlemler kuşkulu gözlem olarak tespit edilmiştir.

Tekli kuşkulu gözlemlerin belirlenmesinde kullanılan yöntemler ve bu yöntemlerin belirledikleri gözlemler toplu olarak Çizelge 4.4'te verilmiştir. Çizelge 4.4'te gözlemlerin yöntemlere göre aldıkları değerlere bakıldığında kritik değere yakın değer alanlar, yani verilerin çoğunluğundan aşırı biçimde ayrı olmayanlar ile kritik değerden uzak değer alanlar farklı olarak işaretlenmiştir. Çizelge 4.4'te, + işareti ile kritik değere yakın olan ve * işareti ile kritik değere uzak olan değerler gösterilmiştir. Kritik değerden daha uzakta değer alan gözlemlerin etkileri daha fazla olacağı düşünülmektedir. Artıklara dayalı yöntemlere göre 71 ve 7 nolu gözlemler, gözlem uzaklıkları matrisine dayalı yöntemlere göre 6, 34, 18 ve 62 nolu

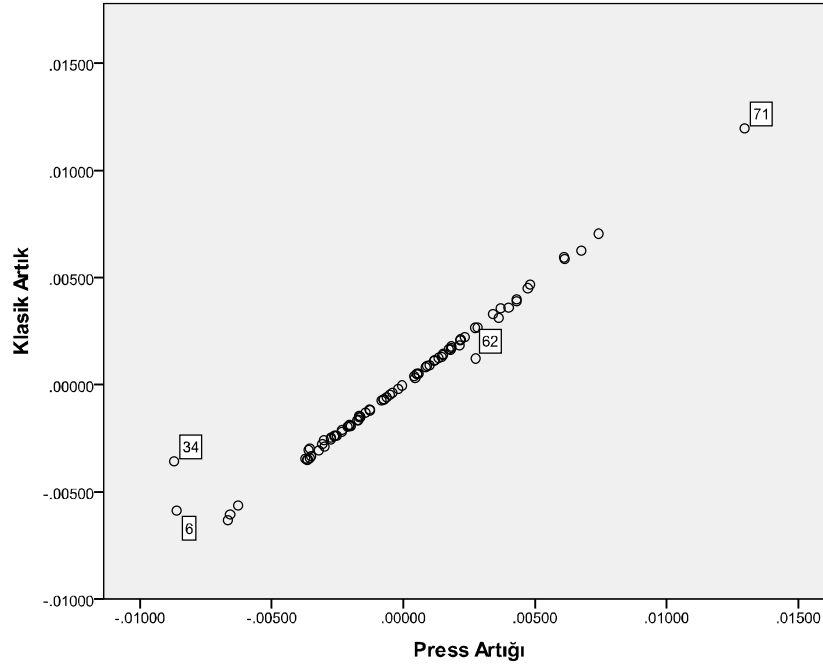
Çizelge 4.4 Tekli Kuşkulu Gözlem Belirleme Yöntemlerine Göre Belirlenen Kuşkulu Gözlemler

No	bi	ri	ri^*	Ci	$hü$	WK_i	Wi	Ci^*	AP_i	VR_i
6					*	*	*	*	*	
34					*	*	*	*	*	*
71	*	*	*			*	*	*	+	*
18					*				*	*
62					*	+	+	+	*	*
7	+									+
61						+		+		

*: Kritik değerden uzakta + : Kritik değere yakın

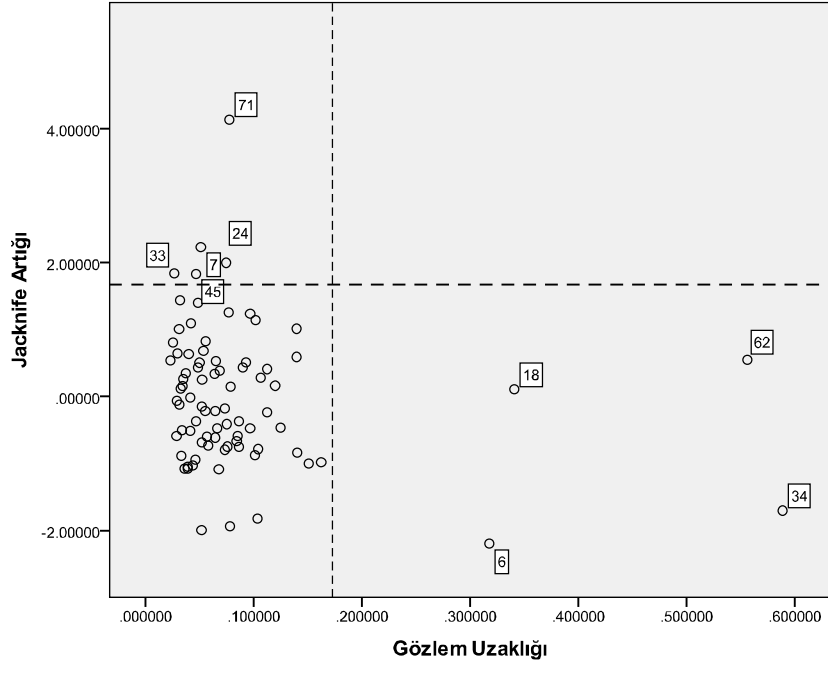
gözlemler, güven elipsoidinin merkezindeki değişime dayalı yöntemlere göre 6, 34, 71, 62 ve 61 nolu gözlemler ve güven elipsoidinin hacmine dayalı yöntemlere göre 6, 34, 71, 18, 62 ve 7 nolu gözlemler kuşkulu gözlem olarak belirlenmiştir. Cook uzaklığı yöntemi herhangi bir kuşkulu gözlem belirleyememiştir. Genel olarak tekli kuşkulu gözlem belirleme yöntemleri ile 6, 34, 71, 18 ve 62 nolu gözlemler modeli en çok etkileyen gözlemler olarak tespit edilmiştir.

Ayrıca grafik yöntemleri kullanarak da kuşkulu gözlemler belirlenebilmektedir. Alt Bölüm 2.4.1'de verilen Press artıklarının klasik artıklara karşı grafiği çizilirse, Şekil 4.1'de verilen grafik elde edilir. Bu grafiğe göre 6 ve 71 nolu gözlemler aykırı değer, 34 ve 62 nolu gözlemler X ekseninde aykırı değer olarak tespit edilmiştir.

