

**REGRESYONDA KALINTI (RESİDUAL)
ANALİZİ**

ARŞ.GÖR. SUAT ŞAHİNLER

**Ç.Ü.
FEN BİLİMLERİ ENSTİTÜSÜ
ZOOTEKNİ ANABİLİM DALI
YÜKSEK LİSANS TEZİ**

ADANA
Eylül , 1993

Çukurova Üniversitesi Fen Bilimleri Enstitüsü Müdürlüğüne,

Bu Çalışma, Jürimiz Tarafından Zootekni Anabilim Dalında
YÜKSEK LİSANS TEZİ Olarak Kabul Edilmiştir.

BAŞKAN : Prof.Dr. Mustafa AKAR

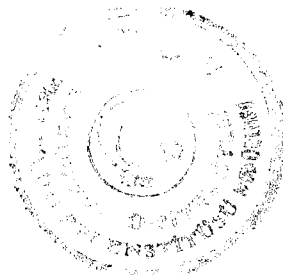
ÜYE : Prof.Dr. Yüksel BEK

ÜYE : Prof.Dr. Erdoğan PEKEL

KOD NO : 728

Yukarıdaki imzaların, adı geçen öğretim üyelerine ait olduğunu
onaylarım.


Prof.Dr. Ural DİNÇ
Enstitü Müdürü



**BU ÇALIŞMA ÇUKUROVA ÜNİVERSİTESİ
ARAŞTIRMA
FONU TARAFINDAN DESTEKLENMİŞTİR.**

İÇİNDEKİLER

Sayfa No

ÇİZELGE LİSTESİ	I
ŞEKİL LİSTESİ	II
ÖZ	IV
ABSTRACT	V
1. GİRİŞ	1
2. ÖNCEKİ ÇALIŞMALAR	8
2.1. Basit (Ordinary) Kalıntılar	8
2.2. Standart Kalıntılar	11
2.3. Kısmi(Partial) Kalıntılar	13
2.4. Yapılan Diğer Çalışmalar ve Uygulamaları	14
3. MATERYAL ve YÖNTEM	18
3.1. Materyal	18
3.2. Yöntem	19
3.2.1. Basit Kalıntılar	20
3.2.1.1. Regresyon Fonksiyonunun Doğrusal Olmaması	22
3.2.1.2. Hata Varyansının Sabit Olmaması	24
3.2.1.3. Hatanın Bağımsız Olmaması	27
3.2.1.4. Anormal Gözlemlerin (Outlier) Olmaması	28
3.2.2. Standart Kalıntılar	30
3.2.3. Kısmi (Partial) Kalıntılar	34
3.2.3.1. Kısmi Kalıntı Grafiklerinde Doğrusal Olmayan Bağımlılığın Belirlenmesi	37
3.2.3.1. Kısmi Kalıntı Grafiklerinde Tahmin Edilen Eğimin Yapısı	40

4. ARAŞTIRMA BULGULARI.....	42
4.1. Basit (Ordinary) Kalıntılar.....	46
4.1.1. Basit Kalıntıların Tüminün Grafiği.....	47
4.1.2. Basit Kalıntıların Zaman Serisi Grafiği..	49
4.1.3. y_i 'e Karşı Basit Kalıntı Grafiği.....	50
4.1.4. x_{ji} 'lere Karşı Basit Kalıntı Grafiği.....	51
4.2. Standart (Standardized) Kalıntılar.....	52
4.2.1. Standart Kalıntıların Serpme Grafiği....	55
4.2.2. Standart Kalıntıların Normal Olasılık Kâğıtları Üzerinde Normal Grafiği.....	56
4.3. Kısmi (Partial) Kalıntılar.....	57
5. TARTIŞMA ve SONUÇLAR.....	63
ÖZET.....	66
SUMMARY.....	71
KAYNAKLAR.....	75
TEŞEKKÜR.....	77
ÖZGEÇMİŞ.....	78

ÇİZELGE LİSTESİ

Sayfa No

Çizelge 3.1. Solunum sayısı(dk), Nabız Sayısı(dk), Süt Verimleri(kg/koy.).....	18
Çizelge 4.1. Varyans Analiz tablosu.....	42
Çizelge 4.2. Solunum sayısı(dk), Nabız Sayısı(dk), Süt Verimleri(kg/koy.), (4.1) Nolu Eşitlik Kullanılarak Hesaplanan Tahmini Değerler ve (3.2.5) nolu Eşitlik Kullanılarak Hesaplanan Kalıntı Değerleri.	43
Çizelge 4.3. Koyunların Süt Verimi ile solunum sayısı Arasındaki Regresyona Ait Varyans Analiz Tablosu	44
Çizelge 4.4. Solunum Sayısı(dk), Süt Verimleri(kg/koy.), (4.2) nolu Eşitlik Kullanılarak Hesaplanan Tahmini Değerler ve (3.2.4) Nolu eşiklik Kullanılarak Hesaplanan Basit Kalıntı Değerleri.....	45
Çizelge 4.2.1. Bağımsız x , M Matrisinin Köşegen Elemanları(m_{ii}) ve Standart Sapma Değerleri(Se_i).....	53
Çizelge 4.2.2. Koyunlara Ait Solunum sayısı(dk), Süt Verimleri(kg/koy.) y , e_i , Standart Kalıntılar, Sıralı Standart Kalıntılar ve Yıgmal Nisbi Frekansları.....	54
Çizelge 4.3.1. Koyunlara Ait Solunum sayısı(dk), Nabız Sayısı(dk), Süt Verimleri(kg/koy.), y , e_i , Sırasıyla (3.3.2) ve (3.3.4) nolu Eşitlikler Yardımı ile Hesaplanan e_1^* ve e_2^* Kısmi Kalıntı Değerleri.	59

ŞEKİL LİSTESİ**Sayfa No**

Şekil 1.1. Kalıntı kareler toplamının uyum eksikliği ve salt hata kareler toplamına ayrışması.....	7
Şekil 3.2.1. Yeterli ve modeli doğrulayıcı kalıntı grafığı.....	21
Şekil 3.2.2a. Doğrusal olmayan regresyon fonksiyonunu gösteren serpmne grafığı.....	23
Şekil 3.2.2b. Doğrusal olmayan regresyon fonksiyonunu gösteren kalıntı grafığı.....	23
Şekil 3.2.3. Sabit olmayan varyansı gösteren kalıntı grafığı.....	24
Şekil 3.2.4. Sabit olmayan varyansı gösteren Uydurulan ^ y değerlerine karşı kalıntı grafığı.....	25
Şekil 3.2.5. Hata teriminin bağımsız olmadığını gösteren kalıntı grafığı.....	28
Şekil 3.2.6. Anormal gözlemi gösteren kalıntı grafığı...	29
Şekil 3.2.7. Birinci regresyon örneğinde bağımsız değişken (x)'e karşı çizilen basit kalıntı grafığı.....	37
Şekil 3.2.8. İkinci regresyon örneğinde bağımsız değişken (x)'e karşı çizilen basit kalıntı grafığı.....	38
Şekil 3.2.9. Birinci regresyon örneğinde bağımsız değişken (x)'e karşı çizilen kısmi kalıntı grafığı.....	38
Şekil 3.2.10. İkinci regresyon örneğinde bağımsız değişken (x)'e karşı çizilen kısmi kalıntı grafığı.....	39

Sayfa No

Şekil 4.1. Verilere ait serpm ve uydurulan regresyon doğrusu grafiği.....	46
Şekil 4.1.1. Kalıntıların tümünün grafiği.....	47
Şekil 4.1.2. Kalıntıların zaman serisi grafiği.....	50
Şekil 4.1.3. Tahmini y_i değerlerine karşı kalıntı grafiği.	51
Şekil 4.1.4. Bağımsız x değişkenine karşı kalıntı grafiği.	52
Şekil 4.2.1. e_i/Se_i 'nin serpm grafiği.....	55
Şekil 4.2.2. $y = 223 - 1.22x_1$ modeli için standart kalıntıların normal grafiği.....	56
Şekil 4.3.1. Solunum sayısına(x_1) karşı çizilen basit kalıntı grafiği.....	60
Şekil 4.3.2. Nabız sayısına(x_2) karşı çizilen basit kalıntı grafiği.....	61
Şekil 4.3.3. Solunum sayısına(x_1) karşı çizilen kısmi kalıntı grafiği.....	61
Şekil 4.3.4. Nabız sayısına(x_2) karşı çizilen kısmi kalıntı grafiği.....	62

ÖZ

Bu çalışmanın amacı, çeşitli şekillerde hesaplanan kalıntıların grafik analizlerini yaparak gözlemleri temsil eden regresyon eşitliğinin yeterliliğini belirlemektir. Bunun için aşağıdaki üç farklı eşitlikle hesaplanan kalıntıların grafik analizleri incelenmiştir.

1. Basit (ordinary) kalıntılar ; $e_i = y_i - \hat{y}_i$,

2. Standart kalıntılar ; $r_i = e_i / S_{e_i}$,

3. Kısmi (partial) kalıntılar ; $e_i^* = e_i + x_i \beta_i$.

Araştırmada 9 hafta süreyle 36 Ostfirizxİvesi melezi koyundan alınan solunum sayısı(dk), nabız sayısı(dk) ortalamaları ve laktasyon süt verimleri kullanılmıştır. Bu verilere ait regresyon doğrusu bulunarak yukarıdaki üç ayrı şekilde hesaplanan kalıntıların grafik analizleri yapılmış ve analizler sonucunda en küçük kareler regresyon doğrusunun yeterli olmadığı fakat yeterli olabilmesi için gözlemde ya bir transformasyona gidilmesi veya tartılı en küçük kareler yönteminin uygulanması gerektiği belirlenmiştir.

ABSTRACT

The aim of this study is to determine sufficiency of the regression model by using graphical analysis method of residuals which were calculated different ways. For this purpose, the following types of residuals were examined.

1. Ordinary residuals ; $e_i = y_i - \hat{y}_i$,
2. Standardized residuals ; $r_i = e_i / Se_i$,
3. Partial residuals ; $e_i^* = e_i + \hat{x}_i \beta_i$

The data of this study was the lactation milk yield (kg), the average of respiration (min) and pulse rate (min) which were obtained during the nine weeks in one week interval from 36 East-Friesian x Awassi crossbreed sheep. The regression model was fitted to the data and residuals were evaluated by using the above formulas. After that the residuals were examined by using the graphical analysis method, we concluded that our least square regression equation was insufficient, but it might be sufficient if we transformed the data or applied the weighted least squares method.

1. GİRİŞ

Regresyon analizi, sebep-sonuç ilişkisi bulunan iki veya daha fazla değişken arasındaki ilişkiyi saptamada uygulanan istatistik yöntemlerden birisi olup Ziraat , Ekonomi, Tıp, Biyoloji vb. bilim dallarında geniş uygulama alanı bulmaktadır.

Regresyon analizinin en basit şekli iki değişkenli veya iki parametrelili olmaktadır. Ayrıca basit regresyon analizi, doğrusal regresyon ve doğrusal şekle dönüştürülebilen regresyon analizini kapsar. Eğer regresyon üç veya daha fazla değişken içeriyorsa bu çoklu regresyon olarak adlandırılır.

Gözlem çiftlerini en iyi temsil eden regresyon doğrusunun aşağıdaki gibi olduğunu varsayalım.

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i \quad (1.1)$$

Burada α ve β sırasıyla α ve β parametrelerinin tahminlerini, x bağımsız değişkeni ve y 'de bağımlı değişkeni göstermekte olup x 'in değeri verildiğinde y 'nin ortalama olarak değeri bulunabilmektedir.

Regresyon analizinde amaç tahmin yapmanın yanı sıra gözlem çiftlerini en iyi temsil eden regresyon eşitliğini bulmaktır. Bu da kalıntı (residual) analizinin konusunu oluşturur. Çünkü bir kalıntı analizi , tahmin edilen eşitlik ile veri arasındaki uyumu veya uyumsuzluğu ortaya koyan bir analiz şeklidir. Aynı zamanda kalıntı analizi, tahmin eşitliğinin verilere uygunluğunun istatistik olarak belirlenmesinden sonra tahmini değer ile gözlemler arasındaki farklılıkların dikkatli bir şekilde incelenmesini ve verilerdeki zayıf

parametre tahminlerine neden olan herhangi bir anormalliğin tespit edilmesini de sağlar (Cook ve Weisberg, 1982).

Kalıntı analizi doğrusal regresyon modelleri yanında doğrusal olmayan regresyon modelleri ile varyans analizinde de kullanılabilir. Gerçekte uydurulmuş herhangi bir modelin uygulanması ve bir grup veri için açıklanamayan varyasyonun ölçülmesi konularının incelenmesi için de yararlıdır (Draper ve Smith, 1981).

Kalıntı analizi açısından önemli olan kalıntıların bazı önemli belirgin özelliklerini ve (1.1) nolu eşitlikte kalıntıyı (e_i) oluşturan ayrılış tiplerini belirlemek ve bunları bir kalıntı analizinde göz önünde bulundurmak önemlidir.

Kalıntıların bazı belirgin özellikleri şunlardır;

a) $\{e_i\}$ lerin ortalaması sıfırdır.

$$\bar{e} = \left(\sum_{i=1}^n e_i \right) / n = 0$$

b) n adet kalıntının örnek varyansı ,

$$S^2 = \left(\sum_{i=1}^n e_i^2 \right) / \{n - (k+1)\} = K.K.T / \{n - (k+1)\}$$

eşitliği ile hesaplanır ; Bu eşitlikte,

n = kalıntı sayısını

k = bağımsız değişken sayısını

e_i = i'nci kalıntı değerini

KKT = kalıntı kareler toplamını

gösterir.

Kalıntı kareler ortalaması, regresyon analizi kullanarak elde edilir ve eğer model doğru ise bu σ^2 'nin sapmasız bir tahmini verir (Kleinbaum ve Kupper, 1978).

c) Kalıntılar $\{e_i\}$ bağımsız şans değişkenleri

değildir. Çünkü tahmin eşitliği kullanılarak elde edilirler ve bu e_j 'lerin toplamının sıfır olmasından açıkça görülür. Pratik olarak eğer kalıntı sayısı (n), bağımsız değişken sayısına(k) kıyasla çok büyükse e_j 'ler arasındaki bağımlılık etkisi herhangi bir kalıntı analizinde genel olarak ihmal edilebilmektedir (Kleinbaum ve Kupper, 1978).

Kalıntıların bu önemli özelliklerini verdikten sonra eşitlik (1.1)'de kalıntıyı oluşturan önemli ayrılış tiplerinin meydana gelmesine neden olan etmenler şunlardır.

1) Anormal gözlemler (Outlier)'in varlığı; çeşitli özel testlerle incelenebilir.

2) İlgisiz faktörlerin bulunması; model (1.1)'den çıkarılır ve bu faktörlerin seviyelerine karşı kalıntı grafikleri oluşturularak incelenebilir.

3) Model (1.1)'e dahil edilen bir faktörle doğrusal olmayan bir ilişkinin bulunması; Yani $\hat{y} = \alpha + \beta x$ gibi bir ilişki yerine $\hat{y} = \alpha + \beta x^2$ gibi bir ilişkinin olması; Bu ilişki, faktörün seviyelerine karşı çizilen kalıntı grafiği ve elde edilen ilişki eğrisi ile incelenebilir.

4) Farklı e_j 'ler arasında bağımlılık (korelasyon) bulunması; Bu faktör kısmi kalıntı analizi veya kalıntı periyod grafiklerinde incelenebilir.

5) Varyansın sabit olmaması; Bu durum kalıntıların veya karelerinin, uydurulmuş değerlere veya varyansı etkileyen faktörlere karşı grafikleri çizilerek incelenebilir.

6) e_j 'lerin normal dağılım göstermemesi; bu durum sıralı kalıntıların standart normal dağılımdan beklenen sıralı istatistiklere karşı grafikleri oluşturularak incelenebilir.

Kalıntı değeri, uyum eksikliği (Lack of fit) ve tekerrürler arası varyasyondan kaynaklanan salt hata (Pure Error) olmak üzere iki değerden oluşmaktadır. Bu durumda

salt hatayı şöyle açıklayabiliriz; X'in, m tane farklı değeri olduğunu varsayalım ve bu m değerlerin j'incisindeki x_j 'de ($j=1,2,\dots,m$) n_j gözlemleri olsun. Bu durumda,

$y_{11}, y_{12}, \dots, y_{1n_1}, x_1$ 'de n_1 tekerrürlü gözlemler
 $y_{21}, y_{22}, \dots, y_{2n_2}, x_2$ 'de n_2 tekerrürlü gözlemler
 $y_{31}, y_{32}, \dots, y_{3n_3}, x_3$ 'de n_3 tekerrürlü gözlemler
 $\dots$
 $y_{m1}, y_{m2}, \dots, y_{mn_m}, x_m$ 'de n_m tekerrürlü gözlemler

olur ve y_{ju} x_j 'de u'uncu gözlemdir ($u=1,2,\dots,n_j$).

Hepsi birlikte,

$$n = \sum_{j=1}^m \sum_{u=1}^{n_j} 1 = \sum_{j=1}^m n_j$$

gözlem vardır.

x_1 'deki n_1 gözlemden $\bar{y}_1$ ortalama için salt hata kareler toplamına yazılımı, y_{1u} 'nin içinde bulunduğu kareler toplamıdır. Yani,

$$\begin{aligned} \sum_{u=1}^{n_1} (y_{1u} - \bar{y}_1)^2 &= \sum_{u=1}^{n_1} y_{1u}^2 - n_1 \bar{y}_1^2 \\ &= \sum_{u=1}^{n_1} y_{1u}^2 - (\sum_{u=1}^{n_1} y_{1u})^2 / n_1 \end{aligned}$$

olur.

Tüm tekerrürlerin kareler toplamının dahil edilmesi ve

toplanması sonucu toplam salt hata kareler toplamı şu şekilde elde edilir.

$$\sum_{j=1}^m \sum_{u=1}^{n_j} (y_{ju} - \bar{y}_j)^2$$

ve bunun serbestlik derecesi ise,

$$n_e = \sum_{j=1}^m (n_j - 1) = \sum_{j=1}^m n_j - m = n - m$$

olur. Böylece salt hata kareler ortalaması,

$$S_e^2 = \left[\sum_{j=1}^m \sum_{u=1}^{n_j} (y_{ju} - \bar{y}_j)^2 \right] / \left(\sum_{j=1}^m n_j - m \right)$$

olur ve S_e^2 , σ^2 'nin tahmini değeri olup bu, modelin doğru veya yanlışlığı göz önünde bulundurulmaksızın hesaplanmıştır. Bu değer (S_e^2) tekerrürler içi kareler toplamlarının toplamını toplam serbestlik derecesine bölmek suretiyle de bulunur.

Eğer X_j noktasında sadece y_{j1}, y_{j2} gibi iki gözlem varsa o zaman,

$$\sum_{u=1}^2 (y_{ju} - \bar{y}_j)^2 = \{(y_{j1} - y_{j2})^2\} / 2$$

şeklindeki bir formül ile hesaplamak daha kolaydır ve 1 serbestlik derecesine sahiptir (Draper ve Smith, 1981). Gerçekte salt kareler toplamı, kalıntı kareler toplamının bir bölümüdür.

