

T.C.
BİNGÖL ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

**JAPON BILDİRCİNLERİNDE (Coturnix coturnix japonica) CANLI AĞIRLIK
PERFORMANSINA AİT VERİLERİN AYKIRI DEĞERLERİNİN
BELİRLENMESİ**

YÜKSEK LİSANS TEZİ

Burhan BAHADIR

Enstitü No: (.....)

Tezin Enstitüye Verildiği Tarih: 26.03.2013
Tezin Savunulduğu Tarih: 12.04.2013

Tez Danışmanı: Yrd. Doç. Dr. Hakan İNCİ (B.Ü.)
İkinci Danışman: Doç. Dr. Ufuk KARADAVUT (A.E.Ü)
Diğer Jüri Üyeleri: Prof. Dr. Turgay ŞENGÜL (B.Ü)
Doç. Dr. Hasan KILIÇ (B.Ü)
Yrd. Doç. Dr. Bünyamin SÖĞÜT(B. Ü)

MAYIS- 2013

Yrd. Doç. Dr. Hakan İNCİ ve Doç. Dr. Ufuk KARADAVUT danışmanlığında, Burhan BAHADIR 'ın “ **Japon Bildirecilerinde (Coturnix Coturnix Japonica) Canlı Ağırlık Performansına Ait Verilerin Aykırı Değerlerinin Belirlenmesi** ” konulu bu çalışma 12 / 04/ 2013 tarihinde aşağıdaki jüri tarafından Bingöl Üniversitesi Ziraat Fakültesi Zootekni Anabilim Dalı'nda Yüksek Lisans tezi olarak kabul edilmiştir.

Danışman :Yrd. Doç. Dr. Hakan İNCİ

İkinci Danışman:Doç. Dr. Ufuk KARADAVUT

Üye :Prof. Dr. Turgay ŞENGÜL

Üye :Doç. Dr. Hasan KILIÇ

Üye :Yrd. Doç. Dr. Bünyamin SÖĞÜT

Bu tezin Zootekni Anabilim Dalı'nda ve Enstitümüz Kurallarına göre düzenlendiğini onaylıyorum.

Enstitü Müdürü

Not: Bu tezde kullanılan özgün ve başka kaynaktan yapılan bildirişlerin, çizelge, şekil ve fotoğrafların kaynak gösterilmeden kullanımı, 5846 sayılı Fikir ve sanat Eserleri Kanunundaki hükümlere tabidir.

ÖNSÖZ

Bu çalışmanın yapılması için gerekli ortamı sağlayan Bingöl Üniversitesi Ziraat Fakültesi Zootekni Bölümü ve Ahi Evran Üniversitesi Ziraat Fakültesi Zootekni Bölümüne teşekkür ederim.

Çalışmalarım boyunca bana yol gösteren, her konuda ilgisini ve desteğini esirgemeyip yardımcı olan çok değerli danışmanlarım, Sayın Hocam Doç.Dr.Ufuk KARADAVUT, Yrd. Doç. Dr. Hakan İNCİ' ye, Ziraat Fakültesi Dekanımız Sayın Prof. Dr. Turgay ŞENGÜL'e ve Yrd. Doç. Dr. Bünyamin SÖĞÜT'e, Tarla Bitkileri Bölümü Öğretim Üyelerinden Sayın Doç.Dr.Kağan KÖKTEN'e, Ayrıca manevi desteğini ve dualarını esirgemeyen sevgili annem Tayyube BAHADIR'a, babam Ali BAHADIR'a Bingöl Gıda Tarım ve Hayvancılık İl Müdürü Sayın Sedat ILDIZ'a ve değerli mesai arkadaşım Şebnem İNCİ'ye sonsuz teşekkürlerimi sunarım.

Burhan BAHADIR

BİNGÖL-2013

İÇİNDEKİLER

	<u>SAYFA NO</u>
ÖNSÖZ.....	II
İÇİNDEKİLER.....	III
ÖZET.....	IV
ABSTRACT.....	V
ÇİZELGELER DİZİNİ.....	VI
ŞEKİLLER DİZİNİ.....	VII
1.GİRİŞ	1
1.1 ÖNCEKİ ÇALIŞMALAR	3
2. MATERYAL VE METOD	8
2.1. Aykırı Değer Belirleme Yöntemleri	8
2.1.1. DFBETA ve DFBETAS ölçüleri.....	14
2.1.2. DFFİT ve DFFITS ölçüleri.....	14
2.1.2.1. Cook Uzaklığı	15
2.1.2.2. Atkinson (1986) Algoritması	15
3.BULGULAR	16
4. SONUÇLAR VE ÖNERİLER.....	36
5. KAYNAKLAR	37
ÖZGEÇMİŞ	40

ÖZET

Bu çalışmanın amacı Japon bildircinlarında canlı ağırlık performansına ait verilerin aykırı olan değerlerini belirlemektir. Japon bildircinleri aynı koşullar altında yetiştirilmişlerdir. Çalışmada bildircinların cinsiyetlerine göre ayrılarak 56 günlük yaşa kadar (7 hafta) olan canlı ağırlık verileri toplanmıştır. Her ölçümde hem erkek hem de dişiler için 50'şer hayvan ölçüme alınmış ve bu değerler üzerinden ölçüm yapılmıştır. Ölçülen her hafta için ayrı ayrı aykırı değerler belirlenmiştir. Aykırı değer belirleme işlemi yapılırken, Cook Uzaklığı, Atkinson Ölçüsü, DFFİTS ve DFBETA ölçütleri kullanılmıştır. Her ölçümdeki aykırı değerler bu yöntemlere göre belirlenmiştir. Buna göre her ölçümde erkekler , dişiler ve sürü toplamı olarak değerlendirilmiş ve her ölçümde mutlaka aykırı değere rastlanmıştır. Ancak özellikle dişilerde daha fazla aykırı değer gözlenmiştir. Aynı koşullar altında yetiştirilmelerine rağmen dişilerdeki değişkenliğin olası çevresel etkilere erkeklere göre daha hızlı tepki göstermelerinden kaynaklandığı görülmüştür.

Anahtar kelimeler: Bildircin, Canlı Ağırlık, Aykırı Değer

ABSTRACT

Determination of Live Weight Determination of the Performance of outlier data in Japanese Quails (*Coturnix coturnix japonica*)

Master Thesis

Burhan BAHADIR

Supervisor: Asst.Prof.Dr.Hakan İNCİ

Assoc.Dr.Ufuk KARADAVUT

This study was aimed to determine the outlier values of the data of Japanese quail to the performance of live weight. Japanese quails are grown under the same conditions. In this study, divided according to gender quail up to 56 days of age (7 weeks), the live weight data were collected. Each measurement for both male and females were 50-animal measurements, and these values were measured over. For measured each week, the values are separately determined. During the process of determining outlier value, Cook distance, Atkinson measur, DFFİTS and DFBETA criteria were used. Outlier values are determined in Each measurment by this method. Accordingly, men, females, and in all the flock are strongly founded outlier value. However, especially in females were observed more outlier. Under the same conditions, in spite of the training, the potential environmental effects of the variability in females was originated to react more quickly than men.

Keywords: Quail, Live Weight, The Outlier

ÇİZELGELER DİZİNİ

Sayfa No

Çizelge 3.1. DFBETA İle Elde Edilen Erkek, Diři ve Sürü Toplamı Deęerleri	18
Çizelge 3.2. DFBETAS İle Elde Edilen Erkek, Diři ve Sürü Toplamı Deęerleri	22
Çizelge 3.3. DFFİTS İle Elde Edilen Erkek, Diři ve Sürü Toplamı Deęerleri	26
Çizelge 3.4. COOK İle Elde Edilen Erkek, Diři ve Sürü Toplamı Deęerleri	30

ŞEKİLLER DİZİNİ

Sayfa No

Şekil 1.1. Örnek Bir Aykırı Değer Durum.....	4
Şekil 1.2. Aykırı Değerle İlgili Bazı Testler	5
Şekil 2.1. Altı Gözlemden Oluşan ve Aykırı Gözlem Barındırmayan Regresyon Doğrusu.....	10
Şekil 2.2. Altı Gözlemden Oluşan ve X Yönünde Aykırı Gözlem Barındıran Regresyon Doğrusu.....	11
Şekil 2.3. Altı Gözlemden Oluşan ve Y Yönünde Aykırı Gözlem Barındıran Regresyon Doğrusu	12
Şekil 2.4. Altı Gözlemden Oluşan ve XY Yönünde Aykırı Gözlem Barındıran Regresyon Doğrusu	13
Şekil 2.5. Maskeleye ve Süpürme Etkisi.....	14
Şekil 3.1. Erkek Bildiricilerde DFBETA İçin Elde Edilen Aykırı Değer Grafiği....	19
Şekil 3.2. Dişi Bildiricilerde DFBETA İçin Elde Edilen Aykırı Değer Grafiği.....	20
Şekil 3.3. Bildiricilerde Sürü Toplamı DFBETA İçin Elde Edilen Aykırı Değer Grafiği.....	21
Şekil 3.4. Erkek Bildiricilerde DFBETAS İçin Elde Edilen Aykırı Değer Grafiği .	23
Şekil 3.5. Dişi Bildiricilerde DFBETAS İçin Elde Edilen Aykırı Değer Grafiği....	24
Şekil 3.6. Bildiricilerde Sürü Toplamı DFBETAS İçin Elde Edilen Aykırı Değer Grafiği.....	25
Şekil 3.7. Erkek Bildiricilerde DFFİTS İle Elde Edilen Aykırı Değer Grafiği.....	27
Şekil 3.8. Dişi Bildiricilerde DFFİTS İle Elde Edilen Aykırı Değer Grafiği.....	28
Şekil 3.9. Bildiricilerde Sürü Toplamı DFFİTS İçin Elde Edilen Aykırı Değer Grafiği	29
Şekil 3.10. Erkek Bildiricilerde COOK İle Elde Edilen Aykırı Değer Grafiği.....	31

Şekil 3.11. Dişli Bıldırcınlarda COOK İle Elde Edilen Aykırı Değer Grafiği.....	32
Şekil 3.12. Bıldırcınlarda Sürü Toplamı İçin COOK İle Elde Edilen Aykırı Değer Grafiği.....	33

1.GİRİŞ

Aykırı gözlem sorunu, temelde eski bir sorun olarak, önceleri hangi gözlem veya gözlemlerin aykırı olabileceği şeklinde öznel bir yapıya sahipken, bu sorun bilgisayar teknolojilerinin gelişmesiyle çok boyutlu verilerde ve regresyon analizinde verinin genel yapısına uymayan gözlem veya gözlem gruplarının teşhis edilmesine yönelik algoritmaların tanımlanması gibi bir soruna dönüşmüştür. Bilimsel çalışmalardan elde edilen veriler genellikle normal dağılım gösterirler. İstatistiksel analiz yöntemlerinin büyük çoğunluğu da üzerinde çalışılan verilerin normal dağılmış olduğu varsayımı üzerine geliştirilmişlerdir. Bu nedenle asıl analizlere geçmeden önce verilere ilişkin normallik testlerinin yapılması gerekmektedir (Bek ve Efe, 1987; Akdeniz, 1998). Bilimsel çalışmalar sonunda elde edilen sayısal değerlerin incelenmesinde zaman zaman bir veya daha fazla gözlemin, diğer gözlemlerin oldukça uzağında kalması ile karşılaşılır. Böyle gözlemlere uç değer, derbeder değer, uyumsuz gözlem, şüpheli gözlem değeri, sürpriz değer, kirliliği, kirlenici, sapan değer (Outlier) v.b. denile gelmiştir (Çil, 1990; Billor et. al., 2000). Bunlar doğal rastgelelik sonucunda ortaya çıkabildikleri gibi kişilerin, makinelerin hatalarından veya benzer nedenlerden oluşabilmektedir (Kaya, 1999).

Dağılımın normal olması beklenen veri setlerinin normal dağılımı göstermemesi durumunda ilk akla gelen nedenlerden biri de aykırı gözlemlerinin varlığıdır. Bu çalışmada aykırı gözlemlerin belirlenmesi için kullanılan bazı yöntemler incelenecektir. Aykırı gözlemler, veri setinin ortalamasının çok uzağına düşen gözlemler olarak ifade edilir. Bu değerler, bir tane olabileceği gibi birden fazla da olabilir. Bu değerler, verilerin standart sapmasını artırmanın dışında, dağılımın şeklini de değiştirebilir ve istatistik karar süreci sonucunda hatalı kararlar verilmesine neden olabilirler (Alpar, 1997). Aykırı gözlemler, verinin çoğunluğundan belirgin bir şekilde sapan değerlerdir. Verinin çoğunluğunun uyduğu mekanizmadan farklı bir şekilde oluşmuş veya kayıt hatası, üretim aşamasında bir bozukluk veya insanlarla ilişkili sebeplerden meydana gelebilirler. Aykırı gözlemler model kurma hatasına, yanlış parametre tahminlerine veya yanlış analiz sonuçlarına sebep olur (Liu, ve ark., 2004).

Bazı durumlarda verilerimizde eksik veriler bulunabilir. Eksik verilerin bulunması verilerde aykırı değerlerin var olmasına neden olabilir. Bu nedenle öncelikle eksik

gözlemlerin belirlenmesi ve ona göre analiz edilmeleri gereklidir. Eksik verilerin birkaç nedeni olabilir: a- Alınan örnek istenilen cevabı vermemiş olabilir, b- Alınan örnek gerçekten konuyu açıklamaya yardımcı olmayan bir örnek olabilir. Yani yanlış örnek seçilmiş olabilir. Bunu da bizleri yanlış sonuçlara götüreceğinden sonuçlar işleme konulmaz. c- Veri alınmıştır ama bilgisayara girilmemiştir. Eksik verinin sebebi ne olursa olsun istenmeyen bir durumdur. Hatta veri eksikliği olan değişken esas araştırma konumuz ise bu durumda durum daha da ciddidir. Hatta lojistik regresyon gibi bazı analizler yapılamaz veya sonuçların güvenilirliği etkilenir. Eksik veri konusuna özellikle dikkat edilmelidir.

Doğrusal regresyon analizi, hatalara ilişkin standart varsayımlar dediğimiz varsayımların sağlanması durumunda En Küçük Kareler (EKK) tahmin yöntemleri ile en iyi yansız tahminleri verir. Ancak veride aykırı değerlerin bulunması durumunda bu varsayımlarda bozulmalar meydana gelebilir (Aydın, 2006). Örneğin; hata dağılımının normalliği bozulabilir; değişen varyans problemi ortaya çıkabilir ve bu yüzden tahminler yanlış ve büyük varyanslı olabilirler (Rousseeuw ve Zomeron, 1990). Demek ki güvenilir tahminlere ulaşmak için veri içerisinde bu tür problemlere sebep olabilecek gözlemlerin tespit edilmesi gerekmektedir. Veri seti içerisinde sadece bir tane aykırı değer bulunması durumunda bunları tespit etmek için güvenilir ve kolay uygulanabilir teknikler mevcuttur. Fakat birden çok aykırı değer mevcut olması durumunda aykırı değerler bazen birbirlerinin varlıklarını gizleyebilmekte ve hatta bu aykırı değerler klasik tahmin yöntemlerinde herhangi bir sorun teşkil etmeyen gözlemlerin bile aykırı değerler olarak görünmesine neden olabilmektedir (Hadi ve Simonoff, 1993).

