DOĞRUSAL REGRESYONDA GÖZLEMLERİN KÜMELEMEDEN GELECEK ETKİLERİ ÜZERİNE BİR ÇALIŞMA

Gülay Başarır

Hacettepe Üniversitesi
Fen Bilimleri Enstitüsü Yönetmeliği'nin
İstatistik Anabilim Dalı için öngördüğü
BİLİM UZMANLIĞI TEZİ
Olarak Hazırlanmıştır

Temmuz – 1987

Fen Bilimleri Enstitüsü Müdürlüğü'ne

İşbu çalışma, jürimiz tarafından
Ana Bilim Dalında Bilim UZMANLIĞI TEZİ olarak
kabul edilmiştir.

Başkan : Prof. Dr. Soner GÖNEN

Üye : Yard. Doç. Dr. Hüseyin TATLIDİL

Üye : Yard. Doç. Dr. Aydın ERAR

ONAY

Yukarıdaki imzaların, adı geçen öğretim üyelerine ait
olduğunu onaylarım.

/ /1987

Prof. Dr. ACAR IŞIN
Fen Bilimleri Enstitüsü Müdürü

Ö Z E T

Bu çalışmanın amacı, kümelenebilir veriler üzerinden önkestirim yapılması istendiğinde, kümelemeden kaynaklanabilecek etkin gözlemin modeli ve önkestirimi nasıl etkilediğini ve hangi etkenlere bağlı olduğunu araştırmaktır.

Birinci Bölümde, etkin gözlem incelemesinde ki bazı sorunlara değinilerek çalışmanın amacı belirtildi.

İkinci Bölümde, etkin gözlemi belirlemeye yönelik grafiksel yöntemler ve ölçütlere ilişkin bir derlemenin sunulduğu genel bilgiler verildi.

Üçüncü Bölümde, amaca yönelik olarak yapılan benzeşim çalışması anlatılarak bulguları verildi. Benzeşim çalışmasının sonuçlarına göre bir kümeden tek olarak alınan gözlemin doğrusal regresyonda etkin gözlem durumuna gelmesi nedeniyle kümelemenin bir etkinlik kaynağı olabileceği görüldü. Kümelemeden kaynaklanan etkin gözlemin özellikle Cook ölçütünü ve kestirimleri etkilediği, önkestirim üzerinde ise etki yaratmadığı gösterildi. Bu bölümde gerçek verilerle bir uygulama da yapıldı.

Son Bölümde, elde edilen sonuçlara göre, kümelenebilir verilerle yapılan çalışmalarda, önkestirimin güvenle yapılabileceği ancak kümelemenin bazı ölçütler ve kestirimler üzerindeki etkisi nedeniyle her kümeden en az iki gözlem alınması gerektiği vurgulandı.

S U M M A R Y

The purpose of this study is to investigate influential observation due to clustering or to find out the factors which it is connected with, when intended to make a prediction on the basis of clusterable data.

In the First Section, some problems encountered in the influential observation has been pointed out.

In the Second Section, various graphical methods and criteria which are used for determining influential observation has been examined.

In the Third Section, a simulation study has been performed for examining effects of the several factors. According to results, clustering could be a source of influence when a single observation was taken from a cluster. It has been also pointed out that an influential observation due to clustering had effect especially on Cook criteria and estimation but had no effect on prediction. At least the criteria and graphical methods have been employed on the numerical example.

In the Last Section, it has been emphasized that, prediction can safely be made in the studies whith clusterable data but owing to the influence of clustering on certain criteria and estimation, at least two observation should be taken from each cluster.

TEŞEKKÜR

Bu çalışmanın her aşamasında değerli katkı ve eleştirileri ile bana yön veren danışmanım Sayın Yrd.Doç.Dr. Aydın Erar'a, çalışmama yapıcı uyarılarda bulunan Sayın Yrd. Doç.Dr. Hüseyin Tatlıdil'e ve Sayın Prof.Dr. Alaettin Kutsal'a ve çalışmam boyunca yardımlarını esirgemeyen tüm bölüm arkadaşlarıma ve aileme teşekkür ederim.

İÇİNDEKİLER DİZİNİ

İÇİNDEKİLER DİZİNİ (devam ediyor)

ŞEKİLLER DİZİNİ

ÇİZELGELER DİZİNİ

BİRİNCİ BÖLÜM

GİRİŞ

Doğrusal regresyonda gözlem etkilerini incelemede ki amaç, aykırı değer, uç değer, etkin gözlem (influential observation) gibi gözlemleri tanımlamak, gözlemlerin model üzerindeki etkilerini incelemek, model yapısı hakkında bilgi edinmek ve değişkenlerin model üzerindeki katkılarının gözlem etkilerine göre bir ön incelemesini yapmaktır. Böylece değişen varyanslılık, otokorelasyon, normal dağılmama gibi durumlarda çözüm için bazı önbilgilerin yanısıra değişken seçimi öncesinde ve model geçerliliğinin incelenmesinde bir önbilgi elde edilir. Gözlem etkilerinin incelenmesi veri çözümlemesinin bir aşaması olarak da düşünülebilir.

Öte yandan veri kümesi incelenirken etkin gözlem ortaya çıkması durumunda bu gözlemin ne yapılacağına karar verilmesi gerekir. Çoğu yazarların araştırmalarında bu sorun, etkin gözlemin veri kümesinden çıkartılıp kalan gözlemler üzerinden araştırmanın tekrar yapılması biçiminde giderilmektedir (Cook, 1979).

Pratikte karşılaşılan bir sorun, küçük örneklemlerde bulunan etkin gözlemin ölçütleri etkileyerek araştırıcıyı yanlış sonuçlara götürmesi ve gözlem sayısının azlığı nedeni ile veri kümesinden çıkartılmak istenmemesidir. Bu durumda, etkin gözlemin gerek modelde bulunmasının gerekse veri kümesinden çıkartılmasının etkilerini incelemek ve etkinliğin kaynağını araştırmak sorunu çözümlemeye yardımcı olur.

Çalışmada, bu sorundan hareketle, kümelenebilir veriler üzerinde yapılan uygulamalarda karşılaşılacak etkin gözlemin kümelemeden kaynaklanıp kaynaklanmadığı araştırılacaktır.

Buna göre çalışmanın ana amacı, verilerin kümelenebildiği durumlarda, özellikle dışarıdan verilen gözlemler üzerinden önkestirim istenildiğinde, etkin gözlemin önkestirimi nasıl etkileyeceğinin ve hangi etkenlere bağlı olduğunun araştırılmasıdır.

Bununla birlikte çalışma, uygulamalardan çıkan bazı sorun ve önsezilere dayanılarak, kümelenebilir verilerde bir kümeden tek bir gözlemin örnekleme seçilmesinin etkisi üzerinde yoğunlaşacaktır. Böylece araştırıcının etkilerini ve kaynağını bildiği etkin gözlemin ne yapılması gerektiği konusunda daha kolay karar verebileceği düşünülmüştür.

Bu amaçlar çerçevesinde çalışmanın ikinci bölümü, etkin gözlemi belirlemeye yönelik grafiksel yöntemler ve ölçütlere ilişkin bir derlemenin sunulduğu genel bilgilere ayrılmıştır. Üçüncü bölümde ölçütlerin ve grafik yöntemlerinin irdelenmesi, benzeşim çalışması ve bulguları ile gerçek veriler üzerinde yapılan bir uygulama yer almaktadır. Sonuçların değerlendirilmesi ve tartışma son bölümdedir.

İKİNCİ BÖLÜM

GENEL BİLGİLER

Üzerinde çalışılan veri kümesindeki etkin gözlemlerin araştırılmasına yönelik çalışmalar önce etkinliğinden kuşku duyulan gözlem değerlerinin test edilmesi ile başlamıştır. Bu konuda ileri sürülen çeşitli ölçütler, gözlemlerin bir alt kümesinin etkin olduğu durumlara genelleştirilmiştir. Çalışmada tek bir etkin gözlem olması durumu araştırılacak olup bu bölümde ilk olarak etkinlik kavramı ve etkinliğin kaynakları anlatılacak, daha sonra etkin gözlemin ölçütler üzerindeki etkileri ve etkinliği ortaya çıkartıcı ölçütler ile grafiksel yöntemler tanıtılacaktır.

2.1. Doğrusal Regresyon Modeli ve Etkin Gözlemler

Alışılagelen bir doğrusal regresyon modeli,

$$\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon} \qquad (2.1)$$

biçiminde verilir. Gözlem sayısı, n; açıklayıcı değişken sayısı, k olmak üzere $\underline{X}$, $n \times (k+1)$ boyutlu ve $k+1$ ranklı açıklayıcı değişkenler matrisi; $\underline{Y}$, $n \times 1$ boyutlu yanıt vektörü; $\underline{\beta}$, $(k+1) \times 1$ boyutlu bilinmeyen katsayılar vektörü; $\underline{\varepsilon}$ ise, $n \times 1$ boyutlu ve $E(\underline{\varepsilon}) = \emptyset$, $V(\underline{\varepsilon}\,\underline{\varepsilon}') = \sigma^2 \underline{I}_n$ koşullarını sağlayan yanılgı (hata) vektörüdür. İstatistiksel çıkarsamalarda $\underline{\varepsilon} \sim N(0, \sigma^2 \underline{I}_n)$ istenilir. Modelin en küçük kareler (EKK) kestiricisi, $\hat{\underline{\beta}} = (b_o, b_1, \ldots, b_k)'$ olmak üzere, $\hat{Y} = \underline{X}\,\hat{\underline{\beta}}$ ile gösterilir (Draper and Smith, 1981, p.22-24).

Öte yandan (2.1) modelinin EKK kestiriminden elde edilen artık vektörü (residual vector), $\hat{\underline{\beta}} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{Y}$ olmak üzere,

$$\underline{e} = \underline{Y} - \underline{X}\,\hat{\underline{\beta}} \qquad (2.2)$$

biçiminde; herbir artık teriminin kendi standart sapmasına bölünmesiyle elde edilen i. studentize artık da (studentized residual),

$$r_i = \frac{e_i}{\hat{\sigma}\sqrt{1-v_{ii}}} \quad , \quad i=1,\ldots,n \qquad (2.3)$$

biçiminde verilir (Cook, 1977). Burada $\hat{\sigma}$, yanılgı varyansının kestiricisidir; v_{ii}, $\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'$ nün i. köşegen ögesi olup; i. gözlemin, veri kümesinin merkezine uzaklığının bir ölçüsüdür (Cook, 1977).

Etkin gözlem, veri kümesinden çıkartıldığında kestirim sonuçlarını büyük ölçüde değiştiren gözlem olarak tanımlanır (Cook, 1979; Cook and Weisberg, 1980; Beckman and Cook, 1983). Ancak bu tanım, çoğu kez, aykırı değer (outlier) ve uç değer (extreme value, high leverage point) kavramları ile birlikte incelenir.

Aykırı değer, en büyük artıklı değer ya da veri kümesindeki öteki gözlemlere uzak olan gözlem değeri olarak verilir (Cook, 1977; Draper and John, 1981; Beckman and Cook, 1983; Hocking, 1983). Bununla birlikte Weisberg, e_i'ler raslantı değişkeni olduğu için bazı gözlemlerin, gerçekten aykırı olmamalarına karşın büyük artıklı çıkabileceklerine dikkat çekmiştir (Weisberg, 1980, p.114).

Beckman ve Cook ile Stevens, aykırı değerin, etkin gözlemden farklı olarak veri kümesinden çıkarıldığında çözümleme sonuçlarını değiştirmeyeceğini belirtmişler; etkin gözlemin, aykırı değerin özel bir türü olduğunu vurgulamışlardır (Beckman and Cook, 1983; Stevens, 1984).

Uç değer, veri kümesi hacminden uzak olan gözlem değeri olarak tanımlanır (Hocking,1983). Tatlıdil (1981), uç değerin, aykırı değerden farkını, gözlem kümesinde yer alması ve çalışmaya katılması olarak belirtmiştir.

Bu tanımlara açıklık getiren yorumları Belsley ve diğerleri grafikler üzerinde vermişlerdir (Belsley, et.al., 1980, p.7-9).

2.2. Etkinliğin Kaynakları

Belsley ve diğerleri (1980, p.6). Beckman ve Cook (1983) ile Stevens (1984), etkin gözlemin kaynaklarını aşağıdaki gibi özetlemişlerdir :

(1) Verilerde önemli miktarda ölçüm, tartım ve kaydetme hatalarının ortaya çıkması,
(2) Veriler için kestirilen model yapısının yanlışlığı,
(3) Aykırılığı yaratan gözlemlerin uç değerler olabilmesi. Bu tür bir gözlem, tüm gözlemlerin yoğunlaştığı bölgenin dışındaki bir bölgede model hakkında tek başına bilgi sağlayabileceği için veri kümesinde en önemli gözlem olma eğilimindedir.

Ayrıca, Gunst ve Mason (1980, p.252), tek tek gözlemlerin değişkenler üzerinde anlamlı etkilerinin olmasını da bir etkinlik kaynağı olarak vermişlerdir.

2.3. Etkin Gözlemin Etkileri

Etkin gözlemin modelde bulunması önkestirim (prediction) değerlerini etkileyebileceği gibi katsayı kestirimlerini de işaret ve büyüklük olarak etkileyecektir. Ayrıca bu tür gözlemin varlığı denklemin Artık Kareler Ortalamasını ($\hat{\sigma}^2$) şişirecektir (Weisberg, 1980, p.116-117).

Cook (1979), etkin gözlemin veri kümesinden çıkartılmasının studentize artıklar, artık varyansları ve kısmı F (partial F) değerleri üzerindeki etkisini kuramsal olarak göstermiştir: $\underline{\hat{\beta}}(i)$, i. gözlemin veri kümesinden çıkartılmasından sonra elde edilen kestirim vektörü ve $\underline{x}_i = (1\ x_{i1}, x_{i2}, \ldots, x_{ik})'$ olmak üzere,

$$\underline{\hat{\beta}} - \underline{\hat{\beta}}(i) = (\underline{X}'\underline{X})^{-1}\underline{x}_i e_i/(1-v_{ii}) \qquad (2.4)$$

ve $\hat{\sigma}^2(i)$, i. gözlem çıkartıldıktan sonra ki σ^2 kestirimi iken,

$$(n-k)\ \hat{\sigma}^2 = (n-k-1)\ \hat{\sigma}^2(i) + \frac{e_i^2}{1-v_{ii}} \tag{2.5}$$

biçiminde verilir.

$\underline{X}(i)$, $\underline{X}$ matrisinden i. satırın çıkartılması ile elde edilen matris olmak üzere, r'inci ve ℓ'inci satırlar için,

$$\underline{X}(i)'\ \underline{X}(i) = \underline{X}'\underline{X} - \underline{x}_i\underline{x}_i' \tag{2.6}$$

$$v_{r\ell}(i) = \underline{x}_r'\ (\underline{X}(i)'\underline{X}(i))^{-1}\underline{x}_\ell \tag{2.7}$$

$$v_{r\ell} = v_{r\ell}(i) - \frac{v_{ri}(i)\ v_{\ell i}(i)}{(1+v_{ii}(i))} \tag{2.8}$$

$$v_{r\ell}(i) = v_{r\ell} + \frac{v_{ri}\ v_{\ell i}}{(1-v_{ii})} \tag{2.9}$$

eşitlikleri yazılabilir (Cook, 1979).

(1) önkestirim varyansı üzerindeki etkileri: φ_{ij}, i. ve j. artıklar arasındaki korelasyonu göstermek üzere Cook, (2.9) eşitliğinden,

$$\frac{v_{jj}(i)}{(1-v_{jj}(i))} = \frac{v_{jj}(1-\varphi_{ij}^2)+\varphi_{ij}^2}{(1-v_{jj})(1-\varphi_{ij}^2)}, \quad j \neq i \tag{2.10}$$

oranını vermiştir. Bu oran, v_{jj} ya da φ_{ij}^2 'nin büyük olduğu durumda büyük olacaktır. Eğer φ_{ij}^2 büyük ise, i. gözlem veri kümesinden çıkartıldığında j. gözlemin ön kestirim varyansı artar. Eğer bu korelasyon ihmal edilebilirse, herhangi bir gözlemin çıkartılması önkestirim varyansını değiştirmeyecektir (Cook, 1979).

(2) studentize artık üzerindeki etkileri: (2.5) ve (2.9) eşitlikleri kullanarak, i.gözlem veri kümesinden çıkartıldıktan sonra j. gözlemin studentize artığı,

$$r_j(i)^2 = \frac{(n-k-1)(r_j-\varphi_{ij}r_i)^2}{(n-k-r_i^2)(1-\varphi_{ij}^2)} \tag{2.11}$$

biçiminde elde edilir.

Eğer artık korelasyonu φ_{ij}^2, ihmal edilebilirse, studentize artığı 1'den büyük olan gözlemin çıkartılması, kalan tüm studentize artıkları artıracaktır. Büyük korelasyon değerleri, i. gözlem çıkartıldığında j. gözlemin studentize artığının artmasına neden olur (Cook, 1979).

(3) kısmi-F'ler üzerindeki etkileri: b_j, $\underline{\hat{\beta}}$'nın j. bileşenini ve d_j, $(\underline{X}'\underline{X})^{-1}$ matrisinin j. köşegen ögesini göstermek üzere,

$$T_j = \frac{b_j}{\hat{\sigma} d_j^{\frac{1}{2}}} \tag{2.12}$$

ölçütü, $\beta_j = 0$ hipotezinin kısmi-F testi için, $T_j^2 = F_j$ olarak kullanılır.

(2.4) ve (2.5) eşitlikleri kullanılarak $F_j(i)$, i. gözlem çıkartıldıktan sonraki kısmi F ölçütünü göstermek üzere,

$$F_j(i) = \frac{(n-k-1)(T_j - \gamma\ r_i w_i^{\frac{1}{2}})^2}{(n-k-r_i^2)(1+\gamma^2 w_i)} \tag{2.13}$$

olarak bulunmuştur. Burada $w_i = v_{ii}/(1-v_{ii})$ ve γ, b_j ile $\underline{x}_i'\ \underline{\hat{\beta}}$ arasındaki korelasyonu gösterir. Eğer $\gamma w_i^{\frac{1}{2}}$ ihmal edilebilirse, $r_i^2 > 1$ olan gözlemin veri kümesinden çıkartılması tüm kısmi-F değerlerini artıracaktır. Eğer $r_i^2 < 1$ ve $F_j > 1$ iken $\gamma w_i^{\frac{1}{2}}$ yeterince büyükse, i. gözlemin çıkartılması kısmi-F değerlerini küçültecektir (Cook, 1979).

2.4. Gözlem Etkilerinin Ortaya Çıkartılması

Gözlem etkilerini ortaya çıkartmada kullanılan grafikler, yorum kolaylığı getirirler ancak verdikleri bilginin soyut olması nedeni ile genellikle istatistiksel ölçütlerle birlikte incelenirler.