Biz kalıntıyı, X_j 'deki u 'ncü gözlem için

$$(y_{ju} - \hat{y}_j) = (y_{ju} - \bar{y}_j) - (\hat{y}_j - \bar{y}_j)$$

şeklinde yazabiliriz. Bu eşitliğin her iki tarafının karesi alınır ve u ve j üzerinden toplanırsa,

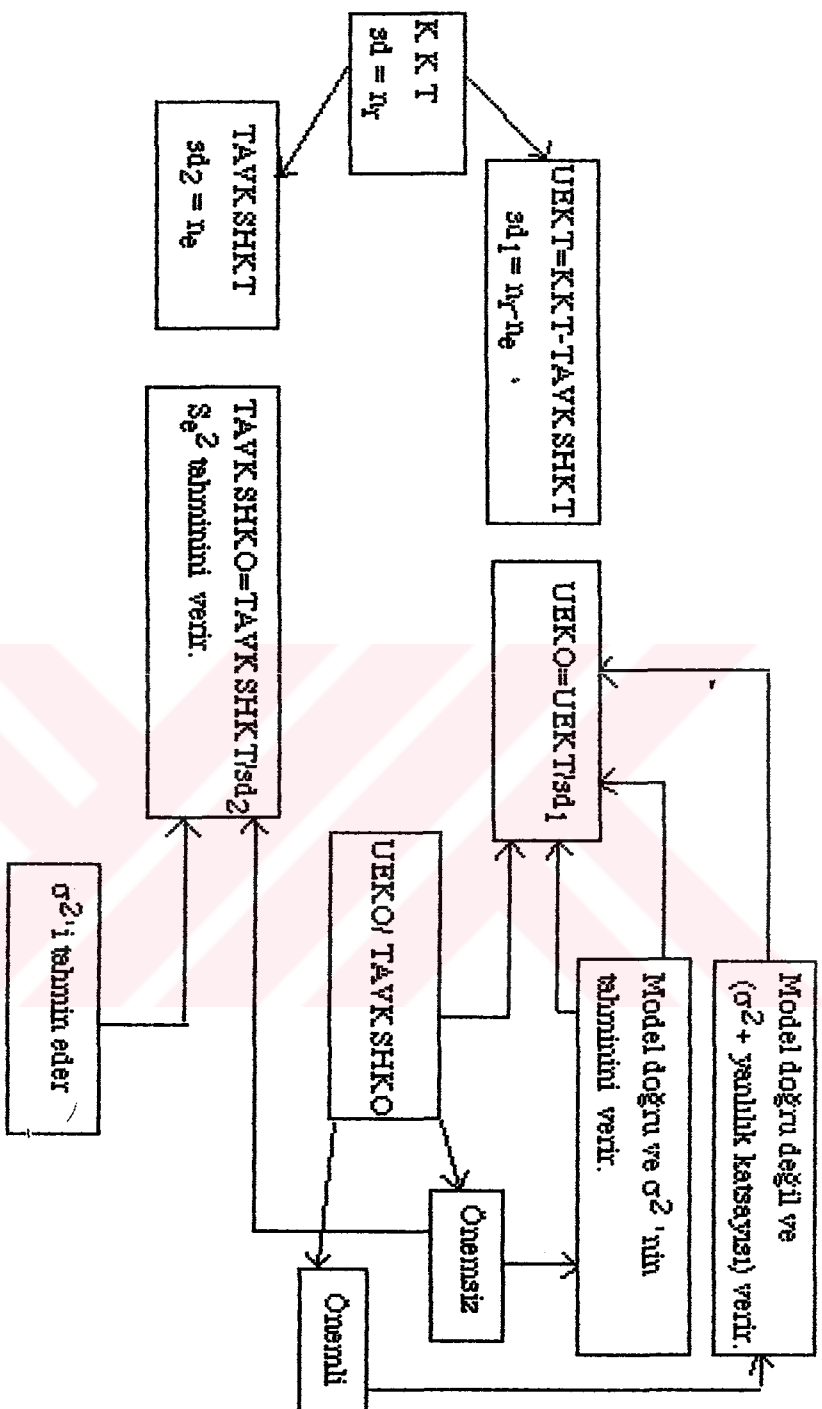
$$\sum_{j=1}^m \sum_{u=1}^{n_j} (y_{ju} - \hat{y}_j)^2 = \sum_{j=1}^m \sum_{u=1}^{n_j} (y_{ju} - \bar{y}_j)^2 + \sum_{j=1}^m n_j (\hat{y}_j - \bar{y}_j)^2$$

elde edilir. Bu eşitlikteki çarpımlar toplamı her bir j için u üzerinden toplandığından ortadan kalkar. Yukarıdaki eşitliğin sol tarafı kalıntı kareler toplamı, sağdaki ilk terim salt hata kareler toplamı, geriye kalan ise uyum eksikliği (lack of fit) kareler toplamıdır. Buradan da anlaşılacağı üzere kalıntı kareler toplamı, salt hata kareler toplamı ile uyum eksikliği kareler toplamından oluşmaktadır. Eğer uyum eksikliği kareler toplamı sıfır ise o zaman salt hata kareler toplamı kalıntı kareler toplamına eşit olur. Bu durum Şekil 1.1'de gösterilmiştir (Draper ve Smith, 1981).

Genel bir regresyon analizinde kalıntı ile ilgili bazı varsayımlar yapılır. Genel varsayımlar; kalıntılar bağımsızdırlar, ortalaması sıfır, varyansı σ^2 olan normal dağılış gösterirler.

Eğer modelimiz doğru ise, kalıntılar varsayımları doğrulayıcı bir temayül ortaya koymalıdırlar veya en azından yapılan varsayımlara ters düşmemelidirler.

Bu çalışmanın amacı, kalıntı analizinde uygulanan özellikle grafiksel metotlar kullanılarak, verileri temsil edebilecek en uygun modeli, daha gerçekçi bir yaklaşım ve grafiksel analizler kullanılarak kalıntı analizleri ile birlikte regresyon uygulamalarını yapmanın sağladığı faydaları karşılaştırmalı olarak göstermektir.



KKT = Kalın Kareler Toplamı, UEKT= Uyum Eksikliği Kareler Toplamı
 TAVKSHKT = Tekerütler Arası Varyasyondan UEKO= Uyum Eksikliği Kareler Orlanması
 Kaynaklanan Salt Hata Kareler Toplamı,
 TAVKSHKO = Tekerütler Arası Varyasyondan
 Kaynaklanan Salt Hata Kareler Orlanması

Şekil 1.1. Kalın Kareler Toplamının Uyum Eksikliği ve Salt H.K.T'na Ayrılması(Draper ve Smith,1961)

2. ÖNCEKİ ÇALIŞMALAR

2.1. Basit (Ordinary) Kalıntılar

Bir regresyon modeli genelde,

$$Y = X\beta + \epsilon \quad (2.1.1)$$

şeklindedir. Burada;

$Y = (n \times 1)$ boyutlu şans değişkeni vektörü,

$X = (n \times (k+1))$ boyutlu bilinen katsayı matrisi,

$\beta = ((k+1) \times 1)$ boyutlu bilinmeyen parametre vektörü,

$\epsilon = (n \times 1)$ boyutlu şans değişkeni vektörü olup,

ortalaması sıfır ($E(\epsilon) = 0$), varyansı ($\text{var}(\epsilon) = \sigma^2 I$) sabittir.

Ascombe ve Tukey (1963), sözel bir ifade ile basit kalıntıyı

$$\text{Kalıntı} = (\text{Gözlem Değeri}) - (\text{Uyum Değeri})$$

olarak tanımlamışlardır. Ascombe ve Tukey (1963), yaptıkları bu çalışmada, daha çok en-küçük kareler uyumu veya varyans analizinden sonra yapılabilecek kalıntı analiz yöntemleri üzerinde durmuşlardır. Ayrıca uygun geleneksel kalıntı analiz yöntemlerinin seçimi veya geçerliliğini belirlemek amacıyla bazı grafiksel, sayısal ve karışık analiz tekniklerini kalıntılara dayalı satır sütun uyumundan çıkarak iki yönlü tabloları inceleyerek uygulama sonuçlarını sunmuşlardır.

Cox ve Snell (1968), Eğer β , β 'nın en küçük kareler tahmincisi ise ve (2.1.1) nolu eşitlikte yerine konursa o zaman Ascombe ve Tukey(1963)'in kullandığı bu sözel ifadeyi

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

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

$$e = Y - \hat{Y} \quad (2.1.2)$$

şeklinde bir formül ile tanımlanmışlardır. Burada;

$Y = (n \times 1)'$ boyutlu gözlem vektörünü,

$X = (n \times (k+1))$ boyutlu bilinen katsayı matrisini,

$\hat{Y}$

$Y = (n \times 1)$ boyutlu Y 'nin tahmin vektörünü

$e = (n \times 1)$ boyutlu kalıntı tahminleri vektörünü

göstermektedir.

Durbin (1969)'ın yaptığı çalışmada regresyonda seri korelasyonlar için en küçük kareler kalıntılarının periyod grafiklerine dayanan testler üzerinde durarak, bu amaç için bilinen en iyi metodun örneğin yığılmalı periyod grafiğini çıkarmak ve elde edilen grafiği Kolmogorov-Smirnov limitleri ile karşılaştırmak olduğunu bildirmiştir. Bunun için periyod grafiği en küçük kareler kalıntılarından hesaplanır, regresyon modelindeki hatalardaki düşük frekanslı varyasyonun yüksek frekanslı varyasyonla bağlantısını test için doğrunun bir parçası grafikte çizilir. Eğer çizginin ucu üste geçerse seri için bağımsızlık hipotezleri reddedilir. Eğer çizgi alta geçerse hipotezler red edilmez. Ortada kalırsa bir karar verilemez.

Kleinbaum ve Kupper (1978)' da Cox ve Snell(1968) ile aynı ifadeyi kullanarak i'nci kalıntıyı(e_i), i'nci gözlem değeri (y_i) ile i'nci tahmin değeri($\hat{y}_i$) arasındaki fark olarak tanımlamışlar ve elde ettikleri bu değerlerin her bir j için x_{ji} 'lere , zamana, y 'lara ve bize yardımcı olabilecek diğer değişkenlere karşı grafiklerini oluşturarak bazı özel varsayımların mevcudiyetini araştırmışlardır.

Draper ve Smith (1981), ise kalıntı kareler toplamının iki kısımdan oluştuğunu belirterek bunun ,

$$\sum \sum e_i^2 = \sum_{j=1}^m \sum_{u=1}^{n_j} (y_{ij} - \hat{y}_j)^2 = \sum_{j=1}^m \sum_{u=1}^{n_j} (y_{ju} - \bar{y}_j)^2 + \sum_{j=1}^m n_j (\bar{y}_j - \bar{y})^2 \quad (2.1.3)$$

şeklindeki bir eşitlikle göstermişlerdir.

Bu eşitliğin sol tarafı kalıntı kareler toplamını , sağdaki ilk terim salt hata kareler toplamını (pure error) , geriye kalan terim ise uyum eksikliği kareler toplamını (lack of fit) göstermektedir. Eğer uyum eksikliği kareler toplamı sıfır ise kalıntı kareler toplamı salt hata kareler toplamına eşit olur.

Cook ve Weisberg (1982), ise (2.1.2) nolu eşitliği,

$$e = (I - V)Y \quad (2.1.4)$$

şeklinde ifade etmişlerdir. Bu eşitlikte,

$V = (V_{ij}) = X(X'X)^{-1} X'$ olup simetrik ve idempotent bir matristir .

2.2. Standart Kalıntılar

Ellenberg (1973) , (2.1.2) nolu eşitlikle elde edilen değerleri .

$$r_i = \frac{e_i}{S\sqrt{m_{ii}}} \quad (2.2.1)$$

şeklinde ifade ederek elde ettiği değeri standart kalıntı olarak adlandırmıştır.

Bu eşitlikte;

r_i = standart kalıntıyı,

e_i = basit kalıntıyı,

$m_{ii} = (I-V)$ matrisinin köşegen elemanları olup V matrisi Cook ve Weisberg(1982)'nin (2.1.4) nolu eşitlikte kullandığı matristir.

S = Varyans Analiz Tablosundaki Hata kareler Ortalamasının karekökünü ifade etmektedir.

Ellenberg bu çalışmasında genel doğrusal bir regresyondan elde edilen n adet standart en küçük kareler kalıntısının, k tane alt kümesi için k değişkenli ortak dağılımını (joint distrubution) çıkarmıştır.

Kleinbaum ve Kupper(1978) ise standart kalıntı değerlerini

$$r_i = \frac{e_i}{S} \quad (2.2.2)$$

formülü ile hesaplamıştır. Bu eşitlikte,

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

olup kalıntıların standart sapmasını göstermektedir. Ayrıca bu çalışmada Kleinbaum ve Kupper (2.2.2) nolu eşitlik kullanılarak elde edilen değerlerin normal olasılık kağıtları üzerinde normal veya yarı normal grafiklerini çizerek özellikle hatanın normallik varsayımının mevcut olup olmadığının belirlenmesinde kullanılmışlardır.

Cook ve Weisberg (1981), (2.2.1) nolu eşitliği,

$$r_i = e_i / \hat{\sigma} \sqrt{(I - v_{ii})} \quad (2.2.3)$$

şeklinde ifade ederek elde edilen bu değerleri içsel (internal) standart kalıntılar olarak isimlendirmişlerdir.

Burada,

$$\hat{\sigma}^2 = \sum e_i^2 / (n - k - 1) \text{ olup Varyans Analiz Tablosundaki Hata Kareler Ortalamasıdır.}$$

$n - k - 1$ = serbestlik derecesini

e_i = kalıntı değerlerini gösterir.

$$t_i = e_i / (\hat{\sigma}_{(i)} \sqrt{(I - v_{ii})}) \quad (2.2.4)$$

formülü ile elde edilen standart kalıntı değerlerini ise dışsal (external) standart kalıntılar olarak adlandırmışlardır.

Burada,

$$\hat{\sigma}_{(i)} = [((n - p)\hat{\sigma}^2 - (e_i^2 / (I - v_{ii}))) / (n - k - 1)]$$

dır ve normallik özelliğinden e_i 'ler ve $\hat{\sigma}_{(i)}$ 'ler bağımsızdırlar.

Thomas ve Ryan (1982) ise (2.2.2) nolu eşitlikteki S yerine Se_i kullanarak bunu, aynı x değerleri için elde edilen Y değerleri arasındaki varyasyona göre ayrı ayrı hesaplayarak i'nci standart kalıntıyı,

$$r_i = \frac{e_i}{Se_i} \quad (2.2.5)$$

şeklinde ifade etmişler ve bu formül ile hesaplama yapan bir de bilgisayar paket programı (Minitab software, 1982) yazmışlardır. Burada,

$$Se_i = \sqrt{H.K.O - \hat{Var}(Y_i)} \quad (2.2.6)$$

yardımla, burada ise

$$\hat{Var}(Y_i) = H.K.O. X_i(X'X)^{-1}X_i' = S^2V$$

2.3. Kısmi (Partial) Kalıntılar

Larsen ve McCleary (1972), yaptıkları çalışmada çoklu regresyon analizinde kullanılmak üzere kısmi kalıntı analizi olarak adlandırılan grafik analiz yöntemi üzerinde durmuşlardır. Bu yöntemde i'nci kısmi kalıntı değeri,

$$\begin{aligned} e_i^* &= e_i + X_i\beta_i \\ &= Y_i - \sum_{j \neq i} X_j\beta_j \end{aligned} \quad (2.3.1)$$

formülü ile hesaplanıp, i'nci bağımsız değişken için düzeltilmiş bağımlı değişken vektörü olarak düşünülmektedir. Larsen ve McCleary bu eşitlikle elde edilen değerlerin grafikleri çizildiğinde; doğrusallıktan sapmaları, bu sapmaların büyüklüğünü (outlier gibi) ve yönünü, varyansların homojen olup olmama durumunu ve doğrusal olmayan ilişkileri kolayca belirlenebileceğini bildirmişlerdir.

Cook ve Weisberg (1982), Larsen ve McCleary (1972)'nin (2.3.1) nolu eşitlikteki ifadesini,

$$e(i) = y_i - \mathbf{x}'_i \hat{\beta}(i), \quad i=1,2,3,\dots,n \quad (2.3.2)$$

şeklinde göstererek, i'nci x değişkeni atılarak uydurulan modelden hesaplanan kalıntı değerlerini tahmini (Predicted) kalıntılar olarak adlandırmışlardır. Tahmini kalıntıların doğru modelin seçiminde çok önemli bir kriter olduğunu ifade ederek, daha iyi olan bir modelin diğerine nisbeten tahmini kalıntı kareler toplamı (PRESS) değerlerinin daha küçük olan model olduğunu vurgulamışlardır.

2.4 . Yapılan Diğer Çalışmalar ve Uygulamaları

Cox ve Snell (1968), kalıntıların bilinen doğrusal hipotezlerden çok daha genel durumlardaki tanımlarını vermişler, kısaca kalıntıların dağılımları ve transformasyonları üzerinde durmuşlardır. Sonuçta kalıntıların tahminde etkili olan şans değişkenleri ile yaklaşık olarak aynı dağılıma sahip olduklarını göstermişlerdir.

Loynes , M.R., (1969)' de bu çalışması, Cox ve Snell (1968)'in çalışmalarını tamamlayıcı bir nitelik göstererek,

$$Y_i = g_i(\beta, e_i), \quad i=1,2,\dots,n \quad (2.4.1)$$

şeklinde Y_i 'lerin β ve e_i 'lerin bir fonksiyonu olarak ifade eden modelde

$$e_i = h_i(Y_i, \beta) \quad (2.4.2)$$

eşitliğini elde etmiş ve buradan kalıntıyı,

$$R_i = \hat{h}_i(Y_i, \hat{\beta}) \quad (2.4.3)$$

şeklinde tanımlamıştır. Burada, $\hat{\beta}$, β 'nın maksimum Likelihood tahmincisidir.

R_i - E_i farklarının ortalaması ve varyansları $n-1$ sıralı olup bu yaklaşımlar sonucu elde edilirler.

Behnken ve Draper (1972), yaptıkları çalışmada regresyon varsayımlarını doğrulayıcı ideal bir kalıntı grafiği örneğinin düz bir band oluşturduğunu, bunun Fisherian sabit etkili faktöriyel planlarda daima doğru olduğunu bildirmişlerdir. Bu durumun genellikle , sürekli değişkenli modellerde birçok grafikte pratik limitler içerisinde doğru olduğunu fakat kalıntıların hepsinin aynı varyansa sahip olmaması nedeniyle/teorik olarak kesin doğru olmadığını ifade etmişlerdir. Behnken ve Draper (1972)'in burada iki amacı vardır;

1) Bu gibi grafiklerde bazı kalıntıların normal dağılım bandının dışında görünse dahi bunların önemsiz olduğunu ortaya koymak,

2) Kalıntı varyanslarının sabitliğini, herhangi bir anormal veri(outlier) atılmadan veya kalıntı grafiği varsayımına göre yetersiz bir model düzeltilmeden önce kontrol etmektir.

Habermann (1973), frekans tablolarında log-lineer modellerle birleştirilmiş kalıntı analizleri için kendisinin genelleştirdiği yöntemleri sunmuş, sonuçları iki yönlü tablolar ve logit modellere uygulamıştır.

Farebrother (1976)'nın bildirdiğine göre Theil(1965) 'x'ler bağımlı olduğunda varyasyon matrisi $\sigma^2 M$ 'in dağılım hakkındaki varsayımların testleri için gerekli pensiplere uygun olmadığını , bu problemi çözmek amacıyla doğrusal, sapmasız fakat skalar (BLUS) bir varyans matrisine sahip olan dağılım tahmincilerini önerdiğini, bu nedenle bu tahminlere BLUS tahmincileri adı verildiğini ve eğer,

$$X = \begin{bmatrix} X_0 \\ X_1 \end{bmatrix}, \quad \epsilon = \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \end{bmatrix} \quad \text{ve} \quad \hat{\epsilon} = \begin{bmatrix} \hat{\epsilon}_0 \\ \hat{\epsilon}_1 \end{bmatrix}$$

şeklinde kolayca bölünebiliyorsa ve X_0 , $k \times k$ boyutlu tekil olmayan matris ise, ϵ_1 'in BLUS tahmincileri,

$$\hat{\epsilon}_1 = \epsilon_1 - X_1 X_0^{-1} \left[\sum_{j=1}^h d_j / (1+d_j) \right] q_j q_j^T \epsilon_0 \quad (2.4.4)$$

eşitliği ile elde edilebileceğini bildirmiştir. Bu eşitliğin çözümünü için $X_0(X'X)^{-1}X_0'$ QD^2Q olarak köşegenleştirilir. Burada,

Q : $k \times k$ boyutlu j 'nci sütunu q_j olan ortonormal karakteristik vektörü,

D^2 : j 'nci köşegen elemanı d_j^2 olan köşegen matrisi göstermektedir, $0 < d_j < 1, i=1,2,\dots,h$.

