

HARRAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BASİT DOĞRUSAL REGRESYON MODELLERİNDEKİ
TEK BİR ANORMAL GÖZLEMİN TESTİ

Zeki DOĞAN

Danışman
Prof.Dr.Necati YILDIZ

YÜKSEK LİSANS TEZİ
ZOOTEKNİ ANABİLİM DALI

1995
ŞANLIURFA

HARRAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BASİT DOĞRUSAL REGRESYON MODELLERİNDEKİ
TEK BİR ANORMAL GÖZLEMİN TESTİ

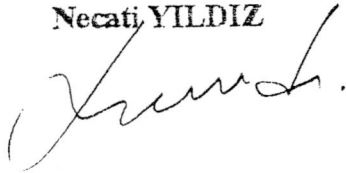
Zeki DOĞAN

YÜKSEK LİSANS TEZİ
ZOOOTEKNİ ANABİLİM DALI

Bu tez 23/01/1995 tarihinde aşağıdaki juri tarafından değerlendirilerek oy birliği/ çokluğu ile kabul edilmiştir.

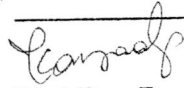
(İmza)

Danışman
Prof.Dr.
Necati YILDIZ



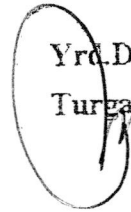
(İmza)

Yrd.Doç.Dr
Tamer KAYAALP



(İmza)

Yrd.Doç.Dr
Turgay SENGÜL



HARRAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

BASİT DOĞRUSAL REGRESYON MODELLERİNDEKİ
TEK BİR ANORMAL GÖZLEMİN TESTİ

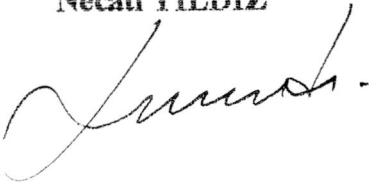
Zeki DOĞAN

YÜKSEK LİSANS TEZİ
ZOOTEKNİ ANABİLİM DALI

Bu tez 23/01/1995 tarihinde aşağıdaki juri tarafından değerlendirilerek oy birliği/ çokluğu ile kabul edilmiştir.

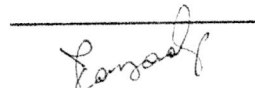
(İmza)

Danışman
Prof.Dr.
Necati YILDIZ



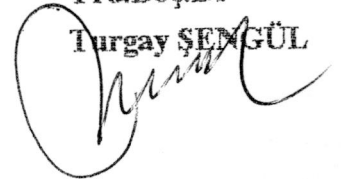
(İmza)

Yrd.Doç.Dr
Tamer KAYAALP



(İmza)

Yrd.Doç.Dr
Turgay ŞENGÜL



İÇİNDEKİLER	Sayfa
SİMGELER	IV
ÇİZELGELER.....	V
ŞEKİLLER	VI
<i>TEŞEKKÜR</i>	VII
ÖZET.....	VIII
ABSTRACT.....	IX
1. GİRİŞ	1
2. ÖNCEKİ ÇALIŞMALAR.....	5
2.1. RESİDUALLERİN İNCELENMESİ.....	12
2.1.1. <i>Residuallerin Özellikleri</i>	14
2.2. HAT MATRİS.....	15
2.3. ANORMAL GÖZLEMLER (OUTLIERLER).....	17
3. MATERYAL VE YÖNTEM.....	20
3.1. MATERYAL.....	20
3.2. YÖNTEM.....	21
3.2.1. <i>Standardize Edilmiş Residual Grafikleri Yöntemi İle Outlierlerin Teşhisi (Plot Yöntemi)</i> ..	21
3.2.2. <i>Cook İstatistiği</i>	23
3.2.3. <i>Katkı Kareler Toplamı Testi</i>	28
4. ARAŞTIRMA SONUÇLARI.....	30
4.1. PLOT YÖNTEMİ.....	31
4.1.1. <i>Tahmin Edilen \hat{Y}_i Değerlerine Karşı Student Türü Residuallerin Grafikleri</i>	33
4.1.2. <i>Bağımsız X_i Değişken Değerlerine Karşılık Student Türü Residuallerin Grafikleri</i>	33
4.2. COOK İSTATİSTİĞİNİN UYGULAMASI.....	34
4.3. KATKI KARELER TOPLAMI TESTİNİN UYGULAMASI.....	36
5. TARTIŞMA.....	40
EK 1. Cook İstatistiğinde kullanılan F Cetveli.....	43

III

EK 2.1 ve n-k-2 Serbestlik Dereceli α/n Yanılma Düzeyindeki F Cetveli	45
KAYNAKLAR.....	47

SİMGELER

"Kısaltmalar"

AP	: Andrews-Pregibon İstatistiği
α	: Regresyon doğrusunun eğimi
β	: Regresyon katsayısı
e	: Residual
EKK	: En Küçük Kareler
gr	: Gram
k	: Bağımsız değişken sayısı
mm	: Milimetre
N	: Normal dağılım
n	: Örnekleme büyüklüğü
p	: Parametre sayısı
r_i	: Standardizet residual
R, V, M	: Hat matris
RKO-MSE	: Residual kareler ortalaması
RKT-SSE	: Residual kareler toplamı
R^2	: Belirtme Katsayısı
S^2	: Örneklerin varyansı
σ^2	: Popülasyon varyansı
t_i	: Studentized residual
X	: Bağımsız değişken sayısı
Var	: Varyans
Y	: Regresyon işlemlerindeki bağımlı değişken
\hat{Y}_i	: Regresyon modelindeki EKK tahmin değeri
%	: Yüzde

Çizelge 1. Bildircinlerden Elde Edilen Yumurta Boy Değerleri (X bağımsız değişkeni) ve Yumurta Ağırlıkları (Y bağımlı Değişkeni).....	20
Çizelge 2. Bildircinların yumurta boyları ile yumurta ağırlıkları arasındaki ilişkiye ait varyans analiz tablosu.....	30
Çizelge 3. Bildircinlerden elde edilen X_i (yumurta boyu), Y_i (gözlem değeri), \hat{Y}_i (tahmin değeri) ve bunlardan hesaplanan e_i (basit residualler), t_i (student türü residual) ve V_{ii} (hat matrisin diagonal elemanları) değerleri	32
Çizelge 4. Bildircinlerden elde edilen 30 gözleme ait (t_i) student türü residualler, (V_{ii}) hat matrisin diagonal elemanları ve (D_i) Cook İstatistiği kritik değerleri	35
Çizelge 5. 28. Gözlem çıkarıldıktan sonraki 29 gözleme ait regresyon analizi için varyans analiz tablosu.....	37
Çizelge 6. 30 Gözlem ait X_i , Y_i , t_i ve kuşkulanılan gözlemlere ait F_i kritik değerleri.	39

ŞEKİLLERSayfa

Şekil 1. $y=f(x)$ fonksiyonunu gösteren doğrunun grafiği.....	2
Şekil 2. Ekstrem değerlerin regresyon doğrusuna etkileri.....	19
Şekil 3. a) hiçbir anormal gözlemin bulunmadığı residual grafiği, b) tek bir anormal gözlemin bulunduğu residual grafiği.....	22
Şekil 4. Bildiricilerden elde edilen verilere ait serpm diyagramı.....	31
Şekil 5. Tahmin değerlerine karşılık student türü residuallerin grafiği.....	33
Şekil 6. X_i değerlerine karşılık student türü residuallerin grafiği.....	34
Şekil 7. 28. Gözlem çıkarıldıktan sonraki X_i ve Y_i değerlerine ait serpm diyagramı.....	41

TEŞEKKÜR

Bana bu araştırma konusunu veren ve çalışmalarımın yürütülmesi süresince yardımlarını esirgemeyen çok kıymetli sayın hocam Prof. Dr. Necati YILDIZ'a, bölümümüz imkanlarından yararlanmamı sağlayan Bölüm Başkanımız Yrd.Doç. Dr. Rahmi KANAT'a ve bana her zaman destek olan Yrd. Doç. Dr. Turgay ŞENGÜL, Yrd. Doç. Dr. Ramazan ŞEVİK, Yrd. Doç. Dr. Tamer KAYAALP, Arş. Gör. Ahmet YILDIZ, Arş. Gör. İrfan ÖZTÜRK, Arş. Gör. Yusuf KONCA, Arş. Gör. Aslan YÜKSEL ve Arş. Gör. Sabri YURTSEVEN'e teşekkürleri bir borç bilirim.

Bu çalışmanın konu ile ilgilenen araştırmacılara faydalı olması en büyük dileğimdir. Çalışmamda kullandığım verilerin analizinde yardımcı olan Arş. Gör. Hikmet ORHAN'a ve bu tezin yazımında büyük emeği geçen İnş. Müh. Mehmet YILDIRIM'a teşekkür ederim.

ÖZET

Yüksek Lisans Tezi

BASİT DOĞRUSAL REGRESYON MODELLERİNDEKİ TEK BİR
ANORMAL GÖZLEMİN TESTİ*Zeki DOĞAN*

Harran Üniversitesi

Fen Bilimleri Enstitüsü

Zootekni Anabilim Dalı

1995

Doğrusal Regresyon modellerinde anormal gözlemlerin bulunması işlemi son yıllarda önemli bir merak konusu olmuştur. Bu çalışmada; basit doğrusal regresyon problemlerindeki anormal gözlemlerin testinde kullanılan üç yöntem önerilmiştir.

- 1) Student türü residuallerin noktalanması ile outlierlerin teşhisi;
 - a) \hat{Y}_i tahmin değerlerine karşılık plotlar,
 - b) X_i bağımsız değişkenlerine karşılık plotlar,
- 2) Cook istatistiği,
- 3) Katkı kareler toplamı testi.

Anormal gözlemler ekstrem değerlerdir. Student türü residual plotlarında anormal gözlemler, sıfırdan yaklaşık olarak dört, beş veya daha fazla standart sapma uzaklıktadırlar. Bu gözlemler literatürde; anormal değerler, düzensiz gözlemler, kuraldışı değerler, başıboş değerler ve çirkin veriler olarak adlandırılmışlardır.

Bu çalışmada bildirecilerden elde edilen yumurta boyları X_i bağımsız değişkeni olarak ve yumurta ağırlıklarında Y_i gözlem değeri olarak kullanılmıştır. Sonuç olarak; bir gözlem her üç metodluda anormal gözlem olarak teşhis edildi.

ANAHTAR KELİMELER: Regresyon Analizi, Residual Analizi, Etkili Gözlemler, Anormal Gözlemler, Anormal Gözlemlerin Bulunması.

ABSTRACT

Master Thesis

TESTING FOR A SINGLE OUTLIER FROM A SIMPLE LINEAR
REGRESSION*Zeki DOĞAN*

Harran University
Graduate School of Natural and Applied Science
Department of Animal Science
1995

The detecting outliers has considerable attracted on the linear regression models in recent years. In this study three methods which are used in detecting of outliers, in a simple linear regression have been performed.

1) Plotting of studentized residuals(t_i);

a) Plots against \hat{Y}_i

b) Plots against the independent variables X_i

2) Cook's statistics

3) Test of contribution sum of squares.

Outliers are extreme observations. In a studentized residual plot, outliers are far from zero about four, five or more standart deviations. These observations have been called as rogue values, contaminants, surprising values, mavericks and dirty data in the literature.

In this study the length of eggs that was obtained the quails was used as X_i independent variables and the weights of eggs was used as Y_i observation values. As a result; one observation was detected as an outlier in this data all three methods.

KEY WORDS: Regression Analysis, Residual Analysis, Influential Observations, Outliers, Detection of Outliers.

1. GİRİŞ

Bir değişkenin, diğer bir veya bir kaç değişkenle ilişkisini ifade eden matematiksel ifadeye *Regresyon Denklemi* denir. Regresyon analizi ise; bu değişkenler arasındaki ilişkiyi belirlemede uygulanan istatistik yöntemlerdir. Eğer regresyon analizinde bir bağımlı (Y) birde bağımsız (X) değişken kullanılıyorsa; *Basit Regresyon*, eğer üç veya daha fazla değişken kullanılıyorsa da; *Çoklu Regresyon* adını alır. Eğer X bağımsız değişkeni arttıkça Y bağımlı değişkeni doğrusal olarak artıyor veya azalıyor *Basit Doğrusal Regresyon* denir. Bu ilişki matematiksel olarak $Y=f(x)$ şeklinde gösterilir. $Y = f(x)$ fonksiyonunu gösteren doğru Şekil 1.'de verilmiştir.

En basit regresyon denklemi $Y=\beta x$ şeklinde tanımlanır. Bu ilişkiye ait regresyon doğrusu;

$$Y = \alpha + \beta x \quad (1)$$

şeklindedir. Değişkenler arasındaki ilişki genel olarak;

$$Y_i = \alpha + \beta x_i + \varepsilon_i \quad (2)$$

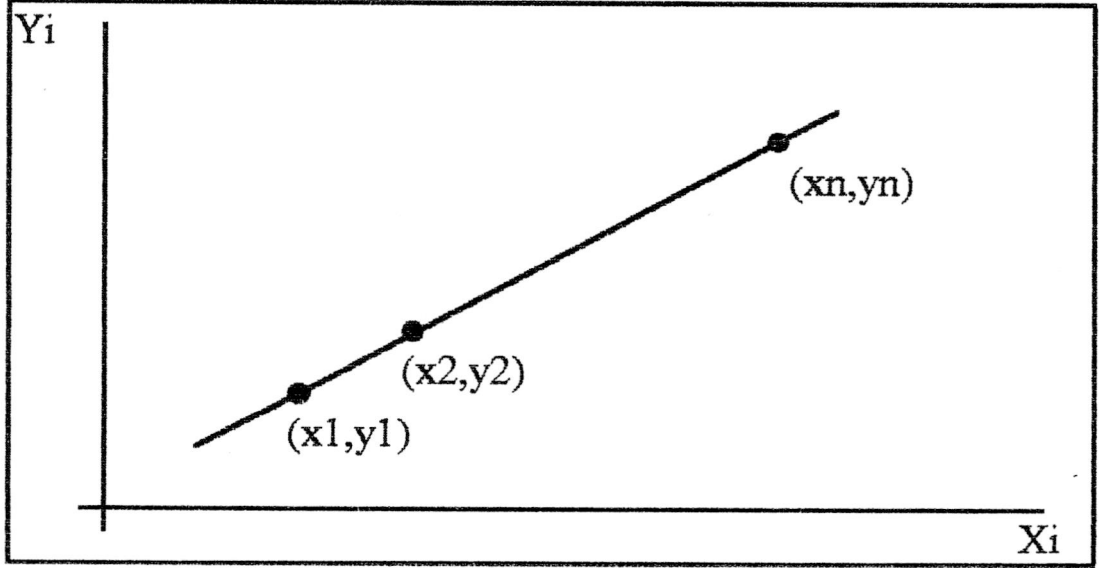
şeklindedir. Burada;

β : Regresyon katsayısıdır. X_i 'deki bir birimlik artışa karşılık gelen Y_i 'deki artıştır.

α : Regresyon doğrusunun eğimi olup, X_i 'nin sıfır olduğu durumda Y_i 'nin alacağı değerdir. Yani regresyon doğrusunun Y eksenini kestiği noktanın orjine olan uzaklığıdır.

ε_i : Hata terimidir. Y_i ile \hat{Y}_i arasındaki fark olarak da ifade edilebilir. Yani regresyon denkleminin açıklayamadığı miktardır.

Regresyon analizinin temeli bu hata teriminin incelenmesine dayanır. En Küçük Kareler metoduna (EKK) göre a ve b öyle bir şekilde tahmin edilmelidirki, hata teriminin kareler toplamı minimum olsun.



Şekil 1. $y=f(x)$ fonksiyonunu gösteren doğrunun grafiği.

Dolayısıyla kesin bağıntı;

$$Y = \alpha + \beta x + e \quad (3)$$

şeklinde gösterilir. $\sum e_i^2$ en küçük olacak şekilde EKK'ya göre katsayılar tahminlenir.