Ortalamadan uzak bir noktada bulunan bu gözlemlerin aykırı olup olmadığını belirleyebilmek amacıyla geliştirilmiş birçok istatistik test mevcuttur. Bu testlerden bazıları sadece bir, bazıları da aynı anda birden fazla gözlemin istatistiksel olarak aykırı gözlem olup olmadığını belirleyebilirler. Teşhis edilen gözlem gerçekten aykırı gözlem midir, aykırı gözlem veriden ihraç edilmeli midir ya da veriden ayrı mı incelenmelidir, gözlemler hangi sebeplerle veriden farklılık gösterirler, bu doğal bir fark mıdır? yoksa kayıt hatası mıdır? İşte bu tür soruların cevapları tamamen araştırmacı tarafından belirlenmelidir ve aykırılığın öznel doğası yüzünden kesin bir şey söylemek mümkün olmamaktadır.

Bu çalışmanın amacı, aykırı gözlemlerin belirlenmesinde literatürdeki yöntemleri tanıtmak ve bıldircılara ait verileri kullanarak belirtilen metotları uygulamalı olarak gösterebilmektir.

1.1. Önceki Çalışmalar

Gentlemen ve Wilk (1975), aykırı gözlemleri bulma işleminin aslında bir karşılaştırma olduğunu ve bir aykırı gözlem sadece bir modele veya teoriye göre aykırı gözlem olduğunu ifade etmişlerdir.

Jiang ve ark. (2004), aykırı gözlemler için, verinin çoğunluğundan belirgin bir şekilde sapan değerler olduğu şeklinde bir tanımlama yapıştır. Verinin çoğunluğunun uyduğu mekanizmadan farklı bir şekilde oluşmuş veya kayıt hatası, üretim aşamasında bir bozukluk veya insanlarla ilişkili sebeplerden meydana gelebilirler. Aykırı gözlemlerin model kurma hatasına, yanlış parametre tahminlerine veya yanlış analiz sonuçlarına sebep olabileceği belirtilmiştir.

Aykırı gözlemler verinin çoğunluğundan uzakta bulunan gözlemlerdir, Verinin çoğunluğunun sahip olduğu dağılımdan farklı bir dağılıma ya da aynı dağılıma fakat farklı parametrelere sahip oldukları düşünülür. Rousseeuw ve Zomeren'a (1990) göre aykırı gözlemler, verideki toplam gözlem sayısının yarısından daha az sayıda olmasına rağmen, o verideki gözlemlerin çoğunun vermek istediği bilgiye engel olan ve sonuçlar üzerinde yanıltıcı bir etki yaratabilen gözlemlerdir.

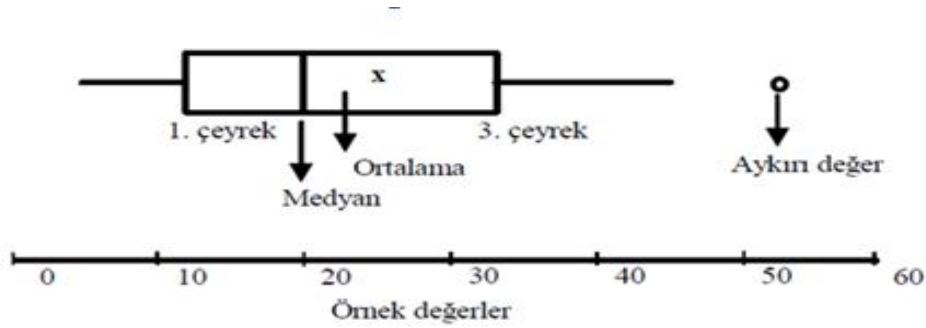
Pearson ve Sekar (1936), Thompson (1935) 'e atıfta buldukları makalelerinde, örneklemeden hesaplanan ortalama ve standart sapma değerlerinin birden fazla aykırı gözlemin varlığı altında bu gözlemlerce etkileneceğini, bu yüzden Thompson'un test istatistiğinin küçük örneklemelerde ancak tek bir aykırı gözlemin varlığı durumunda başarılı olabildiğini göstermiştir.

Collett ve Lewis (1976), tek değişkenli aykırı gözlemlerin teşhisinde başka bir sorun olarak da, aykırı gözlem teşhis sürecinin öznel doğası olduğunu öne sürmüştür. İlk etapta araştırmacı tarafından görüldüğünde verinin çoğunluğuna uymadığı düşüncesini doğuran gözlemlere şüpheli gözlem, istatistiksel bir kritere uygunluk sonucu verinin çoğunluğuna uymayan gözlemlere de bozucu gözlem denmiştir. Bu şekilde bozucu gözlemin teşhisi için iki aşama vardır. Araştırmacının kişisel tercihi olarak bir gözlemi veya gözlemler kümesini şüpheli bulmak ve sonra da şüpheli gözlemlerin aykırılığını test

etmektedir. İkinci bir öznel karar, aykırı gözlemin teşhisi için hangi yöntemin kullanılacağı ve son olarak da hipotezin hangi hata düzeyinde reddedileceğidir.

Aykırı değerlerin varlığını teşhis etmede çeşitli tanısal yöntemler geliştirilmiştir. Bunların içerisinde standartlaştırılmış ve jackknife artıklara dayalı tanısal yöntemler aykırı değerlerin tespiti için sıkça başvurulan yöntemlerdir (Cook, 1979).

Aykırı gözlemler, veri setinin ortalamasının çok uzağına düşen gözlemler olarak ifade edilir. Bu değerler, bir tane olabileceği gibi birden fazla da olabilir. Bu değerler, verilerin standart sapmasını artırmanın dışında, dağılımın şeklini de değiştirebilir ve istatistik karar süreci sonucunda hatalı kararlar verilmesine neden olabilirler. Şekil 1.1’de aykırı bir değer durumu görülmektedir.



Şekil 1.1. Örnek bir aykırı değer durumu (West, 1999a).

Üçkardeş ve ark, (2006), ortalamadan uzak bir noktada bulunan bu gözlemlerin aykırı olup olmadığını belirleyebilmek amacıyla geliştirilmiş birçok istatistik test içerisinde bazılarını ele alarak incelemişlerdir. Bu testlerden bazıları sadece bir, bazıları da aynı anda birden fazla gözlemin istatistiksel olarak aykırı gözlem olup olmadığını belirleyebilirler. Bu durum Şekil 1.2’de gösterilmiştir.

Testler	Normallik varsayımı	Örnek büyüklüğü	Aykırı gözlem sayısı	Test sınıfı
Dixon Test	Normal	$3 \leq n \leq 25$	Tek	Parametrik
Rosner Test	Normal	$n \geq 25$	Çok	Parametrik
Discordance Test	Normal	$n \leq 50$	Tek	Parametrik
Grubbs Test	Normal	$n \leq 50$	Tek/Çok	Parametrik
Walsh Test	Normal olmayan	$60 \leq n \leq 220 \quad \alpha = 0.10$ $n > 220 \quad \alpha = 0.05$	Çok	Parametrik olmayan

Şekil 1.2. Aykırı değerle ilgili bazı testler (Üçkardeş, 2006).

Seber (1984), tek bir aykırı gözlemin bulunabilmesi hem regresyon analizinde hem de çok değişkenli verilerde cebirsel işlemlerle mümkün olabilmektedir. Birden fazla aykırı gözlemin teşhis yöntemleri ise gerek regresyon analizinde gerekse çok değişkenli verilerde iteratif algoritmalara ve tüm mümkün alt kümelerin taranmasına dayanmaktadır. Öte yandan aykırı gözlem sayısı ne olursa olsun teşhis edilebilmeleri iki boyutta çok kolayken, boyut arttıkça verinin grafiklerle gösterilmesi imkanı olmadığından aykırı gözlem teşhisinin zorlaştığını belirtmiştir.

Satman (2005), yapmış olduğu çalışmada çeşitli örnekleme büyüklüğü, parametre sayısı, kirlenme yüzdesi ve kirlenme türlerinin kombinasyonları olarak her algoritma için 16 konfigürasyon, her konfigürasyon için 250 deneme olmak üzere toplam 4000'er deneme üzerinden aykırı gözlemleri teşhis edememe (masking) ve aykırı gözlem olmayan gözlemleri aykırı gözlem olarak teşhis etme (sweeping) gibi başarısızlıklar baz alınarak bir performans değerlendirilmesi yapmıştır.

Wisnowski, 1999 aykırı gözlemlerin tespitinde maskeleye üzerinde çalışmıştır. Maskeleye (aykırı gözlemlerin temiz gözlem gibi görünmesi), süpürme (temiz gözlemlerin aykırı gözlem gibi görünmesi) birden fazla aykırı gözlemin bulunduğu örneklemlerde meydana gelebilir ve aykırı gözlemlerin bulunduğu verilerde kontrol edilmesi güç sorunlar oluşturabileceği belirtilmiştir.

Rousseeuw, P.J. (1984), aykırı gözlemlerin teşhisinde en küçük kareler regresyonunun kalıntılarına bakmayı bir çözüm yolu olarak görmüşse de durumun her zaman böyle olmadığını belirlemiştir.

Pearson ve Sekar (1936), Thompson (1935) 'e atıfta buldukları makalelerinde, örneklemeden hesaplanan ortalama ve standart sapma değerlerinin birden fazla aykırı gözlemin varlığı altında bu gözlemlerce etkileneceğini, bu yüzden Thompson'un test istatistiğinin küçük örneklemelemlerde ancak tek bir aykırı gözlemin varlığı durumunda başarılı olabildiğini göstermiştir.

Grubbs (1950) kendisinin belirlediği bir test istatistiğini kullanmış, belirli hata düzeyleri için kritik değerlerin olduğu tablolar da düzenlemiştir. Test istatistiğini benzer şekilde, verideki en büyük iki ve en küçük iki gözlemin aykırı gözlem olup olmadığını sınıyacak şekilde uyarlamıştır, ve belirli hata düzeylerinde kritik değerlerin bulunduğu tabloları hazırlamıştır. Fakat açıktır ki, bu test istatistiği de ortalamanın bir fonksiyonu olarak testi yapılacak aykırı gözlemlerin etkisinde kalmaktadır.

Dixon (1950), normal dağılımdan gelen bir veride, aykırı gözlemlerin teşhis yöntemi olarak kullanmak üzere, varyansın bilindiği ve bilinmediği durumlar için çeşitli kriterler öne sürmüştür. Varyansın bilindiği durumda, verideki en büyük ve en küçük değerli gözlemlerin farkının standart sapma içindeki payı da aykırılık için bir değerlendirme yöntemi olabilir. Varyans parametresinin bilinmediği durumda ise verideki ilk iki gözlemin değerinin farkının verinin değişim aralığındaki payı hesaplanarak, ilk gözlemin aykırılığı konusunda bir fikir elde edilebilir.

Collett ve Lewis (1976), tek değişkenli aykırı gözlemlerin teşhisinde başka bir sorun olarak da, aykırı gözlem teşhis sürecinin öznel doğası olduğunu öne sürmüştür. İlk etapta araştırmacı tarafından görüldüğünde verinin çoğunluğuna uymadığı düşüncesini doğuran gözlemlere şüpheli gözlem, istatistiksel bir kritere uygunluk sonucu verinin çoğunluğuna uymayan gözlemlere de bozucu gözlem denmiştir. Bu şekilde bozucu gözlemin teşhisi için iki aşama vardır. Araştırmacının kişisel tercihi olarak bir gözlemi veya gözlemler kümesini şüpheli bulmak ve sonra da şüpheli gözlemlerin aykırılığını test etmektir. İkinci bir öznel karar aykırı gözlemin teşhisi için hangi yöntemin kullanılacağı ve son olarak da hipotezin hangi hata düzeyinde reddedileceğidir.

Görüldüğü gibi uzun bir dönem boyunca tek değişkenli aykırı gözlemlerin varlığı bilim adamlarının dikkatini çekmiş, bilgisayar destekli karmaşık hesaplara girilmeden, matematiksel istatistiğin araçlarıyla aykırı gözlemlerin teşhisi yoluna gidilmiştir. Tek değişkenli gözlem değerleri içinde verinin çoğunluğundan uzakta kalan gözlemler, aykırılık için bir şüphe yaratmakla beraber, bu yöntemlerin asıl amacı bir dağılım

varsayımı altında şüphe edilen gözlemlerin de aynı dağılımdan geldiği hipotezinin testine dayanmaktadır. Normallik varsayımı ve parametrelerin bilinmediği durum söz konusu olduğundan ve geliştirilen test istatistiklerinin bir çoğu örneklem ortalamasının bir fonksiyonu olduğundan, Maskeleye (Masking) ve Süpürme (Sweeping) etkileriyle karşılaşılmaktadır. Birden fazla aykırı gözlemin varlığı durumunda örneklemeden hesaplanan ortalama, aykırı gözlemlere doğru çekileceğinden, aykırı gözlemlerin ortalamaya uzaklıkları küçülmekte ve normal veri gibi gözükmektedir. Bu maskeleye etkisidir. Diğer taraftan aykırı gözlemler ortalamaı kendilerine doğru çekerken, normal gözlemlerin de ortalamaya olan uzaklıkları artmaktadır. Bu durumda normal olan gözlemlerin aykırı gözlem şeklinde teşhis edilmesi durumu ortaya çıkar. Bu da süpürme etkisidir. Maskeleye ve Süpürme birden fazla aykırı gözlemin bulunduğu örneklemelede meydana gelebilir ve aykırı gözlemlerin bulunduğu verilerde kontrol edilmesi güç sorunlar oluşturabilir (Meloun, 2001).

2.MATERYAL VE METOT

Araştırma Bingöl Üniversitesi Ziraat Fakültesi Zootekni Bölümü Kanatlı Üretim Birimi ve Ahi Evran Üniversitesi Ziraat Fakültesi Zootekni Bölümü Kanatlı Üretim Biriminde bulunan Bildircin ünitelerinde yürütülmüştür. Çalışmaya Japon Bildircinlerinde canlı ağırlık verileri yumurtadan çıkıştan itibaren 10. haftaya kadar haftada iki kez olmak üzere 100 denek hayvan üzerinde her hayvandan 20 kez ölçülmüş ve kaydedilmiştir. Çalışma 10'ar hayvanlı 5 tekerrürlü olarak yürütülmüştür. Her hayvan grubundan toplam 50 hayvanda ölçüm yapılmıştır. Araştırma, kafes sistemli bir kümeste yürütülmüştür. Araştırmada 0-10 haftalık bildircin büyütme yemi kullanılarak ve yemleme her dönemde Ad-libitum olarak yapılmıştır. Rasyonların besin madde içerikleri NRC 1994 (Nutrient Requirements of Poultry) verilerine göre hazırlanmıştır. Araştırmaya eş zaman çıkışlı 150 bildircin içinden 4.haftadan sonra cinsiyete göre seçilen 100 (50 Erkek, 50 Dişi) bildircin kullanılmış ve kuluçkadan çıkan civcivlere kanat numaraları verilmiştir. Bildircinlerin 0-10 hafta boyunca 0,01 hassasiyetli dijital terazi ile haftada iki kez olmak üzere her hayvanın tartımları yapılmıştır. Böylece elde edilen verilerle birlikte Japon Bildircinlerde canlı ağırlığa ait veriler ile aykırı gözlemler belirlenmiştir. Ölçümler yapılırken erkek, dişi ve sürü toplamı olarak değerlendirilmiştir.