Schwarzmann (1991), Beckmann ve ark.(1987)'nin genel kalıntı analizlerine benzeyen karışık ve sabit etkili modeller için kısmi etki analizleri üzerinde durmuşlar ve sonuçta kısmi etki analizlerinin bir istatistikçiyi genel kalıntı analizi metotlarına götürdüğünü ifade ederek sabit etkili modellerde h_{max} ve r vektörlerinin birbirleriyle orantılı olduğunu ve böylece benzer çözümlere götüreceğini belirtmiştir. Burada h_{max} verilen modelde likelihood'un maksimum eğrisi olup lokal etki analizi metodunda ana unsurdur. Ayrıca Schwarzmann kalıntı teşhis metotlarının en büyük lokal duyarlılığın yönünün kontrolü ve hatalı verinin özel bir durumu olduğunu gösterdiği için lokal etki analizinde çok daha geçerli bir durum oluşturan bir de teorem ispatlamışlardır.

3. MATERİYAL ve YÖNTEM

3.1. Materyal

Bu çalışmada materyal olarak Ç.Ü. Ziraat Fakültesi Döner Sermaye İşletmesindeki Ost-Friz x İvesi ırkı 36 koyuna ait birer hafta ara ile 9 hafta boyunca alınan solunum sayısı(dk) (SolS)nabız Sayısı (dk)(NaS) değerlerinin ortalamaları ve bu koyunlara ait laktasyon süt verimleri (kg/koyun) kullanılmıştır (Çizelge 3.1).

Çizelge 3.1. Solunum Sayısı(dk),Nabız sayısı (dk), Süt Verimleri(kg/koyun).

Koy.No	SolS	NaS	SütV	Koy.No	SolS	NaS	SütV
1	62	100	169	19	71	80	152
2	61	96	181	20	93	80	105
3	44	80	187	21	79	35	112
4	43	76	187	22	78	96	102
5	65	88	178	23	46	102	155
6	37	92	133	24	81	114	121
7	78	84	107	25	74	75	124
8	82	90	111	26	65	81	107
9	86	86	124	27	54	90	178
10	82	92	100	28	46	63	180
11	56	106	153	29	57	70	150
12	86	104	110	30	36	75	160
13	70	100	105	31	71	75	145
14	61	102	178	32	73	66	140
15	47	75	175	33	99	75	121
16	41	90	181	34	80	90	132
17	61	93	160	35	84	75	130
18	29	100	175	36	49	75	155

3.2. Yöntem

Regresyon analizinde hata (ϵ_i) ile ilgili olarak ortalaması sıfır, varyansı σ^2 olan normal dağılım gösterdiği varsayılır. Normal dağılım göstermesi varsayımı, parametrik önem testleri için mutlak gereklidir (Kleibaum ve Kupper, 1978).

Eğer model gerçekten verilere uyum sağlıyorsa yani model doğru ise kalıntılar sözü edilen tüm varsayımları sağlaması gerekir. Başka bir ifade ile varsayımları doğrulayıcı bir temayül ortaya koymalıdır veya en azından yapılan varsayımlara ters düşmemelidir. Bu durumun bir kalıntı analizinde daima göz önünde tutulması gerekir..

Uydurulan bir modelin doğru olup olmadığını kontrol için bazı kalıntı test yöntemleri uygulanır. Bu yöntemler içinde çoğunlukla tahminler doğru olmadığında bunu genelde açığa çıkaran ve yapılması kolay olan grafik metotlar kullanılır. Bu nedenle bizde bu çalışmada grafik metotları inceleyeceğiz.

En basit regresyon modeli genelde,

$$Y = XB + \epsilon \quad (3.2.1)$$

şeklindedir. Burada ;

Y: $(n \times 1)$ boyutlu bağımlı şans değişkeni vektörünü,
X: $(n \times (k+1))$ boyutlu bağımsız değişken matrisini,
B: $((k+1) \times 1)$ boyutlu bilinmeyen parametre vektörünü,
 ϵ : $(n \times 1)$ boyutlu gözlenmeyen şans değişkeni vektörünü gösterir.

Bu durumda y gözlem değeri

$$y_i = \alpha + \beta_1 x_1 + \beta_2 x_2 + \varepsilon_i \quad i = 1, 2, \dots, n. \quad (3.2.2)$$

olur. Bu modelde parametre tahminleri için en - küçük kareler yöntemi uygulanırsa ,

$$(i) \hat{\beta} = (X'X)^{-1} X'Y \quad (3.2.3)$$

olur ve (3.2.2) nolu eşitlikte yerine yazılırsa uydurulmuş model,

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}_1 \hat{x}_1 + \hat{\beta}_2 \hat{x}_2 + \varepsilon_i \quad i = 1, 2, \dots, n. \quad (3.2.4)$$

olur.

Bu çerçeve içinde bu çalışmada şu analizler yapılmıştır.

3.2.1. Basit Kalıntılar

i 'nci kalıntıyı , y_i gözlem değeri ile $\hat{y}_i$ tahmini değeri arasındaki fark olarak tanımlanır (Kleinbaum ve Kupper, 1978).

Eğer β_1, β_2 'nın en küçük kareler tahmincisi ise ve model (3.2.1) 'de yerine konursa kalıntı,

$$\begin{aligned} Y &= X\hat{\beta} + e \\ e &= Y - X\hat{\beta} \\ e &= Y - \hat{Y} \end{aligned} \quad (3.2.5)$$

şeklinde hesaplanır (Cox ve Snell, 1968). Bu şekilde hesaplanan kalıntıların grafikleri çeşitli şekillerde çizilebilir. Bunlar;

1. Kalıntıların tümünün grafiği,
2. Eğer verilerle ilgili sıralama biliniyorsa kalıntıları zaman serisinde gösterme,
3. Uydurulan y_i değerlerine karşı grafik çizme,
4. Bağımsız x_{ji} ($j=1,2,...,k$ için) değişkenlerine karşı grafik çizme,
5. Üzerinde çalışılan problemin özelliğine uygun bir yol izleyerek grafik çizme.

Bu grafikler çizildiğinde yeterli ve modelimizi doğrulayıcı bir kalıntı grafiği Şekil 3.2.1'deki gibi bir izlenim verir(Draper ve Smith,1981).



Şekil 3.2.1. Yeterli ve Modeli Doğrulayıcı Kalıntı Grafiği.

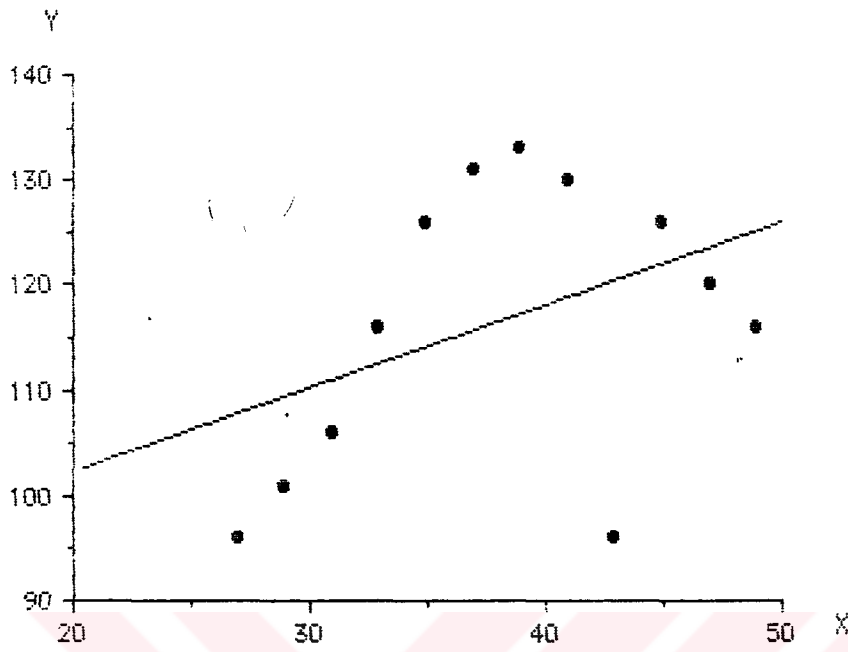
Yukarıda verilen basit kalıntı grafikleri çizildiğinde kullanıcı için önemli olan birçok bilgiler edinilebilir. Genel olarak karşılaşılabilecek durumlar şunlardır.

3.2.1.1. Regresyon Fonksiyonun Doğrusal Olmaması

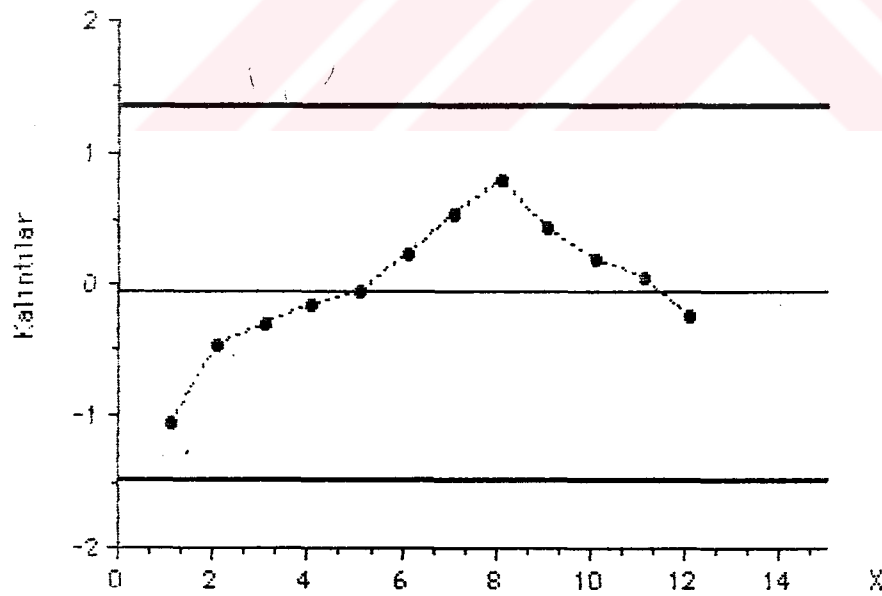
Doğrusal regresyon fonksiyonunun incelenen veriler için uygun olup olmadığı, çoğu zaman gözlemlerin serpmeye diyagramı üzerine, uydurulan regresyon doğrusunun çizilmesi ile anlaşılır. Şekil 3.2.2a Uydurulan regresyon doğrusu ve gözlemlerin serpmeye grafiğini göstermektedir. Bu grafik incelenen verilerin doğrusal bir regresyon fonksiyonuna uygun olmadığını göstermektedir.

Şekil 3.2.2b, aynı örneğin x_1 bağımsız değişkenine karşı çizilmiş basit kalıntı grafiğidir. Bu grafik de doğrusal regresyon fonksiyonunda uyum eksikliği olduğu ifade etmektedir. Çünkü kalıntılar sıfırdan sistematik bir şekilde ayrılışlar göstermektedir. Küçük x değerleri için negatif, ortada pozitif ve büyük x değerleri için yine negatif değerler olmaktadır ve bu bize doğrusal olmayan bir regresyon fonksiyonuna ihtiyaç olduğunu gösterir.

Bu durumda, 3.2.2a ve 3.2.2b Şekilleri regresyon fonksiyonunun yeterliliğinin incelenmesinde aynı derecede etkilidirler. Kalıntı grafiği Şekil 3.2.2b, serpmeye grafiği Şekil 3.2.2a'ya göre bazı üstünlüklere sahiptir ; kalıntı grafiği modelin yeterliliğinden başka diğer yönlerin (varsayımlar gibi)incelenmesi içinde kolayca uygulanabilir. Serpmeye grafiğinde örneğin dik bir eğim varsa serpmeye grafiğinden doğrusal regresyon fonksiyonun yeterliliğinin belirlenmesi oldukça zordur. Diğer yandan bir kalıntı grafiği bu durumlarda regresyon doğrusu etrafındaki herhangi bir sistematik ayrılışı daha açık bir şekilde gösterilebilir.



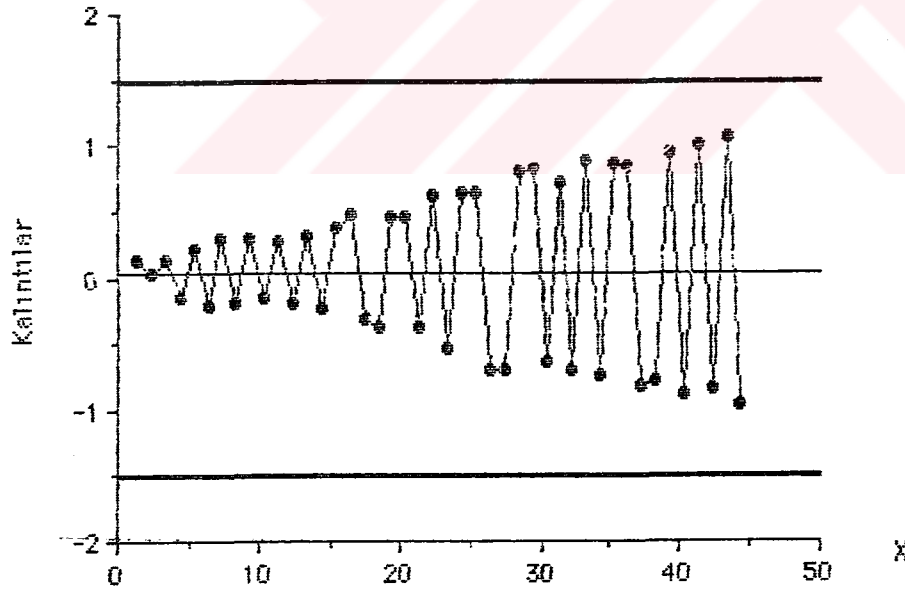
Şekil 3.2.2a. Doğrusal Olmayan Regresyon Fonksiyonunu Gösteren Serpme Grafiği(Netter ve Wassermann, 1973).



Şekil 3.2.2b. Doğrusal Olmayan Regresyon Fonksiyonunu Gösteren Kalıntı Grafiği(Netter ve Wassermann, 1973).

3.2.1.2. Hata Varyansının Sabit Olmaması

Bağımsız bir değişkene karşı çizilen bir kalıntı grafiği sadece doğrusal bir regresyon fonksiyonunun uygun olup olmadığının incelenmesi için faydalı değildir, aynı zamanda hata varyansının sabit olup olmadığının belirlenmesi için de faydalıdır. Şekil 3.2.3. bir regresyon uygulaması için x bağımsız değişkenine karşı kalıntı grafiğini göstermektedir. Bu grafikte x değerleri yükselirken kalıntıların sıfırdan ayrılışları da artmaktadır. Bu durum hata varyansının x değeri büyüdükçe artmakta olduğunu gösterir. Birçok işletme ve ekonomi uygulamalarında kalıntılar Şekil 3.2.3'deki tipte trapezoidal bir dağılım gösterirler (Netter ve Wassenman, 1973). Bu durumda tartılı en küçük kareler analizinin uygulanması önerilir.

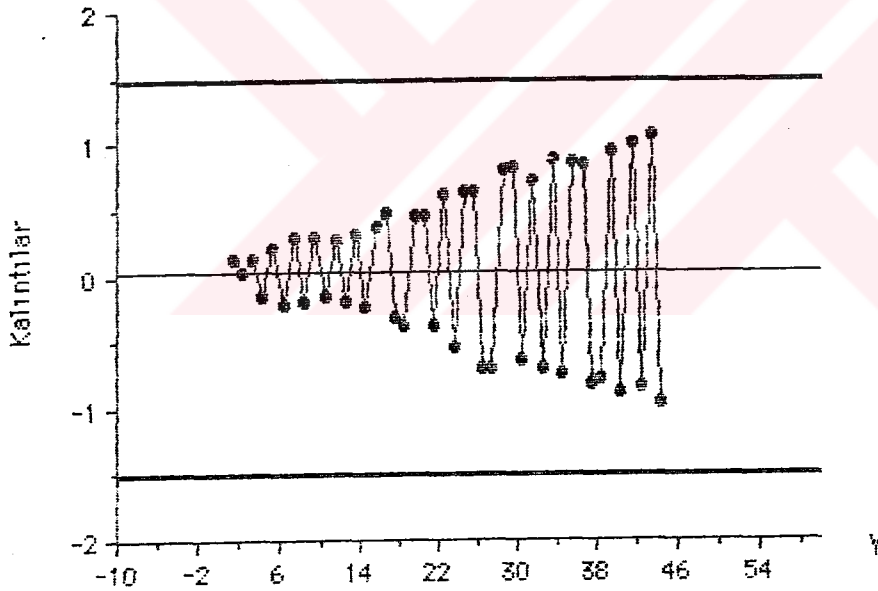


Şekil 3.2.3. Sabit Olmayan Varyansı Gösteren Kalıntı Grafiği.

y_i uydurulan değerlerine karşı çizilen kalıntı grafikleri, özellikle regresyon fonksiyonu doğrusal olmadığında veya çoklu regresyon modeli ile çalışıldığında kalıntı varyansının

sabitliğinin incelenmesinde etkilidir. Şekil 3.2.4. y_i uydurulan değerlerine karşı kalıntı (e_i) grafiğini göstermektedir. Bu

grafikte de hata varyansının y_i değerleri artıkça arttığı görülmektedir. Bu durumda da yine tartılı en küçük kareler analizi yapılması veya regresyon analizi yapılmadan önce y_i gözlemlerinde bir transformasyon yapılması önerilir. Sabit bir kalıntı varyansı için bu grafik Şekil 3.2.1. deki gibi düz bir band eğilimindedir.



Şekil 3.2.4. Sabit Olmayan Varyansı Gösteren Uydurulan y Değerlerine Karşı Kalıntı Grafiği(Netter ve Wassermann, 1973).

Burada şöyle bir soru akla gelebilir. "Neden $\hat{Y}_i$ 'lere karşı değilde Y_i 'lara karşı kalıntı grafikleri oluşturulur?" Çünkü, kalıntılar ve Y_i 'ler genelde birbiriyle ilişkilidir. Fakat kalıntılarla $\hat{Y}_i$ 'lar ilişkili değildir. İlişki olmaması demek, doğrunun eğiminin sıfır olması demektir. Bunu görmek için, kalıntı grafiklerini (i) y_i ve (ii) $\hat{Y}_i$ değerlerine karşı gelen ordinat olarak düşünmek ve noktalara doğru en küçük kareler doğrusunu bulmak yeterlidir.