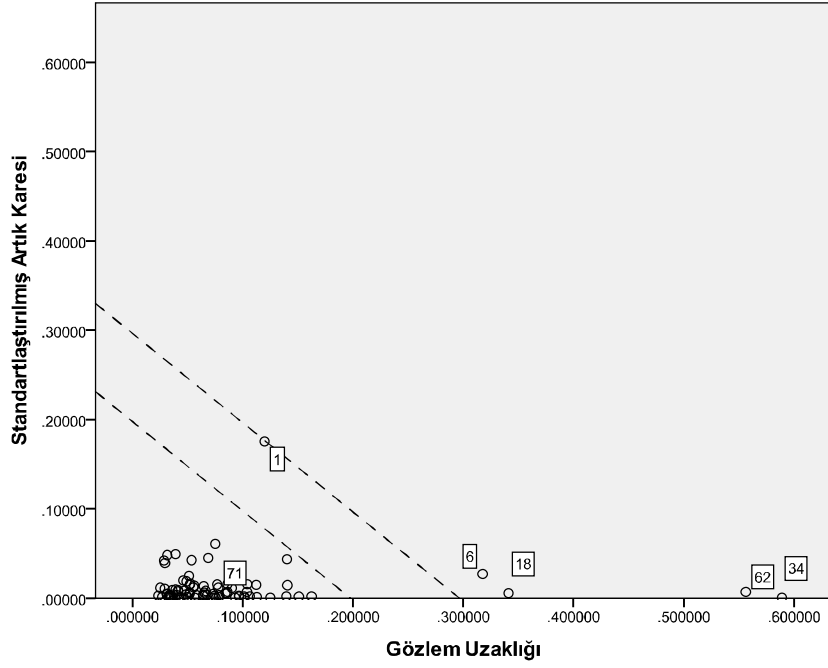


Şekil 4.1 Press Artığı ile Klasik Artık grafiği

Alt Bölüm 2.4.2'de verilen Williams grafiği ve Pregibon grafiği, Şekil 4.2'de ve Şekil 4.3'teki gibi elde edilmiştir. Şekil 4.2'de verilen Williams grafiğine göre, 71, 24, 7, 33 ve 45 nolu gözlemler aykırı değer, 62, 34, 18 ve 6 nolu gözlemler X ekseninde aykırı değer olarak tespit edilmiştir.



Şekil 4.2 Williams Grafiği



Şekil 4.3 Prejibon Grafiği

Şekil 4.3'te verilen Pregibon grafiğine göre ise, 34, 62, 18 ve 6 nolu gözlemler çok etkili ve 1 nolu gözlem az etkili gözlem olarak belirlenmiştir. Grafik yöntemler kullanılarak 6, 34, 71, 62 ve 18 nolu gözlemlerin verilerin çoğunluğundan uzakta yer alan gözlemler olarak saptanmıştır.

Sonuç olarak tekli kuşkulu gözlem tespit yöntemleri ve grafik yöntemlerine bakıldığında ortak olarak 6, 18, 34, 62 ve 71 nolu gözlemler tekli kuşkulu gözlemler olarak tespit edilmişlerdir.

4.4. Çoklu Kuşkulu Gözlemlerin Belirlenmesi

Bölüm 3'de çoklu kuşkulu gözlemlerin etkisinin tekli kuşkulu gözlemlerin etkisinden daha fazla olduğundan bahsedilmişti. Çoklu kuşkulu gözlemlerin modele olan etkilerini tekli kuşkulu gözlemlerin etkileri ile karşılaştırabilmek için öncelikle Bölüm 3'te verilen çoklu kuşkulu gözlem belirleme yöntemleri ile çoklu kuşkulu gözlemlerin belirlenmesi gerekmektedir.

4.4.1. İkili kuşkulu gözlemlerin belirlenmesi ve modele etkileri

Bu uygulamada gözlem çiftlerinden oluşan ikili altkümelerin çoklu kuşkulu gözlem tespit yöntemlerinden Alt Bölüm 3.2.1'de verilen genelleştirilmiş Cook uzaklığı yöntemi ile aldıkları değerler karşılaştırılacaktır. İkili altkümelerin sayısı oldukça fazla olacağı için MINITAB yazılımında çalıştırılan MULTDIST makrosu yardımı ile tüm ikili altkümelerin aldıkları genelleştirilmiş Cook uzaklık değerleri hesaplanmıştır. $81!/[2!(81-2)!]=3240$ adet ikili altkümenin C_1 değerleri hesaplanarak, en yüksek değerleri alan ikililer Çizelge 4.5'te verilmiştir.

Çizelge 4.5 En Yüksek Genelleştirilmiş Cook Uzaklığı Alan İkili Gözlemler

i	j	Cook uzaklığı
6	34	2.36
34	59	0.84
34	62	0.82
11	34	0.77
34	58	0.73
9	34	0.71
...

Çizelge 4.5'te verilen tabloya göre, genelleştirilmiş Cook uzaklığı yöntemine göre 6 ve 34 nolu gözlemler diğer 3239 ikili alt kümeden daha yüksek C_i değerine sahiptir. 6 ve 34 nolu gözlemler birlikte kuşkulu gözlemlerdir. 6 ve 34 nolu gözlemler tekli kuşkulu tespit yöntemlerinden Cook uzaklığı yöntemi ile ayrı ayrı kuşkulu gözlem olarak tespit edilememişlerdir. Bunun nedeni ise gizleme etkisi olabilir. Bunu belirlemek için birlikte kuşkulu olan gözlemlerden birisini veri setinden çıkarıp diğerinin yeni kurulan modelde alacağı C_i değerine bakmaktır. Eğer C_i değeri, önceden aldığı değerden fazla ise çıkarılan gözlemin varlığı diğer gözlemin etkisini gizlediği söylenebilir. Ek 1'de verilen tablodan tekli kuşkulu gözlemler için Cook uzaklıkları $C_6 = 0,305$ ve $C_{34} = 0,578$ olarak verilmişti. 6 nolu gözlem veri setinden çıkarılıp model tekrar kurulduğu zaman 34 nolu gözlemin aldığı Cook uzaklığı değeri 0,578'den 1,39'a çıkmaktadır. 34 nolu gözlem veri setinden çıkarıldığında yeni kurulan modelde 6 nolu gözlemin Cook uzaklığı değeri 0,305'ten 0,644'e çıkmaktadır. Görüldüğü gibi 6 ve 34 nolu gözlemlerden birinin modelden çıkması, diğerinin modeldeki etkisinin artmasına neden olmaktadır. Bu durum gözlemlerin birbirlerinin etkilerini gizlemelerinden kaynaklanmaktadır.

Çoklu kuşkulu gözlemlerin modele olan etkilerinin tekli kuşkulu gözlemlerin etkisinden daha fazla olduğunu göstermek amacı ile, 6 ve 34 nolu gözlemlerin ayrı ayrı ve birlikte olduklarında modele yaptıkları etkileri karşılaştıralım. Bir gözlemin modelde neleri etkileyebileceğini, o gözlemin veri setinde bulunduğu oluşturulan modelin bilgileri ile çıktıktan sonra oluşturulan modelin bilgilerini karşılaştırılarak belirlenebilir

6 nolu gözlem veri setinden çıkarıldığı zaman, modelin belirtme katsayısı % 48,2 ve modelin standart hatası 0,003238 olmaktadır. Model parametrelerinde meydana gelen değişiklikler Çizelge 4.6'da verilen tabloda gölgeli olarak işaretlenmiştir.

Çizelge 4.6 {6} Nolu Gözlem Çıkarıldığındaki Model Parametrelerine Ait Bilgiler

	Sabit	ERK	15-25YAS	ISZ	EGT	S.END	GOC
Katsayı	0.195	-0.081	-0.013	0.000	-0.150	0.000	0.003
Std. Hata	0.045	0.046	0.031	0.000	0.041	0.001	0.019
t	4.316	-1.760	-0.405	1.269	-3.697	0.029	0.175
Sig.	0.000	0.083	0.687	0.208	0.000	0.977	0.861

Çizelge 4.6'da verilen tabloya göre 6 nolu gözlemin modelden çıkarılması üç regresyon katsayısının değerlerini bir miktar değiştirmiştir.