α ve β parametrelerinin nokta tahminleri olan a ve b 'nin EKK tahminleri;

$$a = \bar{Y} - b\bar{x}$$

$$b = \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sum x_i^2 - \frac{(\sum x_i)^2}{n}} = \frac{S_{xy}}{S_{xx}}$$

şeklinde bulunur [14]. Matris cebirinde ise;

$$Y_i = x_i \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \varepsilon_i \quad (4)$$

şeklinde gösterilir. Burada n sayıda ve tek bir bağımsız değişken için model ;

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & X_n \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \cdot \\ \cdot \\ \varepsilon_n \end{bmatrix}$$

olur. Bunun kapalı şekli $Y = X\beta + \varepsilon_i$ dir.

$$\beta = (X'X)^{-1}X'Y \quad (5)$$

$$X'X = \begin{bmatrix} n & \sum x \\ \sum x & \sum x^2 \end{bmatrix}$$

$$X'Y = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

Bulunan bu değerler (5)'de yerine konursa sonuçta;

$$\beta = \begin{bmatrix} a \\ b \end{bmatrix}$$

şeklinde bir matris elde edilecektir. Bulunan bu değerler de (3)'de yerine konduğunda regresyon modeli ortaya çıkmış olur.

Regresyon modelindeki e_i , gözlem değerleri ile tahmin değerleri arasındaki farklardır[9].

$$e_i = Y_i - \hat{Y}_i \quad i=1,2,\dots,n$$

e_i modelin açıklayamadığı kısım olarak bilinmekle birlikte çalışmalarda farklı isimlerde kullanılmaktadır. Bunlar;

- hata terimi,
- kalıntılar,
- artıklar,
- yanılğı terimi,
- residuallerdir.

Hata terimleri normal dağılış gösteren bir şans değişkeni olup $e_i \sim N(0, \sigma^2)$ 'dir[18].

Regresyon analizinde en önemli işlemlerden birisi; residuallerin analizidir. Residuallerin analizi sayesinde verilerde zayıf parametre tahminlerine neden olan herhangi bir anormallik ortaya çıkarılmakta ve anormal gözlemler teşhis edilebilmektedir [18].

Beckman ve Cook (1983)'e göre outlierler veri setlerinde ortalamadan önemli derecede sapan ekstrem noktalardır. Outlierler çalışmalarda farklı isimlerle ifade edilmektedirler. Bunlar;

- Outlierler,
- Anormal gözlemler,
- Aykırı değerlerdir.

Bu çalışmada genellikle hata terimleri için *Residualler*, anormal gözlemler için ise *Outlierler* terimleri kullanılmıştır.

Residuallerin standardize edilmiş halleri, hat matris olarak bilinen bir izdüşüm matrisi ile birlikte kullanıldığında outlierlerin teşhisini sağlar[17].

Doğrusal regresyon modellerinde outlierlerin bulunması işlemi son yıllarda önemli bir merak konusu olmuştur[4]. Outlierlerin teşhisi konusunda bir çok istatistikler geliştirilmiştir. Bunlardan bazıları ;

- Standardize edilmiş residual grafikleri,
- Cook İstatistiği,
- Dffits İstatistiği,
- Dfbetas İstatistiği,
- Andrews-Pregibon İstatistiği,
- Katkı Kareler Toplamı testidir.

Bu çalışmada; bunlardan daha elverişli olan Residuallerin Noktalanması, Cook İstatistiği ve Katkı Kareler Toplamı Testi metodları karşılaştırmalı olarak değerlendirilmiştir.

2. ÖNCEKİ ÇALIŞMALAR

Regresyon analizi ile ilgili, oldukça önemli çalışmalar yapmış olan Draper ve Smith (1966), basit doğrusal regresyon modelini;

$$Y = x\beta + \varepsilon \quad (1)$$

şekinde göstermiştir. Burada;

Y : $(n \times 1)$ boyutlu gözlem vektörü matrisi,

X : $(n \times p)$ boyutlu tam ranklı şans değişkeni matrisi,

b : $(p \times 1)$ boyutlu bilinmeyen parametreler vektörü,

ε : $(n \times 1)$ boyutlu gözlenmeyen hata terimleri vektörü.

Eğer b 'nin EKK tahmini (1)'de yerine konunca residual terimi;

$$Y = x\hat{\beta} + e$$

$$e = Y - x\hat{\beta}$$

$$\hat{Y} = x\hat{\beta} \text{ olduğundan;}$$

$$e = Y - \hat{Y}_i \quad (2)$$

Y : gözlem değeri,

\hat{Y} : tahmin değeri,

şeklinde tanımlanmaktadır. e_i basit residuallerinin tümünün grafikleri, \hat{Y}_i değerlerine karşı ve X_i bağımsız değişken değerlerine karşılık residualleri noktalamışlar; böylece modelin uyumunu hatanın normallik varsayımını bozup bozmadığını ve gözlemlerde anormal gözlem bulunup bulunmadığını incelemişlerdir.

Freund ve Wail (1961), basit doğrusal modeli (1)'deki gibi göstermişlerdir. Bu basit ilişkiden residuali \tilde{Y}_i olarak ifade edip;

$$\tilde{Y}_i = Y - x\hat{\beta}$$

şeklinde hesaplamışlardır.

Mason ve Gunst (1978), regresyon analizinde en önemli işlerden birisinin residuallerin incelenmesi olarak belirlerler. Residual analizi verilerin tahmin denklemlerindeki değişkenlerin, gözlem ve tahmin değerleri arasındaki farkın incelenmesini gerektirip, yetersiz parametre tahminlerine neden olan anormal noktaların meydana çıkarılmasında residuallerin önemli bir yer tuttuğunu ve anormal

gözlemlerin teşhisinde standart residuallerin noktalanmasını tavsiye etmişlerdir. Herhangi bir i . residuali,

$$r_i = Y_i - \hat{Y}_i \quad i=1,2,\dots,n$$

şeklinde, i . standart residuali;

$$S_i = \frac{r_i}{\hat{\sigma}} \quad (3)$$

$$\hat{\sigma}^2 = \text{MSE} = \frac{r_i^2}{n-p-1}$$

MSE : Residual kare ortalaması

P : Parametre sayısı

n : Örnek büyüklüğüdür.

olarak hesaplayıp, hat matrisine bağlı olarak student türü residuali;

$$t_i = \frac{r_i}{\hat{\sigma} \sqrt{1-h_{ii}}} \quad i=1,2,3\dots n \quad (4)$$

şeklinde hesaplanmıştır. Burada;

h_{ii} : i . gözlem uzaklığı olarak adlandırılır ve gözlem noktasının değişkenlerin oluşturduğu elipsoid merkeze olan uzaklıktır. Outlierlerin teşhisinde student türü residuallerin kullanılmasını tavsiye etmişlerdir.

Hoaglin ve Welsch(1978), \hat{Y} değerlerinde önemli etkisi olan Y verisinin etkisinin anlaşılmasının EKK'de önemli bir işlem olduğunu belirtmişler ve bu bilgileri hat matris olarak bilinen izdüşüm matrisinin içerdiğini, ayrıca student türü residuallerle birlikte anormal veri noktalarının teşhisini sağlayacağını belirtmişlerdir.

$$Y = X\beta + \varepsilon$$

şeklindeki doğrusal bir modelde;

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

$$\hat{Y} = X(X'X)^{-1} X'Y \quad (5)$$

$$\beta = (X'X)^{-1} X'Y \quad (6)$$

$$V = X(X'X)^{-1} X' \quad (7)$$

(6) ve (7)'deki deęerler (5)'te yerine konduęunda \hat{Y} deęerini;

$$\hat{Y} = V \cdot Y$$

olarak bulmuřlardır.

Montgomery ve Peck (1982), residualleri dięer alıřmalarda olduęu gibi;

$$e_i = Y_i - \hat{Y}_i$$

řeklinde ifade edip residuallerin varyansını;

$$S_e^2 = \frac{\sum_{i=1}^n (e_i - \bar{e})^2}{n-2} = \frac{\sum_{i=1}^n e_i^2}{n-2} = \frac{SSE}{n-2} = MSE$$

řeklinde, standart residuali de;

$$d_i = \frac{e_i}{\sqrt{MSE}} \quad i=1,2,\dots,n$$

olarak gstermiřlerdir. rnekler zerinde hem e_i hem de d_i residual deęerlerini \hat{Y}_i , X_i , ve zaman serilerine karřılık noktalayıp outlierleri teřhis edip bir outlierin bir u deęer ve en geniř residuele sahip olan 0'dan 3 veya 4 standart sapma uzaklıkta olan kuvvetli bir gzlem olduęunu belirtmiřlerdir.

Ellenberg(1976), genel doęrusal regresyon modellerinde tek bir outlier iin bir test kullanılmıřtır.

$$Y_i = X_i\beta + U_i \quad i=1,2,\dots,n$$

gibi bir modelde hata terimini U_i olarak gstermiřtir. Standart residuali, EKK tahminlerini kullanarak;

$$\varepsilon_{si} = \frac{U_i}{((n-k)\text{Var}(U_i))^{1/2}} \quad i=1,2,\dots,n$$

$$U_i = Y - X\hat{\beta}_i$$

$$E(UU') = \sigma^2 M$$

$$M_{n \times n} = I - X(X'X)^{-1}X'$$

$$t_i = \frac{U_i}{Sm_{ii}}$$

$$S^2 = U'U$$

$$Sm_{ii}^2 = \sqrt{\text{Var}(U_i)}$$

şeklinde bulunan herbir e_i değeri C_α gibi bir kritik değerle karşılaştırılmıştır. Eğer;

$$\text{Max}|e_i| \geq C_\alpha$$

ise 0 hipotezinin reddedildiğini,

$$\text{Max}|e_i| < C_\alpha$$

olduğunda ise 0 hipotezinin kabul edileceğini belirtmektedir.

Bertram ve Price (1991), normal basit residuali;

$$\varepsilon_i = Y_i - \hat{Y}_i$$

olarak, standart residuali de;

$$\varepsilon_{si} = \frac{\varepsilon_i}{\left[1 - \frac{1}{n} - \frac{(X_{ii} - \bar{X}_1)^2}{\sum (X_{ii} - \bar{X}_1)^2} \right]^{1/2}}$$

formülü ile ifade etmişlerdir. Bulunan bu ε_{si} değerlerini (\hat{Y}_i) tahmin değerlerine, (X_i) bağımsız değişkenlerine ve zamana bağlı olarak noktalanması ile, regresyon doğrusundaki yetersizliği, sürekli varyans eksikliğini, outlierlerin varlığını ve tahmin modellerinin yetersizliğini ortaya çıkarabileceğini belirtmişlerdir. Bu bilgiler doğrultusunda bir televizyon kanalının izleme oranıyla ilgili bir çalışmadan elde edilen gözlemlerin standart residuallerini X_i bağımsız değişkenine karşılık noktalandıklarında parametre tahminlerine etki eden outlierlerin bulunduğunu ortaya çıkarmışlardır.

Netter ve Wasserman(1973), outlierleri şöyle tarif ederler: outlierler uç gözlemler olup, bir residual dağılımını gösteren noktalamada orjinden yatay veya

düşey uzaklaşan noktalar. Yaklaşık olarak 0'dan dört veya daha fazla standart sapma uzaklıktadırlar. Standart residualleri ;

$$r_i = \frac{e_i}{\sqrt{\text{MSE}}} \quad i=1,2,\dots,n$$

şeklinde tarif etmişlerdir.

Cook (1977), outlierlerin teşhisinde standart residuallerden ziyade student türü residuallerin kullanılmasının daha uygun olacağını belirtmişlerdir. Student türü residualleri;

$$t_i = \frac{e_i}{S\sqrt{1-V_i}} \quad i=1,2,\dots,n$$

şeklinde gösterip, outlierlerin teşhisinde kullanılan t_i 'ye ve V_{ii} 'ye bağlı olarak;

$$D_i = \frac{t_i^2(V_{ii})}{p(1-V_{ii})} \quad i=1,2,\dots,n$$

kritik değerinin kullanılmasını tavsiye etmiştir. Bulunan bu D_i değerini $F(1-\alpha, p, n-p)$ cetvel değeri ile karşılaştırmıştır.

Beckman ve Cook(1983), outlierler alanında çok geniş bir çalışma yapmışlardır. Outlierlerin ne olduğu, niçin outlierlerle çalışıldığı, outlierlerin tarihçesi, outlierlerin sebepleri ve normal doğrusal modellerde outlierler gibi konularda açıklayıcı bilgi sunmuşlardır.

Andrews ve Pregibon(1978), yaptıkları çalışmada parametre tahminlerinin sonuçları üzerinde büyük etkisi olan bir veya birden fazla outlierleri ortaya çıkaran bir test sunmuşlardır. Bu test istatistiğine isimlerinin ilk harfleri olan AP İstatistiği denmektedir.

Barnett (1978), örnekler üzerinde outlierlerle ilgili yapılan çalışmalarını incelemiştir. Outlierleri tanımlarken en basit sayı dizilerindeki anormalliklerden başlayarak, normal örneklere doğru konuyu açmışlardır. Outlierleri basit bir örnek olan Finney (1974) değerlerinde incelemiştir. Kütmes hayvanlarından sayılan bir kuşun düzenli zaman aralıklarındaki ortalama ağırlıkları şöyledir;

1.20 1.60 1.90 1.55 2.20 2.25

Buradaki 1.55 deęerinin örnek daęılımını bozduęunu ve outlier olabileceęini belirtmiřtir. Yine başka bir örnekte; 10 tane zar arka arkaya atılmıř ve çıkan sonuçlar řöyledir;

2, 3, 12, 2, 1, 3, 4, 5, 6

Buradaki 12'nin kesinlikle outlier olduęunu ve sebebinin kayıtlamalardaki hata olduęunu, çünkü zarın 12 sayısı bulunan bir yüzü olmadıęını belirtmiřtir.

Draper ve John (1981), yaptıkları çalışmalarda outlierleri; kare toplamları metodu, Andrews-Pregibon istatistięini ve Cook'un (1977) sunduęu istatistięi karşılařtırmalı olarak incelemiřlerdir. Uyguladıkları rakamlı örnekte tek bir outlier teřhis etmiřlerdir.

Balasooria ve Tse (1986), outlierlerin teřhisinde student türü residualleri temel alan istatistikleri tavsiye edip student türü residualleri;

$$r_i = \frac{e_i}{\hat{\sigma}(1-V_i)^{1/2}} \quad i=1,2,\dots,n$$

řeklinde göstermiřlerdir. Cook ve Weisberg'de (1982) verilen $n-k-1$ serbestlik dereceli student türü, k deęiřkenli bir istatistik;

$$t_i = \left(\frac{r_i(n-k-1)}{(n-k-r_i-2)} \right)^{1/2} \quad i=1,2,\dots,n$$

tavsiye ettiler. Outlierleri teřhis etmek için $S = \text{Max}|t_i|$ gibi bir deęeri nazara aldılar. Eęer α önem seviyesinde;

$$S > t_i(\alpha/2n, n-k-1)$$

ise i . gözlemin outlier olduęunu belirtmiřlerdir.

Akar ve řahinler (1993), standart residuallere grafik analizi uygulayarak serpme grafiklerini normal olasılık kaęıtları üzerindeki noktalarını, normal ve yarı normal grafiklerini oluřturmuřlar ve hatanın normallik varsayımını bozan iki gözlemin outlier olduęunu teřhis etmiřlerdir. Standart residualleri bulurken;

$$r_i = e_i / \text{Se}_i$$

$$\text{Se}_i = \sqrt{\text{HKO} - \text{Var}(\hat{Y}_i)}$$

ve,

$$r_i = \frac{e_i}{\sqrt{Sm_{ii}}}$$

formüllerini kullanmışlardır.

Erar (1979), student türü residuallerden, tahmin değerinin varyansından ve residuallerin varyansından oluşan bir test üzerinde durmuşlardır. Herbir i . gözlem için;

$$D_i = \frac{t_i^2}{p} \cdot \frac{V(\hat{Y}_i)}{V(R_i)}$$

kritik değerini bulmuşlar, uyguladıkları rakamlı örnekte en büyük D_i değerine sahip gözlemin outlier olduğu sonucuna varmışlardır.

