

Doğrusal olmayan regresyon analizi ve biyoistatistikte kullanımı

Mehmet N. ORMAN, İ. Safa GÜRCAN

Ankara Üniversitesi, Veteriner Fakültesi, Biyometri Anabilim Dalı, Ankara

Özet: Doğrusal olmayan regresyon analizinin biyoistatistikte kullanımı örnekler üzerinde açıklanmıştır. Doğrusal regresyon analizinde kullanılan kısıtlamaların bu analizde de kullanılabilirliği üzerinde durulmuştur. Veri yapısının doğrusal olmadığı durumlarda kullanılan bu tahmin ve analizde iki farklı örnek üzerinde çalışılmıştır. Bunlar ineklerde bir laktasyonda elde edilen süt miktarı (kg) ve yumurtacı tavukların yumurta verimleri (%)'dir. Süt verimi için kullanılan modele ait parametreler, doğrusal olmayan ve logaritmik dönüşüm uygulayarak doğrusal regresyon analizleri ile tahmin edilmiştir. Sonuçta doğrusal olmayan iteratif tekniğin daha uygun sonuçlar verdiği belirlenmiştir. Yumurta verimi için kullanılan modelde ise parametreler üzerinde hipotezler kurulmuş ve kısıtlı ile kısıtlı olmayan analizler karşılaştırılmıştır. Kısıtlı modelin daha uygun olduğu belirlenmiştir. Veri yapısı gereği doğrusal olmayan bir yapı oluşturan sayısal değerlere ait modellerde doğrusal olmayan regresyon analizi uygulamanın gerekli olduğu sonucuna varılmıştır. Eğer parametre uzayı hakkında bir varsayımda bulunuluyorsa hipotezlerinde kurulabileceği görülmüştür.

Anahtar kelimeler: Doğrusal olmayan regresyon, laktasyon eğrisi, yumurta verimi eğrisi

Application of nonlinear regression in biostatistics

Summary: Application of nonlinear regression methods in biostatistics was studied. Two approaches were used, nonlinear regression and restricted nonlinear regression. The restrictions used for the nonlinear regression were the same as used in linear regression. Cows' lactation and layer egg production curves were the two sets of data being used. Wood model (incomplete gamma model) was used as a lactation model. Logarithmic linear (transformed) and nonlinear regression methods were two different approaches for the study of lactation curve. Mean sum of squared values were estimated 10.21 and 10.29 for nonlinear and transformed case, respectively. The coefficient of determination values were found 0.873 and 0.872 for nonlinear and transformed case, respectively. According to the results obtained, nonlinear method found to be better. For the egg production curve, two different estimation approaches were used. These approaches were restricted nonlinear and unrestricted nonlinear regression methods. Lastly, two approaches being tested. Mean sum of squared values were estimated 2593.42 and 165.69 for unrestricted and restricted case, respectively. The coefficient of determination values were found 32.40 and 95.68 for unrestricted and restricted case, respectively. As the result, restricted nonlinear method found to be better to use. It is concluded that nonlinear regression method is suitable for nonlinear data sets and restrictions can be used.

Key words: Egg production curve, lactation curve, nonlinear regression

Giriş

İstatistikte iki ya da ikiden fazla değişken arasındaki ilişkinin matematiksel yapısının hesaplanmasında regresyon analizi kullanılır. Regresyon analizinde çoğunlukla temel amaç, ele alınan değişkenlerle matematiksel model oluşturmak ve bağımlı değişkende meydana gelen değişimin ne kadarının bağımsız değişken(ler) tarafından oluşturulduğunun belirlenmesidir. Basit bir regresyon modeli eşitlik 1.1 de verildiği gibidir.

$$y_i = f_i(x, \theta) + \epsilon_i \quad 1.1$$

Bu eşitlikte;

y_i : bağımlı değişkeni,

x : k-boyutlu bağımsız değişken vektörünü,

$f_i(x, \theta)$: regresyon modelini (doğrusal ya da doğrusal olmayan),

θ : p-boyutlu bilinmeyen parametre vektörünü,

ϵ_i : deneysel (rastgele) hatayı (artık),

göstermektedir (1).