(i) için eğim $1-R^2$ 'dir.

$$\sum (e_i - \bar{e})(Y_i - \bar{Y}) = \sum e_i(Y_i - \bar{Y}), \text{ (Eğer } \alpha \text{ katsayısı modelde}$$

$$\text{var ise } \bar{e} = 0 \text{ 'dır)}$$

$$= \sum e_i Y_i$$

$$= e'Y$$

$$= e'e,$$

$$\{ e'e = Y'(I-V)'(I-V)Y$$

$$= Y'(I-V)Y$$

$$= Y'e$$

$$= \text{Kalıntı K.T.}$$

$$= e'Y \text{ 'dir. Burada}$$

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

$$\sum (e_i - \bar{e})^2 = \sum e_i^2 = e'e$$

$$\sum (Y_i - \bar{Y})^2 = \text{Düzeltilmiş Genel K.T.}$$

$$r_{ey} = (e'e) / \{ (e'e) \sum (Y_i - \bar{Y})^2 \}^{1/2}$$

$$= \text{Kalıntı K.T.} / \text{Düzeltilmiş Genel K.T.}$$

$$r_{ey} = \{1-R^2\}^{1/2}$$

şeklinde bulunur. Bu, ancak model mükemmel bir şekilde uydurulursa sıfır olacaktır. Diğer durumlarda $\hat{Y}_i$ 'lere karşı çizilen kalıntı grafikleri $1-R^2$ eğimini göstereceklerdir.

$$\begin{aligned} \text{(ii) } \sum (e_i - \bar{e})(\hat{Y}_i - \bar{Y}) &= \sum e_i \hat{Y}_i, \quad (\text{yukarıdaki benzer işlemler} \\ &= e'Y \quad \text{yapılırsa}) \\ &= Y'(I-R)'RY, \quad \left\{ \begin{aligned} \hat{Y} &= XB \\ &= X(X'X)^{-1}X'Y \\ &= RY \end{aligned} \right\} \\ &= Y'(R-R^2)Y = 0 \\ &\quad \{e_{ey} = 0 \text{ olduğu için}\} \end{aligned}$$

böylece,

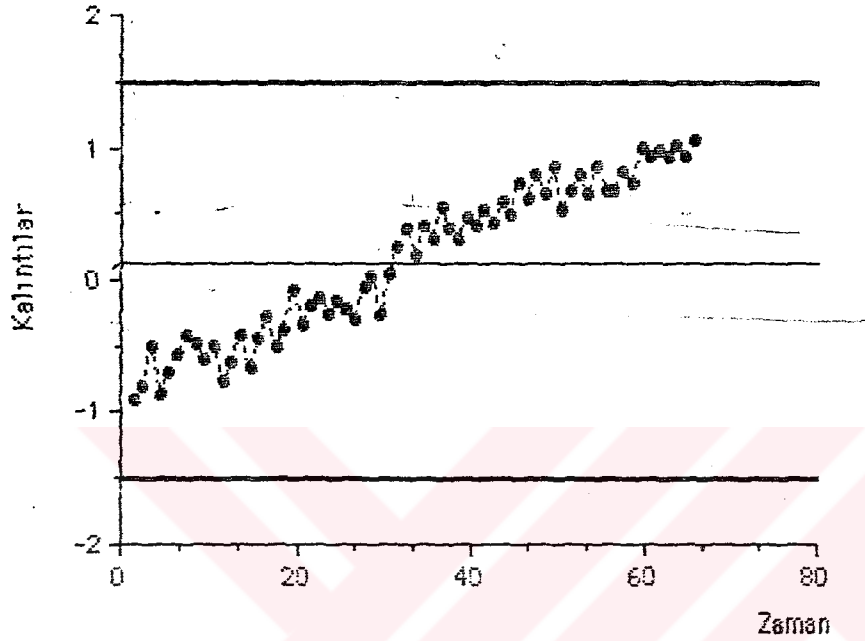
$$r_{ey} = 0$$

olur.

3.2.1.3. Hatanın Bağımsız Olmaması

Veriler zamana bağlı elde edildiğinde, model zamanı açıkça bir değişken olarak kapsamasa bile kalıntıların zamana karşı grafiklerini oluşturmak gerekir. Bunda amaç, zaman ile hatalar arasında herhangi bir ilişkinin olup olmadığını görmektir. Zamana karşı kalıntı grafikleri çizildiğinde, kalıntılar arasında belirgin bir korelasyon varsa Şekil 3.2.5'dekine benzer bir dağılım elde edilir. Başlangıçta çoğunlukla negatif kalıntılar varken sonraları pozitif kalıntılar ortaya çıkmıştır. Bu da bize açıkça, zamanla ilişkili bazı etkilerin var olduğunu gösterir. Bu durum zamana eklenmesi gereken bazı doğrusal terimlerin olduğunu ve bu terimlerin modele dahil edilmesi gerektiğini gösterir. Bazen hatanın

bağımsızlığı ile ilgili problemlerde modelden çıkarılan önemli bir değişkenin (zaman gibi) incelenmesi faydalıdır.



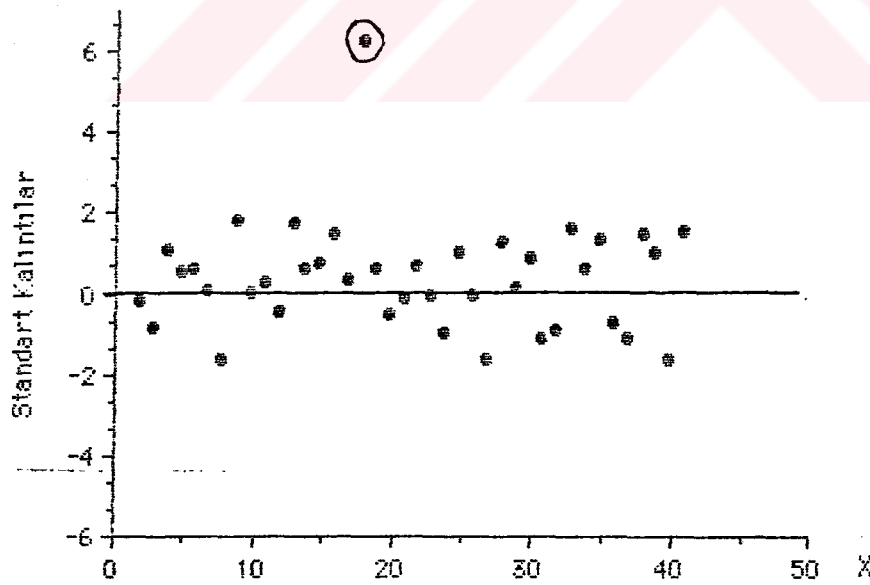
Şekil 3.2.5. Hata Teriminin Bağımsız Olmadığını Gösteren Kalıntı Grafiği(Netter ve Wassermann,1973).

Hatalar bağımsız olduğu zaman kalıntıların sıfır çizgisi etrafında Şekil 3.2.1' de olduğu gibi az veya çok şansa bağlı olarak bir dağılım göstermesi yani düz bir band eğilimi içerisinde olması beklenir.

3.2.1.4. Anormal Gözlemlerin (Outlier) Olması

Anormal gözlemler ekstrem gözlemlerdir. Bir kalıntı grafiğinde sıfır çizgisinden dört veya daha fazla standart sapma kadar uzakta olan noktalardır(Netter ve Wasserman,1973). Şekil 3.2.6. daire içerisine alınmış bir anormal gözlem içeren standart kalıntı grafiğini göstermektedir. Bu kalıntı uydurulan

değerden 6 standart sapma fark göstermektedir. Anormal gözlemlerin olması ortaya büyük zorluklar çıkarır. Bir anormal gözlemle karşılaşıldığında şüphelenilecek ilk şey, gözlemin bir ölçme hatası veya konu dışı bir etkinin sonucu olduğudur ve bu gözlem hesaplarda kullanılmamalıdır. Böyle bir gözlemin atılmasının sebebi, en küçük kareler yöntemi, sapmalar kareler toplamını minimize ettiği için bu anormal gözlem uydurulan doğruyu kendine doğru çekmesidir. Eğer anormal gözlem bir ölçme hatası veya konu dışı bir sebepten olmuşsa ortaya çıkan bu durum yanıltıcı olabilir. Diğer yandan anormal gözlemler, modelden çıkarılan herhangi bir bağımsız değişkenle interaksiyon sonucu oluşmuş ise bu gözlemler önemli bilgiler ifade ederler. Çoğu zaman önerilen güvenilir bir kural vardır; Bir anormal gözlem; eğer kayıt tutarken yanlış bir hesaplama, alet bozukluğu veya benzer bir etki ile oluşmuş ise ve bunun için bir delil varsa atılabilir (Netter ve Wasserman, 1973).



Şekil 3.2.6. Anormal Gözlemi Gösteren Kalıntı Grafiği (Netter ve Wasserman, 1973).

3.2.2. Standart Kalıntılar

Yukarıda verdiğimiz (3.2.1) nolu denklem yeniden yazılır , parametre tahminleri için en küçük kareler yöntemi uygulanırsa,

$$(i) \quad Y = X \hat{\beta} + \epsilon \quad (3.2.6)$$

$$(ii) \quad e = (Y - X \hat{\beta}) \quad (3.2.7)$$

olur . Burada $\hat{\beta}$ 'nın değeri yerine yazılırsa,

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

$$= [I - X(X'X)^{-1} X']Y$$

$$e = MY$$

elde edilir ve

$$e = M\epsilon$$

olur. Burada da,

$$(iii) \quad M = I - X(X'X)^{-1} X' \quad (3.2.8)$$

(n×n)

ve M, {n-(k+1)} ranklı idempotent bir matristir.

$$(iv) \quad E(e'e) = \sigma^2 M \quad (3.2.9)$$

(v) Model için hata kareler toplamının tahmini değeri

$$S^2 = e'e = e'Me \quad (3.2.10)$$

elde edilir.

Böylece standart en küçük kareler kalıntı tahmini,

$$r_i = \frac{e_i}{S\sqrt{m_{ii}}} \quad (i=1,2,\dots,n) \quad (3.2.11)$$

formülü ile hesaplanır ve burada,

r_i = i' nci standart kalıntıyı,

e_i = i' nci kalıntıyı,

m_{ii} = M' nin i' nci köşegen elemanını,

M = σ^2 ile çarpılsın veya çarpılmasın varyans - kovaryans matrisini (Ellenberg, 1973),

S = Varyans analiz tablosundaki hata kareler ortalamasının karekökünü gösterir.

Draper ve Smith (1981) ve Kleinbaum ve Kupper (1978), standart kalıntıyı,

$$r_i = e_i / S \quad (3.2.12)$$

formülü ile hesaplamışlardır. Bu eşitlikte ,

r_i = i nci standart kalıntıyı

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

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

göstermektedir. Burada k, bağımsız değişken sayısıdır.

Thomas ve Ryan (1982), ise standart kalıntıyı,

$$r_i = e_i / Se_i \quad (3.2.13)$$

formülü ile hesaplamışlar ve bu formülü kullanarak hesaplamaları yapan bir de bilgisayar programı (Minitab Statistical Software Version II,1982) yazmışlardır. Bu formülde,

$$Se_i = \sqrt{\hat{H.K.O.} - \text{Var}(\hat{Y}_i)} \quad (3.2.14)$$

$$\text{Var}(\hat{Y}_i) = \hat{H.K.O.} * X_i(X'X)^{-1}X_i'$$

göstermektedir. Burada kullanılan H.K.O. , varyans analiz tablosundaki hata kareler ortalamasıdır. $\sqrt{\hat{S}m_{ii}}=Se_i$ olduğu için (3.2.11) ve (3.2.13) sayılı eşitlikler yardımıyla elde edilen standart kalıntı değerleri aynı sonucu vermektedirler ve bu çalışmada bu iki eşitlik yardımıyla elde edilen standart kalıntı değerleri kullanılacaktır. Çünkü diğer formülde kullanılan S değeri , tahmini değerdeki tüm varyasyonu dikkate aldığından daha küçük standart kalıntı değerleri vermektedir. Bu da anormal gözlemlerin belirlenmesini nisbeten daha zor hale getirmektedir. Se_i değeri ise aynı x değerinden hesaplanan tahmini değerlerin standart sapmasıdır. Hesaplanan bu değerler ilk olarak serpmne diyagramında daha sonra da normal olasılık kağıtları üzerinde normal grafikleri oluşturularak varsayımların kontrolü yapılacaktır. Bu grafikte X eksen standart kalıntı değerlerini , Y eksen ise yığılmalı nisbi frekansları göstermektedir.

Bu grafiğin çizimi için;

- a) Önce standart kalıntıları küçükten büyüğe doğru sıralanır,
- b) Sıralanan bu değerlerin yığılmalı nisbi frekansları bulunur,
- c) Bulunan nisbi frekanslar Y eksenine, standart kalıntı değerleri ise sıralı olarak X eksenine yazılır,
- d) Daha sonra her bir standart kalıntı değeri ile buna karşılık gelen yığılmalı nisbi frekans değerinin kesişim noktaları belirlenir.
- e) Standart kalıntıları bağımsız değişken (x), yığılmalı frekanslarını bağımlı değişken (y) olarak alıp bu noktaları temsil eden en-küçük kareler regresyon doğrusu bulunarak bunun temsil ettiği doğru aynı grafik üzerinde gösterilir.

Bu tip özel bir grafik ile ilgili olan önemli bir özellik, doğrunun elde edilmesinde yığılmalı nisbi frekansların kullanılmasıdır. Elde edilen bu doğru grafikte gösterilir ve daha sonra $N(0,1)$ için noktaların normallikten herhangi bir sapmanın olup olmadığının belirlenmesinde bir ölçü olarak kullanılır. Varsayımın mevcut olduğunu, bu grafikteki noktaların yaklaşık aynı doğru üzerinde olmasıyla anlarız.

Normallik varsayımının geçerliliğini belirlemede daha çok sayıya dayalı bir kriter olarak standart test istatistikleri de kullanılabilir (Kleinbaum ve Kupper, 1978). Burada (e_i/S) 'lerin standart normal dağılımdan alınan bir örnek olduğu varsayılır. Bu durumda örneğimizdeki standart kalıntının %95'nin -1.96 ile 1.96 arasında veya %99'nun -2.576 ile 2.576 arasında olması beklenir. Eğer $n-(k+1)$ çok küçükse bu durumda $t_{n-(k+1)}$ dağılımının %95 ve %99 limitleri kullanılır (Kleinbaum ve Kupper, 1978).

3.2.3. Kısmi Kalıntılar

Bu çalışmada çoklu regresyonda kullanılmak ve bir araştırmacıya aşağıdaki konularda yardımcı olabilecek kalıntı analiz yöntemi üzerinde durulacaktır. Kısmi kalıntıların grafik analizi olarak adlandırılan bu yöntem araştırmacıya;

1) i'nci bağımsız değişkenin diğer tüm bağımsız değişkenler mevcut olduğunda bağımlı değişken için tahmindeki önemini belirlenmesinde,

2) Doğrusal olmayan herhangi bir i'nci değişkenin doğrusallıktaki önemini tayin etmede ve bu değişkende hangi tür bir transformasyonun yapılabileceğinin belirlenmesinde yardımcı olur. Diğer bir deyişle kısmi kalıntı grafikleri, doğrusallıktan sapmaları ve miktarlarını gösterirler.

Bu grafikler, uydurulan en küçük kareler doğrusu hakkında çok daha net ve uygun bilgileri özet bilgi olarak verirler. Değişkenlerin doğru bir şekilde belirlenmesi veya değişkenlere hangi transformasyon uygulamamın gerekli olduğunun bilinmediği erken safhalarda modelin kullanılabilirliği belirlenebilmektedir.

Kalıntı değerleri (3.2.5) nolu eşitlik kullanılarak

$$\hat{e} = (Y - X \hat{B})$$

hesaplanır. Eğer X_i , X 'in i'nci sütunu ise o zaman i'nci kısmi kalıntı vektörü,

$$e_i^* = \hat{e} + X_i \hat{B}_i \quad (3.2.15)$$

olarak tanımlanır (Larsen ve McCleary, 1972). Bu eşitlikte $\hat{e}$ 'nin değeri yerine yazılırsa,

$$e_i^* = (Y - \hat{\alpha} - \hat{\beta}_1 X_1 - \hat{\beta}_2 X_2 - \dots - \hat{\beta}_k X_k) + X_i \hat{\beta}_1$$

örneğin ,

$$e_1^* = (Y - \hat{\alpha} - \hat{\beta}_1 X_1 - \hat{\beta}_2 X_2 - \dots - \hat{\beta}_k X_k) + X_1 \hat{\beta}_1$$

ise,

$$e_i^* = Y - \sum_{j=1}^k X_j \hat{\beta}_j \quad (3.2.16)$$

olur (Larsen ve McCleary, 1972). Burada $\hat{\beta}_1$, (3.2.3) nolu eşitlik yardımıyla tüm modelden elde edilen değerdir .

Şimdi X_i 'de e_i^* 'nin basit doğrusal regresyonunu düşünelim. Bu regresyon k adet değişkenin tümünü içeren çoklu doğrusal regresyon modeli için $\hat{\beta}_1$ 'nin bir tahmini değeri olarak

elde edilen en küçük kareler doğrusunun eğimi $\hat{\beta}_1$ değeri ile aynı olacaktır. e_i^* 'in X_i 'deki regresyonunda en küçük kareler işlemi,

$$\| e_i^* - \hat{\beta}_1 X_i \|^2 \quad (3.2.17)$$

ifadesinin minimize eden $\hat{\beta}_1$ 'yi seçer. (3.2.17) nolu eşitlik herhangi bir Z vektörü için,

$$\| Z \|^2 = \sum Z_i^2 \quad (3.2.18)$$

şeklinde hesaplanır. Böylece kısmi kalıntının tanımından (3.2.17) nolu eşitlik,

$$\| Y - \sum_{i \neq j} \hat{\beta}_i X_i - \tilde{\beta}_j X_j \| \quad (3.2.19)$$

olarak yazılabilir. Burada, $\hat{\beta}_i$, tüm model uydurulduğunda

$\hat{\beta}_i$ 'nin en küçük kareler tahmincisidir. $\tilde{\beta}_j$ (3.2.19) nolu eşitliği en iyi biçimde minimize eder. Bu nedenle tüm

$\tilde{\beta}_j = \hat{\beta}_j$ için,

$$\| Y - \sum_{i \neq j} \hat{\beta}_i X_i - \tilde{\beta}_j X_j \| < \| Y - \sum_{i \neq j} \hat{\beta}_i X_i - \hat{\beta}_j X_j \| \quad (3.2.20)$$

eşitsizliği daima sağlar. Burada $\tilde{\beta}_j$, $\hat{\beta}_j$ 'ya eşit olmayan herhangi bir katsayıdır. Fakat (3.2.20) nolu eşitsizlik

$\tilde{\beta}_j = \hat{\beta}_j$ da doğru olmaz. O zaman tüm model için en küçük kareler tahmincileri kalıntı kareler toplamını minimize

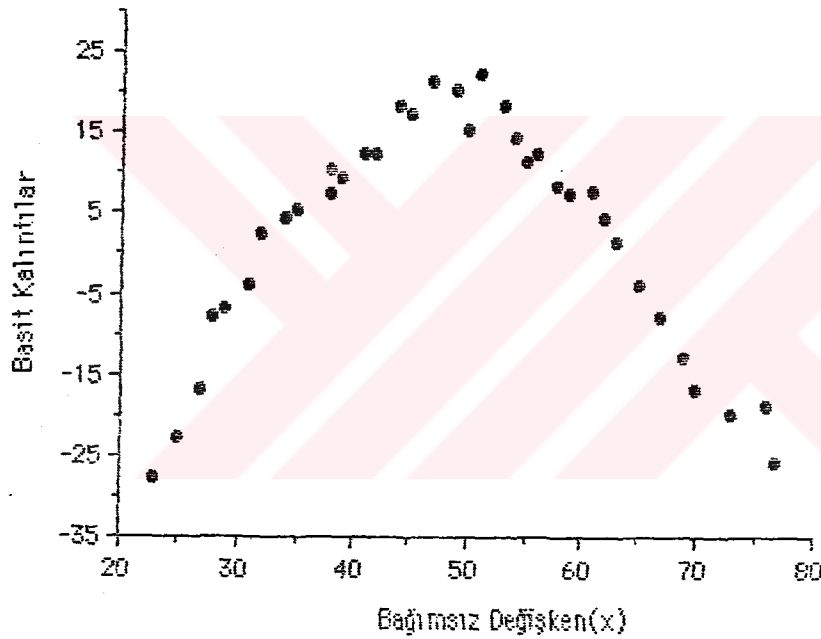
etmez. Tersı durumunda ise bu $\tilde{\beta}_j = \hat{\beta}_j$ olmasını sağlar.

