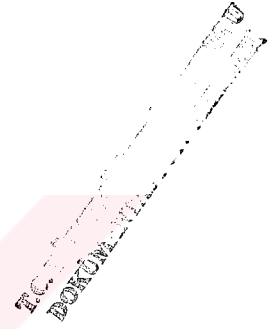


68001

**ÇOKLUBAĞLANTIYA NEDEN OLAN GÖZLEMLERİN
BELİRLENMESİ**

Ufuk EKİZ

**YÜKSEK LİSANS TEZİ
(İSTATİSTİK)**



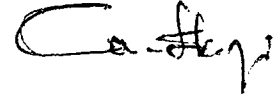
**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ**

Haziran 1997

ANKARA

Ufuk EKİZ tarafından hazırlanan ÇOKLU BAĞLANTIYI ETKİLEYEN ETKİLİ GÖZLEMLERİN BELİRLENMESİ adlı bu tezin Yüksek Lisans tezi olarak uygun olduğunu onaylarım.

Doç. Dr. Müslim EKNİ
Tez Yöneticisi



Bu çalışma, jürimiz tarafından İstatistik Anabilim Dalında Yüksek Lisans tezi olarak kabul edilmiştir.

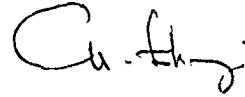
Başkan : Prof.Dr.Soner GÖNEN



Üye : Doç.Dr.Hamza GANGAM



Üye : Doç.Dr.Müslim EKNİ



Üye :

Üye :

Bu tez, Gazi Üniversitesi Fen Bilimleri Enstitüsü tez yazım kurallarına uygundur.



**ÇOKLU BAĞLANTIYA NEDEN OLAN GÖZLEMLERİN
BELİRLENMESİ
(Yüksek Lisans Tezi)**

Ufuk EKİZ

**GAZİ ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
Haziran 1997**

ÖZET

Regresyon analizinde kullanılan en küçük kareler tahmin edicileri üzerinde, olumsuz etkilere sebep olan ve bağımsız değişkenler arasındaki sıkı ilişki olarak tanımlanan çoklu bağlantının varlığı çoğu çalışmada gözden kaçırılmaktadır. Bu sebeple çalışmada, çoklubağlantının en küçük kareler tahmin edicileri üzerindeki etkileri incelenmiştir. Daha sonra, çoklubağlantının şiddetinin önemli olup olmadığının belirlenmesine yönelik çeşitli yöntemler incelenmiştir. Bundan sonraki aşamalarda problem yaratan çoklubağlantının ortadan kaldırılmasına yönelik değişken eliminasyon yöntemleri, Sağlam(Robust) tahmin yöntemleri ve çoklubağlantıya neden etkili gözlemlerin ortaya çıkarılmasına ilişkin yöntemler üzerinde durulmuştur. Çoklu bağlantıya neden olan etkili gözlemlerin belirlenmesine yönelik olan dik izdüşüm bileşenleri yöntemine benzer bir yöntem önerilmiş, önerilen bu yöntem ile dik izdüşüm bileşenleri yöntemi uygulama sonunda birbirleri ile karşılaştırılmıştır.

Bilim Kodu : 406.01.01

Anahtar Kelimeler: Regresyon, En küçük kareler, Çoklubağlantı, Etkili gözlem, Dik
İzdüşüm bileşenleri

Sayfa Adedi : 74

Tez Yöneticisi : Doç. Dr. Müslim EKNİ

INVESTIGATION OF MULTICOLLINEARITY-INFLUENTIAL OBSERVATIONS

M.Sc. Thesis

Ufuk Ekiz

GAZİ UNIVERSITY
INSTITUTE OF SCIENCE AND TECHNOLOGY
June 1997

ABSTRACT

Multicollinearity is caused negative influences about the least squares method in regression analysis. For this reason in this thesis influences of multicollinearity are investigated about the least squares method. And then the violence of multicollinearity are investigated by several methods. Variables elimination methods, Robust guess methods and methods of the multicollinearity influential observations are determined. Method of the orthogonal leverage component determines multicollinearity-influential observations. We suggested the new method which was similar method of the orthogonal leverage component. This method is compared with the method of orthogonal leverage component in applied.

Science code	:406.01.01
Key words	:Regression, The Least Squares, Multicollinearity Influential Observation, Orthogonal Leverage Component.
Page number	:74
Adviser	:Associated Prof.Dr.Müslim EKNİ

TEŞEKKÜR

Çalışmalarım boyunca değerli yardım ve katkılarıyla beni yönlendiren Hocam Doç. Dr. Müslim EKNİ 'ye ve çalışma arkadaşlarıma ayrıca manevi desteklerinden dolayı sevgili eşim Meltem EKİZ 'e ve aileme teşekkürü bir borç bilirim.



ÇİZELGELERİN LİSTESİ

Çizelge	Sayfa
Çizelge 4.1. İller bazında 1996 yılı bağımlı ve bağımsız değişken değerleri	49
Çizelge 4.2. Çizelge 4.1 deki verilere ilişkin çoklubağlantı ölçüleri	51
Çizelge 4.3. Doğrusal regresyon modelinin parametre tahminleri	52
Çizelge 4.4. Çizelge 4.1 deki bağımsız değişkenlere ilişkin dik izdüşüm bileşenlerinin değerleri	53
Çizelge 4.5. Etkili gözlemleri belirlemekte kullanılan başlıca istatistiklerin sonuçları	56
Çizelge 4.6. $\cos\theta_{\ell i}$, i. gözlemin $(X'X)$ matrisinin ℓ . özvektörüyle yaptığı açılarının cosinüs değerleri	58
Çizelge 4.7. Veride ayrı ayrı çoklubağlantıyı maskeleyen ve şiddetlendiren gözlemler yer almaması durumunda koşullu indeks ve VIF değerleri	61
Çizelge 4.8. Doğrusal regresyon modellerine ait parametre tahminlerine ilişkin t değerleri	63

ŞEKİLLERİN LİSTESİ

Şekil

Sayfa

Şekil 1.1. Bağımsız değişkenin boyu(normu) ile bağımlı değişken arasındaki ilişki	11
Şekil 1.2. Serpme diyagramı	11
Şekil 3.1. Çoklubağlantıyı maskeleyen ve şiddetlendiren noktalar	28
Şekil 3.2. Etkili gözlemlerin çoklubağlantıya etkisi	36
Şekil 3.3. i.gözlemin v_i ile yaptığı açının cosinüsü	47



İÇİNDEKİLER

	Sayfa
ÖZET	i
ABSTRACT	ii
TEŞEKKÜR	iii
ÇİZELGELERİN LİSTESİ	iv
ŞEKİLLERİN LİSTESİ	v
1. GİRİŞ	1
2. GENEL BİLGİLER	2
2.1. Çok Değişkenli Doğrusal Regresyon Modeli	2
2.2. Çok Değişkenli Doğrusal Regresyon Modelinde En Küçük Kareler Parametre Tahminleri	5
2.3. En Küçük Kareler Yöntemi İçin Gerekli Varsayımlar	6
2.4. En Küçük Kareler Tahmin Edicilerinin Karşılaştığı Sorunlar	10
2.5. Çoklubağlantının Tanımı	14
2.6. Çoklubağlantının En Küçük Kareler Tahmin edicileri Üzerindeki Etkisi ...	17
2.7. Çoklubağlantı Durumunda Parametre Tahmini ve Kestirim	19
2.8. Çoklubağlantının Belirlenmesi.....	21
2.9. Çoklubağlantının Etkisinin Ortadan Kaldırılması	25
3. ÇOKLUBAĞLANTIYI ŞİDDETLENDİREN ETKİLİ GÖZLEMLER	27
3.1 Etkili Gözlemin Tanımı	27
3.2. Etkili Gözlemin Çoklubağlantıya Etkisi	27
3.2.1. Çoklubağlantıyı maskeleyen veya şiddetlendiren etkili gözlem	27
3.2.2. Bir gözlemin merkezden uzaklaştıkça çoklubağlantıya olan etkisi	29

3.2.3. Çoklubağılantıya neden olan gözlemin özdeğer analizi ile incelenmesi	32
3.3. Çoklubağılantıya Neden Olan Etkili Gözlemin Belirlenmesi	36
3.3.1. Etkili gözlemin belirlenmesinde kullanılan istatistikler	36
3.3.2. Etkili gözlemin, özvektörlerle yaptığı açılara göre bu gözlemin çoklubağılantıyı şiddetlendiren veya maskeleyen gözlem olup olmadığının incelenmesi	45
4. UYGULAMA	48
5. SONUÇ VE ÖNERİLER	64
KAYNAKLAR	66
EK - 1.BİR GÖZLEMİN ÇOKLUBAĞLANTIYA OLAN ETKİSİ	69
ÖZGEÇMİŞ	74

1. GİRİŞ

Çok değişkenli regresyon çalışmasında optimal özellikleri taşıdığı bilinen en küçük kareler parametre tahminleri için gerekli şartlar çoğu kez gerçekleşmeyebilir. Örneğin , homojen varyans varsayımının sağlanmıyor olması, veride aykırı gözlemin veya uç gözlemin olması ve bağımlı değişkeni açıklama gücüne sahip bağımsız değişken yada değişkenlerin işleme dahil edilmemesi veya gereği kadar önemli bilgi sağlamadıkları için bu bağımsız değişkenlerin elimine edilmesi gibi sorunlarla karşı karşıya kalınmaktadır.

İki yada daha çok bağımsız değişken arasında tam yada yakın doğrusal ilişkinin bulunması olarak tanımlanan "çoklubağlantı" problemi de en küçük kareler parametre tahminlerini optimallikten uzaklaştıran başka bir etkendir.

Bu çalışmadan amaç , çoklubağlantının , en küçük kareler tahmin edicilerine etkilerini, varlığının nasıl belirleneceğini , etkisinin nasıl arındırılacağını , gözlem veya gözlemlerin çoklubağlantıya olan etkileri hakkında yapılan bazı çalışmaları düzenli bir akış içinde sunmak ve bunları gerçek bir uygulamayla pekiştirmektir.

2. GENEL BİLGİLER

2.1. Çok Değişkenli Doğrusal Regresyon Modeli

Bağımlı değişkeni açıklamak için kullanılan çok değişkenli doğrusal regresyon modeli,

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i \quad (2.1)$$

şeklinde gösterilir. Burada $i = 1, 2, \dots, n$ gözlem sayısını, p de bağımsız değişken sayısını ifade etmektedir. Bu modelde β_j ($j=0,1,2,\dots,p$) bilinmeyen parametrelerdir.

Doğrusal model matris biçiminde ifade edilirse;

Y ; $n \times 1$ boyutlu bağımlı değişken vektörü,

X ; $n \times p'$ boyutlu ilk sütunu 1 vektörü , geri kalan p sütunu bağımsız değişkenlerden oluşan matris,

β ; $p' \times 1$ boyutlu bilinmeyen parametreler vektörü,

ε ; $n \times 1$ boyutlu rasgele hata vektörü,

olmak üzere

$$Y = X \beta + \varepsilon \quad \text{yada}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

şeklinde yazılabilir. Bu model yapısında

X matrisi;

X in her bir sütunu bir bağımsız değişkeni ifade etmektedir. X in her bir satırının elemanları (r. satır) uygun β parametre katsayıları ile $E(y_r)$ yi belirler. X 'in birinci satırı ile β 'yı çarpıp ε 'un ilk elemanını ekleyerek yukarıdaki model için ilk gözlem değeri

$$y_1 = \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \dots + \beta_p x_{1p} + \varepsilon_1$$

şeklinde elde edilmiş olur.

Y ve ε vektörleri rasgele vektörlerdir. X matrisi bilinen sabitler matrisi olarak gözönüne alınmaktadır.

β vektörü;

β vektörü veriden tahmin edilecek olan bilinmeyen sabitler vektörüdür. Her bir elemanı, β_j , j. bağımsız değişkendeki birim değişimin bağımlı değişkendeki değişimi veren kısmi regresyon katsayısıdır. Tanımlanan her bir kısmi regresyon katsayısı modeldeki bağımsız değişkenler setine bağlıdır.

Rasgele ε (Hata) Vektörü;

ε , ortalaması sıfır olan çok değişkenli normal dağılıma sahip rasgele hata vektörüdür. n elemanlı rasgele bir vektörün varyans-kovaryans matrisi nxn boyutlu simetrik bir matris olup köşegen elemanları rasgele değişkenlerin (ε_i) varyanslarına ve köşegen dışı elemanları da i. ve j. için ε_i ile ε_j arasındaki kovaryansa eşittir. ε için varyans-kovaryans matrisi $\sigma^2 I$ dir. Burada I (nxn) boyutlu birim matristir. σ^2 , ε_i lerin hepsi için ortak varyanstır. ε 'un dağılımı kısaca şöyle yazılabilir.

$$\varepsilon \sim N(0; \sigma^2 I)$$

ε 'un varyans-kovaryans matrisi için iki varsayım vardır.

- i) ε_i 'ler ortak varyansa (σ^2) sahip olmalıdır (homojen varyans durumu).
- ii) ε_i 'ler istatistiksel olarak bağımsızdır (bağımsızlık için kovaryansların sıfır olması gerekir).

Y vektörü;

Modeldeki $X\beta$ sabitler seti ile ε rasgele hatalar vektörü toplamı olan Y, ortalaması $X\beta$ ve varyans-kovaryans matrisi $\sigma^2 I$ olan rasgele gözlemler vektörüdür.

$$E(Y) = E(X\beta + \varepsilon) = E(X\beta) + E(\varepsilon) = X\beta \quad \text{ve}$$

$$\text{Var}(Y) = \text{Var}(X\beta + \varepsilon) = \text{Var}(X\beta) + \text{Var}(\varepsilon) = \text{Var}(\varepsilon) = \sigma^2 I \quad \text{dir.}$$

$\text{Var}(Y)$ ile $\text{Var}(\varepsilon)$ aynıdır. Çünkü rasgele değişken ile sabit terimin toplamı varyansı değiştirmez. ε ' un dağılımı normal iken Y de çok değişkenli normal dağılmaktadır. Böylece

$$Y \sim N(X\beta; \sigma^2 I) \quad \text{dir.}$$

Bu sonuç tam doğru (yanlışsız) oluşturulan doğrusal regresyon modelindeki varsayımlara bağlıdır. Eğer önemli bağımsız değişkenler modele alınmamışsa yada model yanlış fonksiyonel formda oluşturulmuş ise $X\beta$ nın Y'yi açıklaması beklenemez.

2.2. Çok Değişkenli Doğrusal Regresyon Modelinde En Küçük Kareler Parametre Tahminleri

Bölüm 2.1 de tanımlanan doğrusal regresyon modeli, gözlenmiş verilerden tahmin edilecek p' tane parametreye sahiptir. Eğer hatalar rasgele olmasaydı, herhangi iki nokta kullanılarak parametre değerlerine doğrudan ulaşılabilirdi. Y 'deki rasgele değişim nedeniyle, gözlenmiş veri noktalarının her biri farklı sonuçlar verir. (Eğer gözlenmiş veriler tam bir doğru oluşturursa ancak o zaman i.kestirim ile i.bağımlı değişken değerleri aynı olabilirler). Bazı kriterler içinde en iyi olan ve bütün bilgileri kapsayarak tek bir sonuç verebilecek bir metoda ihtiyaç vardır.

En küçük kareler tahmin yöntemi gözlenmiş Y_i ile bunların tahminleri arasındaki farkların kareler toplamını, bir başka deyişle hata kare toplamını sıfır yapacak çözümü bulmaktır. $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$; $\beta_0, \beta_1, \dots, \beta_p$ parametrelerinin tahminleri olmak üzere

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2} + \dots + \hat{\beta}_p x_{ip}$$

her bir x_i (i; 1, 2, ... , n) için Y 'nin ortalamasının tahminidir. Dikkat edilirse $\hat{Y}_i$, parametreler yerine tahmin değerlerinin kullanılması ile bulunmaktadır. En küçük kareler yönteminde , hata kareler toplamını sıfır yapacak $\hat{\beta}_0, \hat{\beta}_j$ (j; 1, ..., p) leri bulmak temel arzudur.

KT (Hata) → Hata kareler toplamı olmak üzere;

$$KT(Hata) = \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 = \sum_{i=1}^n e_i^2$$

dır. Burada $e_i = (Y_i - \hat{Y}_i)$, i. gözlem için gözlenmiş artıktır. $\beta_0, \beta_1, \dots, \beta_p$ nin tahminleri KT(Hata) 'ı en küçükleyen değerlerdir.

2.3. En Küçük Kareler Yöntemi İçin Gerekli Varsayımlar

En küçük kareler yöntemi hataların karelerinin toplamsallığına ve bunların varyanslarının σ^2 olmasına dayandırılmaktadır. Bağımsızlık ve ortak varyans varsayımları sağlandığında en küçük kareler tahmin edicileri , doğrusal yansız tahmin ediciler arasında en küçük varyansa sahip tahmin edicilerdir. Normallik varsayımı da sağlanıyorsa, en küçük kareler tahminleri aynı zamanda En Çok Olabilirlik tahmin edicileri olurlar (Rawlings, 1982).