34 nolu gözlem veri setinden çıkarıldığı zaman, modelin belirtme katsayısı % 47,4 ve modelin standart hatası 0,003279 olmaktadır. Model parametrelerindeki meydana gelen değişiklikler Çizelge 4.7'de verilen tabloda gölgeli olarak işaretlenmiştir.

Çizelge 4.7 {34} Nolu Gözlem Çıkarıldığında Model Parametrelerine Ait Bilgiler

	Sabit	ERK	15-25YAS	ISZ	EGT	S.END	GOC
Katsayı	0.119	-0.068	-0.021	0.000	-0.074	0.001	0.014
Std. Hata	0.048	0.046	0.031	0.000	0.045	0.001	0.020
t	2.459	-1.487	-0.672	0.833	-1.639	1.220	0.718
Sig.	0.016	0.141	0.503	0.408	0.106	0.226	0.475

Çizelge 4.7'de verilen tabloya göre, 34 nolu gözlem modelden çıkarıldığında dört regresyon katsayısının değerleri ve bir regresyon katsayısının standart hatası bir miktar değişmektedir.

6 ve 34 nolu gözlemler aynı anda veri setinden çıkarılıp model yeniden kurulduğu zaman, modelin belirtme katsayısı % 52,5'e çıkmakta ve modelin standart hatası 0,003124'e düşmektedir. 6 ve 34 nolu gözlemler modelden birlikte çıkarıldığı zaman model bilgileri üzerinde meydana gelen değişiklikler Çizelge 4.8'de verilen tabloda gölgeli olarak işaretlenmiştir.

Çizelge 4.8 {6, 34} Nolu Gözlemler Çıkarıldığındaki Model Parametrelerine Ait Bilgiler

	Sabit	ERK	15-25YAS	ISZ	EGT	S.END	GOC
Katsayı	0.150	-0.098	0.019	0.000	-0.098	0.002	0.016
Std. Hata	0.047	0.045	0.033	0.000	0.044	0.001	0.019
t	3.178	-2.191	0.569	0.526	-2.235	1.962	0.832
Sig.	0.002	0.032	0.571	0.601	0.028	0.054	0.408

Çizelge 4.8'de verilen tabloya göre, 6 ve 34 nolu gözlemlerin birlikte modelden çıkarılması, modele ait bilgilerin çoğunu etkilemiştir. Dört regresyon katsayısının değerleri ve iki tanesinin standart hatası değişmiştir. Ayrıca önceden sadece bir değişkenin modele katkısı anlamlı iken 6 ve 34 nolu gözlemler birlikte çıkarıldığında üç değişkenin modele katkısı anlamlı olmaktadır. Görüldüğü gibi çoklu kuşkulu gözlemlerin modele etkisi, tekli kuşkulu gözlemlerin modeldeki etkilerinden daha fazladır. Bu nedenle çoklu kuşkulu gözlemlerin belirlenmesi oldukça önemlidir. Bu durumda model denklemi aşağıdaki biçimde elde edilmiştir.

$$ASA = 0,15 - 0,098 \times ERK + 0,019 \times 15-25YAS + 0,000 \times ISZ - 0,098 \times EGT + 0,002 \times S.END + 0,016 \times GOC + \varepsilon \quad (3.21)$$

4.4.2. Geri eleme yöntemi ile çoklu kuşkulu gözlemlerin belirlenmesi

Çoklu kuşkulu gözlemlerin tespit edilmesi işlemine alt küme büyüklüğü artırılarak devam edilebilmektedir. Fakat alt küme büyüklüğü arttıkça yapılması gereken işlem sayısı artacağı için hesaplama yapmak bir süre sonra mümkün olmayacaktır. Uygulamada 81 gözlemden oluşan $81!/[3!(81-3)!] = 85320$ tane üçlü altkümenin genelleştirilmiş Cook uzaklığı değerleri, MINITAB yazılımında yer alan MULTDIST makrosu yardımı ile güçlü bir bilgisayar üzerinde hesaplatılmıştır. Yapılan hesaplama işlemleri 24 saat sürdükten sonra işlemler tamamlanamadan yazılım hata vermiştir. Bu durumdan da anlaşılacağı gibi uygulamada olası bütün altkümelerin çoklu kuşkulu gözlem tespit yöntemleri ile hesaplanması çok zordur. Bu nedenle işlem yükünü hafifletmek amacı ile geliştirilen yöntemler kullanılmalıdır. Bu yöntemlerden birisi Alt Bölüm 3.2'de bahsedilen geri eleme yöntemidir.

MINITAB yazılımında bulunan MULTCASE makrosu yardımı ile geri eleme yöntemi kullanılarak istenilen alt küme büyüklüğünde en yüksek genelleştirilmiş Cook uzaklığı değerine sahip alt kümeler bulunabilmektedir. Üçlü ve dördü kuşku gözlemleri bulmak amacı ile yapılan yinelemelerde elde edilen sonuçlar Çizelge 4.9'da verilmiştir.

Çizelge 4.9 Geri Eleme Yöntemi İle Hesaplanan Çoklu Kuşku Gözlemlerin Genelleştirilmiş Cook Uzaklığı Değerleri

<i>I</i>	<i>C_i</i>	<i>I</i>	<i>C_i</i>
6 34 41	3.04	6 34 41 59	3.97
6 34 59	2.86	6 9 34 80	2.85
6 34 62	2.56	6 18 34 62	2.66
6 34 18	2.41	6 34 42 60	2.23
6 34 71	1.77	6 34 41 71	2.12
6 32 61	0.97	24 34 62 67	1.57
18 71 79	0.46	26 33 35 71	0.46
6 18 62	0.38	18 29 57 77	0.41

Çizelge 4.9'da verilen tabloda geri eleme yöntemi ile yapılan yinelemelerde elde edilen en yüksek genelleştirilmiş Cook uzaklığına sahip üçlü ve dördü altkümeler verilmiştir. Üçlü kuşku gözlem olarak 6, 34, 41 altkümesi tespit edilmiştir. Dördü kuşku gözlem tespiti yapıldığında 6, 34, 41 ve 59 nolu gözlemlerin oluşturduğu alt kümenin çoklu kuşku olduğu tespit edilmiştir. Dördü kuşku gözlem altkümesinin elemanlarından 6 ve 34 nolu gözlemler daha önce tekli kuşku gözlem tespiti yöntemleri ile kuşku olarak tespit edilmişlerdi. Ancak, Çizelge 4.2'de verilen tabloya bakıldığında 41 ve 59 nolu gözlemlerin tekli kuşku gözlem tespit yöntemleri ile kuşku gözlem olarak belirlenmedikleri görülmektedir. 41 ve 59 nolu gözlemler ayrı ayrı olarak modeli önemli derecede etkilemediği halde, 34 ve 6 nolu gözlemler ile birlikte olduklarında modeli önemli derecede etkilemektedirler. Ayrıca Çizelge 4.9'a bakıldığında, sadece tekli kuşku gözlem belirleme yöntemleri ile kuşku olarak belirlenen gözlemlerden oluşan dördü alt kümelerin daha düşük C_i değerine sahip oldukları görülmektedir. Bu uygulamadan da anlaşılacağı gibi çoklu kuşku gözlemler sadece tekli kuşku gözlem belirleme yöntemleri ile kuşku gözlem olarak belirlenen gözlemlerden oluşmamaktadır. Dördü kuşku gözlemin etkisi, gözlemler modelden birlikte çıkarılarak, modeldeki