Bağımsız değişkenlere karşı çizilen kısmi kalıntı grafikleri incelendiğinde değişkenler arasındaki önemli ilişkileri ve eğimin yapısını ortaya çıkarırlar. Şimdi bunları sırasıyla inceleyelim;

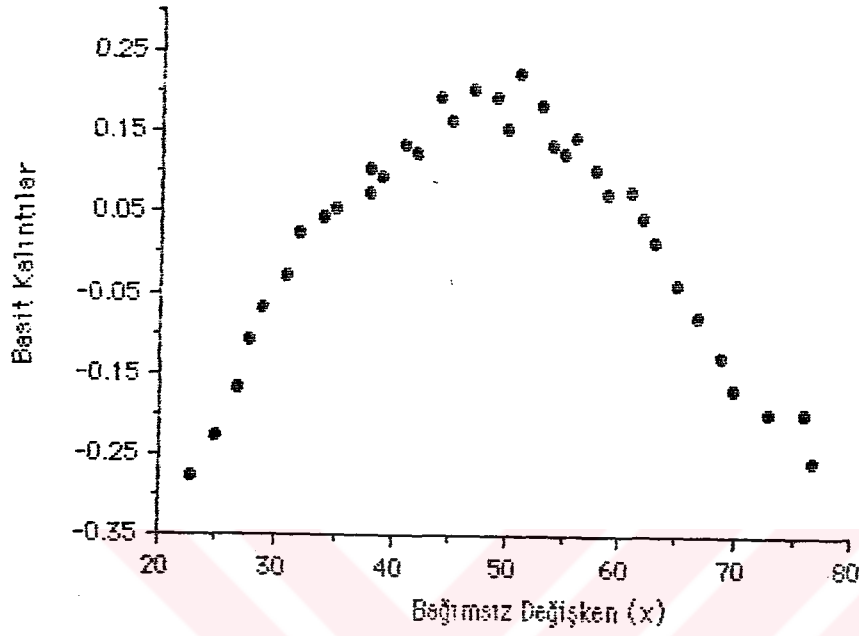
3.2.3.1. Kısmi Kalıntı Grafikleri

Kısmi kalıntı grafikleri özellikle bağımlı ile bağımsız değişkenler arasındaki ilişki doğrusal olmadığı zaman çok faydalı bilgiler verebilmektedirler. Bu amaçla Larsen ve

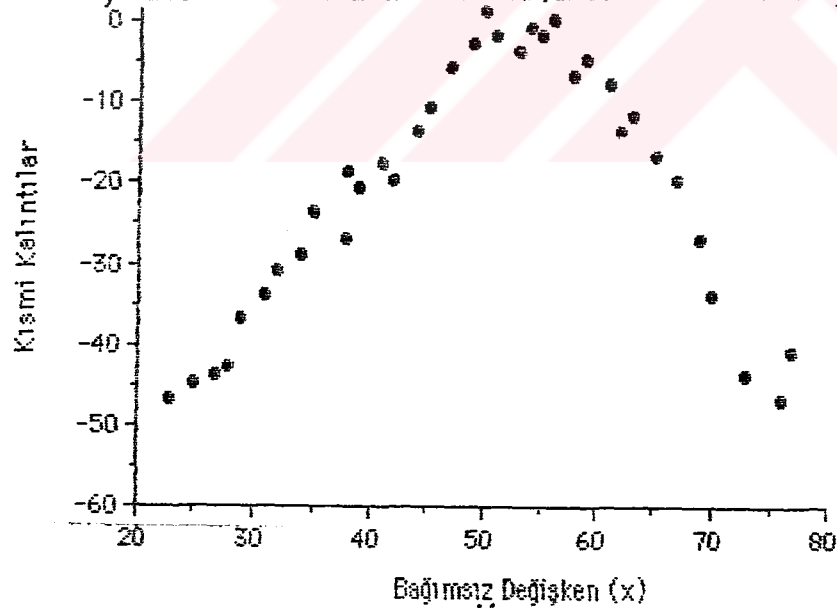
McCleary,1972'den alınan iki farklı regresyon örneği incelenmiştir. Şekil 3.2.7 ve 3.2.8 sırasıyla birinci ve ikinci örneğe ait basit kalıntı grafikleri , Şekil 3.2.9 ve 3.2.10 ise birinci ve ikinci örneklerin kısmi kalıntı grafikleridir. Burada ilk göze çarpan durum, iki kısmi kalıntı grafiği Şekil 3.2.9 ve 3.2.10 birbirinden oldukça farklı bir görünüm arz ederken, basit kalıntı grafikleri Şekil 3.2.7 ve 3.2.8 hemen hemen benzerdirler.



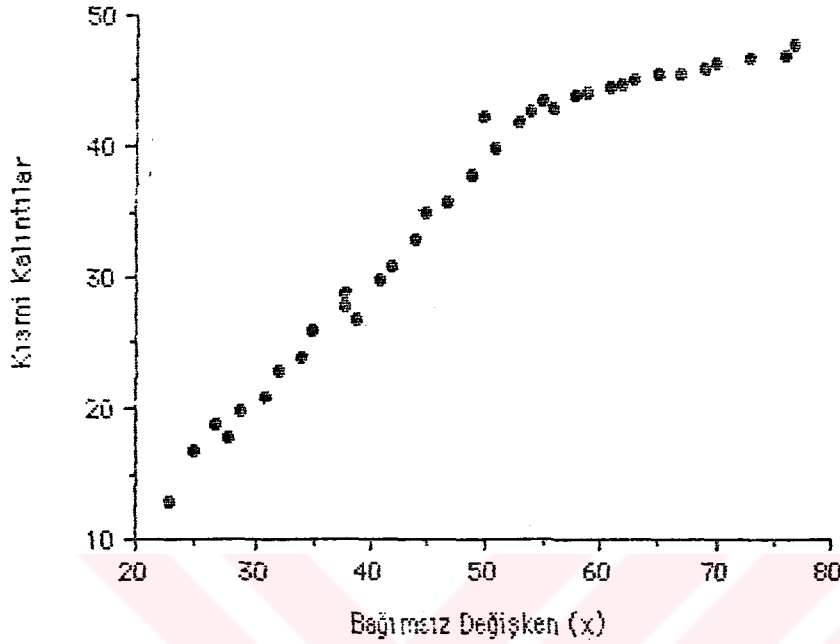
Şekil 3.2.7. Birinci Regresyon Örneğinde Bağımsız Değişken (X)'e Karşı Çizilen Basit Kalıntı Grafiği(Larsen ve McCleary, 1972).



Şekil 3.2.8. İkinci Örnekte Bağımsız Değişken (X)'e Karşı Çizilen Basit Kalıntı Grafiği (Larsen ve McCleary, 1972).



Şekil 3.2.9. Birinci Regresyon Örneğinde Bağımsız Değişken (X)'e Karşı Çizilen Kısmi Kalıntı Grafiği (Larsen ve McCleary, 1972).



Şekil 3.2.10. İkinci Örnekte Bağımsız Değişken (X)'e Karşı Çizilen Kısmi Kalıntı Grafiği(Larsen ve McCleary, 1972).

Şekil 3.2.7 ve 3.2.8 her ikisi de modele quadratik bir terimin eklenmesi gerektiğini ifade etmektedirler. Kısmi kalıntı grafikleri Şekil 3.2.9 ve 3.2.10 ise, quadratik terim eklenmesinin birinci örnek için ikincisinden daha kritik bir durum arz ettiğini ifade etmektedir. Diğer bir deyişle birinci örneğe quadratik terim eklenmesi ikinci örneğe göre daha önemlidir. Bundan başka Şekil 3.2.10'un incelenmesi sonucunda modele quadratik terim eklenmesinden ziyade Y'de logaritmik bir transformasyon yapılması gerektiğini görülmektedir. Bu durum doğrusal olmayan bir ilişkiyi doğrusallaştırmayı gösterir, aynı zamanda doğrusal olma ve olmama durumunun açıkça görülmesi istenen bir durumdur. İşte kısmi kalıntı grafikleri yukarıda verilen sonuçların çıkarılmasında yardımcı olur.

3.2.3.2. Kısmi Kalıntı Grafiklerinde Tahmin Edilen Eğimin Yapısı

Bu bölümde sadece kısmi kalıntı grafikleri yardımıyla

(2.2.8) nolu eşitlikteki β_i 'nin yapısı araştırılmıştır.

Doğrusal regresyonun genel varsayımları altında,

$$e_i^* = \alpha + \beta_i x_i + e \quad (3.2.21)$$

burada, e 'nin elemanları bağımsız dağılış gösterirler, β_i 'nin tahmini varyansı,

$$\|e\|^2 / (n-2) \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2 \quad (3.2.22)$$

olur. Bununla beraber tüm model uydurulduğunda β_i 'nin gerçek minimum varyanslı sapmasız tahmini,

$$\|e\|^2 / (n-2) \sum_{j=1}^n (X_{ij} - \bar{X}_i)^2 (1-R_i^2) \quad (3.2.23)$$

olur. Burada R_i^2 , X_i ile diğer bağımsız değişkenler arasındaki çoklu korelasyon katsayısıdır. Bundan dolayı araştırıcı kısmi grafikleri incelemekle β_i 'nin varyans tahmininin beklenenden yüksek olup olmadığını belirleyebilecektir. Tüm modelde β_i 'nin varyansının yüksek tahmin edilmesi R_i^2 büyük değilse önemli değildir. Yüksek olduğunda kısmi grafikler, i 'nci bağımsız değişkenin diğer bağımsız değişkenler mevcut olduğundaki öneminin incelenbilmesinde yetersiz kalır. Diğer özellikler üzerinde etkisi yoktur (Larsen ve McCleary, 1972). Bu

ölçeklendirilir fakat bu durumda ise sonuç grafiği mevcut olan herhangi bir doğrusal olmama durumunu fazla abartacaktır. Diğer bir deyişle , kısmi kalıntıları yeniden ölçeklendirmek yukarıda verilen özelliği koruyabilir fakat doğrusal olmamanın önemini saptama, doğru x_i 'nin ve doğru transformasyonun kesin bir seçimini yapmada yetersiz bilgi verir. Bu son özellik diğer iki özellikten daha faydalı olduğu için daha önemlidir. Bu nedenle de kalıntıların ölçeklendirilmemesi önerilir.

Parametre tahminleri için çoklu regresyon analizi yapılırken MINITAB STATISTICAL SOFTWARE VERSION II, 1981, paket programı kullanılmıştır.

Araştırmada kullanılan grafikler ise Macintosh Classic bilgisayarda CRICKET GRAPH VERSION 1.2, 1987, paket programı ile çizdirilmiştir.

4. ARAŞTIRMA BULGULARI

Materyalde (Çizelge 3.1) verilen değerler kullanılarak en küçük kareler regresyon doğrusu tahmin edilirse,

$$\hat{y}_i = 226 - 1.22 x_1 - 0.034 x_2 \quad (4.1)$$

olur ve

$$y_i = 226 - 1.22 x_1 - 0.034 x_2 + e_i$$

eşitliği elde edilir. Burada x_1 ve x_2 sırasıyla dakikadaki solunum sayısı ve nabız sayısıdır. Bu eşitlik kullanılarak hesaplanan tahmini değerler ve (3.2.5) nolu eşitlik kullanılarak hesaplanan kalıntı değerleri, Nabız sayısı (dk), solunum sayısı (dk), süt verimleri(kg/koyun) Çizelge 4.2'de verilmiştir.

Örneğe ait regresyon katsayısının önem kontrolü için yapılan varyans analiz sonuçları ise Çizelge 4.1'de verilmiştir.

Çizelge 4.1. Koyunların Süt Verimleri ile Solunum Sayısı ve Nabız Sayısı Arasındaki Regresyona Ait Varyans Analiz Tablosu.

V.K.	SD	KT	KO	F	P
Regresyon	2	16539.2	8269.6	20.63**	0.000
Hata	33	13227.8	400.8		
Genel	35	29747.0			

Buna göre, koyunların süt verimleri ile solunum sayısı veya nabız sayısı arasındaki azalış istatistik bakımdan önemlidir ($p < 0.01$). Fakat örneğe ait $R^2 = \% 55.69$ olup yüksek bir değer sayılmaz. Yani koyunların dakikadaki solunum sayıları ve nabız sayıları laktasyon süt verimlerindeki varyasyonun ancak

% 55.69'ünü açıklayabilmektedir. Regresyon katsayıları önemli olduğu halde R^2 'nin yüksek çıkmaması, gözlemler içerisinde anormal gözlemlerin olabileceğini gösterir.

Çizelge 4.2. Solunum sayısı(dk), nabız sayısı (dk), süt verimleri (kg/koyun), (4.1) nolu eşitlik kullanılarak hesaplanan tahmini değerler ve (3.2.5) nolu eşitlik kullanılarak hesaplanan kalıntı değerleri

SolS(dk)	NaS(dk)	SütV(kg/koy.)	$\hat{Y}$	e_i
62	100	169	146.70	22.30
61	96	181	148.06	32.94
44	80	187	169.41	17.59
43	76	187	170.78	16.22
65	88	178	143.44	34.56
37	92	133	177.57	-44.57
78	84	107	127.66	-20.66
82	90	111	122.56	-11.56
86	86	124	117.80	6.20
82	92	100	122.49	-22.49
56	106	153	154.84	-0.84
86	104	110	117.19	-7.19
70	100	105	136.91	-31.91
61	102	178	147.86	30.14
47	75	175	165.91	9.09
41	90	181	172.75	8.25
61	93	160	148.16	11.84
29	100	175	187.09	-12.09
71	80	152	136.37	15.63
93	80	105	109.44	-4.44
79	35	112	128.11	-16.11
78	96	102	127.25	-25.25
46	102	155	166.22	-11.22
81	114	121	122.97	-1.97
74	75	124	132.87	-8.87
65	81	107	143.68	-36.68
54	90	178	156.83	21.17
46	63	180	167.55	12.45
57	70	150	153.84	-3.84
36	75	160	179.38	-19.38
71	75	145	136.54	8.46
73	66	140	134.40	5.60
99	75	121	102.27	18.73
80	90	132	125.01	6.99
84	75	130	120.63	9.37
49	75	155	163.47	-8.47

Eğer gözlemler içerisinde anormal gözlemler varsa bunu ortaya çıkarmamanın bir yolu kalıntı analizi yapmaktır(Bölüm 4.2).

Yapılan değişken eleme (Backward Elimination) testi sonucu koyunlarda dakikadaki nabız sayısının süt verimine önemli bir etkisi olmadığı yani süt verimlerindeki varyasyonun önemsiz bir kısmını oluşturduğu görülmüştür. Diğer bir deyişle sadece dakikadaki solunum sayısı Y'deki açıklanabilen varyasyonun büyük bir kısmını açıklayabilmektedir. Bu nedenle bu çalışmada dakikadaki solunum sayısı (x) ile laktasyon süt verimleri (y) arasındaki ilişkiyi gösteren regresyon modeli üzerinde çalışılmıştır. Uydurulan bu yeni model ise,

$$\hat{y}_i = 223 - 1.22 x_1 \quad (4.2)$$

dir ve yapılan varyans analizi sonucuna göre yine solunum sayısı ile laktasyon süt verimleri arasındaki regresyon katsayısı önemli bulunmuştur ($p < 0.01$)(Çizelge 4.3). $R^2 = \% 55.5$. Bu eşitlik kullanılarak hesaplanan tahmini değerler ve (3.2.5) nolu eşitlik kullanılarak hesaplanan kalıntı değerleri, solunum sayısı (dk), süt verimleri(kg/koy.) Çizelge 4.4'de verilmiştir ve Bölüm 4.1 ve 4.2'de (4.2) nolu eşitlik kullanılarak hesaplanan değerler kullanılmıştır.

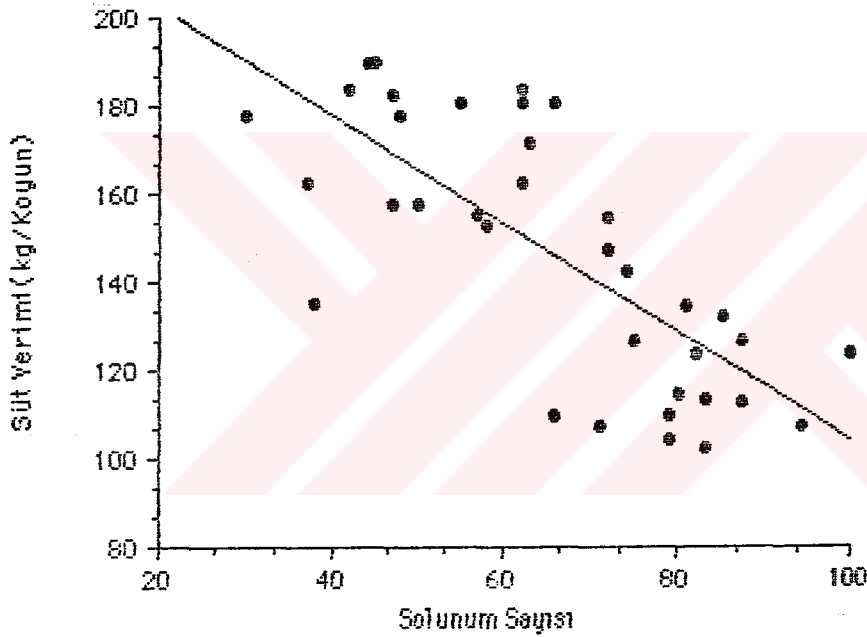
Çizelge 4.3. Koyunların Süt Verimleri ile Solunum Sayısı Arasındaki Regresyona Ait Varyans Analiz Tablosu.

V.K.	SD	KT	KO	F	P
Regresyon	1	16530	16530	42.46**	0.000
Hata	34	13237	389		
Genel	35	29747			

Çizelge 4.4. Solunum Sayısı(dk), Süt Verimleri(kg/koy.)(4.2)
Nolu Eşitlik Kullanılarak Hesaplanan Tahmini
Değerler ve (3.2.5) Nolu Eşitlik Kullanılarak hesap-
lanan Kalıntı Değerleri,

SolS(dk)	SütV(kg/lak.)	$\hat{Y}$	e_i
62	169	147.20	21.80
61	181	148.42	32.58
44	187	169.20	17.80
43	187	170.43	16.57
65	178	143.53	34.47
37	133	177.76	-44.76
78	107	127.64	-20.64
82	111	122.75	-11.75
86	124	117.86	6.14
82	100	122.75	-22.75
56	153	154.53	- 1.53
86	110	117.86	- 7.86
70	105	137.42	-32.42
61	178	148.42	29.58
47	175	165.54	9.46
41	181	172.87	8.13
61	160	148.42	11.58
29	175	187.54	-12.54
71	152	136.20	15.80
93	105	109.30	- 4.30
79	112	126.42	-14.42
78	102	127.64	-25.64
46	155	166.76	-11.76
81	121	123.97	- 2.97
74	124	132.53	- 8.53
65	107	143.53	-36.53
54	178	156.98	21.02
46	180	166.76	13.24
57	150	153.31	- 3.31
36	160	178.98	-18.98
71	145	136.20	8.80
73	140	133.75	6.25
99	121	101.97	19.03
80	132	125.19	6.81
84	130	120.30	9.70
49	155	163.09	- 8.09

Doğrusal regresyon fonksiyonunun verilere uygun olup olmadığının analizi için çoğu zaman uydurulan regresyon doğrusu grafiği ile verilerin serpmeye diyagramı üzerinde çalışılır(Netter ve Wasserman,1973). Şekil 4.1 koyunlarda dakikadaki solunum sayısı ve laktasyon süt verimleri arasındaki ilişkiyi gösteren uydurulan regresyon doğrusu ($y=223-1.22x_1$) ve verilerin serpmeye grafiğini göstermektedir.



Şekil 4.1. Verilere Ait Serpmeye ve Uydurulan Regresyon Doğrusu Grafiği

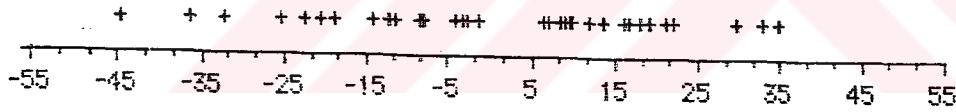
4.1. Basit (Ordinary) Kalıntılar

Şimdi yöntem kısmında verildiği şekilde basit kalıntıların grafiklerini teker teker çizerek inceleyelim.

4.1.1. Kalıntıların Tümünün Grafiği

Bu grafik, kalıntı değerlerinin doğal sayılar doğrusu üzerinde serpmeye diyagramı şeklinde işaretlenmesiyle elde edilir. Örnekte kalıntıların tümünün bir grafiği çizildiğinde Şekil 4.1.1'deki diyagramı elde ederiz.