En küçük kareler analizinde iki temel problemin ortak varyans ve artıkların bağımsız olmasının gerekliliğinden bahsedilmişti. Ortaya çıkan diğer problemler etkili gözlemler, model fonksiyonunun yetersiz tanımlanması, bağımsız değişkenler arasında doğrusallığa yakın bir ilişkinin bulunması (Çoklubağlantı) veya bağımsız değişkenlerin artığa dahil edilmesidir. Problemlerin tanımları bu bölümde kısaca özetlenecek , en küçük kareler tahmin edicileri üzerindeki etkileri ve çarelerinin neler olduğu ve problemin en aza indirilmesine değinilecektir. Bütün regresyon analizlerinde önemli adım modelin yeterliliği ve verilerin davranışlarıdır. Bazı teknikler varsayımların doğruluğunu kontrol etmek için kullanılmaktadır. Bununla beraber bağımsız veri setine doğrusal regresyon modelinin uygun olarak karşılık gelip gelmediğine dikkat edilmelidir. Burada önemli olan husus modelde yada veride ortaya çıkan problemin farkına varabilmek , mümkün problemleri ortadan kaldırmaktır. Varsayımlar sağlanmadığında en küçük karelere alternatif sağlam(robust) tahmin ediciler vardır. Sağlam regresyon olarak verilen istatistiksel teknikler parametrik modeldeki varsayımlar doğrultusunda tahminlerin duyarlılığını azaltırlar. Örneğin, en küçük karelerle işlem yapmanın verideki etkili

gözlemlere yada büyük artıklara karşı duyarlı olduğu bilinir. Çünkü çözüm hata kareler toplamalarının en küçüklenmesinden elde edilmiştir. Sağlam regresyon yönteminde büyük artıklara ağırlık vermek vasıtasıyla bu artıkların tahminlere olan etkisi azaltılabilir. Ancak Sağlam regresyona bu tezde yer verilmeyecektir.

a) Normal Dağılımın Sağlanmama Durumu (Nonnormality);

Varsayılan hataların, ε , regresyon parametrelerinin ve varyanslarının tahmini için normal dağılıyor olması şart değildir. Normallığe sadece parametrelere ilişkin güven aralıkları ve anlamlılığın testi için ihtiyaç duyulur. t, F, χ^2 testlerinde rasgele değişkenlerin normal dağılıyor olmasına ihtiyaç vardır. Aynı şekilde, alışıla gelmiş parametreye ilişkin güven aralığı da normal dağılıma bağlıdır (student t dağılımından dolayı).

b) Ortak Olmayan Varyans Durumu (Heterogeneous Varyans);

Ortak varyans varsayımı alışılmış en küçük karelerde anahtar rolü oynar. Bu varsayımla ifade edilen bağımsız değişkendeki her bir gözlemin aynı miktarda bilgiyi kapsamasıdır. Netice olarak, alışılmış en küçük karelerdeki bütün gözlemler aynı ağırlığı almaktadır. Diğer yandan, ortak olmayan varyans durumunda bazı gözlemler diğer gözlemlerden daha çok bilgiyi kapsarlar. Kullanılan veride çok bilgi kapsayan böyle gözlemlere yüksek ağırlık vermeye ihtiyaç duyulabilir. Alışılmış en küçük kareler tahmin edicisinin en küçük varyanslı olma özelliği doğrudan bu varsayıma bağlıdır.

c) Hataların İlişkili Olması;

Artıklar arasındaki korelasyon, ilişkili veriler ile ortaktır. Zaman serilerindeki hatalar arasındaki ilişki doğrudan doğruya daha sonraki gözlemlerin hataları ile ilişkiye neden olmaktadır. Muhtemelen bazı fiziksel projelerdeki bu süreklilik işlem sonrasında seri korelasyonlar gösterecektir. Örneğin, fabrika bacasından çıkan dumandaki kirlilik miktarının saat başı ölçümleri arasında oldukça yüksek bir ilişki(correlation) vardır. Biyoloji derslerindeki tekrarlı

ölçümlerde bazı hassasiyet gösteren durumlar vardır, şöyleki, bitki ve hayvanların gelişmelerini incelemekte yada klinik denemelerde çoğunlukla artıklar arasında ilişkiye sebebiyet vermektedir. Bazı deney tasarımlarında, deney bloklarındaki gözlemler arasında ilişkili hatalar ortaya çıkmaktadır. Bununla beraber, bazı durumlarda deneyin türünü belirlemek arzu edilmesede ilişkileri ortaya çıkabilir. Örneğin, deney birimlerini gruplamada onların hangi gruba uygun olduğunu belirlemekte yapılan işlemleri uygulamak, besleyici çözümlerin uygulanması, ölçümlerin yapılması v.s, gruplar içindeki gözlemler arasında hataların ilişkili olarak ortaya çıkmasına neden olacaktır. Bu ilişkiler uygun analizlerde sıkça gözden kaçırılır ve hesaba katılmaz.

d) Etkili Gözlemler (Influential Data Points);

Alışılmış en küçük kareler yönteminde her gözleme eşit ağırlık verilir. Bununla beraber, her gözlem, en küçük kareler tahmin edicilerinin sonuçları üzerinde aynı etkiye sahip değildir. Örneğin, basit doğrusal regresyonda bağımsız değişkenin ortalamasından oldukça uzakta olan gözlem değerleri regresyon doğrusunun yönü üzerinde oldukça etkilidir. Verideki diğer noktalarla ilgisi olmayan, onlardan uzak olan tek bir nokta, regresyonun sonuçları üzerinde diğer noktaların tümünden daha büyük etkiye sebep olabilir. Bu gözlemler etkili yada izdüşüm değeri (leverage) yüksek gözlemler olarak isimlendirilirler. Etkili gözlemler üzerinde daha ayrıntılı bilgiler bölüm 3 'te yer almaktadır.

e) Modelin Yetersizliği (Model Inadequacies);

Bağımlı değişkeni açıklamakta gerekli tüm bağımsız değişkenler modele alınmış ise alışılmış en küçük kareler tahmin edicileri yansız olurlar. Modelin tam modelden çeşitli farklılıkları varsa yansız olmayacaklardır. Örneğin, önemli bir bağımsız değişken modelde bulunmuyorsa, hata kare ortalaması σ^2 nin yanlış tahmini ve ayrıca bütün değişkenlere ilişkin regresyon katsayıları da yanlış olur.

f) Bağımsız Değişkenlerdeki Hatalar (Error in the Independent Variables);
En uygun (Tam) modelde varsayılan bağımsız değişkenler hatasız ölçülmüşlerdir. Bağımsız değişkenlerin ölçümlerinden doğan hatalar, hata terimi içinde gizlenir. Oysa ki, gözlenmiş x_i ;

$$x_i = \eta_i + \delta_i$$

şeklinde dir. Burada η_i , x_i nin doğru olarak gözlenemeyen değeri , δ_i 'de hata ölçüsüdür. δ_i hatasının varyansının σ_δ^2 ve ortalamasının sıfır olduğu varsayılmaktadır. Varsayılan regresyon modelindeki Y_i , η_i değerinin bir fonksiyonudur. Buna göre model,

$$Y_i = \mu + \beta \eta_i + \varepsilon_i$$

dir. $\hat{\beta}$ deki yanlılık; alışılmış en küçük kareler uygulandığında bu problem (gözlenemeyen η_i yerine x_i nin kullanılması) β nin yanlış tahmin edicisini verir. Genelde $\hat{\beta}$ 'ne, tam değere doğru meydana bir değer biçilir. Eğer δ_i hatalarının ölçümleri η_i den bağımsız olursa, $\hat{\beta}$ nin beklenen değeri yaklaşık olarak,

$$E(\hat{\beta}) = \frac{\beta}{1 + n\sigma_\delta^2 \sum_{i=1}^n x_i^2}$$

olur (Rawlings, 1982). $E(\hat{\beta})$ daima 1 den büyüktür, $\sigma_\delta^2 > 0$. Eğer σ_δ^2 , $\sum x_i^2/n$ 'e nazaran küçük olursa, yanlılık küçük olur.

g) Çoklubağlantı Problemi (The Multi-Collinearty Problem);

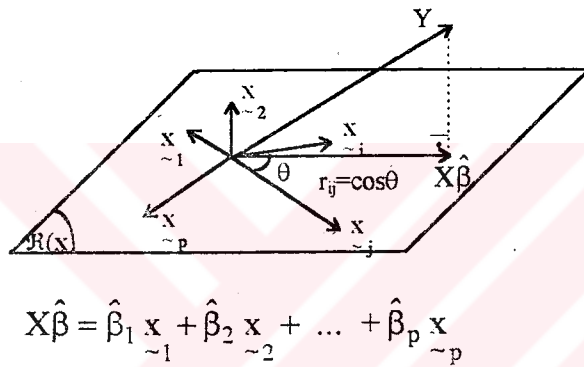
X' in bazı sütunlarının doğrusal fonksiyonları sıfıra oldukça yakın olduğunda X de tekil(singular) olma durumu ortaya çıkmaktadır. X' in tekil olması durumunda en küçük kareler analizinde $(X'X)^{-1}$ bulunamaz. Böyle tekil bir matrisin olması oldukça zahmetli bir durum ortaya çıkarır. Çalışılmakta olan sistemde birbirine oldukça yakın olan bağımsız değişkenler X de yakın tekilliğe sebep olabilir. Tekillığe yakın olan durumda normal eşitliklerden tek bir çözüm bulunur. Bu sağlıklı olmayan bir çözümdür. Değişkenlerdeki küçük değişiklikler (Y yada X) regresyon katsayılarının tahmininde şiddetli değişikliğe sebep olabilir. Regresyon katsayılarının varyansları (bağımsız değişkenler tekilliği gerektiriyorsa) çok büyük olacaktır. Sonuçta, tekilliğe-yakınlık içinde bulunan değişkenler birbirleri yerine geçebilirler, genellikle bağımsız değişkenlerin farklı kombinasyonları Y yerine hemen hemen eşit olacak değeri verebilirler. Tekillığe yakın olmakla X de ortaya çıkan zorluklara ortak olarak "Çoklubağlantı" (Multi-Collinearty) problemi adı verilir.

2.4. En Küçük Kareler Tahmin Edicilerinin Karşılaştığı Sorunlar

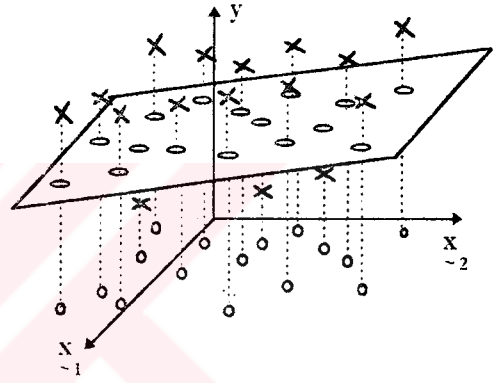
En küçük karelerdeki sorunlar dört temel başlık altında toplanmaktadır.

i) $\hat{Y} = X\hat{\beta}$ vektörü $Y \in \mathcal{R}^n$ olmak üzere X (bağımsız) değişkenlerinin gerdiği alt uzaya Y 'nin dik izdüşümüdür. $\hat{Y}$, X değişkenlerinin doğrusal fonksiyonudur. Bu doğrusal fonksiyondaki bağımsız değişkenlere ilişkin katsayıların tahmini en küçük kareler yöntemi ile bulunurken bu çözüm, X değişkenlerinin yapısına bağlı olacaktır. X değişkenleri arasında, bir yada bir kaçının doğrusal bağımlı olması durumunda ($X'X$ matrisinin tam ranklı olmaması durumunda) en küçük kareler çözümü tek olarak belirlenemeyecektir. X değişkenlerinin doğrusal bağımsız olması durumunda

çözüm tek olmakla birlikte değişkenlerin birbirine karşı diklikten uzaklaşması halinde ,Y (bağımlı değişken) deki küçük değişmeler , doğrusal regresyon modelindeki katsayıların tahminlerinin önemli derecede değişmesine neden olabilecektir. Bu değişmelerin büyüklüğü, bağımlı değişkenin konumuna ve bağımsız değişkenlerin boylarına (normlarına) bağlı olacaktır. Doğrusal regresyon modelindeki boyu küçük değişkenlere karşılık gelen katsayıların bağımlı değişkendeki değişimlere daha duyarlı olacağı açıktır (Şekil 1.1).



Şekil 1.1. Bağımsız değişkenin boyu ile bağımlı değişken arasındaki ilişki



Şekil 1.2. Serpme diyagramı

ii) Bağımsız X değişkenlerine ilişkin gözlem noktalarının serpme diyagramını göz önüne alalım (Şekil 1.2). Bu gözlemlere karşılık gelen bağımlı Y_i değişkeninin ölçüm değerlerine en yakın düzlemin (modelin) bulunması , bizi en küçük kareler çözümüne götürmektedir. Bağımsız değişkenlere ilişkin gözlem noktalarının daha geniş bir alana yayılmış olması, düşünülen düzlemi ve dolayısıyla en küçük kareler çözümlerini daha kararlı kılacağı açıktır. Bağımsız değişkenlere ilişkin gözlem noktalarının aynı doğrultuda uzayıp gitmesi istenmeyen bir durum olmaktadır. Ayrıca bazı sapan değerler kararsız bir düzlemi kararlı veya kararlı bir düzlemi kararsız yapabilir.

iii) Belirli bir problem için verilerdeki küçük bir bozulmanın etkisi çözümde küçük bir değişmeye neden oluyorsa probleme "iyi koşullu", büyük bir değişmeye neden oluyorsa "zayıf koşullu" problem denir (Öztürk, 1992) .

Amaç $Ax = b$ kararlı denklem sisteminin çözümü olduğunda sağ taraftaki terimin değişmesi ile çözümdeki değişme arasındaki ilişkinin bir göstergesi fonksiyondaki matrisin

$$k(A) = \|A\| \|A^{-1}\|$$

koşul sayısıdır. A matrisinin birden küçük olmayan koşul sayısı $k(A)$ bire yakın olduğunda, denklem sistemini sağ taraftaki terimde yapılan değişikliklere karşı kararlıdır. Bu durumda A matrisine "iyi koşullu" matris denir.

Bir matrisin özdeğerlerinin mutlak değerce en büyüğü o matrisin boyunu aşmadığından

$$k(A) \geq \frac{\max_i \{|\lambda_i|\}}{\min_i \{|\lambda_i|\}}$$

dır. Burada λ_i , A'nın i. özdeğeridir.

β parametre vektörünün en küçük kareler tahmin edicisi $X'X\beta = X'Y$ normal denklemlerinin bir çözümü olmak üzere Y' nin gözlenen değerleri bir olasılık dağılımından gelmiş olmaları nedeni ile $X'X$ matrisinin zayıf koşullu olması durumunda en küçük kareler çözümleri kararsız olacaktır. En küçük kareler çözümlerindeki kararsızlık, kendini parametrelerin en küçük kareler tahmin edicilerinin varyansları üzerinde gösterecektir. Doğal olarak parametreye ilişkin tahmin edicilerin varyansları büyük olacaktır. Varyansların büyük

olması bir taraftan uyarıcı olmakla beraber, en küçük kareler tahmin edicilerinin en küçük varyansa sahip olma özelliğine güvenerek elde edilen istatistiksel sonuçlar bizi yanıltmış olacaktır.

iv) Buraya kadar en küçük kareler tahmin edicilerinin kararsızlığı sezgisel olarak açıklanmaya çalışılmıştır. En küçük kareler tahmin edicisi biraz daha yakından incelenirse,

$$\hat{\beta} = (X'X)^{-1} X'Y = \sum_{j=1}^p \lambda_j^{-1} \left(V_j' X'Y \right) \tilde{V}_j \quad (2.2)$$

$$\text{Cov}(\hat{\beta}) = \sigma^2 (X'X)^{-1}, \text{Var}(\hat{\beta}_j) = \sigma^2 (1 - R_j^2)^{-1}, j; 1, 2, \dots, p \quad (2.3)$$

$$\text{HKO}(\hat{\beta}) = E\left(\left\|\hat{\beta} - \beta\right\|^2\right) = \sigma^2 \text{tr}(X'X)^{-1} = \sigma^2 \sum_{j=1}^p \lambda_j^{-1} \quad (2.4)$$

$$E\left(\left\|\hat{\beta}\right\|^2\right) = \left\|\beta\right\|^2 + \text{HKO}(\hat{\beta}) \quad (2.5)$$

$$\text{Var}(a' \hat{\beta}) = a' (X'X)^{-1} a \sigma^2 = \sigma^2 \sum_{j=1}^p \lambda_j^{-1} \left(V_j' a \right)^2 \quad (2.6)$$

eşitliklerden de görüldüğü gibi, bazı R_j^2 (Çoklu korelasyon katsayısının karesi) değerleri bire yakın veya $X'X$ çok küçük özdeğerlere sahip olduğunda sorunlar ortaya çıkmaktadır. Eş. 2.2 de $V_j' X'Y$ nin değerleri sıfıra yakın değilse, $X'X$ in küçük özdeğerlerine karşılık gelen özvektörlerin önünde büyük çarpanlar ortaya çıkaracaktır. Böylece parametreye ilişkin tahmin değerleri mutlak değerce büyük ve yanlış işaretli olabileceklerdir. $X'X$, küçük özdeğerlere sahip olduğunda parametre tahminine ilişkin hata kare ortalaması

ve $\|\beta\|^2$ deki yansılık büyük olacaktır. Eş. 2.6 dan bazı doğrultularda öngörülenlerin güvenli olmayacağı görülmektedir.

En küçük kareler tahmin edicisi ve genelde doğrusal modellerle ilgili bu sorunlar "Çoklubağlantı" problemi ile ilgilidir.

2.5. Çoklubağlantının Tanımı

Çoklubağlantı, standartlaştırılmış bağımsız değişkenler ile oldukça kolay tanımlanabilir. Çok değişkenli doğrusal regresyon modeli

$$Y = \beta_0 \mathbf{1} + Z \beta + \varepsilon$$

şeklinde yazılabilir.

Burada Y , $n \times 1$ boyutlu bağımlı değişken vektörü, $Z = (Z_1, Z_2, \dots, Z_p)$, tam ranklı standartlaştırılmış bağımsız değişkenler matrisidir. Z matrisi, $Z'Z$ değişkenler arasındaki ilişki matrisini verecek şekilde standartlaştırılmıştır. β_0 bilinmeyen sabit, $\mathbf{1}$, $n \times 1$ boyutlu birler, β $p \times 1$ boyutlu bilinmeyen regresyon katsayıları ve ε , $n \times 1$ boyutlu gözlenemeyen rasgele hata terimleri vektörleri, $\varepsilon_i \sim N(0; \sigma^2 I)$ dir. (Gunst ,1983), tanımlamasında standartlaştırılmış Z matrisindeki çoklubağlantıyı Z' nin sütunları arasındaki ilişki olarak ifade etmektedir.

Z matrisinin sütun vektörlerinin , diklikten uzaklaşması durumunda çoklu bağlantıdan söz edilebilir. Sütun vektörleri doğrusal bağımlı olduğunda "Tam Çoklubağlantı" vardır denir. İki aşırı durum olan diklik ve bağımlılığın dışındaki durumların tümünde çoklubağlantı söz konusudur. Ancak bir varlık sorunundan fazla, var olması halinde bir derece sorunu olacağı açıktır. Bu

yüzden çoklubağlantı, Z nin sütun vektörlerinin yaklaşık olarak sıfır olan doğrusal fonksiyonu biçiminde ifade edilir.

Tanım: Eğer $\eta > 0$ (yeterince küçük) ve elemanlarının tümü sıfır olmayan $c' = (c_1, c_2, \dots, c_p)$ sabitler vektörü ,

$$\sum_{j=1}^p c_j Z_j = \delta \quad \|\delta\| < \eta \|c\| , \quad (\|c\| = (c'c)^{1/2} \text{ dir}).$$

eşitliğini sağlayacak şekilde bulunabiliyorsa Z nin sütunları arasında çoklubağlantı olduğu söylenebilir (Mason, Gunst, 1985).

Bu tanımlamanın yararlı olması için η 'in yeterince küçük seçilmesi gerekmektedir. Genelde tüm durumlar için sabit bir η seçimi uygun olmayacaktır. η 'nın değeri modelin uygulandığı araştırmaya bağlı olmakla beraber seçiminde kesin bir kural yoktur. Ancak

U; XX' matrisinin $n \times n$ lik özvektörler matrisi,

V; $X'X$ matrisinin $p \times p$ lik özvektörler matrisi,

D; $X'X$ matrisinin $p \times p$ lik köşegen özdeğerler matrisi,

λ_j ; $X'X$ matrisinin j. özdeğeri, olmak üzere

$$X = UD^{1/2}V'$$

yazılabilir. Burada i.gözlem

$$x_i = \sum_{j=1}^p \lambda_j^{1/2} u_{ij} v'_j$$

şeklinde gösterilebilir(Mandel, 1982).