2.1. Aykırı Değer Belirleme Yöntemleri:

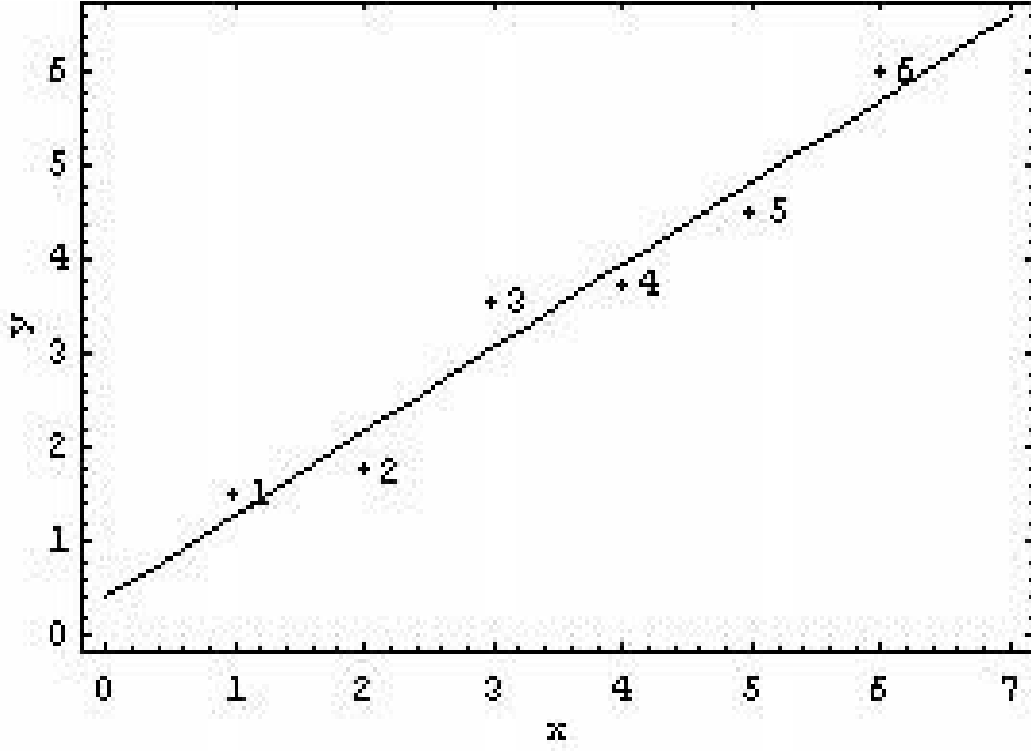
Tek değişkenli verilerde aykırı gözlem teşhis yöntemleri genel yapı ve çıkış noktası olarak modern çok değişkenli verilerde aykırı gözlem teşhis yöntemlerine temel teşkil etmekle birlikte, aralarındaki en büyük farkın da gelişen bilgisayar teknolojisiyle karmaşıklığı (Complexity) yüksek algoritmaların kullanılabilir hale gelmesi olarak söylenebilir. Matematiksel istatistik yöntemleri ile tek değişkenli bir verideki en büyük veya en küçük gözlemin değerinin, verideki diğer gözlemlerin oluşturduğu yapıdan sapıp sapmadığının testinde kullanmak üzere geliştirilen test istatistikleri bu dönemin genel çalışma sınırlarını çizmiştir. İleriki dönemlerde, bilgisayar teknolojisinin gelişimiyle, aykırı gözlem teşhis yöntemleri analitik ifadelerden çıkıp, iteratif yöntemlerin ağırlık kazandığı bir alt alan haline gelmiştir.

Bir gözlemin aykırı olup olmadığına karar vermek analizden bağımsız düşünülebilecek bir durum değildir. Şöyle ki, verinin ortalamasına uzak düşen bir gözlem

çok deęişkenli analiz için bir aykırı gözlem gibi düşünülebilirken, aynı gözlem, veriden hesaplanan regresyon hiper düzleminin yakınına düşüyor olabilir. Öte yandan analizde kullanılan uzaklık kriteri de verinin aykırılığı için farklı kararlara sebep olabilir. Örneęin, bir gözlem Öklit uzaklığına göre verinin merkezinden çok sapmış gibi görünebilir fakat verinin ortalama ve varyansını da hesaba katan Mahalanobis uzaklığına göre veri tamamen normal olabilir.

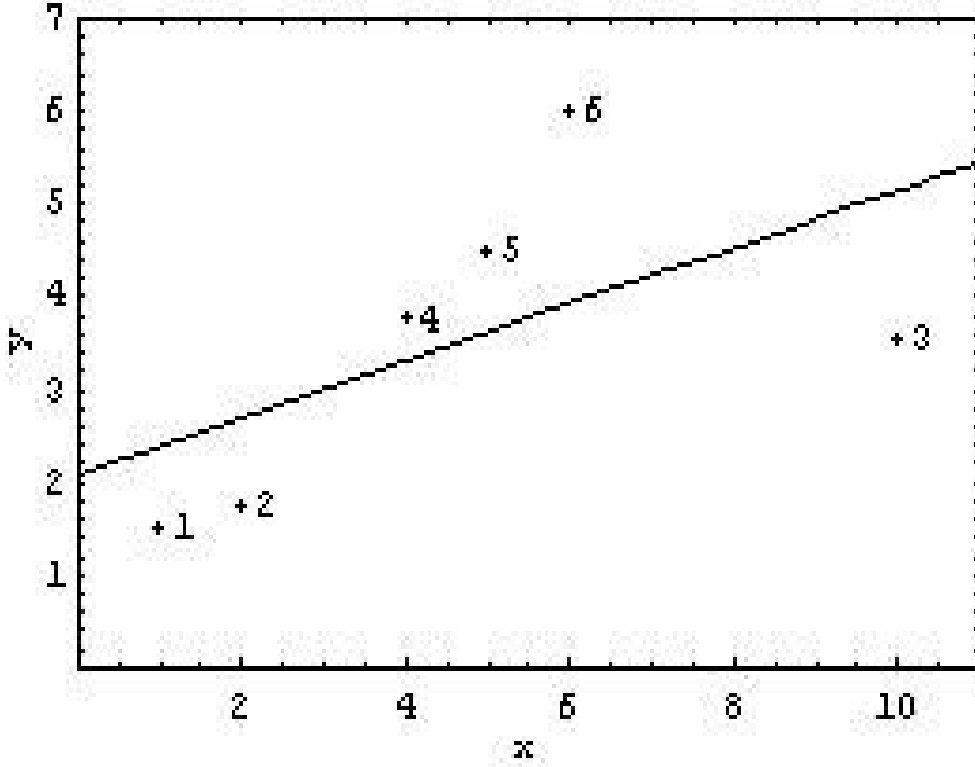
Bazı aykırı gözlem teşhis yöntemleri, gerek regresyon analizinde gerekse çok deęişkenli verilerde, tek bir aykırı gözlemi teşhis edebilecek şekilde geliştirilmişlerdir. Sonraki dönemlerde, yeni algoritmaların yazılabilme imkanı doğdukça, bir veya birden fazla aykırı gözlemi teşhis edebilecek yöntemler geliştirilebilmiştir. Daha sonra da görülebileceęi gibi tek bir aykırı gözlemin bulunabilmesi hem regresyon analizinde hem de çok deęişkenli verilerde cebirsel işlemlerle mümkün olabilmektedir. Birden fazla aykırı gözlemin teşhis yöntemleri ise gerek regresyon analizinde gerekse çok deęişkenli verilerde iteratif algoritmalara ve tüm mümkün alt kümelerin taranmasına dayanmaktadır. Öte yandan, aykırı gözlem sayısı ne olursa olsun teşhis edilebilmeleri iki boyutta çok kolayken, boyut arttıkça verinin grafiklerle gösterilmesi imkanı olmadığından aykırı gözlem teşhisi zorlaşır (Seber, 1984).

Aykırı gözlem türlerinin ne şekilde meydana geldiklerini anlamak için aşağıdaki grafikleri incelemek doğru olacaktır. Şekil 2.1’de 6 gözlemden oluşan ve aykırı gözlem barındırmayan bir veriden hesaplanan regresyon doğrusu gösterilmiştir (Satman, 2005). Şekil incelendiğinde doğru üzerinde hiç bir aykırı gözlemin olmadığı görülecektir. Eğer aykırı gözlem olsaydı regresyon doğrusuna göre uzaklaşma olacaktı. Görülen uzaklıklar aykırı gözlemlerden kaynaklanmaktadır. Görülen farklılık gözlenen deęerler ile beklenen deęerler arasındaki farklığın ifadesi olarak görülen hata miktarını göstermektedir.



Şekil 2.1. Altı gözlemden oluşan ve aykırı gözlem barındırmayan regresyon doğrusu

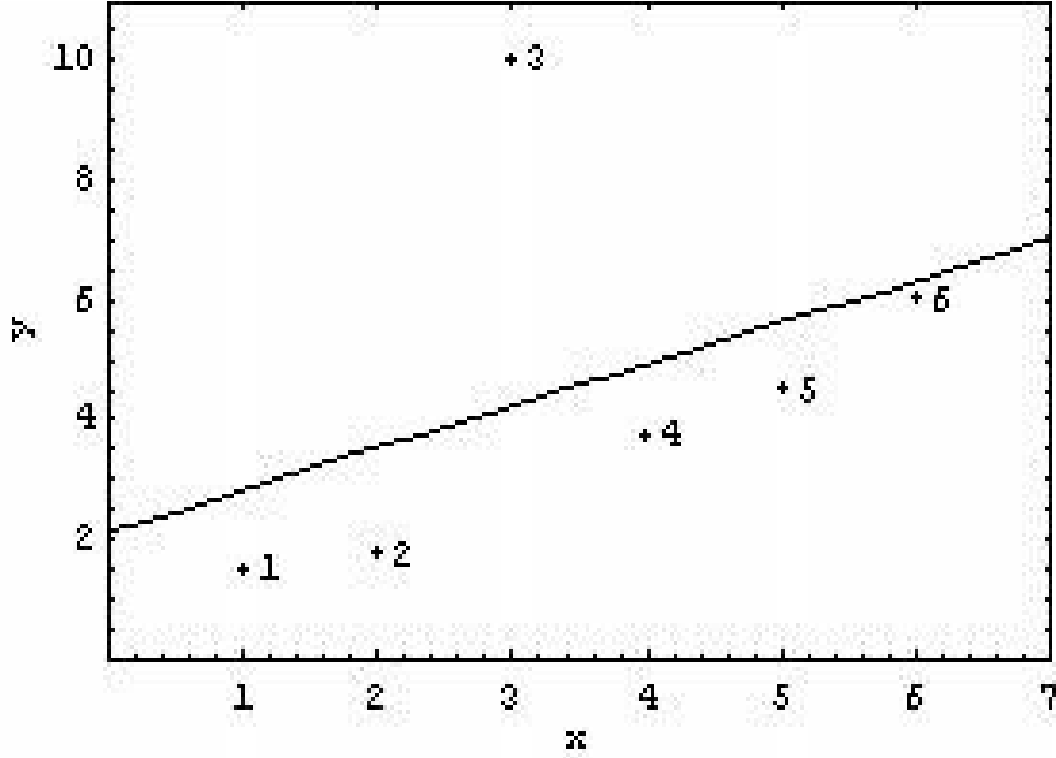
Çeşitli sebeplerle verilerdeki gözlemlerden bir veya birkaçının apsisi, diğer gözlemlerden uzağa düşmüş olabilir. Bu tür bir durum Şekil 2.2.'de gösterilmiştir ve böyle gözlemler x yönünde (x uzayında) meydana gelen aykırı gözlemlerdir. X yönündeki aykırı gözlemler, regresyon doğrusunun eğimini çok fazla değiştirebilir. Bu yüzden bu tür gözlemlere kötü kaldıraç noktası (Bad Leverage) adı verilmiştir (Penny, 1996).



Şekil 2.2. Altı gözlemden oluşan ve X yönünde aykırı gözlem barındıran regresyon doğrusu

Şekil 2.2 incelendiğinde bir önceki şekle göre aykırı gözlemler içerdiği görülmektedir. Burada aykırı gözlemimiz 6 numaralı ölçüm değeridir. Bu regresyon hattından oldukça uzaklaşmış ve artık hata olmaktan ziyade aykırı değer haline gelmiştir.

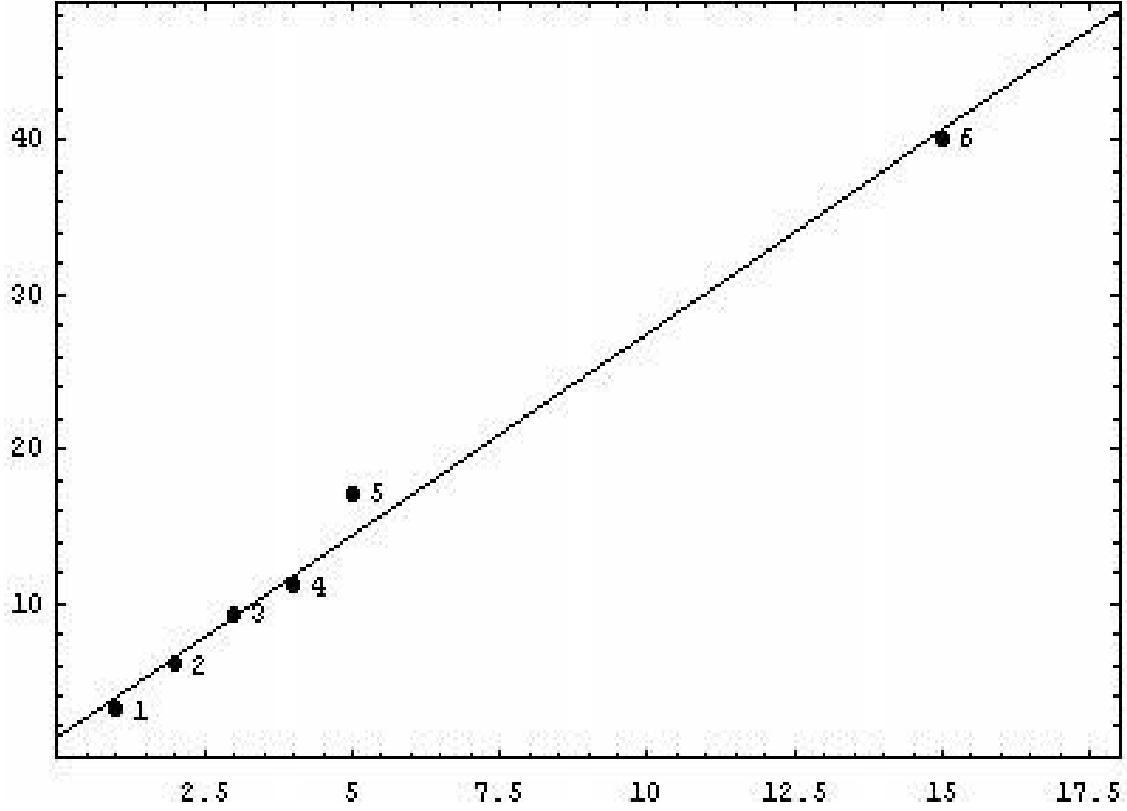
Bazen de gözlemleri oluşturan noktaların ordinatları, diğer bir deyişle açıklanan değişkene ilişkin gözlemler, diğer gözlemlerin çoğunluğunun oluşturduğu yapıdan farklılık gösterebilir. Bu tür aykırı gözlemler y yönünde (dikey) aykırı gözlemler olarak adlandırılır. Şekil 2.3'te, Şekil 2.2'deki veride (3, 3.55) olan gözlem değeri (3,10) olarak değiştirilip yeniden hesaplanan regresyon doğrusu gösterilmiştir. Yeni veriden hesaplanan regresyon denkleminin sabit parametresindeki artış y yönündeki aykırı gözlemin varlığı sebebiyle olmuştur.



Şekil 2.3. Altı gözlemden oluşan ve Y yönünde aykırı gözlem barındıran regresyon doğrusu

Şekil 2.3 incelendiğinde yine bir aykırı gözlemin olduğu görülmektedir. Ancak bu sefer aykırı değer X eksenine göre değil Y eksenine göre olmuştur. Yani aykırı değer karakterin almış olduğu değere göre değişim gösterebilmektedir.