değişimden gözlemlenebilmektedir. 6, 34, 41 ve 59 nolu gözlemler birlikte veri setinden çıkarılarak model tekrar kurulduğunda modelin belirtme katsayısı %53.3'e yükselmekte ve modelin standart hatası 0,003123'e düşmektedir. 6, 34, 41 ve 59 nolu gözlemler modelden birlikte çıkarıldığı zaman model bilgileri üzerinde meydana gelen değişiklikler Çizelge 4.10'da verilen tabloda gölgeli olarak işaretlenmiştir.

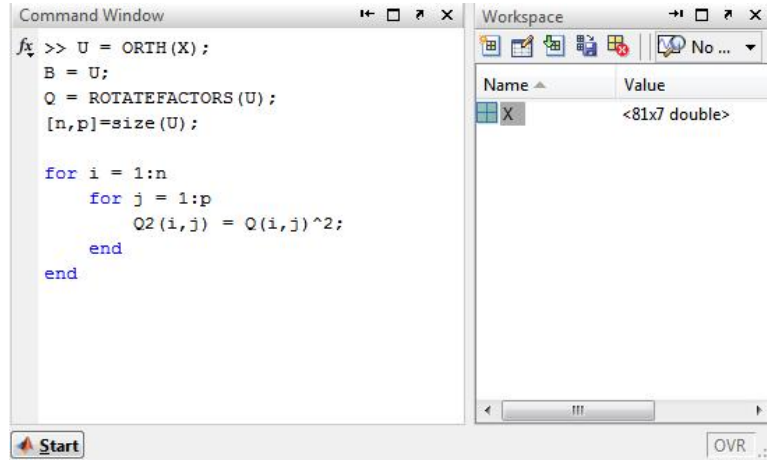
Çizelge 4.10 {6,34,41,59} altkümesi çıkarıldığında model parametrelerine ait bilgiler

	Sabit	ERK	15-25YAS	ISZ	EGT	S.END	GOC
Katsayı	0.121	-0.091	0.022	0.000	-0.073	0.003	0.023
Std. Hata	0.052	0.045	0.033	0.000	0.048	0.001	0.020
t	2.350	-2.002	0.659	0.295	-1.513	2.332	1.156
Sig.	0.022	0.049	0.512	0.769	0.135	0.023	0.252

Çizelge 4.10'da verilen tabloda görüldüğü gibi 6, 34, 41 ve 59 nolu gözlemler modelden birlikte çıkarılması modelin genelinde büyük bir değişime neden olmakla birlikte, 15-25YAS değişkenine ait katsayı kestirimi negatif bir sayıdan pozitif bir sayıya değişmiştir. 6, 34, 41 ve 59 nolu gözlemlerin modelden çıkarılması, daha önce modele anlamlı katkısı olan EGT değişkeninin modeldeki katkısının anlamsız olmasına, ERK ve S.END değişkenlerinin modele katkılarının anlamlı olmasına yol açmıştır. Uygulama sonuçlarından da görüleceği gibi çoklu kuşkulu gözlemlerin modele etkileri tahmin edilememekte ve tekli kuşkulu gözlemlerin modele etkilerinden daha fazla bir etkiye sebep olmaktadır.

4.4.3. Rankı etkileyen gözlemlere dayalı yöntem ile çoklu kuşku gözlemlerin belirlenmesi

Alt Bölüm 3.6'da ayrıntılı olarak anlatılan rankı etkileyen gözlemlere dayalı yöntem ile, I altkümesinde yer alan gözlemlerin sayısını belirlemeden çoklu kuşku gözlemler belirlenebilmektedir. Yöntemin uygulanabilmesi için MATLAB yazılımı kullanılmıştır. Alt Bölüm 3.6'da ifade edilen yöntemin aşamaları Şekil 4.4'te verilen kod yardımı ile X matrisi üzerine uygulanmıştır.



```
Command Window
>> U = ORTH(X);
B = U;
Q = ROTATEFACTORS(U);
[n,p]=size(U);

for i = 1:n
    for j = 1:p
        Q2(i,j) = Q(i,j)^2;
    end
end

Workspace
Name Value
X <81x7 double>
```

Şekil 4.4 Rankı Etkileyen Gözlemlere Dayalı Yönteme Ait Kod Penceresi

MATLAB yazılımında çalıştırılan kod yardımı ile, öncelikler X matrisinin bir birim temel tabanı olan U matrisi bulunmuştur. U matrisi B matrisi ile eşitlenmiştir. Daha sonra X matrisinin elde edilen birim temel tabanı olan U matrisi, varimax dönüştürme yöntemi ile kolonları mümkün olduğunca çok sayıda sıfır ya da sıfıra yakın değerler ile az sayıda büyük değerler içeren ve bir birim dikey taban olan Q matrisi elde edilmeye çalışılmıştır. Elde edilen Q matrisi Çizelge 4.12'de verilmiştir. Daha sonra ise Q matrisinin elemanlarının kareleri alınarak $Q2$ matrisi elde edilmiştir. Çoklu kuşku gözlemlerin belirlenmesi amacı ile $Q2$ matrisinde çok sayıda sıfır ya da sıfıra yakın değerler ile toplamları bire yakın olan az sayıda değerlerin bulunduğu kolon belirlenmelidir. Elde edilen $Q2$ matrisi Çizelge 4.13'te verilmiştir.

Çizelge 4.13'te verilen $Q2$ matrisine bakıldığında, kolonları çok sayıda sıfır ve sıfıra yakın değerler içeren bir kolon ilk bakışta tespit edilememektedir. Bu nedenle

Q2 matrisinin elemanları bir ondalık basamaklı sayılar haline yuvarlanarak, kolonlara tekrar bakılmıştır. Bu durumda 2 numaralı kolondaki 6, 34, 41, 35 ve 16 numaralı gözlemler haricindeki gözlemlerin sıfır ve sıfıra yakın değerler içerdiği görülebilmektedir. 6, 34, 41, 35 ve 16 numaralı gözlemler $(q_{ij})^2 > 0$ koşulunu sağladığı için I alt kümesinin elemanlarını oluştururlar. $\sum_{i \in I} q_{ij}^2 \cong 0,817 \approx 1$ değerinin bire yakın olduğu varsayımı altında I alt kümesinde bulunan gözlemler çoklu kuşkulu gözlemlerdir.