$X = UD^{1/2}V'$ eşitliğinin her iki tarafı sağdan V ile çarpılırsa ($V'V=I$)

$$XV = UD^{1/2} \quad \text{olur. } j.\text{özvektör değeri}$$

$$X \underset{\sim}{v}_j = \lambda_j^{1/2} \underset{\sim}{u}_j \quad (2.7)$$

dir.Eş. 2.7. de $\lambda_j^{1/2}$ küçükse bu eşitlik bir çoklubağlantı tanımlar.

$V_j = (v_{1j}, v_{2j}, \dots, v_{pj})$ olmak üzere;(Tanımdan dolayı)

$$\sum_{k=1}^p v_{kj} x_k = \delta \quad \text{olur. Açılırsa}$$

$$v_{1j} \underset{\sim}{x}_1 + v_{2j} \underset{\sim}{x}_2 + \dots + v_{pj} \underset{\sim}{x}_p = \delta_{(n \times 1)} \quad (2.8)$$

yazılabilir.Eş. 2.7 ve Eş. 2.8 eşitlenirse

$$\delta = \lambda_j^{1/2} u_j \quad \text{elde edilir.}$$

$$\|\delta\| = (\delta' \delta)^{1/2} = (\lambda_j^{1/2} u_j' u_j \lambda_j^{1/2})^{1/2} \quad \text{olur.}$$

U matrisi dik olduğundan $(u_j' u_j) = 1$ dir.Böylece,

$$\|\delta\| = \lambda_j^{1/2} \quad \text{olur. Tanımda yer alan}$$

$$\|\delta\| < \eta \quad \text{koşulunda } c = v_j \text{ alınırsa}$$

$$\|\delta\| < \eta \quad \|v_j\| \quad X'X \text{ matrisinin } j. \text{ özvektörü olan } v_j \text{ nin boyu 1 olduğundan}$$

$$\|\delta\| < \eta \quad \text{olur ve}$$

$$\lambda_j^{1/2} < \eta$$

elde edilir.. η nin küçük ve sıfıra yakın olması tercih edildiğinden $\lambda_j^{1/2}$ nin çok yakınında bir değer olarak seçilmesi tercih edilmelidir.

2.6. Çoklu Bağlantının En Küçük Kareler Tahmin Edicileri Üzerindeki Etkisi

Çoklubağlantının en küçük kareler tahmin edicileri üzerindeki etkileri

i) Şişirilmiş Varyanslar

ii) Şişirilmiş Tahminler

şeklinde iki başlık altında özetlenebilir. Ayrıca bir regresyon çalışmasında çoklubağlantının varlığı , modelin yanlış tanımlanmasına sebebiyet verebilir. Amacımız model belirlemekten çok, çoklubağlantının en küçük kareler tahmin edicilerine etkilerini incelemektir. Şişirilmiş varyanslar ve tahminler iki bağımsız değişken için aşağıdaki gibi gösterilebilir. Bağımsız değişkenler ve bağımlı Y $\sum_{i=1}^n Z_{ij} = 0$, $\sum_{i=1}^n Z_{ij}^2 = 1$ ve $\sum_{i=1}^n Y_i = 0$, $\sum_{i=1}^n Y_i^2 = 1$ olacak biçimde standartlaştırılır iseler, r ilişki katsayısını göstermek üzere;

$$r_{x_j x_\ell} = \sum_{i=1}^n Z_{ij} Z_{i\ell} = r_{j\ell} \quad \text{ve}$$

$$r_{x_{jy}} = \sum_{i=1}^n Z_{ij} y_i = r_{jy} \quad \text{yazılabilir. Veya matris biçiminde}$$

$$r_{i\ell} = Z_j' X_\ell$$

$$r_{jy} = Z_j' Y$$

olur. Böylece $p = 2$ için

$$Z'Z = \begin{bmatrix} 1 & r_{12} \\ r_{21} & 1 \end{bmatrix} \text{ matrisinin tersi}$$

$$(Z'Z)^{-1} = \begin{bmatrix} 1/(1-r_{12}^2) & -r_{12}/(1-r_{12}^2) \\ -r_{12}/(1-r_{12}^2) & 1/(1-r_{12}^2) \end{bmatrix}$$

olur. Buradan,

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

$$\hat{\beta} = \begin{bmatrix} (r_{1y} - r_{12}r_{2y})/(1-r_{12}^2) \\ (r_{2y} - r_{12}r_{1y})/(1-r_{12}^2) \end{bmatrix}$$

olarak bulunur. $(Z'Z)^{-1}$ ve $\hat{\beta}$ eşitliklerinde $r_{12} \rightarrow \pm 1$ olması $1-r_{12}^2$ den dolayı tüm değerleri büyütecektir. $\hat{\beta}$ de $r_{1y} - r_{12}r_{2y}$ den dolayı işaret farklılaşmasına, $V(\hat{\beta}) = \sigma^2 (Z'Z)^{-1}$ den dolayı büyük varyanslara (şişirilmiş) yol açacaktır. p tane bağımsız değişkenin olması durumunda $(Z'Z)^{-1} = C$ matrisinin köşegen elemanları

$$C_{jj} = (1-R_j^2)^{-1}, \quad j = 1, 2, \dots, p$$

olacaktır. R_j^2 , j. bağımsız değişkenin diğer bağımsız değişkenlerce açıklanan miktarıdır. $R_j^2 \rightarrow 1$ olması C_{jj} değerlerinin şişmesine yol açar. Bu durum tahmin ve varyanslarının önemli ölçüde, yapay olarak büyümesine neden olur.

Diğer yandan $\lambda_1 \leq \lambda_2 \dots \leq \lambda_p$, $Z'Z$ nin özdeğerleri ve $V_{p \times p} Z'Z$ nin özvektörler matrisi olmak üzere ;

$$(Z'Z)^{-1} = \sum_{j=1}^p \lambda_j^{-1} V_j V_j'$$

biçiminde de yazılabilir. Böylece

$$\hat{\beta} = \sum_{j=1}^p \lambda_j^{-1} V_j Z' Y V_j' \quad \text{ve}$$

$$V(\hat{\beta}) = \sigma^2 \sum_{j=1}^p \lambda_j^{-1} V_j V_j'$$

bulunur. $\lambda_j \rightarrow 0$ için $\hat{\beta}$ ve $V(\hat{\beta})$ nin şişirilmiş olarak elde edileceği açıktır.

2.7. Çoklubağlantı Durumunda Parametre Tahmini ve Kestirim

Kestirimler, özdeğerler türünden tanımlandığında bir z_i gözlemi için $V(\hat{y})$,

$$\hat{y}_i = z_i' \hat{\beta}, \quad (i; 1, 2, \dots, n)$$

$$\hat{y}_i = z_i' (Z'Z)^{-1} Z' Y \quad \text{ve}$$

$$V(\hat{y}) = z_i' \left\{ \sum_{j=1}^p \lambda_j^{-1} V_j V_j' \right\} z_i \sigma^2$$

olarak yazılır. İşlemler sırasında $\lambda_{enk} \approx 0$ 'a karşılık gelen V_1 özvektörü için

$$z_i' V_1 \approx 0$$

olacaktır. Çoklubağlantıdaki değişkenler için doğrusal bağıntı yaklaşık sıfır, çoklubağlantı dışındaki değişkenler için v_{li} ler yaklaşık sıfırdır. Böylece, $\hat{\beta}$ için ortaya çıkan sorunlar kestirim için söz konusu olmayacaktır. Tek tek

parametre tahminleri doğru olmasa bile doğru bir kestirim çoklubağlantı durumunda yapılabilir. Bununla birlikte kestirimin kabul edilebilirliği çoklubağlantının tanımlandığı bölge için de geçerli olacaktır. "Ön kestirilebilirlik bölgesi" adı verilen bu bölge,

$$\mathcal{R} = \left\{ Z : \left| \frac{Z}{\tilde{V}_j} \right| \leq c_j, \quad j=1, 2, \dots, p \text{ ve } a_\ell \leq Z_\ell \leq b_\ell \right\}$$

olarak verilir. a_ℓ ve b_ℓ , ℓ . değişkeninin en büyük ve en küçük standartlaştırılmış değerleridir. c_j için iki yol gösterilir.

i) $c_j = \sqrt{\lambda_j}$, j indisi en küçük özdeğere karşılık gelir. Ve

ii) $c_j = \text{enb} \left\{ \left| \frac{Z_i}{\tilde{V}_j} \right|, \quad i = 1, 2, \dots, n \right\}$ dır. Burada Z_i , Z nin i . satırıdır.

V_j yukardaki gibi en küçük özdeğere karşılık gelen özvektördür.

(i) ve (ii) iç değer bulma amacı ile kullanılabilir. Ancak (ii). daha dar bir bölge tanımlar.

Diğer taraftan son yıllarda kullanılan bir başka yöntem de, $A = [X:Y]$ için

$$A'A = \left[\begin{array}{c|c} X'X & X'Y \\ \hline X'Y & 1 \end{array} \right]$$

parçalı matrisinin özdeğer ve özvektörlerinin kullanımınıdır. Özdeğer regresyonu (Latent Root Regression) adı verilen bu yöneme göre,

ℓ_j ; $A'A$ nın özdeğerleri,

W_j ; ℓ_j ye karşılık gelen özvektör,

olmak üzere

$$\ell_{p-1} \ll \lambda_p \text{ ve } |W_{p+1,p-1}| \neq 0$$

ise en küçük kareler tahmini kestirim amacı ile kullanılabilir.

2.8. Çoklubağlantının Belirlenmesi

Kısaca iki yada daha çok bağımsız değişken arasındaki ilişki olarak tanımlanan çoklubağlantının belirlenmesi , aşağıdaki yaygın gösterimler ile açıklanmaktadır.

$$R = (r_{ij}) \text{ ve}$$

$$R^{-1} = (r^{ij})$$

bağımsız değişkenler arasındaki basit korelasyon matrisini ve bunun tersini , λ_i , $i = 1, 2, \dots, p$ ($\lambda_p \leq \lambda_{p-1} \leq \dots \leq \lambda_1$), R nin özdeğerleri V_i ($i = 1, \dots, p$), R ' nin ortonormal özvektörlerini , D , köşegen elemanları λ_i ler olan köşegen matrisi V ise sütunları V_i ler olan özvektörler matrisini ifade etmek üzere,

$$RV = VD \tag{2.9}$$

yazılabilir. Burada $V'V = VV' = I$ dır.

Çoklubağlantıyı teşhis etmekte kullanılan bazı yaygın gösterimler

i) Basit ikili korelasyonlar,

$$|r_{ij}| = 1$$

ii) Çoklu korelasyon katsayısının karesi,

$$R_i^2 = 1 - \frac{1}{r^{ii}} > 0.9$$

iii) Varyans artırıcı faktörler (VIF),

$$VIF = r^{ii} > 10$$

iv) Özdeğerler,

$$\lambda_i \cong 0$$

dir(Walker,1989).Bu gösterimlerden ilki, ikili bağımsız değişkenler arasındaki basit ilişki katsayısı r_{ij} , x_i ve x_j arasındaki basit ilişkiyi ortaya çıkarabilir. Böylece $|r_{ij}| = 1$ olması durumunda i. ve j. değişkenler tam orantılı demektir.

Gösterimlerden ikincisi R_i^2 , x_i ile geri kalan X değişkenleri arasındaki çoklu ilişki katsayısının karesidir. Gösterimin derecesi, X_i 'nin diğer bağımsız değişkenlerin hepsi ile olan doğrusal ilişkinin gücünü açıklar.

Gösterimlerden üçüncüsü, standartlaştırılmış değişkenlere ilişkin korelasyon matrisinin tersinin, köşegen elemanlarıdır. Marquardt (1970) tarafından VIF (Variance Influence Factors) olarak adlandırılmıştır. Değişkenlerin standartlaştırılması, toplamı sıfır, kareler toplamı bir olacak şekilde olmaktadır. Regresyon modelinde i. parametrenin en küçük kareler tahminin varyansı r^{ii} ile orantılıdır. i. dik değişken göz önünde bulundurulursa $r^{ii} = 1$ olması şişme için göreceli bir miktardır. Pratik olarak $VIF_i > 10$ olması, R_i ile VIF_i arasındaki basit ilişkiye dayanmaktadır. $VIF_i > 10$ 'a $R_i^2 > 0.9$ uygun düşmektedir. $X'X$, ilişki matrisi R 'yi verecek formda iken i. bağımsız değişkenin varyans artırıcı faktörü VIF_i

$$VIF_i = \frac{1}{1 - R_i^2} \quad (2.10)$$

yazılabilir.Burada, R_i i. bağımsız değişkeni ile geri kalan bağımsız değişkenler arasındaki çoklu ilişki katsayısıdır. V ve D matrislerinin elemanlarından VIF,

$$VIF_i = \sum_{j=1}^p V_{ij}^2 / \lambda_j \quad (2.11)$$

şeklinde de ifade edilmektedir(Walker,E.,1989).Burada Küçük özdeğerlerin VIF_i değerlerini şişirebileceği görülmektedir.

Gösterimlerden dördüncüsü; Bağımsız değişkenler arasındaki doğrusal ilişki, R nin özdeğer ve özvektör analizinden ortaya çıkarılabilir. Özdeğerin sıfıra çok yakın olması Çoklubağlantının belirlenmesinde önemlidir. Bu değere uygun öz vektör doğrusal ilişkinin varlığını belirleyecektir.Tekil Değer Ayrışımı (Singular Value Decomposition) , (Mandel, 1982; Searle, et. al., 1981) örneklerinde görüldüğü gibi kullanılırsa, X matrisi

$$X = UDV'$$

şeklinde de yazılabilir. u_{ij} , XX' matrisinin özvektörler matrisi olan U'nun i.satır j.sütun elemanıdır.U ve V aşağıdaki özellikleri sağlarlar,

$$U'U = VV' = V'V = I \quad \text{ve}$$

$$UU' = H = X (X'X)^{-1} X' \quad \text{dır.} X'X \text{ Tekil Değer Ayrışımın 'dan}$$

$$X'X = VD^2 V' \text{ ve } V'X'XV = D^2 \quad \text{olur.}$$

X ilişki matrisini ($R = Z'Z$) verecek şekilde standartlaştırıldıktan sonra $R = Z'Z$ matrisinden

$$V_i'RV_i = (Z_{(-i)}V_i)' (Z_{(-i)}V_i) = \lambda_i \quad (2.12)$$

yazılabilir.

Burada $Z_{(-1)}$ nxm boyutlu standartlaştırılmış değişkenler matrisidir (1 sabitler sütunu çıkarılmış iken). Eş. 2.12 de $\lambda_i = 0$ dan anlaşılan $Z_{(-1)}V_i = 0$ olmasıdır. Böylece, $V_i = \{v_{ij}, j = 1, 2, \dots, m, p = m+1\}$, olmak üzere öne sürülen ilişki;

$$\sum_{j=1}^m v_{ij} Z_{(-1)j} = 0$$

dır. (Hocking, Pendleton, 1983). Burada Z_j standartlaştırılmış j. değişkendir. V_i 'nin büyük elemanları doğrusal ilişki katsayıları olarak öne sürülmektedir.

Belsley (1980), öne sürdüğü analizi pxp boyutlu $R = Z'Z$ 'ye uygulamıştır. Burada Z , X in herbir sütununun karelerinin toplamı bir olacak şekilde standartlaştırılması ile elde edilmiştir. Ancak değişkenler ortalamaları etrafında merkezileştirilmemiş olup Z sabit sütunu da kapsamaktadır. R 'nin sıfıra yakın özdeğeri çoklubağlantının varlığını göstermektedir. $V_i = (v_{ij}, j = 0, 1, \dots, p)$ için,

$$\frac{1}{\sqrt{n}} V_{i0} + \sum_{j=1}^m v_{ij} Z_j = 0$$

olmaktadır. R nin özdeğerlerinin sıfır olması değişkenlerden bazıları arasında tam bir ilişkinin olduğu anlamına gelir. Sıfırdan farklı (sıfıra çok yakın) özdeğerin büyüklüğü, çoklubağlantının şiddetini gösterir. Çoklubağlantı içinde bulunan değişkenlere karşılık, sıfırdan farklı v_{ij} ler, çoklubağlantı dışında bulunanlara ise sıfır yada sıfıra yakın v_{ij} ler karşılık gelir.

Çoklubağlantının belirlenmesinden önce X veri matrisinin sütunları merkezileştirilebilir. Merkezileştirme hakkında öne sürülen farklı fikirler için Belsley (1984) ve Myers (1986) 'e bakılabilir. İyi kabul edilen çoklubağlantı ölçüsü "Koşullu İndis",

$$n_k = \frac{\lambda_1}{\lambda_k}, \quad k = 2, 3, \dots, p \quad (2.13)$$

şeklinde gösterilir(Montgomery,Peck,1982).(2.13) teki değerlerin büyük çıkması çoklubağlantıyı göstermektedir. Her bir $n_p = \lambda_1/\lambda_p$ koşul numarası olarak adlandırılır. En küçük kareler tahmin edicilerinin kararlı olup olmadığının önemli belirtisidir. n_p , $100 < n_p < 1000$ ise olduğunda güçlü çoklubağlantı , $n_p > 1000$ iken şiddetli çoklubağlantının olduğu öne sürülmüştür. Belsley (1980) de çoklubağlantı ölçüsünü

$$n_i = \sqrt{\frac{\lambda_{\max}}{\lambda_i}} > 30$$

olarak öne sürmüştür.

R nin Eş. 2.9 da olduğu gibi özdeğer ve özvektörlerine ayrıştırılmasının bazı dezavantajları bulunmaktadır.R için X de düzenlemeye (standartlaştırma) ihtiyaç duyulmasından dolayı, X 'te az da olsa bozulma olmuşsa özdeğerler durağan olmazlar. Bir dereceye kadar daha durağan olabilecek bir alternatif $Z_{(-1)}$ yada Z'nin tekil değer ayrıştırımının tercih edilmesidir (Belsley,1980).

Z'nin T.D.A mı

$$Z = UDV'$$

dır.Burada $U'U = V'V = I$ ve $D =$ Köşegen (μ_i) dir. μ_i ler Z nin tekil değerleri olarak adlandırılır ve bunların karekökleri $R = Z'Z$ nin özdeğerleridir. $(\mu_i = \sqrt{\lambda_i})$, V , R 'nin , U ise ZZ' nin özvektörler matrisidir.

2.9. Çoklu Bağlantının Etkisinin Ortadan Kaldırılması

Çoklubağlantının , en küçük kareler tahmin edicileri üzerindeki etkileri, β parametrelerine ilişkin tahmin edicilerin varyanslarının şişirilmiş , β parametrelerine ilişkin tahmin edicilerin yanlış işaretli ve kestirimlerinin şişirilmiş olabileceği şeklinde tanımlanmaktadır.