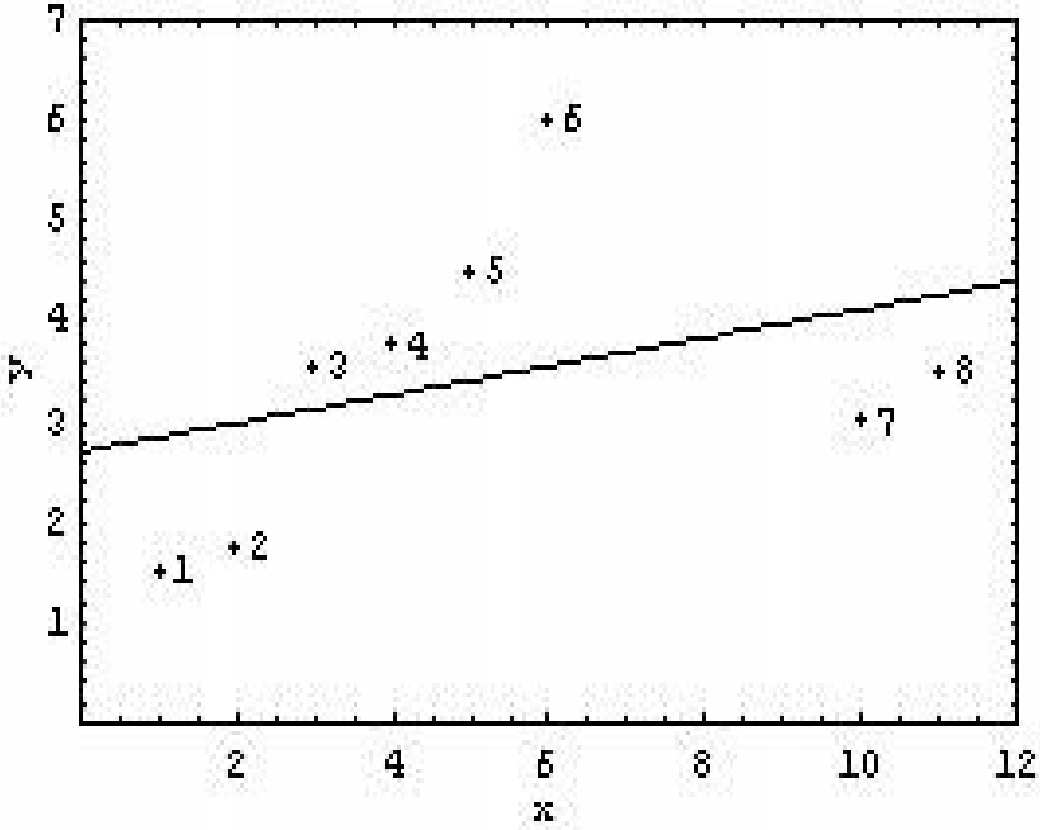
Diğer taraftan bir veya birden fazla gözlem hem açıklayıcı değişkenler hem de açıklanan değişken bakımından verinin çoğunluğunun oluşturduğu yapıya uzak düşebilirler. Bu tür gözlemler hem x hem de y yönünde aykırı gözlemlerdir. Şekil 2.4'te 6 nolu gözlem hem x hem de y yönünde aykırı gözlemdir. Fakat burada diğer türden aykırı gözlemlerden farklı bir durum söz konusudur. Altıncı gözlem veriye uzak olduğu halde, veriden hesaplanan regresyon doğrusunun yakınından geçmekle birlikte, güven bandını da daraltmaktadır. Bu yüzden böyle gözlemlere iyi kaldıraç noktaları (Good Leverage) da denmektedir (Prescott, 1975; Maronna ve Yohai, 2002).



Şekil 2.4. Altı gözlemden oluşan ve XY yönünde aykırı gözlem barındıran regresyon doğrusu

Şekil 2.4 incelendiğinde yine bir aykırı gözlemin olduğu ancak X ve Y yönlerinden ziyade XY yönüne doğru bir aykırılığın olduğu görülmektedir. Altı numaralı gözlem XY yönünde aykırılık göstermektedir.

Bazı araştırmacılar aykırı gözlemlerin teşhisinde en küçük kareler regresyonunun kalıntılarına bakmayı bir çözüm yolu olarak görmüşler de durum her zaman böyle değildir (Woodruff ve Rocke, 1994). Şekil 2.5'te 7 ve 8 numaralı gözlemler x yönünde aykırı gözlemler olarak verilmiştir. Kalıntılara bakıldığında bu iki gözlemin de kalıntıları ufak bulunacaktır. Öte yandan verinin yapısına çok iyi uyan 1 ve 6 numaralı gözlemler ise büyük kalıntı değerlerine sahip olacaklardır. Bu yüzden 7 ve 8 numaralı gözlemler birbirlerini maskeleymiştir. Bu maskeleyme etkisidir. Öte yandan 1 ve 6 numaralı gözlemler aykırı gözlem olmadığı halde büyük kalıntıya sahip olmuştur. Bu da süpürme etkisidir.



Şekil 2.5. Maskeleye ve Süpürme etkisi

2.1.1 DFBETA ve DFBETAS ölçüleri; i . gözlem, veriden çıkarıldığında hesaplamak üzere kullanılırlar. DFBETA, X açıklayıcı değişken matrisi, r kalıntı vektörü, i h şapka matrisinin i . köşegen elemanı ve i x , X matrisinin i . satırı olmak üzere $DFBETA = ((X'X)^{-1}X'r_i) / (1-h_i)$ şeklinde tanımlanmıştır. Benzer bir şekilde i . gözlemin j . parametre değerini ne kadar değiştirdiğinin ölçüsü olarak DFBETAS kullanılabilir. DFBETAS, DFBETA'dan farklı olarak i . gözlemin çıkartılmasıyla elde edilen standart hatayı da göz önüne alır. DFBETA değeri yüksek olan değer aykırılık için bir belirti olurken, gözlem sayısı büyüdükçe, gözlemlerden hesaplanan DFBETA değerleri aynı oranda küçülmektedir. Öte yandan DFBETAS değerleri, n 'le orantılı şekilde küçülürler (Belsley ve ark., 1980). Büyük DFBETAS değeri i . gözlemin j . parametre üzerindeki etkisini göstermektedir. n gözlem sayısı olmak üzere $DFBETAS_{ij} > 2$ / veya $DFBETA_i > 2/n$ olan gözlemler aykırı gözlem olarak düşünülebilir.

2.1.2 DFFIT ve DFFITS ölçüleri; i . gözlem, veriden çıkarıldığında hesaplanacak olan yeni regresyondaki tahmini y değerlerinin değişimini hesaplamak üzere kullanılırlar. DFFIT ölçütü regresyon denkleminin kurulduğu koordinat sisteminden bağımsızdır. Büyük

DFFIT değerlerine karşılık gelen gözlemler aykırı gözlem olarak düşünülebilir. i . gözlemin çıkartılmasıyla elde edilen standart hatayı da göz önüne alan form DFFITS ile gösterilir. DFBETA değeri yüksek olan değer aykırılık için bir belirti olurken, gözlem sayısı büyüdükçe, gözlemlerden hesaplanan DFBETA değerleri aynı oranda küçülmektedir. Öte yandan DFBETAS değerleri, n 'le orantılı şekilde küçülürler. Büyük DFBETAS değeri i . gözlemin j .parametre üzerindeki etkisini göstermektedir. n gözlem sayısı olmak üzere $DFBETAS_{ij} > 2$ / veya $DFBETA_{i} > 2 / n$ olan gözlemler aykırı gözlem olarak düşünülebilir.

2.1.2.1 Cook Uzaklığı: Cook uzaklığı, $\hat{\beta}$ parametre vektörünün tahmini, (i) $\hat{\beta}_i$ gözlem veriden çıkartıldıktan sonra hesaplanan parametre vektörü, \mathbf{X} açılıcı değişken matrisi, p parametre sayısı, \mathbf{R} kalıntı vektörü, $n p s - 2 = \mathbf{R}'\mathbf{R}$ olmak üzere, D istatistiği sıfır hipotezi altında p ve $n-p$ serbestlik dereceli merkezi F dağılımına uymaktadır. i D değerleri belirli bir α düzeyinde $p, n-p$ F değerini geçen gözlemler aykırı gözlem olarak kabul edilebilir.

2.1.2.2 Atkinson (1986) Algoritması: Atkinson (1986) algoritması iki aşamadan oluşmaktadır. p , parametre sayısı olmak üzere öncelikle tüm mümkün p 'li alt kümelerden geçen regresyon denklemleri hesaplanır. Oluşturulan denklemler içinde, p gözleme ait kalıntılar hesaba katılmamak üzere, kalıntı karelerinin medyanını en küçük yapan parametre tahminleri bulunur. Daha sonra standart kalıntılar içinden büyük olanlara ait gözlemler veriden çıkartılarak ikinci aşamaya geçilir. Bu haliyle algoritma, LMS algoritmasının bir türevidir. LMS'te olduğu gibi tüm alt kümelerin incelenmesi yerine sadece p 'li gruplarla hesap yapılmaktadır. İlk aşamada aykırı olduğu teşhis edilen m adet gözlem veriden çıkartılır. Üzerinde işlem yapılacak veri artık $(n-m)$ gözlem içermektedir. Gözlem sayısı $(n-m)$, \mathbf{H} şapka matrisi, i h şapka matrisinin i . köşegen elemanı, $s^2(i)$, i . gözlem olmadan hesaplanan regresyonda hata teriminin varyansının tahmincisi ve i r i . Gözlemin kalıntısı olmaktadır. Birinci aşamada m gözlemin atılmasıyla oluşturulan yeni $(n-m)$ adet gözlemden hesaplanan Cook uzaklıkları aykırılık için bir göstergedir. Sonuç olarak medyana dayalı dayanıklı bir teşhisçi olan LMS sadece bir başlangıç verisi için kullanılmış, bu sayede aykırı gözlemlerin bir kısmı ilk aşamada elenmiş, ikinci aşamada daha az aykırı gözlem içeren gözlemler üzerinden hesaplanan en küçük kareler regresyonu ile aykırı gözlem teşhisi yapılmış olmaktadır.

3. BULGULAR

Aykırı gözlemlerin teşhis edilmesi olasılıkları bir çok faktöre bağlıdır. Bunlar bilgisayar hızındaki gelişme, gözlem sayısı, kirlenme oranı, kirlenmenin tipi, ve algoritmanın parametreleridir (Woodruff and Rocke, 1993; Woodruff and Rocke, 1994). Bu çalışmada yapılan simülasyonlarda; gözlem sayıları, regresyon parametrelerinin sayıları, kirlenme oranları ve hangi yönde kirlenileceği gibi bazı kombinasyonlar literatürdeki simülasyonlarla paralellik gösterecek şekilde seçilmiştir.

Simülasyon sonuçları her bir konfigürasyon için tablo halinde sunulmuştur. Başarı ve başarısızlık durumları ise çift taraflı değerlendirilmiştir. Birinci başarı kriteri maskeleyedir ve bir algoritmanın aykırı gözlemlerin yüzde kaçını teşhis edemediğini gösterir. Bu kriterin sıfıra yakın çıkması bir başarı kriteri olacaktır. Diğer bir kriter ise süpürme etkisi olacaktır. Bu ise bir algoritmanın temiz gözlemlerin yüzde kaçını aykırı gözlem olarak yanlış teşhis ettiğinin ölçüsüdür ve sıfıra yakın çıkması bir başarı kriteri olacaktır. O halde hem maskeleye hem de süpürme oranlarını en küçük seviyede tutan algoritma simülasyona göre başarılı olacaktır.

Bazı algoritmaların boyut ve gözlem sayısı sınırından bazılarının ise amacından dolayı simülasyonları farklı konfigürasyonlarda yapılmıştır. Atkinson (1994), Marchette ve Jeffrey, (2003) gibi tamamen görsel olan ve başarı değerlendirilmesi bilgisayara bırakılamayacak algoritmalar da simülasyonlara dahil edilmemiştir. Tablolarda n gözlem sayısını, p parametre sayısını, σ_x x yönünde sapmayı, σ_y y yönünde sapmayı, c kirlenme oranını, M maskeleye oranını, S de süpürme oranını göstermektedir. Herhangi bir algoritmanın amacı dışındaki bir konfigürasyon için veya imkansız durumlar için “-“ işareti kullanılmıştır. Örneğin şapka matrisi yapısı gereği içinde bağımlı değişkeni bulundurmadiğinden Y yönünde aykırı gözlemler için bir sınaama yapmak anlamsızdır. Gentleman ve Wilk (1975) algoritması büyük örneklerde ve parametre sayısı arttığında işlemleri imkansız hale getirmektedir. Bu haliyle algoritma için sadece belirli konfigürasyonlarda simülasyon yapılmıştır.

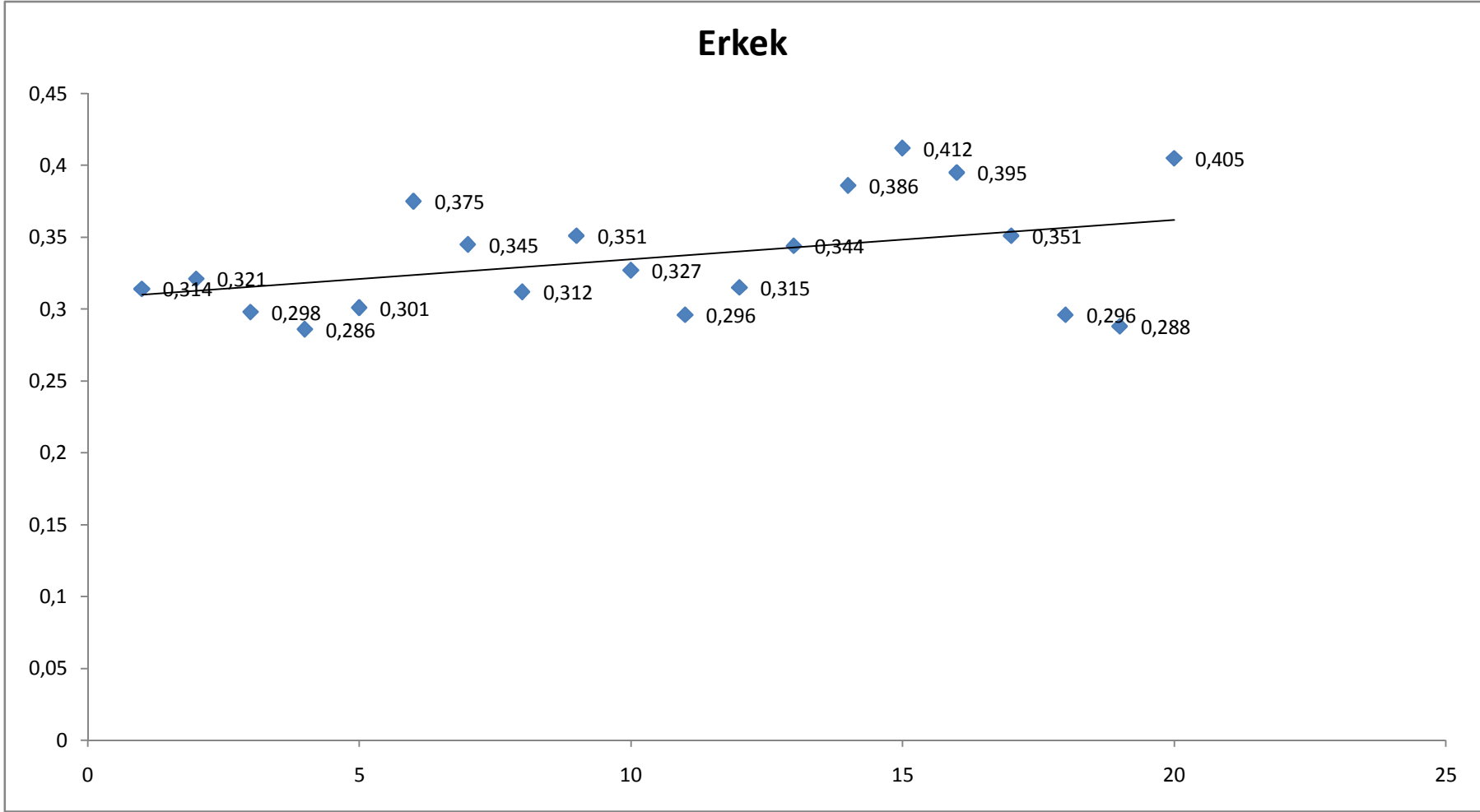
Yapılan çalışmada, erkekler, dişiler ve sürü ortalamaları olarak üç farklı değerlendirme yapılmıştır. Yapılan değerlendirmelerde de elde edilen sonuçlar buna göre üç grup altında verilmiştir. DFBETA yöntemi sonucunda elde edilen veriler Çizelge