I alt kümesinde bulunan gözlemlerin model üzerindeki etkisini belirlemek için, gözlemler veri setinden çıkarılarak model yeniden kurulur. Yeni kurulan modelin belirtme katsayısı % 48,7'e yükselmiş ve modelin standart hatası 0,003175' düşmüştür. Gözlemler modelden birlikte çıkarıldığı zaman model bilgileri üzerinde meydana gelen değişiklikler Çizelge 4.11'te verilen tabloda gölgeli olarak işaretlenmiştir.

Çizelge 4.11 {6, 34, 41, 35, 16} Nolu Gözlemler Çıkarıldığındaki Model Parametrelerine Ait Bilgiler

	Sabit	ERK	15-25YAS	ISZ	EGT	S.END	GOC
Katsayı	0.134	-0.103	0.030	0.000	-0.081	0.003	0.020
Std. Hata	0.052	0.046	0.037	0.000	0.049	0.002	0.020
t	2.601	-2.231	0.807	0.278	-1.636	1.941	0.994
Sig.	0.011	0.029	0.422	0.782	0.106	0.056	0.324

Çizelge 4.13'te verilen tabloya göre 6, 34, 41, 35 ve 16 numaralı gözlemlerin modelden birlikte çıkarılması modelin genelinde büyük bir değişime neden olmakla birlikte, 15-25YAS değişkenine ait katsayı kestirimi negatif bir sayıdan pozitif bir sayıya değişmiştir. 6, 34, 41, 35 ve 16 numaralı gözlemlerin modelden çıkarılması, daha önce modele anlamlı katkısı olan EGT değişkeninin modeldeki katkısının anlamsız olmasına, ERK ve S.END değişkenlerinin modele katkılarının anlamlı olmasına yol açmıştır.

Çizelge 4.12 Rankı Etkileyen Gözlemlere Dayalı Yöntem İle Elde Edilen Q Matrisi

	1	2	3	4	5	6	7
1	0.1690	-0.1160	-0.0406	0.0650	0.2440	-0.1153	0.0059
2	0.2355	0.0152	0.0419	-0.0058	0.1014	0.0386	0.0791
3	0.0365	-0.0008	0.1642	-0.0241	-0.0235	-0.0050	0.0468
4	0.1227	-0.0176	0.0885	-0.0990	-0.3219	0.0318	-0.0514
5	-0.0080	0.0437	0.1118	0.0440	0.0100	-0.0821	0.0690
6	-0.0965	-0.2593	-0.1199	-0.0181	0.0983	-0.0142	0.4654
7	-0.0568	-0.0587	0.0354	0.0141	-0.0108	-0.2087	0.0108
8	-0.0566	0.0433	0.1319	0.0511	-0.0719	-0.1225	-0.0046
9	0.0214	-0.0033	0.0730	-0.0046	0.0974	-0.1842	0.0171
10	-0.0563	-0.0175	0.1238	-0.0538	-0.0069	-0.1776	0.0282
11	-0.0117	-0.1092	-0.0472	-0.0032	-0.2366	-0.1233	0.0048
12	0.2274	0.0241	-0.0109	0.0233	-0.0286	0.0560	0.0796
13	0.1952	0.0217	-0.0428	0.0902	-0.0794	-0.0042	-0.0084
14	0.0110	-0.0502	0.0226	0.1237	0.0791	-0.0641	0.0848
15	-0.0517	0.0125	0.1370	-0.0219	-0.0421	-0.1153	0.0547
16	0.0055	-0.2264	0.0863	-0.0391	0.0177	-0.0666	-0.0181
17	-0.0835	0.0315	0.0694	-0.0795	-0.0829	-0.2285	0.0848
18	-0.0588	0.0171	-0.0182	0.5576	-0.0147	-0.0427	-0.1533
19	0.0151	0.0078	0.1741	0.0149	-0.0236	-0.0645	-0.0439
20	-0.0082	-0.0668	0.1291	-0.0901	0.0096	-0.1401	0.0127
21	0.2362	-0.0285	0.0193	-0.0450	0.0092	0.0451	0.0614
22	0.0237	-0.0071	-0.0372	0.0489	0.0780	-0.2316	0.0365
23	0.1378	0.0257	0.0173	0.0369	0.1625	-0.0168	0.1770
24	-0.0153	0.0407	-0.0095	0.1222	-0.1622	0.0123	0.1768
25	0.0560	0.0021	0.0821	-0.0206	-0.1407	0.1298	0.2270
26	-0.0351	-0.0376	-0.0049	0.0209	0.0763	-0.0062	0.3594
27	0.1848	-0.1597	0.0580	-0.0126	0.0794	-0.0657	-0.1494
28	-0.0168	0.0405	0.1291	0.0852	-0.0288	-0.0086	0.0856
29	-0.0059	0.0228	0.0216	0.3502	-0.1427	0.1209	0.0482
30	0.2942	0.0235	-0.1331	-0.1284	-0.1154	-0.0370	0.1223
31	0.1618	-0.0514	0.0207	-0.0347	0.1194	-0.1687	-0.0723
32	-0.0194	0.0259	-0.0354	-0.0111	-0.0910	-0.0301	0.3207
33	0.0601	-0.0492	0.0308	-0.0015	0.0691	-0.1005	0.0737
34	-0.0182	-0.7392	-0.0413	0.0470	-0.1023	0.1554	-0.0798
35	-0.0370	-0.2599	-0.0088	-0.0308	0.0569	-0.0657	0.1579
36	0.0332	-0.0257	0.0677	-0.0469	-0.3383	0.0243	0.0196
37	-0.0366	0.0348	0.2138	0.0603	-0.0146	-0.0589	-0.0380
38	0.0386	-0.0436	0.0665	-0.0751	0.0138	-0.0978	0.0950
39	-0.0224	-0.0754	0.0209	0.0069	-0.0115	-0.2546	-0.0800
40	0.0360	0.0380	0.0252	0.2201	0.0940	-0.0049	0.1110
41	0.0076	-0.2907	-0.0145	0.0639	-0.0413	-0.0464	-0.0424

Çizelge 4.12 (devam) Rankı Etkileyen Gözlemlere Dayalı Yöntem İle Elde Edilen
Q Matrisi