En küçük kareler yöntemini uygularken bu istenmeyen durumlardan kaçınabilmek için çoklubağlantı durumunun ortadan kaldırılması gerekmektedir. Çoklubağlantının ortadan kaldırılması için önerilen yöntemlerden biri değişken eliminasyon yöntemidir. Bu yöntem, aralarında kuvvetli doğrusal ilişki bulunan bir yada bir kaç değişkenden Y yi açıklama gücü en az olan bir yada bir kaçının modelden çıkarılması işlemini kapsamaktadır. Çoklubağlantının ortadan kaldırılması için önerilen bazı yanlış tahmin ediciler de vardır. Bunlardan en çok kullanılanlardan ikisi Ridge regresyon ve Temel Bileşenler regresyonudur. Ridge regresyonu, çoklu bağlantıdan dolayı ortaya çıkan $X'X$ matrisindeki tekilliğe yakın durumu bu matrisin köşegenine sabit eklemek suretiyle tekil olmaktan uzak bir matris elde etmek amacı ile ortaya atılmıştır. Yani, eğer $X'X$ tekil ise, o zaman $(X'X + kI)$ tekil değildir (burada k pozitif bir sayıdır). Köşegen elemanlarına yapılan bu ekleme, bağımsız değişkenler arasındaki tekilliğe yakın olma durumunu önlemesinden dolayı önemli bir sorun olarak görülmemektedir. Temel bileşenler regresyon analizinde çoklubağlantı problemine, X uzayında çoklubağlantıya sebep olan boyutların elimine edilmesi itibarıyla yaklaşmaktadır. Bağımsız değişkenin modelden çıkarılması modelin Y 'yi açıklamakta yetersiz kalmasına sebep olur. Bununla beraber, temel bileşenler regresyonunda, boyut düşürmek itibarıyla bu tek bağımsız değişkenin dışında, geri kalan değişkenlerin doğrusal kombinasyonu olarak tanımlanır.

Çoklubağlantının etkisinin ortadan kaldırılması için düşünülen diğer bir yöntem de etkili değer analizidir. Bir yada daha fazla etkili gözlem çoklu bağlantıya sebep veriyor olabilir. Bu amaçla bir sonraki bölümde etkili gözlemler, etkili gözlemlerin çoklubağlantıya etkisi ve bu gözlemlerin belirlenmesi ayrıntılı olarak incelenecektir.

3. ÇOKLUBAĞLANTIYI ŞİDDETLENDİREN ETKİLİ GÖZLEMLER

3.1 Etkili Gözlemin Tanımı

Alışılmış en küçük kareler yönteminde bütün gözlemlere aynı ağırlık verilmektedir. Bununla beraber, bütün gözlemler en küçük karelerin sonuçları üzerinde aynı oranda etkiye sahip değildir. Örneğin, basit bir doğrusal regresyon probleminde bağımlı değişkenin ortalamasından çok uzakta olan gözlem yada gözlemler oldukça etkilidir. Diğer gözlemlerden uzakta olan tek bir nokta regresyon sonuçları üzerinde diğer noktaların tümünden daha büyük etkiye sebep olabilir. Böyle gözlemler etkili (influential) noktalar yada izdüşüm değeri yüksek noktalar olarak adlandırılırlar.

Muhtemelen en küçük karelerin sonuçlarında etkili olan gözlem, X uzayındaki diğer gözlemlerle olan pozisyonundan belirlenir. Genelde, böyle gözlemler X uzayındaki gözlemlerin merkezinden uzaktadır. Regresyon sonuçlarını etkileyeceği için muhtemelen bu gözlem yüksek etkiye sahip olacaktır.

3.2. Etkili Gözlemin Çoklubağlantıya Etkisi

3.2.1. Çoklubağlantıyı maskeleyen veya şiddetlendiren etkili gözlem

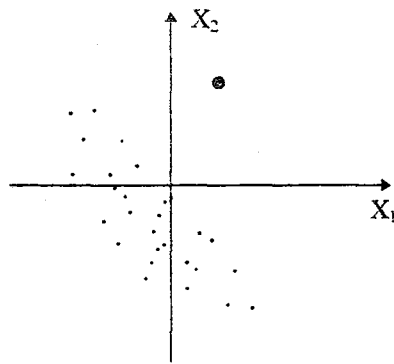
Çoklubağlantıyı daha önce X 'in sütunları arasındaki ilişki olarak tanımlamıştık. Yalnız son zamanlarda, verilerdeki bir kaç gözlemin çoklu bağlantının yapısını kuvvetli bir şekilde etkileyebileceği öne sürülmektedir (Belsley, et al., 1980; Dorsett, et al., 1983; Mason, Gunst, 1985).

Verinin çoklubağlantı yapısında yüksek etkiye sahip gözlemler çoklu bağlantıyı etkileyen (Collinearity-Influential) gözlemler olarak adlandırılmaktadır. Daha açık tanımlanırsa,

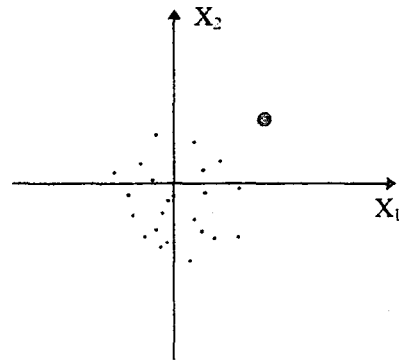
Tanım: $x_{\sim i}$, X matrisinin i. satırı ($i = 1, 2, \dots, n$) ve n_k ile $n_k(i)$ de sırasıyla $x_{\sim i}$ dahil olduğunda ve olmadığında hesaplanan koşul sayıları (conditional index) olmak üzere, eğer $x_{\sim i}$ çoklubağlantıyı etkileyen nokta ise, önceden kararlaştırılan bir δ değeri için;

$$|n_k - n_k(i)| > \delta n_k, \quad k = 2, 3, \dots, p$$

dir. Çoklubağlantıyı etkileyen gözlemler iki grupta sınıflandırılmaktadırlar, çoklubağlantıyı maskeleyen (Masking-Collinearity) ve çoklubağlantıyı şiddetlendiren (Induce-Collinearity) gözlemler. Bu iki durumda aşağıdaki şekilde izah edilmektedir.



a) Çoklubağlantıyı maskeleyen
($n_2 \ll n_2(i)$)



b) Çoklubağlantıyı şiddetlendiren
($n_2 \gg n_2(i)$)

Şekil 3.1. Çoklubağlantıyı maskeleyen ve şiddetlendiren noktalar

Şekil 3.1 de tanımlanan çoklubağlantıyı etkileyen noktalar genel bir gösterimdir. Her ikiside verinin hacmi ile son derece ilgilidir. i. gözlemin izdüşüm değeri h_i ile gösterilir, h_i , $H = X(X'X)^{-1}X'$ izdüşüm (hat) matrisinin i. köşegen elemanıdır. Hoaglin ve Welsch (1978) , eğer $h_i > 2p/n$ ise noktanın yüksek izdüşüm değerine sahip olduğunu ileri sürmektedirler.

Çoklubağlantıyı etkileyen gözlem için gerekli koşul , bunun izdüşüm değeri yüksek gözlem olmasıdır. Mason ve Gunst (1985) , bir gözlemin izdüşüm değerinin arttığında çoklubağlantı ölçüsünün arttığını ispatlamışlardır. Dorsett (1983) te çoklubağlantıyı etkileyen gözlemlerin $X'X$ matrisinin özdeğer yapısında ortaya çıkardığı sonuçlar üzerinde durmuştur. Bu iki yaklaşıma, çoklubağlantıyı etkileyen gözlemlerin , çoklubağlantının yapısı üzerinde ne denli etkili olduklarını göstermek bakımından bir sonraki bölümde ayrıntılı bir şekilde yer verilecektir.

3.2.2. Bir gözlemin merkezden uzaklaştıkça çoklubağlantıya olan etkisi

Çoklubağlantı, standartlaştırılmış bağımsız değişkenler üzerinden oldukça kolay irdelenebilir , bu sebeple çoklu doğrusal regresyon modeli,

$$Y = \beta_0 \mathbf{1} + Z\beta + \varepsilon$$

şeklinde ifade edebilir.

Burada Y , $n \times 1$ boyutlu bağımlı değişken vektörü, $Z = (Z_1, Z_2, \dots, Z_p)$ tam ranklı sütun toplamı sıfır, kareleri toplamı bir olacak şekilde standartlaştırılmış bağımsız değişkenler matrisi ($Z'Z$ korelasyon matrisini verir), β_0 bilinmeyen sabit, $\mathbf{1}$, $n \times 1$ boyutlu elemanları 1 olan vektör, β , $p \times 1$ boyutlu bilinmeyen regresyon katsayıları vektörü ve ε , $n \times 1$ boyutlu gözlenemeyen rasgele hata terimleri vektörü ($\varepsilon_i \sim \text{NID}(0; \sigma^2)$) olmak üzere,

Gunst (1983) teki tanımlamasında sabitlenmiş (fixed) Z matrisindeki çoklubağlantıyı, Z 'nin sütunları arasındaki ilişki olarak ifade etmiştir.

Z nin sütunları arasındaki çoklubağlantıya izdüşüm değeri yüksek noktaların sebep olduğuna dair bir imaya muhakkak ihtiyaç yoktur. Bununla beraber, izdüşüm değeri yüksek noktalar çoklubağlantıya sebebiyet verebilirler, Mason ve Gunst (1985) teki ıspatlarında görüldüğü gibi, $p = 2$ için model;

$$Y = \beta_0 + \beta_1 Z_1 + \beta_2 Z_2 + \varepsilon$$

olur. Burada sabitleştirilmiş gözlem değerleri $s_i = (s_{i1}, s_{i2})$ $i = 1, 2, \dots, n$ olmak üzere

$i = 1$ için, $\theta s = (\theta s_{11}, \theta s_{12})$, $(\theta; s_{11}, s_{12} > 0)$ ve

$i \neq 1$ için, $S = [s_{i1}, s_{i2}]$

olsun. Burada S , $(n-1) \times 2$ boyutlu matristir . $\theta \rightarrow \infty$ için geri kalan gözlemlere göre, ilk gözlem değerleri daha büyük olacaktır. 1 sabitler vektörünü içermeyen X matrisi;

$$X = \begin{bmatrix} \theta s \\ S \end{bmatrix}$$

şeklinde yazılır . O zaman $X'X = \{(n-1)/n\} \theta^2 s' s + S' S$ olur . Teknik olarak $S' S$ 'in bileşenleri θ 'yı içermektedir. Fakat bu bileşenler n ile bölündüklerinden , büyük örnek hacmi için $S' S$ teki bu bileşenlere önem verilmemiştir. Buna göre sabit terimide içeren bağımsız değişkenler matrisinin birinci gözlemine ilişkin izdüşüm değeri;

$$h_1 = 1/n + \{(n-1)/n\}^2 \theta^2 s' (X'X)^{-1} s$$

şeklinde ifade edilir. Parçalanmış matrislerin tersini almak için gerekli formül kullanılırsa;

$$h_1 = 1/n + \{(n-1)/n\}^2 \theta^2 s' \left\{ \frac{(S'S)^{-1} - \{(n-1)/n\} \theta^2 (S'S)^{-1} s' s (S'S)^{-1}}{(1 + \{(n-1)/n\} \theta^2 s' (S'S)^{-1} s)} \right\} s$$

olur ve $d = \{(n-1)/n\} s' (S'S)^{-1} s$ (Sadece s_i nin fonksiyonu olduğundan $d > 0$ dır) kabul edildiğinde;

$$h_1 = n^{-1} \{1 + (n-1) \theta^2 d / (1 + \theta^2 d)\}$$

şeklinde elde edilir. Böylece $\theta \rightarrow \infty$ iken, $h_1 \rightarrow 1$, özel olarak θ keyfi derecede büyükse $h_1 > 2(p+1)/n$ olur. Yani bir gözlem diğerlerinden uzaklaştıkça izdüşüm değeri 1'e yaklaşmaktadır. Bu ise bizim için, o gözlemin etkili bir gözlem olması anlamına gelmektedir.

Ayrıca, $\theta \rightarrow \infty$ iken, standartlaştırılmış veri matrisinin elemanları da

$$Z_{1j} \rightarrow \left\{ \frac{n-1}{n} \right\}^{1/2}, j = 1, 2$$

$$Z_{ij} \rightarrow -\{n(n-1)\}^{-1/2}, i \neq 1, j = 1, 2$$

değerlerine yakınsamaktadır (sonuçların elde edilmesi Ek 1 'de verilmiştir.). Böylece, θ keyfi derecede büyük iken $Z_1 - Z_2$ sonucu keyfi olarak sıfır yapılmış olur. Bu ise etkili bir gözlem olan ilk gözlemin çoklubağlantıyı şiddetlendirebileceğinin bir göstergesidir.

3.2.3. Çoklubağlantıya neden olan gözlemin özdeğer analizi ile incelenmesi

Dorsett (1983) makalesinde, çoklubağlantıya neden olan (Collinearity-Influential) gözlemler sebebiyle $X'X$ matrisinin özdeğer yapısında oluşan farklılığı aşağıdaki teorem ile vermektedir.

Teorem 1: x_i , X matrisinin i . satırı, V_j , $X'X$ in j . özvektörü ve $X_{(i)}$ de X matrisinin i . satırı çıkartıldıktan sonraki veri matrisi olsun. Bu durumda eğer,

$$x_{\sim i} = \alpha V_{\sim j}$$

ise $X'X$ matrisinin özdeğerleri $\lambda_1, \lambda_2, \dots, \lambda_{j-1}, \lambda_j, \lambda_{j+1}, \dots, \lambda_p$ olmak üzere $X'_{(i)} X_{(i)}$ nin büyükten küçüğe sıralanmış özdeğerleri $\lambda_1, \lambda_2, \dots, \lambda_{j-1}, \lambda_j - \alpha^2, \lambda_{j+1}, \dots, \lambda_p$ olmaktadır.

Böylece, j . özvektör yönünde yer alan gözlem j . özdeğeri şişirir. Eğer α küçük ise o zaman çıkartılan nokta , çoklubağlantının yapısında önemli etkiye sahip olmayacaktır.

Gözlemin izdüşüm değeri h_i ile α arasında doğrusal bir ilişki vardır. Gerçekten

$X=UDV'$ nün tekil değer ayrışımından izdüşüm matrisi $H = X(X'X)^{-1}X'$ U , XX' matrisinin özvektörler matrisi olmak üzere.

$$H=UU'$$

şeklinde yazılabilir. Burada i . gözleme ilişkin dik izdüşüm değeri

$$h_i = \sum_{k=1}^p u_{ik}^2$$

dir. Aynı şekilde i.gözlem;

$$\underset{\sim}{x}_i = \sum_{k=1}^p u_{ik} \lambda_k^{1/2} V_k \text{ olur.}$$

Eğer $\underset{\sim}{x}_i = \alpha V_j$ ise ;

$k \neq j$ için $u_{ik} = 0$ ve

$k = j$ için $u_{ij} = \alpha / \lambda_j^{1/2}$ olur. Yani,

$$\begin{aligned} \underset{\sim}{x}_i = \alpha V_j &= u_{i1} \lambda_1^{1/2} V_1 + \dots + u_{ip} \lambda_p^{1/2} V_p \Rightarrow k \neq j \text{ olduğunda } u_{ik} = 0 \text{ olduğu için} \\ &= u_{ij} \lambda_j^{1/2} V_j \end{aligned}$$

dir. Buradan,

$$\alpha V_j = u_{ij} \lambda_j^{1/2} V_j$$

$$u_{ij} = \alpha / \lambda_j^{1/2}$$

olarak elde edilir. Ayrıca

$$h_i = \sum_{k=1}^p u_{ik}^2 \text{ olduğundan}$$

$$h_i = \frac{\alpha^2}{\lambda_j} \text{ olur.}$$

Bu teoreme ilişkin aşağıda verilecek sonuç, α ve h_i arasındaki doğrusal ilişkiyi kullanmakta olup bir çoklubağlantı ölçüsü olan "Koşullu İndis" in değerinde (sonucunda), izdüşüm değerinin (h_i) önemini açıkça göstermektedir.

Sonuç 1:

$$x_i = \alpha V_j \quad \text{iken}$$

İ) Eğer $j = 1$ ise,

$$|n_k - n_k(i)| = h_i n_k \quad ; k = 2, 3, \dots, p \text{ için}$$

İİ) Eğer $j \neq 1$ ise,

$$|n_k - n_k(i)| = \begin{cases} n_k (h_i / (1 - h_i)) & ; k = j \text{ için} \\ 0 & ; k \neq j \text{ için} \end{cases}$$

dir.

İspat:

$$\text{İ) } n_k(i) = \frac{\lambda_1(i)}{\lambda_k(i)} = \frac{\lambda_1 - \alpha^2}{\lambda_k} = \frac{(1 - h_i)\lambda_1}{\lambda_k} = n_k(1 - h_i) \quad \text{olduğundan}$$

$$|n_k - n_k(i)| = h_i n_k \quad \text{olur.}$$

$$\text{İİ) } n_k(i) = \frac{\lambda_1(i)}{\lambda_k(i)} = \frac{\lambda_1}{\lambda_k - \alpha^2} = \frac{\lambda_1}{\lambda_k(1 - h_i)} = n_k(1 / (1 - h_i)) \quad \text{olduğundan}$$

$k=j$ için $|n_k - n_k(i)| = n_k (h_i / (1 - h_i))$ elde edilir, diğer taraftan $k \neq j$ ise $n_k(i) = n_k$ olur. Böylece, koşullu indis üzerinde izdüşüm değerinin rolü açıkça görülmektedir.

Diğer taraftan, $x_{\sim i}$ nin V_j lerin herhangi birinin tam doğrultusunda olmadığını düşünürsek, böyle bir $x_{\sim i}$

$$x_{\sim i} = \sum_{j=1}^p \alpha_j V_j$$

şeklinde gösterilir ise, x_i gözlemine ilişkin izdüşüm değeri h_i (u_{ij} , XX' matrisinin özvektörler matrisi olan U nun i,j . elemanı olmak üzere) de

$$h_i = \sum_{j=1}^p u_{ij}^2 \quad (3.1)$$

şeklinde ifade edilebilir.

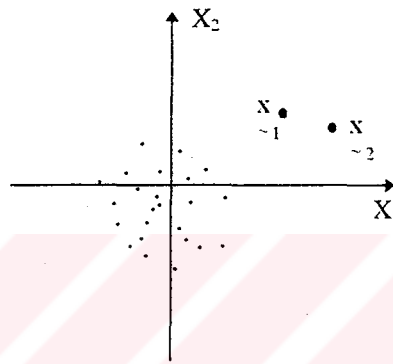
Velleman ve Ypelear (1980), u_{ij} 'leri dik izdüşüm bileşenleri(orthogonal leverage component) olarak adlandırmışlardır. Büyük değerli u_{ij} değerlerinden herhangi birinin, (u_{ik}), çıkarılan $x_{\sim i}$ gözleminin k . koşullu indisi etkileyebileceğinin, bir göstergesi olduğunu ileri sürmektedirler. u_{ij}^2 değeri p/n den büyük ise $x_{\sim i}$ gözleminin çıkarılması k . koşullu indisi etkileyecektir. u_{ij}^2 için p/n sınır değeri h_i 'nin sınır değeri olan $2p/n$ 'e dayandırılmaktadır.

Şimdiye kadar tek bir etkili gözlemin var olması durumu üzerinde duruldu. Birden fazla etkili gözlemin olması durumunda Dorsett (1983) tanımladığı teoremi çoklu duruma genişletmiştir.