3.1’de, DFBETAS yöntemi sonucunda elde edilen veriler Çizelge 3.2’de, DFFİTS yöntemi sonucunda elde edilen veriler Çizelge 3.3’de ve COOK yöntemi sonucunda elde edilen veriler Çizelge 3.4’de gösterilmiştir. Ayrıca, aykırı değerlere ait grafikler DFBETA yöntemi sonucunda elde edilen veriler için Şekil 3.1’de, DFBETAS yöntemi sonucunda elde edilen veriler için Şekil 3.4’de, DFFİTS yöntemi sonucunda elde edilen veriler için Şekil 3.7’de ve COOK yöntemi sonucunda elde edilen veriler ise Şekil 3.10’da gösterilmiştir.(Erkek, Dişi, Sürü toplamı)

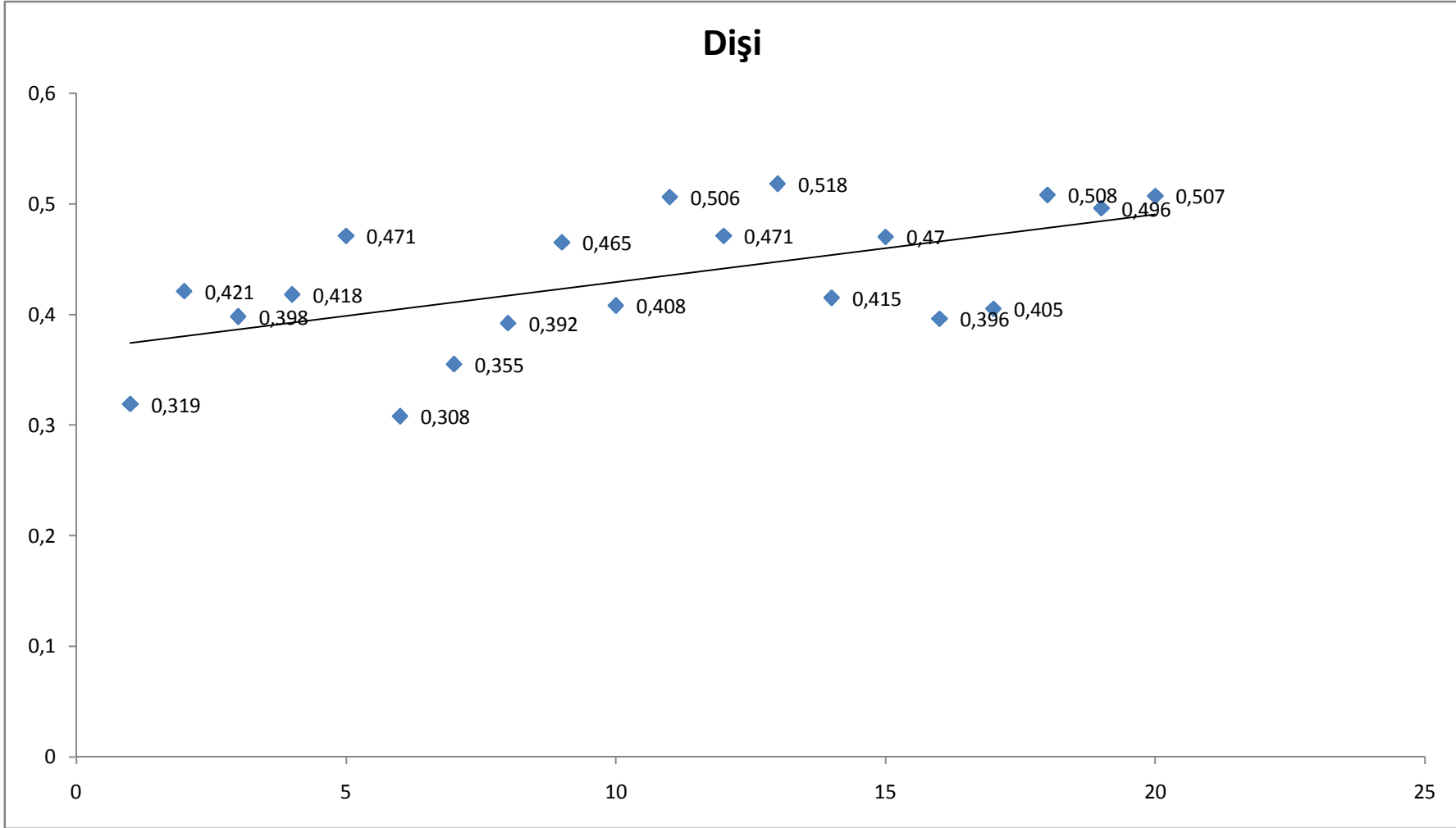
DFBETA, bazı konfigürasyonlarda maskeleye açısından başarılı gibi görünse de çok fazla temiz gözlemi aykırı gözlem olarak teşhis etmiştir. DFBETAS'ın ise y yönündeki ufak oranlarda kirlenme karşısında başarılı olduğu söylenebilir. DFFITS ve Cook ise x yönünde aykırı gözlemlere karşı ufak örneklemelerde ve düşük boyutta kısmen başarılı olabilmışlerdir. Cook uzaklığı, x yönündeki sapmalarda başarılı gibi gözükse de, temiz gözlemleri aykırı gözlem olarak teşhis etme oranı çok yüksek çıkmıştır.

Çizelge 3.1. DFBETA ile elde edilen erkek, dişi ve sürü toplamı değerleri

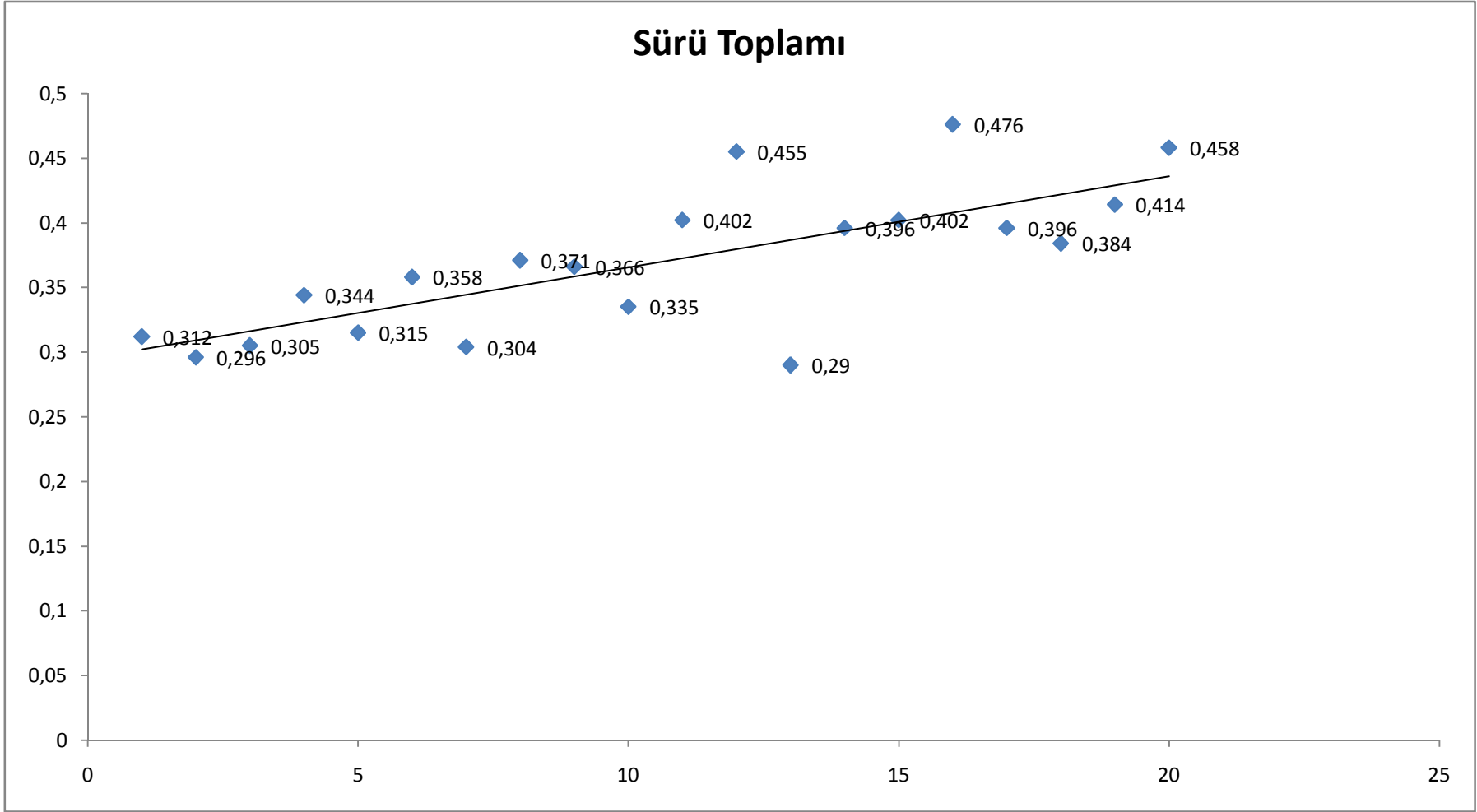
DFBETA			
Ölçüm Sırası	Erkek	Dişi	Sürü Toplamı
1	0,314	0,319	0,312
2	0,321	0,421	0,296
3	0,298	0,398	0,305
4	0,286	0,418	0,344
5	0,301	0,471	0,315
6	0,375	0,308	0,358
7	0,345	0,355	0,304
8	0,312	0,392	0,371
9	0,351	0,465	0,366
10	0,327	0,408	0,335
11	0,296	0,506	0,402
12	0,315	0,471	0,455
13	0,344	0,518	0,290
14	0,386	0,415	0,396
15	0,412	0,470	0,402
16	0,395	0,396	0,476
17	0,351	0,405	0,396
18	0,296	0,508	0,384
19	0,288	0,496	0,414
20	0,405	0,507	0,458



Şekil 3.1. Erkek Bıldırcınlarda DFBETA için elde edilen aykırı değer grafiği



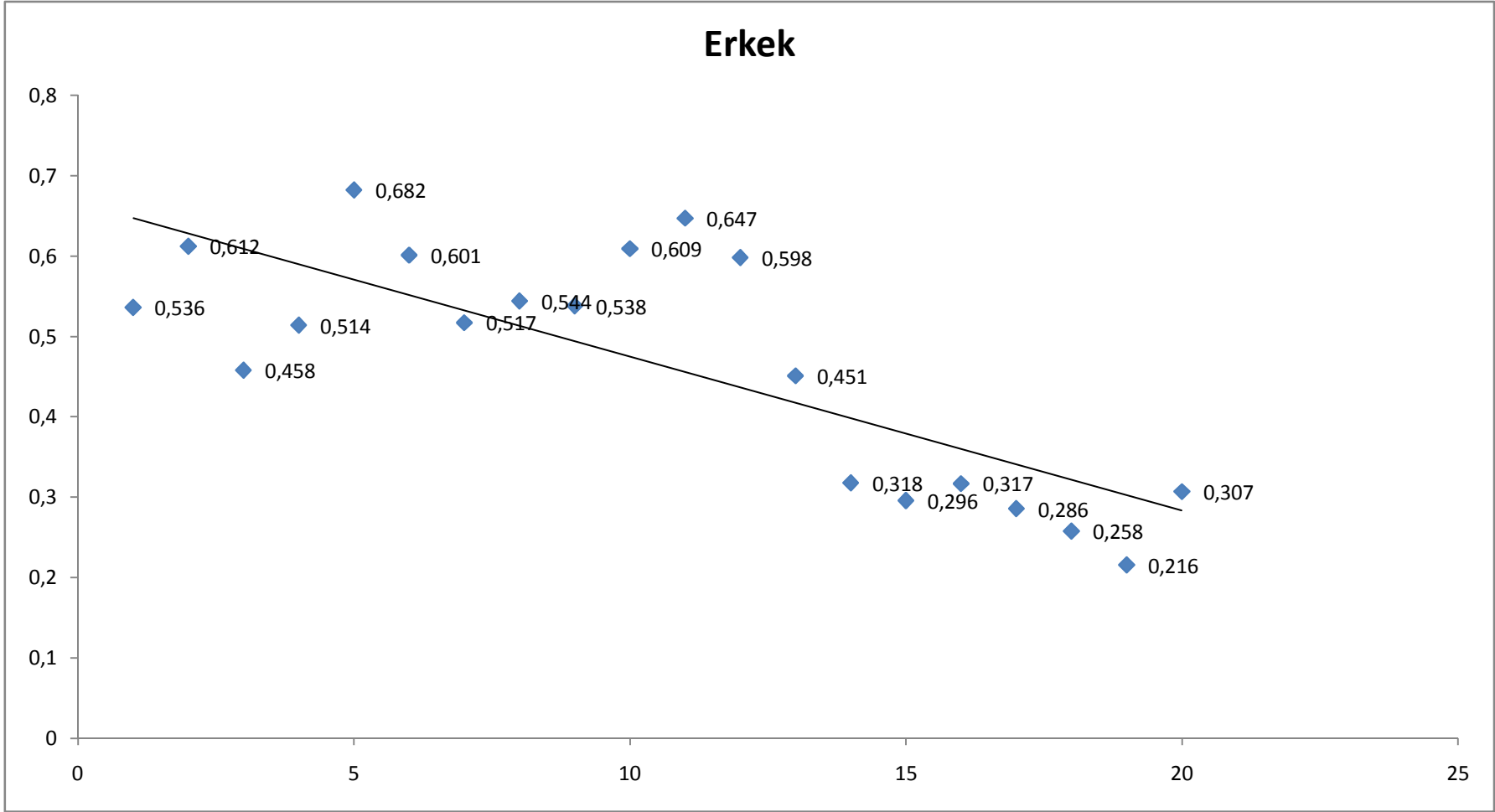
Şekil 3.2. Dişi Bıldırırcınlarda DFBETA için elde edilen aykırı değer grafiği