Tatlıdil (1979), doğrusal regresyon problemlerindeki tek bir outlier değerinin orataya çıkarılmasını sağlayan bir test tavsiye etmişlerdir. Standartlaştırılmış residuallerin karelerinin beta dağılışı göstereceği farzedilmiş ve buna uygun bir test geliştirilmiştir. Yöntem kısmında Ellenberg'dan (1976) alınan;

$$F(Z_i) = \frac{T\left(v + \frac{1}{2}\right)}{T\left(\frac{1}{2}\right)T(v)} \cdot (1 - Z_i^2)^{v-1}$$

kullanılmıştır.

$-1 < Z_i < 1$ olup, buradan $Z_i^2 = U_i^2$ şeklinde yeni bir değişken tanımlanmıştır.

$$F(U_i) = \frac{T\left(v + \frac{1}{2}\right)}{T\left(\frac{1}{2}\right)T(v)} \cdot (U_i)^{\frac{1}{2}} \cdot (1 - U_i)^{v-1} \quad i=1,2,\dots,n$$

$0 < U_i < 1$ olur. Burada;

$$F(U_i) = \frac{T(a+b)}{T(a) \cdot T(b)} \cdot U_i^{a-1} \cdot (1 - U_i)^{b-1}$$

biçiminde $a=1/2$ ve $b = \frac{n-p-1}{2}$ parametrelili bir beta dağılımıdır. Daha sonra beta dağılımı kritik değerlerinin üst sınırlarını kullanarak C kritik değerini bulmuştur. Eğer; $\max U_i > C$ ise yokluk hipotezinin reddedileceği, $\max U_i < C$ ise de yokluk hipotezinin kabul edileceğini belirtmektedir.

2.1. RESİDUALLERİN İNCELENMESİ

Basit doğrusal bir regresyon modeli genellikle;

$$Y = X\beta + \varepsilon$$

şeklinde dir. Burada;

Y : (nx1) boyutlu şans değişkeni vektörü,

X : (nxp) boyutlu bilinen katsayı vektörü,

β : (px1) boyutlu bilinmeyen parametre vektörü,

ε : (nx1) boyutlu şans değişkeni olup, model tarafından açıklanmayan, \hat{Y}_i ile Y_i değerleri arasındaki farklar yani farklı işaretli residuallerdir[15].

Regresyon analizinde en önemli işlemlerden birisi residuallerin incelenmesidir. Residual analizi; verilerin tahmin denklemindeki değişkenleri, gözlem ve tahmin değerleri arasındaki farkın dikkatle incelenmesi gerektirir. Ayrıca verilerdeki yetersiz parametre tahminlerine neden olabilen anormal noktaları teşhis eder. Anormal verileri analize dahil edip tahmin yapan istatistikler olduğu gibi, hazırlanacak residual noktalarıyla bunların analize katılıp katılmayacağına karar verilebilir. Residuallerin analizinde birçok metodlar vardır. Fakat bunlar içinde en çok kullanılanı grafiksel metodlardır[18].

Draper ve Smith (1966), b'nin EKK tahminini yerine koyduklarında basit residualleri şu şekilde tanımlarlar;

$$Y = X\beta + e$$

$$e = Y - \hat{Y} \quad (1)$$

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

$$e = Y - X\beta \quad (2)$$

Basit residualler (1) ve (2)'de olduğu gibi her iki şekilde de ifade edilebilirler.

Cox, D.R. ve Shell E.J.'e (1968) göre basit residualler;

Residual = Gözlem Değeri - Tahmin Değeri
şeklindedir[1].

Yine [9], (1) nolu eşitlikteki basit residualleri;

$$e=(1-R)Y$$

şeklinde ifade etmişlerdir. Bu eşitlikte;

$$R =(R_{II})= X(X'X)^{-1}X'$$

olup, hat matris olarak adlandırılan simetrik ve idempotent bir matristir.

Residualler $\sigma^2(S)$ 'ye bağlı olduklarından; residuallerin incelenmesi ya S 'ye yada V 'ye göre standartlaştırılarak yapılır.

Standart residualleri;

$$r_i = \frac{e_i}{S} \quad (3)$$

$$S = \frac{\sum_{i=1}^n e_i^2}{[n - (k+1)]}$$

veya

$$r_i = \frac{e_i}{Se_i}$$

$$Se_i = \sqrt{RKO - \text{Var}(\hat{Y}_i)}$$

$$\text{Var}(\hat{Y}_i) = RKO \cdot X(X'X)^{-1}X_i$$

şeklinde ifade etmişlerdir[7],[8],[9],[18] ve [20].

e/σ standart normal dağılım göstermesine karşın, e_i/S 'nin aynı dağılıma uygun olduğu söylenemez. Standart normal dağılım gösteren e/σ elemanlarının % 95'inin -

1.96 ile +1.96 sınırları arasında olduğu ve buna benzer olarak e_i/S 'nin elemanlarının genellikle -2 ile +2 arasında olabileceği belirtilir[9].

Örnek hacminin küçük olduğu durumlarda ve outlierlerin teşhisinde hat matrisle birlikte elde edilen student türü residuallerin kullanılması daha uygundur[4].

$$t_i = \frac{e_i}{S\sqrt{1-V_{ii}}}$$

$$V_{ii} = V = X(X'X)^{-1}X'$$

matrisinin diagonal elemanlarıdır.

$$S^2 = \frac{\sum_{i=1}^n e_i^2}{n-k}$$

k : bağımsız değişken sayısıdır.

Student türü residual (3) nolu formüldeki standart residuele bağlı olarak;

$$t_i = \frac{r_i}{\sqrt{1-V_{ii}}} \quad i=1,2,\dots,n$$

şeklinde gösterilebilir[4].

2.1.1. Residuallerin Özellikleri

1. Residuallerin toplamı 0 (sıfır)'dır.

$$\sum (Y_i - \hat{Y}_i) = 0 \quad i=1,2,\dots,n$$

2. Residuallerin ortalaması 0 (sıfır)'dır.

$$\bar{e} = \frac{\sum e_i}{n} = 0 \quad i=1,2,\dots,n$$

3. n sayısındaki residualin örnek varyansı S^2 'dir.

$$S^2 = \frac{\sum_{i=1}^n e_i^2}{n-(k+1)}$$

veya

$$S^2 = \frac{RKT}{(n - k - 1)}$$

şeklinde ifade edilebilir.

4. Residuailler bağımsız şans değişkeni değillerdir. Çünkü tahmin denklemine dayanılarak elde edilebilirler[20].

5. Residuailler birbirinden bağımsızdırlar. Yani kovaryans 0 (sıfır)'dır.

2.2. HAT MATRİSİ

EKK metodunda uyumu yapılmış her \hat{Y}_i değerinde etkisi olan Y_i verisinin etkisinin anlaşılması önemli bir işlemdir. Hat matris olarak bilinen izdüşüm matrisi bu bilgileri ve student türü residuaillerle birlikte kullanıldığında anormal veri noktalarının bulunmasını sağlar. EKK ile doğrusal model tahminlerinde Y_i 'lerin \hat{Y}_i 'ler üzerindeki etkisinin hesaplanması sık sık yapılır. Gözlem değerleri \hat{Y}_i 'lere karşılık gelen tahmin edilmiş \hat{Y}_i değerleri için aralarındaki ilişki kolaylıkla hesaplanabilir. Buda X değişkenleri arasındaki çok varyanslı anormal noktaları ortaya çıkartır. İşte bu bilgileri hat matris verebilir[17].

Doğrusal bir model;

$$Y_{n \times 1} = X_{n \times p} b_{p \times 1} + e_{n \times 1} \quad (1)$$

şeklindedir.

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

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

$$V = X(X'X)^{-1} X' \quad (2)$$

olduğunda; tahmin değeri;

$$\hat{Y} = V \cdot Y \quad (3)$$

$$e = Y - \hat{Y}$$

V : Hat matris olarak bilinen izdüşüm matrisi.

Analizlerde V matrisinin köşegen elemanları V_{ii} 'ler kullanılır. Hat matris (2)'den de anlaşılacağı üzere bağımsız değişkenlere bağlı olarak bulunur. Bir izdüşüm matrisi olan V , simetrik ve idempotenttir.

$$(V^2=V)$$

Hat matrisin V_{ii} değerleri, $0 \leq V_{ii} \leq 1$ 'dir. Hat matrisin köşegen elemanları olarak adlandırılan V_{ii} değerleri gözlem uzaklığı olarak adlandırılır ve i . gözlem noktasının değişkenlerin oluşturduğu elipsoid merkezine uzaklığının bir ölçüsüdür. Residualer her zaman gerçek sapmayı vermezler. Bu sebeple residual terimlerinin S^2 ve $(1-V_{ii})$ değerleri ile standartlaştırılması gerekir. İşte bu şekilde student türü residualer ortaya çıkar.

$$t_i = \frac{e_i}{S\sqrt{1-V_{ii}}}$$

e : Basit residualer

S^2 : Residualerin kare ortalamasıdır.

Bir çok araştırmacı V_{ii} 'nin model (1)'in üzerine kurulmuş bir analizi anlamada rol oynadığını belirtmişlerdir. Örneğin; eğer $\max(V_{ii})$ 1'den dikkate alınmayacak kadar küçük ise residualer incelendiğinde bir anormal gözlemi belirtmeyecektir. Box ve Draper'e göre V_{ii} 'nin ortalaması p/n 'dir. Dolayısıyla;

$$\max(V_{ii}) \geq \frac{p}{n} \text{ 'dir.}$$

Buna göre n ; p 'den dikkate alınabilecek büyüklükte olmadıkça anormal gözlem durumlarını tanımlamak zordur[17].

Hat matrise bağlı olarak residualerin ve \hat{Y}_i tahmin değerlerinin varyansıda bulunabilir[17].

$$\text{Var}(\hat{Y}) = \sigma^2 V = \sigma^2 (X(X'X)^{-1}X')$$

$$\text{Var}(e) = \sigma^2(1-V) = \sigma^2(1-(X(X'X)^{-1}X'))$$

2.3. ANORMAL GÖZLEMLER (OUTLIERLER)

Verilerin değerlendirilmesinde kullanılan metodlar ne kadar sağlam olursa olsun, analizlerin uygulanacağı rakamlar güvenilir olmadıkça verilecek hükümler hatadan arınmış olamaz. Bazen elde edilen verilerden bir veya birkaçı diğerlerine nazaran oldukça sapmalı bir görünümde olabilirler. Gerek orijinal verilerin içerisinde ve gerekse de uydurulan matematik modele göre tahmin edilen hata terimlerinin mutlak değerinin çok büyük olması, gözlemin anormal olabileceğini gösterir. Ancak bu gözlemin kesinlikle anormal olup olmadığına karar vermek üzere bir kısım testler yapılmalıdır[24].

Anormal gözlemler (outlierler) ekstrem (uç) değerlerdir. Outlierler en geniş residuale sahip ve residual grafiklerinde sıfır çizgisinden 3-4 veya daha fazla standart sapma kadar uzaklıkta olan gözlemlerdir[7],[20],[21].

Outlierlerin bulunması konusu regresyon eşitliğinin açıklamadığı miktar anlamına gelen hata terimleriyle yakından ilgilidir. Dolayısıyla residual veya standart residual grafiklerinde anormal gözlemler teşhis edilebilir.

Araştırmacılar veri setlerindeki geniş hacimli anormal gözlemler için tarihsel gelişim içerisinde;

- outliers : dış değer
- discordant observations : uyumsuz gözlemler
- rogue value : kural dışı değer
- contaminants : lekeliler
- surprising values : şüpheli değerler
- mavericks : başıboşlar
- dirty data : çirkin veri gibi terimler kullanılmıştır[6].

Diğer analizlerde olduğu gibi doğrusal regresyon problemlerinde de gözlemler arasında anormal gözlemlerin bulunması araştırmacılar için istenmeyen bir durumdur. Anormal gözlemler regresyon işlemleri için; “beklenen değerinden çok farklı bir gözlem değerine sahip veya en büyük residual değerine sahip gözlem” şeklinde tanımlanabilir[23].

Veri setleri içerisindeki şüpheli gözlemlerin anormal gözlem olup olmadıklarının teşhisi konusunda birçok metodlar geliştirilmiştir. Bunlar içerisinde; standardize edilmiş residual grafikleri, Cook istatistiği, Dfits istatistiği, Dfbetas istatistiği, A-P istatistiği, katkı kareler toplamı testi kullanılan testler arasındadır. Teşhis edilen bu outlierlerin analizden çıkarılması gerekmektedir. Fakat otomatik

yoketme her zaman akılcı bir prosedür değildir. O sebeple yukarıda sayılan istatistik testlerden biri uygulanmalıdır. Bir veri setinde bir veya daha fazla parametre az sayıda gözleme dayanılarak tahmin edilirse ortaya problemler çıkabilir. Eğer bir veri setinin modeli uydurulmak ve parametre tahminleri yapılmak istenirse, denemelerde genellikle ortaya çıkan bir veya iki gözlemin iptal edilmesi gerekir. Böyle bir durumda örnek büyüklüğü düşeceğinden, denemelerde bu tür problemlerle karşılaşmamak için örnek büyüklüğü mümkün olduğu kadar geniş tutulmalıdır.

Outlierlerin ortaya çıkış sebepleri şöyle sıralanabilir :

1. Regresyon tahmin modelinin yetersiz olması,
2. Ölçüm, tartım ve bazı kayıtlama hataları,
3. Gerekli transformasyonların yapılmaması,
4. Örneklemenin tesadüfiliğinden,
5. Örneklemedeki bazı gözlemlerin, diğer gözlemlerden farklı etkenlerin etkisinde kalmış olmasındandır[13].

Eğer veriler içerisinde anormal konumda olan gözlem sayısı 2-3-4 gibi ise bu gözlemlerdeki farklılık gerçek varyasyonuda yansıtabilir[5].

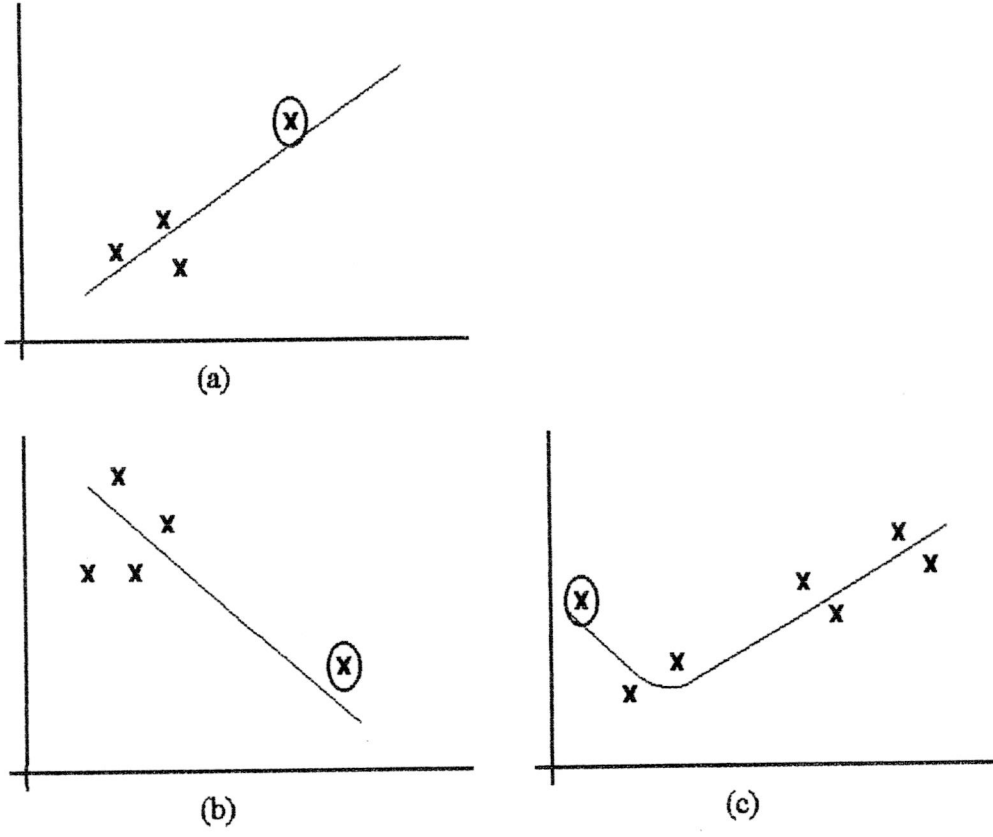
Yabancı bir etkenin etkisinde kalan veya bir yanlışlıktan dolayı ortaya çıkan bir gözlem outlier ise yanılta bir uyuma sebep olabilir. Ayrıca modelde belirtilmeyen başka bir bağımsız değişkenin interaksyonu ile bir outlier meydana geldiği zaman, bu bilinmeyen bağımsız değişkenin teşhisinde outlierler önemli bir bilgi ifade ederler[21].