Biz genelde $\epsilon_i \sim N(0, \sigma^2)$ olduğunu varsayalım. Eğer kalıntılar normal dağılım gösteriyorsa ve modelimiz doğru ise diyagramda sıralanmış kalıntılar, ortalaması sıfır varyansı σ^2 olan bir normal dağılımdan rastgele alınmış 36 gözlemden oluşan bir örnek izlenimi vermelidirler (Draper ve Smith, 1981).



Şekil 4.1.1. Kalıntıların Tümünün Grafiği.

Burada kalıntıların ortalaması sıfırdır, çünkü herhangi bir regresyon modelinin α sabit terimi içermesi, kalıntıların ortalamasının sıfır olmasını gerektirir. Bu, hata kareler toplamının α 'ya göre kısmi türevi alınarak elde edilen birinci derecedeki normal eşitlikte kolayca görülür.

Eğer uydurulan model,

$$E(Y) = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$$

ise, normal eşitlik,

$$-2\sum(Y_i - \hat{\alpha} - \hat{\beta}_1 X_1 - \hat{\beta}_2 X_2 - \dots - \hat{\beta}_k X_k) = 0$$

şeklinde yazılabilir. Burada $i=1,2,\dots,n$ üzerinden toplam alınır ve gerekli işlemler yapılırsa,

$$\sum(Y_i - \hat{Y}_i) = 0$$

olur. Böylece,

$$\sum e_i = \sum e_i / n = 0$$

olur. Şekil 4.1.1'deki grafik az da olsa düzensiz bir görünüm ortaya koymasına karşın örnekteki gözlemlerin gerçekten normal dağılım gösterip göstermedikleri hakkında yeterli bilgi vermez. Bu grafikler hakkında karar vermeye yardımcı olacak diğer bir yol ileride açıklanacak olan standart olasılık kağıtları üzerinde kalıntıların normal veya yarı normal grafiklerini oluşturmaktır. Bu grafiklerde noktalar aşağı yukarı düz bir doğru üzerine düşmelidir (Draper ve Smith, 1981). Bazı istatistikçilerin tercih etmesine karşılık bu yolun kullanılması ile herhangi bir avantaj sağlanamayacağını bildirmişlerdir. Kalıntı sayısı çok fazla ise nokta diyagramdan ziyade histogram çizilebilir.

4.1.2. Kalıntıların Zaman Serisi Grafiği

Bu grafikte , X eksenı sıralı zaman aralıklarını, Y eksenı kalıntı deęerlerini gösterir.

Örneęimize ait zaman serisi grafięi çizilirse Şekil 4.1.2'deki diyagram elde edilir. Eęer biz bu diyagramdan bir adım geriye gidersek kalıntıların Şekil 3.2.1'de verilen düz bir bandının elde edildięini görürüz. Bu bize , burada uzun dönem (-40 ve 40) eğiliminin verilere etkili olmadıęını gösterir (Draper ve Smith ,1981).

Yukarıda söylendięi gibi kalıntılar uzun dönem eğilimini ortaya çıkarırlar. Zaman grafięinin iyice incelenmesi yorumun kısa dönem eğilimine yayılamayacaęını gösterir. Draper ve Smith (1981), bu grafiklerde belli bir yöne doğru eğilimin olduęu durumlarda bunun zamana baęlı olarak düzeltilmiş modele eklenmesi gereken bazı terimlerin olduęunu ifade ettięini belirtmişlerdir. Örneęimize ait grafięin (Şekil 4.1.2) belli bölgelerinde aşıęı doğru bir eğim olduęundan deneysel olarak doğrusal bir görünüm için modele,

$$\delta\{(t-1) \bmod 12\}$$

şeklinde bir deęişken eklenebilir. Burada,

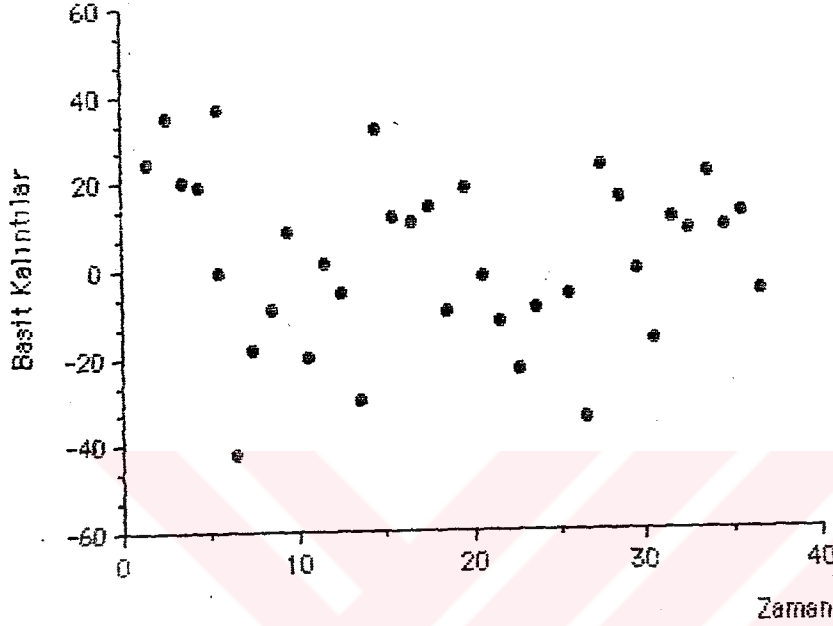
δ =tahmin edilen regresyon katsayısını,

$(t-1) \bmod 12 = (t-1)$ 'i 12 ile böldükten sonra kalan deęişkeni gösterir. Alınan bu yeni deęişkenin deęerleri,

Eski deęişken (t) 1 2 3 4 5 6 7 8 9 10 11 12 ... 35 36

$(t-1) \bmod 12$ 0 1 2 3 4 5 6 7 8 9 10 11 ... 10 11

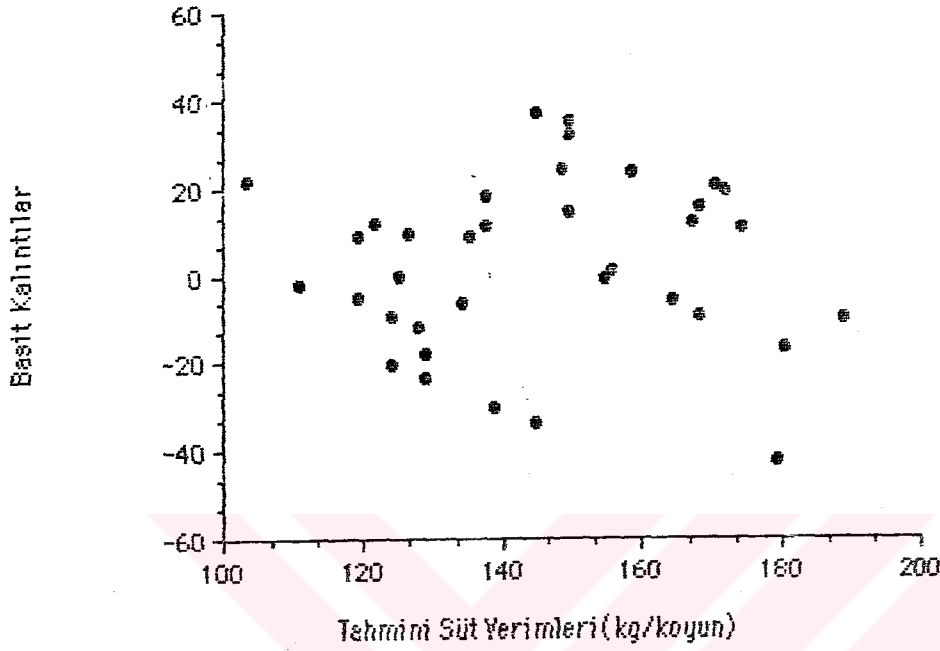
olur. Dięer bir alternatif yol ise $\{(t-1) \bmod 12-1\}$ gibi bir deęişken kullanmaktır. bu ise 0,1,2,...,11 yerine -1,0,1,...,10 gibi bazen daha uygun olan deęerler elde edilmesini sağlar.



Şekil 4.1.2. Kalıntıların Zaman Serisi Grafiği.

4.1.3. y_i 'e Karşı Kalıntı Grafiği

Bu grafikte X eksenini y_i değerlerini, Y eksenini ise kalıntıları gösterir. örneğimize ait grafik çizildiğinde varyansın belli bir bölgeye kadar y_i değerleri arttıkça arttığı daha sonra ise azaldığı izlenimi vermektedir(Şekil 4.3.1) bu durumda burada en küçük kareler analizi uygulanmadan önce y_i gözlemlerinde bir transformasyon yapılması önerilir.



Şekil 4.1.3. Tahmini y_i Değerlerine Karşı Kalıntı Grafiği.

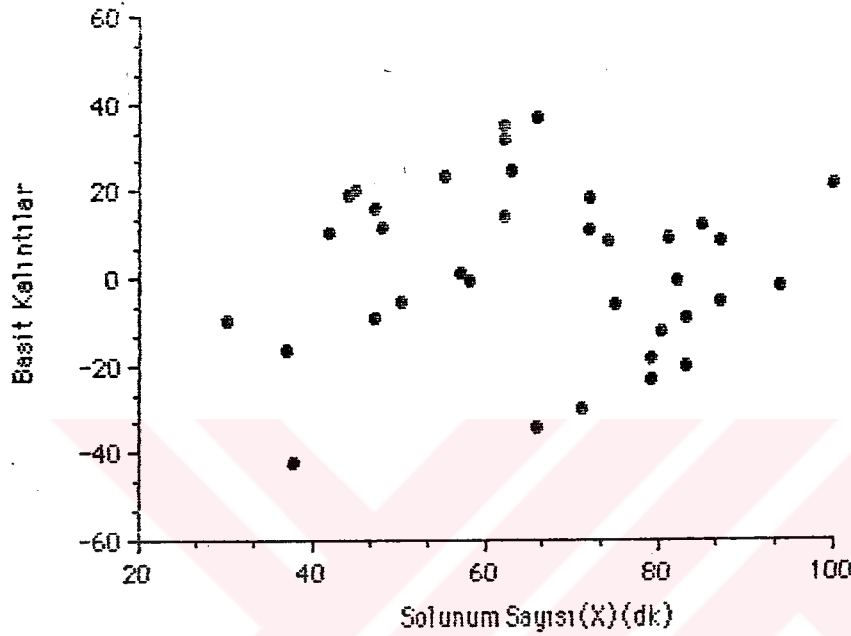
4.1.4. x_{ji} 'lere Karşı Kalıntı Grafiği

Bu grafiklerin şekli y_i 'lara karşı çizilen grafiklerle

aynıdır , fakat burada y_i değerleri yerine $x_{j1}, x_{j2}, \dots, x_{jn}$, gibi x_{ji} 'lere bağlı değerler kullanılır (Şekil 4.1.6.). Bu grafikte de

y_i 'ya karşı çizilen kalıntı grafiğinde olduğu gibi varyansın x değerleri arttıkça önce arttığı daha sonra azaldığı görülmektedir. Bu durum bize tartılı en küçük kareler regresyon analizinin uygulanması gerektiğini göstermektedir. x_{ji} 'lere karşı çizilen grafiklerde sadece 2 veya 3 x değişkeni içeren regresyon problemlerinde 2 veya 3 boyutlu bir diyagram çizmek mümkündür. Fakat 3'den fazla x değişkeni bulunan

modellerde ait gruplar oluşturarak diyagram çizmek daha uygundur. Diğerleri 2 boyutlu grafiklerle ifade edilebildiği için buna gerek kalmaz.



Şekil 4.1.4. Bağımsız x Değişkenine Karşı Kalıntı Grafiği.

4.2. Standart (Standardized, Studentized) Kalıntılar

Örneğimize ait en küçük kareler regresyon doğrusu tahmin edildiğinde ,

$$\hat{y}_1 = 223 - 1.22 x_1 \quad (4.2.1)$$

olur ve

$$y_1 = 223 - 1.22 x_1 + e_i$$

eşitliği elde edilir.

Yöntem kısmında verilen (3.2.8) nolu eşitlikteki M matrisinin köşegen elemanları (m_{ii}) ve (3.2.14) nolu eşitlikle hesaplanan Se_i değerleri Çizelge 4.2.1'de verilmiştir. Hesapladığımız bu değerlerin ilk önce serpmе grafiğini daha sonra da normal olasılık kağıtları üzerinde normal veya yarı normal grafiklerini çizerek inceleyeceğiz.

(4.2.1) nolu eşitlik kullanılarak her yaş grubu ortalamasına ait tahmini süt verimleri, (3.2.5) nolu formül yardımıyla kalıntılar ve (3.2.12) nolu (burada $S = 19.7$) ve (3.2.13) nolu formül yardımıyla hesaplanan standart kalıntı, sıralı standart kalıntı değerleri ve yığılmalı nisbi frekanslar Çizelge 4.2.2'de verilmiştir.

Çizelge 4.2.1. Bağımsız x_i M Matrisinin Köşegen Elemanları (m_{ii}) ve Standart Sapma Değerleri(Se_i).

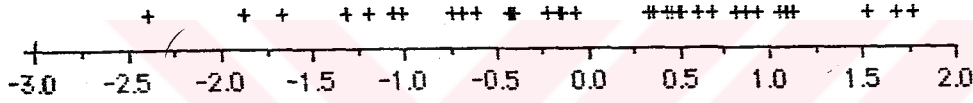
x_i	m_{ii}	Se_i	x_i	m_{ii}	Se_i
62	0.971593	19.44	71	0.968564	19.41
61	0.971025	19.43	93	0.899498	18.70
44	0.933709	19.05	79	0.95357	19.26
43	0.929887	19.02	78	0.956082	19.28
65	0.972210	19.45	46	0.940812	19.13
37	0.903155	18.74	81	0.948020	19.20
78	0.956082	19.28	74	0.964299	19.37
82	0.944971	19.17	65	0.972210	19.45
86	0.930967	19.03	54	0.961989	19.34
82	0.944971	19.17	46	0.940812	19.13
56	0.965475	19.37	57	0.966946	19.39
86	0.930967	19.03	36	0.898066	18.69
70	0.969624	19.42	71	0.968564	19.41
61	0.971025	19.43	73	0.965902	19.38
47	0.944092	19.16	99	0.865472	18.35
41	0.921699	18.93	80	0.950888	19.20
61	0.971025	19.43	84	0.938331	19.10
29	0.857385	18.26	49	0.950109	19.22

Çizelge 4.2.2. Koyunlara Ait Solunum Sayısı(dk), Süt Verimleri(kg/koyun), $\bar{Y}$, e_i , $r_i=e_i/S$, $r_i=e_i/Se_i$, Sıralı Standart Kalıntılar ($r_i()$) ve Yığılmalı Nisbi Frekansları.

x_i	Y	$\hat{Y}$	e_i	$r_i = \frac{e_i}{S}$	$r_i()$	$r_i = \frac{e_i}{Se_i}$	$r_i()$	Yığ. Nis. Frek.
62	169	147.20	21.8	1.11	-2.27	1.12	-2.39	0.027
61	181	148.42	32.6	1.65	-1.85	1.68	-1.88	0.055
44	187	169.20	17.8	0.90	-1.65	0.93	-1.67	0.083
43	187	170.49	16.6	0.84	-1.30	0.87	-1.33	0.110
65	178	143.59	34.5	1.75	-1.15	1.77	-1.19	0.138
37	133	177.76	-44.8	-2.27	-1.04	-2.39	-1.07	0.166
78	107	127.64	-20.6	-1.04	-0.96	-1.07	-1.02	0.194
82	111	122.75	-11.8	-0.60	-0.73	-0.61	-0.75	0.222
86	124	117.86	6.1	0.31	-0.64	0.32	-0.69	0.250
82	100	122.75	-22.8	-1.15	-0.60	-1.19	-0.61	0.277
56	153	154.53	-1.5	-0.08	-0.60	-0.08	-0.61	0.305
86	110	117.86	-7.9	-0.40	-0.43	-0.41	-0.44	0.333
70	105	137.42	-31.4	-1.65	-0.41	-1.67	-0.42	0.361
61	178	148.42	29.6	1.50	-0.40	1.52	-0.41	0.388
47	175	165.54	9.7	0.48	-0.22	0.49	-0.23	0.416
41	181	172.87	8.1	0.41	-0.17	0.43	-0.17	0.444
61	160	148.42	11.6	0.59	-0.15	0.60	-0.15	0.472
29	175	187.54	-12.5	-0.64	-0.08	-0.69	-0.08	0.500
71	152	136.20	15.8	0.80	0.31	0.81	0.32	0.527
93	105	109.30	-4.3	-0.22	0.32	-0.23	0.32	0.555
79	112	126.42	-14.4	-0.73	0.35	-0.75	0.35	0.583
78	102	127.64	-25.6	-1.30	0.41	-1.33	0.43	0.611
46	155	166.76	-11.8	-0.60	0.45	-0.61	0.45	0.638
81	121	123.97	-3.0	-0.15	0.48	-0.15	0.49	0.666
74	124	132.59	-8.5	-0.43	0.49	-0.44	0.51	0.694
65	107	143.59	-36.5	-1.85	0.59	-1.88	0.60	0.722
54	178	156.98	21.0	1.07	0.67	1.09	0.69	0.750
46	180	166.76	13.2	0.67	0.80	0.69	0.81	0.777
57	150	153.31	-3.3	-0.17	0.84	-0.17	0.87	0.805
36	160	178.98	-19.0	-0.96	0.90	-1.02	0.93	0.833
71	145	136.20	8.8	0.45	0.97	0.45	1.04	0.861
73	146	133.75	6.3	0.32	1.07	0.32	1.09	0.888
99	121	101.97	19.0	0.97	1.11	1.04	1.12	0.916
80	132	125.19	6.9	0.35	1.50	0.35	1.52	0.944
84	130	120.30	9.7	0.49	1.65	0.51	1.68	0.972
49	155	169.09	-8.1	-0.41	1.75	-0.42	1.77	1.000

4.2.1. Standart Kalıntıların Serpme Grafiği

Bu grafiğin şekli e_j 'lerin kullanımı ile elde edilen grafik şekli ile aynı olabilir. Fakat burada sadece ölçekte bir değişiklik söz konusudur.



Şekil 4.2.1. e_j / S_{e_j} 'nin Serpme Grafiği.

Bu şekildeki bir grafik bize ne ifade eder? Biz genelde $e_j \sim N(0, \sigma^2)$ olduğunu varsayalım, öyleyse $e_j / \sigma \sim N(0, 1)$ olur. Eğer standart kalıntılar normal dağılım gösteriyorlarsa, e_j / S 'nin serpme grafiğinin ortalaması sıfır, varyansı 1 olan bir dağılımdan şansa bağlı olarak alınan 36 gözlem örneğine benzemesini bekleyebiliriz. Bu şekildeki bir grafiği standart normal sapma tablolarından örnek genişliği 36 olacak şekilde şans örnekleri seçilerek bunların serpme grafiği çizilir ve örneğimizin serpme grafiğiyle karşılaştırılır. Böylece $N(0, 1)$ varsayımı altında örnek noktalarındaki varyasyon için daha önceden bir fikir edinilebilir (Kleinbaum ve Kupper, 1978).

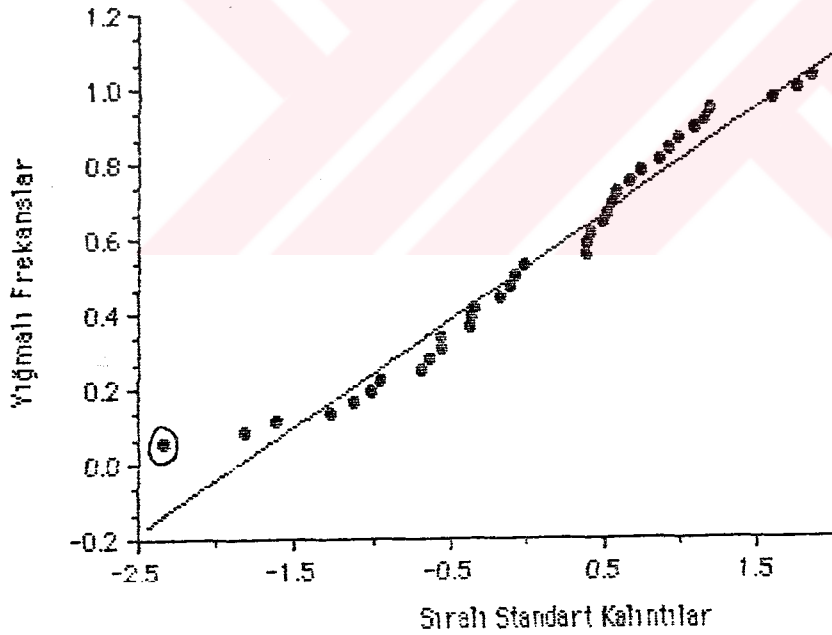
4.2.2. Standart Kalıntıların Normal Olasılık Kağıtları Üzerinde Normal Grafiği

Çizelge 4.2.2'de verilen sekizinci sütunu bağımsız (x), dokuzuncu sütunu bağımlı (y) değişken olarak alıp en-küçük kareler doğrusu bulunursa,

$$y = 0.512 + 0.285 r_i$$

eşitliği elde edilir.