Parametre " θ "nın ve hata " ϵ "un modelde alabileceği değerler önceden bilinmemektedir. Her bir y_i

ve x değerine bağlı olarak değişen hatanın (artıkların) tahmin edilebilmesi oldukça zordur. Bu zorluğa rağmen y_i ve x ye ait veriler kullanılarak θ tahmin edilebilir. Bu durumda 1.1'deki eşitlik aşağıdaki gibi olur.

$$y_i = f_i(x, \hat{\theta}) \quad 1.2$$

" $\hat{\theta}$ " işareti ilgili özelliğe ait tahmin değerini belirtmektedir. Bu tahminlerde hata kareler toplamı (HKT)'nin en küçük olması gerekir (1).

$$\text{Min HKT}(\theta) = \sum_{i=1}^n \epsilon^2 = \sum_{i=1}^n (y_i - f_i(x, \theta))^2 \dots \dots \dots 1.3$$

Eşitlik 1.3'de " θ "nın yerine " $\hat{\theta}$ " yazılacak olursa HKT değeri en küçük olur. Bu değer ise en küçük kareler tahmin edicisidir (1).

Modellerde bağımsız değişkenler " x " ile bağımlı değişken " y " arasındaki ilişki doğrusal ya da doğrusal olmayabilir. Bu da modelde yer alan parametrelerin tahmin yöntemlerinde farklılaşmalara neden olacaktır. Doğrusal modellerde sonuç tek aşamada hesaplanır. Doğrusal olmayan modeller ise farklı iki yaklaşımla çözümleri belirler. Birinci yöntemde bir dönüşümle (logaritmik, karekök, tersini alma,.... gibi) doğrusal hale getirilir

doğrusal teknikle çözülür ve sonra ters dönüşümle sonuçlar tahmin edilir. İkinci yöntemi ise doğrusal olmayan tahmin yöntemi kullanarak iteratif yöntemlerle parametre tahminleri elde etmektir.

Doğrusal olmayan regresyon analizinde kullanılan yöntemler aşağıdaki gibi sınıflandırılabilir (9).

- "Gauss Newton" metodu ve farklı düzenlemeleri,
- "Gradient" metodu,
- Bunların bileşimlerinin oluşturduğu metotlar,
- "Türev gerektirmeyen" metotlar

Metotların tamamında iterasyonlar kullanılarak tahmin değerlerine ulaşılır. Yaklaşımları farklı olan bu metotların teorik olarak yeterlilik, gereklilik, güvenilirlik, normallik, vs. testleri yapılmış ve sonuçlarının geçerliliği kabul edilmiştir (3-6,11). Doğrusal olmayan regresyon analizinde de parametreler üzerinde hipotezler kurulabilmektedir. Böylece, doğrusal regresyon modellerindeki kısıtlamalar benzer şekilde kullanılmaktadır.

$$H_0 : \beta = a$$

$$H_1 : \beta \neq a \quad 1.4$$

Burada:

β : hipotez altındaki herhangi bir parametreyi,

a : hipotez altındaki parametrenin alabileceği herhangi bir değeri,

H_0 ve H_1 : boş ve alternatif hipotezleri, temsil etmektedir.

İneklerin herhangi bir laktasyonda süt verimleri düzenli aralıklarla kontrol edilmektedir. Bu aralık Türkiye'de genellikle 15 ya da 30 gündür. Kültür ırkı bir inek için ideal laktasyon süresinin 305 gün olduğu varsayılırsa toplam 10 ile 20 arasında kontrol verim değeri elde edilmektedir. Laktasyon eğrisi oluşturmak için yaygın olarak logaritmik dönüşümün uygulanıp doğrusal regresyon analizi kullanılmaktadır. Bunun dışında doğrusal ve doğrusal olmayan farklı modeller de kullanılmaktadır.