Uydurulan modelimizde bir adet bağımsız değişken var, fakat basit residualler yüksek çıktı. Birden fazla anormal gözlemler meydana çıkar. Böyle bir durumda gözlemler ikinci bir bağımsız değişkenin etkisinde kalmış demektir.

Bu çalışmada outlierlerin teşhisinde kullanılmak üzere üç metod incelenmiştir. Bunlar;

1. Standardize Edilmiş Residual Grafikleri Yöntemi (Plot Yöntemi),
2. Cook İstatistiği,
3. Katkı Kareler Toplamı Testi.

Teşhisi yapılan anormal gözlemler bir tane dahi olsa analiz dışına atılmalıdır. Tek bir değer dahi regresyon doğrusunun yönünü ve şeklini değiştirebilmektedir. Bu durum şekil 2.'de açıkça görülmektedir.



Şekil 2. Ekstrem değerlerin regresyon doğrusuna etkileri.

3. MATERYAL VE YÖNTEM

3.1. MATERYAL

Bu çalışmada materyal olarak Harran Üniversitesi Ziraat Fakültesi; Döner Sermaye İşletmesindeki Zootekni Anabilim Dalı denemelerinde kullanılan bildircinlerin yumurtalarına ait; yumurta boyu (mm) ve yumurta ağırlıkları (gr) kullanılmıştır.

Çizelge 1. Bildircinlerden Elde Edilen Yumurta Boy Değerleri (X bağımsız değişkeni) ve Yumurta Ağırlıkları (Y bağımlı Değişkeni).

Gözlem No	Yumurta Boyları (mm)	Yumurta Ağırlıkları (gr)	Gözlem No	Yumurta Boyları (mm)	Yumurta Ağırlıkları (gr)
1	28.04	8.92	16	32.02	11.03
2	28.10	9.01	17	32.12	11.27
3	28.64	9.13	18	32.24	11.27
4	29.50	10.21	19	32.26	11.61
5	29.62	10.23	20	32.64	11.81
6	29.78	10.27	21	32.87	11.96
7	30.42	10.11	22	33.06	11.64
8	30.68	10.17	23	33.15	11.92
9	30.94	10.68	24	33.19	11.82
10	31.36	10.40	25	33.38	11.83
11	31.38	10.55	26	33.43	12.08
12	31.45	10.61	27	33.64	12.55
13	31.62	10.98	28	34.94	15.04
14	31.78	10.70	29	34.97	12.82
15	31.98	10.82	30	35.91	13.13

3.2. YÖNTEM

Veri setlerindeki geniş residuallere sahip gözlemlerin outlier olup olmadığının teşhisi konusunda birçok yöntemler geliştirilmiştir.

Bu çalışmada pratikliği ve kolaylığı bakımından kullanışlı olan “*Plot Yöntemi*”, outlierlerin teşhisinde güvenilirliğin bir göstergesi olan student türü residualleri(t_i) temel alan istatistiklerden olan “*Cook İstatistiği*” ve bu iki metodun yanında residuallerin kareleri toplamında gözlemlerin etki paylarını temel alan “*Katkı Kareler Toplamı Testi*” de üçüncü bir yöntem olarak incelenmiştir.

3.2.1. Standardize Edilmiş Residual Grafikleri Yöntemi İle Outlierlerin Teşhisi (Plot Yöntemi)

Outlierlerin bulunması konusu regresyon eşitliğinin açıklayamadığı miktar anlamına gelen residuallerle yakından ilgilidir. Dolayısıyla outlierlerin teşhisinde residual analizlerinden yararlanılacaktır.

Residual analizlerinde genellikle, kolaylığı ve parametre tahminlerine neden olan, verilerdeki herhangi bir anormalliği ortaya çıkarması nedeniyle mümkün olduğunca grafiksel metodlar kullanılır. Fakat çizilen grafiklerin hangisinde olursa olsun, verilerdeki anormallikleri belirleyebilmek için biraz tecrübeye ihtiyaç vardır[1].

Ascombe ve Tukey'e (1963) göre residualler;

Residual = Gözlenen Değer - Tahmin Değeri

şeklinde tanımlanır[2].

$$e_i = Y_i - \hat{Y}_i \quad i=1,2,\dots,n \quad (1)$$

e_i değerleri S^2 ile alakalı olduklarından; e_i değerleri S^2 'ye bağlı olarak standartlaştırılırlar.

Balasoorya ve Tse'da (1986) outlierlerin incelenmesinde standart residuallerden ziyade student türü residuallerin kullanılmasının daha avantajlı olduğu belirtilmiştir.

$$t_i = \frac{e_i}{S\sqrt{1-V_{ii}}} \quad i=1,2,\dots,n \quad (2)$$

- e_i : basit residualler,
 S : residuallere ait standart sapma,
 V_{ii} : V matrisinin diagonal elemanları,
 t_i : i. student türü residualler.

Eğer gözlemler belirli bir zaman süreci içerisinde elde edilmişlerse; student türü residualler zaman serilerine karşılıklı noktalanabilir.

Montgomery ve Peck (1982), student türü residualleri değil de standart residualleri kullanmışlardır.

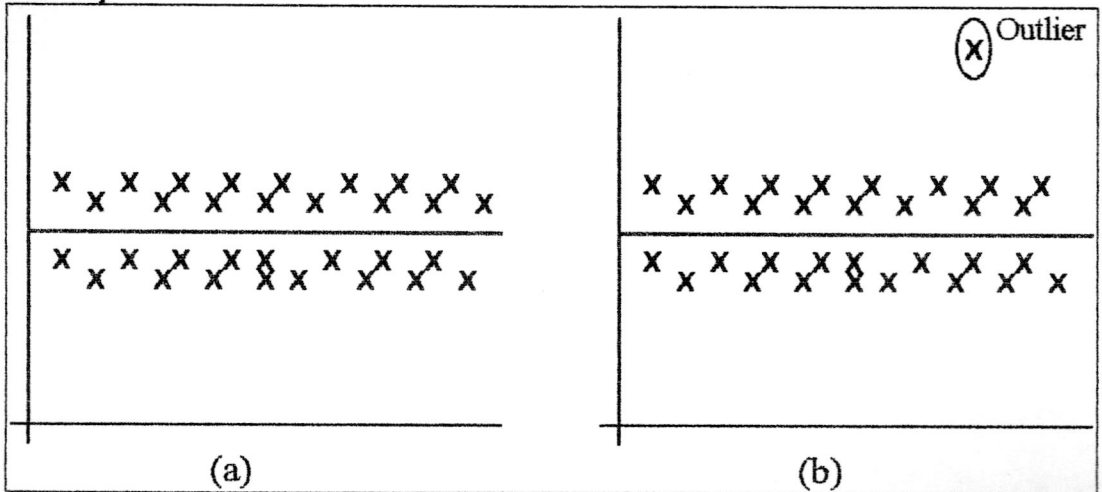
Residual grafikleri genellikle bağımsız değişken (X_i) ve tahmin değerine (\hat{Y}_i) karşı noktalanırlar. Bu noktalar; düz tahmin çizgisi yetersizliği, sürekli varyans eksikliği, outlierlerin varlığı ve karşılıklı etkileşen hatalar gibi işlemlerde hareket noktasıdır[7].

(2) nolu formüle göre student türü residualleri hesaplayan (*Minitab 8.2*) bir bilgisayar paket programı da hazırlanmıştır. Bu çalışmada bu hesaplanan student türü residualler;

- 1 - Tahmin edilen (\hat{Y}_i) değerlerine karşı,
- 2 - Bağımsız (X_i) değişken değerlerine karşı,

noktalanmıştır[9].

Hazırlanan bu noktalarla hangi gözlemlerin outlier olduğu teşhis edilebilir. Fakat, yukarıda da belirtildiği gibi bu teşhis yöntemi biraz tecrübe gerektirmektedir. Şekil 3.'de anormal veri noktalarının bulunduğu ve bulunmadığı residual grafikleri verilmiştir.



Şekil 3. a) hiçbir anormal gözlemin bulunmadığı residual grafiği,

b) tek bir anormal gözlemin bulunduğu residual grafiği.

3.2.2. Cook İstatistiği

Tam ranklı doğrusal modellerde, parametre vektörlerinin EKK tahminlerinin belirlenmesinde her veri noktasının katkısının anlaşılabilmesi için geliştirilen yeni bir ölçümdür[8].

Genellikle outlierlerin teşhisinde student türü residualleri temel alan istatistikler daha kullanışlıdır[4]. Cook İstatistiği bunlardan biridir.

Genel regresyon modeli olarak;

$$Y = X\beta + e \quad (1)$$

modeli kabul edilir.

Y : (nx_1) boyutlu gözlemler boyutlu gözlemler vektörü,

x : $(n \times p)$ boyutlu tam ranklı matris,

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

e : $(n \times 1)$ boyutlu hata terimleri vektörü.

EKK metoduna göre β 'nin tahmin değeri;

$$\hat{\beta} = (X'X)^{-1} X'Y \quad (2)$$

şeklindedir.

$$e = Y_i - \hat{Y}_i$$

$$e = Y - X\hat{\beta} \quad (3)$$

şeklinde yazılabilir. β 'nin (2)'deki değeri (3)'de yerine konursa;

$$e = Y - X \cdot (X'X)^{-1} X'Y$$

$$e = (I - X \cdot (X'X)^{-1} X')Y \quad (4)$$

olur. \hat{Y}_i ile e_i 'lerin kovaryans matrisleri;

$$V(\hat{Y}) = X(X'X)^{-1}X' \sigma^2 \quad (5)$$

$$V(e_i) = (1 - X(X'X)^{-1}X')\sigma^2 \quad (6)$$

şeklinde gösterilir.

Normal dağılımın gereği olarak, bilinmeyen β için % $(1-\alpha)$ 100'lük güven elipsoidi β^* ile gösterilir ve aşağıda ifade edildiği gibi vektörler kümesidir[4].

$$\frac{(\beta^* - \hat{\beta})' X'X(\beta^* - \hat{\beta})}{PS^2} \leq F(1-\alpha, p, n-p) \quad (7)$$

$$S^2 = e'e / n-p$$

olup, burada;

p : parametre sayısı,

n : örnek büyüklüğüdür.

$F(1-\alpha, p, n-p)$ ise p ve $n-p$ serbestlik dereceli F tablosunda $1-\alpha$ güven düzeyindeki cetvel değeridir[16].

β 'nin tahmininde herhangi bir verinin etki derecesini bulmak için yine (7) nolu eşitlikten faydalanılır. İlk işlem β 'yi tahmin edip kaydetmektir. Daha sonra etkisi araştırılacak gözlem silinir ve $\hat{\beta}_{(-i)}$ tahmin edilir. $\hat{\beta}_{(-i)}$ ile $\hat{\beta}$ arasındaki uzaklık i . gözlemin etki payıdır. Bu yolla herbir gözlemin $\hat{\beta}$ ile arasındaki uzaklık yani herbir veri noktasının kritik yapısının değeri bulunmuş olur[8]. Bu kritik yapı;

$$D_i = \frac{(\hat{\beta}_{(-i)} - \hat{\beta})' X'X(\hat{\beta}_{(-i)} - \hat{\beta})}{PS^2} \quad i=1,2,\dots,n \quad (8)$$

şeklinde ifade edilir. D_i değeri i . gözlemin çıkarılmasından dolayı regresyon katsayılarının oluşturduğu elipsoitteki değişimi yani $\hat{\beta}_{(-i)}$ ile $\hat{\beta}$ arasındaki uzaklığın önem seviyesi cinsinden değerini verir. Ayrıca, $\hat{\beta}_{(-i)}$ ile $\hat{\beta}$ arasındaki uzaklık;

$$\hat{\beta} - \hat{\beta}_i = (X_{-i}' X_{-i})^{-1} X_{-i}' [Y_i - X_i \hat{\beta}] \quad (9)$$

şekinde gösterilebilir.

X_{-i} : i. sıranın çıkarılmasıyla elde edilen matris

X_i : çıkarılan i. sırayı gösterir.

EKK'ya göre uydurulan modelde her \hat{Y}_i tahmin değerine etki eden herbir Y_i gözlem değerinin etkisinin anlaşılması önemli bir konudur. Hat matris bu etkilerin anlaşılmasında önemli yer tutar.

$$V_{ii} = X_i' (X'X)^{-1} X_i \quad (10)$$

koşulunda X_i 'nin $(n \times 1) \times p$ boyutlu tam ranklı olduğu düşünülürse;

$$(X_{-i}' - X_{-i}')^{-1} = (X'X)^{-1} + (X'X)^{-1} X_i X_i' (X'X)^{-1} / (1 - V_{ii}) \quad (11)$$

olacaktır. Her iki taraf X_i ile çarpılırsa;

$$(X_{-i}' - X_{-i}')^{-1} X_i = (X'X)^{-1} X_i + (X'X)^{-1} X_i X_i' (X'X)^{-1} X_i / (1 - V_{ii}) \quad (12)$$

(10) nolu formüldeki V_{ii} , (12)'deki eşitinin yerine konursa;

$$(X_{-i}' - X_{-i}')^{-1} X_i = (X'X)^{-1} X_i + (X'X)^{-1} V_{ii} X_i / (1 - V_{ii})$$

olur. Eşitliğin sağ tarafında paydalar eşitlenirse;

$$(X_{-i}' - X_{-i}')^{-1} X_i = (X'X)^{-1} X_i + \frac{(X'X)^{-1} V_{ii} X_i}{(1 - V_{ii})}$$

$$(X_{-i}' - X_{-i}')^{-1} X_i = \frac{(1 - V_{ii})(X'X)^{-1} X_i + (X'X)^{-1} V_{ii} X_i}{(1 - V_{ii})}$$

V_{ii} 'lerin çarpanları aynı olduğundan ortak paranteze alınabilir;

$$(X_{-i}' - X_{-i}')^{-1} X_i = \frac{(X'X)^{-1} X_i (1 - V_{ii} + V_{ii})}{(1 - V_{ii})}$$

$$(\mathbf{X}_{-i}' - \mathbf{X}_{-i})^{-1} \mathbf{X}_i = \frac{(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_i}{(1-V_{ii})} \quad (13)$$

Elde edilen (13) nolu denklem (9) nolu denklemdeki eşitinin yerine konulursa;

$$\hat{\beta} - \hat{\beta}_i = \frac{(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_i}{(1-V_{ii})} [\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}]$$

olur. Oluşan bu denklem (8) nolu D_i formülündeki yerine konur;

$$D_i = \frac{\left[\frac{(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_i}{(1-V_{ii})} [\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}] \right]' \mathbf{X}'\mathbf{X} \left[\frac{(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_i}{(1-V_{ii})} [\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}] \right]}{PS^2}$$

köşeli parentezler kaldırılıp \mathbf{X} vektörleri yanyana getirilirse;

$$D_i = \frac{[\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}] \mathbf{X}_i' (\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{X}) (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_i [\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}] / (1-V_{ii})^2}{PS^2}$$

ve

$$(\mathbf{X}'\mathbf{X})^{-1} (\mathbf{X}'\mathbf{X}) = 1$$

$$V_{ii} = \mathbf{X}_i' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}_i$$

olduğundan;

$$D_i = \frac{[\mathbf{Y}_i - \mathbf{X}_i \hat{\beta}]^2 V_{ii}}{PS^2 (1-V_{ii})^2} \quad (14)$$

(14) nolu denklem;

$$D_i = \left[\frac{[Y_i - X_i \hat{\beta}]}{S\sqrt{1-V_{ii}}} \right]^2 \cdot \frac{V_{ii}}{P(1-V_{ii})}$$

şeklinde yazılabilir.

$$e_i = [Y_i - X_i \hat{\beta}]$$

olduğu hatırlanırsa;

$$D_i = \left[\frac{e_i}{S\sqrt{1-V_{ii}}} \right]^2 \cdot \frac{V_{ii}}{P(1-V_{ii})}$$

Student türü residualler;

$$t_i = \frac{e_i}{S\sqrt{1-V_{ii}}}$$

olduğundan[4];

$$D_i = \frac{t_i^2 V_{ii}}{P(1-V_{ii})} \quad (15)$$

olur. Denklemi daha da basite indirgersek;

$$D_i = \frac{t_i^2 X_i' (X'X)' X_i}{P(1-X_i' (X'X)' X_i)}$$

pay ve payda S^2 ile çarpılınca (5) ve (6) nolu formullerdeki kovaryans matrisleri elde edilir.

$$D_i = \frac{t_i^2 (V(\hat{Y}))}{P(V(e_i))}$$

p : parametre sayısı,
 t_i : Student türü residual,
 $p=k+1$ olduğundan

$$D_i = \frac{t_i^2 (V(\hat{Y}))}{(k+1)(V(e_i))}$$

olarak yazılabilir.