Teorem 2: Eğer iki gözlem $x_{\sim 1} = \alpha V_{\sim i}$ ve $x_{\sim 2} = \beta V_j$ X matrisinden çıkarılırsa, özdeğerler $\lambda_1, \lambda_2, \dots, \lambda_i - \alpha^2, \dots, \lambda_j - \beta^2, \dots, \lambda_p$ olur. Burada $x_{\sim 1}$ ve $x_{\sim 2}$ gözlemleri farklı özvektörler yönündedirler. Aynı özvektör yönünde olması durumunda ise,

Sonuç 2: Eğer iki gözlem, $\tilde{x}_1 = \alpha V_i$ ve $\tilde{x}_2 = \beta V_i$, X matrisinden çıkarılırsa, özdeğerler $\lambda_1, \lambda_2, \dots, \lambda_i - \alpha^2 - \beta^2, \dots, \lambda_p$ olur.

Eğer noktaların ikisinde aynı özvektör yönünde ise bunlardan birini çıkarmak çoklu bağlantı ölçüsünü etkilemeyebilir. Her iki noktayı çıkarmak ise sonuç üzerinde oldukça etkili olabilir. Aşağıdaki şekilde bu durum açıkça görülmektedir.



Şekil 3.2. Etkili gözlemlerin çoklubağlantıya etkisi

3.3. Çoklubağlantıya Neden Olan Etkili Gözlemin Belirlenmesi

3.3.1. Etkili gözlemin belirlenmesinde kullanılan istatistikler

Etkili gözlem yada gözlemlerin belirlenmesinde kullanılan istatistiklerden bazıları şunlardır.

- | | |
|---|------------------------------|
| 1. İzdüşüm Değerleri, | h_i |
| 2. Artıklar, | $e_i = y_i - \hat{y}_i$ |
| 3. İkili Grafikler, | (x_i, x_j) ve (y_i, y_j) |
| 4. Student Tipi Artıklar | t_i ve t_i^* |
| 5. Cook 'un Uzaklık Ölçüsü, | D_i |
| 6. DFFITS _i ve DFBETAS _i | |
| 7. COVRATIO | |
| 8. WSSD _i | |
| 9. Ortalamayı-Saptıran Aykırı Gözlem(Mean-Shift Outlier) Modeli | |

1. İzdüşüm Değerleri (h_i);

Hoagling ve Welsch (1978) etkili gözlemi tespit etmekte "İzdüşüm" matrisini göz önüne almışlardır. X , $n \times p$ boyutlu bağımsız değişkenler matrisini göstermek üzere ; İzdüşüm matrisi

$$H = X(X'X)^{-1}X'$$

şeklinde , kestirim değerleri vektörünü de

$$\hat{Y} = X\hat{\beta} = HY$$

şeklinde ifade etmişlerdir. H matrisinin i . satır j . sütun elemanı olan h_{ij} 'ler bağımsız değişkenlere ilişkin j . gözlemin i . kestirim değerine etkisini ifade eder. Yani $\hat{Y}_i$ 'ne j . gözlemin (x_j , $j = 1, 2, \dots, n$) etkisini ifade eder. Özellikle h_{ii} köşegen elemanları, bizzat Y_i 'nin kestirimine aksetmektedir (Gerçekten $h_{ii} = 1$ olması durumunda, $i \neq j$ için h_{ij} lerin hepsi sıfır olur).Gözlemin etkili olup olmamasının yorumunda H 'ın bazı özellikleri önemlidir. Simetrik ve idempotent olan H matrisinin özelliklerinde,

i) Simetrik ve idempotent olmasından dolayı,

$$\text{rank}(H) = \text{tr}(H) = p \text{ dir. Dolayısı ile } \sum_{i=1}^n h_{ii} = p \text{ dir.}$$

ii) H matrisinin özdeğerleri 0 yada 1 dir ve sıfır olmayan özdeğerlerin sayısı matrisin rankına eşittir.

$$\text{iii) } h_{ii} = \sum_{j=1}^n h_{ij}^2 = h_{ii}^2 + \sum_{i \neq j} h_{ij}^2 \text{ dir.}$$

iv) $0 \leq h_{ii} \leq 1$ arasında deęer alır.

$h_{ii} = 0$ ise, h_{ij} ler sıfır olur, $i \neq j$ için

$h_{ii} = 1$ ise, h_{ij} ler sıfır olur, $i \neq j$ için

İzdüşüm(Hat) matrisinin köşegen elemanlarının ortalaması p/n dir. Tecrübelerle göre h_i bakımından bir gözlemin etkili sayılması o gözleme ilişkin h_i (izdüşüm değeri) nin $2p/n$ den büyük olmasıdır ($h_i > 2p/n$).

2. Artıklar ($e_i = y_i - \hat{y}_i$);

Doğrusal regresyon modeli $Y = X\beta + \varepsilon$ şeklinde tanımlansın, $E(\varepsilon) = 0$, $\text{Var}(\varepsilon) = \sigma^2 I$ olmak üzere, ε , $n \times 1$ boyutlu bilinmeyen rasgele hatalar vektörüdür.

En küçük kareler yöntemi için doğrusal regresyon modelinde bilinen artıklar vektörü e .

$$\begin{aligned} e &= Y - \hat{Y} \\ &= (I - H)Y \end{aligned}$$

dir. Burada $H = X(X'X)^{-1}X'$ izdüşüm matrisidir, $\hat{Y}$ 'de kestirim değeri vektörüdür.

e ve ε arasındaki ilişki,

$$\begin{aligned} e &= (I - H)Y, (Y = X\beta + \varepsilon) \\ &= (I - H)(X\beta + \varepsilon) \\ &= (I - H)\varepsilon \end{aligned}$$

gösterilebilir. Bunu skaler olarak ifade edecek olursak, $i = 1, 2, \dots, n$ için,

$$e_i = \varepsilon_i - \sum_{j=1}^n h_{ij} \varepsilon_j \quad (3.2)$$

dir. e ve ε arasındaki ilişkinin sadece H 'a bağlı olduğu Eş. 3.2 den kolayca görülmektedir. Eğer h_{ij} ler yeterince küçük iseler ε nun yerine geçmek için e uygun olacaktır. Etkili gözlemin bilirlenmesinde büyük e_i değeri bize bir fikir verebilir (Cook, Weisberg, 1982).

3. İkili Grafikler $((x_i, x_j)$ ve $(y_i, x_i))$:

Uç noktalar x değişken çiftlerinin hepsi için serpm diyagramlarına bakılarak ile belirlenebilir. Bununla beraber, çok değişken olması durumunda uç değer olabilecek gözlem yada gözlemler serpm diyagramlarından açıkça görülmeyebilirler.

4. Student Tipi Artıklar (t_i ve t_i^*):

Etkili gözlemin belirlenmesinde kullanılan bir başka istatistikte "student tipi artıklar" dır. t_i ,student tipi artıkları ifade etmek üzere,

$$t_i = r_i = \frac{e_i}{s\sqrt{1-h_i}}$$

şeklinde gösterilmektedir. Burada s^2 hata kare ortalamasıdır (Diğer bir gösterim ile $\hat{\sigma}^2$ dir). Belsley (1980) ve Velleman ve Welsch (1981), s yerine $s_{(i)}$ yi koymak vasıtasıyla değişiklik yapmayı tercih etmişlerdir.

$$s_{(i)}^2 = \frac{n-p-1}{n-p-2} s^2 - \frac{e_i^2}{(1-h_{ii})(n-p-2)}$$

Burada $s_{(i)}^2$, i. birim çıkarıldığında regresyon için hata kare ortalamasıdır. Buna göre Belsley, Velleman ve Welsch tarafından etkili gözlemin belirlenmesi amacı ile oluşturulan diğer bir "student" tipi artık t_i^* dır. Bu ise,

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

şeklinde ifade edilmiştir.

t_i , ortalaması sıfır varyansı bir olan normal dağılıma ($t_i \sim N(0,1)$), t_i^* da $(n-p-1)$ serbestlik dereceli t dağılımına sahiptir.

5. Cook 'un Uzaklık Ölçüsü (D_i);

D_i istatistiği Cook (1977,1979) tarafından ileri sürülmüştür. Bu uzaklık ölçüsü, çıkarılan i. gözlemin etkisini yansıtmaktadır. $\hat{\beta}$, β nın tahminini, $\hat{Y}_{(i)}$, i. birim çıkarılmış iken $(n-1) \times 1$ boyutlu kestirim değerleri vektörünü ifade etmek üzere, i. gözlemin etkinliğinin ölçüsü,

$$\Delta_i = (\hat{Y} - \hat{Y}_{(i)})' (\hat{Y} - \hat{Y}_{(i)}) = (\hat{\beta} - \hat{\beta}_{(i)})' X'X (\hat{\beta} - \hat{\beta}_{(i)})$$

dir. Böylece Δ_i , $\hat{Y}$ ile $\hat{Y}_{(i)}$ arasındaki uzaklığın karesi olmuş olur.

Hipotezlerin ($\beta = \beta_0$) testi için kareler toplamı değeri ile Δ_i arasındaki benzerliği Cook aşağıdaki gibi ifade etmiştir.

$$D_i = \Delta_i / (p+1) s^2$$

Cook tarafından ileri sürülen D_i istatistiği, $\{(p+1), (n-p-1)\}$ serbestlik dereceli F dağılımına yakınsamaktadır.

n tane D_i 'nin kolayca hesaplanması bakımından aşğıdaki formül kullanılmaktadır.

$$D_i = \frac{e_i^2 h_{ii}}{(1 - h_{ii})^2 (p + 1) s^2} = \frac{t_i^2 h_{ii}}{(p + 1)(1 - h_{ii})}$$

6. DFFITS_i ve DFBETAS_i İstatistikleri;

Belsley (1980) etkili gözlemin belirlenmesinde kullanılmak üzere iki istatistik öne sürmüştür. Birincisi, etkili olabileceğı düşünölen gözlem için

$$\hat{Y}_{(i)} - \hat{Y}_{i(i)} = \frac{h_{ii} e_i}{1 - h_{ii}}$$

ile verilen $\hat{Y} - \hat{Y}_{(i)}$ yi $\hat{Y}_i$ nin standart sapmasının tahminine bölerek standartlaştırmışlar ve

$$\begin{aligned} \text{DFFITS}_i &= \frac{\sqrt{h_{ii}} e_i}{s_{(i)} (1 - h_{ii})} \\ &= \frac{\sqrt{h_{ii}}}{\sqrt{(1 - h_{ii})}} t_i^* \end{aligned}$$

şeklinde tanımlamışlardır. Hesaplamalarda $\hat{Y} - \hat{Y}_{(i)}$ nin i. bileşenini göz önünde bulundurmamışlardır.

Dikkat edilirse, DFFITS_i Cook uzaklığı D_i nin kare kökünün p katıdır. Buna göre, $| \text{DFFITS}_i | \geq 2(p/n)^{1/2}$ olması etkili gözlemi belirlemede kullanılmaktadır.

İkinci olarak, Belsley (1980) doğrusal regresyon modelindeki parametrelerin tahminlerini kullanarak

$$DFBETA_i = \hat{\beta} - \hat{\beta}_{(i)}$$

ölçüsünü j. bileşeni için $\hat{\beta}_j$ nin standart sapmasının tahmini ile standartlaştırarak

$$DFBETAS_{ij} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{s_{(i)} \sqrt{s^{jj}}}$$

istatistiğini öne sürmüştür. Burada $s^{jj} (X'X)^{-1}$ matrisinin j. köşegen elemanıdır. Etkili gözlemin belirlenmesinde DFBETAS_{ij} istatistiği için sınır değeri $2/\sqrt{n}$ dir. ($|DFBETAS_{ij}| > 2/\sqrt{n}$).

7. COVRATIO İstatistiği;

Belsley (1980) in öne sürdüğü bu istatistik

$$COVRATIO_i = \left(\frac{s_{(i)}^2}{s^2} \right) \frac{1}{1 - h_{ii}}$$

dir. Bu, i. durum çıkarılmış iken $\hat{\beta}$ nin varyansının tahminini belirleyen orandır. Etkili gözlemin belirlenmesinde bu istatistik için aşağıdaki koşul verilmektedir.

$$|COVRATIO_i - 1| > 3p/n$$

8. WSSD_i İstatistiği;

Daniel ve Wood (1971) i. gözlem için standartlaştırılmış uzaklığın karesini ağırlıklandırmayı ileri sürmüşlerdir.

$$WSSD_i = \sum_{j=1}^{p-1} \left(\frac{\hat{\beta}_j (x_{ji} - \bar{x}_j)}{s_y} \right)^2$$

Böylece, WSSD_i bağımsız değişkenlere ilişkin verilerin ağırlıklandırılması ile i. gözlemin merkeze olan uzaklığının ölçüsü olmaktadır. Bu istatistik, bazı örnekler için etkili gözlemlere karşı duyarsız bulunmaktadır.

Etkili gözlemin (lerin) belirlenmesinde kullanılan bu istatistiklerden, İzdüşüm Matrisi, COVRATIO, WSSD, t_i , t_i^* ve e_i etkili olabilecek gözlemleri, Cook uzaklığı D_i , DFFITS_i ve DFBETAS_{ij} istatistikleri ise doğrudan etkili gözlemleri belirlerler.

9. Ortalamayı-Saptıran Aykırı Gözlem Modeli

(Mean-Shift Outlier Model);

i. gözlemin etkili olup olmadığının test edilmesinde kullanılan bir başka yöntemde Ortalamayı-saptıran aykırı gözlem modelidir. Cook ve Weisberg (1982) tarafından önerilen bu model,

$$Y = X\beta + d_i\phi + \varepsilon \quad (3.3)$$

$$E(\varepsilon) = 0, \text{ Var}(\varepsilon) = \sigma^2 I$$

şeklinde oluşturulmuştur.

Burada, d_i i. elemanı 1, diğer elemanları sıfır olan $n \times 1$ boyutlu bir vektördür. $H_0 : \phi = 0$ hipotezinin testi sonucunda, sıfırdan farklı ϕ değeri i. durumun

etkili olduğunu ima etmektedir. Normallik varsayımı altında, $\phi = 0$ hipotezinin testi için kullanılan t istatistiği

$$t_i = \frac{e_i (1 - h_{ii})^{1/2}}{\left\{ \left[Y'(I - H)Y - e_i^2 / (1 - h_{ii}) \right] (n - p' - 1) \right\}} \quad (3.4)$$

dir. Yokluk hipotezi altında bu $(n - p' - 1)$ serbestlik dereceli t dağılımıdır. Ortalamayı-Saptıran Aykırı Gözlem modeli için uygun görülen bu test istatistiği Dışsal Student Artık ile özdeştir (aynıdır). Bu modelde, $Z = (x; d_i)$ şeklinde oluşturulmuş olan bağımsız değişkenler matrisine d_i 'yi eklemekteki amaç i. gözlemin $|Z'Z|$ ile $|X'X|$ arasındaki farka ne denli etkili olduğunu görmek istenmesinden kaynaklanmaktadır. Ortalamayı-Saptıran Aykırı Gözlem modelindeki mantıktan hareketle Styán ve Simo(1993) , Eş. 3.4 de gösterilen t_i istatistiğinin karesini

$$t_i^2 = \frac{Y' P_{Md_i} Y}{Y'(M - P_{Md_i})Y / (n - p' - 1)}$$

şeklinde ifade etmişlerdir.

Burada $M = I - H$, P_{Md_i} de Md_i vektörüne ilişkin izdüşüm (hat) matrisidir ($P_{Md_i} = (Md_i) [(Md_i)'(Md_i)]^{-1} (Md_i)'$). Ayrıca $Y' P_{Md_i} Y$, Eş. 3.3 de verilen modeldeki değişkenler arasında dikliğin sağlanması için yapılan dönüşümden sonra elde edilen

$$Y = X\beta^* + (I - H)d_i\phi + \varepsilon$$

modelindeki $(I - H)d_i$ değişkenine ilişkin regresyon kareler toplamı ve $Y'(M - P_{Md_i})Y$ de bu modele ilişkin hata kareler toplamıdır.

$$\underbrace{Y'(M - P_{Md_i})Y}_{\text{KT(i), i. gözlem olmadan ki yada Ortalamayı Saptıran Aykırı Gözlem modeline ilişkin hata kare toplamı}} = \underbrace{Y'MY}_{\text{X bağımsız değişkenleri ne ilişkin hata kare toplamı}} - \underbrace{Y'P_{Md_i}Y}_{\text{(I-H)d_i bağımsız değişkenine ilişkin regresyon kare toplamı}}$$

Styan ve Simo (1993) etkili gözlemin belirlenmesinde kullanılan $DFBETA_i$ istatistiğini de

$$DFBETA_i = \hat{\beta} - \hat{\beta}_{(i)} = (X'X)^{-1}X'd_i \frac{e_i}{m_{ii}}$$

şeklinde ifade etmiştir. Burada m_{ii} , $(I - H) = M$ matrisinin i. köşegen elemanıdır.

3.3.2. Etkili gözlemin, özvektörlerle yaptığı açılara göre bu gözlemin çoklu bağlantıyı şiddetlendiren veya maskeleyen gözlem olup olmadığının incelenmesi

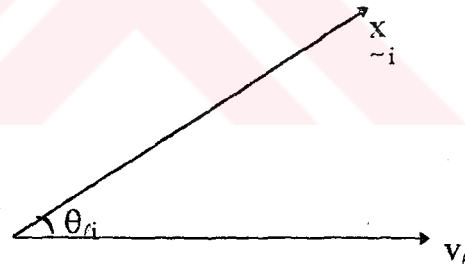
Bundan önceki bölümde, Velleman ve Ypeelar 'ın XX' matrisinin özvektörler matrisi U olmak üzere (u_{ij} , U nun i. satır ve j. sütun elemanı) , u_{ij} 'leri dik izdüşüm bileşenleri olarak adlandırdıklarını ve büyük değerli u_{ij} 'lerden herhangi birinin (u_{ik}) çıkarılan x_i gözleminin k. koşullu indisi etkileyebileceğini söylediklerini belirtmiştik. Yani u_{ik} nın büyük çıkması, i. gözlemin k. özvektör doğrultusunda bulunan izdüşüm değeri yüksek bir gözlem olduğu anlamına gelir. Dolayısı ile bu gözlemin çıkarılması k. özdeğerin değerinin küçülmesine ve doğal olarak k. koşullu indisin değerinin küçülmesine sebep olacaktır. Bu yöntemle izdüşüm değeri yüksek olan bir gözlemin çoklu bağlantıyı şiddetlendiren mi yoksa maskeleyen mi olduğuna karar verebiliriz (Muhakkakki izdüşüm değeri yüksek olan gözlem etkili bir gözlem demek değildir; Bölüm 3.3).