	1	2	3	4	5	6	7
42	0.0709	-0.0446	0.1170	-0.0587	0.0482	0.0036	0.1200
43	-0.0014	0.0216	0.1509	-0.0700	-0.0662	-0.0474	0.0928
44	0.0634	0.0406	0.0438	0.0423	0.0524	-0.0444	0.1404
45	0.0292	-0.0632	0.1293	-0.0787	-0.0241	-0.1292	-0.0523
46	0.1177	-0.0027	0.0915	-0.0856	-0.0035	-0.1294	-0.0529
47	0.2589	-0.0223	0.0010	0.1455	0.0677	0.0768	-0.0387
48	-0.0785	0.0054	0.0237	-0.0024	-0.0044	-0.3171	-0.0151
49	0.2208	0.0214	0.0251	0.0475	-0.0914	0.0491	-0.0331
50	0.0254	-0.0141	0.1556	0.0522	0.0363	-0.0549	-0.0338
51	0.0581	-0.0007	0.0809	0.0884	-0.0139	0.0242	0.0687
52	-0.0338	0.0479	0.1753	0.0689	-0.0926	-0.0805	-0.0652
53	-0.0284	0.0044	0.1562	0.0703	-0.0470	0.0235	0.0803
54	0.0121	-0.0754	0.1307	-0.0495	-0.0413	-0.0741	-0.0089
55	0.0202	-0.0293	0.1431	-0.0132	-0.0091	0.0088	0.0914
56	0.2471	0.0202	-0.0819	0.0243	-0.0193	-0.0662	-0.0049
57	-0.0260	0.0661	0.1428	0.1531	0.0251	-0.0516	0.0154
58	0.0627	0.0428	0.0213	0.0638	-0.0054	-0.0256	0.1403
59	-0.0382	-0.1649	-0.0013	0.0828	-0.1226	-0.1490	-0.1042
60	-0.0010	0.0179	0.0965	0.1788	-0.0820	0.0479	0.0425
61	-0.0324	0.0745	0.0904	0.0315	-0.0104	0.0291	0.2934
62	0.0805	0.0640	-0.4965	0.1083	-0.3300	-0.4110	0.0873
63	0.2192	-0.0468	0.1433	-0.0889	-0.0232	0.0678	-0.0471
64	-0.0011	0.0066	0.1441	-0.0374	0.0392	-0.1312	0.0202
65	0.2115	0.0052	0.0564	-0.0918	-0.0932	0.0166	0.0118
66	0.1189	0.0059	0.0591	0.1594	0.0244	-0.0082	-0.0598
67	-0.0096	-0.0812	0.1739	-0.0169	-0.0218	-0.0136	0.0165
68	0.0971	-0.0142	0.1364	-0.0015	-0.0077	0.0095	0.0018
69	0.0301	-0.0016	0.0517	0.0817	-0.2324	0.1085	0.1187
70	0.0305	-0.0304	0.1414	0.0239	-0.0274	0.0204	0.0478
71	0.0435	-0.0377	-0.0142	0.2557	0.0392	0.0368	0.0814
72	0.2120	-0.0172	0.0447	0.0731	0.0140	0.0686	-0.0092
73	0.3231	0.0354	-0.1183	-0.0315	0.0374	-0.1302	-0.0382
74	-0.0257	0.0251	0.2126	0.0637	-0.0264	-0.0301	-0.0319
75	-0.0009	0.0193	0.0723	0.0605	-0.3138	-0.0354	-0.0678
76	0.0798	-0.0124	0.0758	-0.0499	-0.2488	0.0141	0.0043
77	-0.0226	-0.0937	-0.0352	0.2783	0.0891	-0.0976	-0.0065
78	-0.0471	0.0071	0.0904	0.0591	-0.0135	-0.1027	0.0660
79	0.1176	-0.0530	0.0899	0.1415	0.0136	0.0893	-0.0101
80	0.1292	0.0424	0.0194	0.0782	0.1466	-0.1515	-0.0219
81	0.0368	0.0330	0.1255	0.0015	0.0281	-0.1328	-0.0299

Çizelge 4.13 Rankı Etkileyen Gözlemlere Dayalı Yöntem İle Elde Edilen Q2 Matrisi

	1	2	3	4	5	6	7
1	0.0286	0.0135	0.0016	0.0042	0.0595	0.0133	0.0000
2	0.0555	0.0002	0.0018	0.0000	0.0103	0.0015	0.0063
3	0.0013	0.0000	0.0269	0.0006	0.0006	0.0000	0.0022
4	0.0151	0.0003	0.0078	0.0098	0.1036	0.0010	0.0026
5	0.0001	0.0019	0.0125	0.0019	0.0001	0.0067	0.0048
6	0.0093	0.0672	0.0144	0.0003	0.0097	0.0002	0.2166
7	0.0032	0.0034	0.0013	0.0002	0.0001	0.0436	0.0001
8	0.0032	0.0019	0.0174	0.0026	0.0052	0.0150	0.0000
9	0.0005	0.0000	0.0053	0.0000	0.0095	0.0339	0.0003
10	0.0032	0.0003	0.0153	0.0029	0.0000	0.0315	0.0008
11	0.0001	0.0119	0.0022	0.0000	0.0560	0.0152	0.0000
12	0.0517	0.0006	0.0001	0.0005	0.0008	0.0031	0.0063
13	0.0381	0.0005	0.0018	0.0081	0.0063	0.0000	0.0001
14	0.0001	0.0025	0.0005	0.0153	0.0063	0.0041	0.0072
15	0.0027	0.0002	0.0188	0.0005	0.0018	0.0133	0.0030
16	0.0000	0.0512	0.0074	0.0015	0.0003	0.0044	0.0003
17	0.0070	0.0010	0.0048	0.0063	0.0069	0.0522	0.0072
18	0.0035	0.0003	0.0003	0.3109	0.0002	0.0018	0.0235
19	0.0002	0.0001	0.0303	0.0002	0.0006	0.0042	0.0019
20	0.0001	0.0045	0.0167	0.0081	0.0001	0.0196	0.0002
21	0.0558	0.0008	0.0004	0.0020	0.0001	0.0020	0.0038
22	0.0006	0.0001	0.0014	0.0024	0.0061	0.0536	0.0013
23	0.0190	0.0007	0.0003	0.0014	0.0264	0.0003	0.0313
24	0.0002	0.0017	0.0001	0.0149	0.0263	0.0002	0.0312
25	0.0031	0.0000	0.0067	0.0004	0.0198	0.0168	0.0515
26	0.0012	0.0014	0.0000	0.0004	0.0058	0.0000	0.1292
27	0.0341	0.0255	0.0034	0.0002	0.0063	0.0043	0.0223
28	0.0003	0.0016	0.0167	0.0073	0.0008	0.0001	0.0073
29	0.0000	0.0005	0.0005	0.1226	0.0204	0.0146	0.0023
30	0.0865	0.0006	0.0177	0.0165	0.0133	0.0014	0.0150
31	0.0262	0.0026	0.0004	0.0012	0.0143	0.0285	0.0052
32	0.0004	0.0007	0.0013	0.0001	0.0083	0.0009	0.1028
33	0.0036	0.0024	0.0009	0.0000	0.0048	0.0101	0.0054
34	0.0003	0.5464	0.0017	0.0022	0.0105	0.0241	0.0064
35	0.0014	0.0676	0.0001	0.0010	0.0032	0.0043	0.0249
36	0.0011	0.0007	0.0046	0.0022	0.1145	0.0006	0.0004
37	0.0013	0.0012	0.0457	0.0036	0.0002	0.0035	0.0014
38	0.0015	0.0019	0.0044	0.0056	0.0002	0.0096	0.0090
39	0.0005	0.0057	0.0004	0.0000	0.0001	0.0648	0.0064
40	0.0013	0.0014	0.0006	0.0484	0.0088	0.0000	0.0123
41	0.0001	0.0845	0.0002	0.0041	0.0017	0.0022	0.0018

Çizelge 4.13 (devam) Rankı Etkileyen Gözlemlere Dayalı Yöntem İle Elde Edilen
Q2 Matrisi