(15) nolu formüle göre hesaplanan D_i kritik değerleri $F(1-\alpha, p, n-p)$ cetvel değeri ile karşılaştırılır. Bu çalışmada kritik değerlerin hesaplanmasında (15) nolu formül esas alınmıştır.

3.2.3. Katkı Kareler Toplamı Testi

Anormal ve etkili gözlemlerin bulunması konusu, regresyon eşitliğinin açıklayamadığı miktar anlamına gelen residual terimleriyle yakından ilgilidir. Dolayısıyla residuallerin kareler toplamı her gözlemin etki paylarının toplamından meydana gelmiştir.

$$RKT = \sum_{i=1}^n e_i^2$$

$$RKO = S^2 = \frac{RKT}{n-k-1}$$

Outlier değer olup olmadığı konusunda kuşkulanılan her bir i . gözlemin çıkarılmasıyla RKT'de belirli bir azalma olacaktır. Bu azalma;

$$RKT - RKT_{-i}$$

RKT_{-i} : i . gözlem çıkarıldıktan sonra residual kare toplamı olup;

$$RKT_{-i} = \hat{\sigma}_{-i}^2 (n-k-2) \quad (1)$$

$$\hat{\sigma}_{-i}^2 = S_{-i}^2 = S^2 (n-k-1-t_i^2) / (n-k-2) \quad (2)$$

S_{-i}^2 : i . gözlem çıkarıldıktan sonraki örnek varyansdır.

t_i : i . Student türü residual

böylece herhangi bir i . gözlemin etki payı;

$$F_i = \frac{RKT - RKT_{-i}}{S_{-i}^2} \quad (3)$$

olarak ifade edilir[13].

Bu yöntemde (3) nolu formüle göre bulunan kritik F_i değeri 1 ve $n-k-2$ serbestlik dereceli α/n yanılma düzeyindeki F cetvel değeri [22] ile karşılaştırılır. Eğer birden fazla outlierin varlığından şüphe ediliyorsa herbir şüpheli gözlem için işlemler tekrarlanır.

4. ARAŞTIRMA SONUÇLARI

Materyal kısmında verilen (Çizelge 1.) değerler kullanılarak EKK tahmin modeli tahmin edilirse;

$$\hat{Y} = -8.46 + 0.616X_i$$

olur. Buradan;

$$\hat{Y} = -8.46 + 0.616X_i + e_i$$

elde edilir. Burada;

X_i : yumurta boyu,

Y_i : yumurta ağırlığı,

e_i : modelin açıklayamadığı kısım yani residuallerdir.

Verilere ait varyans analizi sonuçları Çizelge 2.'de verilmiştir. Elde edilen analiz sonuçlarına göre bıldırcınların yumurta boyu ile yumurta ağırlığı arasındaki ilişki istatistik bakımından önemlidir ($P < 0.01$). Örneğe ait gerek belirtme katsayısı gerekse de düzeltilmiş belirtme katsayıları oldukça yüksek çıkmıştır.

$$R^2 = \% 87.0$$

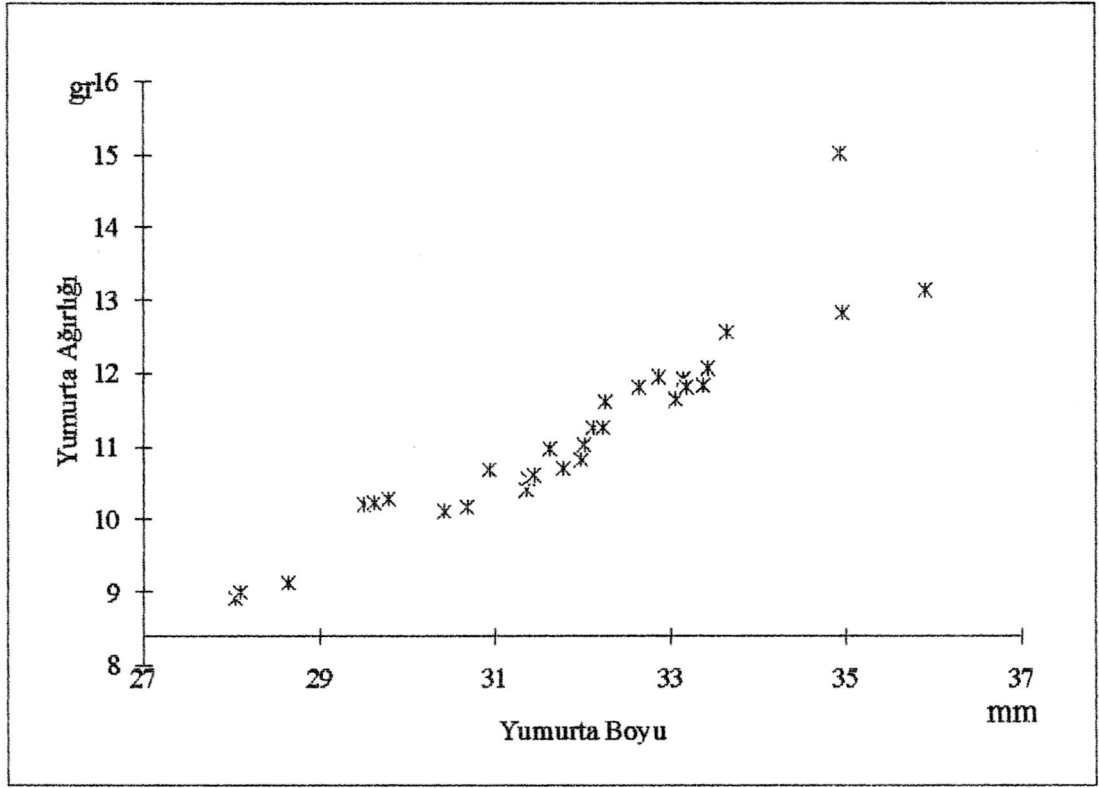
$$(\text{düz}) R^2 = \% 86.6$$

Buradan anlaşılan yumurta boyları, yumurta ağırlıklarındaki varyasyonun % 87'sini açıklayabilmektedir.

Çizelge 2. Bıldırcınların yumurta boyları ile yumurta ağırlıkları arasındaki ilişkiye ait varyans analiz tablosu.

V.K.	S.D.	K.T.	K.O.	F	P
Regresyon	1	41.528	41.528	188.05**	0.000
Hata	28	6.183	0.221		
Genel	29	47.718			

$$S = 0.4699 \quad R^2 = \% 87.0 \quad \text{düz } R^2 = \% 86.6$$



Şekil 4. Bildircinlerden elde edilen verilere ait serpme diyagramı.

Doğrusal regresyon denkleminin değişkenlere uygunluğunu anlamak için regresyon doğrusu grafiği veya verilerin serpme diyagramı hazırlanır. Şekil 4.'de bildircinlardaki yumurta boyu ile yumurta ağırlığı arasındaki ilişkiyi ifade eden noktaların dağılımı görülmektedir.

4.1. PLOT YÖNTEMİ

Yöntem kısmında belirtildiği gibi outlierin teşhisinde standardize edilmiş residuallerin grafikleri en çok kullanılan ve pratik olan yöntemdir. Grafiklerde 3.2.1.'de verilen (2) nolu formülde verilen student türü residualler kullanılmıştır.

$$t_i = \frac{e_i}{S\sqrt{1-V_{ii}}} \quad i=1,2,\dots,n$$

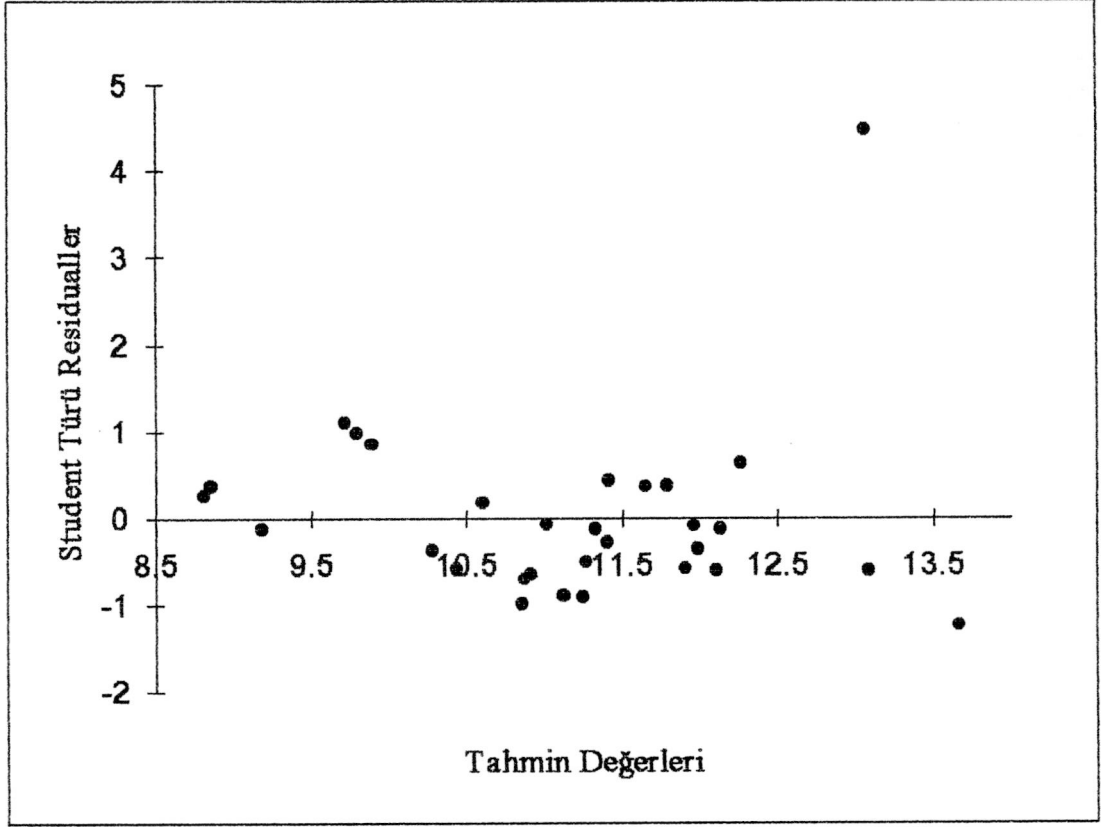
Elde edilen bu student türü residualler \hat{Y}_i tahmin değerlerine ve X_i bağımsız değişkenine karşılık noktalanır. Grafikler hazırlanırken X ekseninde X_i değerleri, Y ekseninde student türü residualler kullanılır.

Bu grafikler hazırlanırken kullanılacak olan X_i , \hat{Y}_i , t_i ve V_{ii} değerleri Çizelge 3.'de verilmiştir.

Çizelge 3. Bildiricilerden elde edilen X_i (yumurta boyu), Y_i (gözlem değeri), \hat{Y}_i (tahmin değeri) ve bunlardan hesaplanan e_i (basit residualler), t_i (student türü residual) ve V_{ii} (hat matrisin diagonal elemanları) değerleri

Gözlem No	X_i	Y_i	\hat{Y}_i	e_i	t_i	V_{ii}
1	28.0	8.9200	8.8134	0.1066	0.25	0.16507
2	28.1	9.0100	8.8503	0.1597	0.37	0.16094
3	28.6	9.1300	9.1830	-0.0530	-0.12	0.126726
4	29.5	10.2100	9.7127	0.4973	1.11	0.083238
5	29.6	10.2300	9.7867	0.4433	0.98	0.078245
6	29.8	10.2700	9.8852	0.3848	0.85	0.071996
7	30.4	10.1100	10.2795	-0.1695	-0.37	0.05168
8	30.7	10.1700	10.4396	-0.2696	-0.59	0.045565
9	30.9	10.6800	10.5998	0.0802	0.17	0.040685
10	31.4	10.4000	10.8585	-0.4585	-0.99	0.035412
11	31.4	10.5500	10.8708	-0.3208	-0.70	0.035242
12	31.5	10.6100	10.9139	-0.3039	-0.66	0.034702
13	31.6	10.9800	11.0187	-0.0387	-0.08	0.033764
14	31.8	10.7000	11.1172	-0.4172	-0.90	0.033363
15	32.0	10.8200	11.2404	-0.4204	-0.91	0.03352
16	32.0	11.0300	11.2651	-0.2351	-0.51	0.033639
17	32.1	11.2700	11.3267	-0.0567	-0.12	0.034065
18	32.2	11.2700	11.4006	-0.1306	-0.28	0.034817
19	32.3	11.6100	11.4129	0.1971	0.43	0.034968
20	32.6	11.8100	11.6470	0.1630	0.35	0.039225
21	32.9	11.9600	11.7887	0.1713	0.37	0.043084
22	33.1	11.6400	11.9057	-0.2657	-0.58	0.047001
23	33.2	11.9200	11.9611	-0.0411	-0.09	0.049086
24	33.2	11.8200	11.9858	-0.1658	-0.36	0.05006
25	33.4	11.8300	12.1028	-0.2728	-0.60	0.055088
26	33.4	12.0800	12.1336	-0.0536	-0.12	0.056521
27	33.6	12.5500	12.2630	0.2870	0.63	0.063038
28	34.9	15.0400	13.0638	1.9762	4.49R	0.121315
29	35.0	12.8200	13.0823	-0.2623	-0.60	0.123024
30	35.9	13.1300	13.6613	-0.5313	-1.25	0.184918

Çizelge 3.'de verilen değerlere göre yöntem kısmında verilen grafikleri çizelim.



Şekil 5. Tahmin değerlerine karşılık student tüür residuallerin grafiđi.