Bu yöntemdeki düşünceye benzer bir başka yöntem de, izdüşüm değeri yüksek olan bir gözlemin özvektörlerle yaptığı açılarının cosinüslerinin değerlendirilmesi ile elde edilebilir. Bunun için, Cook ve Weisberg (1980) in herhangi bir $\mathbf{x}_{\sim i}$ gözleminin h_i 'sinin ne sebepten dolayı nispeten büyük yada küçük olduğunu gösterebilmek için tanımladıkları

$$h_{ii} = \frac{1}{n} + \mathbf{x}_{\sim i}' \mathbf{x}_{\sim i} \sum_{\ell=1}^p \frac{\cos^2 \theta_{\ell i}}{v_{\ell}}$$

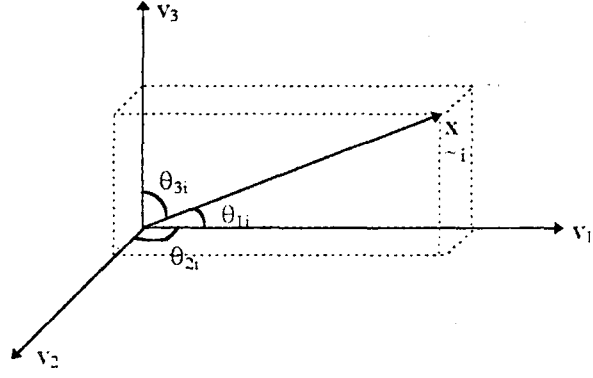
ifadesindeki $\cos \theta_{\ell i}$ dikkate alınırsa, $\lambda_1, \lambda_2, \dots, \lambda_p$ $\mathbf{X}'\mathbf{X}$ in özdeğerlerini ve $v_1, v_2, \dots, v_p$ de bunlara karşılık gelen özvektörleri ifade etmek üzere, i . gözlem $\mathbf{x}_{\sim i}$ ile ℓ . özvektör v_{ℓ} arasındaki $\theta_{\ell i}$ açısının cosünüsü

$$\cos \theta_{\ell i} = \frac{\mathbf{v}_{\ell}' \mathbf{x}_{\sim i}}{(\mathbf{x}_{\sim i}' \mathbf{x}_{\sim i})^{1/2}}$$



olur.

Buna göre Şekil 3.3 den de kolayca anlaşılacağı gibi i . gözlemin v_{ℓ} ile yaptığı açının cosünüsü 1 e ne kadar yakınsa gözlem v_{ℓ} yönünde demektir. Eğer izdüşüm değeri yüksek olan gözlemin, en büyük özdeğer (λ_1) e karşılık gelen özvektör, v_1 ile yaptığı açının cosünüsü 1 e yakınsa bu gözlemin çoklu bağlantıyı şiddetlendiren bir gözlem olduğu söylenebilir. Eğer izdüşüm değeri yüksek olan gözlem v_p (en küçük özdeğere karşılık gelen özvektör) yönünde ise bu gözlemi çoklu bağlantının şiddetini maskeleyen bir gözlem olarak tanımlayabiliriz.



Şekil 3.3. i.gözlemin v_i ile yaptığı açının cosinüsü

4. UYGULAMA

Regresyon analizinde çokça karşılaşılan sorunlardan biri çoklubağlantıdır. Çoklubağlantının varlığı değişkenler arasındaki sıkı ilişkiden kaynaklanmaktadır. Her durumda gerçekleşmede, bu sıkı ilişkiye etkili gözlem yada gözlemler neden olabilir. Yada etkili gözlem veya gözlemler çoklubağlantıyı gizliyor olabilir. Bu tür gözlemlerin belirlenmesine yönelik yöntemleri bir uygulamada kullanmak amacı ile Devlet İstatistik Enstitüsü (DİE) yayın şubesinden aşağıda açıklanan değişkenlere ilişkin değerler elde edilmiştir.

İller bazında 1996 yılına ilişkin kişi başına gayri safi yurtiçi hasıla (İİKBGSYH) (\$), bağımlı değişkeni, bir hektarlık tarımsal alan başına düşen nüfus (BHTABN), illere göre bir yaşından küçük bebek ölümlerinin sayısı (İGBKBÖS), iller itibarıyla doktor başına düşen hasta sayısı (İİDBDHS), iller itibarı ile kilometre kare başına düşen nüfus (İİKKBN) , iller itibarı ile imalat sanayiinin deki işyeri sayısı (İŞYERİ), iller itibarı ile öğretmen başına düşen öğrenci sayısı (ÖBDÖS), iller itibarı ile üniversite (lisans, ön lisans) öğrencisi başına düşen nüfus (ÜÖBDN) da bağımsız değişkenleri ifade etmek üzere, 70 il üzerinden bu değişkenlere ilişkin değerler Çizelge 4.1 de verilmiştir. Gerekli hesaplamalar için SPSS ve MAT LAB paket programları kullanılmıştır.

Çizelge 4.1. İller bazında 1996 yılı bağımlı ve bağımsız değişken değerleri

	İller	iikbgsh	igkbös	işyeri	öbdös	üöbdn	bhtabn	iikkbn	iidbdhs
1	Adana	2844	958	243	34.12	169.99	66.43	114.86	2783.04
2	Adıyaman	1222	113	13	29.39	2006.75	122.25	75.47	14849.95
3	Afyon	1453	154	123	26.59	403.91	126.93	54.20	7165.87
4	Ağrı	666	23	4	43.97	450.26	973.58	38.64	15855.46
5	Amasya	1951	58	27	19.57	189.10	58.96	60.49	4684.09
6	Ankara	3996	1772	682	27.78	30.17	106.31	133.02	797.17
7	Antalya	3079	293	64	22.81	266.61	48.55	61.92	2356.23
8	Artvin	2560	17	12	20.19	1332.21	359.25	23.37	5642.29
9	Aydın	2810	144	112	22.16	816.91	59.52	107.72	3100.87
10	Balıkesir	2744	158	123	23.34	208.34	34.76	67.47	3017.83
11	Bilecik	3868	22	47	23.26	323.56	20.54	41.46	8159.43
12	Bingöl	849	34	2	24.94	1077.33	1211.35	28.41	9376.78
13	Bitlis	765	29	6	31.35	2640.74	436.32	41.38	8229.28
14	Bozü	2916	24	111	18.14	104.99	186.15	49.96	4025.72
15	Burdur	2352	25	37	13.07	102.66	42.56	34.33	5449.81
16	Bursa	3660	786	558	28.90	100.89	36.04	164.01	2301.81
17	Çanakkale	3664	59	36	15.87	187.31	28.55	42.36	3246.75
18	Çankırı	1373	31	13	20.51	196.98	27.95	31.54	12529.70
19	Çorum	2201	110	99	27.12	525.31	167.88	45.72	6524.04
20	Denizli	2434	135	168	20.28	335.37	70.31	66.84	3578.77
21	Diyarbakır	1696	670	34	32.79	203.62	85.25	81.17	4789.28
22	Edirne	2689	58	54	15.12	47.73	44.11	65.52	2141.46
23	Elazığ	2092	125	34	25.14	77.03	200.17	50.43	2951.95
24	Erzincan	1603	43	14	16.10	168.02	85.42	23.88	6455.78
25	Erzurum	1312	261	25	31.05	54.52	1066.54	30.73	3048.38
26	Eskişehir	2885	407	160	23.17	1.87	133.75	48.95	1813.48
27	Gaziantep	2085	1269	194	47.28	371.59	114.42	159.03	4744.53
28	Giresun	1425	15	37	20.01	291.62	457.78	63.48	5846.37
29	Gümüşhane	1178	4	2	19.78	183.82	280.84	21.61	9776.69
30	Hatay	2268	198	71	31.56	328.90	48.25	205.46	4080.58
31	Isparta	1748	81	61	16.81	148.02	160.57	52.13	3498.73
32	İçel	3167	425	108	25.15	761.98	52.08	92.69	4033.18
33	İstanbul	4037	4809	4560	38.56	74.37	821.48	1523.56	1197.32
34	İzmir	4093	1264	1160	30.31	71.37	92.55	239.43	1212.55

35	Kars	1071	30	9	29.53	1214.00	1890.89	31.59	10064.42
36	Kastamonu	1964	50	33	17.74	172.14	154.40	27.88	5787.61
37	Kayseri	1986	273	152	27.32	104.05	313.09	58.91	3050.53
38	Kırklareli	3945	59	48	16.90	468.31	109.60	48.19	3287.92
39	Kırşehir	1687	26	17	22.05	109.24	190.92	36.89	4553.69
40	Kocaeli	7349	241	310	26.72	285.24	399.56	306.53	3466.49
41	Konya	1995	446	194	35.40	99.10	137.54	46.35	3768.42
42	Kütahya	2429	104	53	24.69	359.35	132.61	49.70	4583.72
43	Malatya	1879	216	35	23.40	115.26	266.42	58.92	4561.04
44	Manisa	2878	196	225	24.52	409.36	90.48	90.37	3145.69
45	K. Maraş	1790	164	57	32.93	946.27	203.45	59.77	8858.88
46	Mardin	1229	49	5	32.93	3182.54	37.40	69.88	18805.94
47	Muğla	3707	36	21	18.44	614.40	20.21	48.00	3382.48
48	Muş	660	15	3	38.36	3353.29	141.56	46.95	17787.00
49	Nevşehir	2371	19	38	28.94	569.32	101.50	52.20	8207.64
50	Niğde	1968	50	13	23.10	113.36	59.62	39.48	6631.63
51	Ordu	1128	84	49	30.85	1189.45	2987.74	134.30	7799.84
52	Rize	2085	14	76	21.88	269.38	8627.72	71.76	3982.03
53	Sakarya	2585	74	114	27.15	217.45	167.13	149.20	3798.18
54	Samsun	2026	498	129	26.80	143.55	30.85	115.37	2705.87
55	Siirt	1326	45	2	29.38	270.75	138.93	38.70	12093.43
56	Sinop	1548	22	26	22.17	1223.85	141.32	39.47	6516.19
57	Sivas	1494	181	20	25.52	124.49	348.41	25.17	3828.93
58	Tekirdağ	3363	96	139	149.80	186.75	122.95	82.28	2846.21
59	Tokat	1635	157	65	25.28	260.32	68.42	71.23	8195.66
60	Trabzon	2050	216	68	24.14	67.99	1648.10	159.20	2977.81
61	Tunceli	1413	6	2	18.12	810.54	84.34	12.71	6124.06
62	Ş. Urfa	1238	387	15	41.24	754.32	87.85	61.76	13312.43
63	Van	1015	320	8	35.94	273.93	628.47	33.18	12271.12
64	Yozgat	1245	47	20	31.77	457.33	124.87	42.94	13394.76
65	Zonguldak	2349	82	70	29.72	351.82	630.70	127.00	3759.82
66	Aksaray	1291	47	21	40.01	964.95	116.54	43.86	6754.63
67	Bayburt	1034	12	4	31.37	6395.25	805.70	26.12	6395.25
68	Karaman	2799	20	24	29.99	821.01	69.23	23.60	6519.82
69	Kırıkkale	3343	88	23	32.43	214.33	15.70	71.29	4412.95
70	Batman	1157	31	2	35.95	1237.56	147.19	81.70	11865.97

Bu değişkenlere ilişkin doğrusal regresyon modelinin parametrelerini tahmin etmeden önce, bağımsız değişkenler arasında anlamlı bir çoklubağlantının olup olmadığına karar vermek gerekir. Çizelge 4.1 deki bağımsız değişkenler arasındaki ikili korelasyonların örneğin, İŞYERİ değişkeni ile ÜÖBDN değişkeni arasındaki korelasyonun 0.9009, İŞYERİ değişkeni ile BHTABN değişkeni arasındaki korelasyonun 0.9396, BHTABN değişkeni ile ÜÖBDN değişkeni arasındaki korelasyonun 0.9671 gibi büyük değerler almış olması bağımsız değişkenler arasındaki çoklubağlantının varlığının en basit belirtisidir. Ayrıca çoklubağlantının belirlenmesinde en çok kullanılan Denklem (2.8.2) deki VIF (varyans artırıcı faktör) ve Denklem (2.8.4) deki koşullu indeks değerleri aşağıdaki çizelgede yer almaktadır.

Çizelge 4.2. Çizelge 4.1 deki verilere ilişkin çoklubağlantı ölçüleri

Koşullu İndeks	Değişkenler	VIF
1	BHTABN	1.036
5.21	İGBKBÖS	14.037
17.37	İİDBDHS	1.865
95.21	İİKKBN	16.865
101.52	İŞYERİ	31.598
128.08	ÖBDÖS	2.190
833.64	ÜÖBDN	1.325

Koşullu indisin 100 den büyük olması bağımsız değişkenler arasında şiddetli çoklubağlantının varlığını göstermektedir. Aynı şekilde İGBKBÖS, İİKKBN ve İŞYERİ değişkenleri için VIF değerlerinin 10 dan büyük olması çoklubağlantının varlığını göstermektedir.

Bu değişkenlere ilişkin doğrusal regresyon modelinin parametre tahminleri ve bu tahminlere ilişkin t ve p değerleri aşağıdaki çizelgede yer almaktadır.

Çizelge 4.3. Doğrusal regresyon modelinin parametre tahminleri

Değişkenler	β tahmini	t değerleri	p değerleri
BHTABN	-0.1570	-1.655	0.1030
İGBKBÖS	-0.1987	-0.327	0.7445
İİDBDHS	-0.1354	-3.881	0.0003
İİKKBN	3.8273	1.636	0.1070
İŞYERİ	-0.5363	-0.518	0.6066
ÖBDÖS	-13.6399	-0.641	0.5238
ÜÖBDN	-0.0868	-0.708	0.4816
Sabit	3309.8406	7.652	0.0000

Bu sonuçlara göre sadece sabit terim ve İİDBDHS değişkenlerinin parametre tahminleri anlamlı bulunmuştur. Ayrıca bu modele ilişkin belirleme katsayısıda 0.67691 olarak elde edilmiştir. Bu ise çoklubağlantıdan dolayı, modelin bağımlı değişkeni açıklamakta pekte yeterli olmadığını göstermektedir. Veri yapısındaki şiddetli çoklubağlantının ortadan kaldırılması için değişken eliminasyon yöntemleri veya yanlış tahminler veren yöntemler uygulanabilir. Ayrıca, çoklubağlantıya sebebiyet veren etkili gözlemler varsa bunların elimine edilmesiyle çoklubağlantı problemi önemini kaybetmektedir.

Çoklubağlantıyı maskeleyen veya şiddetlendiren gözlemlerin belirlenmesinde Velleman ve Ypeeler (1980) tarafından "dik izdüşüm bileşenleri" olarak adlandırılan u_{ij}^2 değerleri kullanılmaktadır (Denklem 3.1). u_{ij}^2 değeri p/n den büyük olması i. gözlemin j. koşullu indisi etkileyebileceğinin bir göstergesidir. Çizelge 4.1 deki değişkenlere ilişkin dik izdüşüm bileşenleri Çizelge 4.4 de yer almaktadır.

Çizelge 4.4. Çizelge 4.1 deki bağımsız değişkenlere ilişkin dik izdüşüm bileşenlerinin değerleri

	İller	Leverage Değerleri (h _i)	u_{i1}^2	u_{i2}^2	u_{i3}^2	u_{i4}^2	u_{i5}^2	u_{i6}^2	u_{i7}^2
1	Adana	.1114	.0101	.0016	.0001	.0025	.0947	.0023	.0001
2	Adıyaman	.1033	.0678	.0006	.0024	.0036	.0008	.0038	.0242
3	Afyon	.0126	.0009	.0011	.0016	.0002	.0016	.0056	.0015
4	Ağrı	.2000	.0610	.0026	.0361	.0002	.0001	.0020	.0779
5	Amasya	.0191	.0019	.0017	.0005	.0031	.0026	.0008	.0086
6	Ankara	.4108	.0258	.0014	.0003	.00263	.2516	.0576	.0479
7	Antalya	.0214	.0124	.0016	.0006	.0016	.0027	.0005	.0021
8	Artvin	.0292	.0001	.0000	.0133	.0007	.0015	.0041	.0094
9	Aydın	.0294	.0077	.0011	.0078	.0009	.0034	.0073	.0012
10	Balıkesir	.0179	.0085	.0018	.0000	.0021	.0030	.0008	.0017
11	Bilecik	.0193	.0035	.0022	.0040	.0011	.0046	.0025	.0015
12	Bingöl	.0262	.0094	.0077	.0000	.0001	.0000	.0034	.0054
13	Bitlis	.0724	.0048	.0004	.0617	.0006	.0001	.0003	.0046
14	Bolu	.0299	.0039	.0007	.0006	.0034	.0131	.0062	.0020
15	Burdur	.0681	.0004	.0019	.0022	.0032	.0047	.0029	.0528
16	Bursa	.0453	.0132	.0019	.0002	.0033	.0007	.0150	.0110
17	Çanakkale	.0356	.0072	.0018	.0000	.0041	.0031	.0006	.0187
18	Çankırı	.1104	.0348	.0026	.0242	.0001	.0011	.0030	.0446
19	Çorum	.0153	.0001	.0008	.0001	.0006	.0024	.0059	.0054
20	Denizli	.0225	.0057	.0014	.0002	.0014	.0080	.0057	.0002
21	Diyarbakır	.0985	.0017	.0015	.0011	.0001	.0773	.0152	.0014
22	Edirne	.0449	.0139	.0017	.0000	.0052	.0056	.0005	.0179
23	Elazığ	.0182	.0089	.0006	.0001	.0038	.0004	.0000	.0044
24	Erzincan	.0491	.0001	.0016	.0031	.0024	.0016	.0038	.0366
25	Erzurum	.0479	.0084	.0045	.0008	.0025	.0060	.0016	.0241
26	Eskişehir	.0305	.0165	.0011	.0001	.0008	.0038	.0082	.0000
27	Gaziantep	.3293	.0018	.0013	.0006	.0094	.2543	.0523	.0096
28	Giresun	.0170	.0001	.0000	.0011	.0024	.0050	.0003	.0081
29	Gümüşhane	.0579	.0112	.0005	.0126	.0011	.0018	.0048	.0258
30	Hatay	.1600	.0037	.0016	.0000	.0010	.0008	.1465	.0064
31	Isparta	.0289	.0061	.0009	.0001	.0034	.0033	.0005	.0147
32	İçel	.0320	.0037	.0013	.0036	.0000	.0106	.0048	.0080