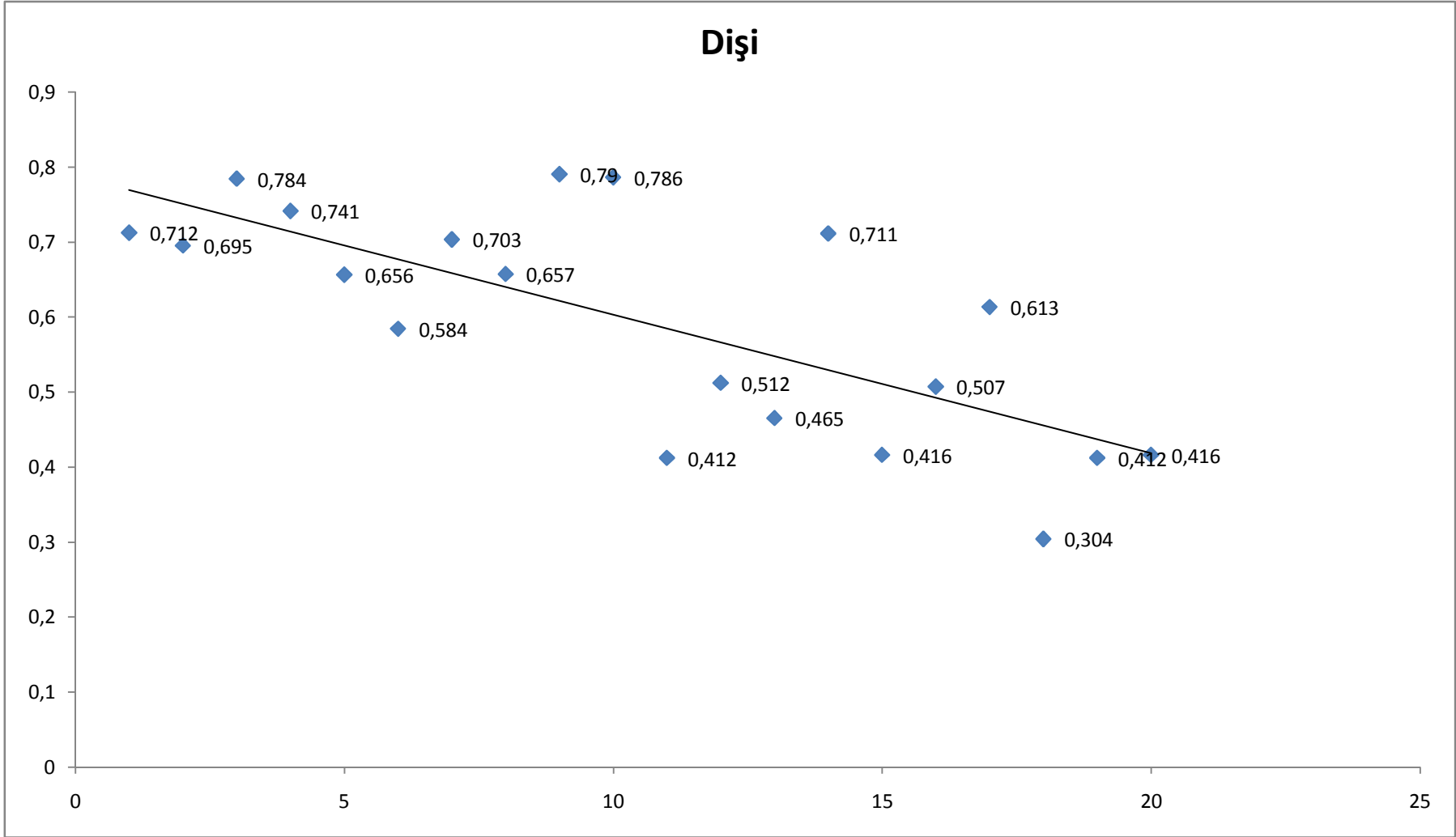
Şekil 3.3. Bıldırcınlarda Sürü toplamı DFBETA için elde edilen aykırı değer grafiği

Çizelge 3.2. DFBETAS ile elde edilen erkek, dişi ve sürü toplamı değerleri

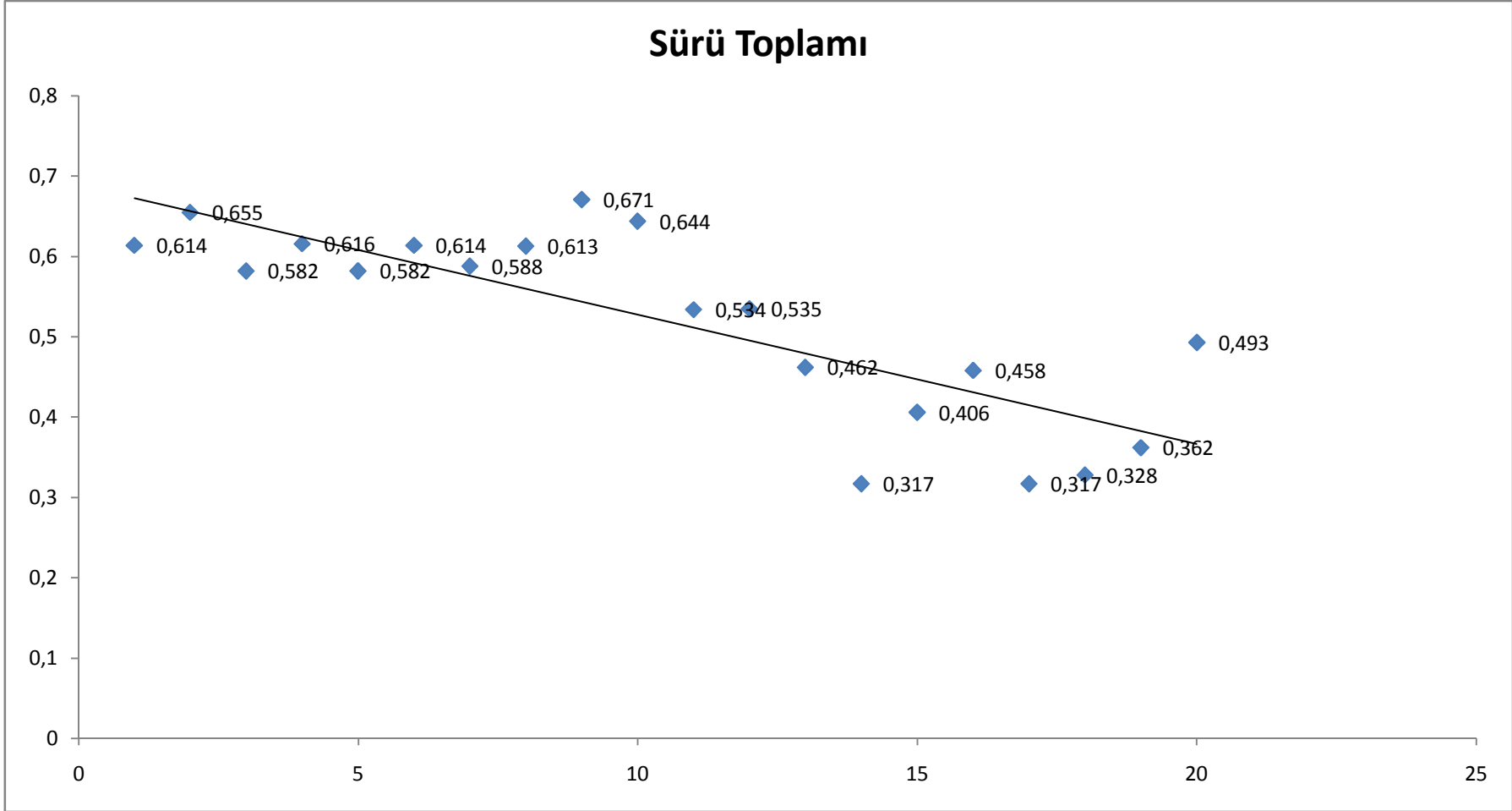
DFBETAS			
Ölçüm Sırası	Erkek	Dişi	Sürü Toplamı
1	0,536	0,712	0,614
2	0,612	0,695	0,655
3	0,458	0,784	0,582
4	0,514	0,741	0,616
5	0,682	0,656	0,582
6	0,601	0,584	0,614
7	0,517	0,703	0,588
8	0,544	0,657	0,613
9	0,538	0,790	0,671
10	0,609	0,786	0,644
11	0,647	0,412	0,534
12	0,598	0,512	0,535
13	0,451	0,465	0,462
14	0,318	0,711	0,317
15	0,296	0,416	0,406
16	0,317	0,507	0,458
17	0,286	0,613	0,317
18	0,258	0,304	0,328
19	0,216	0,412	0,362
20	0,307	0,416	0,493



Şekil 3.4. Erkek Bıldırcınlarda DFBETAS için elde edilen aykırı değer grafiği



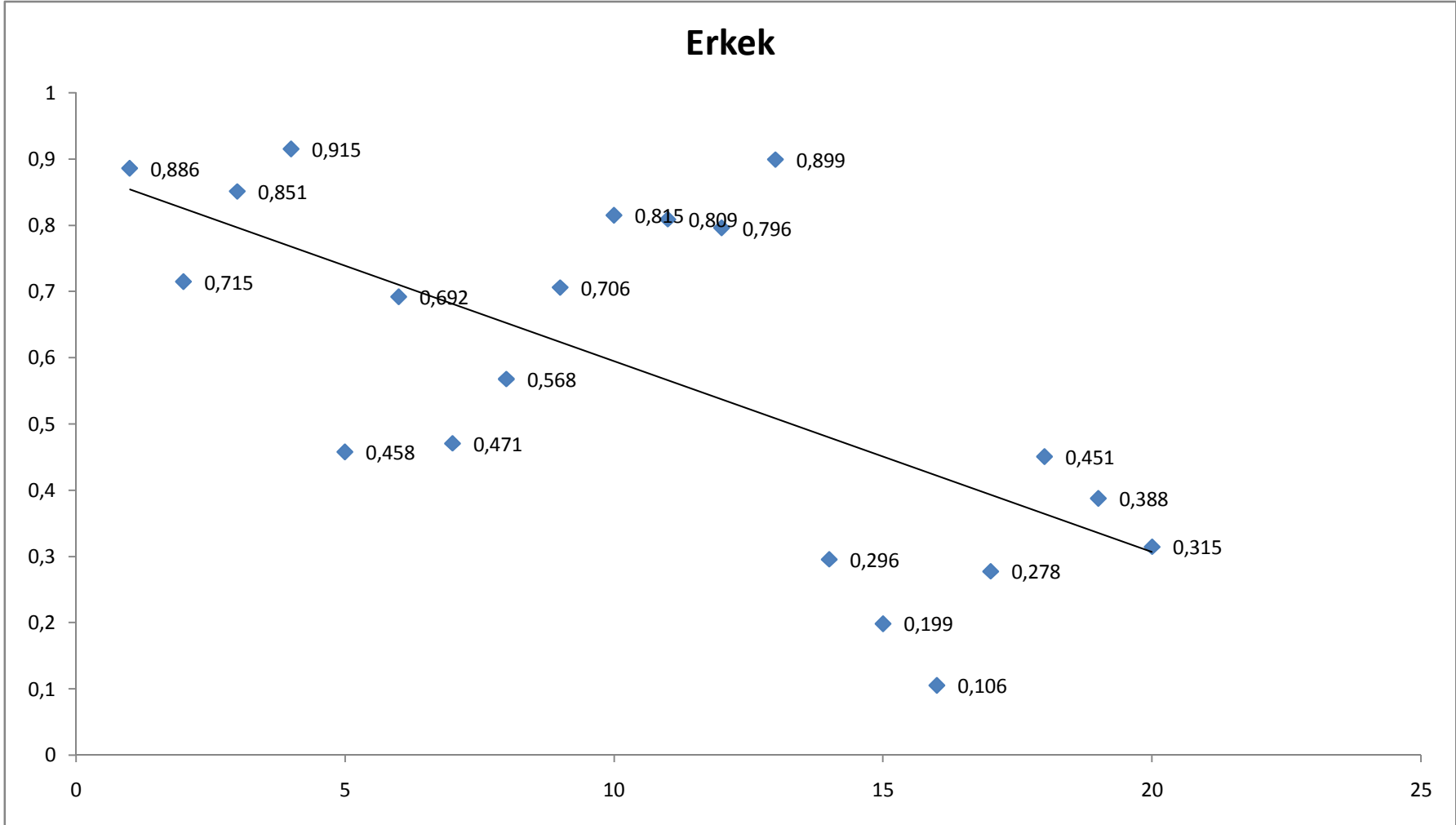
Şekil 3.5. Dişi Bıldırıcınlarda DFbetas için elde edilen aykırı değer grafiği



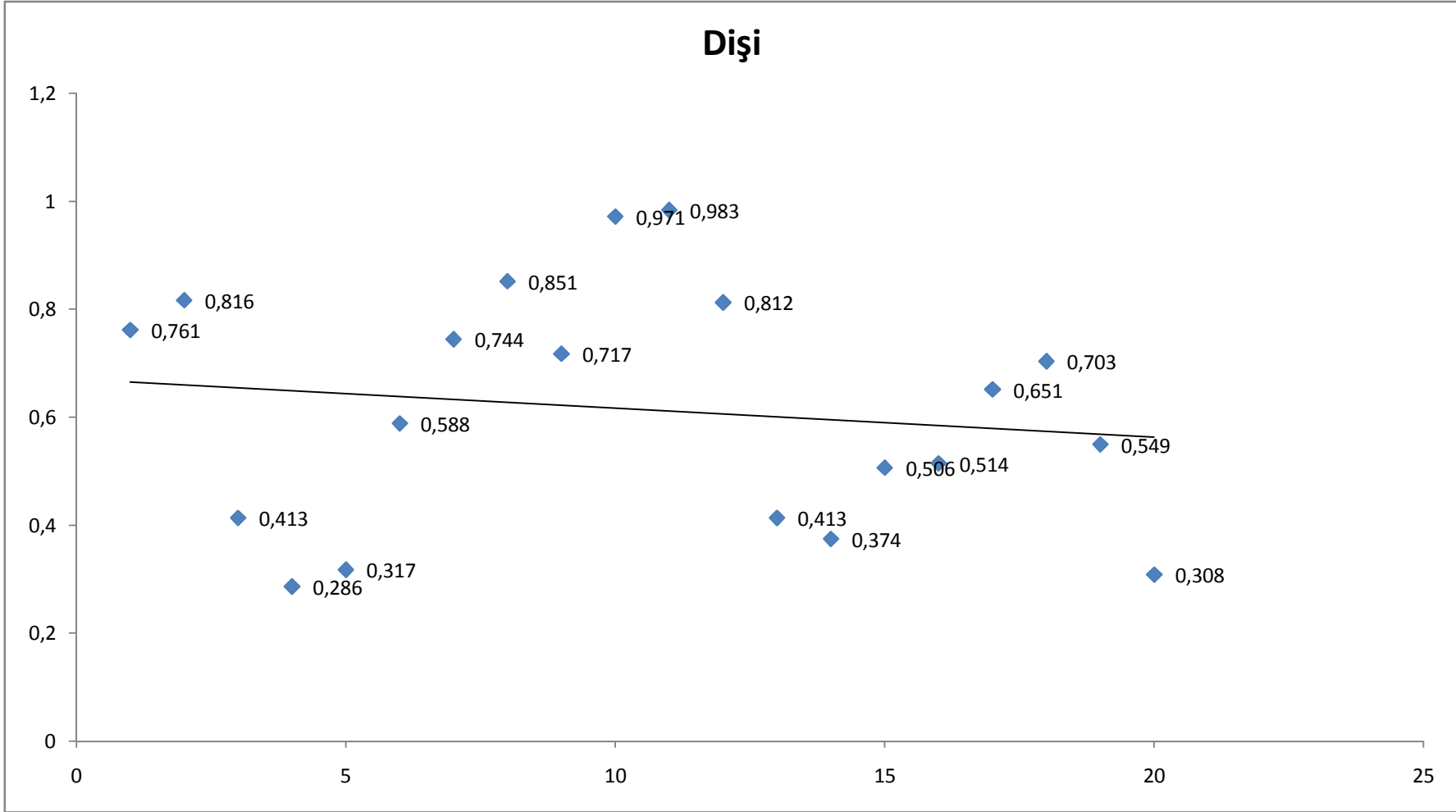
Şekil 3.6. Bıldırcınlarda Sürü toplamı içim DFBETAS ile elde edilen aykırı değer grafiği