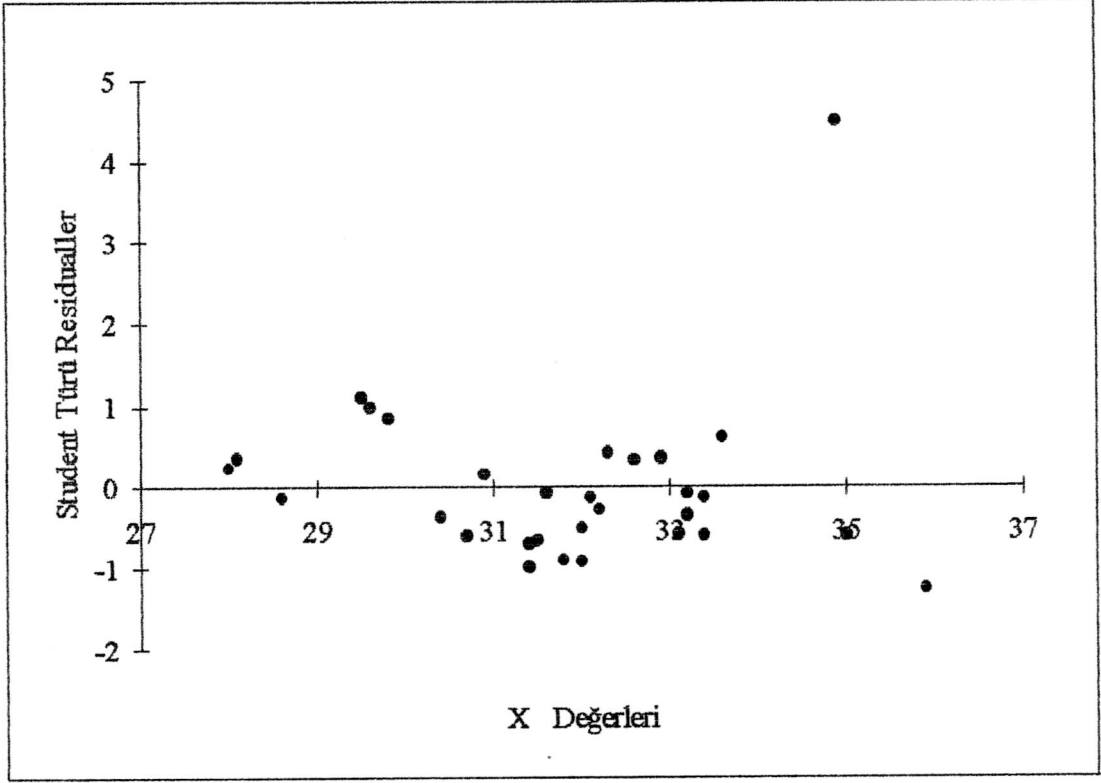
4.1.1. Tahmin Edilen \hat{Y}_i Deđerlerine Karşı Student Tüür Residuallerin Grafikleri

Bu grafikte Y eksenini (t_i) student tüür residualleri, X ekseninde (\hat{Y}_i) tahmin deđerlerini gösterir. Bu tür noktaların düz bir bant görünümünde olması beklenir. Noktalar 0 çizgisi üzerinde paralel bir şekilde uzanırlar. Örneđimize ait ; \hat{Y}_i deđerlerine karşı student tüür residuallerin grafikleri Şekil 5.'de verilmiştir.

4.1.2. Bađımsız X_{ij} Deđerliken Deđerlerine Karşılık Student Tüür Residuallerin Grafikleri

Bu tür bir grafikte; Y eksenini (t_i) student tüür residualleri, X eksenini de X_{ij} bađımsız deđerlikenleri gösterir. Bu grafiklerin de \hat{Y}_i deđerlerine karşılık hazırlanan plotlarda olduđu gibi düz bir bant görünümünde olması beklenir. Örneđimize ait; X_{ij}

bağımsız değişkenlerine karşılık student türü residuallerin grafikleri Şekil 6.'da verilmiştir.



Şekil 6. X_i değerlerine karşılık student türü residuallerin grafiği.

4.2. COOK İSTATİSTİĞİNİN UYGULAMASI

Student türü residuallerin kısım 3.2.1.'de gösterilen grafiğinde 28. gözlemin anormal olabileceği ortaya çıkmıştır. Aynı zamanda gözlemler için student türü residualleri ve hat matrisleri kullanarak elde edilen istatistiklerde kısım 3.2.2.'de verildiği gibi kritik değerler elde etmek mümkündür. Çizelge 4.'de Cook istatistiğinin kritik değerlerini (1) hesaplamada kullanılacak; (t_i) student türü residualler(2), hat matrisin diagonal elemanları (3) ve hesaplanan Cook istatistik değerleri verilmiştir.

$$D_i = \frac{t_i^2 V_{ii}}{P(1-V_{ii})} \quad (1)$$

$$t_i = \frac{e_i}{S\sqrt{1-V_{ii}}} \quad (2)$$

$$V_{ii} = X_i' (X'X)^{-1} X_i \quad (3)$$

D_i : Cook istatistiğinin kritik değeri

t_i : Student türü residualleri

p : Parametre sayısı

V_{ii} : Hat matrisin diagonal elemanları

Çizelge 4. Bıldırıcınlardan elde edilen 30 gözleme ait (t_i) student türü residualler, (V_{ii}) hat matrisin diagonal elemanları ve (D_i) Cook İstatistiği kritik değerleri

Gözlem No	t_i	V_{ii}	D_i
1	0.25	0.16507	0.00610
2	0.37	0.16094	0.01320
3	-0.12	0.126726	0.00106
4	1.11	0.083238	0.05545
5	0.98	0.078245	0.04098
6	0.85	0.071996	0.02803
7	-0.37	0.05168	0.00374
8	-0.59	0.045565	0.00823
9	0.17	0.040685	0.00064
10	-0.99	0.035412	0.01812
11	-0.7	0.035242	0.00882
12	-0.66	0.034702	0.00779
13	-0.08	0.033764	0.00012
14	-0.9	0.033363	0.01407
15	-0.91	0.03352	0.01436
16	-0.51	0.033639	0.00451
17	-0.12	0.034065	0.00027
18	-0.28	0.034817	0.00144
19	0.43	0.034968	0.00330
20	0.35	0.039225	0.00256
21	0.37	0.043084	0.00313
22	-0.58	0.047001	0.00827
23	-0.09	0.049086	0.00021
24	-0.36	0.05006	0.00345
25	-0.6	0.055088	0.01040
26	-0.12	0.056521	0.00041
27	0.63	0.063038	0.01339
28	4.49	0.121315	1.38932
29	-0.6	0.123024	0.02491
30	-1.25	0.184918	0.17790

Çizelge 4.'deki gözlemler içinde en geniş residuele ve kritik değere sahip olan 28. gözlemin outlier olup olmadığını anlamak için $F(1-\alpha, p, n-p)$ cetvel değeri ile karşılaştırılır. $p=2$, $n=30$, $\alpha=0.10$ alınmıştır. $\alpha=0.10$ için $F(0.90, 2, 28)=0.105775$ olduğundan $D_{28}>0.105775$ 'dir. Böylece 28. değer outlier olduğuna karar verilir. 30. gözlem olan $D_{30}=0.17790$ değeri de dikkat çekmektedir. Bunun anormal gözlem olup olmadığını anlamak için 28. gözlem çıkarıldıktan sonra tekrar Cook istatistikleri hesaplanır. Fakat 28. gözlem çıkarıldığında, analiz sonucunda en geniş residuele sahip gözlem olarak 27. gözlem dikkat çekmektedir.

4.3. KATKI KARELER TOPLAMI TESTİNİN UYGULAMASI

Çizelge 3.'deki gözlemler içinde en geniş residuele sahip gözlem 28. gözlemdir. Bu değer outlier olup olmadığını anlamak için bu değeri analizden çıkardıktan sonra tekrar model uydurup parametre tahminlerini yapmak gerekir. Dolayısıyla 28. gözlemin çıkarılmasıyla RKT'de belirli bir sapma olacaktır. Bu sapma 28. gözlemin outlier olup olmadığı belirleyecektir. Bu sapmadan faydalanarak kısım 3.2.3.'deki (3) nolu formül kullanılarak 28. gözlemin outlier olup olmadığı anlaşılacaktır. 28. gözlem çıkarıldıktan sonraki veriler kullanıldığında EKK regresyon doğrusu tahmini;

$$\hat{Y}_i = -6.50 + 0.55X_i$$

olur ve buradan da;

$$Y_i = -6.50 + 0.55X_i + e_i$$

eşitliği elde edilir. Bu yeni örneğe ait varyans analiz tablosu Çizelge 5.'de verilmiştir.

28. gözlem çıkarıldıktan sonra yine regresyon katsayısı önemli bulunmuştur. Geniş residuele sahip olan ve outlier olmaları muhtemel bazı gözlemlere ait kritik F_i değerleri Çizelge 6.'da verilmiştir.

Çizelge 5. 28. Gözlem çıkarıldıktan sonraki 29 gözleme ait regresyon analizi için varyans analiz tablosu.

V.K.	S.D.	K.T.	K.O.	F	P
Regresyon	1	30.337	30.337	471.05**	0.000
Hata	27	1.739	0.064		
Genel	28	32.076			

$$S=0.2538 \quad R^2=\% 94.6 \quad \text{düz } R^2= \% 94.4$$

Bu gözlemler içerisinde diğerlerine nazaran geniş residuele sahip olan birkaç tanesinin F_i değerleri şu şekilde hesaplanır :

F_{28} için :

$$RKT = 6.183$$

$$RKT_{-i} = S_{-i}^2(n-k-2)$$

$$S_{-i}^2 = S^2(n-k-1-t_i^2)/(n-k-2)$$

$$F_i = (RKT - RKT_{-i}) / S_{-i}^2$$

$$S_{-28}^2 = 0.221(30-1-1-4.49^2)/(30-1-2) = 0.064$$

$$RKT_{-28} = 0.064(30-1-2) = 1.728$$

$$F_{28} = (6.183 - 1.728) / 0.064 = 69.61$$

F_{30} için :

$$S_{-30}^2 = 0.221(30-1-1-(-1.25)^2)/(30-1-2) = 0.209$$

$$RKT_{-30} = 0.209(30-1-2) = 5.643$$

$$F_{30} = (6.183 - 5.643) / 0.064 = 2.58$$

F₄ için :

$$S_{.4}^2 = 0.221(30-1-1-1.11^2)/(30-1-2) = 0.219$$

$$RKT_{.4} = 0.219(30-1-2) = 5.913$$

$$F_4 = (6.183 - 5.913) / 0.064 = 1.233$$

F₅ için :

$$S_{.5}^2 = 0.221(30-1-1-0.98^2)/(30-1-2) = 0.221$$

$$RKT_{.5} = 0.221(30-1-2) = 5.967$$

$$F_5 = (6.183 - 5.967) / 0.064 = 0.97$$

Diğer gözlemlere ait kritik değerler 0.97'den daha küçük çıkacaklarından hesaplamaya gerek yoktur.

Bulunan F_i değerleri içerisinde en büyük olan F₂₈ değeri F(α/n, 1, n-k-2) ile karşılaştırılır. α=0.10 için;

F(0.0033, 1, 27) ≅ 8.866'dır. Dolayısıyla F_{enb} olan F₂₈ > 8.866 olduğundan F₂₈ outlierdir.

Çizelge 6. 30 Gözlem ait X_i , Y_i , t_i ve kuşkuilanılan gözlemlere ait F_i kritik değerleri.

Gözlem No	X_i	Y_i	t_i	F_i
1	28.0	8.9200	0.25	**
2	28.1	9.0100	0.37	**
3	28.6	9.1300	-0.12	**
4	29.5	10.2100	1.11	1.233
5	29.6	10.2300	0.98	0.970
6	29.8	10.2700	0.85	**
7	30.4	10.1100	-0.37	**
8	30.7	10.1700	-0.59	**
9	30.9	10.6800	0.17	**
10	31.4	10.4000	-0.99	**
11	31.4	10.5500	-0.70	**
12	31.5	10.6100	-0.66	**
13	31.6	10.9800	-0.08	**
14	31.8	10.7000	-0.90	**
15	32.0	10.8200	-0.91	**
16	32.0	11.0300	-0.51	**
17	32.1	11.2700	-0.12	**
18	32.2	11.2700	-0.28	**
19	32.3	11.6100	0.43	**
20	32.6	11.8100	0.35	**
21	32.9	11.9600	0.37	**
22	33.1	11.6400	-0.58	**
23	33.2	11.9200	-0.09	**
24	33.2	11.8200	-0.36	**
25	33.4	11.8300	-0.60	**
26	33.4	12.0800	-0.12	**
27	33.6	12.5500	0.63	**
28	34.9	15.0400	4.49R	69.438
29	35.0	12.8200	-0.60	**
30	35.9	13.1300	-1.25	2.580

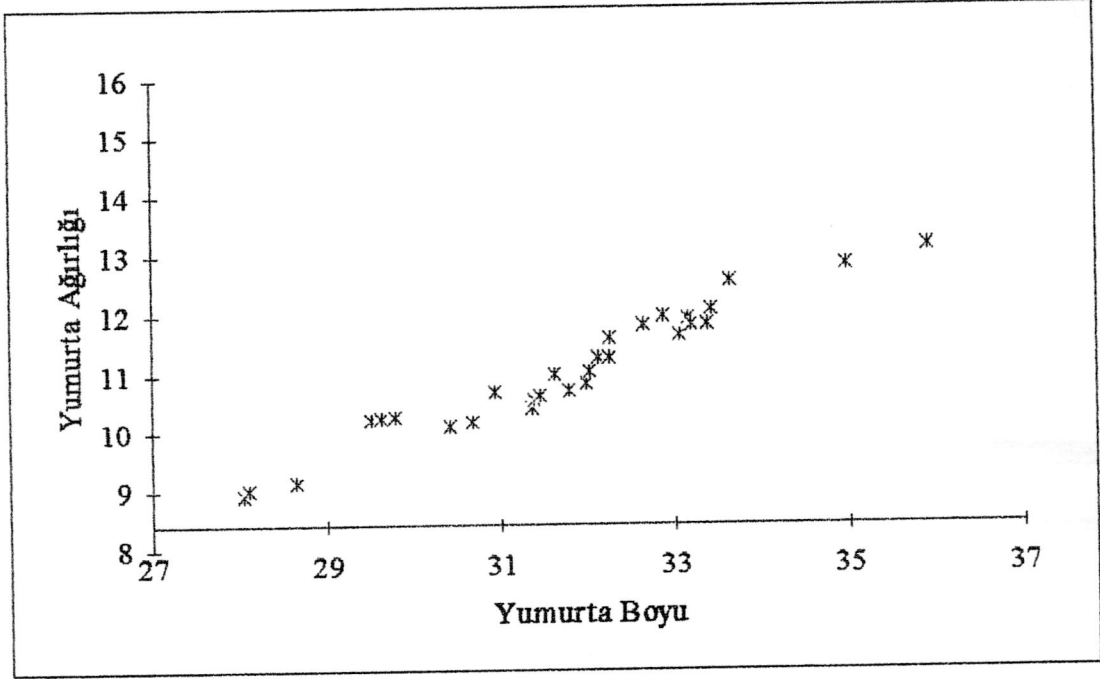
5. TARTIŞMA

Bu çalışmanın başında da belirtildiği gibi, verilerin değerlendirilmesinde kullanılan metodlar ne kadar sağlam olursa olsun, analizlerin uygulanacağı rakamlar güvenilir olmadıkça verilecek hükümler hatadan arınmış olamazlar. Verilecek hükümlerin hatadan arınmış olması için gözlemler içindeki anormal değerlerin (outlierler) teşhis edilmesi ve analizden çıkarılması gerekmektedir. Bu çalışmada outlierlerin teşhis edilmesinde üç yöntem, örnek üzerinde incelenmiştir.

Bıldırıncılardan elde edilen gözlemlere ait student türü residual grafiklerinde 30 gözlemden 28. gözlem, sıfır çizgisinden yaklaşık olarak $t_i/S = 4.49/0.4693 = 9.56$ standart sapma uzaklıktadır. Bu durum 28. gözlemin outlier olarak karar verilmesi için yeterlidir. Cook'a (1977) göre hesaplanan D_i değerleri içinde de en yüksek etkiye ve student türü residuele sahip olan gözlem yine 28. gözlemdir. Yalnız elde edilen değerler içerisinde geniş residuele sahip bir kaç gözlem daha bulunmaktadır. Bu gözlemlerin de outlier olup olmadığını anlamak için, outlier olduğu teşhis edilen gözlem çıkarıldıktan sonra aynı işlemlerin tek tek şüpheli gözlemler için uygulanması gerekmektedir. Bu yöntem *Adımsal Yöntem (İterasyon Yöntemi)* denir.