33	İstanbul	.9300	.0250	.0014	.0212	.8021	.0676	.0091	.0037
34	İzmir	.3015	.0219	.0014	.0003	.0256	.0040	.1901	.0581
35	Kars	.0449	.0139	.0262	.0000	.0000	.0000	.0044	.0004
36	Kastamonu	.0300	.0001	.0010	.0020	.0025	.0023	.0045	.0176
37	Kayseri	.0278	.0084	.0002	.0002	.0011	.0000	.0052	.0127
38	Kırklareli	.0303	.0069	.0010	.0015	.0031	.0035	.0004	.0139
39	Kırşehir	.0120	.0022	.0007	.0010	.0040	.0033	.0008	.0000
40	Kocaeli	.2160	.0063	.0000	.0000	.0001	.0242	.1851	.0003
41	Konya	.0951	.0050	.0011	.0008	.0001	.0050	.0167	.0664
42	Kütahya	.0078	.0021	.0010	.0000	.0020	.0012	.0004	.0011
43	Malatya	.0106	.0022	.0004	.0013	.0017	.0012	.0007	.0032
44	Manisa	.0299	.0077	.0012	.0008	.0006	.0085	.0036	.0075
45	K. Maraş	.0174	.0066	.0005	.0000	.0000	.0001	.0000	.0101
46	Mardin	.2042	.1446	.0007	.0208	.0116	.0004	.0017	.0245
47	Muğla	.0260	.0063	.0016	.0036	.0031	.0032	.0000	.0081
48	Muş	.1686	.1232	.0002	.0341	.0102	.0001	.0003	.0005
49	Nevşehir	.0178	.0038	.0013	.0010	.0007	.0040	.0002	.0068
50	Niğde	.0117	.0002	.0019	.0043	.0023	.0015	.0004	.0012
51	Ordu	.1102	.0026	.0795	.0005	.0000	.0000	.0270	.0005
52	Rize	.8133	.0041	.7832	.0113	.0042	.0005	.0076	.0024
53	Sakarya	.0627	.0048	.0008	.0000	.0023	.0119	.0342	.0087
54	Samsun	.0416	.0104	.0019	.0000	.0001	.0137	.0140	.0015
55	Siirt	.0522	.0303	.0014	.0194	.0001	.0004	.0006	.0000
56	Sinop	.0170	.0002	.0006	.0077	.0005	.0024	.0011	.0045
57	Sivas	.0129	.0047	.0001	.0005	.0026	.0009	.0025	.0016
58	Tekirdağ	.0237	.0094	.0011	.0000	.0027	.0101	.0001	.0003
59	Tokat	.0150	.0036	.0018	.0057	.0003	.0001	.0005	.0031
60	Trabzon	.0951	.0088	.0171	.0011	.0022	.0001	.0618	.0040
61	Tunceli	.0314	.0000	.0011	.0015	.0016	.0021	.0067	.0184
62	Ş. Urfa	.0976	.0444	.0016	.0100	.0022	.0270	.0024	.0099
63	Van	.0778	.0320	.0002	.0235	.0004	.0190	.0008	.0019
64	Yozgat	.0679	.0452	.0015	.0186	.0000	.0006	.0010	.0009
65	Zonguldak	.0577	.0049	.0005	.0001	.0024	.0045	.0264	.0090
66	Aksaray	.1358	.0004	.0008	.0023	.0006	.0013	.0002	.1301
67	Bayburt	.6497	.0008	.0095	.6243	.0113	.0005	.0031	.0003
68	Karaman	.0336	.0002	.0012	.0011	.0011	.0026	.0052	.0221
69	Kırıkkale	.0568	.0026	.0020	.0002	.0029	.0012	.0042	.0436
70	Batman	.0519	.0291	.0007	.0000	.0001	.0007	.0079	.0133

Çizelge 4.4 deki izdüşüm değerlerine göre $h_i = 2p/n = 14/70 = 0.20$ den büyük olan iller Ankara, Gaziantep, İstanbul, İzmir, Kocaeli, Mardin, Rize, Bayburt'tur. Bu illere ilişkin dik izdüşüm bileşenlerinden (öz değerler $\lambda_1 > \lambda_2 > \lambda_3 > \lambda_4 > \lambda_5 > \lambda_6 > \lambda_7$ olmak üzere) Ankara'nın λ_5 'i, Gaziantep'in λ_5 'i, İstanbul'un λ_4 'ü, İzmir'in λ_6 'y1, Kocaeli'nin λ_6 'y1, Rize'nin λ_2 'yi, Bayburt'un λ_3 'ü ve Mardin'in λ_1 'i büyük ölçüde etkileyeceğini $u_{ij}^2 = \frac{p}{n} = \frac{7}{70} = 0.10$ kritik değerine göre yapılan değerlendirmeden söylenebilir.

Ayrıca Muş ilinin izdüşüm değeri 0.20 den büyük olmasada , $u_{ij}^2 = 0.1232$ değerinin 0.10 kritik değerinden büyük olması λ_1 'i etkileyebileceğini göstermektedir. Veriden çıkması durumunda büyük özdeğerin küçülmesine neden olacak gözlemler çoklubağlantıyı şiddetlendiren, küçük özdeğerin küçülmesine neden olacak gözlemler ise çoklubağlantıyı maskeleyen gözlemler olacaktır. Dik izdüşüm bileşenlerine göre, İzmir ve Kocaeli illeri çoklubağlantıyı maskeleyen, Mardin ve Muş illeri de çoklubağlantıyı şiddetlendiren gözlemler olarak görünmektedir.

Çoklubağlantıyı etkileyen gözlemlerin belirlenmesine yönelik benzer bir yöntem (3.3.1) ve (3.3.2) bölümlerinde anlatılmıştır. Bu yöntem iki aşamadan oluşmaktadır. Birinci aşama etkili yada aykırı gözlemlerin belirlenmesinde kullanılan çeşitli istatistikler yardımı ile diğer gözlemlerden farklı (uzak) olan gözlemlerin belirlenmesidir. İkinci aşamada, farklı olan bu gözlemlerin, $(X'X)$ matrisinin özvektörleri ile yaptığı açılarının cosinüs değerlerine göre, çoklu bağlantıyı şiddetlendiren veya maskeleyen gözlem olup olmadığına karar verilmesidir.

Bu bağlamda, Çizelge 4.1 deki değişkenlere ilişkin olarak, etkili gözlemleri belirlemekte kullanılan başlıca istatistiklerin sonuçları Çizelge 4.5 te ve herbir gözlemin özvektörlerle yapmış oldukları açıların cosinüs değerleri de Çizelge 4.6 da yer almaktadır.

Çizelge 4.5. Etkili gözlemleri belirlemekte kullanılan başlıca istatistiklerin sonuçları

	r_i	t_i	D_i	$Covratio_i$	$DFFITS_i$	$SDFFITS_i$
1	.32919	.35205	.00223	1.28190	-40.64	.13
2	.30138	.32082	.00171	1.27359	34.49	.12
3	-.6742	-.68347	.00161	1.10163	-15.98	-.11
4	.18010	.20318	.00141	1.44178	42.19	.11
5	-.7417	-.75435	.00245	1.09444	-21.99	-.14
6	1.6310	.15108	.42771	1.06519	1036.15	1.91
7	.33271	.33882	.00053	1.16372	10.59	.06
8	.44522	.45522	.00118	1.15922	17.38	.10
9	-.0242	-.02470	.00000	1.19082	-.95	-.01
10	.02803	.02850	.00000	1.17668	.80	.01
11	2.1918	2.22961	.02162	.60391	65.51	.43
12	-.7768	-.79305	.00332	1.09400	-28.16	-.16
13	.9938	-1.0399	.01284	1.08332	-81.08	-.32
14	.36127	.36954	.00079	1.17080	14.37	.08
15	-.1539	-.16065	.00029	1.23712	-11.88	-.05
16	1.0458	1.07839	.00921	1.04087	56.95	.27
17	1.0478	1.07491	.00758	1.03125	47.27	.25
18	-.0532	-.05689	.00006	1.30061	-6.51	-.02
19	.13606	.13812	.00007	1.17078	3.57	.02
20	-.2480	-.25269	.00030	1.17267	-8.13	-.05
21	-.7519	-.79832	.01013	1.18189	-82.14	-.28
22	-.3768	-.38845	.00119	1.18721	-20.37	-.10
23	-.6831	-.69450	.00203	1.10588	-19.71	-.13
24	-.7682	-.79374	.00533	1.12064	-44.69	-.21
25	-1.212	-1.2517	.01299	.98956	-69.07	-.32
26	.15986	.16357	.00016	1.18825	6.44	.04
27	-.1623	-.20034	.00263	1.72600	-72.98	-.14
28	-1.097	-1.1148	.00502	.99986	-30.45	-.20

29	-.6499	-.67474	.00443	1.15721	-43.42	-.19
30	-.8512	-.93675	.02316	1.23071	-154.38	-.43
31	-1.131	-1.1557	.00753	.99997		-.25
32	.70834	.72533	.00319	1.11550	29.54	.16
33	-1.023	-4.3335	39.7757	1.13982	-14892.0	-21.1
34	1.5581	1.88365	.20473	1.03890	617.88	1.31
35	-.2097	-.21623	.00037	1.20328	-11.33	-.05
36	-.4187	-.42830	.00106	1.16384	-16.69	-.09
37	-.6628	-.67725	.00252	1.12044	-25.05	-.14
38	1.4225	1.45531	.01237	.90274	57.09	.32
39	-.9226	-.93501	.00296	1.04411	-21.44	-.15
40	4.6580	5.30915	1.05390	.01158	1197.00	3.90
41	-.3212	-.34035	.00178	1.25988	-33.91	-.12
42	-.0141	-.01427	.00000	1.16464	-.27	.00
43	-.7050	-.71394	.00163	1.09334	-15.46	-.11
44	.22380	.22891	.00030	1.18350	8.88	.05
45	.09084	.09231	.00003	1.17484	2.55	.02
46	1.0980	1.24206	.05393	1.19135	263.79	.66
47	1.1618	1.18592	.00737	.98769	41.87	.24
48	.49047	.54258	.00824	1.34174	94.30	.26
49	.53265	.54139	.00121	1.13285	15.15	.10
50	-.2832	-.28692	.00027	1.15697	-6.49	-.05
51	-.7016	-.74983	.00999	1.20944	-85.71	-.28
52	.88541	2.13252	2.72911	3.59186	3651.94	4.81
53	-.3376	-.35137	.00129	1.21436	-24.18	-.10
54	-.9403	-.96774	.00693	1.06806	-47.83	-.24
55	-.0436	-.04518	.00002	1.21977	-2.67	-.01
56	-.6761	-.68691	.00190	1.10598	-18.74	-.12
57	-1.086	-1.1011	.00423	.99962	-26.05	-.18
58	.60892	.62082	.00190	1.12626	20.65	.12
59	-.4574	-.46426	.00081	1.14109	-11.87	-.08
60	-.9222	-.97715	.01466	1.12946	-97.28	-.34
61	-.9111	-.93263	.00521	1.06591	-37.48	-.20
62	.25853	.27434	.00119	1.27005	27.99	.10
63	-.0916	-.09618	.00012	1.25289	-7.98	-.03
64	.11427	.11928	.00016	1.23868	8.79	.04
65	-.4059	-.42140	.00172	1.19949	-27.07	-.12
66	-.7020	-.76150	.01280	1.24300	-106.51	-.32

67	-.4597	-.79314	.15542	3.12440	-780.64	-1.11
68	.91983	.94267	.00558	1.06571	39.72	.21
69	.99097	1.02818	.01011	1.06848	65.15	.28
70	-.2676	-.27689	.00068	1.20759	-16.28	-.07

r_i : standartlaştırılmış artıklar

t_i : student tipi artıklar

D_i : Cook uzaklığı

$SDF\hat{F}ITS_i$: standartlaştırılmış $DF\hat{F}ITS_i$ istatistiği

Çizelge 4.6. $\cos\theta_{1i}$, i. gözlemin $(X'X)$ matrisinin ℓ . özvektörüyle yaptığı açıların cosinüs değerleri

No	$\cos\theta_{1i}$	$\cos\theta_{2i}$	$\cos\theta_{3i}$	$\cos\theta_{4i}$	$\cos\theta_{5i}$	$\cos\theta_{6i}$	$\cos\theta_{7i}$
1	.9468	-.0236	.2765	-.0786	.1426	.0068	.0003
2	.9987	.0391	-.0175	-.0279	-.0011	.0000	-.0011
3	.9984	.0296	.0325	.0355	-.0050	-.0026	.0009
4	.9969	-.0075	.0141	.0764	-.0050	-.0026	.0001
5	.9974	.0375	.0283	.0546	-.0044	.0060	.0006
6	.4043	-.1744	.7780	-.3130	.3176	-.0448	-.0023
7	.9936	.0099	.0863	-.0490	.0520	.0120	.0025
8	.9916	-.0182	-.0677	-.1088	-.0027	-.0015	.0006
9	.9867	.0129	-.0199	-.1601	-.0050	.0169	.0019
10	.9975	.0280	.0648	.0069	-.0044	.0044	.0038
11	.9969	.0489	.0220	.0576	-.0092	-.0014	.0003
12	.9966	-.0768	-.0290	.0053	-.0017	-.0026	-.0003
13	.9786	-.0111	-.0984	-.1804	-.0005	-.0001	.0006
14	.9967	.0033	.0373	.0675	-.0239	-.0013	.0022
15	.9956	.0442	.0320	.0761	-.0092	-.0006	-.0003
16	.9256	-.0322	.3616	-.1024	.0298	-.0110	.0054
17	.9982	.0391	.0286	.0354	-.0042	.0046	.0014
18	.9950	.0509	.0287	.0808	-.0058	-.0025	-.0009
19	.9995	.0215	.0171	.0165	-.0060	-.0025	.0014
20	.9985	.0202	.0464	-.0114	-.0172	-.0011	.0028
21	.9902	.0164	.1134	.0152	.0776	.0091	-.0008
22	.9959	.0255	.0588	.0609	-.0126	.0170	.0025
23	.9964	-.0206	.0508	.0629	.0109	.0081	.0039
24	.9963	.0390	.0266	.0711	-.0038	-.0017	-.0003
25	.9531	-.2841	.0391	.0859	.0437	.0003	.0040

26	.9707	-.0549	.2218	.0119	.0724	-.0063	.0055
27	.9690	-.0107	.1996	-.0614	.1306	.0137	-.0013
28	.9980	-.0261	.0086	.0562	-.0100	.0034	.0003
29	.9961	.0251	.0218	.0818	-.0061	-.0027	-.0005
30	.9976	.0291	.0494	.0014	.0033	.0379	.0004
31	.9980	.0020	.0371	.0502	-.0056	.0041	.0013
32	.9919	.0166	.0426	-.1105	.0412	.0088	-.0001
33	.1956	-.2754	.8624	-.3725	-.0580	.0021	-.0018
34	.5890	-.1530	.7320	-.2983	-.0221	-.0651	.0127
35	.9904	-.1326	-.0390	.0067	-.0019	-.0031	-.0001
36	.9970	.0249	.0267	.0676	-.0050	-.0018	.0003
37	.9926	-.0646	.0957	.0355	.0147	-.0019	.0045
38	.9991	.0111	-.0076	-.0389	-.0057	.0049	.0015
39	.9968	.0106	.0245	.0756	-.0062	.0020	.0018
40	.9898	-.0833	.0946	-.0163	-.0409	.0502	-.0001
41	.9909	-.0034	.1271	.0269	.0334	-.0086	.0049
42	.9995	.0180	.0202	.0177	-.0005	.0023	.0019
43	.9964	-.0113	.0529	.0626	.0176	.0050	.0005
44	.9962	.0036	.0619	-.0570	-.0199	.0009	.0045
45	.9997	.0239	.0028	-.0056	.0016	-.0001	.0004
46	.9968	.0443	-.0344	-.0571	-.0032	-.0010	-.0011
47	.9962	.0377	-.0294	-.0723	-.0041	.0072	.0017
48	.9956	.0378	-.0444	-.0727	-.0037	-.0020	-.0006
49	.9987	.0381	.0077	.0322	-.0084	.0002	.0007
50	.9955	.0433	.0317	.0783	-.0037	.0006	.0005
51	.9498	-.3060	-.0639	-.0044	-.0013	.0066	-.0005
52	.4621	-.8717	-.1164	.1141	.0009	-.0041	-.0003
53	.9981	.0011	.0402	.0322	-.0223	.0235	.0022
54	.9832	.0111	.1653	-.0202	.0716	.0208	.0007
55	.9960	.0415	.0251	.0756	-.0043	-.0016	-.0003
56	.9962	.0237	-.0415	-.0726	-.0053	.0000	.0004
57	.9961	-.0428	.0427	.0608	.0212	-.0003	.0024
58	.9980	-.0018	.0550	.0163	-.0237	.0079	.0034
59	.9967	.0411	.0375	.0597	-.0007	.0013	-.0002
60	.8925	-.4397	.0164	.0926	.0170	.0306	-.0006
61	.9989	.0348	-.0224	-.0224	-.0043	-.0026	.0003
62	.9980	.0415	.0295	.0362	.0120	-.0005	-.0005
63	.9966	-.0003	.0349	.0734	.0107	-.0025	-.0005

64	.9968	.0430	.0211	.0644	-.0052	-.0019	-.0003
65	.9926	-.1184	.0023	.0159	-.0094	.0206	.0025
66	.9987	.0297	-.0208	-.0342	-.0030	.0010	.0028
67	.7766	-.0781	-.2787	-.5596	.0064	-.0009	.0007
68	.9990	.0374	-.0158	-.0187	-.0057	-.0020	.0019
69	.9975	.0450	.0316	.0438	.0004	.0096	.0032
70	.9993	.0370	-.0089	.0017	-.0046	.0020	-.0001

Çizelge 4.5 te sonuçları verilen test istatistikleri için geçerli sınır değerleri Bölüm 3.3 de verildiği gibi r_i için 1.64, t_i için $t_{(n-p-1),0.05} = t_{62; 0.05} = 1.671$, D_i için $F(\alpha; p+1; n-p-1) = F(0.05; 8; 62) = 2.9$ Covratio_i için $|Covratio_i - 1| > 3p/n = \frac{21}{70} = .30$, SDFİTS_i için $|SDFİTS_i| \geq 2(p/n)^{1/2} = 0.6324$ olarak belirlenmiştir. Kritik değerleri verilmiş olan test istatistiklerinden

r_i istatistiğine göre; Ankara, Bilecik, Kocaeli

t_i istatistiğine göre; Ankara, Bilecik, İstanbul, İzmir, Kocaeli, Rize

D_i istatistiğine göre; İstanbul, Rize

Covratio_i istatistiğine göre; Ağrı, Bilecik, Gaziantep, Kocaeli, Rize, Bayburt

SDFİTS_i istatistiğine göre; Ankara, İstanbul, Kocaeli, Mardin, Bayburt

illeri etkili olabilecek gözlemler olarak belirlenmiştir.

Bu aşamadan sonra, etkili olan bu gözlemlerin çoklubağlantıyı maskeleyen mi yoksa şiddetlendiren mi olduklarını belirlemek için Çizelge 4.6 daki $\cos\theta_{fi}$ değerlerinden faydalanmak mümkündür.

Örneğin, $\cos\theta_{3,6} = 0.7780$ gibi 1 'e yakın bir değer olması Ankara 'nın çoklubağlantıyı maskeleyen bir gözlem olduğunu gösterir. Aynı şekilde $\cos\theta_{1,40} = 0.9898$, 1 'e çok yakın bir değer olduğundan Kocaeli 'nin çoklubağlantıyı şiddetlendiren bir gözlem olduğunu gösterir. Bu

değerlendirmeye göre, Ankara, İstanbul, İzmir, Rize, Bayburt çoklubağlantıyı maskeleyen; Bilecik, Kocaeli, Ağrı, Gaziantep ve Mardin çoklubağlantıyı şiddetlendiren gözlemler olarak belirlenmiştir. Bu aşamadan sonra, çoklubağlantıyı maskeleyen yada şiddetlendiren gözlemlerin, çoklubağlantıyı belirlemekte sıkça kullanılan koşullu indeks ve VIF üzerindeki etkilerini görmek gerekmektedir. Veride ayrı ayrı çoklubağlantıyı maskeleyen ve şiddetlendiren gözlemlerin yer almaması durumunda koşullu indeks ve VIF değerleri Çizelge 4.7 de yer almaktadır.