Çizelge 3.3. DFFİTS ile elde edilen erkek, dişi ve sürü toplamı değerleri

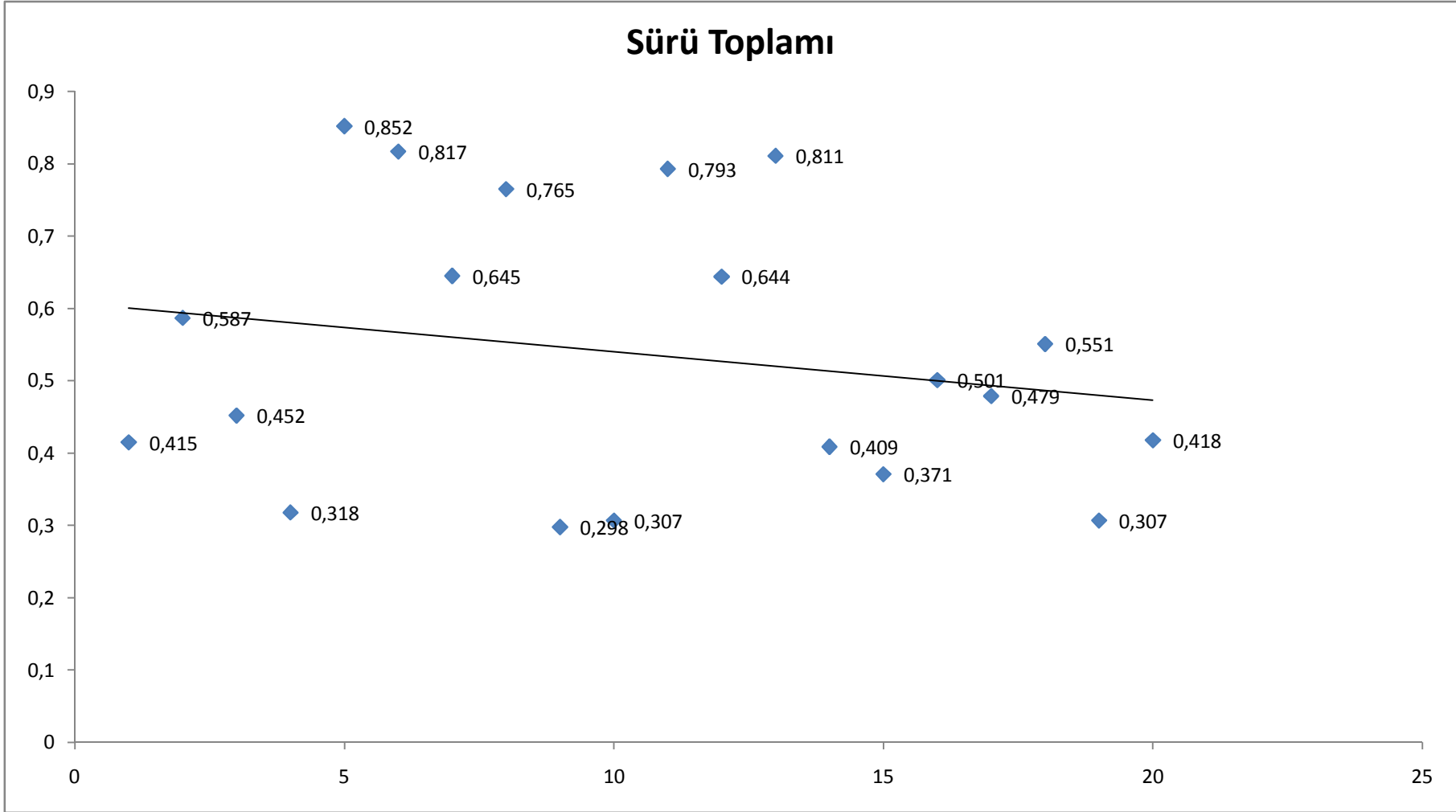
DFFİTS			
Ölçüm Sırası	Erkek	Dişi	Sürü Toplamı
1	0,886	0,761	0,415
2	0,715	0,816	0,587
3	0,851	0,413	0,452
4	0,915	0,286	0,318
5	0,458	0,317	0,852
6	0,692	0,588	0,817
7	0,471	0,744	0,645
8	0,568	0,851	0,765
9	0,706	0,717	0,298
10	0,815	0,971	0,307
11	0,809	0,983	0,793
12	0,796	0,812	0,644
13	0,899	0,413	0,811
14	0,296	0,374	0,409
15	0,199	0,506	0,371
16	0,106	0,514	0,501
17	0,278	0,651	0,479
18	0,451	0,703	0,551
19	0,388	0,549	0,307
20	0,315	0,308	0,418



Şekil 3.7. Erkek Bıldırıcılarda DFFİTS ile elde edilen aykırı değer grafiği



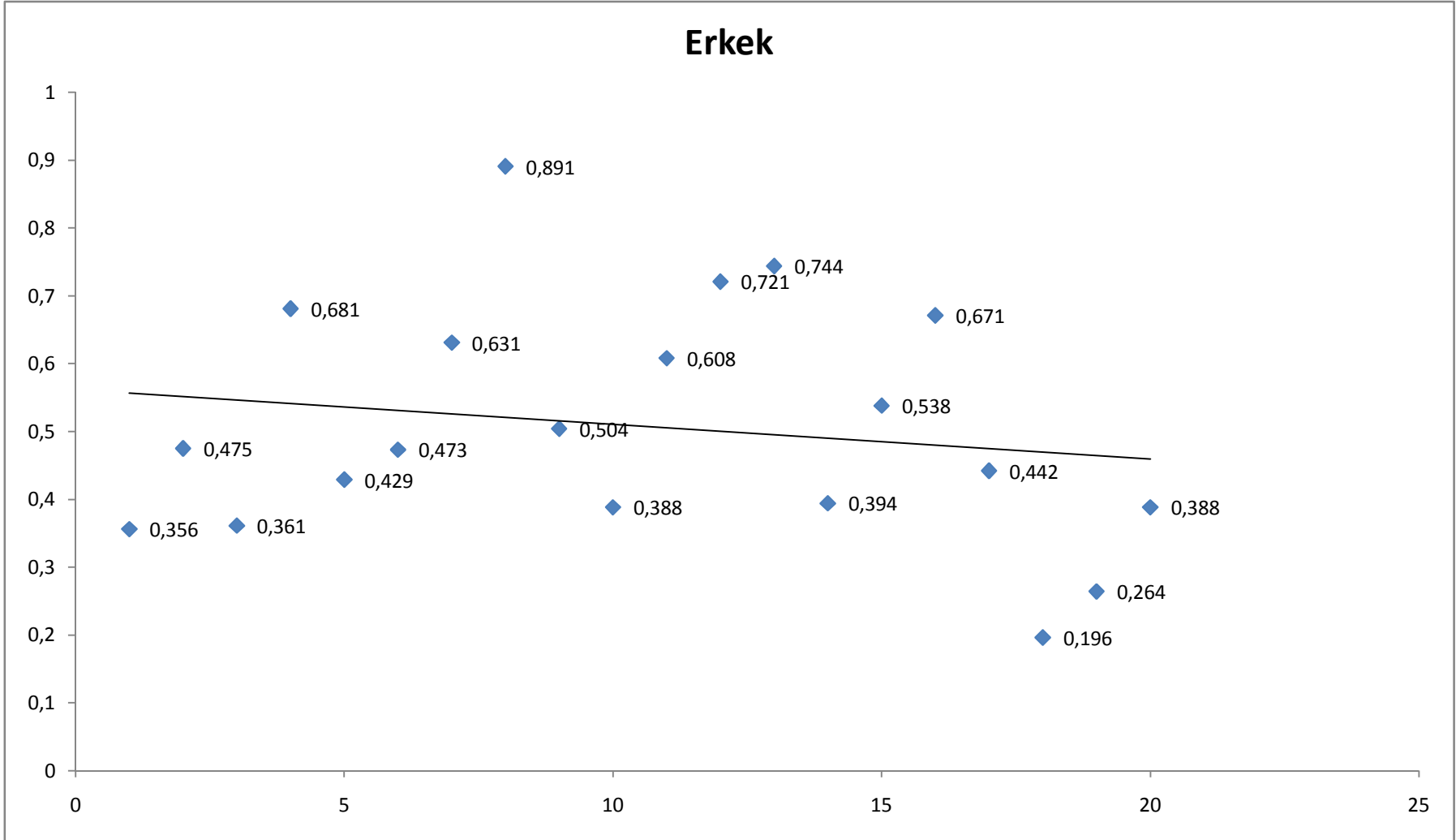
Şekil 3.8. Dişi Bıldırıcınlarda DFFİTS ile elde edilen aykırı değer grafiği



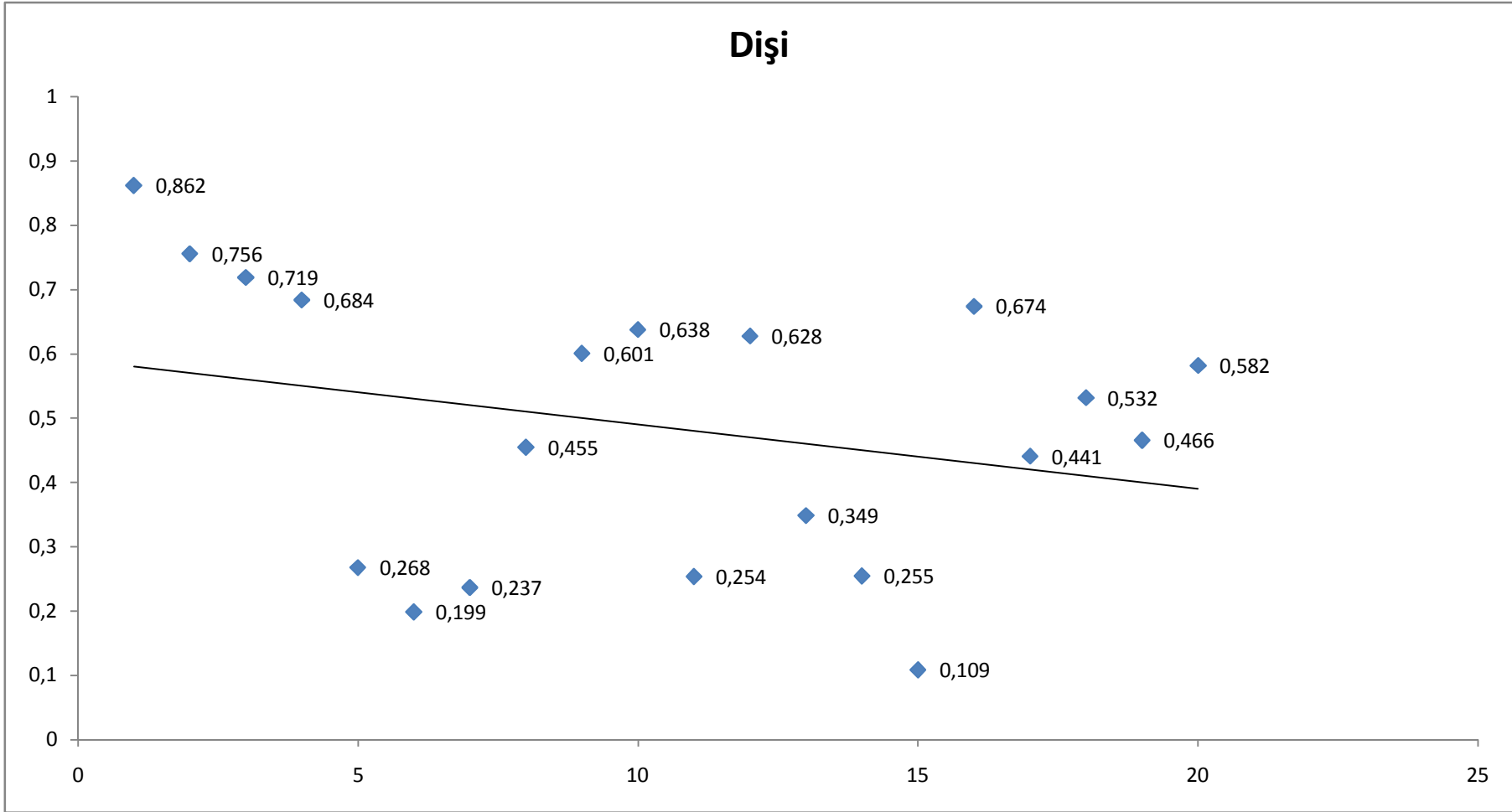
Şekil 3.9. Bildiricilerde Sürü toplamı için DFFİTS ile elde edilen aykırı değer grafiği

Çizelge 3.4. COOK uzaklığı ile elde edilen erkek, dişi ve sürü toplamı değerleri

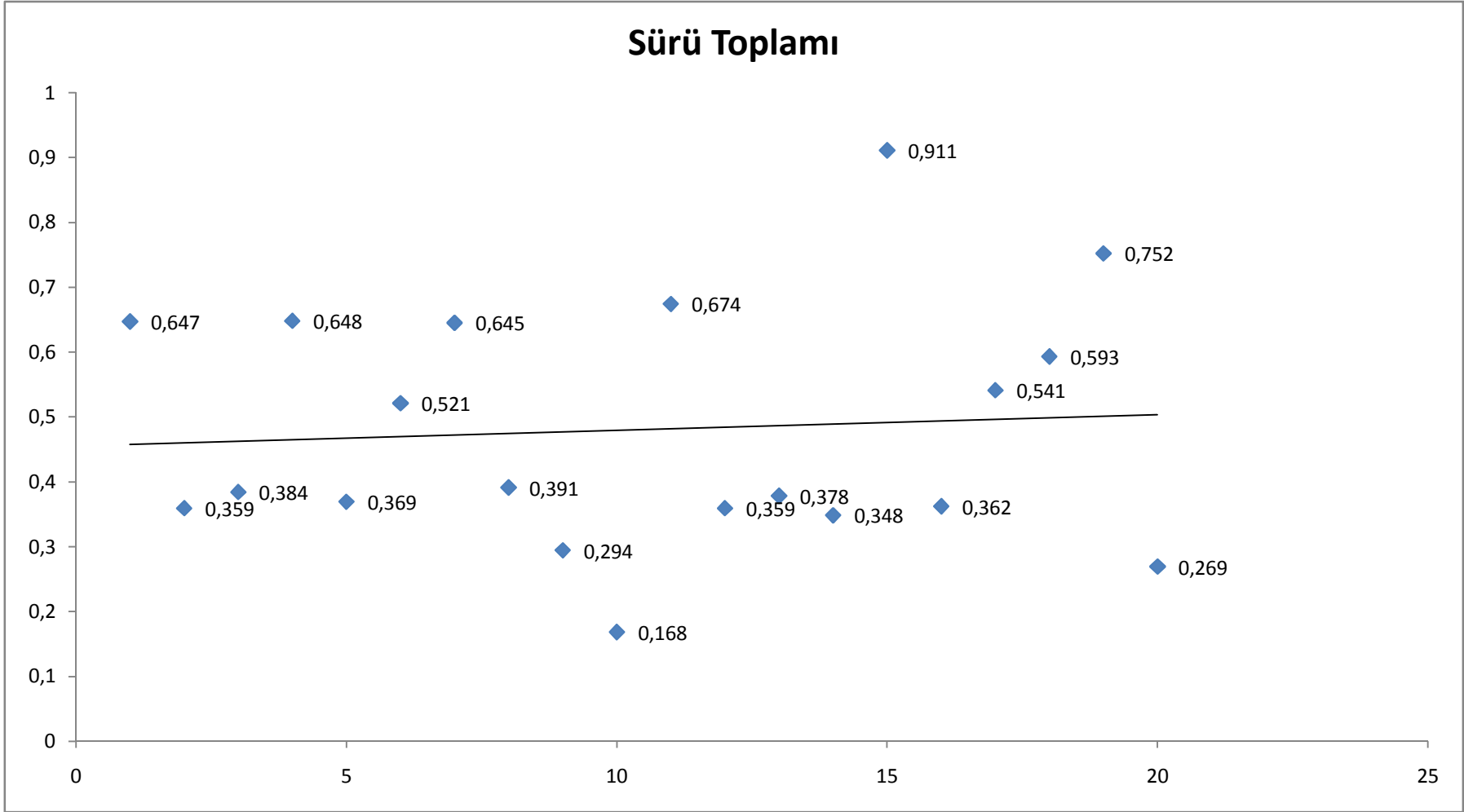
COOK			
Ölçüm Sırası	Erkek	Dişi	Sürü Toplamı
1	0,356	0,862	0,647
2	0,475	0,756	0,359
3	0,361	0,719	0,384
4	0,681	0,684	0,648
5	0,429	0,268	0,369
6	0,473	0,199	0,521
7	0,631	0,237	0,645
8	0,891	0,455	0,391
9	0,504	0,601	0,294
10	0,388	0,638	0,168
11	0,608	0,254	0,674
12	0,721	0,628	0,359
13	0,744	0,349	0,378
14	0,394	0,255	0,348
15	0,538	0,109	0,911
16	0,671	0,674	0,362
17	0,442	0,441	0,541
18	0,196	0,532	0,593
19	0,264	0,466	0,752
20	0,388	0,582	0,269