Adımsal yöntemin bir sakıncası; outlier olup olmadığından kuşkulanan gözlemler birbirine yakın olmaları halinde, gerçekten outlier olan bir gözlemin outlier çıkmaması ihtimalidir. Ayrıca örnek içinde genel olarak düşünüldüğünde outlier olmayan bir gözlem adımsal yöntemde outlier çıkma ihtimali yüksektir. Örnek büyüklüğünün küçük olduğu durumlarda, örnek büyüklüğü daha da küçüleceğinden tavsiye edilemez. O sebeple birden fazla outlierden kuşkulunalınan örneklerde; aynı anda birden fazla outlieri test edebilen AP İstatistiği [3] gibi istatistikler kullanılmalıdır. Ayrıca gözlem sayısının ($n < 20$) küçük olduğu durumlarda birden fazla outlier olup olmadığının test edilmesi güçtür. Uygulama kısmında kullanılan üçüncü yöntem olan, Katkı Kareler Toplamı Testinde de 28. gözlem outlier çıkmıştır.

28. gözlem çıktıktan sonra tekrar EKK modeli uydurulduğunda ve analiz yapıldığında, 28. gözlemin parametre tahminleri üzerinde ne kadar etkiye sahip olduğu açıkça görülmektedir. Şekil 7.'de 28. gözlem çıkarıldıktan sonraki X_i ve Y_i değerlerinin dağılımı görülmektedir.



Şekil 7. 28. Gözlem çıkarıldıktan sonraki X_i ve Y_i değerlerine ait serpmne diyagramı.

30. gözleme ait S ve R^2 değerleri sırasıyla; 0.4699 ve % 87'dir. Yani yumurta boyları yumurta ağırlıklarındaki varyasyonun % 87'sini açıklayabilmektedir. 28. gözlem çıkarıldıktan sonraki S ve R^2 değerleri; 0.25 ve % 94'dür. Görüldüğü gibi standart sapma daha da düşmüş ve yumurta ağırlıklarındaki varyasyonun % 94'ü yumurta boyları ile açıklanabilmektedir. Bu sonuçlar gösteriyor ki; uygulanan outlier teşhisi isabetlidir.

Çalışmalar sonunda bir yada daha fazla gözlemin outlier olduğu sonucuna ulaşırsa, araştırmacının aklına şu sorular gelmelidir;

- acaba model mi yanlışdır?
- bazı transformasyonlar mı gereklidir?
- ölçü, tartım veya kayıtlama hatası mı yapılmıştır?
- acaba örneklemedeki bazı gözlemler diğerlerinden farklı etkenlerin etkisinden mi kalmıştır?
- yoksa bu anormallik tesadüfiliğe mi kaynaklanmıştır?

Ancak bütün bu soruların cevabı araştırmacıdan beklenmemelidir. Anormal gözlemin sebebinin araştırılması; araştırmacıyla birlikte bu tür çalışmalar yapmış ve uzmanlaşmış kişilerin de yardımları ile mümkün olacaktır.

Eğer arařtırmacılar analizde kullandıkları veriler içindeki outlierleri kabataslak teşhis etmek istiyorlarsa; daha önce bu metodu kullanmış bir uzmanın yardımına başvurarak plot yöntemini uygulamalıdır. Eğer arařtırmacı student türü residuallere dayanarak daha hassas bir test yapmak ve outlierleri teşhis etmek istiyorsa Cook istatistiğini veya residuallerin kare toplamında gözlemlerin etki paylarını temel alan katkı kareler toplamı testini kullanmalıdır. Arařtırmacı veri setinde birden fazla gözlemden şüpheleniyorsa adımsal yöntem dediğimiz sırasıyla şüpheli gözlemlerin çıkartılıp test edildiği yöntemi kullanmalıdır. Fakat bu yöntemin küçük örneklerde uygulaması güçtür. Ayrıca örneği bütün olarak ele almadığından gerçekte outlier olan bir gözlem outlier çıkmayabilir veya bunun tam tersi de olabilir. Bu problemlerle karşılaşmamak için aynı anda birden fazla outlier teşhisi yapılabildiği, bağımsız deęişken matrisinin Y vektörü ile genişletilmesi üzerine kurulan AP istatistiği kullanılmalıdır.

Sonuç olarak; eęer bir çalışmada outlier deęerler teşhis edilmişse bu gözlemlerin analizden çıkartılıp yeni bir model uydurulması gerekir. Eğer analizden çıkartılmayacaksa da; EKK yönteminin dışında alternatif tahmin yöntemlerinden olan "*Robust Tahmin*" yöntemleri kullanılarak outlierlerin etkileri azaltılmalıdır. Ayrıca çalışmalardaki birkaç tane outlierin teşhisi, tesbit edilmemiş ve etkisi araştırılmamış dięer bir bağımsız deęişkenin varlığından da haber verir.

TUTANAK

Zootekni Anabilim Dalında Yüksek Lisans öğrencisi Zeki DOĞAN'ın "Basit Doğrusal Regresyon Modellerindeki Tek Bir Anormal Gözlemin Testi" adlı Yüksek Lisans tezi Harran Üniversitesi Fen Bilimleri Enstitüsünün lisans üstü öğretim yönergesinin ilgili maddesi uyarınca 23 / 01 / 1995 tarihinde tez savunmasına tabi tutulmuş ve sınav sonucunda başarılı olduğuna jürimizce karar verilmiştir.

(Jüri Başkanı)

(Jüri Üyesi)

(jüri üyesi)

Prof.Dr.

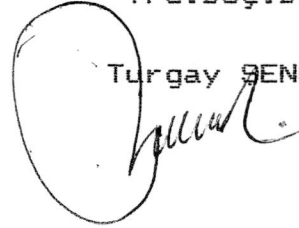
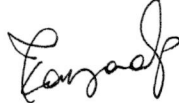
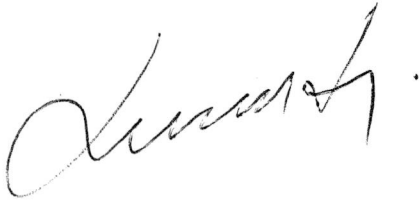
Yrd.Doç.Dr

Yrd.Doç.Dr

Necati YILDIZ

Tamer KAYAALP

Turgay BENGÜL



1 ve n-k-2 serbestlik dereceli χ^2/n yanılma düzeyindeki F cetveli

f_2	p	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	0.250	5.83	7.50	8.20	8.58	8.82	8.98	9.10	9.19	9.26	9.32	9.41	9.49	9.58	9.63	9.67	9.71	9.76	9.80	9.85
1	0.100	39.86	49.50	53.59	55.83	57.24	58.20	58.91	59.44	59.86	60.20	60.70	61.22	61.74	62.00	62.26	62.53	62.79	63.06	63.33
1	0.025	64.8	800	864	900	922	937	948	957	963	969	977	985	992	997	1.001	1.004	1.007	1.010	1.013
1	0.005	16.211	20.000	21.615	22.500	23.056	23.437	23.715	23.925	24.091	24.224	24.426	24.650	24.836	24.940	25.044	25.148	25.253	25.359	25.465
2	0.250	2.57	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.37	3.38	3.39	3.41	3.43	3.43	3.44	3.45	3.46	3.47	3.48
2	0.100	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.40	9.42	9.44	9.45	9.46	9.47	9.47	9.48	9.49
2	0.025	38.51	39.00	39.16	39.25	39.30	39.33	39.36	39.37	39.39	39.40	39.42	39.43	39.45	39.46	39.46	39.47	39.48	39.49	39.50
2	0.005	19.8	19.9	19.9	19.9	19.9	19.9	19.9	19.9	19.9	19.9	19.9	19.9	19.9	19.9	19.9	19.9	19.9	19.9	200
3	0.250	2.02	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44	2.44	2.45	2.46	2.46	2.46	2.46	2.47	2.47	2.47	2.47
3	0.100	5.54	5.46	5.39	5.34	5.31	5.27	5.25	5.24	5.24	5.23	5.22	5.22	5.22	5.18	5.18	5.17	5.16	5.15	5.13
3	0.025	17.44	16.04	15.44	15.10	14.88	14.62	14.54	14.47	14.42	14.34	14.34	14.34	14.17	14.12	14.08	14.04	13.99	13.95	13.90
3	0.005	55.55	49.80	47.47	46.20	45.39	44.84	44.43	44.13	43.88	43.69	43.59	43.59	43.08	42.78	42.62	42.47	42.31	42.15	41.83
4	0.250	1.81	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08
4	0.100	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.92	3.90	3.87	3.84	3.83	3.82	3.80	3.79	3.76
4	0.025	12.22	10.65	9.98	9.56	9.36	9.20	9.07	8.98	8.90	8.84	8.75	8.66	8.56	8.51	8.46	8.41	8.36	8.31	8.26
4	0.005	31.33	26.28	24.26	23.16	22.46	21.98	21.62	21.35	21.14	20.97	20.70	20.44	20.17	20.03	19.89	19.75	19.61	19.47	19.32
5	0.250	1.69	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.88	1.88	1.88	1.87	1.87	1.87
5	0.100	4.06	3.78	3.52	3.37	3.25	3.15	3.07	3.00	2.96	2.92	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
5	0.025	10.01	8.43	7.79	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.57	6.52	6.43	6.33	6.28	6.18	6.12	6.07	6.02
5	0.005	22.78	18.31	16.53	15.56	14.94	14.51	14.20	13.96	13.77	13.62	13.38	13.15	12.90	12.78	12.66	12.53	12.40	12.27	12.14
6	0.250	1.62	1.76	1.78	1.79	1.79	1.78	1.78	1.78	1.77	1.77	1.77	1.77	1.76	1.75	1.75	1.75	1.74	1.74	1.74
6	0.100	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.90	2.87	2.84	2.82	2.80	2.78	2.76	2.74	2.72
6	0.025	8.81	7.29	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.44	5.37	5.27	5.17	5.12	5.07	5.01	4.96	4.90	4.85
6	0.005	18.64	14.54	12.92	12.03	11.46	11.07	10.79	10.57	10.39	10.25	10.03	9.81	9.59	9.47	9.36	9.24	9.12	9.00	8.88
7	0.250	1.57	1.70	1.72	1.72	1.71	1.71	1.70	1.70	1.69	1.69	1.68	1.68	1.67	1.67	1.66	1.66	1.65	1.65	1.65
7	0.100	3.50	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.67	2.63	2.59	2.58	2.56	2.54	2.51	2.49	2.47
7	0.025	8.54	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.67	4.57	4.47	4.42	4.38	4.31	4.25	4.20	4.14
7	0.005	16.24	12.40	10.88	10.05	9.52	9.16	8.89	8.68	8.51	8.38	8.18	7.97	7.75	7.64	7.53	7.42	7.31	7.19	7.08
8	0.250	1.54	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.64	1.63	1.62	1.62	1.62	1.60	1.60	1.60	1.59	1.58	1.58
8	0.100	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.50	2.46	2.42	2.40	2.39	2.36	2.34	2.32	2.29
8	0.025	8.07	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.20	4.10	4.00	3.95	3.89	3.84	3.78	3.73	3.67
8	0.005	14.09	11.04	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.01	6.81	6.61	6.50	6.40	6.29	6.18	6.06	5.95

KAYNAKLAR

- 1) AKAR, M. ve ŞAHİNLER, S., "Kalıntıların Grafik Analizleri Yardımıyla Regresyon Modelinin Yeterliliğinin Saptanması", Araştırma 92' Sempozyumu, 1992, Ankara.
- 2) AKAR, M. ve ŞAHİNLER, S., "Standart Kalıntılar ve Hatanın Normallik Varsayımının İncelenmesi", 1993, Adana.
- 3) ANDREWS, D.F. ve PREGIBON, D., "Finding the Outliers That Matter" Journal of the Royal Statistical Society, Ser. B,40, 87-93, 1978.
- 4) BALASOORIYA ,W., ve TSE, Y.K., "Outlier Detection in Linear Models : A Comparative Study in Simple Linear Regression", Commun. Statis. Theory. Meth., National Universty of Singapore, Singapore, 12, 3589-3597, 1986.
- 5) BARNETT, V. , "The Study of Outliers: Purpose end Models", Applied Statistics, 27, 242-250, 1978.
- 6) BECKMAN, J. R. ve COOK, R.D., "Outlier.....s", Technometrics, 25, 119-149, 1983
- 7) BERTRAM, S., ve PRICE, C., "Regression Analysis By Example" Wiley, 24-30, 1991.
- 8) COOK, R.D., "Detection of Influential Observations In Linear Regression", Technometrics, 19, 15-18, 1977.
- 9) DRAPER, N.R., ve SMITH, H., "Applied Regression Analysis", Wiley, New York, 1966.

- 10) DRAPER, N.R., ve JOHN, J.A., "Influential Observations and Outliers In Regression", *Technometrics* , 23, 21-26, 1981.
- 11) ELLENBERG ,J.H., "Testing for a Single Outlier from a General Linears Model",*Biometrics*, 32, 637-645, 1976.
- 12) ERAR, A., ÇETİNEL, B., ve TATLIDİL, H., "Doğrusal Regresyon Modelinde Sapan Gözlemlerin Bulunması", Hacettepe Ün., 1979, Ankara.
- 13) ERAR, A., "Regresyon (Bağlanım) Çözümlemesi, Ders Notları" Hacettepe Ün., 1985, Ankara.
- 14) ERTEK, T., "Ekonometriye Giriş", Beta Basım Yayım Dağıtım A.Ş., İstanbul, 1987.
- 15) FREUND, R.J., WAİL, R.W., ve CLUNIES-ROSS, C.W., "Residual Analysis", *American Statistical Association Journal*, 3,98-104, 1961.
- 16) GRAYBILL, F.A., "An Introduction to Linear Statistical Models", Vol. 1., Mc Graw-Hill, New York, 425-443, 1961.
- 17) HOAGLIN, D.C., ve WELSCH, R.E., "The Hat Matrix in Regression and Anova", *The American Istatistician*, 32, 17-32,1978.
- 18) MASON, L.R., ve GUNST, F.R., "Residual Analysis and Its Application", Chapter 7, San Antonia, Dallas, Texas, 220-230, 1978.
- 19) MINITAB Reference Manual Version 8.2, "Release 7", 1989, U.S.A.
- 20) MONTGOMERY, D.C., ve PECK, E.A., "Introductions to Linear Regression Analysis", 1982, New York.
- 21) NETTER, J., ve WASSERMAN, W., "Applied Linear Statistical Models", Georgetawn Ontaria, 106-109, 1973.

- 22) SNEDECOR, G.W., ve COCHRAN, W.G., "Statistical Methods", Iowa State Universty Press, Iowa, 564-567, 1971.
- 23) TATLIDİL, H., "Genel Doğrusal Regresyonda Tek Sapan Değerin Testi", Hacettepe Ün., 1979, Ankara.
- 24) YILDIZ, N., ve BİRCAN, H., "Araştırma ve Deneme Metodları", Atatürk Ün. Ziraat Fak., 1991, Erzurum.