Tüm bu değerler göz önüne alınarak normal grafiği çizilirse Şekil 4.2.2'de verilen durum elde edilir.



Şekil 4.2.2. $y=223 - 1.22x_i$ Modeli İçin Standardize Kalıntıların Normal Grafiği.

Şekil 4.2.2 incelendiğinde tüm noktalar yaklaşık aynı doğru üzerindeyken daire içerisine alınmış gözlem değeri (-2.39) doğrudan diğerlerine göre oldukça uzak bir bölgededir. Bu durumda bu gözlem değerinin normalliği bozduğu düşünülebilir. Bu gözlem aynı zamanda yapılan %95'lik standart test istatistiği limitleri (-1.96,1.96) dışındadır ve anormal gözlem olabilir. Öyleyse bu gözlem değeri normalliği bozduğu için atılarak tekrar yeni bir doğru tahmin edip bu yeni modelin kullanılması daha doğru olacaktır.

Standart kalıntıların grafiksel analizlerinde diğer bir yol ise sıradan kalıntılarda olduğu gibi bağımsız x değişkenine karşı grafiklerini oluşturmaktır. Çoklu regresyonda bu grafikler modeldeki her bir x değişkeni için çizilir. Bu grafikler çizildiğinde mümkün olabilen örnek tipleri yöntem kısmında verilen x_{ji} değişkenlerine karşı basit kalıntıların grafiklerine benzer sonuçlar verirler.

4.3. Kısmi (Partial) Kalıntılar

Materyal kısmında verilen koyunlara ait solunum sayısı (x_1), nabız sayısı (x_2) bağımsız değişkenler ve laktasyon süt verimleri bağımlı değişken (y) olarak alınmıştır. Bu verilere en küçük kareler yöntemi uygulanarak ,

$$\hat{y} = 226 - 1.22x_1 - 0.034x_2 \quad (4.3.1)$$

eşitliği elde edilmiştir. Daha sonra (3.2.15) veya (3.2.16) nolu eşitlikler yardımıyla ,

$$\begin{aligned} e_1^* &= e + X_1\beta_1 \\ &= e + X_1 * (-1.22) \end{aligned} \quad (4.3.2)$$

veya

$$\begin{aligned} e_1^* &= Y - (\alpha + X_2\beta_2) \\ &= Y - (226 + X_2 (-0.034)) \end{aligned} \quad (4.3.3)$$

$$\begin{aligned} e_2^* &= e + X_2\beta_2 \\ &= e + X_2 * (-0.034) \end{aligned} \quad (4.3.4)$$

veya

$$\begin{aligned} e_2^* &= Y - (\alpha + X_1\beta_1) \\ &= Y - (226 + X_1 (-1.22)) \end{aligned} \quad (4.3.5)$$

eşitlikleriyle e_1^* ve e_2^* kısmi kalıntı değerleri hesaplanmıştır. Hesaplanan bu değerler Çizelge 3.2.17'de verilmiştir.

e_1^* 'in X_1 'deki regresyonunda en küçük kareler

işlemi uygulanırsa e_1^* ve e_2^* için (3.2.17) nolu eşitlikteki β_1 'yi gösteren aşağıdaki eşitlikler bulunur.

Çizelge 4.3.1. Koyunlara Ait Solunum Sayısı(dk), Nabız Sayısı(dk) Süt Verimleri (kg/koyun), y , e_1 , Sırasıyla(4.3.2) ve (4.3.4)Nolu Eşitlikler Yardımı ile Hesaplanan e_1^* ve e_2^* Değerleri.

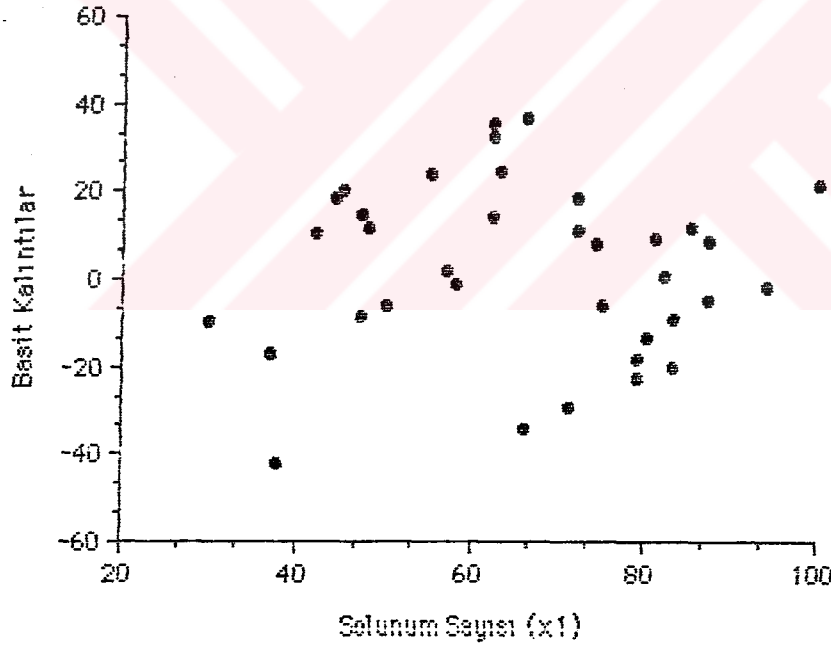
x_1	x_2	y	e_1	e_1^*	e_2^*
62	100	169	22.30	- 53.34	18.90
61	96	181	32.94	- 41.48	29.68
44	80	187	17.59	- 36.09	14.87
43	76	187	16.22	- 36.24	13.64
65	88	178	34.56	- 44.74	31.57
37	92	133	- 44.57	- 89.71	-47.69
78	84	107	- 20.66	-115.82	-23.52
82	90	111	- 11.56	-111.60	-14.62
86	86	124	6.20	- 98.72	3.27
82	92	100	- 22.49	-122.53	-25.61
56	106	153	- 0.84	- 69.16	- 4.44
86	104	110	- 7.19	-112.11	-10.73
70	100	105	- 31.91	-117.31	-35.31
61	102	178	30.14	- 44.28	26.67
47	75	175	9.09	- 48.25	6.54
41	90	181	8.25	- 41.77	5.19
61	93	160	11.84	- 62.58	8.68
29	100	175	- 12.09	- 47.47	-15.49
71	80	152	15.63	- 70.99	12.91
93	80	105	- 4.44	-117.90	- 7.16
79	35	112	- 16.11	-112.49	-20.70
78	96	102	- 25.25	-120.41	-28.51
46	102	155	- 11.22	- 67.34	-14.69
81	114	121	- 1.97	-100.79	- 5.85
74	75	124	- 8.87	- 99.15	-11.42
65	81	107	- 36.68	-115.98	-39.43
54	90	178	21.17	- 44.71	18.11
46	63	180	12.45	- 43.67	10.31
57	70	150	- 3.84	- 73.38	- 6.22
36	75	160	-19.38	- 63.30	-21.93
71	75	145	8.46	- 78.16	5.91
73	66	140	5.60	- 83.46	3.36
99	75	121	18.73	-102.05	16.18
80	90	132	6.99	- 90.61	3.93
84	75	130	- 9.37	- 93.11	6.82
49	75	155	- 8.47	- 68.25	-11.02

$$e_1^* = 226 - 1.22X_1$$

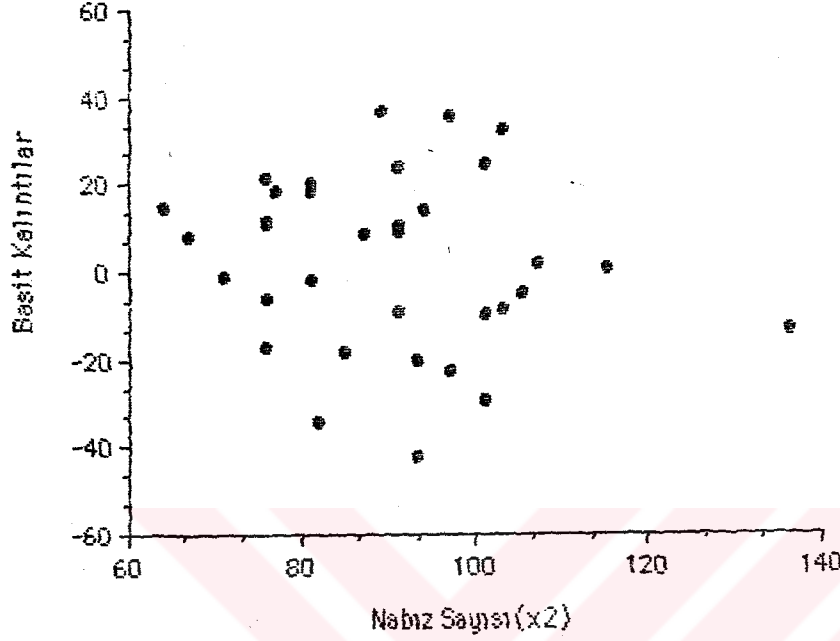
$$e_2^* = 226 - 0.034 X_2$$

Yukarıda da görüldüğü gibi $\hat{\beta}_1 = \hat{\beta}_1$ ve $\hat{\beta}_2 = \hat{\beta}_2$ olur.

Şekil 4.3.1 ve 4.3.2 Çizelge 4.3.1'deki sırasıyla x_1 ve x_2 'ye karşı sıradan kalıntı grafiklerini, Şekil 4.3.3 ve 4.3.4. ise kısmi kalıntı grafiklerini göstermektedir.

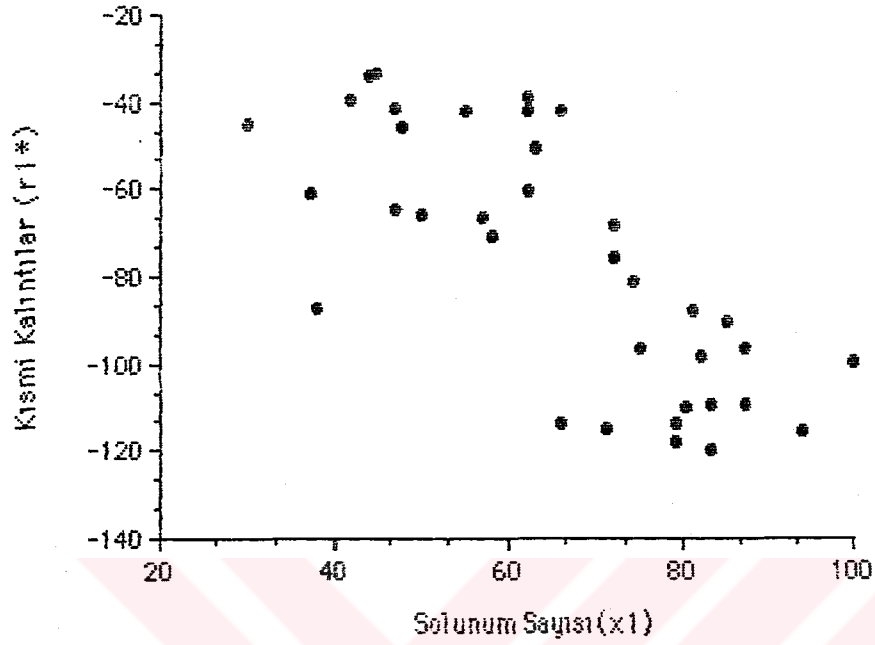


Şekil 4.3.1. Solunum Sayısına(x_1)Karşı Çizilen Basit Kalıntı Grafiği.

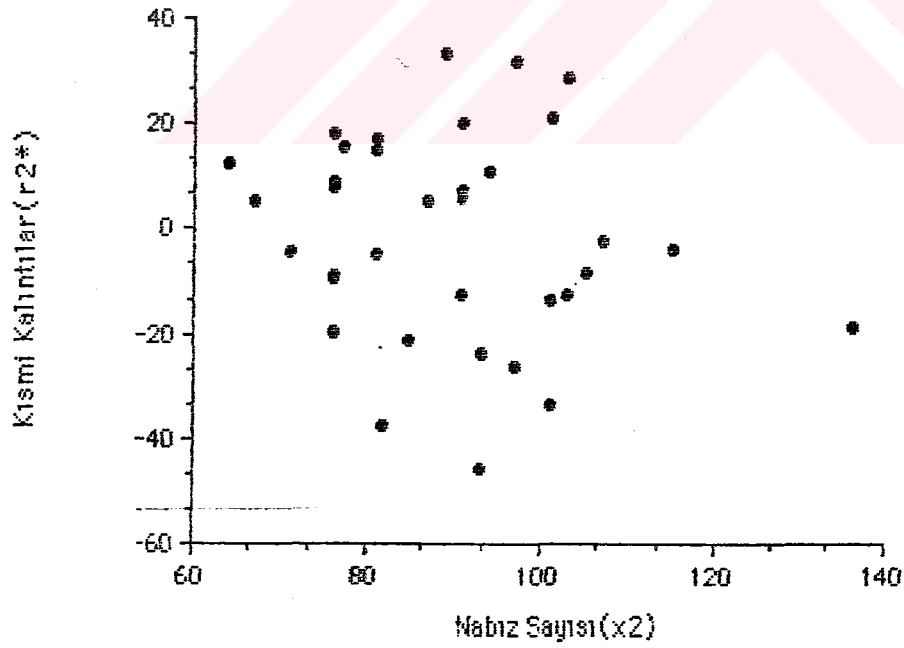


Şekil 4.3.2. Nabız Sayısına(x_2) Karşı Çizilen Basit Kalıntı Grafiği.

Şekil 4.3.1 ve 4.3.2. dikkatli bir şekilde incelendiğinde Şekil 4.3.1'de varyansın x_1 ile birlikte önce artıp sonra azaldığı, Şekil 4.3.2'de ise x_2 ile birlikte sürekli arttığı görülmektedir. Dolayısıyla varyans sabit olmayıp bu varsayım bu örnek için varsayım sağlanmaz. Bu durumda ise tartılı en küçük kareler yöntemi uygulanmalı veya gözlem değerlerinde bir transformasyon yapılmalıdır. Şekil 4.3.4. de Şekil 4.3.2 ile aynı sonuçları ifade ederken Şekil 4.3.3'de, x_2 için düzeltilen bağımlı değişken y 'nin, x_1 ile kesin bir negatif korelasyon içerisinde olduğu görülmektedir. x_1 için düzeltilen y ise x_2 ile sadece hafif bir negatif korelasyona sahip olduğu görülmektedir. Diğer bir ifade ile Şekil 4.3.3 ve Şekil 4.3.4 x_1 'in y üzerinde önemli bir negatif etki, x_2 'nin ise önemli bir etkisinin olmadığını göstermektedirler. Dolayısıyla x_1 , x_2 'ye göre Y için daha önemli bir değişkendir.



Şekil 4.3.3. Solunum Sayısına(x_1) Karşı Çizilen Kısmi Kalıntı Grafiği.



Şekil 4.3.4. Nabız Sayısına(x_2) Karşı Çizilen Kısmi Kalıntı Grafiği

5. TARTIŞMA ve SONUÇ

Kalıntı analizlerinde, genelde kolaylığı ve zayıf parametre tahminlerine neden olan verilerdeki herhangi bir anormallığı kolayca ortaya çıkarması nedeniyle, mümkün olduğunca grafiksel metotlar kullanılır. Fakat çizilen grafiklerin hangisi olursa olsun verilerdeki anormallikleri kolayca belirleyebilmek için biraz tecrübe birikimine ihtiyaç vardır. Bu ise zamanla grafiklerin birkaç defa çizilmesiyle kazanılabilir veya konuyu bilen bir kişiden bilgi almak, araştırmacıyı daha iyi sonuca götürebilir.

x_{ji} 'lere karşı çizilen grafiklerde sadece 2 veya 3 x değişkeni içeren regresyon problemlerinde 2 veya 3 boyutlu bir diyagram çizmek mümkündür. Fakat 3 'den fazla x değişkeni bulunan modellerde alt gruplar oluşturarak diyagram çizmek daha uygundur. Diğerleri iki boyutlu grafiklerle ifade edilebildiği için buna gerek kalmaz.

Örnekte verilen süt verimlerini tahmin etmek amacıyla uydurulan en küçük kareler regresyon modeli yeterli bir model olmayıp bunun yerine tartılı en küçük kareler regresyon modeli uydurulmalıdır. Bunu, yöntem kısmında verilen

kalıntı-zaman, kalıntı - y ve kalıntı- x_{ji} grafiklerindeki kalıntı dağılımları açıkça ortaya koymuştur. Eğer tartılı en küçük kareler modeli uydurularak hesaplanacak kalıntılarla aynı grafikler çizilirse, yeterli bir modeldeki kalıntı davranışları ile yetersiz bir modeldeki kalıntı davranışları mukayese edilebilir ve aralarındaki fark kolayca görülebilecektir.

Kalıntıların normalden sapmaları durumu diğer bir ifadeyle normallik varsayımı, uydurulan modelin yeterliliği hakkında oldukça önemli bilgiler sağlamaktadır. Fakat normalden sapmalarla ilgili kalıntı analizi yapmak oldukça zor

bir iştir. Çünkü bir grup kalıntının dağılımını birçok faktör etkiler. Örneğin uygun olmayan bir regresyon modeli kullanıldığında kalıntılar normal olmayan bir görünüm ortaya koyabilirler veya varyanslar homojen olmayabilir (ki bu durumda çizilen grafik eğer varyasyon çok fazla ise yanıltıcı olabilir ve eğer durum uygunsa e_i/S' 'nin grafiği yerine $e_i / \{(1-v_{ii})S^2\}^{1/2}$ 'nin grafiğini çizmek daha iyidir (Draper ve Smith , 1981). Burada v_{ii} , $V=X(X'X)^{-1}X'$ matrisinin i'nci köşegen elemanına ait değerdir.) veya kalıntı sayısı olasılık dağılımlarının normal durumu ortaya koyacak ve yeterli bir istatistik bilgi verecek genişlikte olmayabilir (Kleinbaum ve Kupper ,1978). Böyle bir durumda ortaya herhangi bir iddia atmadan önce bu gibi varsayımlardan ayrışlarla ilgili birçok örnek inceleyerek iyi bir pratik yapılmalıdır. Bu strateji dahilinde normallik varsayımını bozan standart kalıntı değerleri anormal gözlemler de olabileceği için bu durum incelendikten sonra karar vermek daha uygundur.

Örneğimizde serpmne grafiği normallik varsayımından açık olarak göze çarpan herhangi bir sapma ifade etmemekle birlikte standart kalıntıların normal olasılık kağıtları üzerine çizilen grafiği(Şekil 4.2.2) ve yapılan z standart test istatistiği, normalliği bozan bir gözlemin mevcut olduğunu göstermektedir. Bu durumda bu gözlemin atılarak yeniden bir model uydurularak bu yeni modelin kullanılması daha uygun olur.

i'nci kısmi kalıntı grafiği, i'nci bağımsız değişkene karşı çizilen genel kalıntı grafiklerinde olduğu gibi kullanıma homojen olmayan varyansı, anormal gözlemleri, doğrusal olmayan ilişkilerde her bir bağımsız değişken için doğrusallığın yönünü ve miktarını izah eden bir eğime sahiptir. Bir çoklu regresyon analizinde amaçlardan birisi, her bir bağımsız değişkenin diğer bağımsız değişkenler mevcut olduğunda

bağımlı değişkeni açıklama durumunu belirlemektir. Çoğunlukla x_i 'ye karşı çizilen y orijinal veri grafikleri bize x_i ve $(y|x_1, x_2, \dots, x_{i-1}, x_{i+1}, \dots, x_k)$ arasındaki ilişkiler hakkında ya biraz veya hiç bir bilgi vermezler. Kısmi kalıntı grafikleri standart varsayımlardan sapmaları gösterirken aynı zamanda sözü edilen bu ilişkileri de açıkça gösterirler.