2.4.1. grafik yöntemleri

2.4.1.a. klasik çizimler

Klasik çizimler içerisinde en basit olanı, açıklayıcı değişkenlerin biribirlerine karşı ve açıklanan değişkene karşı olan çizimleridir. Bu çizimlerle, değişkenlere uydurulan modelin geçerliliği yanında veri kümesindeki farklı gözlemler de incelenebilir (Daniel and Wood, 1980, p.51-52). Herbir açıklayıcı değişkene karşı studentize artığın (r_i) çizimi, artıkların dağılımını açıklayıcı değişkenin bir fonksiyonu olarak verir. Bu çizimlerle, ilgili açıklayıcı değişkenin doğru fonksiyonel şekli seçilebilir; karesel ya da etkileşim terimlerine gerek duyulup duyulmadığına karar verilebilir. Bu tür çizimler, r_i'ye karşı $\hat{y}$'nın çiziminde olduğu gibi, varsayım bozulumlarını, aykırı ve uç değerleri araştırmada kullanılabilir (Daniel and Wood, 1980, p.27-28; Gunst and Mason, 1980, p.242-246; Weisberg, 1980, p.121; Draper and Smith, 1981, p.149-151). Ancak klasik çizimler ile gözlemlerin tek tek etkileri görülemiyebilir. Bu çizimler, özellikle, gözlemlerin biribirlerinin etkilerini maskeledikleri durumda yetersiz kalırlar. Ayrıca başka değişkenlerin etkisi söz konusu olduğunda, değişken sayısı arttıkça grafiklerdeki netlik kaybolur. Bu durumlarda yetersiz kalan klasik çizimler dışında başka grafik yöntemlerine gereksinim duyulur (Gunst and Mason, 1980, p.247; Belsley, et.al, 1980, p.30).

2.4.1.b. kısmi artık çizimleri

Kısmi artık çizimleri, temelde, Draper ve Smith'in değindikleri kısmi artık ilişkilerine dayanarak farklı amaçlarla değişik biçimlerde verilmiştir:

(1) Gunst ve Mason'un kısmi-artık çizimleri: Gunst ve Mason, bir x_j değişkeni için kısmi-artık terimini, y vektörünün i. öğesi y_i olmak üzere,

$$e_i^{\star} = y_i - (\hat{y}_i - b_j x_{ij})$$

$$= e_i + b_j x_{ij} \quad , \quad i=1,\ldots,n \qquad (2.14)$$
$$j=1,\ldots,k$$

biçiminde tanımlamışlardır. $b_j x_{ij}$ terimi, (2.1) modelinin kestiriminden elde edilen terimdir. (2.14) eşitliği, $\underline{X}$ matrisinden j. değişken çıkartıldıktan sonra bulunan matris $\underline{X}_{(-j)}$ olmak üzere,

$$\underline{Y} = f(\underline{X}_{(-j)}) + \underline{u}$$

ve

$$\underline{X}_j = f(\underline{X}_{(-j)}) + \underline{v} \qquad (2.15)$$

bağıntılarından elde edilen $\underline{u}$ ve $\underline{v}$ artıkları arasındaki

$$\underline{u} = b_j \underline{v} + \underline{e} \qquad (2.16)$$

ilişkisi yardımıyla gösterilebilir (Draper and Smith,1980, p.196-201). Grafik, $e_i^{\star}$ 'a karşı $\underline{X}_j$'nin çizimi biçiminde olup öteki açıklayıcı değişkenlerin etkisi arıtıldıktan sonra Y-$\underline{X}_j$ arasındaki ilişkiyi belirler. Gunst ve Mason, bu çizimlerle, X_j değişkeninin doğrusallığının yönü ve derecesine ilişkin bilgi edinilebileceğini ve dönüşüm türüne karar verilebileceğini belirtmişlerdir (Gunst and Mason, 1980, p.247-252).

(2) Belsley ve diğerleri'nin kısmi artık çizimleri: Belsley ve diğerleri, artık çizimlerinin (2.15) eşitliğinde verilen $\underline{u}$ ve $\underline{v}$ kismi artıkları arasında yapılmasını önermişlerdir.

Bu çizimler, (2.16)'dan yararlanarak $\underline{v} = \underline{X}_j - \hat{\underline{X}}_j$ bulunduktan sonra $\underline{u} = b_j(\underline{X}_j - \hat{\underline{X}}_j) + \underline{e}$'den $\underline{u}$'nun elde edilmesi ile yapılabilir.

Kısmi artıklar arasında yapılan bu çizimlerle gözlemlerin, katsayı kestirimlerini, birlikte nasıl etkiledikleri gözlenebilir; aykırı ve uç değerler belirlenebildiği gibi maskeleme durumları hakkında da bir önbilgi elde edilebilir.

Karşılaşılan maskeleme durumları grafiksel olmayan ileri inceleme tekniklerini gerektirir (Belsley, et.al., 1980, p.30, 51-54).

(3) bileşen ve bileşen + artık çizimleri: Daniel ve Wood herbir açıklayıcı değişkene ilişkin bileşen (component) etkisini, $\bar{x}_j$, j. açıklayıcı değişkenin ortalaması olmak üzere,

$$\begin{aligned} c_{ij} &= b_j(x_{ij} - \bar{x}_j) \\ &= b_j x_j^{\star} \qquad i=1,\ldots,n\ ;\ j=1,\ldots,k \end{aligned} \tag{2.17}$$

biçiminde tanımlamışlardır. Bileşen terimine her bir gözleme ilişkin artık terimini ekleyerek, bileşen + artık ifadesi,

$$c_{ij} + e_i = b_j x_j^{\star} + e_i \tag{2.18}$$

olarak elde edilir. Bu eşitlik, (2.14)'deki kısmi artık terimiyle aynı olup, tek farkı, burada $\underline{x}_j$'lerin merkezileştirilmiş olmasıdır. Grafik, c_{ij}'ye karşı X_j'nin çizimi şeklinde olup, buradan elde edilen bileşen doğrusunun eğimi, x_j değişkeninin etkisinin önemliliğini gösterir. Aynı grafik üzerinde bileşen + artık değerleri de işaretlenerek artıkların dağılımı incelenebilir. Bileşen doğrusu çevresinde kümelenen artıklardan uzak bir gözlemin etkinliğinden kuşku duyulur. Daniel ve Wood, ilgili değişkenin bu gözlemden ne kadar etkilendiğini araştıran uygulamalar yapmıştır.

Bu tür çizimlerle uygun denklem biçimi araştırılabilir, herbir açıklayıcı değişken aralığı üzerinden gözlemlerin dağılımı gözlenebilir ve gözlemlerin denklemdeki herbir açıklayıcı değişken üzerinde ki etkisi araştırılabilir (Wood, 1973; Daniel and Wood, 1980, p.124-125).

2.4.1.c. duyarlılık çizimleri

Baskerville ve Toogod'un incelediği duyarlılık çizimleri, herbir gözlemin, i. gözlem veri kümesinden çıkartıldıktan sonraki parametre kestirim vektörü $\hat{\underline{\beta}}(i)$'ye karşı çizimi şeklindedir. Bu çizimlerle, herbir gözlemin çıkartılmasının değişkenler üzerindeki etkisi yani regresyon katsayılarının her veri noktasına karşı duyarlılığı incelenebilir.

Herbir açıklayıcı değişken için yapılan duyarlılık çizimleri aynı grafikte yer alır. Çizimler irili ufaklı dişlerden oluşur ve $\hat{\underline{\beta}}(i)$ üzerideki sıfıra yakın dişlere karşılık gelen gözlemin ilgili açıklayıcı değişkenin katsayısını şişireceğini gösterir. Sıfırdan uzak dişler için tersi yorum yapılır (Baskerville and Toogood, 1982).

Baskerville ve Toogood, duyarlılık çizimlerinin, modeldeki tüm açıklayıcı değişkenleri tek grafikte içermesi, katsayı kestirimlerini etkileyen gözlemlere ilişkin bilgi edinerek model uyumu çözümlemesinin yapılabilmesi açılarından yararlı çizimler olduklarını belirtmişlerdir (Baskerville and Toogood, 1982).

2.4.2. istatistiksel ölçütler

Bu kesimde, gözlem etkilerini ortaya çıkartan istatistiksel ölçütler tanıtılacaktır.

2.4.2.a. temel ölçütler: Artık, studentize artık, V_{ii}, t_i ölçütleri

Bu kesimde incelenen ölçütler daha çok aykırı ve uç değer hakkında bir önbilgi sağlarlar ve öteki ölçütler için bir temel oluştururlar:

Artık: (2.1) modelinden artık vektörü,

$$\begin{aligned} \underline{e} &= \underline{Y} - \hat{\underline{Y}} \\ &= (\underline{I} - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}')\underline{Y} \\ &= (\underline{I} - \underline{V})\ \underline{Y} \end{aligned} \tag{2.19}$$

biçiminde de yazılır.

Weisberg, aykırı değer için $E(\hat{y}_i) \neq \underline{x}_i' \underline{\beta}$ olup, aykırı değere en iyi adayın $|\hat{e}_i|$ 'nın en büyük değerli gözlem olacağını belirtmiştir (Weisberg, 1980, p.114).

v_{ii} ölçütü: Kesim 2.1'de tanımlandığı gibidir. Çeşitli özellikleri Weisberg (1980, p.101-105) tarafından verilmiştir.

Bu ölçüt, $\underline{x}_j$ değişken değerleri merkezileştirildikten sonra elde edilen çarpımlar matrisi $\underline{X}'\underline{X}$ olarak gösterildiğinde,

$$v_{ii} = \frac{1}{n} + (\underline{x}_i - \bar{\underline{x}})' \, (\underline{X}'\underline{X})^{-1} (\underline{x}_i - \bar{\underline{x}}) \qquad (2.20)$$

biçiminde de verilir (Weisberg, 1980, p.101-104).

Belsley ve diğerleri, v_{ii} ölçütü için 2 k/n değerini kriter olarak önermişler ve bu kriteri aşan v_{ii} değerine sahip gözlemlerin uç değer olabileceğini belirtmişlerdir (Belsley, et.al., 1980, p.17).

Öte yandan i. artık değerinin varyansı (2.19) eşitliğinden,

$$\mathrm{Var}(\hat{e}_i) = \hat{\sigma}^2 (1 - v_{ii}) \qquad (2.21)$$

yazılır. Buradan, büyük v_{ii} değerli gözlemlerin hata varyansının küçük olacağı, v_{ii} değeri 1'e yaklaştıkça bu varyansın sıfıra yaklaşacağı ve $\hat{y}_i \cong y_i$ olacağı söylenebilir (Weisberg, 1980, p.103).

Studentize artık: Artık incelemesini kolaylaştırmak amacı ile her artığın standart hatasına bölünmesi ile elde edilen studentize artık ölçütü (2.3)'de verildiği gibidir.

$E(r_i) = 0$ ve model doğru olduğunda $\mathrm{Var}(r_i) = 1$ dir. Aksi takdirde $\mathrm{Var}(r_i)$, tüm i'ler için sabit olmayacaktır. Studentize artıklar, student - t gibi dağılım gösterdiklerinden aykırı değer testi için kullanılabildikleri gibi varsayım bozulumlarını incelemek için artık grafiklerinde de kullanılırlar (Ellenberg, 1970; Weisberg, 1980, p.105).

t_i ölçütü: Aykırı değeri test etmede kullanılan bir ölçüttür. Aykırılığından kuşku duyulan i. gözlem çıkartıldıktan sonra bu gözlem için, $\tilde{y}_i = \underline{x}_i' \,\hat{\underline{\beta}}(i)$ iken,

$$\mathrm{Var}(\tilde{y}_i) = \sigma^2(i)\ \underline{x}_i'\ (\underline{X}(i)'\underline{X}(i))^{-1}\underline{x}_i \qquad (2.22)$$

yazılır. y_i bir aykırı değer ise $E(y_i-\tilde{y}_i) \neq 0$ olacaktır ve $\mathrm{Var}(y_i-\tilde{y}_i) = \sigma^2\ (1+\underline{x}_i'(\underline{X}(i)'\ \underline{X}(i))^{-1}\ \underline{x}_i)$ bulunacaktır. $E(y_i-\tilde{y}_i) = 0$ hipotezinin student-t testi için ölçüt,

$$t_i = \frac{y_i - \tilde{y}_i}{\hat{\sigma}(i)\sqrt{1+\underline{x}_i'(\underline{X}(i)'\underline{X}(i))^{-1}\underline{x}_i}} \qquad (2.23)$$

biçiminde verilip, aykırı değer testi için özel olarak hazırlanmış $t(\alpha/n,k+1)$ çizelge değerleri ile karşılaştırılır. (2.23) eşitliğinde,

$$\hat{\sigma}^2(i) = \frac{1}{n-k-2}\hat{\sigma}^2(n-k-1-r_i^2) \qquad (2.24)$$

ifadesi yerine konarak,

$$t_i = r_i\left(\frac{n-k-2}{n-k-1-r_i^2}\right)^{1/2} \qquad (2.25)$$

biçiminde yazılır (Weisberg, 1980, p.115).

Bu ölçüt başka yazarlarca F dağılımlı olarak da elde edilmiştir (Besley, et.al., 1980, p. 115).

2.4.2.b. Cook uzaklığı

Cook (1977,1979) tarafından önerilen bu ölçüt, $k'=k+1$ iken,

$$D_i = \frac{(\hat{\underline{\beta}} - \hat{\underline{\beta}}(i))'\ \underline{X}'\underline{X}(\hat{\underline{\beta}} - \hat{\underline{\beta}}(i))}{k'\ \hat{\sigma}^2} \qquad (2.26)$$

biçiminde verilir; veri kümesinden bir gözlem çıkartıldığında regresyon katsayısındaki değişimin bir ölçüsüdür. Böylece regresyon katsayısını etkilemede hangi gözlemin en etkin olduğu görülebilir. Ölçüt,

$$D_i = \frac{r_i^2 w_i}{k'} \qquad (2.27)$$

biçiminde de yazılabilir Burada $w_i = v_{ii}/(1-v_{ii})$ ve $0 \leqslant w_i \leqslant \infty$ dur. r_i^2 ya da w_i büyük ise D_i değeri de büyüktür. r_i^2 bileşeni i. gözlemin aykırılığını test ederken, w_i oranı, $\hat{\beta}$'nın belirlenmesinde i. gözlemin ağırlığını gösterir. Böylece Cook uzaklığı, y değerlerinin oluşturduğu uzay ile x_j'lerin oluşturduğu uzayda aykırı olan gözlemlerin birleşik etkisinin bir ölçüsüdür (Gunst and Mason, 1980, p.256; Cook,1979). Bu ölçütün $F_{(k',n-k',1-\alpha)}$ kritik değeri ile karşılaştırılması, i. gözlemin çıkartılmasının $\hat{\beta}$ kestirimini tüm veriye dayanan % (1-α) güven bölgesinin kenarına taşıması ile sonuçlanacağını gösterir. $D_i > 1$ durumu β kestirimini %50 güven bölgesinin kenarına taşımaya karşılık gelir ve etkin gözlem olarak yorumlanır (Cook, 1977, 1979).

Cook ve Weisberg (1980), Cook ölçütünü genelleştirmişler ve ayrıntılı çıkarsamalar vermişlerdir.

Atkinson, Cook ölçütünde bazı değişiklikler yaparak değiştirilmiş Cook uzaklığı adı verilen,

$$T_i = \left(\frac{n-k}{k} \; \frac{v_{ii}}{1-v_{ii}} \right)^{1/2} |t_i| \qquad (2.28)$$

ölçütünü elde etmiştir. Büyük T_i değeri etkin gözlemin göstergesidir (Atkinson, 1981).

2.4.2.c. AP ölçütü

Nokta kümelerinin X'lerin oluşturduğu uzayda uzak olup olmadığını belirlemek için Andrews ve Pregibon, v_{ii}'nin genellemesi olan AP ölçütünü önermişlerdir (Draper and John, 1981). Bir gözlem veri kümesinde aykırı ise, bu gözlemin veri kümesinden çıkartılması veri kümesi hacmini küçültecektir. Bu düşünceye dayanarak Andrews ve Pregibon,

$$R_{ij\ldots}^{(\ell)}(\underline{X}^{\star}) = \frac{D_{ij\ldots}^{(\ell)} |\underline{X}^{\star\prime}\underline{X}^{\star}|}{|\underline{X}^{\star\prime}\underline{X}^{\star}|} \tag{2.29}$$

oranının küçük değerlerinin aykırı gözlemi göstereceğini belirtmişlerdir. Burada $\underline{X}^{\star} = (\underline{X},\underline{Y})$ eklemeli matrisini ve $D_{ij\ldots}$ simgesi, veri kümesinden çıkartılan i'inci, j'inci gibi ℓ sayıda gözlemi göstermektedir.

AP ölçütü genel olarak etkin gözleme değil, $\underline{X}$'lerin uzayında uzak olan gözlemlere ağırlık verir.

Draper ve John, bu ölçütü,

$$R_{ij\ldots}^{(\ell)} = (1 - Q_{\ell}/AKT)\, |\underline{I}-\underline{R}_{22}| \tag{2.30}$$

biçiminde iki bileşenin çarpımı olarak yorumlamışlardır. Burada Q_{ℓ}, $E\begin{pmatrix}\underline{Y}_1\\ \underline{Y}_2\end{pmatrix} = \begin{pmatrix}\underline{X}_1 & \emptyset\\ \underline{X}_2 & \underline{I}\end{pmatrix}\begin{pmatrix}\beta\\ \gamma\end{pmatrix}$ modelinde γ'nin kestirimi için "ekstra" kareler toplamı olup,

$$Q_{\ell} = \underline{e}_2'\,(\underline{I}-\underline{R}_{22})^{-1}\underline{e}_2 \tag{2.31}$$

biçiminde verilir. $\underline{R}_{22}$ ise,

$$\underline{R}_{22} = \underline{X}_2\,(\underline{X}'\underline{X})^{-1}\underline{X}_2' \tag{2.32}$$

olarak tanımlanır.

(2.30) eşitliğindeki Q_{ℓ} 'nin büyüklüğü aykırı değer kümesini belirtir. İkinci bileşenin küçük değerleri uç değerlere karşılık gelir. Ancak yazarlar ölçütün, uç değerlerin etkin olup olmadığını göstermediğini belirtmişlerdir (Draper and John, 1981).

Draper ve John (1981), tek bir gözlemin aykırılığını test etmek için de kullanılabilen AP ölçütünü Cook ölçütü ile karşılaştırmışlar ve farklı sonuçlar verebildiklerini örneklerle göstermişlerdir.