Bu çalışmanın amacı doğrusal olmayan regresyon analizinin biyoistatistikte farklı alanlarda kullanılabilir olduğunu göstermektir. Bu amaçla iki farklı konu belirlenmiştir. Bunlar laktasyon eğrisi ve yumurta verim eğrisidir.

Materyal ve Metot

Bu çalışmada Karaköy Tarım İşletmesi'nde yetiştirilen Jersey ırkı bir ineğe ait süt verim kayıtları ile Ankara Üniversitesi, Veteriner Fakültesi, Araştırma, Uygulama ve Eğitim Çiftliği'nden alınan yumurta verimine ait kayıtlar kullanılmıştır (8,9).

Çalışmada Marquardt metodu kullanılmıştır (7). Bu metodun uygulama şekli aşağıda verildiği gibidir.

Eşitlik 1.1'de verilen modelin birinci dereceden Taylor serisi açılımı aşağıdaki gibidir.

$$f(x, \theta) = f(x, \theta_0) + (\partial f(x, \theta) / \partial \theta_1) (\theta_1 - \theta_{10}) + (\partial f(x, \theta) / \partial \theta_2) (\theta_2 - \theta_{20}) + \dots \quad 1.5$$

Eşitlik 1.1 ve 1.5 birlikte düzenlenmesiyle aşağıdaki eşitlik elde edilir.

$$y_i - f(x, \theta) = (\partial f(x, \theta) / \partial \theta_1) (\theta_1 - \theta_{10}) + (\partial f(x, \theta) / \partial \theta_2) (\theta_2 - \theta_{20}) + \dots \quad 1.6$$

Burada

$$W_i = y_i - f(x, \theta) \text{ ve } P_i = (\partial f(x, \theta) / \partial \theta_1) (\theta_1 - \theta_{10}) \text{ ve } (\theta_1 - \theta_{10}) = [\theta] \text{ olarak belirtilirse, eşitlik 1.6'yı şu şekilde düzenlenir.}$$

$$W_i = P_{i1} [\theta]_1 + \dots + P_{ip} [\theta]_p + e_i \quad 1.7$$

Vektörel gösterimi de

$$W = P_o [\theta] + e \quad 1.8$$

Bu eşitlik kullanılarak yeni W^* ve P^* değerleri hesaplanır.

$$P^* = (P_o^*) = P_o / \sqrt{(P_o^* P_o^*)} \text{ ve}$$

$$W^* = (W_o^*) = W_o / \sqrt{(P_o)} \quad 1.9$$

Böylece θ_{ek} değeri aşağıdaki eşitlik kullanılarak hesaplanır.

$$(P^{*(r)} + I^{(r)} I) \theta_{ek}^{(r)} = W^{*(r)} \quad 1.10$$

λ sabit bir değerdir ve başlangıç değeri olarak 10 alınır (7).

Buradan bir sonraki iterasyon için

$$\theta^{(r+1)} = \theta^{(r)} + \theta_{ek}^{(r)} \quad 1.11$$

değeri hesaplanır ve bu eşitlik kullanılarak yeni hata kareler toplamı hesaplanır. Bu yeni hata kareler toplamı aşağıdaki şartı yerine getirebilecek I değeri ile hesaplanır.

$$HKT(\theta^{(r+1)}) < HKT(\theta^{(r)})$$

Herhangi bir iterasyonda ikinci sabit değer " v " yardımıyla, r 'inci iterasyondaki yeni $HKT(\theta_i)$, λ^r / v ile hesaplanır; aşağıdaki karşılaştırmalar yapılır

i) Eğer $HKT(\theta_{r+1}) \leq HKT(\theta^r)$ ise yeni $\lambda^{(r)} = \lambda^{(r-1)} / v$

ii) Eğer $HKT(\theta_{r+1}) > HKT(\theta^r)$ ise ve aynı zamanda

$$\text{Eğer } HKT(\theta_{r+1}) \leq HKT(\theta^r) \text{ ise yeni } \lambda^{(r)} = \lambda^{(r-1)}$$

iii) Eğer $HKT(\theta_{r+1}) > HKT(\theta^r)$ ise ve aynı zamanda

Eğer $HKT(\theta_{r+1}) > HKT(\theta^r)$ ise yeni $\lambda^{(r)}$ değeri aşağıdaki eşitsizlik sağlanıncaya kadar " v " değeri ile çarpılır. " v " herhangi çok küçük bir değer (10^{-5} gibi) dir (7).