$V_1 = 1$

v_2	$p=0.0001$	$p=0.001$	$p=0.005$	$p=0.01$	$p=0.025$	$p=0.05$	$p=0.1$	$p=0.25$	$f=0.5$	$p=0.75$	$f=0.7$	$p=0.95$	$p=0.975$	$p=0.99$	$p=0.995$	$p=0.999$	
1	(+7) 4.0528 (+5)	4.0528 (+4)	1.6211 (+3)	4.0522 (+3)	4.4177 (+2)	1.6145 (+2)	3.9804 (+1)	5.8285	1.0000	(-1) 1.7157	(-2) 2.5085	(-3) 6.1239	(-3) 1.5437	(-4) 4.9730	(-5) 6.1687	(-6) 7.1674	(-8) 7.5779
1.2	(6) 2.3821 (4)	5.1319 (3)	3.5094 (2)	1.1048 (3)	2.3927 (2)	7.4802 (1)	2.3000 (1)	4.3669	(-1) 8.7148	(-1) 1.5811	(-2) 2.3124	(-3) 5.7938	(-3) 1.4146	(-4) 2.3097	(-5) 5.7736	(-6) 2.3094	(-8) 7.3694
1.5	(5) 1.4790 (3)	6.8637 (2)	8.0184 (2)	3.1756 (2)	9.2839 (1)	3.6200 (1)	1.3728 (1)	3.3235	(-1) 7.6142	(-1) 1.4540	(-2) 2.1794	(-3) 5.3994	(-3) 1.3468	(-4) 2.1516	(-5) 5.3834	(-6) 2.1531	(-8) 7.1531
2	(+3) 9.9985 (+2)	9.9850 (+2)	1.9850 (+2)	9.8503 (+1)	3.8506 (+1)	1.8513 (+1)	8.5263	2.5714	(-1) 6.6667	(-1) 1.3333	(-2) 2.0402	(-3) 5.9125	(-4) 1.2508	(-4) 2.0002	(-5) 5.0000	(-6) 2.0000	(-8) 2.0000
3	(+2) 7.8401 (+2)	1.6703 (+1)	5.5521 (+1)	3.4116 (+1)	1.7443 (+1)	0.1228 (+1)	5.5383	2.0239	(-1) 5.8506	(-1) 1.2195	(-2) 1.6559	(-3) 4.6359	(-4) 1.1572	(-4) 1.8507	(-5) 4.6264	(-6) 1.8506	(-8) 1.8506
4	(+2) 2.4162 (+1)	7.4137 (+1)	3.1333 (+1)	2.1198 (+1)	1.2218 (+1)	7.7086	4.5448	1.8074	(-1) 5.4801	(-1) 1.1654	(-2) 1.7911	(-3) 4.4528	(-4) 1.1116	(-4) 1.7779	(-5) 4.4444	(-6) 1.7779	(-8) 1.7779
5	(+2) 1.2494 (+1)	4.7181 (+1)	2.2785 (+1)	1.6258 (+1)	1.0007 (+1)	6.6079	4.0604	1.6925	(-1) 5.2807	(-1) 1.1350	(-2) 1.7470	(-3) 4.3448	(-4) 1.0948	(-4) 1.1500	(-5) 4.3373	(-6) 1.1500	(-8) 1.1500
6	(+1) 8.2489 (+1)	3.5507 (+1)	1.8635 (+1)	1.3745 (+1)	8.8131	5.9874	3.7750	1.6214	(-1) 5.1489	(-1) 1.1132	(-2) 1.7181	(-3) 4.2737	(-4) 1.0671	(-4) 1.7068	(-5) 4.2668	(-6) 1.7068	(-8) 1.7068
7	(+1) 6.2167 (+1)	2.9245 (+1)	1.6236 (+1)	1.2246 (+1)	8.0727	5.5914	3.5904	1.5732	(-1) 5.0572	(-1) 1.0946	(-2) 1.6975	(-3) 4.2235	(-4) 1.0546	(-4) 1.6808	(-5) 4.2167	(-6) 1.6808	(-8) 1.6808
8	(+1) 5.0694 (+1)	2.5415 (+1)	1.4688 (+1)	1.1259 (+1)	7.5709	5.3177	3.4579	1.5384	(-1) 4.9898	(-1) 1.0879	(-2) 1.6825	(-3) 4.1862	(-4) 1.0453	(-4) 1.6718	(-5) 4.1797	(-6) 1.6718	(-8) 1.6718
9	(+1) 4.3477 (+1)	2.2857 (+1)	1.3614 (+1)	1.0561 (+1)	7.2091	5.1174	3.3601	1.5121	(-1) 4.9382	(-1) 1.0796	(-2) 1.6706	(-3) 4.1573	(-4) 1.0391	(-4) 1.6604	(-5) 4.1509	(-6) 1.6604	(-8) 1.6604
10	(+1) 3.8577 (+1)	2.1040 (+1)	1.2826 (+1)	1.0044 (+1)	6.9367	4.9846	3.2850	1.4915	(-1) 4.8873	(-1) 1.0729	(-2) 1.6613	(-3) 4.1343	(-4) 1.0224	(-4) 1.6513	(-5) 4.1281	(-6) 1.6513	(-8) 1.6513
12	(+1) 3.2427 (+1)	1.8643 (+1)	1.1754 (+1)	9.3302	6.5538	4.7472	3.1765	1.4611	(-1) 4.8369	(-1) 1.0631	(-2) 1.6471	(-3) 4.0979	(-4) 1.0236	(-4) 1.6377	(-5) 4.0948	(-6) 1.6378	(-8) 1.6378
15	(+1) 2.7448 (+1)	1.6487 (+1)	1.0798 (+1)	8.6831	6.1995	4.5401	3.0732	1.4321	(-1) 4.7775	(-1) 1.0534	(-2) 1.6335	(-3) 4.0639	(-4) 1.0154	(-4) 1.6241	(-5) 4.0601	(-6) 1.6240	(-8) 1.6240
20	(+1) 2.3399 (+1)	1.4814 (+1)	9.9437	8.0960	5.8715	4.3513	2.9747	1.4037	(-1) 4.7192	(-1) 1.0437	(-2) 1.6197	(-3) 4.0321	(-4) 1.0069	(-4) 1.6106	(-5) 4.0264	(-6) 1.6105	(-8) 1.6105
24	(+1) 2.1665 (+1)	1.4028 (+1)	9.5513	7.8220	5.7167	4.2597	2.9271	1.3898	(-1) 4.6992	(-1) 1.0389	(-2) 1.6129	(-3) 4.0153	(-4) 1.0029	(-4) 1.6040	(-5) 4.0036	(-6) 1.6036	(-8) 1.6036
30	(+1) 2.0092 (+1)	1.3293 (+1)	9.1797	7.5625	5.5675	4.1709	2.8807	1.3761	(-1) 4.6616	(-1) 1.0341	(-2) 1.6060	(-3) 3.9986	(-4) 0.9860	(-4) 1.5973	(-5) 3.9940	(-6) 1.5972	(-8) 1.5972
40	(+1) 1.8668 (+1)	1.2609 (+1)	8.8278	7.3141	5.4239	4.0848	2.8354	1.3626	(-1) 4.6330	(-1) 1.0294	(-2) 1.5993	(-3) 3.9819	(-4) 0.9743	(-4) 1.5885	(-5) 3.9763	(-6) 1.5885	(-8) 1.5885
60	(+1) 1.7377 (+1)	1.1973 (+1)	8.4946	7.0771	5.2857	4.0012	2.7914	1.3493	(-1) 4.6053	(-1) 1.0247	(-2) 1.5925	(-3) 3.9651	(-4) 0.9630	(-4) 1.5840	(-5) 3.9597	(-6) 1.5830	(-8) 1.5830
120	(+1) 1.6204 (+1)	1.1380 (+1)	8.1750	6.8510	5.1324	3.9201	2.7470	1.3362	(-1) 4.5774	(-1) 1.0200	(-2) 1.5856	(-3) 3.9487	(-4) 0.9619	(-4) 1.5793	(-5) 3.9414	(-6) 1.5794	(-8) 1.5794
∞	(+1) 1.5137 (+1)	1.0828 (+1)	7.8794	6.6349	5.0219	3.8415	2.7055	1.3231	(-1) 4.5494	(-1) 1.0153	(-2) 1.5793	(-3) 3.9321	(-4) 0.9601	(-4) 1.5748	(-5) 3.9356	(-6) 1.5749	(-8) 1.5749

$V_1 = 2$

v_2	$p=0.0001$	$p=0.001$	$p=0.005$	$p=0.01$	$p=0.025$	$p=0.05$	$p=0.1$	$p=0.25$	$p=0.5$	$p=0.75$	$p=0.9$	$p=0.95$	$p=0.975$	$p=0.99$	$p=0.995$	$p=0.999$	$p=0.9999$	v_2
1	(+7) 5.0000(+5)	5.0000(+4)	2.0000(+4)	4.9995(+3)	7.9950(+2)	1.9950(+2)	4.9500(+1)	7.5000	1.5000	(-1) 3.8889(-1)	1.1728(-2)	5.4016(-2)	(-2) 2.5970(-2)	(-2) 1.0152(-2)	(-3) 5.0378(-3)	1.0015(-4)	1.0002(-4)	1
1,2	(6) 2.7850(4)	5.9999(3)	4.1033(3)	1.2921(2)	2.8011(2)	8.7817(1)	2.7250(1)	5.4476	1.3049	(-1) 3.6913(-1)	1.1518(-2)	5.3550(-2)	(-2) 2.5860(-2)	(-2) 1.0135(-2)	(-3) 5.0335(-3)	1.0013(-4)	1.0001(-4)	1,2
1,5	(5) 1.6158(3)	7.4992(2)	8.7646(2)	3.4737(2)	1.0185(2)	3.9966(1)	1.5408(1)	4.0122	1.1399	(-1) 3.5064(-1)	1.1312(-2)	5.3088(-2)	(-2) 2.5750(-2)	(-2) 1.0118(-2)	(-3) 5.0293(-3)	1.0012(-4)	1.0001(-4)	1,5
2	(+3) 9.9990(+2)	9.9900(+2)	1.9900(+2)	9.9000(+1)	3.9000(+1)	1.9000(+1)	9.0000	3.0000	1.0000	(-1) 3.3333(-1)	1.1111(-2)	5.2632(-2)	(-2) 2.5641(-2)	(-2) 1.0101(-2)	(-3) 5.0251(-3)	1.0010(-4)	1.0001(-4)	2
3	(+2) 6.9474(+2)	1.4850(+1)	4.9799(+1)	3.0817(+1)	1.6044(+1)	9.5521	5.4624	2.2798	(-1) 8.8110(-1)	3.1712(-1)	1.0915(-2)	5.2181(-2)	(-2) 2.5583(-2)	(-2) 1.0084(-2)	(-3) 5.0208(-3)	1.0008(-4)	1.0001(-4)	3
4	(+2) 1.9800(+1)	6.1246(+1)	2.6284(+1)	1.8000(+1)	1.0649(+1)	6.9443	4.3246	2.0000	(-1) 8.2843(-1)	3.0941(-1)	1.0819(-2)	5.1956(-2)	(-2) 2.5479(-2)	(-2) 1.0076(-2)	(-3) 5.0188(-3)	1.0008(-4)	1.0001(-4)	4
5	(+1) 9.7027(+1)	3.7122(+1)	1.8314(+1)	1.3274(+1)	8.4336	5.7861	3.7797	1.8528	(-1) 7.9877(-1)	3.0489(-1)	1.0761(-2)	5.1824(-2)	(-2) 2.5447(-2)	(-2) 1.0071(-2)	(-3) 5.0176(-3)	1.0007(-4)	1.0001(-4)	5
6	(+1) 6.1633(+1)	2.7000(+1)	1.4584(+1)	1.0925(+1)	7.2598	5.1433	3.4633	1.7622	(-1) 7.7976(-1)	3.0192(-1)	1.0723(-2)	5.1733(-2)	(-2) 2.5425(-2)	(-2) 1.0067(-2)	(-3) 5.0168(-3)	1.0007(-4)	1.0001(-4)	6
7	(+1) 4.5132(+1)	2.1689(+1)	1.2404(+1)	9.5466	6.5415	4.7374	3.2574	1.7010	(-1) 7.6655(-1)	2.9983(-1)	1.0696(-2)	5.1672(-2)	(-2) 2.5410(-2)	(-2) 1.0065(-2)	(-3) 5.0161(-3)	1.0006(-4)	1.0001(-4)	7
8	(+1) 3.6000(+1)	1.8494(+1)	1.1042(+1)	8.6491	6.0595	4.4590	3.1131	1.6569	(-1) 7.5683(-1)	2.9828(-1)	1.0676(-2)	5.1624(-2)	(-2) 2.5398(-2)	(-2) 1.0063(-2)	(-3) 5.0158(-3)	1.0006(-4)	1.0001(-4)	8
9	(+1) 3.0342(+1)	1.6387(+1)	1.0107(+1)	8.0215	5.7147	4.2565	3.0065	1.6236	(-1) 7.4938(-1)	2.9708(-1)	1.0660(-2)	5.1586(-2)	(-2) 2.5389(-2)	(-2) 1.0062(-2)	(-3) 5.0153(-3)	1.0006(-4)	1.0001(-4)	9
10	(+1) 2.6548(+1)	1.4905(+1)	9.4270	7.5594	5.4564	4.1028	2.9245	1.5975	(-1) 7.4349(-1)	2.9612(-1)	1.0646(-2)	5.1557(-2)	(-2) 2.5382(-2)	(-2) 1.0060(-2)	(-3) 5.0150(-3)	1.0006(-4)	1.0001(-4)	10
12	(+1) 2.1856(+1)	1.2974(+1)	8.5096	6.9266	5.0959	3.8853	2.8068	1.5595	(-1) 7.3477(-1)	2.9469(-1)	1.0629(-2)	5.1512(-2)	(-2) 2.5371(-2)	(-2) 1.0059(-2)	(-3) 5.0145(-3)	1.0006(-4)	1.0001(-4)	12
15	(+1) 1.8109(+1)	1.1339(+1)	7.7008	6.3589	4.7650	3.6823	2.6952	1.5227	(-1) 7.2619(-1)	2.9327(-1)	1.0610(-2)	5.1469(-2)	(-2) 2.5361(-2)	(-2) 1.0057(-2)	(-3) 5.0143(-3)	1.0006(-4)	1.0001(-4)	15
20	(+1) 1.5119	9.9526	6.9865	5.8489	4.4613	3.4928	2.5893	1.4870	(-1) 7.1773(-1)	2.9186(-1)	1.0592(-2)	5.1424(-2)	(-2) 2.5350(-2)	(-2) 1.0055(-2)	(-3) 5.0138(-3)	1.0006(-4)	1.0001(-4)	20
24	(+1) 1.3853	9.3394	6.6610	5.6136	4.3187	3.4028	2.5383	1.4695	(-1) 7.1356(-1)	2.9116(-1)	1.0583(-2)	5.1403(-2)	(-2) 2.5345(-2)	(-2) 1.0054(-2)	(-3) 5.0135(-3)	1.0005(-4)	1.0001(-4)	24
30	(+1) 1.2718	8.7734	6.3547	5.3904	4.1821	3.3158	2.4887	1.4524	(-1) 7.0941(-1)	2.9046(-1)	1.0571(-2)	5.1382(-2)	(-2) 2.5339(-2)	(-2) 1.0054(-2)	(-3) 5.0133(-3)	1.0005(-4)	1.0001(-4)	30
40	(+1) 1.1692	8.2508	6.0664	5.1785	4.0510	3.2317	2.4404	1.4355	(-1) 7.0531(-1)	2.8976(-1)	1.0564(-2)	5.1358(-2)	(-2) 2.5334(-2)	(-2) 1.0053(-2)	(-3) 5.0133(-3)	1.0005(-4)	1.0001(-4)	40
60	(+1) 1.0781	7.7678	5.7950	4.9774	3.9253	3.1504	2.3932	1.4186	(-1) 7.0122(-1)	2.8907(-1)	1.0555(-2)	5.1337(-2)	(-2) 2.5329(-2)	(-2) 1.0052(-2)	(-3) 5.0130(-3)	1.0005(-4)	1.0001(-4)	60
120	9.9549	7.3211	5.5393	4.7865	3.8046	3.0728	2.3473	1.4024	(-1) 6.9717(-1)	2.8838(-1)	1.0545(-2)	5.1316(-2)	(-2) 2.5323(-2)	(-2) 1.0051(-2)	(-3) 5.0128(-3)	1.0005(-4)	1.0001(-4)	120
∞	9.2104	6.9076	5.2983	4.6052	3.6889	2.9957	2.3026	1.3863	(-1) 6.9315(-1)	2.8768(-1)	1.0536(-2)	5.1293(-2)	(-2) 2.5316(-2)	(-2) 1.0050(-2)	(-3) 5.0123(-3)	1.0005(-4)	1.0001(-4)	∞