Tatlıdil, (2.29) eşitliğinden yararlanarak, AKT(i), veri kümesinden i. gözlem çıkartıldıktan sonra ki artık kareler toplamı (AKT) olmak üzere,

$$u = \frac{AKT(i)}{AKT} \tag{2.33}$$

biçiminde, veri kümesinden gözlem çıkartmanın AKT üzerindeki etkisinden yararlanarak aykırılığı test eden bir ölçüt önermiştir. Yazar, bu ölçütü, ℓ tane gözlemin aykırılığının testi için genelleştirerek incelemiş ve $(n-k-\ell)/2$ ve $\ell/2$ serbestlik dereceleri ile beta dağılımlı olduğunu göstererek, ölçüte ilişkin çizelge değerlerini ℓ'nin çeşitli değerleri için vermiştir (Tatlıdil, 1981).

2.4.2.d. DFBETA ölçütü

i. gözlemin veri kümesinden çıkartılmasının katsayı kestirimleri üzerindeki etkisini gösteren bu ölçüt,

$$\begin{aligned} DFBETA_i &= \underline{\hat{\beta}} - \underline{\hat{\beta}}(i) \\ &= \frac{(\underline{X}'\underline{X})^{-1}\,\underline{x}_i'\,e_i}{1 - v_{ii}} \end{aligned} \tag{2.34}$$

biçiminde verilir. Belsley ve diğerleri, i. gözlemi çıkartmanın $\underline{\hat{\beta}}$ vektörünün b_j bileşeninde yarattığı değişimi, $\underline{C} = (\underline{X}'\underline{X})^{-1}X'$ olmak üzere,

$$b_j - b_j(i) = \frac{c_{ji}e_i}{1-v_{ii}} \tag{2.35}$$

biçiminde göstermişlerdir. Ancak yazarlar, b_j'deki değişimi, $\underline{C}'\underline{C} = (\underline{X}'\underline{X})^{-1}$ iken,

$$Var(b_j) = \hat{\sigma}^2 \sum_{r=1}^{n} C_{jr}^2 \tag{2.36}$$

yardımıyla DFBETA'nın ölçeklenmiş biçimi olarak,

$$\begin{aligned} DFBETAS_{ij} &= \frac{b_j - b_j(i)}{\hat{\sigma}(i)\sqrt{(\underline{X}'\underline{X})^{-1}_{jj}}} \\ &= \frac{C_{ji}}{\sqrt{\sum_{r=1}^{n} C_{jr}^2}} \cdot \frac{e_i}{\hat{\sigma}(i)(1-v_{ii})} \end{aligned} \tag{2.37}$$

eşitliğini elde etmişlerdir. Bu ölçüt için kritik değer olarak $2/\sqrt{n}$ değeri önerilmiştir. $|DFBETAS| > 2/\sqrt{n}$ ise ilgili gözlemin, j. değişkeninin katsayı kestirimi üzerinde etkin olduğu söylenir (Belsley, et.al., 1980, p.13-14).

2.4.2.e. DFFIT ölçütü

Belsley ve diğerleri veri kümesinden gözlem çıkartmanın önkestirim üzerindeki etkisinin bir ölçüsü olarak, $\hat{y}_i(i)$, i. gözlem çıktıktan sonraki kestirim değeri olmak üzere,

$$DFFIT = \hat{y}_i - \hat{y}_i(i)$$

$$= \underline{x}_i(\hat{\underline{\beta}} - \hat{\underline{\beta}}(i))$$

$$= \frac{v_{ii}}{1-v_{ii}} e_i \qquad (2.38)$$

oranını vermişlerdir.

Ölçüt, $\hat{\sigma}_{\hat{y}_i}$ hatası ile bölünerek,

$$DFFITS_i = \frac{v_{ii}e_i/(1-v_{ii})}{\hat{\sigma}(i)\sqrt{v_{ii}}}$$

$$= \left(\frac{v_{ii}}{1-v_{ii}}\right)^{1/2} \frac{e_i}{\hat{\sigma}(i)\sqrt{1-v_{ii}}} \qquad (2.39)$$

olarak elde edilmiştir.

Bu ölçüt için karşılaştırma kriteri $2\sqrt{k/n}$ olup, $|DFFITS| > 2\sqrt{k/n}$ ise ilgili gözlemin etkin olduğu yorumu yapılır (Belsley, et.al., 1980, p.15-16).

Weisberg, bu ölçütü,

$$\frac{1}{k'}(DFFITS_i)^2 = \frac{1}{k'\hat{\sigma}^2(i)}(\hat{\underline{\beta}}(i) - \hat{\underline{\beta}})'(\underline{X}'\underline{X})(\hat{\underline{\beta}}(i) - \hat{\underline{\beta}}) \quad (2.40)$$

olarak vermiş ve Cook ölçütüne çok benzer olduğunu ancak herbirinin ayrı yorumları bulunduğunu vurgulamıştır (Weisberg, 1983).

2.4.2.f. COVRATIO ve FRATIO ölçütleri

Veri kümesinden bir gözlem çıkartmanın, regresyon katsayılarının kovaryans matrisi $V(b_j) = \sigma^2(\underline{X}'\underline{X})^{-1}_{jj}$ üzerindeki etkisinin bir ölçüsü COVRATIO ölçütü, önkestirim varyansı üzerindeki etkisinin bir ölçüsü de FRATIO ölçütüdür.

COVRATIO ölçütü: Belsley ve diğerleri, bir gözlemin veri kümesinden çıkartılması ile $\hat{\sigma}^2$ kestiriminin de değişeceğini düşünerek,

$$\text{COVRATIO} \equiv \frac{|\hat{\sigma}^2(i)(\underline{X}(i)'\underline{X}(i))^{-1}|}{|\hat{\sigma}^2(\underline{X}'\underline{X})^{-1}|}$$

$$= \frac{\hat{\sigma}^{2k}(i)}{\hat{\sigma}^{2k}} \; \frac{|(\underline{X}(i)'\underline{X}(i))^{-1}|}{|(\underline{X}'\underline{X})^{-1}|} \qquad (2.41)$$

oranını önermişlerdir. $\hat{\sigma}^2(i)$ ve $\underline{X}(i)$ önceki kesimlerde tanımlandığı gibidir. Burada,

$$\frac{|(\underline{X}(i)'\underline{X}(i))^{-1}|}{|(\underline{X}'\underline{X})^{-1}|} = \frac{1}{1-v_{ii}} \qquad (2.42)$$

eşitliğinden yararlanarak yazarlar,

$$\text{COVRATIO} = \frac{1}{\left(\frac{n-k-1}{n-k} + \frac{t_i^2}{n-k}\right)^k (1-v_{ii})} \qquad (2.43)$$

ölçütünü elde etmişler; bu ölçütün, r_i ve v_{ii}'nin belirleyici özelliklerini birleştirdiğini ve maskeleme sorununun çözümünde yardımcı olduğunu belirtmişlerdir.
$|\text{COVRATIO}-1| \geq 3k/n$ ise ilgili gözlemin etkin olduğu söylenir (Belsley, et.al.,1980, p.22-23).

FRATIO ölçütü: Belsley ve diğerleri bu ölçütü,

$$Var(\hat{y}_i) = \hat{\sigma}^2 v_{ii} \qquad (2.44)$$

$$\begin{aligned} Var(\hat{y}_i(i)) &= Var(\underline{X}'\hat{\beta}(i)) \\ &= \hat{\sigma}^2(i) \frac{v_{ii}}{1-v_{ii}} \end{aligned} \qquad (2.45)$$

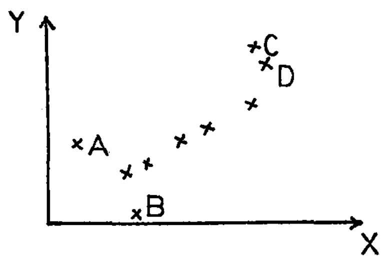
eşitliklerinden yararlanarak, bu iki varyansın birbirine oranından,

$$FRATIO = \frac{\hat{\sigma}^2(i)}{\hat{\sigma}^2(1-v_{ii})} \qquad (2.46)$$

olarak elde etmişlerdir. Etkin gözlem için bu ölçüt değeri oldukça küçük olacaktır (Belsley, et.al.,1980, p.24).

2.4.2.g. birden çok etkin gözlem olması durumundaki ölçütler

Bazı veri kümelerinde gözlemler tek başına değil de bir alt küme şeklinde etkin olabilir. Bu durum Şekil 2.1'de incelenebilir (Cook and Weisberg, 1980):



Şekil 2.1

Burada A ve B gözlemleri tek tek etkin olup birlikte etkin değilken, C ve D gözlemleri tek tek değilde birlikte etkindirler.

Gözlemlerin bir alt kümesinin etkin olduğu durumda buraya kadar anlatılan ölçütler yetersiz kalacaktır. Gözlemlerin biribirinin etkisini maskelemesi söz konusu olduğunda, bu etki incelenen bu ölçütlerle ortaya çıkartılamıyacağı için araştırıcıyı yanlış sonuçlara götürecektir. Bu durumda etkin alt kümeyi belirleyebilecek ve maskeleme sorununu ortaya çıkartacak başka yöntemler önerilmiştir. Bu yöntemler ayrıntılı olarak Belsley ve arkadaşları (1980,p.31-39) ve Tatlıdil (1981) tarafından incelenmiştir.

ÜÇÜNCÜ BÖLÜM

DOĞRUSAL REGRESYON ÇÖZÜMLEMESİNDE GÖZLEMLERİN KÜMELEMEDEN GELEBİLECEK ETKİLERİ ÜZERİNE BİR BENZEŞİM ÇALIŞMASI

Çalışmada ki temel amaç, doğrusal regresyonda kümelenmiş verilerden yapılan örneklemede bir kümeden tek bir gözlem alınmasının ölçütler, kestirim ve önkestirim üzerinde etkinlik yaratıp yaratmayacağının incelenmesi idi.

Bu bölümde, yukarıdaki amaca yönelik olarak, özellikle küçük örneklemler düşünülerek çeşitli etkenlere göre değişimleri incelemek için bir benzeşim çalışması verilecektir. Çalışmada, aykırı ya da uç değer olsun olmasın etkin gözlem üzerinde durulacaktır. Çalışma boyunca Kesim 2.3'de değinilen artıklar arası korelasyonların anlamlı olmadıkları; verilerde, etkin gözlem dışında, herhangi bir varsayım bozulumu bulunmadığı varsayılacaktır.

Benzeşim uygulaması, kullanılacak veriler ve bunların kümelenmesi ile çalışmada ele alıncak ölçütler ve bilgisayar programının tanıtımından sonra verilecektir. Bulgulara esas olan çizelgeler üzerinden yalnızca amaca yönelik yorumlar alınacaktır.

Bu konuda bir örnek sunmak amacı ile, son kesimde, gerçek verilerle bir uygulama yapılacaktır.

3.1. Çalışmada Kullanılan Veriler ve Kümeleme

Benzeşim çalışması için kullanılacak bağımsız değişken değerleri, yapay olarak türetilme yerine, gerçek verilerden alındı. Doğrusal regrasyon sonuçları doğrudan veri yapısına bağlı olduğundan, çalışma bulgularının gerçek bir yapı ile ilişkilendirilmesi amacı ile bu veriler kullanıldı. Veriler, Farmasötik Kimya'nın sık sık başvurduğu Hansch ve diğerleri (1973) tarafından verilmiş olan 236 bileşiğe karşılık gelen değişken değerleridir. Ancak, bu çalışmada kabul edilen evren, bu bileşikler içinde tüm değişken değerleri mevcut olan ve karşılıkları Hansch ve Unger'in (1973) kümeleme çalışmasında bulunabilen 40 gözlemden oluşturuldu ve çalışmaya 6 bağımsız değişken alındı.

Öte yandan, küme (cluster), istatistiksel evrenin yakın öğelerinden oluşan grup; kümeleme (clustering) ise gözlemlerar arası gruplandırma olarak tanımlanmaktadır (Çetinel, 1982). Kümeleme çözümlemesinde, değişkenlere göre birbirine yakın gözlemlerin olabildiğince aynı kümede toplanması istenilir.

Çalışma amacı için veriler, Çetinel'in (1982), verdiği bilgisayar programından yararlanarak, aşamasıralı kümeleme yöntemlerinden ortalama (centroid) tekniği kullanılarak kümelendirildiler. Hansch ve Unger'in (1973) 90 gözlemlik kümeleme ile ilgili çalışmasına parelel olması için, veriler üzerinde küme sayısının 5 ve 10 olduğu iki ayrı kümeleme yapıldı. Bu kümelemeler daha sonra ki model seçiminde kullanıldı. Yapılan kümeleme sonuçları, Hansch ve Unger'in (1973) vermiş oldukları kümelemeye oldukça benzerdir. Yalnızca çalışmaya katılan 10., 22., 23., ve 32. gözlemler yazarların kümelemesinden farklı kümelerde yer almışlardır. 5'lik ve 10'luk kümeler Çizelge 3.1'de verildi. Bu kümelerden yapılan örnekleme Kesim 3.5'de anlatıldı.

Çizelge 3.1 : Verilerin Kümelenmesi

	1.K	2.K	3.K	4.K	5.K	6.K.	7.K	8.K	9.K	10.K
10 Küme	1,2,4,5, 6,9,11, 12,13,14, 15,16,17, 18,19,20, 21,24	3, 7	8	10, 39,	40	22, 23	25, 26, 27	28	29, 30, 31	32,33, 34,35, 36,37,38
5 Küme	1,2,3,4,5 6,7,9,11, 12,13,14, 15,16,17, 18,19,20, 21,24,28	8	10,40, 32,33, 34,35, 36,37, 38,39	22,23, 25,26, 27	29,30, 31					

3.2. Çalışmada Kullanılacak Etkinlik Ölçütleri

Gözlem etkilerinin incelenmesinde kullanılabilecek, etkin gözlemlerin belirlenmesine ilişkin ölçütler Bölüm 2'de verildi. Çalışmaya bunlardan, D_i, DFFITS, t_i ve u_i ölçütleri alınacaktır.

Cook uzaklığı, etkin gözlem araştırmasında temel olması, X'lerin ve y'lerin uzayında aykırı olan gözlemlerin bileşik etkisini vermesi ve pratikte en sık başvurulan ölçüt olması nedeni ile çalışmaya alınacaktır (Cook, 1979).

DFFITS ölçütü, Cook ölçütüne yakın olmakla birlikte Weisberg (1983), bu ölçütün etkinlik üzerine farklı yorumlamaları olabileceğini belirtmiştir. Bu nedenle çalışmaya seçilecek bir diğer ölçüttür.

Etkin gözlemin, uç değer ya da aykırılıktan gelebileceği düşüncesi ile aykırı değer testi için kullanılan ölçütlerden u_i ve t_i ölçütleri de bu tür testlere örnek olarak alınacaktır. Bölüm 2'de incelenen diğer ölçütlerden AP ölçütü, Draper ve John (1981) ile Tatlıdil'in (1981) belirttikleri gibi, etkin gözleme değil, X'lerin uzayında uzak olan gözlemlere ağırlık verdiği için; COVRATIO ölçütü, çalışma amacına yardımcı olmadığından; FRATIO ölçütü, ölçüt değerinin önemli olup olmadığını belirleyebilecek kesin bir kriteri bulunmadığı için; DFBETAS ölçütü ise, Cook ölçütüne benzer olması nedeniyle çalışmaya alınmadı.

3.3. Bilgisayar İzlencesi

Başlangıçta bilgisayar programı gerçek uygulamalara yönelik olarak yapıldı. Bu programda, özellikle etkinliğinden kuşku duyulan tek gözlemin incelenmesine dayalı, Kesim 2.4.1'de verilen tüm grafiksel tekniklerle Kesim 2.4.2'de verilen istatistiksel ölçütler yer aldı. Grafik çiziminde Sparks'ın (1971) vermiş olduğu programdan uyarlama yapıldı. GÖZET adlı program BASIC programlama dilinde yazıldı.

Benzeşim çalışması için, yukarıdaki programdan yararlanılarak, SİMGÖZ adlı ek program yazıldı. Bu program ve genel akış şeması ile gerekli açıklamalar Ek II'de verildi.

Benzeşim çalışması sonuçları AMSTRAD CPC6128 bilgisayarından alındı.

3.4. Benzeşim Çalışması

3.4.1. etkenler

Çalışmada, sonuçların hangi durumlara göre değiştiğini görmek üzere aşağıdaki etkenler alındı:

(1) örneklem genişliği (N): Her bir model için, bütün kümelerden seçilen gözlemlerin sayısıdır. Gerçek uygulamalarda sık sık küçük örneklemler üzerinde çalışıldığından burada N'nin değerleri 9,10,12,15 ve 18 olarak alındı. Bunların herbiri, aynı zamanda, daha sonra tanımlanacak olan modellerin ilk gözlem sayılarıdır.

(2) Kümeleme: Çalışma için, Kesim 3.1'de belirtildiği gibi veriler üzerinde iki ayrı kümeleme yapıldı. Çizelge 3.1'de görüldüğü gibi küme sayısı fazla iken bir kümede tek kalabilen bir gözlem, daha az sayıda küme olduğunda öteki gözlemlerle birleştirmektedir. Bu etken alınırken, bir kümeden tek bir gözlem alınmasının regresyon sonuçlarını kümelenme biçiminde etkileyebileceği düşünüldü. Çalışmada, kümeleme tekniklerinde yapılacak değişikliklerin, sonuçları büyük oranda değiştirmeyeceği varsayılacaktır.

(3) Son kümeden alınan gözlem sayısı (NI): Çalışmada, bir küme dışında, N'ye bağlı olarak öteki kümelerden, en az üçer gözlem alınması düşünüldü. Ancak "son küme" olarak tanımlanan kümeden önce bir, daha sonra iki ve üç gözlem alınarak bu kümeden tek alınan gözlemin etkinliği karşılaştırılmak istendi. Böylece NI = 1,2,3 alındı. Son küme, kümeleme sonucunda yalnızca üç gözlem içeren ve 10'luk kümede 9'uncu, 5'lik kümede ise 5'inci kümedir. Bu seçimi yaparken, çalışmada kümelerdeki değişimin etkisi olmadığı düşünüldü.

3.4.2. karşılaştırma ölçütleri

Benzeşim sonuçlarını karşılaştırmak üzere, çalışmaya alınan ölçütlerin, son kümeden alınan tek gözleme ilişkin

1000 iterasyon üzerinden ortalamaları ve standart sapmaları elde edildi. Bu ortalamalardan bazıları grafikler üzerinde incelenecektir.