Çizelge 4.7. Veride ayrı ayrı çoklubağlantıyı maskeleyen ve şiddetlendiren gözlemlerin yer almaması durumunda koşullu indeks ve VIF değerleri

	Çoklubağlantıyı Maskeleyen Gözlemler Veride Yok İken		Çoklubağlantıyı Şiddetlendiren Gözlemler Veride Yok İken	
Değişkenler	Koşullu İndeks	VIF	Koşullu İndeks	VIF
BHTABN	1	1.135	1	1.041
İGBKBÖS	5.3454	2.654	4.7428	17.121
İİDBDHS	11.5272	2.786	15.6285	1.672
İİKKBN	40.7818	1.821	95.5142	21.521
İŞYERİ	73.5090	2.472	105.8571	42.041
ÖBDÖS	77.4363	2.543	134.6142	1.943
ÜÖBDN	1110.9454	1.769	794.3571	1.316

Çoklubağlantıyı maskeleyen gözlemler veride yer almadığı zaman büyük koşullu indeks değeri 1110.9454 olarak elde edilmiştir. Bu değer Çizelge 4.2 deki en büyük koşullu indeks değeri 833.64 ten çok daha büyük olması, çoklu bağlantıyı maskeleyen gözlemlerin gerçekte çoklubağlantının şiddetini bir ölçüde frenlediğini göstermektedir. Aynı şekilde çoklubağlantıyı

şiddetlendirdiği düşünülen gözlemler veride yer almadığı durumda en büyük koşullu indeks değeri 794.3571 olarak elde edilmiştir. Bu değerin 833.64 değerinden daha küçük çıkması, çoklubağlantıyı şiddetlendirdiği düşünülen gözlemlerin , gerçekten çoklubağlantının şiddetini bir ölçüde artırdığını göstermektedir.

Çoklubağlantıyı maskeleyen gözlemler veride yer almaması durumunda VIF değerlerinin en büyük koşullu indeks değerinin tersine çoklubağlantının varlığını göstermemesi ilginç bir durum yaratmaktadır. Bunun sebebi en büyük özdeğer (λ_1) in büyümesi ve diğer özdeğerlerin birbirine yakın değerler olacak şekilde küçülmeleridir. Bu aşamada Denklem (2.11) de verilen

$$VIF_i = \sum_{j=1}^p v_{ij}^2 / \lambda_j$$

formülünden de görüleceği gibi v_{ij}^2 lerin aldığı değerler de önemlidir. Çoklu bağlantıyı şiddetlendiren gözlemlerin veride yer almaması durumunda, VIF değerleri ile koşullu indekslerin aynı yönlü ilişki içinde olması λ_1 , λ_2 , λ_3 ve λ_4 özdeğerlerinin λ_5 , λ_6 ve λ_7 değerlerinden çok büyük değerler almış olmasından kaynaklanmaktadır. Çoklubağlantıyı şiddetlendiren gözlemlerin veride yer almaması durumunda en büyük özdeğer λ_1 değeri düşmüş, diğer özdeğerlerin değerleri yükselmiştir. Bu ise en büyük koşullu indeks değerinin küçülmesine neden olmuştur.

Aşağıdaki Çizelge 4.8 de çoklubağlantıyı maskeleyen ve şiddetlendiren gözlemlerin veride yer almaması durumunda bu verilere ait doğrusal regresyon modelinin parametre tahminlerine ilişkin t değerleri yer almaktadır.

Çizelge 4.8. Doğrusal regresyon modellerine ait parametre tahminlerine ilişkin t değerleri

	Çoklubağlantıyı Maskeleyen Gözlemlerin Veride Yok İken	Çoklubağlantıyı Şiddetlendiren Gözlemler Veride Yok İken
Değişkenler	t değerleri	t değerleri
BHTABN	-3.0000	-1.598
İGBKBÖS	-1.840	0.174
İİDBDHS	-2.113	-5.264
İİKKBN	4.242	-1.172
İŞYERİ	2.051	1.219
ÖBDÖS	-1.225	-0.678
ÜÖBDN	-0.174	-0.937
Sabit	7.376	10.243

$$t_{57;0.05} = 1.675$$

Bu iki modele ilişkin (çoklubağlantıyı maskeleyen ve şiddetlendiren gözlemler kullanılmadığında) belirleme katsayıları da sırasıyla 0.79641 ve 0.78211 olarak elde edilmiştir. Veride çoklubağlantıyı maskeleyen gözlemlerin yer almaması durumunda en büyük koşullu indeks değerinde büyüme olmuştur. Fakat VIF değerlerinin küçük olması model parametrelerine ilişkin tahminlerin çoğunun anlamlı bulunmasını ve belirleme katsayısının da belirgin şekilde yükselmesini sağlamıştır.

5. SONUÇ VE ÖNERİLER

Bir regresyon çalışmasında bağımsız değişkenler arasındaki şiddetli çoklu bağlantı, model parametrelerinin tahminleri ve bunlara ilişkin varyanslar üzerinde olumsuz etkilere neden olmaktadır. Bu olumsuz durum kestirim sonuçları üzerinde de kendisini gösterir.

Şiddetli çoklubağlantının problem olmaktan çıkarılması için çeşitli değişken eliminasyon yöntemleri veya yanlış tahminler veren çeşitli regresyon yöntemleri kullanılmaktadır. Bunların yanısıra şiddetli çoklubağlantı üzerinde, verideki bazı gözlemlerin etkin olabileceği düşüncesi kaçınılmazdır. Bu tür etkili gözlemler çoklubağlantıyı şiddetlendiriyor veya maskeliyor olabilir. Çoklubağlantıyı maskeleyen veya şiddetlendiren gözlemlerin belirlenmesine yönelik yöntemlerin gösterimi için yapılan uygulamadaki veriler Devlet İstatistik Enstitüsü (DİE) yayın şubesinde elde edilmiştir. Kişi başına gayri safi yurtiçi hasıla değerlerinden oluşan bağımlı değişkeni açıklamakta çeşitli sosyo-ekonomik bağımsız değişkenler kullanılmıştır.

Uygulamada çoklubağlantıyı maskeleyen veya şiddetlendiren gözlemlerin belirlenmesi aşamasında kullanılan dik izdüşüm bileşenlerinin değerleri u_{ij}^2 ler ile bu tezde önerilen $\cos \theta_{fi}$ değerleri doğrusal yönlü bir ilişki sergilemişlerdir. Yani her iki yöntemde de hemen hemen aynı gözlemler çoklubağlantıyı maskeleyebilecek veya şiddetlendirebilecek gözlemler olarak belirlenmiştir. Uygulamada da çoklubağlantıyı maskeleyen gözlemler veriden çıkarıldığında, çoklubağlantının şiddetlendiği, çoklubağlantıyı şiddetlendiren gözlemlerin veriden çıkarıldığında ise çoklubağlantının şiddetinin azaldığı daha açık görülmüştür (en büyük koşullu indeks değerlerine göre).

u_{ij}^2 değerleri için bilinen p/n lik sınır değerine benzer bir sınır değeri, i. gözlemin $(X'X)$ matrisinin özvektörleri ile yaptığı açılarının cosünüsleri olan

$\cos \theta_{\ell i}$ ler için $\ell = 1, 2, \dots, p$ olmak üzere $\sum_{j=1}^p \cos^2 \theta_{\ell i} = 1$ olmasından

faydalanarak elde edilebilir. Belirli bir kritik değerden büyük $\cos \theta_{\ell i}$ değeri i. gözlemin ℓ . özdeğeri dolayısı ile ℓ . koşullu indeksi etkileyeceğini gösterecektir. $\cos \theta_{\ell i}$ 'nin $(-1,1)$ aralığında değer alıyor olması i. gözlemin yerinin belirlenmesinde u_{ij}^2 değerlerine göre daha avatajlı olmasını sağlar. Ayrıca dik izdüşüm bileşenleri yönteminin çoklubağlantı üzerinde etkili olabilecek gözlemleri belirlerken sadece izdüşüm değeri yüksek olan gözlemleri baz alması bir dezavantajdır. Bu dezavantaj etkili olabilecek gözlemleri belirlemekte kullanılan tüm istatistikleri baz almasından dolayı $\cos \theta_{\ell i}$ yönteminde oluşmaz. Sonuç olarak $\cos \theta_{\ell i}$ yönteminin dik izdüşüm değerlerine göre iki avantajı bulunmaktadır. Birincisi, $\cos \theta_{\ell i}$ değerlerinin $(-1,1)$ aralığında değer alıyor olması bu yöntemi daha kullanışlı hale getirmektedir. İkincisi, etkili olabilecek gözlemleri belirlemekte kullanılan tüm istatistikleri kapsamış olmasıdır.

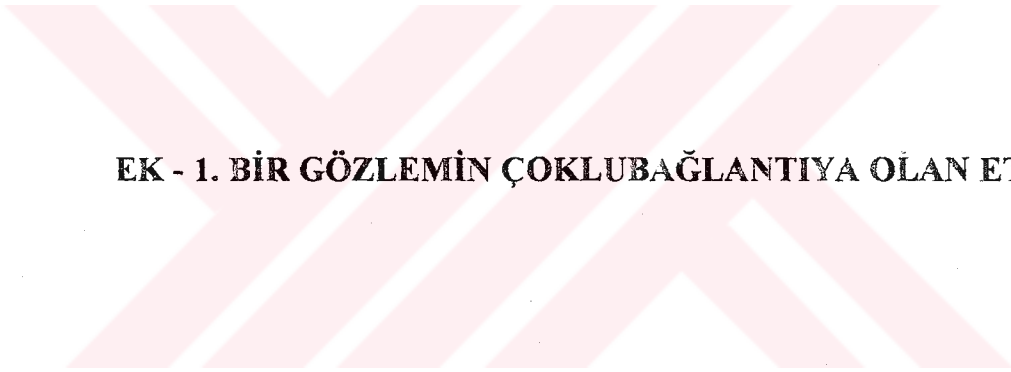
KAYNAKLAR

- Andrews, D.F., Pregibon, D., 1978, Finding the Outliers that Matter: J. Royal Statistical Society, Series B, 40, 85-93.
- Belsley, D.A., Kuh, E., Welsch, R.E., 1980, Regression Diagnostic: John-Wiley Sons, New York.
- Belsley, D.A., 1984, Demaining Conditioning Diagnostics Through Centering (with discussion):The American Statistician,38,73-93.
- Cook,R.D.,1977,Detection of Influential Observations in Linear Regression: Technometrics,19,15-18.
- Cook,R.D.,1979,Influential-Observations in Linear Regression:J.American Statistical Association.,74,169-174.
- Cook. R.D., Weisberg, S., 1982, Residuals and Influence in Regression: New York, Chapman and Hall.
- Daniel, C., Wood, F.S., 1971, Fitting Equations to Data (2 nd edition):, John-Wiley Sons, New York.
- Dorsett, D., Gunst, R.F., Gartland, E.C.Jr.,1983,Multicollinear Effects of Weighted Least Sqaes Regression: Statistics and Probality Letters, 1: 207-211.
- Gunst, R.F., 1983, Regression Analysis with Multicollinear Predictor Variables, Definition, Detection and Effects: Comm. in Stat., 12, 2217-2260.
- Hoaglin, D.C., Welsch, R.E., 1978, The Hat Matrix in Regression and ANOVA: American Statistician, 32, 17-22.
- Hocking, R.R., Pendleton, O.J., 1983, The Regression Dilemma: Commun. Statist.-Theo. Meth., 12(5), 497-527.

- Mandel, J., 1982, Use of the Singular Value Decomposition in Regression Analysis: *The American Statistician*, Vol. 36, No: 1.
- Marquardt, D.W., 1970, Generalized Inverses, Ridge Regression, Biased Linear Estimation and Nonlinear Estimation: *Technometrics*, 12, 591-612.
- Mason, R.L., Gunst, R.F., 1985, Outlier-Induced Collinearities: *Technometrics*, Vol. 27, No: 4.
- Montgomery, D.C., Peck, E.A., 1982, *Introduction to Linear Regression Analysis*: John-Wiley Sons., New York.
- Myers, R.H., 1986, *Classical and Modern Regression with Applications*: Duxbury Press Boston, MA. p. 237-238.
- Öztürk, F., 1992, İstatistiğin Geometrik Bakışı: DİE Araştırma Sempozyumu, Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü, Ankara.
- Rawlings, J.O., 1982, *Applied Regression Analysis*: John-Wiley Sons., New York.
- Styan, G.P.H., Markku, N., Simo, P., 1993. Simplified Matrix Proofs Related to the Deletion of an Observation in the General Linear Model: *Communications in Statistics A-Theory and Methods*, 9.
- Velleman, P.F., Ypelear, M.A., 1980. Constructing Regressions with Controlled Features, A Method of Probing Regression Performance: *JASA*, 75, 839-844.
- Velleman, P.F., Welsch, R.E., 1981, Efficient Computing of Regression Diagnostics: *The American Statistician*, 35, 234-242.
- Walker, E., 1989, Detection of Collinearity-Influential Observations: *Commun. Statist-Theory Meth.*, 18(5), 1675-1690.



EKLER



EK - 1. BİR GÖZLEMİN ÇOKLUBAĞLANTIYA OLAN ETKİSİ

X_j değişkenleri toplamaları sıfır,kareleri toplamı bir olacak şekilde standartlaştırılırsa,i.satı,j,sütundaki gözlem değeri

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

olur.i = 1 için;

$$Z_{11} = \frac{x_{11} - \bar{x}_1}{\sqrt{\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2}} \quad \text{olmalıdır.Buna göre}$$

$$x_{11} - \bar{x}_1 = \theta s_{11} - \frac{(\theta s_{11} + s_{21} + \dots + s_{n1})}{n} = \frac{(n-1)\theta s_{11}}{n} - \underbrace{\frac{(s_{21} + \dots + s_{n1})}{n}}_A \quad \text{ve}$$

$$\sum_{i=1}^n (x_{i1} - \bar{x}_1)^2 = \sum_{i=1}^n x_{i1}^2 - n\bar{x}_1^2$$

$$= \theta^2 s_{11}^2 + s_{21}^2 + \dots + s_{n1}^2 - n \frac{(\theta s_{11} + s_{21} + \dots + s_{n1})^2}{n^2}$$

$$= \theta^2 s_{11}^2 - \frac{\theta^2 s_{11}^2}{n} + \underbrace{s_{21}^2 + \dots + s_{n1}^2 - \frac{(s_{21} + \dots + s_{n1})^2}{n}}_B \quad \rightarrow \text{Bu açılım } \theta_{s11} \text{ 'i kapsamaktadır. Fakat bu göz ardı edilmektedir.}$$

$$= \frac{(n-1)\theta^2 s_{11}^2}{n}$$

elde edilir.

$$Z_{11} = \frac{x_{11} - \bar{x}_{.1}}{\sqrt{\sum_{i=1}^n (x_{i1} - \bar{x}_{.1})^2}} = \frac{\left(\frac{n-1}{n}\right)\theta s_{11} - A}{\sqrt{\left(\frac{n-1}{n}\right)\theta^2 s_{11}^2 + B}}$$

olur. θ dan bağımsız olan ve önemsenmeyen A ve B ifadeleri çıkarılırsa

$\theta, s_{11} > 0$ olduğundan

$\theta \rightarrow \infty$ için;

$$Z_{11} \rightarrow \frac{(n-1)\theta s_{11}}{(n-1)^{1/2}\theta s_{11}} = \left(\frac{n-1}{n}\right)^{1/2} \text{ değerine yakınsar}$$

Aynı şekilde $i=1, j=2$ için.

$$Z_{12} = \frac{x_{12} - \bar{x}_{.2}}{\sqrt{\sum_{i=1}^n (x_{i2} - \bar{x}_{.2})^2}}$$

$$\sum_{i=1}^n (x_{i2} - \bar{x}_{.2})^2 = \sum_{i=1}^n x_{i2}^2 - n\bar{x}_{.2}^2$$

$$= \theta^2 s_{12}^2 + s_{22}^2 + \dots + s_{n2}^2 - n \frac{(\theta s_{12} + s_{22} + \dots + s_{n2})^2}{n^2}$$

$$= \theta^2 s_{12}^2 - \frac{\theta^2 s_{12}^2}{n} + s_{22}^2 + \dots + s_{n2}^2 - \frac{(s_{22} + \dots + s_{n2})^2}{n}$$

$$= \frac{(n-1)\theta^2 s_{12}^2}{n} + C$$

$$Z_{12} = \frac{\left(\frac{n-1}{n}\right)\theta s_{12} - D}{\sqrt{\left(\frac{n-1}{n}\right)\theta^2 s_{12}^2 + C}} \text{ elde edilir.}$$

$\theta, s_{12} > 0$ olduğundan

$\theta \rightarrow \infty$ için;

$$Z_{12} \rightarrow \frac{(n-1)\theta s_{12}}{(n-1)^{1/2}\theta s_{12}} = \left(\frac{n-1}{n}\right)^{1/2} \text{ olur.}$$

Görüldüğü gibi Z_{1j} ; $j=1,2$ için $\left(\frac{n-1}{n}\right)^{1/2}$ değerine yakınsamaktadır. Diğer taraftan Z_{ij} ; $i \neq 1$, $j=1,2$ için , gözlemlerin hepsi aynı değere yakınsamaktadır. Bu nedenle bir gözlem için $-(n(n-1))^{-1/2}$ değerine yakınsadığını göstermek yeterli olacaktır.

$i=2$, $j=1$ için ;

$$Z_{21} = \frac{x_{21} - \bar{x}_{.1}}{\sqrt{\sum_{i=1}^n (x_{i1} - \bar{x}_{.1})^2}}$$

yazılır.

$\sum_{i=1}^n (x_{i1} - \bar{x}_{.1})^2 \rightarrow Z_{11}$ de gösterilmişti. $(x_{21} - \bar{x}_{.1}) = \frac{-s_{11}}{n} \theta$ olarak elde edilir.

Yukarıda yapılan işlemlere benzer şekilde devam edilirse

$\theta, s_{11} > 0$ ve $\theta \rightarrow \infty$ için;

$$Z_{21} \rightarrow \frac{-\frac{s_{11}}{n}\theta}{\left(\frac{n-1}{n}\right)^{1/2} s_{11}\theta} = -\frac{1}{n} \frac{n^{1/2}}{(n-1)^{1/2}} = -\frac{1}{n^{1/2}(n-1)^{1/2}} = -(n(n-1))^{-1/2}$$

olur.



ÖZGEÇMİŞ

Bu tezi yazan Ufuk EKİZ 17.02.1970 tarihinde Ankara 'da doğmuştur. İlk öğrenimini Ankara Açıkalın İlköğretim okulunda, orta öğrenimini Cebeci Ortaokulu ve Ankara Lisesinde, Yüksek öğrenimini Gazi Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümünde tamamlanmıştır. Şu anda Gazi Üniversitesi Fen Edebiyat Fakültesi İstatistik Bölümünde Araştırma Görevlisi olarak çalışmaktadır.