	1	2	3	4	5	6	7
42	0.0050	0.0020	0.0137	0.0034	0.0023	0.0000	0.0144
43	0.0000	0.0005	0.0228	0.0049	0.0044	0.0022	0.0086
44	0.0040	0.0016	0.0019	0.0018	0.0027	0.0020	0.0197
45	0.0008	0.0040	0.0167	0.0062	0.0006	0.0167	0.0027
46	0.0139	0.0000	0.0084	0.0073	0.0000	0.0168	0.0028
47	0.0670	0.0005	0.0000	0.0212	0.0046	0.0059	0.0015
48	0.0062	0.0000	0.0006	0.0000	0.0000	0.1005	0.0002
49	0.0488	0.0005	0.0006	0.0023	0.0083	0.0024	0.0011
50	0.0006	0.0002	0.0242	0.0027	0.0013	0.0030	0.0011
51	0.0034	0.0000	0.0065	0.0078	0.0002	0.0006	0.0047
52	0.0011	0.0023	0.0307	0.0048	0.0086	0.0065	0.0043
53	0.0008	0.0000	0.0244	0.0049	0.0022	0.0006	0.0064
54	0.0001	0.0057	0.0171	0.0025	0.0017	0.0055	0.0001
55	0.0004	0.0009	0.0205	0.0002	0.0001	0.0001	0.0084
56	0.0610	0.0004	0.0067	0.0006	0.0004	0.0044	0.0000
57	0.0007	0.0044	0.0204	0.0234	0.0006	0.0027	0.0002
58	0.0039	0.0018	0.0005	0.0041	0.0000	0.0007	0.0197
59	0.0015	0.0272	0.0000	0.0069	0.0150	0.0222	0.0109
60	0.0000	0.0003	0.0093	0.0320	0.0067	0.0023	0.0018
61	0.0011	0.0055	0.0082	0.0010	0.0001	0.0008	0.0861
62	0.0065	0.0041	0.2465	0.0117	0.1089	0.1689	0.0076
63	0.0480	0.0022	0.0205	0.0079	0.0005	0.0046	0.0022
64	0.0000	0.0000	0.0208	0.0014	0.0015	0.0172	0.0004
65	0.0447	0.0000	0.0032	0.0084	0.0087	0.0003	0.0001
66	0.0141	0.0000	0.0035	0.0254	0.0006	0.0001	0.0036
67	0.0001	0.0066	0.0302	0.0003	0.0005	0.0002	0.0003
68	0.0094	0.0002	0.0186	0.0000	0.0001	0.0001	0.0000
69	0.0009	0.0000	0.0027	0.0067	0.0540	0.0118	0.0141
70	0.0009	0.0009	0.0200	0.0006	0.0007	0.0004	0.0023
71	0.0019	0.0014	0.0002	0.0654	0.0015	0.0014	0.0066
72	0.0450	0.0003	0.0020	0.0053	0.0002	0.0047	0.0001
73	0.1044	0.0013	0.0140	0.0010	0.0014	0.0170	0.0015
74	0.0007	0.0006	0.0452	0.0041	0.0007	0.0009	0.0010
75	0.0000	0.0004	0.0052	0.0037	0.0985	0.0013	0.0046
76	0.0064	0.0002	0.0057	0.0025	0.0619	0.0002	0.0000
77	0.0005	0.0088	0.0012	0.0774	0.0079	0.0095	0.0000
78	0.0022	0.0000	0.0082	0.0035	0.0002	0.0106	0.0043
79	0.0138	0.0028	0.0081	0.0200	0.0002	0.0080	0.0001
80	0.0167	0.0018	0.0004	0.0061	0.0215	0.0230	0.0005
81	0.0014	0.0011	0.0158	0.0000	0.0008	0.0176	0.0009

5. SONUÇLAR

Bu çalışmada çoklu doğrusal regresyon modellerinde çoklu kuşkulu gözlemlerin etkilerinin incelenmesi hedeflenmiştir. Bu hedefe ulaşmak için çalışmanın ikinci bölümünde, çoklu kuşkulu gözlemlerin etkilerinin araştırılacağı regresyon modelinin kurulum aşamaları ayrıntılı olarak anlatılmış, regresyon modelini etkileyen gözlemler olan kuşkulu gözlemler (aykırı gözlemler ve etkili gözlemler) ve doğrusal regresyon modellerinde kullanılan tekli kuşkulu gözlem belirleme yöntemleri tanıtılmış. Üçüncü bölümde çoklu kuşkulu gözlemlerin belirlenmesine neden ihtiyaç olduğu anlatılmış, çoklu kuşkulu gözlem belirlemede yaşanan sıkıntılar ve çözüm yolları belirtilmiş ve çoklu kuşkulu gözlem belirlemede kullanılan yöntemlerden artıklara dayalı yöntem, güven elipsoidinin merkezindeki değişime dayalı yöntemler, güven elipsoidinin hacmine dayalı yöntemler ve rankı etkileyen gözlemlere dayalı yöntem ayrıntılı olarak tanıtılmıştır.

Çalışmada ifade edilen durumların uygulama üzerinde görülebilmesi için, beşinci bölümde uygulama verisi olarak ülkemizdeki 81 ile ait asayiş suç oranları alınmış ve *Braithwaite*'ın 2005 yılında önerdiği "Birleştirici Utandırma Teorisi" yardımı ile bir toplumda meydana gelen suçu ve suç oranlarını açıklamak için regresyon çözümlemesi yapılmıştır. Öncelikle regresyon çözümlemesinin varsayımları test edilmiş ve ortaya çıkan çoklu bağlantı problemi, çoklu bağlantıya sebep olan iki açıklayıcı değişkenden birisi modelden çıkarılarak çözümlenmiştir. Daha sonra tekli kuşkulu gözlem belirleme yöntemleri kullanılarak, modeli etkilemesinden kuşkulanılan gözlemler belirlenmiştir. Belirlenen gözlemler modelden çıkarıldığında, modelin belirtme katsayısını artırdıkları, modelin standart hatasını düşürdükleri ve parametre katsayılarını değiştirdikleri gözlemlenmiştir. Uygulamada tekli kuşkulu gözlem belirlemede kullanılan yöntemlerden Cook uzaklığı yöntemi kuşkulu gözlem belirlemede yeterli olamamıştır. Fakat Cook uzaklığının çoklu kuşkulu gözlem belirlemek için uyarlanmış hali olan genelleştirilmiş Cook uzaklığı yöntemi ikili kuşkulu gözlem belirlemede başarılı olmuştur. Ardından bu ikili incelendiğinde, birbirlerinin etkilerini gizlediklerinden dolayı Cook uzaklığı yöntemi ile tespit edilemedikleri anlaşılmıştır. Çoklu kuşkulu gözlem araştırması üçlü ve dörtlü kuşkulu gözlemleri araştırma ile devam etmiş,

bu araştırma da geri eleme yöntemi kullanılmıştır. Tespit edilen üçlü ve dörtlü kuşkulu gözlemlere bakıldığında gözlemlerin bazıları tekli kuşkulu gözlem belirleme yöntemleri ile kuşkulu olarak belirlenen gözlemler iken bazıları tekli kuşkulu gözlem belirleme yöntemleri ile kuşkulu olarak belirlenmedikleri görülmüştür. Ayrıca tekli kuşkulu gözlem belirleme yöntemleri ile kuşkulu olarak belirlenen gözlemlerin oluşturduğu alt kümelerin çoklu kuşkulu olmadıkları tespit edilmiştir. Çoklu kuşkulu gözlemler modelden çıkarıldığında ise, modelin belirtme katsayısını arttığı, modelin standart hatasını düştüğü ve regresyon katsayılarının, varyanslarının ve modeldeki anlamlılıklarının değiştiği gözlemlenmiştir. Çoklu kuşkulu gözlemlerin modele olan etkilerinin tekli kuşkulu gözlemlerin modele olan etkilerinde daha fazla olduğu görülmüştür. Ayrıca uygulamada çoklu kuşkulu gözlemleri belirlemede sıkıntılar yaşandığı gözlemlense de, önerilen çözüm yöntemleri ile çoklu kuşkulu gözlemlerin belirlenebildiği ortaya çıkmıştır.