Etkin gözlemlerin kestirim ve önkestirim değerlerini etkileyebileceği Kesim 2.3'de belirtilmişti. Bu nedenle, gerçek parametreler bilindiğinden, çalışmada kestirimlere ilişkin toplam yanılgı kareler ortalamalarının (varyans + yan^2) ele alınması düşünüldü ve üç ayrı yoldan hesaplandı :

i) Katsayılara ilişkin Yanılgı Kareler Ortalaması: Kestirimlerin gerçek katsayılara uzaklıklarının toplam bir ölçüsü olarak,

$$\text{Beta YKO} = \frac{(\underline{\beta} - \hat{\underline{\beta}})' (\underline{\beta} - \hat{\underline{\beta}})}{\sigma^2} \qquad (3.1)$$

eşitliğinden hesaplandı (Hocking 1976).

ii) Tüm regresyon modeli için toplam önkestirim yanılgı kareler ortalaması (Top.ÖYKO): Bir modelin toplamsal önkestirim gücünü görmek üzere,

$$\text{Top.ÖYKO} = \frac{(\underline{\beta} - \hat{\underline{\beta}})' \underline{X}' \underline{X} (\underline{\beta} - \hat{\underline{\beta}})}{\sigma^2} \qquad (3.2)$$

eşitliğinden yararlanıldı. Eşitlik $\hat{y} = \underline{X}\hat{\underline{\beta}}$ ile $E(\hat{y}) = \underline{X}\underline{\beta}$ arasındaki uzaklığın karesel bir ölçüsüdür (Hocking, 1976).

iii) Dışarıdan gözlemlerin toplam önkestirim yanılgı kareler ortalaması: Etkin gözlemin veri kümesinde bulunmayan ve elde edilen regresyon modelinden önkestirilmek istenen gözlemler üzerindeki etkilerini görmek amacı ile bu tür gözlemler için de önkestirim yanılgı kareler ortalaması bulundu. Veri kümesi dışında 6 gözlem alınarak, bunların herbiri için, dışarıdan alınan i. gözlem vektörü $\tilde{\underline{X}}$ olmak üzere, $\tilde{y}_i = \tilde{\underline{x}}_i' \hat{\underline{\beta}}$ önkestirim değeri ile $E(\tilde{y}_i)$ değeri elde edildi. Toplam önkestirim için,

$$\text{ÖYKO} = \sum_{\ell=1}^{6} V(\tilde{y}_\ell) + \sum_{\ell=1}^{6} [E(\tilde{y}_\ell) - \mu_{\tilde{y}_\ell}]^2 \qquad (3.3)$$

eşitliği kullanıldı. $\mu_{\tilde{y}} = \underline{\tilde{X}}'\underline{\beta}$'dır. $\tilde{X}$, dışarıdan alınan $\underline{X}_j$ değişkenler matrisidir. Eşitlikteki ikinci terim yan^2'yi belirtir. Herbir model için, i ve ii'den bulunan YKO'ların 1000 iterasyon üzerinden ortalamaları alındı. Karşılaştırma sağlayabilmek için ii'den bulanan ortalamalar son kümeden seçilen gözlemlerin sayıları dikkate alınarak N ile standartlaştırıldı.

Bu ölçütlerin yanısıra, etkin gözlemlerin etkilerini ençok belli etkileri ve modellerin geçerlilik ölçütü olarak da başvuralan Press ölçütü, karşılaştırmada 1000 iterasyon üzerinden ortalaması alınıp, N ile standartlaştırılarak elde edildi. Press için,

$$\text{Press} = \sum_{i=1}^{N} \left(\frac{e_i}{1 - v_{ii}} \right)^2 \qquad (3.4)$$

eşitliği kullanıldı (Hocking, 1976).

3.5. Model Oluşturma ve Uygulama

Çalışmanın başlangıcında Kesim 3.1'de değinilen 40 gözlem üzerinden doğrusal regresyon incelemesi yapıldı ve model varsayımlarının sağlandığı görüldü. Ancak bu incelemede, tek başına bir kümede yer alan 8. gözlem etkin bir gözlem olarak bulundu. Bunun yanında 23. gözlemin de bu incelemede az da olsa etkin sayılabilecek bir gözlem olduğu görüldü. Çalışmada bu sonuçlar dikkate alındı.

Çalışmada, birinci etkenin herbir düzeyi için ikinci ve üçüncü etken düzeylerini birarada içeren 4 çalışma modeli oluşturuldu. Bu modellerin herbiri için gözlemler 5 kümeden örneklenmektedir. 1.çalışma modeli, verilerin gerek 5 gerekse 10 kümede kümelendiği durumların her ikisinde de aynı kümede yer alan gözlemlerden, 2.çalışma modeli ise 10'luk kümelemede farklı kümelerde yer alan gözlemleri de örnekleme alarak oluşturuldu. Her iki model için de üçüncü etken olarak son kümeden önce bir, sonra iki ve üç gözlem alınarak örneklemler tamamlandı.

Son kümeden alınan tek gözlemin etkisini sınamak için, son küme olarak, tüm veri kümesinde gerçekte etkin sayılabilecek ve etkinliği uç değer olmasından kaynaklanan 23. gözlemi içeren, çalışma modeli 1 ve 2'dekinden farklı bir küme alındı. Böylece 3. ve 4. çalışma modelleri, 1. ve 2. çalışma modellerine parelel olarak oluşturuldu.

Bu yolla 4 çalışma modelini oluşturma işlemi, herbir örneklem genişliği için tekrarlandı. Herbir çalışma modeli de, son kümeden alınan gözlem sayısına bağlı olarak kendi içerisinde 3 model içermektedir.

Çalışmada ayrıca, son kümeden tek olarak alınan gözlemin etkinliğini, normal çalışma modelleri ile karşılaştırmak için, iki tür denetim modeli oluşturuldu.

İlk denetim modelinde son küme, ilk öğesi birinci kümenin elemanlarından, diğer iki öğesi ise kendi elemanlarından olacak şekilde seçildi. Böylece son kümeden alınan tek gözlemin farklılığı nedeni ile bu gözlemin etkinliğinin çalışma modelleri ile karşılaştırılabileceği düşünüldü.

İkinci denetim modelinde ise son küme, tamamen birinci kümenin elemanlarından oluşan bir küme olarak alındı. Bu kümeden alınan tek gözlemin bir etkinlik yaratmayacağı ve dolayısı ile çalışma modelleri ile karşılaştırılabileceği düşünüldü. Böylece sonuçta herbiri 3 ayrı model içeren 4 model için bir çalışma ve iki denetim olmak üzere 3 regresyon modeli, herbir örneklem genişliği için denendi. Yani $3 \times 4 \times 3 \times 5 = 180$ benzeşim modeli incelendi. Herbir benzeşim modelinin içerdiği gözlemleri gösteren çizelgeler Ek 1'de verilmektedir.

Öte yandan regresyon modelleri yaratılırken birinci aşamada, β_j ve σ^2 gerçek değerleri için, verilerin gerçek alınması nedeni ile önceki bir çalışmada tam küme modelinden elde edilen $\hat{\beta}_j$ ve $\hat{\sigma}^2$ değerleri kullanıldı (Özden, vd., baskıda). Bu değerler yaklaşık olarak,

$$\underline{\beta}' = (0.36 \quad 0.2 \quad -0.03 \quad -0.22 \quad -0.20 \quad -0.01 \quad 0.005)$$

ve $\sigma^2 = 0.01$ olarak alındı.

İkinci aşamada, (2.1) modelini oluşturmak üzere $\varepsilon_i \sim N(0,\sigma^2)$, $i=1,..,n$ değerleri merkezi limit teoremine göre türetildi (Halaç, 1982, s.130-132).

Böylece, türetilen ε_i'ler yardımıyla,

$$Y_i = \beta_o + \beta_i x_{i1} + \ldots + \beta_6 x_{i6} + \varepsilon_i, \quad i=1,\ldots,n \tag{3.5}$$

eşitliğinden Y_i değerleri elde edilerek, (2.1) eşitliğinin kestirimleri ile (2.2) artıkları ve buradan çalışma için gerekli ölçütler hesaplandı. Bu işlemler, herbir benzeşim modeli için 1000'er kez tekrarlandı.

3.6. Bulgular

3.6.1. v_{ii} üzerine irdeleme

Herbir benzeşim modeli için iterasyona girmeden önce elde edilen, son kümeden tek alınan gözleme ilişkin v_{ii} değerleri Çizelge 3.2'de verildi.

Bu çizelge incelendiğinde, son kümeden tek olarak alınan gözlemin denetim modellerine göre v_{ii} üzerinde etkili olduğu görülebilir. Son kümeden alınan gözlem sayısı arttıkça değerler küçülmekte; denetim modelinin değerlerine benzer duruma gelmektedir. Son kümenin tüm veri kümesinde etkin sayılabilecek 23. gözlemi içeren küme olarak alındığı 3. ve 4. çalışma modelleri için, bu kümeden tek alınan gözlemin v_{ii} üzerindeki etkisi denetim modellerine göre, diğer çalışma modellerinden daha büyüktür. $N=9$ durumunda 1. ve 3. çalışma modelleri için görülen istisnai durumlar küçük örneklem genişliğine bağlanabilir. Böylece denebilir ki, kümelemenin v_{ii} ölçütü üzerinde arttırıcı yönde bir etkisi vardır ve örneklemdeki gözlem sayısı arttıkça bu etki küçülmektedir. Ancak örnekleme alınan diğer kümelerde bulunan bir etkin gözlem, kümelemenin ölçüt üzerindeki etkisini daha da arttırmakta ve gözlem sayısında ki artış bu etkiyi azaltamamaktadır.

	N=9			N=10			N=12			N=15			N=18		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
MODEL 1	0.99	0.87	0.86	0.95	0.75	0.66	0.90	0.68	0.61	0.82	0.60	0.54	0.70	0.52	0.46
DENETİM A	0.99	0.86	0.83	0.92	0.91	0.88	0.86	0.77	0.69	0.60	0.53	0.50	0.50	0.49	0.45
DENETİM B	0.99	0.86	0.48	0.92	0.47	0.30	0.86	0.35	0.34	0.60	0.34	0.27	0.50	0.30	0.25
MODEL 2	0.98	0.78	0.75	0.95	0.58	0.44	0.89	0.57	0.41	0.83	0.51	0.38	0.67	0.41	0.34
DENETİM A	0.89	0.72	0.59	0.64	0.44	0.40	0.68	0.52	0.47	0.53	0.39	0.36	0.30	0.29	0.28
DENETİM B	0.89	0.38	0.38	0.64	0.33	0.28	0.68	0.29	0.29	0.53	0.24	0.23	0.30	0.18	0.18
MODEL 3	0.99	0.96	0.95	0.98	0.92	0.80	0.97	0.93	0.84	0.93	0.86	0.73	0.89	0.80	0.69
DENETİM A	0.99	0.96	0.85	0.70	0.70	0.61	0.76	0.76	0.69	0.56	0.55	0.51	0.35	0.31	0.25
DENETİM B	0.99	0.66	0.48	0.70	0.59	0.51	0.76	0.39	0.35	0.56	0.36	0.32	0.35	0.30	0.26
MODEL 4	0.97	0.91	0.91	0.99	0.93	0.83	0.98	0.92	0.87	0.93	0.83	0.73	0.83	0.78	0.71
DENETİM A	0.89	0.89	0.80	0.57	0.55	0.52	0.61	0.59	0.58	0.42	0.41	0.37	0.34	0.34	0.32
DENETİM B	0.89	0.38	0.38	0.57	0.31	0.29	0.61	0.28	0.24	0.42	0.30	0.29	0.34	0.21	0.21

v_{ii}'ler N'ye bağlı olup, etkinlikleri uç değerden kaynaklanır, böylece bu çalışmada etkin gözlem uç değer olarak ortaya çıkmaktadır.

3.6.2. çalışmaya alınan ölçütler için bulgular

Son kümeden tek alınan gözlemin, 1000 ilterasyon üzerinden Cook ölçütü için ortalama ve sapmaları, DFFITS ölçütü içinde ortalamaları Çizelge 3.3 ve 3.4'de verildi.

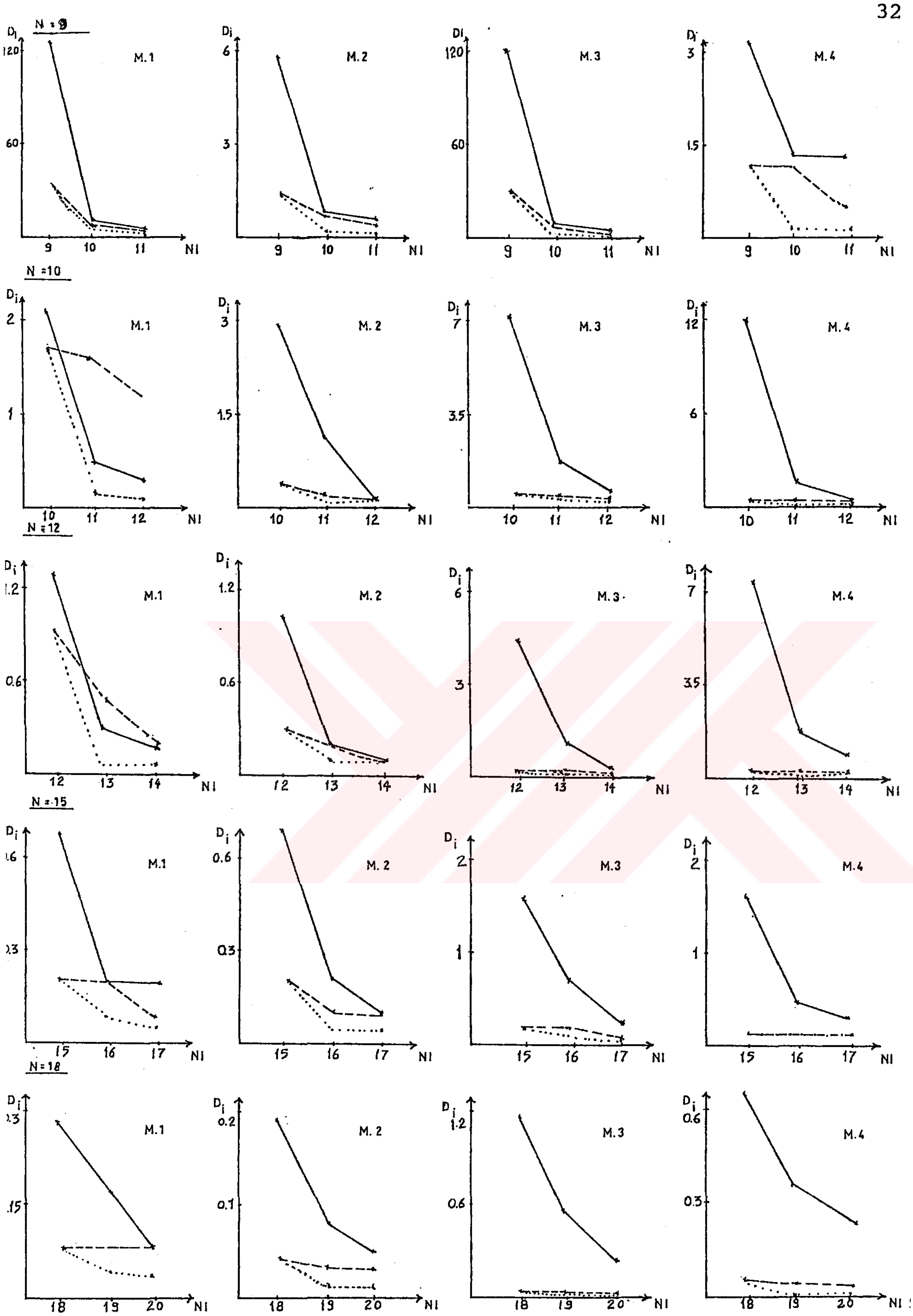
Son kümeden tek ve daha fazla gözlem almanın Cook üzerindeki etkileri denetim modellerine göre herbir benzeşim modeli için, ortalamalar dikkate alınarak Şekil 3.1'de gösterildi. Şekil 3.1'e göre, son kümeden tek alınan gözlemin ortalama Cook uzaklığı denetim modellerine göre çok daha büyük elde edilmekte; bu kümeden alınan gözlem sayısı arttıkça ölçüt değeri denetim modellerinin değerine yaklaşmaktadır. Buna göre, kümelemenin Cook ölçütü üzerinde büyültücü yönde etkisi olduğu ve bu etkinin örneklem genişliğinin büyük olması durumunda da değişmediği söylenebilir.

Çizelge 3.4'de DFFITS ölçütü incelendiğinde bulguların Cook ölçütünde elde edilen bulgulara benzer olduğu görülür. Ancak kümelemenin DFFITS ölçütünü özellikle küçük örneklemlerde büyültücü yönde etkilediği, örneklem genişliği büyüdükçe bu etkinin küçüldüğü, ayrıca 3. ve 4. çalışma modellerinde son kümeden tek alınan gözlemin ölçüt üzerinde etkisinin daha büyük olması nedeni ile diğer kümelerde bulunan etkin gözlemin, ölçüt üzerindeki kümeleme etkisini daha da arttırdığı söylenebilir.