Şekil 3.10. Erkek Bıldırcınlarda COOK ile elde edilen aykırı deęer grafięi



Şekil 3.11. Dişi Bildiricilerde COOK ile elde edilen aykırı değer grafiği



Şekil 3.12. Bıldırcınlarda sürü toplamı için COOK ile elde edilen aykırı değer grafiği

Çizelgeler incelendiğinde nasıl bir değişim olduğu görülebilmektedir. Özellikle ortalama değerler bakımından incelendiğinde erkek, dişi ve sürü ortalamaları arasında değişimlerin olduğu görülmektedir. DFBETA sonucunda elde edilen değerlerin ortalaması erkek için 0,3359, dişi için 0,4323 ve sürü ortalaması için ise 0,3689 olarak belirlenmiştir. DFBETAS sonucunda elde edilen değerlerin ortalaması erkek için 0,4652, dişi için 0,5938 ve sürü ortalaması için ise 0,5195 olarak belirlenmiştir. DFFİTS sonucunda elde edilen değerlerin ortalaması erkek için 0,5807, dişi için 0,6138 ve sürü ortalaması için ise 0,5370 olarak belirlenmiştir. COOK sonucunda elde edilen değerlere bakıldığında ise ortalama değerler erkek için 0,5057, dişi için 0,4578 ve sürü ortalaması için ise 0,4692 olarak belirlenmiştir.

Sonuçlar genel olarak incelendiğinde DFFİTS değerlerinin diğerlerine göre daha yüksek çıktığı görülmektedir. En düşük çıkan ise DFBETA ile elde edilen sonuçlardır. Buna göre DFFİTS değerinin küçük olanını tercih etmek gerekeceğinden en iyi sonucun genel olarak düşük sonuçlar veren DFBETA'nın en iyi sonucu verdiği söylenebilir.

Erkekler için değerlendirme yapıldığında; DFBETA için değer 0,3359, DFBETAS için değer 0,4652, DFFİTS için değer 0,5807 ve COOK için değer ise 0,5057 olarak belirlenmiştir. Buna göre DFBETA değerleri en düşük olurken, DFFİTS değerleri ise en yüksek olmuştur.

Dişiler için değerlendirme yapıldığında DFBETA için değer 0,4323, DFBETAS için değer 0,5938, DFFİTS için değer 0,6138 ve COOK için değer ise 0,4578 olarak belirlenmiştir. Buna göre DFFİTS değerleri en yüksek olurken, DFBETA değerleri ise en düşük olmuştur.

Sürü ortalaması için değerlendirme yapıldığında DFBETA için değer 0,3689, DFBETAS için değer 0,5195, DFFİTS için değer 0,6138 ve COOK için değer ise 0,4692 olarak belirlenmiştir. Buna göre DFFİTS değerleri erkek ve dişilerde olduğu gibi yine en yüksek olurken, DFBETA değerleri ise en düşük olmuştur.

DFBETA sonucunda elde edile aykırı gözlemler şu şekilde ifade edilebilir: Erkekler için aykırı değer tespit edilememiştir. Dişiler için 13. Gözlem değeri ve sürü ortalaması için ise 11. gözlem değeri aykırı değer olarak belirlenmiştir.

DFBETAS sonucunda elde edilen aykırı gözlemler şu şekilde ifade edilebilir; Erkekler için 5. 18. ve 19. gözlem değerleri, dişiler için 18. gözlem değeri ve sürü ortalaması için ise 14. ve 17. gözlem değeri aykırı değer olarak belirlenmiştir.

DFFİTS sonucunda elde edile aykırı gözlemler şu şekilde ifade edilebilir: Erkekler için 1. 13. 14. 15. 16. ve 17. gözlem değerleri, dişiler için 4. 10. ve 11. gözlem değerleri ve sürü ortalaması için ise 5. ve 9. gözlem değerleri aykırı değer olarak belirlenmiştir.

COOK sonucunda elde edilen aykırı gözlemler şu şekilde ifade edilebilir; Erkekler için 8. 18. ve 19. gözlem değerleri, dişiler için 1. 2. ve 15. gözlem değerleri ve sürü ortalaması için ise 10. ve 15. gözlem değeri aykırı değer olarak belirlenmiştir.

Aykırı değer bulma yönünden DFBETA yöntemi daha az bulurken DFFİTS yöntemi ise en çok aykırı değer bulan yöntem olmuştur. Bunu COOK yöntemi izlemiştir. Yapılacak çalışmanın hassasiyetine uygun olarak hangi yöntemin kullanılması gerektiğine karar verilebilir. Hassasiyeti artırmak isteniliyor ise DFBETA yöntemini kullanmamak gerekecektir. Ancak hassasiyetin artması durumunda DFFİTS ve COOK yöntemlerinden birisi kullanılabilir.

4. SONUÇLAR VE ÖNERİLER

Aykırı gözlemler uzun bir dönem araştırma konusu olmuş, önceleri tek boyutlu verilerde hangi tip gözlemlerin aykırı olabileceği üzerine düşünülmüş, daha sonra sorun çok boyutlu veriler düzeyine çıkarılmış ve hızlanan bilgisayarlar ve etkin algoritmalar ile teşhis edilmeleri daha da kolaylaşmıştır. Tek bir aykırı gözlemin teşhisi için geliştirilen yöntemler, birden fazla aykırı gözlemin olduğu verilerde, tüm mümkün alt kümeler için uygulandığında kesin sonuç verebilmektedir. Fakat Gentleman ve Wilk (1975) algoritması gibi bu şekilde çalışan kaba kuvvet tekniklerin gözlem sayısı arttıkça hesaplanması imkansız hale gelmektedir.

Öte yandan tüm mümkün alt kümelerin araştırılmasını gerektirmeyen ve daha etkin olan Billor, Hadi, Velleman (2000) Bacon ve Billor, Chatterjee, Hadi (2005) algoritmaları yapılan simülasyonlarda diğer yöntemlere göre başarılı çıkmıştır. Yine de tüm mümkün alt kümelerin taranması yoluna gidilmediğinden bazı durumlarda başarısız olma şansları olacaktır. O halde sorun, ilk etapta bir teknoloji sorunudur. Büyük örneklemelerde ve büyük boyutta (çoklu regresyonda) tüm mümkün alt kümelerin çekilip ayrı ayrı analiz edilmesi işlemi çok hızlı yapabilecek işlemcilerin tasarlanması ile aykırı gözlem sorunu, teknik bir sorun olmaktan çıkacaktır. Tabii aykırı gözlem teşhisinin teknik boyutunun çözümü yetmeyecek, bir veya birden fazla gözlemin hangi durumlarda aykırı gözlem olacağı yine de öznel bir sorun olarak kalacaktır.

Diğer taraftan, bir şekilde, aykırı gözlem olarak teşhis edilen gözlemler temiz veri üzerinden analizi başlatmak üzere silinmek için teşhis edilmemelidir. Bazen aykırı gözlemler, aykırılıklarıyla daha önemli bilgi veriyor olabilirler ve bazen tek başına verinin çoğunluğu yerine aykırı gözlemlerin kendilerinin çalışma konusu yapılması gerekebilir. Öte yandan aykırı gözlemin aykırılığına karar verirken yapılacak olan analizin türünün de dikkate alınması gereklidir. Çok değişkenli bir veride aykırı gibi görünen gözlem, regresyon analizinde aykırı olmayabilir. Bu haliyle aykırılık salt uzağa düşmek anlamına gelmemektedir.

5. KAYNAKLAR

- Akdeniz, F.** 1998. Olasılık ve İstatistik, Baki Kitapevi, Adana, 546s.
- Alpar, R.** 1997. Uygulamalı Çok Değişkenli İstatistiksel Yöntemlere Giriş I, Spor Kitapevi Ankara, 340s.
- Atkinson, A.,C.** “Masking Unmasked”, *Biometrika*, 1986, Vol. 74, No. 3, pp. 533-541
- Aydın,** 2006. Grafik Yöntemlerle Etkin Gözlemlerin Ve Aykırı Değerlerin Tespiti. 19 Mayıs Üni. Fen Bilim. Ens. Yüksek Lisans Tezi. Samsun.
- Bek, Y. ve EFE, E.** 1987. Araştırma Deneme Metotları 1, Ç.Ü. Ziraat Fakültesi Ofset ve Teksir Atölyesi, Adana, 395s.
- Belsley, D., A., Kuh, E., Welsch, R., E.** “Regression Diagnostics : Identifying Influential Data and Sources of Collinearity”, New York, John Wiley & Sons, 1980
- Billor, N., Hadi, A.,S., Velleman, P., V.** 2000. “BACON: Blocked Adaptive Computationally Efficient Outlier Nominators”, *Computational Statistics & Data Analysis*, 2000, Vol. 34, pp. 279-298.
- Cook, R., D.** 1979. “Influential Observations in Linear Regression”, *Journal of the American Statistical Association*, 1979, Vol. 74, No. 365, pp. 169- 174
- Collett, D., Lewis, T.** 1976. “The Subjective Nature of Outlier Rejection Procedures”, *Applied Statistics*, 1976, Vol.25, No. 3, pp. 228-237.
- Çil, B.** 1990., “Regresyon Analizinde Tek Bir Sapan Değerin “outlier’ın”Belirlenmesine İlişkin Metodların Mukayesesi” , Doktora Tezi, Ankara Üniversitesi, 1990.
- Dixon, W. J.,** “Analysis of Extreme Values”, *The Annals of Mathematical Statistics*, 1950, Vol. 21, No 4, pp. 488-506.
- Gentleman, J. F., Wilk, M. B.** “Detecting Outliers . II. Supplementing The Direct Analysis of Residuals”, *Biometrics*, 1975, 31, pp. 387-410.
- Grubbs, F., E.** “Sample Criteria for Testing Outlying Observations”, *The Annals of Mathematical Statistics*, 1950, Vol. 21, No.1, pp.27-58. Belsley, Kuh, Welsch, 1980. Regression diagnostic. John-Willey&Sons.
- Hadi, A., S., Simonoff, J., S.,** “Procedures for the Identification of Multiple Outliers in Linear Models”, *Journal of the American Statistical Association*, 1993, Vol. 88, No. 424, pp. 1264-1272.

- Jiang, D., Tang, C., Zhang, A.,** 2004. Cluster Analysis For gene Expression Data: A Survey. IEEE Transactions on Knowledge and Data Engineering Vol, 16, No:11 1370 November 2004
- Kaya, A.,** 1999. “Zaman Serilerinde Sapan Değerlerin Analizi Üzerine Bir Araştırma”,Doktora Tezi, Dokuz Eylül Üniversitesi, 1999
- Liu, Hancong, Shah, Sirish, Jiang, Wei,** “On-line Outlier Detection and Data Cleaning”,Computers & Chemical Engineering, 2004, In Press.
- Marchette, D., J., Solka, J., L.,** “Using Data Images for Outlier Detection”,Computational Statistics & Data Analysis, 2003, Vol. 43, pp. 541-552
- Maronna, R.A., and Yohai, V.J.** 2002. “Comment on ‘Inconsistency of Resampling Algorithms for High Breakdown Regression and a New Algorithm’ by D.M.Hawkins and D.J. Olive,” Journal of the American Statistical Association, 97,154-155
- Meloun, M. Militky, J.,** 2001. Detection of Single Influential Points in OLS Regression Model Building. Analitica Chemica Acta, April, 169-191.
- Pearson, E., S., Sekar, C. Chandra** “The Efficiency of Statistical Tools and A Criterion for the Rejection of Outlying Observations”, Biometrika, 1936, Vol. 28, No. 3/4, pp. 308-320.
- Penny, K., I.** “Appropriate Critical Values when Testing for a Single Multivariate Outlier by Using the Mahalanobis Distance”, Applied Statistics, 1996, Vol. 45, No. 1, pp. 73-81
- Prescott, P.** (1975), “An Approximate Test for Outliers in Linear Models,” Technometrics, 17, 129–132.
- Rousseeuw, P.J.** (1984), : “Least Median of Squares Regression,” Journal of the American Statistical Association, 79, 871–880.
- Rousseeuw, P.J. ve Zomeren, B.C.** 1990. Unmasking outliers and leverage points. Journal of the American Statistical Association, 85, 411, s.633-639.
- Satman, M.H.** 2005. Doğrusal Regresyonda Aykırı Gözlemlerin Teşhis Yöntemleri. İstanbul Üniversitesi Sosyal Bilimler Enstitüsü Ekonometri Anabilim Dalı Yüksek Lisans Tezi

- Seber, G.,A.,F.** “Multivariate Observations”, New York, John Wiley & Sons, 1984.
- Thode, C. H. 2001. Testing for Normality, Marcel Dekker Inc., New York, 368p.
- Thompson, William R.** “On a Criterion for the Rejection of Observations and the Distribution of the Ratio of Deviation to Sample Standart Deviation”, The Annals of Mathematical Statistics, 1935, Vol. 6, No. 4, pp. 214-219.
- Üçkardes, F.** 2006. İstatistik Testler Üzerine Bir Çalışma. K.S.Ü. Fen Bil. Ens., Ziraat Fak., Zootekni Bölümü, Yüksek Lisans Tezi, 250s.
- West, S. E.** 1999a. Handbook for Statistical Analysis of Environmental BackGround Data, EPA Company, Washington, 147p.
- Wisnowski, J., W.** “Multiple Outliers in Linear Regression : Advances in Detection Methods, Robust Estimation, and Variable Selection”, Doktora Tezi, Arizona State University, 1999
- Woodruff, D. L., and Rocke D. M.** 1993. “Heuristic Search Algorithms for the Minimum Volume Ellipsoid,” Journal of Computational and Graphical Statistics, 2, 69-95.
- Woodruff, D. L. and Rocke, D. M.** 1994. “Computable robust estimation of multivariate location and shape in high dimension using compound estimators,” Journal of the American Statistical Association, 89, 888-896.

ÖZGEÇMİŞ

1979 yılında Bingöl'de doğdum. İlk orta ve lise öğrenimimi Bingöl'de tamamladım. 1997 yılında yükseköğrenime Atatürk Üniversitesi Ziraat Fakültesi Gıda Mühendisliği Bölümünde başladım ve 2002 yılında mezun oldum. 2004 yılında Bingöl ili Gıda, Tarım ve Hayvancılık Müdürlüğünde Gıda Mühendisi olarak işe başladım. Hala bu görevde devam etmekteyim. 2010 yılında Bingöl Üniversitesi Fen Bilimleri Enstitüsü Zootekni Anabilim Dalında Yüksek Lisans eğitimine başladım.