Genel olarak uygulama sonucunda şu bulgular elde edilmiştir;

- Regresyon çözümlenmesinde gerekli aşamalar yerine getirildikten sonra kuşkulu gözlem analizlerine geçilmesi gerektiği,
- Regresyon çözümlenmesinde çoklu kuşkulu gözlemlerin belirlenmesi aşaması yapılmadan sağlıklı bir kuşkulu gözlem analizinin yapılamayacağı,
- Çoklu kuşkulu gözlemlerin belirlenmesinde zorluklar yaşanabileceği fakat bu zorlukları aşmak için çözüm yollarının var olduğu,
- Çoklu kuşkulu gözlemlerin etkilerinin tekli kuşkulu gözlemlerinin etkilerinden daha fazla olduğu,
- Çoklu kuşkulu gözlemlerin modeli oluşturan parametrelerin anlamlılıkları üzerinde oldukça fazla etkili olabildikleri belirlenmiştir.

KAYNAKLAR

Akar, M. & Şahinler, S., 1993. Çoklu Regresyon Analizinde Kısmi Kalıntılar ve Grafik Analizi. 1. *Ulusal Ekonometri ve İstatistik Sempozyumu*. İzmir.

Alpar, R., 2003. *Uygulamalı çok değişkenli istatistiksel yöntemlere giriş : I*. Ankara: Nobel.

Andrews, D.F. & Pregibon, D., 1978. Finding the outliers that matter. *J. Roy. Statist. Soc., Ser. B(40)*, pp.85-93.

Aşıkıl, B., 2006. *Çoklu Doğrusal Regresyonda Aykırı, Etkili Değerlerin Araştırılması ve Bir Uygulama*. Yüksek Lisans Tezi. İstanbul: Mimar Sinan Güzel Sanatlar Üniversitesi.

Aydın, Y., 2006. *Grafik Yöntemlerle Etkin Gözlemlerin Ve Aykırı Değerlerin Tespiti*. Yüksek Lisans Tezi. Samsun: T.C. Ondokuz Mayıs Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilimi Dalı.

Belsley, D.A., Kuh, E. & Welsch, R.E., 1980. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Wiley.

Braithwaite, J., 2005. *Crime, Shame and Reintegration*. 15th ed. New York: Cambridge University Press.

Chatterjee, S. & Hadi, A.S., 1986. Influential Observations, High Leverage Points, and Outliers in Linear Regression. *Statistical Science*, Vol. 1(3), pp.379-93.

Chatterjee, S. & Hadi, A.S., 1988. *Sensitivity Analysis in Linear Regression*. John Wiley and Sons, Inc., Canada.

Chatterjee, S. & Hadi, A.S., 2006. *Regression Analysis by Example*. New Jersey: John Wiley and Son, Inc.

Cook, R.D., 1977. Detection of influential observation in linear regression. *Technometrics*, (19), pp.15-18.

Cook, R.D. & Weisberg, S., 1982. *Residuals and Influence in Regression*. New York: Chapman and Hall.

Delozier, M.R. & Orlich, S., 2005. Discovering influential cases in linear regression with MINITAB Peeking into multidimensions with a MINITAB macro. *Statistical Methodology*, (2), p.71–81.

Dolu, O., 2010. *Suç Teorileri Teori, Araştırma ve Uygulamada Kriminoloji*. 1st ed. Ankara: Seçkin Yayıncılık.

Draper, N.R. & John, J.A., 1981. Influential Observations and Outliers In Regression. *Technometrics*, Volume 23, February. pp.21-26.

Draper, N.R. & Smith, H., 1966. *Applied Regression Analysis*. John Wiley&Sons Inc.

Hoaglin, D.C. & Welsch, R.E., 1978. The Matrix in Regression and ANOVA. *Am.Statist.* 32, pp.17-22.

Imon, A.H.M.R., 2005. Identifying Multiple Influential Observations in Linear Regression. *Journal of Applied Statistics Vol. 32, No. 9*, November. p.929–946.

Kempthorne, P.J., 1989. *Identifying Rank-Influential Groups of Observations in Linear Regression Modeling*. Cambridge: Sloan School of Management Massachusetts Institute of Technology.

Meloun, M. & Militky, J., 2001. Detection of single influential points in OLS regression model building. *Analytica Chimica Acta*, (439), p.169–191.

Montgomery, D.C. & Peck, E.A., 1992. *Introduction to Linear Regression Analysis*. John Wiley and Sons, Inc., Canada.

Myers, R.H., 1990. *Classical and Modern Regression with Applications. Second Edition*. Duxbury.

Rousseeuw, P.J. & Leroy, A.M., 2003. *Robust Regression and Outlier Detection*. John Wiley&Sons, Inc.

Rousseeuw, P.J. & Zomeren, B.C., 1990. Unmasking Outliers and Leverage Points. *Journal of the American Statistical Association*, pp.633-39.

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

Scanlon, E.S., 1994. Residuals and Influence in Regression. In *CAS Proceedings, Vol. LY,XXI*. The Marriott- Copley Place, Boston, Massachusetts, 1994.

Seber, G.A.F. & Lee, A.J., 2003. *Linear Regression Analysis. Second Edition*. John Wiley and Sons, Inc., Canada.

Shih, W.J. & Weisberg, S., 1986. Assessing Influence in Multiple Linear Regression with Incomplete Data. *Technometrics*,28, pp.231-39.

Ullah, M.A. & Pasha, G.R., 2009. The origin and developments og influence measures in regression. *Pakistan Journal of Statistics*, pp.295-307.

Welsch, R.E., 1982. Influence functions and regression diagnostics. In *Modern Data Analysis*. New York: Academic.

Welsch, R.E. & Kuh, E., 1977. *Linear regression diagnostics*. Technical report 923-77. Solan School of Management, Massachusetts Institute of Technology.

ÖZGEÇMİŞ

Adı Soyadı : Ömer LÜSER

Doğum Yeri : Kocaeli

Doğum Yılı : 1985

Medeni Hali : Bekar

Eğitim ve Akademik Durumu:

Lise 2000-2003 Polis Koleji

Lisans 2003-2007 Hacettepe Üniversitesi İstatistik Bölümü

Yabancı Dil: İngilizce

İş Tecrübesi:

2007-... Emniyet Genel Müdürlüğü