Çizelge 3.5'de son kümeden tek alınan gözleme ilişkin ortalama t değerleri, Çizelge 3.6'da ise ortalama u değerleri verildi. Bu çizelgeler incelendiğinde son kümeden tek alınan gözlemin t ve u ölçütleri üzerindeki etkisinin denetim modeline göre farklı olmadığı görülür. Küçük örneklemlerde son kümeden alınan gözlem sayısının artması u ölçüt değerini büyütmekte, örneklem genişliği büyüdükçe bu etki azalmaktadır. Öte yandan büyük örneklemlerde t ölçüt değeri küçülmektedir. Böylece kümelemenin t_i ve u_i ölçüt-ortalama Cook değerleri ve standart sapmaları(*)

	N=9 (1)	N=9 (2)	N=9 (3)	N=10 (1)	N=10 (2)	N=10 (3)	N=12 (1)	N=12 (2)	N=12 (3)	N=15 (1)	N=15 (2)	N=15 (3)	N=18 (1)	N=18 (2)	N=18 (3)
MODEL 1	122.60 (86.11)	0.94 (0.85)	0.92 (0.89	2.30 (1.11)	0.40 (0.43)	0.27 (0.29)	2.30 (1.33)	0.30 (0.31)	0.20 (0.15)	0.70 (0.79)	0.20 (0.25)	0.20 (0.21)	0.30 (0.43)	0.20 (0.20)	0.10 (0.15)
DENETİM A	32.20 (21.9)	0.93 (0.81)	0.73 (0.69)	1.60 (1.47)	1.50 (1.45)	1.07 (1.18)	0.90 (0.90)	0.50 (0.53)	0.30 (0.35)	0.20 (0.27)	0.20 (0.19)	0.10 (0.18)	0.10 (0.18)	0.10 (0.17)	0.10 (0.17)
DENETİM B	32.20 (21.9)	0.28 (0.25)	0.12 (0.12)	1.60 (1.47)	1.13 (0.13)	0.06 (0.07)	0.90 (0.90)	0.08 (0.09)	0.08 (0.09)	0.20 (0.27)	0.08 (0.09)	0.06 (0.06)	0.10 (0.18)	0.06 (0.08)	0.05 (0.06)
MODEL 2	5.50 (4.25)	0.50 (0.45)	0.40 (0.43)	2.80 (2.54)	1.07 (0.19)	0.10 (0.12)	1.03 (1.15)	0.20 (0.20)	0.09 (0.11)	0.70 (0.85)	0.20 (0.19)	0.10 (0.11)	0.29 (0.22)	0.09 (0.12)	0.07 (0.09)
DENETİM A	1.20 (0.84)	0.37 (0.33)	0.25 (0.25)	0.30 (0.23)	0.11 (0.11)	0.10 (0.10)	0.30 (0.33)	0.15 (0.18)	0.11 (0.13)	0.20 (0.19)	0.10 (0.12)	0.10 (0.09)	0.06 (0.07)	0.05 (0.07)	0.05 (0.07)
DENETİM B	1.20 (0.84)	0.09 (0.08)	0.10 (0.09)	0.30 (0.23)	0.07 (0.07)	0.06 (0.06)	0.30 (0.33)	0.06 (0.06)	0.06 (0.07)	0.20 (0.19)	0.05 (0.06)	0.05 (0.05)	0.06 (0.07)	0.03 (0.04)	0.03 (0.04)
MODEL 3	120.30 (81.79)	3.30 (2.76)	2.50 (2.55)	7.10 (6.40)	1.70 (1.73)	0.50 (0.60)	5.20 (5.57)	1.60 (1.85)	1.62 (0.70)	1.80 (2.25)	0.80 (1.05)	0.30 (0.39)	1.20 (1.50)	0.60 (0.71)	0.30 (0.37)
DENETİM A	32.20 (21.99)	3.20 (2.75)	0.80 (0.81)	0.30 (0.30)	0.30 (0.30)	0.20 (0.23)	0.40 (0.43)	0.40 (0.49)	0.30 (0.33)	0.20 (0.23)	0.20 (0.20)	0.10 (0.18)	0.08 (0.09)	0.07 (0.08)	0.05 (0.06)
DENETİM B	32.20 (21.9)	0.28 (0.25)	0.12 (0.12)	0.30 (0.30)	0.19 (0.20)	0.14 (0.15)	0.40 (0.43)	0.10 (0.11)	0.08 (0.09)	0.20 (0.23)	0.08 (0.09)	0.07 (0.08)	0.08 (0.09)	0.06 (0.08)	0.05 (0.06)
MODEL 4	3.80 (2.88)	1.50 (1.38)	1.50 (1.47)	11.60 (10.67)	1.80 (1.77)	0.64 (0.73)	7.40 (7.85)	1.70 (1.92)	0.86 (1.03)	1.90 (2.34)	0.60 (0.75)	0.40 (0.17)	0.70 (0.98)	0.40 (0.54)	0.30 (0.39)
DENETİM A	1.20 (0.84)	1.15 (1.00)	0.59 (1.11)	0.18 (1.17)	0.18 (0.17)	0.16 (0.17)	0.20 (0.23)	0.20 (0.23)	0.19 (0.22)	0.10 (0.13)	0.10 (0.12)	0.10 (0.11)	0.08 (0.09)	0.07 (0.09)	0.07 (0.09)
DENETİM B	1.20 (0.84)	0.09 (0.08)	0.10 (0.09)	0.18 (1.17)	0.06 (0.07)	0.06 (0.06)	0.20 (0.23)	0.06 (0.06)	0.05 (0.05)	0.10 (0.13)	0.10 (0.07)	0.10 (0.08)	0.08 (0.09)	0.04 (0.05)	0.04 (0.05)

(*): Standart sapma.



Şekil 3.1. Son kümeden tek alınan gözlemin gözlem sayısına göre ortalama Cook değerleri.

NI: Son kümeden alınan gözlem sayısına göre örneklem genişliği.

N : Model.

(——) : Çalışma Modeli

(----) : Denetim A

(....) : Denetim B

DFFITS ölçütü değerleri

		N=9			N=10			N=12			N=15			N=18		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
MODEL	1	5.90	0.48	0.31	1.33	0.35	0.24	0.80	0.15	0.16	0.30	0.19	0.20	0.21	0.21	0.03
DENETİM	A	2.09	0.49	0.40	0.37	0.37	0.07	0.79	0.53	0.22	0.12	0.06	0.05	0.01	0.04	0.04
DENETİM	B	2.09	0.64	0.05	0.37	0.09	0.00	0.79	0.02	0.02	0.12	0.07	0.03	0.01	0.02	0.04
MODEL	2	1.63	0.16	0.00	1.36	0.27	0.21	1.10	0.25	0.19	0.45	0.23	0.14	0.13	0.12	0.12
DENETİM	A	0.06	0.05	0.04	0.02	0.17	0.16	0.24	0.21	0.14	0.25	0.09	0.08	0.12	0.08	0.02
DENETİM	B	0.06	0.06	0.05	0.02	0.01	0.00	0.24	0.03	0.03	0.25	0.06	0.05	0.12	0.03	0.01
MODEL	3	7.23	0.86	0.79	3.96	1.28	0.45	1.47	0.03	0.02	1.00	0.60	0.29	0.99	0.66	0.36
DENETİM	A	2.09	1.63	0.38	0.23	0.14	0.07	0.17	0.16	0.15	0.04	0.05	0.04	0.03	0.02	0.02
DENETİM	B	2.09	0.64	0.05	0.23	0.30	0.16	0.17	0.11	0.05	0.04	0.01	0.02	0.03	0.08	0.01
MODEL	4	0.90	0.69	0.70	0.80	0.59	0.25	0.77	0.53	0.51	0.72	0.50	0.09	0.57	0.49	0.31
DENETİM	A	0.06	0.05	0.04	0.18	0.12	0.02	0.03	0.02	0.03	0.08	0.04	0.09	0.02	0.00	0.01
DENETİM	B	0.06	0.10	0.09	0.18	0.03	0.02	0.03	0.02	0.02	0.08	0.01	0.00	0.02	0.01	0.01

		N=9			N=10			N=12			N=15			N=18		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
MODEL	1	2.20	1.50	1.13	1.57	1.16	0.99	1.02	0.92	0.93	0.94	0.91	0.89	0.89	0.89	0.82
DENETİM	A	2.22	1.42	1.07	1.48	1.13	0.99	0.98	0.98	0.89	0.93	0.89	0.89	0.88	0.86	0.85
DENETİM	B	2.22	1.37	1.01	1.48	1.18	1.07	0.98	0.98	0.96	0.93	0.91	0.90	0.88	0.88	0.88
MODEL	2	2.32	1.35	1.11	1.47	1.07	1.00	0.93	0.91	0.92	0.91	0.89	0.90	0.86	0.84	0.83
DENETİM	A	2.37	1.33	1.14	1.42	1.03	1.04	1.01	0.92	0.88	0.90	0.91	0.88	0.85	0.83	0.83
DENETİM	B	2.37	1.40	1.18	1.42	1.12	1.09	1.01	0.95	0.95	0.90	0.90	0.91	0.85	0.84	0.84
MODEL	3	2.20	1.55	1.13	1.42	1.14	0.95	0.96	0.78	0.79	0.87	0.87	0.82	0.89	0.82	0.79
DENETİM	A	2.22	1.50	1.08	1.37	1.04	1.05	0.91	0.92	0.92	0.95	0.88	0.85	0.87	0.86	0.86
DENETİM	B	2.22	1.37	1.01	1.42	1.05	0.97	0.91	0.92	0.90	0.95	0.91	0.86	0.87	0.86	0.86
MODEL	4	2.22	1.42	1.02	1.28	1.04	0.93	0.98	0.95	0.84	0.90	0.84	0.83	0.81	0.78	0.98
DENETİM	A	2.37	1.33	1.14	1.34	1.10	1.03	0.98	0.98	0.88	0.94	0.91	0.91	0.89	0.86	0.86
DENETİM	B	2.37	1.40	1.80	1.34	1.13	1.04	0.98	0.93	0.93	0.94	0.89	0.89	0.89	0.89	0.85

		N=9			N=10			N=12			N=15			N=18		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
MODEL	1	0.52	0.66	0.75	0.65	0.74	0.80	0.79	0.84	0.86	0.87	0.88	0.89	0.90	0.91	0.93
DENETİM	A	0.52	0.66	0.75	0.66	0.74	0.80	0.80	0.83	0.86	0.87	0.89	0.89	0.91	0.92	0.92
DENETİM	B	0.52	0.66	0.77	0.66	0.74	0.79	0.80	0.82	0.85	0.87	0.88	0.89	0.91	0.91	0.92
MODEL	2	0.54	0.68	0.75	0.67	0.76	0.79	0.82	0.84	0.86	0.87	0.89	0.89	0.91	0.92	0.93
DENETİM	A	0.52	0.67	0.75	0.66	0.76	0.79	0.79	0.84	0.87	0.88	0.88	0.90	0.91	0.92	0.92
DENETİM	B	0.42	0.67	0.73	0.66	0.74	0.78	0.79	0.83	0.85	0.88	0.88	0.89	0.91	0.92	0.92
MODEL	3	0.52	0.64	0.75	0.66	0.75	0.81	0.81	0.87	0.88	0.88	0.89	0.91	0.90	0.92	0.93
DENETİM	A	0.52	0.64	0.75	0.68	0.77	0.79	0.82	0.84	0.86	0.86	0.89	0.90	0.91	0.91	0.92
DENETİM	B	0.52	0.66	0.77	0.67	0.77	0.81	0.82	0.82	0.85	0.86	0.89	0.89	0.91	0.92	0.92
MODEL	4	0.55	0.67	0.74	0.68	0.76	0.82	0.80	0.83	0.87	0.87	0.89	0.91	0.92	0.93	0.94
DENETİM	A	0.52	0.67	0.75	0.67	0.75	0.79	0.80	0.83	0.86	0.87	0.88	0.89	0.91	0.92	0.92
DENETİM	B	0.52	0.67	0.73	0.67	0.74	0.79	0.80	0.84	0.85	0.87	0.89	0.89	0.81	0.91	0.92

leri üzerinde önemli bir etki yaratmadığı söylenebilir.

Çizelge 3.7'de verilen, son kümeden alınan gözlem sayısına göre herbir benzeşim modeli için toplam Press ölçütü değerlerinden yararlanarak Şekil 3.2 oluşturuldu.

Şekil 3.2 incelendiğinde, son kümeden tek gözlem alınmasının Press ölçüt üzerindeki etkisinin denetim modeline göre daha büyük olduğu, özellikle küçük örneklemlerde çok büyük olan bu etkinin örneklem genişliği (N) arttıkça küçüldüğü görülebilir. Ayrıca 3. ve 4. çalışma modellerinde bu etkinin diğer çalışma modellerine göre daha büyük olduğu söylenebilir. Son kümeden alınan gözlem sayısı arttıkça denetim modeline yakın ölçüt değerleri elde edilmektedir. Böylece kümelemenin Press ölçütü üzerinde, özellikle küçük örneklemlerde büyültücü etkisi olduğu, örneklem genişliği büyüdükçe bu etkinin küçüldüğü söylenebilir. Verilerde diğer kümelerde bulunan etkin sayılabilecek bir gözlem, kümelemenin bu ölçüt üzerindeki etkisini daha da arttırmaktadır.

3.6.3. toplam yanılgı kareler ortalamaları için bulgular

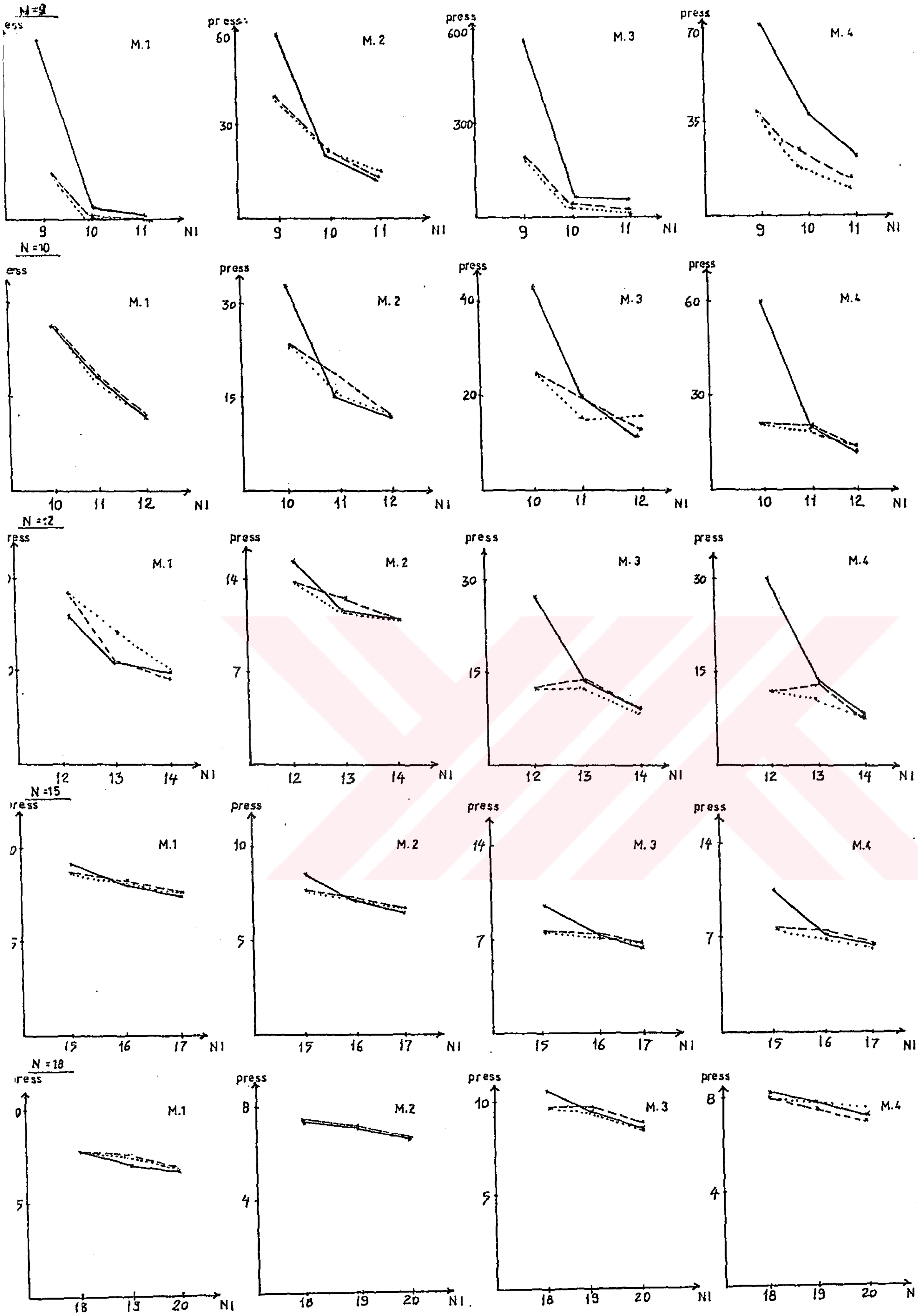
Son kümeden alınan gözlem sayısına göre her benzeşim modeli için 1000 iterasyon üzerinden ortalama Bet YKO ölçütü değerleri Çizelge 3.8'de verildi. Son kümeden tek ve daha fazla gözlem almanın Bet YKO ölçütü üzerindeki etkileri denetim modellerine göre her benzeşim modeli için ortalamaları alınarak Şekil 3.3'de gösterildi.

Şekil 3.3'e göre, benzeşim modelleri için bu ölçüte ilişkin tutarlı yorumlar yapılamamaktadır. Bu bulgulardan Bet YKO ölçütü için beklenildiği gibi yeterli bir sonuç elde edilememektedir.

3.6.4. toplam önkestirim yanılgı kareler ortalaması ve dışardan gözlemlerin yanılgı kareler ortalamalarına ilişkin bulgular

Son kümeden alınan gözlem sayısına göre herbir benzeşim

	N=9			N=10			N=12			N=15			N=18		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
MODEL 1	545.6	30.9	21.1	26.7	19.3	13.7	16.6	12.6	11.7	10.4	9.2	8.6	8.2	7.6	7.2
DENETİM A	176.6	26.3	16.0	27.2	21.3	15.2	18.5	12.9	10.9	9.9	9.6	8.7	8.0	7.6	7.4
DENETİM B	176.6	21.3	19.4	27.2	19.3	15.5	18.5	15.9	12.6	9.9	9.3	8.9	8.0	7.7	7.3
MODEL 2	56.5	20.4	14.7	37.2	16.2	14.7	14.3	11.0	10.5	10.0	8.6	8.3	8.3	8.1	7.7
DENETİM A	34.8	23.7	15.1	26.6	20.6	15.1	13.0	11.8	10.4	9.2	8.9	8.2	8.4	8.2	7.7
DENETİM B	34.8	20.8	16.7	26.6	18.5	14.8	13.0	11.1	10.2	9.2	8.8	8.4	8.4	8.2	7.7
MODEL 3	526.2	42.8	40.7	37.0	17.5	12.7	25.6	13.5	10.3	12.2	9.7	8.2	9.6	8.3	7.5
DENETİM A	176.6	37.0	21.6	20.6	17.5	13.4	12.7	14.1	10.3	9.4	9.6	8.5	8.3	8.3	7.5
DENETİM B	176.6	21.3	19.4	20.6	13.8	14.7	12.7	11.2	9.8	9.4	8.9	8.8	8.3	8.4	7.6
MODEL 4	74.9	41.2	22.6	58.0	19.8	12.7	34.7	15.9	11.4	13.1	9.7	8.5	9.2	8.4	7.7
DENETİM A	34.8	22.9	16.9	22.0	21.4	13.9	13.9	15.3	10.5	9.9	9.8	8.5	8.7	8.2	7.6
DENETİM B	34.8	20.8	16.7	22.0	19.2	14.1	13.9	12.4	10.8	9.9	9.2	8.4	8.7	8.3	7.9