Mükemmele yakın bir uyum olduğu durumlarda kısmi kalıntı grafikleri doğrusallıktan sapmaları gösteremeyip gizleyebilirler. Bu durum daima göz önünde bulundurulmalıdır. Bu ekstrem durumda kısmi kalıntı grafikleri (a) anormal gözlemlerin, (b) varyansın homojen olmaması gibi durumların incelenmesinde yetersiz kalırlar.

X'in birinci sütunu birlerden oluşmuşsa Larsen ve McCleary(1972)'nin bildirdiğine göre Mallows(1970) , (2.1.5) nolu eşitlikte , X_i 'ye karşı ,

$$e_i^* = r + (X_i - \bar{X}_i)\beta_i + \bar{Y}$$

değerlerinin grafiğinin çizilmesini önermiştir. Burada $\bar{X}_i$ ve $\bar{Y}$ sırasıyla, X_i ve Y vektörlerinin ortalamasıdır. Bu

grafiğin orijinini $(\bar{X}_i, \bar{Y})$ noktalarına doğru kaydırır.

Sonuç olarak genelde kalıntı grafikleri istatistikçiye veya araştırmacıya özel bir bilgi sağlayacak mantıklı bir yolla çizilmelidir. Çizilen bu grafikler temel oluştururlar ve analizin tam anlamıyla olabilmesi için daha detaylı bilgi sağlarlar. Diğer bir ifadeyle kalıntıların incelenmesi ile analiz tam olarak yapılmış ve tereddütler ortadan kalkmış olur.

ÖZET

Regresyon analizi, sebep-sonuç ilişkisi bulunan iki veya daha fazla değişken arasındaki ilişkiyi saptamada uygulanan istatistik yöntemlerden birisidir.

Gözlem çiftlerini en iyi temsil eden regresyon doğrusunun aşağıdaki gibi olduğunu varsayalım.

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i \quad (1)$$

Burada α ve β sırasıyla α ve β parametrelerinin tahminlerini, x bağımsız değişkeni ve $\hat{y}$ 'de bağımlı değişkeni göstermekte olup x 'in değeri verildiğinde $\hat{y}$ 'nin ortalama olarak değeri bulunabilmektedir.

Regresyon analizinde amaç tahmin yapmanın yanı sıra gözlem çiftlerini en iyi temsil eden regresyon eşitliğini bulmaktır. Bu da kalıntı analizinin konusunu oluşturur. Çünkü bir kalıntı analizi, tahmin edilen eşitlik ile veri arasındaki uyumu veya uyumsuzluğu ortaya koyan bir analiz şeklidir. Aynı zamanda kalıntı analizi, tahmin eşitliğinin verilere uygunluğunun istatistik olarak belirlenmesinden sonra tahmini değer ile gözlemler arasındaki farklılıkların dikkatli bir şekilde incelenmesini ve verilerdeki zayıf parametre tahminlerine neden olan herhangi bir anormalliğin tespit edilmesini de sağlar (Cook ve Weisberg, 1982).

Kalıntı analizi doğrusal regresyon modelleri yanında doğrusal olmayan regresyon modelleri ve varyans analizinde de kullanılabilir. Gerçekte uydurulmuş herhangi bir modelin uygulanması ve bir grup veri için açıklanamayan

varyasyonun ölçülmesi konularının incelenmesi için de yararlıdır(Draper ve Smith,1981).

Uydurulan bir modelin doğru olup olmadığını kontrol için bazı kalıntı test metotları uygulanır. Bu metotlar içinde çoğunlukla tahminler doğru olmadığında bunu genelde açığa çıkaran ve yapılması kolay olan grafik metotlar kullanılır. Bu nedenle bizde bu çalışmada grafik metotları incelenmiştir.

Regresyon analizinde kalıntıyı oluşturan önemli ayrılış tipleri şunlardır.

1) Anormal gözlemlerin varlığı

2) İlgisiz faktörlerin bulunması ; model (1.1)' den çıkarılır ve bu faktörlerin seviyelerine karşı kalıntı grafikleri oluşturularak incelenebilir.

3) Model (1.1)'e dahil edilen bir faktörle doğrusal olmayan bir ilişkinin bulunması; Yani $\hat{y} = \alpha + \beta x$ gibi bir ilişki yerine $\hat{y} = \alpha + \beta x^2$ gibi bir ilişkinin olması; Bu ilişki , faktörün seviyelerine karşı çizilen kalıntı grafiği ve elde edilen ilişki eğrisi ile incelenebilir.

4) Farklı e_j 'ler arasında ilişki (korelasyon) bulunması; Bu faktör kısmi kalıntı analizi veya kalıntı periyod grafiklerinde incelenebilir.

5) Varyansın sabit olmaması; Bu durum kalıntıların veya karelerinin uydurulmuş değerlere veya varyansı etkileyen faktörlere karşı grafikleri çizilerek incelenebilir.

6) e_j 'lerin normal dağılım göstermemesi ; bu durum sıralı kalıntıların standart normal dağılımdan beklenen sıralı istatistiklere karşı grafikleri oluşturularak incelenebilir.

Bu çalışmanın amacı, kalıntı analizinde uygulanan özellikle grafiksel metotlar kullanılarak, verileri temsil edebilecek en uygun modeli, daha gerçekçi bir yaklaşımla grafiksel analizler kullanılarak kalıntı analizleri ile birlikte

regresyon uygulamalarını yapmanın sağladığı faydaları karşılaştırmalı olarak göstermektir.

Bu çalışmada materyal olarak Ç.Ü. Ziraat Fakültesi Döner Sermaye İşletmesindeki Ost-Frizxİvesi ırkı 36 koyuna ait birer hafta ara ile 9 hafta süreyle alınan solunum sayısı(dk), nabız Sayısı (dk) değerlerinin ortalamaları ve laktasyon süt verimleri (kg/koyun) kullanılmıştır.

Analizlerde aşağıda verilen hesaplama şekillerine göre farklı üç tip kalıntı incelenmiştir. Bunlar;

$$1. \text{ Basit (ordinary) kalıntılar ; } e_i = y_i - \hat{y}_i \quad (2)$$

$$2. \text{ Standart kalıntılar ; } r_i = e_i / Se_i \quad (3)$$

$$3. \text{ Kısmi (partial) kalıntılar ; } e_i^* = e + x_i \beta_i \quad (4)$$

Materyalde verilen solunum sayısı x_1 , nabız sayısı x_2 , süt verimleri ise bağımlı değişken y olarak alınır ve en küçük kareler regresyon doğrusu tahmin edilirse,

$$\hat{y}_i = 226 - 1.22 x_1 - 0.034 x_2 \quad (5)$$

olur ve

$$y_i = 226 - 1.22 x_1 - 0.034 x_2 + e_i \quad (6)$$

eşitliği elde edilir.

Regresyon katsayısının önem kontrolü için yapılan varyans analiz sonuçlarına göre, koyunların süt verimleri ile solunum sayısı veya nabız sayısı arasındaki azalış istatistik bakımdan önemlidir($p < 0.01$). Fakat örneğe ait $R^2 = \% 55.69$ olup yüksek bir değer sayılmaz. Yani koyunların dakikadaki solunum sayıları ve nabız sayıları laktasyon süt verimlerindeki

varyasyonun ancak % 55.6'sını açıklayabilmektedir. Regresyon önemli olduğu halde R^2 'nin yüksek çıkmaması, gözlemler içerisinde anormal gözlemlerin olabileceğini gösterir.

Eğer gözlemler içerisinde anormal gözlemler varsa bunu ortaya çıkarmamanın bir yolu kalıntı analizi yapmaktır.

Yapılan değişken eleme testi sonucu koyunlarda dakikadaki nabız sayısının süt verimine önemli bir etkisi olmadığı yani süt verimlerindeki varyasyonun önemsiz bir kısmını oluşturduğu görülmüştür. Diğer bir deyişle sadece dakikadaki solunum sayısı Y 'deki açıklanabilen varyasyonun büyük bir kısmını açıklayabilmektedir. Bu nedenle bu çalışmada dakikadaki solunum sayısı (x) ile laktasyon süt verimleri (y) arasındaki ilişkiyi gösteren regresyon modeli üzerinde çalışılmıştır. Uydurulan bu yeni model ise,

$$\hat{y}_1 = 223 - 1.22 x_1 \quad (7)$$

dir ve yapılan varyans analizi sonucuna göre yine solunum sayısı ile laktasyon süt verimleri arasındaki regresyon katsayısı önemli bulunmuştur ($p < 0.01$). Uydurulan bu eşitlik yardımıyla,

1) (2) nolu eşitlik kullanılarak basit kalıntı değerleri hesaplanmıştır. Hesaplanan bu değerler daha sonra ,

- Kalıntıların tümünün grafiği,
- Eğer sıralama biliniyorsa kalıntıları zaman serisinde gösterme,

- Uydurulan y_j değerlerine karşı grafik çizme,
- $j=1,2,...,k$ için bağımsız x_{ji} değişkenlerine karşı grafikleri çizilerek incelenmiştir. Sonuçta , örneğe ait varyansın sabit olmadığı, bu nedenle en küçük kareler yöntemi yerine tartılı en küçük kareler yöntemi uygulanması veya gözlemlerde analizden önce bir transformasyon yapılması gerektiği görülmüştür.

2) (7) nolu eşitlikten elde edilen tahmini değerlerden (3) sayılı formül yardımıyla standart kalıntı değerleri hesaplanmış ve bu değerlerin önce serpine grafiği daha sonra da normal grafikleri oluşturularak kalıntıların normallik varsayımının mevcut olup olmadığının kontrolünde kullanılmıştır. Örneğimizde bu grafikler incelendiğinde normalliği bozan bir gözlemin olduğu görülmüştür.

3) (5) nolu eşitlikten elde edilen tahmini değerlerden (4) sayılı formül yardımıyla kısmi kalıntı değerleri hesaplanmış ve grafikleri incelenmiştir. Grafikler dikkatli bir şekilde incelendiğinde varyansın x_1 ile birlikte önce artıp sonra azaldığı, x_2 ile birlikte sürekli arttığı görülmüştür. Dolayısıyla varyans sabit olmayıp bu varsayım bu örnek için mevcut değildir. Bu durumda ise tartılı en küçük kareler yöntemi uygulanmalıdır. Ayrıca x_2 için düzeltilen bağımlı değişken y , x_1 ile kuvvetli bir negatif korelasyon içerisinde olduğu, x_1 için düzeltilen y ise x_2 ile sadece zayıf bir negatif korelasyona sahip olduğu görülmüştür. Diğer bir ifade ile x_1 'in y üzerinde oldukça yüksek bir azalış, x_2 'nin ise oldukça düşük bir azalışa neden olduğunu göstermektedirler.

SUMMARY

The regression analysis is one of the statistical methods which is applied to determination of relationship between one or more variables.

We assume that the regression equation which is the best fitted to the data is

$$\hat{y}_i = \hat{\alpha} + \hat{\beta}x_i \quad (1)$$

Where $\hat{\alpha}$ and $\hat{\beta}$ are the estimation of parameters α and β , x is independent variable, y is dependent variable. When the value of x is given, estimation of y can be found.

The aim of the regression analysis is the prediction and fitting of a regression equation to the pairs of observation which must represent the observations better than others. This is also subject to residual analysis. Because, the residual analysis is a type of analysis which examines disagreement between assumed model and the data. However, after determining the regression equation the residual analysis provides examining the differences between observed and predicted values carefully (Cook and Weisberg, 1982).

Residual analysis is not only a usefull tool for linear regression models, but also for non-linear regression models and analysis of variance models. In fact, residual analysis applies to any situation where a model is fitted and measures of unexplained variation (in the form of a set of residuals) are available for examination.

There are some residuals test methods to test the fitted regression model whether the model fits the data appropriately or not. However, we applied the graphical analysis method because the applicatin of it was easy and uses

extensively.

The main types of departures in regression analysis are,

- 1) The presence of outliers;
- 2) The relevance of a further factor, omitted from model(1), detected by plotting the residuals against the levels of that factor;
- 3) Non-linear regression on a factor already included in model (1), detected by plotting the residuals against the levels of that factor and obtaining a curved relationship;
- 4) Correlation between different ϵ_i 's, detected from partial residual analysis or a periodogram analysis of residuals.
- 5) Non-constancy of variance; detected by plotting residuals or square residuals against factors thought to affect the variance or against fitted values;
- 6) Non-normality of the distribution of the ϵ_i 's, detected by plotting the ordered residuals against the expected order statistics from a standard normal distribution.

The aim of this study is to determine sufficiency of the regression model by using graphical analysis method of residuals which were calculated in different ways.

The data of this study was the lactation milk yield (kg), the average of respiration (min) and pulse rate (min) which were obtained during the nine weeks period in one week interval from 36 East-Friesian x Awassi crossbreed sheep. In this study the following types of residuals were examined.

$$1. \text{ Ordinary residuals } ; e_i = Y_i - \hat{Y}_i \quad (2)$$

$$2. \text{ Standardized residuals } ; r_i = e_i / \text{Se}_i \quad (3)$$

$$3. \text{ Partial residuals } ; e_i^* = e_i + x_i \hat{\beta}_i \quad (4)$$

If the respiration pulse rate and lactation milk yield were

represent by x_1 and x_2 , dependent variable y and the following equations is obtained,

$$\hat{y}_1 = 226 - 1.22 x_1 - 0.034 x_2 \quad (5)$$

$$y_1 = 226 - 1.22 x_1 - 0.034 x_2 + e_i \quad (6)$$

As a result of the analysis of variance which was made for testing of coefficient of regression that between the lactation milk yield with the respiration and pulse rate were significant ($p < 0.01$). But the coefficient of determination (R^2) was found to be 55.69 % which was not so high. This means that the respiration and pulse rate can explain 55.69 % variation in y . The R^2 was insignificant while the coefficient of regression were significant which mean that there were some outlier observations in data. If so, the residual analysis should have been done.

The pulse rate was eliminated by using the Backward Elimination method. Therefore, we defined a regression equation between the lactation milk yield and respiration as follows:

$$\hat{y}_1 = 223 - 1.22 x_1 \quad (7)$$

The regression coefficient between lactation milk yield and respiration were found to be significant by using the method analysis of variance ($p < 0.01$). Therefore, the drove some graphs with respect to the fitted regression equation (7) as follows:

— 1). The ordinary residuals were calculated by using equation (2) after that this calculated values were plotted as follows,

- Overall
- In time squance, if the order is known

- Against the fitted values $\hat{y}$
- Against the independent variable x_{ji} , for $j=1,2,\dots,k$

In the result, it was seen that the variance is not constant. So that, the weighted least squares method must be applied instead of the least squares method or a transformation on the observations before the analysis.

2) Standardized residuals were calculated by using equation (3), examined the scatter diagram and constructed a normal plot of the standardized residuals and was used as a yardstick in assessing whether the scatter of points reflects any obvious deviation from normality. In our case, we saw an observation which departs from the normality.

3) The partial residuals were calculated by using equation (5), plotted the graphics and examined. When the plots were examined carefully we saw the variance is increasing with x_1 and after decreasing. Therefore, the variance of the error terms is not constant. In this case, the weighted least squares method must be applied. The dependent variable corrected for x_2 has a definite negative correlation with x_1 , the dependent variable corrected for x_1 has a slightly negative correlation with x_2 .

KAYNAKLAR

- AKAR,M.; ŞAHİNLER,S., 1992. Kalıntıların Grafik Analizleri Yardımıyla Regresyon Modelinin Yeterliliğinin Saptanması. 'Araştırma 92' Simpozyumu, Ankara.
- ASCOMBE, F.J.; TUKEY J.W.,1963. The Examination and Analysis of Residual , Technometrics,5:141-160.
- BEHNKEN, R.W.; DRAPER, N.R., 1972. Residual and Their Variance Patterns. Technometrics, 14:101-111.
- CRICKET GRAPH Version 1.2, 1987. Malvern, PA , 19355 USA.
- COOK,R.D.; WEISBERG, S., 1982. Residuals and Influence in Regression. School of Statistics University of Minnesota Newyork- London.
- COX,D.R.; SNELL, E.J., 1968 . A General Definition of Residuals. Journal / Royal Statistical Society.Serie,30 : 348-375.
- DRAPER,N.R.;SMITH,H.,1981 .The Exemination of Residuals. Applied Regression Analysis. Chapter 3:141-190. Newyork.
- DURBIN, J., 1969. Tests for Serial Correlation in Regression Analysis Based on The Periodogram of Least Squares Residuals.
- ELLENBERG, J.H.,1973. The Joint Distribution of The Standardized Leased Squares Residuals from a General Linear Regression. Journal of American Statistical Association, 68: 941-943.

FAREBROTHER, R.W., 1976. BLUS Residuals. Applied Statistics, 25:317-319.

HABERMANN, Shelby J., 1973. The Analysis of Residuals in Cross-Classified Tables. Biometrics, 12:205-220.

KLEINBAUM, D.G.; KUPPER, L.L., 1978. Residual Analysis Transformations, and Weighted Least Squares. Applied Regression Analysis and Other Multivariable Methods. Chapter 16: p.235-243. University of North California, Massachusetts.

LARSEN, W. A.; McCLEARY, J., 1972. The Use of Partial Residual Plots in Regression Analysis. Technometrics, 14:781-790.

LOYNES, M.R., 1969. On Cox and Snell's General Definition of Residuals. Technometrics, 66:122-145.

MINITAB Statistical Software Version II, 1982. Pennsylvania State University, PA, U.S.A.

NETTER, J.; WASSERMANN, W., 1973. Applied Linear Statistical Models. Chapter: 4, 97-136. Georgetown, Ontario.

SCHWARZMANN, B., 1991. A Connection Between Local Influence Analysis and Residual Diagnostics. Technometrics, 33:103-104.

THOMAS, A., RYAN, J.R., 1982. Minitab Reference Manual. Cuxbury Press. 20 Providence Street, Boston.

TEŞEKKÜR

Bu araştırmanın yapılabilmesi için her türlü imkanı sağlayan, tüm çalışmalarım boyunca bana yol gösteren ve tezin yazımı sırasında yardımlarını esirgemeyen başta sayın danışman hocam Doç.Dr. Mustafa AKAR'a ve Anabilim Dalı Başkanımız Sayın Prof. Dr. Yüksel BEK'e, bölüm imkanlarından yararlanmama izin veren ve bu konuda yardımcı olan bölüm başkanımız Sayın Prof.Dr. Erdoğan PEKEL'e, çalışmalarım boyunca her türlü desteği veren sevgili eşim Nuray ŞAHİNLER'e ve çalışmaya katkısı olan tüm çalışma arkadaşlarıma teşekkür ederim.

ÖZGEÇMİŞ

1966 yılında Konya ili Ereğli ilçesinde doğdum. 1972 yılında ailemle birlikte Adana'ya yerleştim. İlk ve orta öğrenimimi 1985 yılında burada tamamladım ve aynı yıl Ç.Ü. Ziraat Fakültesi Zootekni Bölümünde Lisans öğrenimime başladım. 1988-1989 öğretim yılında lisans öğrenimimi bölüm birincisi olarak tamamlayarak 1989-1990 öğretim yılında Ç.Ü. Fen Bilimleri Enstitüsünde Yüksek Lisans eğitimine başladım. Halen Yüksek Lisans eğitimine devam etmekte olup , Zootekni Anabilim dalında Araştırma Görevlisi olarak görev yapmaktayım.