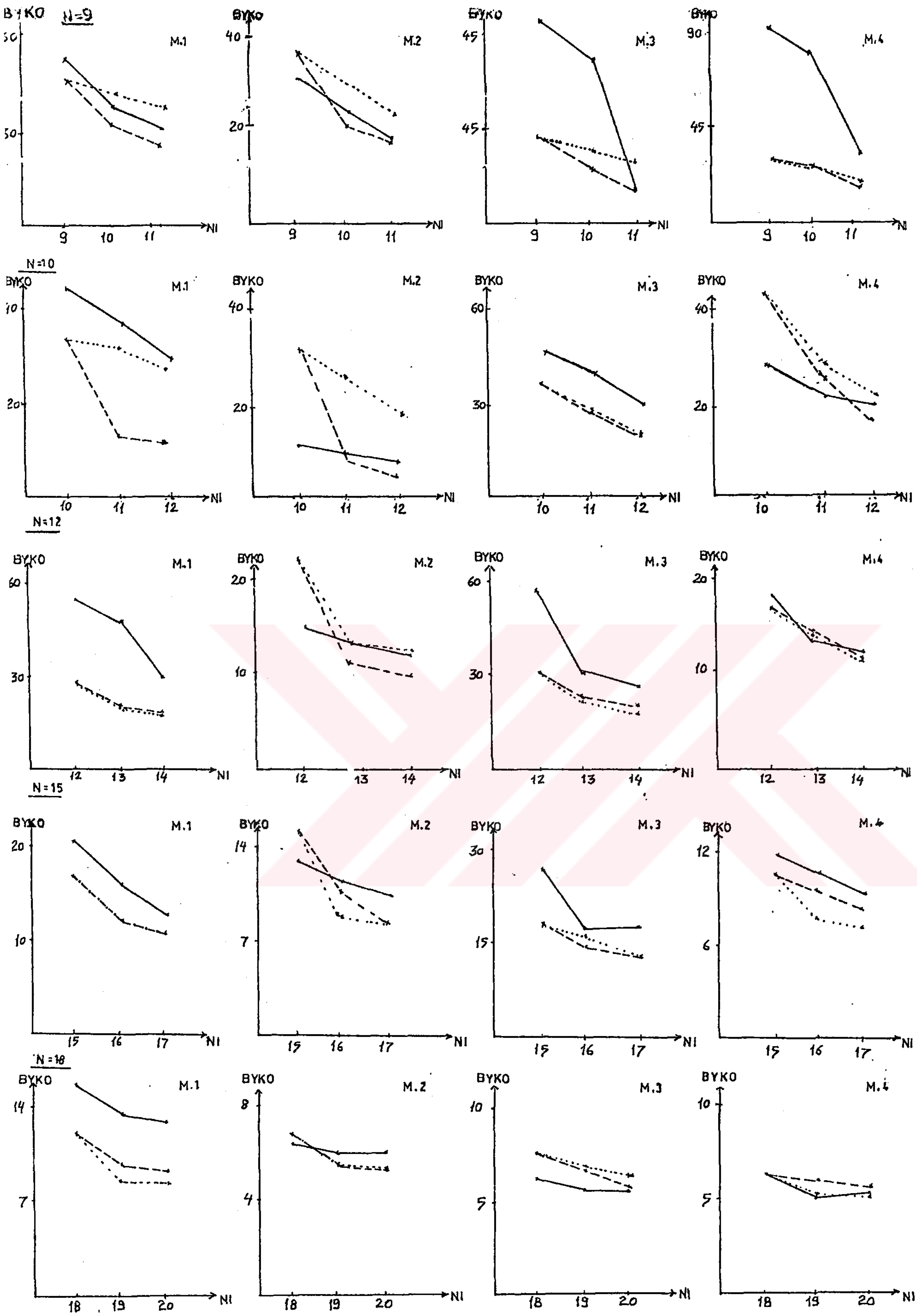


ekil 3.2. Son kümeden alınan gözlem sayısına göre toplam Press ölçütü değerleri.

I: Son kümeden alınan gözlem sayısına göre örneklem genişliği.
N: Model.

(———) : Çalışma Modeli
(----) : Denetim A
(....) : Denetim B

		N=9			N=10			N=12			N=15			N=18		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
MODEL	1	51.2	39.1	31.2	44.5	38.9	31.1	48.9	42.7	26.3	20.5	16.4	12.6	14.9	12.9	12.3
DENETİM	A	46.7	31.9	26.8	33.6	13.8	13.0	24.5	17.8	15.9	16.2	11.6	10.4	11.2	9.0	8.6
DENETİM	B	46.7	41.9	38.2	33.6	32.0	28.9	24.5	17.1	14.2	16.2	11.7	10.8	11.2	8.1	8.1
MODEL	2	31.4	26.5	21.5	11.9	10.0	8.8	11.5	10.8	9.7	10.6	9.6	8.6	6.9	6.4	6.4
DENETİM	A	37.9	20.8	18.6	30.8	9.1	7.5	15.5	8.6	7.6	13.8	8.6	7.3	7.3	5.9	5.8
DENETİM	B	37.9	31.5	25.2	30.8	24.6	20.8	15.5	11.4	10.8	13.8	7.7	7.3	7.3	5.9	5.7
MODEL	3	95.8	86.9	26.9	46.6	30.9	31.1	58.6	32.9	30.1	25.5	18.6	19.2	7.1	6.4	6.4
DENETİM	A	46.7	37.6	22.3	34.8	26.4	22.1	32.5	25.6	21.3	18.7	14.4	13.0	8.5	7.5	6.8
DENETİM	B	46.7	41.9	38.2	34.8	29.2	21.8	32.5	24.4	20.0	18.7	16.2	12.5	8.5	7.9	7.5
MODEL	4	91.3	88.4	40.1	29.5	24.0	22.6	19.1	15.4	13.9	12.0	10.8	9.8	7.4	5.9	6.3
DENETİM	A	37.9	36.1	23.6	42.1	28.2	18.9	18.6	16.4	13.1	10.6	9.7	8.5	7.1	6.8	6.4
DENETİM	B	37.9	31.5	25.2	42.1	31.6	25.3	18.6	16.7	12.8	10.6	7.9	7.5	7.1	6.3	6.1



*Şekil 3.3. Son kümeden alınan gözlem sayısına göre ortalama Bet YKO değerleri.

*NI: Son kümeden alınan gözlem sayısına göre örneklem genişliği.
N: Model.

(———) : Çalışma Modeli
(----) : Denetim A
(....) : Denetim B

modeli için toplam ÖYKO değerleri Çizelge 3.9'da, dışardan gözlemlerin ÖYKO değerleri ise Çizelge 3.10'da verildi.

Çizelge 3.9 incelendiğinde, son kümeden tek gözlem alınmasının denetim modeline göre çalışma modelinin ÖYKO'sunu değiştirmediği görülür. Böylece kümelemenin toplam ÖYKO üzerinde herhangi bir etkisi olmadığı söylenebilir.

Çizelge 3.10'a göre, son kümeden tek alınan gözlem denetim modeline göre dışarıdan alınan gözlemlerin ÖYKO'sunu değiştirmemektedir.

Son kümeden alınan gözlemlere göre her benzeşim modeli için dışardan gözlemlerin toplam yan değerleri Çizelge 3.11'de verildi. Buna göre son kümeden alınan tek gözlemin denetim modeline göre dışardan gözlemlerin yan'ını 1. ve 2. çalışma modelleri için etkilemediği, 3. ve 4. çalışma modelleri için ise bir miktar arttırıcı yönde etkinin söz konusu olduğu söylenebilir.

Böylece kümeleme, dışarıdan alınan gözlemlerin ÖYKO'sunu değiştirmemekte ancak eğer verilerde etkin sayılabilecek bir başka gözlem varsa, önkestirme yan getirmektedir. Bu da, verilerde etkin gözlem varlığında kümelemenin, önkestirim varyansını küçültüp, yan'ı büyülterek ÖYKO'yu değiştirmediği şeklinde yorumlanabilir.

Elde edilen bulgulara göre sonuçta, kümelemenin u ve t ölçütleri dışında çalışmaya alınan D_i, DFFITS, v_{ii} ve Press ölçütlerini etkilerken, modelin gerek ÖYKO'sunu gerekse dışardan alınan gözlemlere ilişkin ÖYKO'yu etkilemediği, Bet YKO için ise bulguların yeterli bilgi vermediği söylenebilir.

3.7. Gerçek Verilerle Uygulama

Bu kesimde çalışmaya açıklık getirmek için gerçek veriler üzerinde bir uygulama yapılacak ve etkin gözlemin çalışmaya alınan ölçütler üzerindeki etkisi araştırılacaktır. Veriler, uygulamanın küçük örneklem üzerinde olması amacı

	N=9			N=10			N=12			N=15			N=18		
	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
MODEL 1	0.085	0.082	0.079	0.80	0.077	0.074	0.074	0.071	0.068	0.066	0.064	0.062	0.060	0.059	0.058
DENETİM A	0.086	0.083	0.078	0.080	0.077	0.074	0.075	0.071	0.069	0.066	0.064	0.063	0.061	0.059	0.058
DENETİM B	0.086	0.083	0.079	0.080	0.077	0.074	0.075	0.071	0.069	0.066	0.064	0.063	0.061	0.059	0.059
MODEL 2	0.086	0.083	0.079	0.081	0.077	0.074	0.074	0.071	0.068	0.065	0.064	0.062	0.061	0.059	0.059
DENETİM A	0.088	0.083	0.079	0.082	0.078	0.074	0.074	0.072	0.069	0.066	0.065	0.063	0.062	0.060	0.059
DENETİM B	0.088	0.083	0.078	0.082	0.078	0.075	0.074	0.072	0.069	0.066	0.065	0.063	0.062	0.060	0.059
MODEL 3	0.085	0.081	0.076	0.081	0.078	0.073	0.075	0.072	0.069	0.067	0.065	0.063	0.061	0.059	0.057
DENETİM A	0.086	0.082	0.077	0.082	0.079	0.074	0.077	0.073	0.070	0.067	0.065	0.062	0.062	0.059	0.058
DENETİM B	0.086	0.083	0.079	0.082	0.079	0.075	0.077	0.073	0.070	0.067	0.065	0.064	0.062	0.059	0.058
MODEL 4	0.087	0.081	0.076	0.082	0.078	0.075	0.074	0.072	0.068	0.066	0.065	0.062	0.061	0.058	0.057
DENETİM A	0.088	0.083	0.078	0.082	0.079	0.074	0.076	0.072	0.069	0.066	0.065	0.063	0.060	0.059	0.057
DENETİM B	0.088	0.083	0.078	0.082	0.079	0.076	0.076	0.072	0.070	0.066	0.064	0.063	0.060	0.060	0.058

toplam ÖYKO değerleri

		N=9			N=10			N=12			N=15			N=18		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
MODEL	1	2.50	2.42	2.05	2.06	1.74	1.67	1.25	1.21	1.12	1.13	1.10	1.08	0.86	0.86	0.85
DENETİM	A	1.96	1.84	1.71	1.31	1.18	1.18	1.10	1.08	1.07	1.09	1.08	1.08	0.85	0.84	0.72
DENETİM	B	1.96	1.92	1.22	1.31	1.31	0.94	1.10	1.08	1.06	1.09	1.07	0.85	0.85	0.84	0.77
MODEL	2	1.31	1.27	1.23	1.05	1.02	1.02	0.94	0.93	0.93	0.95	0.94	0.94	0.81	0.80	0.80
DENETİM	A	1.32	1.25	1.21	1.17	1.02	1.01	0.94	0.92	0.92	0.95	0.91	0.91	0.81	0.80	0.80
DENETİM	B	1.32	1.30	1.23	1.17	1.16	1.00	0.94	0.92	0.91	0.95	0.87	0.78	0.81	0.80	0.75
MODEL	3	3.29	3.23	1.37	1.58	1.45	1.43	1.18	1.08	1.04	1.18	1.10	1.06	0.85	0.83	0.82
DENETİM	A	1.96	1.87	1.51	1.55	1.44	1.42	1.34	1.11	1.03	1.29	1.10	1.02	0.88	0.86	0.84
DENETİM	B	1.96	1.92	1.22	1.55	1.50	0.94	1.34	1.32	1.29	1.29	1.28	0.94	0.88	0.88	0.79
MODEL	4	1.54	1.51	1.07	1.12	1.05	1.00	0.93	0.91	0.88	0.97	0.94	0.91	0.80	0.79	0.77
DENETİM	A	1.32	1.31	1.07	1.39	1.22	1.03	1.06	0.96	0.88	0.99	0.93	0.88	0.81	0.79	0.77
DENETİM	B	1.32	1.30	1.23	1.39	1.34	1.04	1.06	1.04	0.95	0.99	0.93	0.84	0.81	0.80	0.76

toplam yan değerleri

		N=9			N=10			N=12			N=15			N=18		
		(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)	(1)	(2)	(3)
MODEL	1	0.219	0.345	0.568	0.065	0.150	0.184	0.069	0.111	0.063	0.096	0.098	0.114	0.061	0.069	0.084
DENETİM	A	0.207	0.186	0.375	0.169	0.131	0.141	0.068	0.079	0.055	0.132	0.114	0.169	0.085	0.008	0.089
DENETİM	B	0.207	0.248	0.101	0.169	0.234	0.046	0.068	0.085	0.067	0.132	0.109	0.081	0.085	0.090	0.054
MODEL	2	0.022	0.021	0.022	0.083	0.082	0.066	0.095	0.077	0.079	0.102	0.049	0.104	0.034	0.030	0.049
DENETİM	A	0.018	0.031	0.030	0.149	0.076	0.139	0.078	0.071	0.068	0.098	0.063	0.087	0.031	0.032	0.049
DENETİM	B	0.018	0.030	0.032	0.066	0.058	0.097	0.078	0.075	0.064	0.098	0.069	0.049	0.031	0.040	0.028
MODEL	3	0.691	0.619	0.271	0.350	0.340	0.297	0.198	0.087	0.098	0.228	0.144	0.157	0.075	0.078	0.064
DENETİM	A	0.207	0.244	0.448	0.283	0.355	0.401	0.197	0.095	0.100	0.112	0.127	0.181	0.110	0.090	0.114
DENETİM	B	0.207	0.248	0.101	0.283	0.185	0.076	0.197	0.117	0.101	0.112	0.109	0.082	0.110	0.099	0.052
MODEL	4	0.058	0.069	0.039	0.095	0.094	0.063	0.046	0.061	0.032	0.072	0.116	0.102	0.047	0.031	0.034
DENETİM	A	0.018	0.036	0.049	0.050	0.045	0.052	0.011	0.015	0.024	0.041	0.044	0.043	0.040	0.030	0.027
DENETİM	B	0.018	0.030	0.032	0.050	0.100	0.031	0.011	0.023	0.023	0.041	0.043	0.029	0.040	0.019	0.027

ile 6 açıklayıcı değişken ve 11 gözlemli farmasötik kimya çalışmasından alındı (Özden v.d., baskıda).

6 açıklayıcı değişken üzerinden elde edilen regresyon çözümlemesi sonuçları ve artıklarla ilgili ölçütler Çizelge 3.12'de verildi. Şekil 3.4 ve 3.5'de de Kesim 2.4.1.b'de değinilen kısmı artık ve bileşen + artık çizimleri, örnek olması açısından verildi.

Bu sonuçlara göre, 7. gözlem v_{ii}, D_i ve DFFITS ölçütlerini etkilerken u ve t ölçütleri üzerinde büyük bir etki yaratmamaktadır. Böylece, bu gözlemin Cook uzaklığının büyüklüğü nedeni ile etkin gözlem olduğu söylenebilir. Şekil 3.4 incelendiğinde, X_6 değişkeninin etkin olan 7. gözlemin etkisi altında olduğu ve öneminin bu gözlemce belirlendiği görülmektedir. Aynı durum Şekil 3.5'de verilen bileşen + artık çiziminde de gözlenmektedir.

Öte yandan Hansch ve Unger'in (1973) kümelerine bakıldığında, 7. gözlem dışında, kümelemden en az üçer gözlem geldiği görülür. 7. gözleme ilişkin bileşik bu gözlemin bulunduğu kümeden tek olarak alınmıştır. Böylece, benzeşim çalışmasına göre, 7. gözlemin etkinliği uç değer olmasından kaynaklanmakta olup ölçütler ve kestirimler üzerinde etkili iken önkestirimi değiştirmeyeceği yorumu yapılabilir.

Örneklemin küçüklüğü nedeni ile, bu etkin gözlemin veri kümesinden çıkartılması bir çözüm getirmemekte, aksine değişik gözlemleri etkin hale getirmektedir. 7. gözlem çıkartıldıktan sonra yapılan regresyon çözümlemesinden elde edilen sonuçlar Çizelge 3.13'de verildi.

Bu sonuçlara göre, incelenen bu örnek daha geniş bir örneklem olsaydı 7. gözlem yine etkin çıkacak ancak bu gözlemin veri kümesinden çıkartılması olumsuz etkiler yaratmayacaktı denebilir.

Çizelge 3.12. Formaksötik kimya verilerinin regresyon çözümleme sonuçları

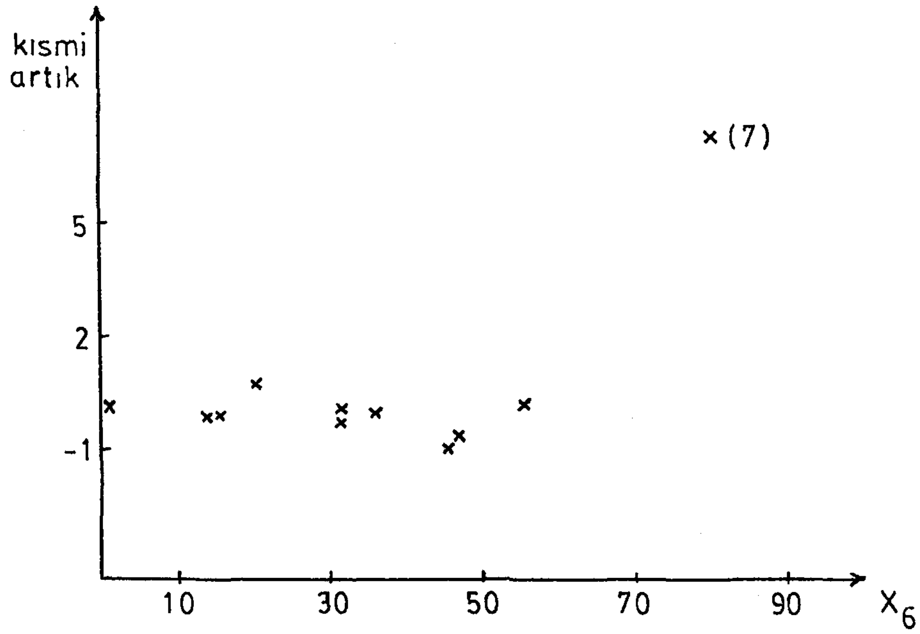
$$\hat{y} = 37.08 + 28.58X_1 - 3.76X_2 - 37.72X_3 - 19.43X_4 - 1.44X_5 + 0.69X_6$$

$\hat{\sigma}^2 = 159.92$ $R^2 = 0.914$

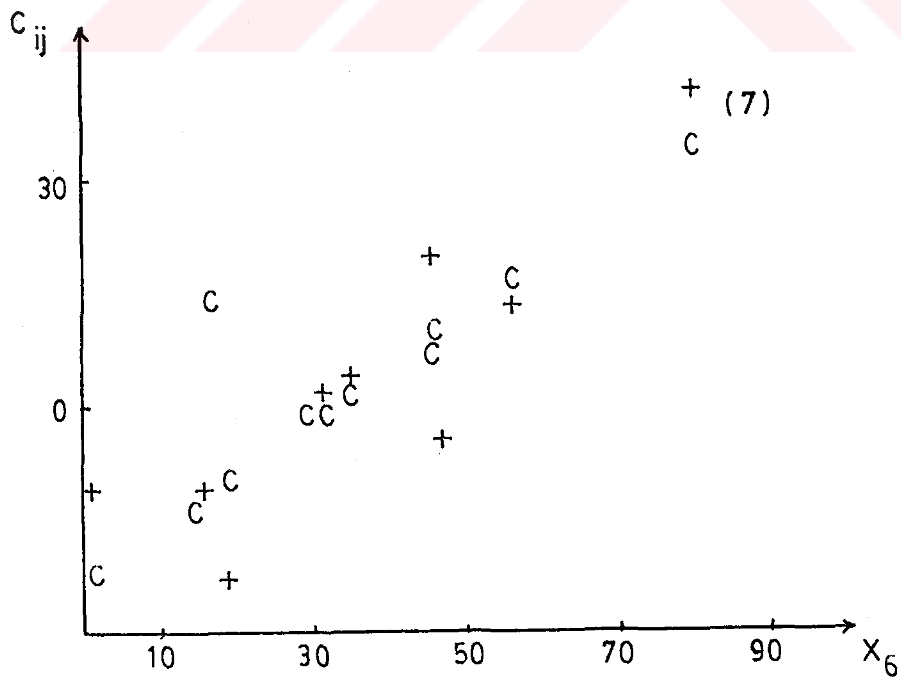
G Ö Z L E M E T K İ L E R İ

I	ARTIK	Vii	Ri	Di	ti	Ui	DFFITS
1	11.19	0.57	1.35	0.35	1.59	0.54	1.84
2	-2.58	0.34	-0.25	0.00	0.22	0.98	-0.16
3	-0.91	0.35	-0.09	0.00	0.08	1.00	-0.06
4	-1.49	0.91	-0.38	0.20	0.34	0.96	-1.05
5	-12.13	0.61	-1.54	0.53	2.08	0.41	-2.61
6	2.68	0.31	0.25	0.00	0.22	0.98	0.15
7	5.93	0.93	1.74	5.48	3.03	0.25	10.81
8	2.08	0.79	0.36	0.07	0.32	0.97	0.62
9	8.54	0.74	1.32	0.71	1.53	0.56	2.57
10	1.98	0.93	0.59	0.67	0.54	0.91	1.96
11	-15.29	0.52	-1.75	0.48	3.13	0.23	-3.28

ENBD= 5.4772 (7) ENBt= 3.1346 (11)



Şekil 3.4. x_6 değişkeni için kısmi artık çizimi.



[1]Şekil 3.5. x_6 değişkeni için bileşen ve bileşen+artık çizimi.

[1]"C": bileşen etkisi, "+": bileşen+artık.

Çizelge 3.13. 7. gözlem çıkarıldıktan sonra ki regresyon çözümleme sonuçları.

$$\hat{y} = 39.78 + 23.66x_1 - 3.29x_2 + 33.79x_3 - 1.07x_4 + 4.97x_5 - 1.31x_6$$

$\hat{\sigma}^2 = 52.46$ $R^2 = 0.971$

G Ö Z L E M E T K İ L E R İ

I	ARTIK	Vii	Ri	Di	ti	Ui	DFFITS
1	3.90	0.68	0.95	0.28	0.93	0.70	1.37
2	-3.68	0.34	-0.63	0.03	0.55	0.87	-0.40
3	-1.57	0.35	-0.27	0.01	0.22	0.98	-0.17
4	-0.63	0.91	-0.28	0.11	0.24	0.97	-0.74
5	-6.64	0.67	-1.60	0.76	3.46	0.14	-4.96
6	8.68	0.38	1.53	0.21	2.64	0.22	2.09
7	1.50	0.79	0.45	0.11	0.38	0.93	0.75
8	-1.78	0.96	-1.23	5.22	1.43	0.49	-7.02
9	0.79	0.93	0.42	0.36	0.36	0.94	1.33
10	-0.57	0.97	-0.47	1.12	0.40	0.93	-2.37

ENBD= 5.2242 (8) ENBt= 3.4556 (5)

DÖRDÜNCÜ BÖLÜM

SONUÇ VE TARTIŞMA

Bu bölüme değin, girişte verilen amaçlar doğrultusunda konu ile ilgili genel bilgiler derlenerek gözlemlerin, kümelemeden kaynaklanabilecek etkilerine ilişkin bulgular incelendi. Çalışmada, kümelenmiş verilerde bir kümeden tek bir gözlem almanın ölçütler, kestirim ve önkestirimler üzerinde etki yaratıp yaratmadığı araştırıldı.

Bir kümeden tek olarak alınan gözlemin,özellikle Cook, v_{ii}, DFFITS ve Press ölçütlerini büyültürken önkestirimler üzerinde anlamlı bir artış yaratmadığı; tek gözleme aynı kümeden başka gözlemler de eklendikçe bu gözlemin ölçütler üzerindeki etkisinin yok olduğu görüldü. Ancak veriler arasında başka kümelerde bulunan bir etkin gözlemin ölçütler ve kestirimler üzerindeki etkisinin kümelemeden bağımsız olarak devam ettiği gözlendi.

Benzeşim uygulamasından elde edilen bulguların geçerliliği gerçek verilerle yapılan uygulamada görüldü.

Sonuç olarak, az sayıda gözlem içeren kümelenebilir veriler üzerinde yapılan çalışmalardan bulunabilecek etkin gözlemin kaynaklarından biri olarak kümeleme önerilebilir. Ayrıca eğer bir gözlemin kümeden tek olarak alınan gözlem olduğuna karar verilmiş ise elde edilen denklemin önkestirim amacı için kullanılabileceği söylenebilir. Bununla birlikte bu tür araştırmalar için her kümeden en az iki gözlemin seçilmesi, kestirim ve ölçütler üzerindeki kümeleme etkilerinin arıtılması yönünden yararlı olacaktır.

Öte yandan çalışmada yapılan bazı zorunlu kısıtlamalar nedeniyle ortaya çıkan bazı eksiklikler ve kuşkular aşağıdaki gibi özetlenebilir:

Kümeleme sonucunda 8. gözlem, tek başına bir kümede yer almıştır. Bu tür bir gözlemin örnekleme sonucu regresyon verisinde yer almasının etkisi burada araştırılmamıştır.

Ancak bu gözlem bir uç değer olarak düşünüldüğünde, etkin gözlemin veri kümesinde bulunması ölçütleri ve kestirimleri etkileyebilecek, önkestirim üzerinde büyük bir etkiye sahip olmayabilecektir. Bunun nedeni, gözlemin farklı bir kümeden gelmesinin veri hacmini büyülterek önkestirim bölgesini genişletmesi; böylece, dışdeğer bulma (extrapolation) tehlikesini azaltmasıdır.

Öte yandan çalışmada tek bir gözlemin etkinliği dikkate alınmıştır. Gözlemlerin bir alt kümesinin etkin olması ve maskeleme etkilerinin ortaya çıkması durumlarında yukarıda elde edilen sonuçların değişebileceği düşünülebilir.

Son olarak çalışmada ki bir başka kısıtlama, etkin gözlem dışında varsayım bozulumu olmaması ve tam küme üzerinden çalışılmasıdır. Çalışmanın hacmini büyük ölçüde arttırması nedeniyle tam küme ve alt kümelerin amacımız doğrultusunda bir karşılaştırması yapılmamıştır. Böyle bir karşılaştırma, gözlemlerin değişkenler üzerindeki etkilerinin varlığı nedeniyle uygulamalarda yararlı olacaktır.

DEĞİNİLEN BELGELER DİZİNİ

Atkinson,A.C.,1981, Two graphical displays for outlying and influential observations in Regression: Biometrica, 68,1,13-20.

Baskerville,J.C. and Toogood, J.H., 1982, Guided Regression modelling for prediction and exploration of structure with many explanatory variables: Technometrics, 24, 1, 9-17.

Beckman,R.J. and Cook, R.D.,1983, Outlier....s: Technometrics, 25,2,120-124, 139-140.

Belsley,D.A., Kuh,E., Welsch,R.E.,1980, Regression diagnostics: Cambridge, Massachusetts, 287 p.

Cook,R.D.,1977, Detection of influential observation in Linear Regression: Technometrics,19,1, 15-18.

Cook,R.D.,1979, Influential observations in Linear Regression: Journal of the American Statistical Association, 74, 365, 169-174.

Cook,R.D. and Weisberg, S.,1980, Characterizations of an emprical influence function for detecting influential cases in Regression: Technometrics, 22,4, 495-508.

Çetinel,B.,1982, Çok değişkenli verilerin kümelendirilmesi için istatistiksel bir yöntem: Doktora tezi, H.Ü.Fen Fakültesi, Beytepe, Ankara, 92 s. (basılmamıştır).

Daniel,C. and Wood, E.S., 1980, Fitting equations to data: Wiley,New York, 447 p.

Draper,N.R. and John,J.A., 1981, Influential observations and outliers in Regression: Technometrics, 23,1, 21-26.

DEĞİNİLEN BELGELER DİZİNİ (devam ediyor)

Draper,N.R. and Smith,H.,1981, Applied Regression Analysis: Wiley, New York, 701 p.

Ellenberg,J.H.,1970, Detection of outliers in multivariate Linear Regression: Ph D thesis, Harvard University, Massachusetts, 121 p.

Gunst,R.F. and Mason,R.L.,1980, Regression Analysis and its application: Marcel Dekkar, New York, 398 p.

Halaç,O.,1982, İşletmelerde simülasyon teknikleri: İstanbul Matbaası, 261 s.

Hansch,C.,Leo,A., Unger,S.H.,Kim,K.H., Nikartani, D., Lien, E.J.,1973, "Aromatic" substituent constant for structure. Activity correlations: Journal of Medical Chemistry, 16,11, 1207-1217.

Hansch,C. and Unger,S.H.,1973, Strategy in drug design. Cluster Analysis as an aid in the selection of substituents, 16,11,1217-1222.

Hocking,R.R.,1976, The analysis and selection of variables in Linear Regression: Biometrics, 32, 1-49.

Hocking,R.R.,1983, Developments in Lineer Regression methodology: 1959-1982: Technometrics, 25,3, 219-249.

Özden,S., Özden, T., Gümüş,F., Erar,A., 1987, Antihistaminik etkili 2-(p-sübstitüefenil)-3H-İmidazo (4,5-c) Pridin türevleri üzerinde kantitatif yapı-etki ilişkisi: Farmasötik Bilimler Dergisi (FABAD), baskıda.

Sparks,D.N.,1971, Algorithm As 44: Journal of Royal Statistical Society, 20,3, 327-331.

DEĞİNİLEN BELGELER DİZİNİ (devam ediyor)

Stevens,J.P.,1984, Outliers and Influential data points in Regression Analysis: Psychological Bulletin, 95,2, 334-344.

Tatlıdil,H.,1981, Doğrusal Regresyonda ve çok değişkenli verilerde kuşkulu gözlemlerin testi: Doktora tezi, H.Ü. Fen Fakültesi, Beytepe, Ankara, 86 s.

Weisberg,S.,1980,Applied Linear Regression: Wiley,New York, 279 p.

Weisberg,S.,1983, Some principles for Regression diagnostics and influence Analysis: Technometrics, 25,3, 240-244.

Wood,F.S., 1973, The use of individual effects and residuals in fitting equations to data: Technometrics, 15,4, 677-695.

E K L E R

EK 1. Benzeşim çalışmasında kullanılan modellerde kümelerden alınan gözlemler

EK 2 . Benzeşim çalışmasında kullanılan bilgisayar izlencesinin işakış şeması ve bazı alt izlenceler

EK-1

BENZEŞİM ÇALIŞMASINDA KULLANILAN MODELLERDE KÜMELERDEN ALINAN GÖZLEMLER

N = 9

Küme	MODEL 1	MODEL 2	MODEL 3	MODEL 4
1	9,11,12,18	3,7,11,18	9,11,12,18	3,7,11,18
2	-	-	-	-
3	32,33,34,36	10,33,34,39	32,33,34,36	10,33,34,39
4	-	-	-	-
5	(29) (30) (31)	(29) (30) (31)	(23) (25) (26)	(23) (25) (26)
[1]Den.A	(13) (30) (31)	(13) (30) (31)	(13) (25) (26)	(13) (25) (26)
Den.B	(13) (2) (4)	(13) (2) (4)	(13) (2) (4)	(13) (2) (4)

N = 10

Küme	MODEL 1	MODEL 2	MODEL 3	MODEL 4
1	9,11,12	3,7,12	9,11,12,18,19,21	3,7,9,12,18,19
2	-	-	-	-
3	32,33,36	10,32,39	32,33,36	10,32,39
4	25,26,27	22,26,27	-	-
5	(29) (30) (31)	(29) (30) (31)	(23) (25) (26)	(23) (25) (26)
Den.A	(13) (30) (31)	(13) (30) (31)	(13) (25) (26)	(13) (25) (26)
Den.B	(13) (2) (4)	(13) (2) (4)	(13) (2) (4)	(13) (2) (4)

[1]Çizgilerin altında verilen Denetim A ve Denetim B modelleri için ilk 4 kümeden alınan gözlemler aynı olup, bu modeller için sadece son kümeden alınan gözlemler verilmiştir.

N =12

Küme	MODEL 1	MODEL 2	MODEL 3	MODEL 4
1	11,12,16,18	3,7,16,18	11,12,14,16,18,21	3,7,12,16,18,21
2	-	-	-	-
3	32,33,34,36	10,32,33,39	32,33,34,36,37	10,32,33,34,39
4	25,26,27	22,25,26	-	-
5	(29)(30)(31)	(29)(30)(31)	(23)(25)(26)	(23)(25)(26)
Den.A	(13)(30)(31)	(13)(30)(31)	(13)(25)(26)	(13)(25)(26)
Den.B	(13) (2)(9)	(13) (2)(9)	(13) (2)(9)	(13) (2)(9)

N=15

Küme	MODEL 1	MODEL 2	MODEL 3	MODEL 4
1	9,11,12,17,18,19	3,7,9,17,18,19	9,11,12,15,17,18, 19,21,24	3,7,9,11,17,18 19,21,28
2	-	-	-	-
3	32,33,34,36,37	10,32,33,37,39	32,33,34,36,37	10,32,33,37,39
4	25,26,27	22,25,26	-	-
5	(29)(30)(31)	(29)(30)(31)	(23)(25)(26)	(23)(25)(26)
Den.A	(13)(30)(31)	(13)(30)(31)	(13)(25)(26)	(13)(25)(26)
Den.B	(13) (2)(4)	(13) (2)(4)	(13) (2)(4)	(13) (2)(4)

N =18

Küme	MODEL 1	MODEL 2	MODEL 3	MODEL 4
1	9,11,12,15,17 19,20,21	3,7,11,12,17, 19,21,28	1,6,9,11,12 15,17,18,19,20,21	3,6,7,11,12,15, 17,18,19,21,28
2	-	-	-	-
3	32,33,34,36 37,38	10,32,33,34, 38,39	32,33,34,36,37,38	10,32,33,34,38, 39
4	25,26,27	22,25,27	-	-
5	(29)(30)(31)	(29)(30)(31)	(23)(25)(26)	(23)(25)(26)
Den.A.	(13)(30)(31)	(13)(30)(31)	(13)(25)(26)	(13)(25)(26)
Den.B	(13)(2)(4)	(13) (2)(4)	(13) (2) (4)	(13) (2)(4)

EK-2

BENZEŞİM ÇALIŞMASINDA KULLANILAN BİLGİSAYAR İZLENCESİNİN İŞAKIŞ ŞEMASI VE BAZI ALT İZLENCELER

İZLENCE ADLARI VE AÇIKLAMALAR :

N : Örneklem genişliği.

NI : Son kümeden alınan gözlem sayısına göre örneklem genişliği. NI $\geqslant$ N olup, son kümeden 1,2, ve 3 gözlem alındığında sıra ile NI = N, NI = N+1 ve NI = N+2 değerini alır.

KK : Herbir benzeşim modeli için, çalışma ve iki denetim modeline ilişkin örneklerin alınmasını kontrol eden değişken. KK = 1 iken çalışma modelinin gözlemleri, KK = 2 ve KK = 3 için ise aynı modele ilişkin denetim modellerinin gözlemleri okunur.

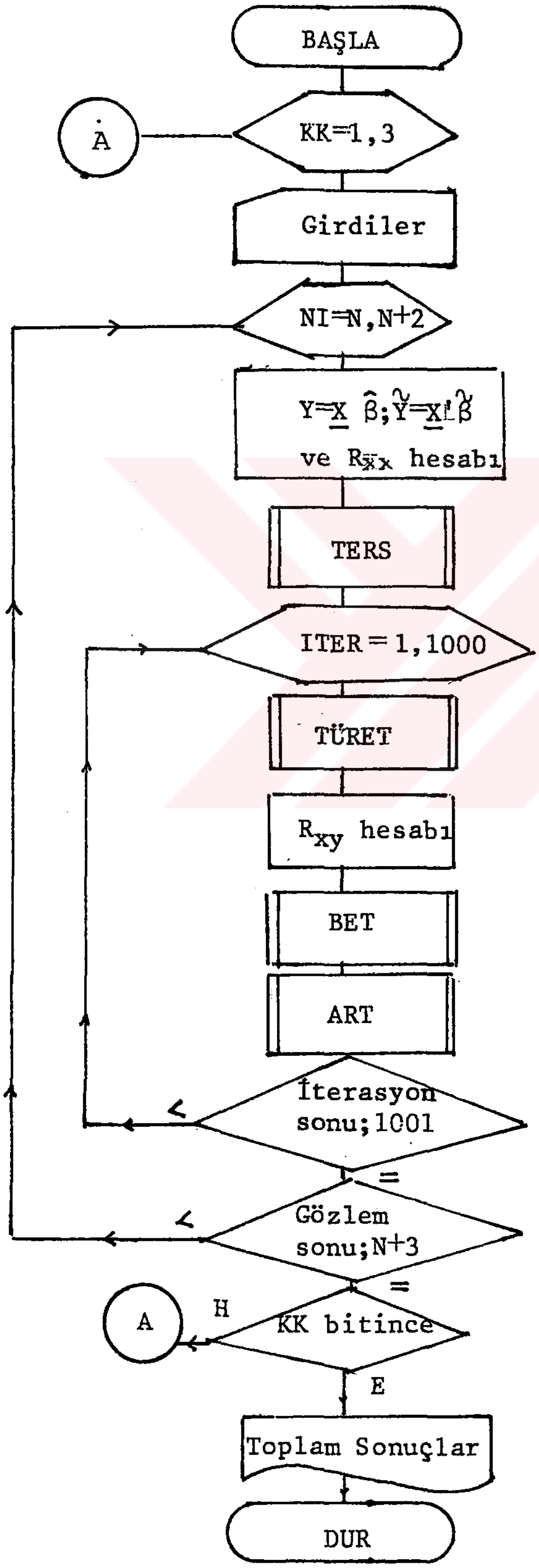
ITER : İterasyon sayısı olup 1000 değerini alır.

TURET : Yapay veri türetir.

BET : $\hat{\beta}$, AKO, Bet YKO ve önkestirim yanılgı kareler ortalamaları hesaplanır.

ART : Artıklara ilişkin ölçüt hesapları, R^2 ve Press bulunur.

TERS : Matris tersini Gauss Yoketme yöntemine göre alır.


BAŞLA
A
KK=1,3
Girdiler
NI=N,N+2
Y=X β̂;Ỹ=XLβ̃
ve R hesabı
TERS
ITER=1,1000
TÜRET
Rxy hesabı
BET
ART
İterasyon
sonu;1001
<
=
Gözlem
sonu;N+3
<
=
A
H
KK bitince
E
Toplam Sonuçlar
DUR

```
!
[ I-N
| #1,1,80,1,1
# #2,1,80,3,22
 #1,1:PEN #1,0
 #1,SPC(30);"GOZLEM ETKILERI ";SPC(33)
 #2
# #3,1,80,24,24
 #3,1:PEN #3,0
?(8,8),TOP(8),TK(8),E(40),X(40,8),D(8),VII(40),BTA(7),Y(40),YTAH(40),DII(100),ART(40)
H(8),W(8),FKIS(8),BET(8),IGOZ(40),YS(40),DD(8),ISAY(40,7),TPL(40,22),IG(22),TT(100)
)FF(100),BFA(10),CAR(10),ITG(6)
UI(100),A(16),F(15),A1(8,8)
)YK(7,7),V12(10),IYOV(15),OYAN(10),OTP(7),OKO(10),YTH(7),TCIZ(10),IGM(2,40)
KKK=1 TO 2
;k=6:JW=8:ITER=1000:L=1:RSAYI=2877.2154:IDG=6:KGS=11:KNI0=8:KNI1=9:KNI2=10:   KNI3=11
="TAB1":ADK$="TAB2"
 .2,-.03,-.22,-.2,-.01,.005,.36
 3.078,1.886,1.638,1.533,1,.816,.765,.741,8,9,10,11
I=1 TO N
J=1 TO K
T X(I,J):NEXT J:NEXT I
I=1 TO 10: IGOZ(I)=0:NEXT I
i=1 TO k: d(i)=1: NEXT i: d(k+1)=0: k1=k+1
I=1 TO KGS: PRINT #2,IG(I);:NEXT I:PRINT #2:PRINT #2
I=1 TO KGS:FOR J=1 TO K:PRINT #2,X(I,J);:NEXT J:PRINT #2:NEXT I
J=1 TO K1
 BTA(J):PRINT #2,BTA(J);:NEXT J:PRINT #2:PRINT #2
I=1 TO IDG:FOR J=1 TO K+1:INPUT OYK(I,J):NEXT J:NEXT I:FOR I=1 TO 8:READ TCIZ(I):NEXT I
I=1 TO IDG:YTH(I)=0 :FOR J=1 TO K:YTH(I)=YTH(I)+OYK(I,J)*BTA(J)
 J:YTH(I)=YTH(I)+BTA(K1):NEXT I
KE=1 TO 4:READ NI:ITG(KE)=NI:RANDOMIZE RSAYI:ISY=0
IT #1,KKK,NI:FOR I=1 TO NI: YTAH(I)=0
J=1 TO K
I(I)=X(I,J)*BTA(J)+YTAH(I)
 J:YTAH(I)=YTAH(I)+BTA(K1)
[ I
):SGMA=0.1:RIR=0:RIR1=0:RIR2=0:RIR3=0:PRES=0:RIR5=0
I=1 TO NI:FOR J=1 TO 7:ISAY(I,J)=0:NEXT J:FOR J=1 TO 22:TPL(I,J)=0:NEXT J:NEXT I
I=1 TO IDG:OTP(I)=0:NEXT I
I=1 TO 12:IYOV(I)=0:NEXT I
J=1 TO K1: TOP(J)=0: TK(J)=0
I=1 TO K1: R(I,J)=0
 I,J
I=1 TO NI
J=1 TO K
(J)=TOP(J)+X(I,J)
L=J TO K
,L)=R(J,L)+X(I,J)*X(I,L)
,J)=R(J,L)
T L,J,I
I=1 TO K:FOR J=1 TO K:A1(I,J)=R(I,J):NEXT J:NEXT I
J=1 TO K:A1(J,K+1)=TOP(J):R(J,K+1)=TOP(J)
```

```
+1,J)=A1(J,K+1):R(K+1,J)=R(J,K+1):NEXT J
+1,K+1)=NI:R(K+1,K+1)=NI
+1:D(K+1)=1:GOSUB 1430:FOR J=1 TO K:TK(J)=A1(J,J)-TOP(J)^2/NI:NEXT J
I=1 TO IDG:FOR J=1 TO K+1:H(J)=OYK(I,J):NEXT J:GOSUB 610:V12(I)=V1:NEXT I
I=1 TO NI:FOR J=1 TO K:H(J)=X(I,J):NEXT J:H(K+1)=1:GOSUB 610
1 >= 1 THEN V1=0.99
I)=V1:NEXT I:GOTO 740
J=1 TO K+1:W(J)=0
(J)=0 THEN 670
M=1 TO K+1
(M)=0 THEN 660
=W(J)+R(J,M)*H(M)
 M
 J

J=1 TO K+1
(J)=0 THEN 720
1+H(J)*W(J)
 J
RN
I=1 TO K+1:FOR J=1 TO K+1
J)=A1(I,J):NEXT J:NEXT I
I=1 TO K-1
J=I+1 TO K
J)=(R(I,J)-TOP(I)*TOP(J)/NI)/SQR(TK(J)*TK(I))
I)=R(I,J)
 J
I)=1
 I
K)=1
:KL=K:D(K+1)=0
 NI-K-1: KOD=1
2*K/NI:DFFSI=2*SQR(K/NI)
B 1430
1

T #2,"ITER=  ";LII
B 2510
K1)=0:TK(K1)=0
J=1 TO K1 : R(J,K1)=0 : NEXT J
I=1 TO NI
K1)=TOP(K1)+X(I,K1)
J=1 TO K1
K1)=R(J,K1)+X(I,J)*X(I,K1)
,J)=R(J,K1)
 J
T I
K1)=R(K1,K1)-TOP(K1)^2/NI
 I=1 TO K
,K1)=(R(I,K1)-TOP(I)*TOP(K1)/NI)/SQR(TK(K1)*TK(I))
1,I)=R(I,K1)
T I
1,K1)=1
```

```
UB 2200
=R(K1,K1)*TK(K1)/(NI-K-1) : AKOP=AKO
UB 1660
KODR=0 THEN 890
=LII+1:IF (LII-1)<ITER THEN 890
 I=1 TO IDG
N(I)=(OTP(I)/ITER-YTH(I))
(I)=SGMA^2*(1+V12(I))+OYAN(I)^2:NEXT I
NI=KNIO THEN 1240
 I=1 TO NI
=(TPL(I,7)-TPL(I,2)^2/ITER)/(ITER-1):IF TP2<0 THEN TPL(I,7)=SQR(0.001) ELSE TPL(I,7)=SQR(TP2)
=(TPL(I,8)-TPL(I,3)^2/ITER)/(ITER-1):IF TP3<0 THEN TPL(I,8)=SQR(0.001) ELSE TPL(I,8)=SQR(TP3)
=(TPL(I,9)-TPL(I,4)^2/ITER)/(ITER-1):IF TP4<0 THEN TPL(I,9)=SQR(0.001) ELSE TPL(I,9)=SQR(TP4)
=(TPL(I,10)-TPL(I,5)^2/ITER)/(ITER-1):IF TP5<0 THEN TPL(I,10)=SQR(0.001) ELSE TPL(I,10)=SQR(TP5)
T I
N=(RIR2/ITER-YTAH(9)):SYKO=SGMA^2*VII(9)+SYAN^2
.(NI+3,3)=SYAN:TPL(NI+3,4)=SYKO:TPL(NI+3,5)=RIR1:TPL(NI+3,6)=RIR
 =(RIR5/(NI*ITER))-(K*SGMA^2)/NI
YAN<0 THEN YAN=YAN*(-1)
.(NI+3,1)=YAN:TPL(NI+3,2)=RIR5
 I=1 TO IDG:TPL(NI+1,I)=OYAN(I):NEXT I
 I=1 TO IDG:TPL(NI+2,I)=OKO(I):NEXT I
 I=1 TO 12:TPL(NI+4,I)=IYOV(I):NEXT I
 I=1 TO KGS:TPL(NI+5,I)=IG(I):NEXT I
T$=KUT$+RIGHT$(STR$(KE),1)+RIGHT$(STR$(KKK),1):SADK$=ADK$+RIGHT$(STR$(KE),1)+RIGHT$(STR$(KKK),1)
NOUT SKUT$
 I=1 TO NI:FOR J=1 TO 7:WRITE #9,ISAY(I,J):NEXT J:NEXT I
SEOUT
NOUT SADK$
 I=1 TO NI+5:FOR J=1 TO 22:WRITE #9,TPL(I,J):NEXT J:NEXT I
SEOUT
T KE
STORE
T KKK
SUB 2650
)
 ******* MATRIS TERSI BULMA **********
=1
 M=1 TO KL
D(M)=0 THEN 1620
=D1*R(M,M)
D1=0 THEN 1640
 I=1 TO KL
(I-M)=0 THEN 1550
 J=1 TO KL
(J-M)=0 THEN 1540
I,J)=R(I,J)-R(I,M)*R(M,J)/R(M,M)
XT J
XT I
R I=1 TO KL
(I-M)=0 THEN 1600
I,M)=-R(I,M)/R(M,M)
M,I)=R(M,I)/R(M,M)
```

```
T I
,M)=1/R(M,M)
T M
URN
NT #JW, "BELIRTEN=0"
URN
 ***** ARTIK , R , D  VE F'LER BULUNUYOR ******
=0:KODR=1
NI=KNI0 THEN 2070
NI=KNI1 THEN KDU=1:GOTO 1710
NI=KNI2 THEN KDU=2 ELSE KDU=3
KDU GOTO 1720,1730,1740
=0.307:DS2=0.588:TS1=63.66:TS2=25.45:UTAB=0.02209:GOTO 1750
=0.325:DS2=0.582:TS1=9.92:TS2=6.21:UTAB=0.08082:GOTO 1750
=0.338:DS2=0.583:TS1=5.84:TS2=4.18:UTAB=0.15156
 I=1 TO NI
P=0
 J=1 TO K
D(J)=0 THEN 1800
P=YSAP+BET(J)*X(I,J)
T J
P=YSAP+BET(K1)
I)=YSAP
(I)=X(I,K1)-YSAP
S=PRES+(ART(I)/(1-VII(I)))^2
ART(I)/SQR(AKOP*(1-VII(I))):E(I)=RI
I=AKOP*((NI-K-1)-E(I)^2)
AKTI < 0.00001 THEN KODR=0:RETURN
T I
 I=1 TO NI
= E(I)^2*VII(I)/((1-VII(I))*(L9+1))
=SQR(ABS(E(I)^2*(NL9-1)/(NL9-E(I)^2)))
I=AKOP*((NI-K-1)-E(I)^2)
KTI/(AKOP*(NI-K-1))
I=AKTI/(NI-K-2)
I=KNI1 THEN  RIR2=RIR2+YS(I)
ITS=ART(I)*SQR(VII(I)/AKOI)/(1-VII(I))
VII(I) > VSI THEN ISAY(I,1)=ISAY(I,1)+1
DI1 > DS1 THEN ISAY(I,2)=ISAY(I,2)+1:IF DI1 > DS2 THEN ISAY(I,3)=ISAY(I,3)+1
TI1 > TS1 THEN ISAY(I,4)=ISAY(I,4)+1:IF TI1 > TS2 THEN ISAY(I,5)=ISAY(I,5)+1
ABS(DFFITS) > DFFSI THEN ISAY(I,6)=ISAY(I,6)+1:IF U>UTAB THEN ISAY(I,7)=ISAY(I,7)+1
(I,1)=TPL(I,1)+VII(I)
(I,2)=TPL(I,2)+DI1:TPL(I,7)=TPL(I,7)+DI1^2
(I,3)=TPL(I,3)+TI1:TPL(I,8)=TPL(I,8)+TI1^2
(I,4)=TPL(I,4)+DFFITS:TPL(I,9)=TPL(I,9)+DFFITS^2
(I,5)=TPL(I,5)+U:TPL(I,10)=TPL(I,10)+U^2
T I
 I=1 TO IDG:YOSAP=0
 J=1 TO K:IF D(J)=0 THEN 2100
AP=YOSAP+BET(J)*OYK(I,J)
T J
AP=YOSAP+BET(K1)
```

```
).1*SQR(1+V12(I))*TCIZ(KE):IF (YOSAP-T9)<=YTH(I) AND YTH(I)<=(T9+YOSAP)THEN IYOV(I)=IYOV(I)+1
).1*SQR(1+V12(I))*TCIZ(KE+4)
:YOSAP-T8) <= YTH(I) AND YTH(I)<=(T8+YOSAP) THEN IYOV(I+6)=IYOV(I+6)+1
(I)=OTP(I)+YOSAP
[ I
|=RIR1+T1:RIR5=RIR5+YKO
=RIR+PRES
JRN
 ***** BETA VE F'LERIN BULUNMASI *****
 I=1 TO K : BET(I)=0
 J=1 TO K : BET(I)=R(I,J)*R(J,K1)+BET(I)
T J,I
RE=0
 I=1 TO K
RE=RKARE+BET(I)*R(I,K1)
T I:R(K1,K1)=1-RKARE
 J=1 TO K
D(J)=0 THEN BET(J)=0:GOTO 2310
(J)=BET(J)*(SQR(TK(K1)/TK(J)))
T J
(K1)=0
 J=1 TO K
(K1)=BET(K1)+BET(J)*TOP(J)/NI
T J
(K1)=TOP(K1)/NI-BET(K1)
 J=1 TO K+1:BFA(J)=(BTA(J)-BET(J)):NEXT J
 J=1 TO K+1:CAR(J)=0
 M=1 TO K+1
(J)=CAR(J)+A1(J,M)*BFA(M)
T M
T J
|=0
 J=1 TO K+1
|=YKO+BFA(J)*CAR(J)
T J
0
 J=1 TO K
T1+(BTA(J)-BET(J))^2:NEXT J:T1=T1+(BTA(K1)-BET(K1))^2
URN
 I=1 TO NI
UB 2590
)=YTAH(I)+E(I)
 Y(I) < 0.01 THEN 2520
 Y(I) > 1 THEN 2520
T I
 I=1 TO NI: X(I,K1)=Y(I):NEXT I
URN
)
 K9=1 TO 12
+RND(5)
T K9
)=EX+SGMA*(A-6)
URN
```

```
NT #2,TAB (30);"RETURN'E BASINIZ....";:INPUT #2,C$
 KKK=1 TO 2:FOR KE=1 TO 4: NI=ITG(KE)
T$=KUT$+RIGHT$(STR$(KE),1)+RIGHT$(STR$(KKK),1):SADK$=ADK$+RIGHT$(STR$(KE),1)+RIGHT$(STR$(KKK),1)
NT #JW,NI;"  GOZLEM ICIN.... ";SKUT$;"  ";SADK$;"  KUTUKLERINDEN... "
NIN SKUT$
 I=1 TO NI:FOR J=1 TO 7:INPUT #9,ISAY(I,J):NEXT J:NEXT I
SEIN
NIN SADK$
 I=1 TO NI+5:FOR J=1 TO 22:INPUT #9,TPL(I,J):NEXT J:NEXT I
SEIN
NT #JW:FOR I=1 TO KGS:PRINT #JW,TPL(NI+5,I);:NEXT I:PRINT #JW
NI=KNIO THEN 2900
NT #JW:PRINT #JW," G-NO";SPC(2);"V ";SPC(1);"D1";SPC(2);"D2";SPC(2);"t1";SPC(2);
NT #JW,"T2";SPC(2);"DFFS";SPC(1);"UI":PRINT #JW,STRING$(79,"-"):PRINT #JW
 I=1 TO NI
NT #JW,USING "####";I;ISAY(I,1);ISAY(I,2);ISAY(I,3);ISAY(I,4);ISAY(I,5);ISAY(I,6),ISAY(I,7)
T I
NT #JW,STRING$(79,"-"):PRINT #JW
NT #JW,"G-NO ";SPC(1);"VOR";SPC(2);" DOR";SPC(4);" tOR";SPC(3);" DFOR";SPC(3);"UOR";SPC(2);
 I=1 TO NI
NT #JW,I;
NT #JW,USING "###.###";TPL(I,1)/ITER;TPL(I,2)/ITER;TPL(I,3)/ITER;TPL(I,4)/ITER;
NT #JW,USING "###.###";TPL(I,5)/ITER;TPL(I,7);TPL(I,8);TPL(I,9);TPL(I,10)
T I
NT #JW,"SAP2";SPC(4);"SAP3";SPC(2);"SAP4";SPC(3);"SAP5":PRINT #JW,STRING$(79,"-"):PRINT #JW
NT #JW:PRINT #JW,"DISARIDAN GOZLEMLERIN ON KES.YAN.K.O VE YANI : "
 I=1 TO IDG
NT #JW,I;
NT #JW,USING "#####.#########  ";SQR(TPL(NI+2,I));TPL(NI+1,I)
T I
NT #JW
NT #JW,"GUVEN SIN. DUSME SAY: ";:FOR I=1 TO 12:PRINT #JW,USING " ####";TPL(NI+4,I);:NEXT I:PRINT #JW
NT #JW:PRINT #JW,"TOP.OYKO ve TOP.YAN= ";:PRINT #JW,USING "#####.###";SQR(TPL(NI+3,2)/(ITER*NI))
NT #JW,USING "#####.########";SQR(TPL(NI+3,1))
NI=KNIO THEN 3040
NT #JW:PRINT #JW,"9.OYKO ve YANI=      ";:PRINT #JW,USING "#####.########  ";SQR(TPL(NI+3,4));TPL(NI+3,3)
NT #JW
NT #JW,"TOP.BET.KO ve PRESS= ";:PRINT #JW,USING "####.###";TPL(NI+3,5)/(SGMA^2*ITER);
NT #JW,TPL(NI+3,6)/(ITER*NI)
NT #JW:PRINT #JW,STRING$(79,"*"):PRINT #JW:NEXT KE
NT #JW:PRINT #JW,STRING$(79,"*"):PRINT #JW
T KKK
URN
```