$$HKT(\theta^{(r)}) > HKT(\theta^r) \quad 1.12$$

Yukarıda açıklanan şartlar sağlanınca aşağıdaki eşitsizlik kontrol edilir. Bu eşitsizlik sağlanınca iterasyon durdurulur ve o aşamadaki θ değeri tahmin değeri olarak alınır (7)

$$[|\theta_{ek}^{(r)}| / (\tau + |\theta^{(r)}|)] < \epsilon \quad 1.13$$

Burada " τ " ve " ϵ " hata paylarıdır ve 10^{-3} ve 10^{-5} olarak alınır (7).

Bu metodun uygulanması değişik araştırmacılar tarafından bilgisayar ortamına aktarılmıştır. Çalışmada SPSS paket programının doğrusal olmayan regresyon kısmı kullanılmıştır.

Verilerin analizinde kullanılan modeller ise aşağıda verilmiştir. Laktasyon eğrisi için Wood modeli olarak da bilinen Gamma modeli (eşitlik 1.14) seçilmiştir (10).

$$y = an^b \exp(-cn) \epsilon \quad 1.14$$

Bu eşitlikte:

y: kontrol günündeki toplam süt verimini (kg),

n: kontrol gününü (gün),

a,b,c: modeldeki regresyon katsayılarını (parametreler),

exp: matematiksel üs fonksiyonunu,

ε: modele ait artık değerleri (hata),

temsil etmektedir.

Eşitlik logaritmik dönüşüm sonrası eşitlik 2.2' deki yapıyı alır.

$$Y = A + b \log(n) - c n + \varepsilon \quad 1.15$$

Bu eşitlikte:

Y: Log(süt) kontrol gününe ait süt değerinin logaritmik değerini,

A: modeldeki 'a' parametresinin logaritmik değerini,

Log: 'e' tabanına göre logaritmayı, göstermektedir.

Yumurta verimi için kullanılan model eşitlik 1.16' da verilmiştir (2).

$$Y = a \exp(-b t) (1 / (1 + \exp(c + d t))) + \varepsilon \quad 1.16$$

Bu eşitlikte

Y: yumurta verimi yüzdesi,

a, b, c, d: modeldeki bilinmeyen parametrelerini,

t: 14'er günlük yumurtlama periyotlarını,

temsil etmektedir.

İkinci parametre olan "b" sıfırdan küçük bir değer almış ve bunun sıfırdan büyük bir değer alıp alamayacağı yönünde hipotez kurulmuştur.

$$H_0: b \geq 0$$

$$H_1: b < 0$$

Burada, b sıfırdan küçük değer alan parametreyi temsil etmektedir (10).

Bulgular

Süt verimi

Laktasyon modeli için kullanılan süt verimine ait analiz sonuçları Tablo 1 ve Şekil 1'de verilmiştir.

Yumurta verimi

Tavuklarda yumurta verimi yüzdeleri de doğrusal olmayan modellerle tanımlanmaktadır (2). Hipotez kurularak ve kurulmadan yapılan analiz sonuçları Tablo 2 ve Şekil 2'de verilmiştir.

Tartışma ve Sonuç

Laktasyon eğrisine ait modelde, dönüşüm uygulamadan hesaplanan hata kareler toplamı 10.21 iken, dönüşüm uygulandığında bu değer 10.29 olmuştur. Yapılan değerlendirmeler dönüşüm uygulanan ve uygulanmayan durumlarda hatanın hemen hemen aynı olduğunu göstermiştir. Bu sonuçlara benzer çalışmaya rastlanılmıştır (10).

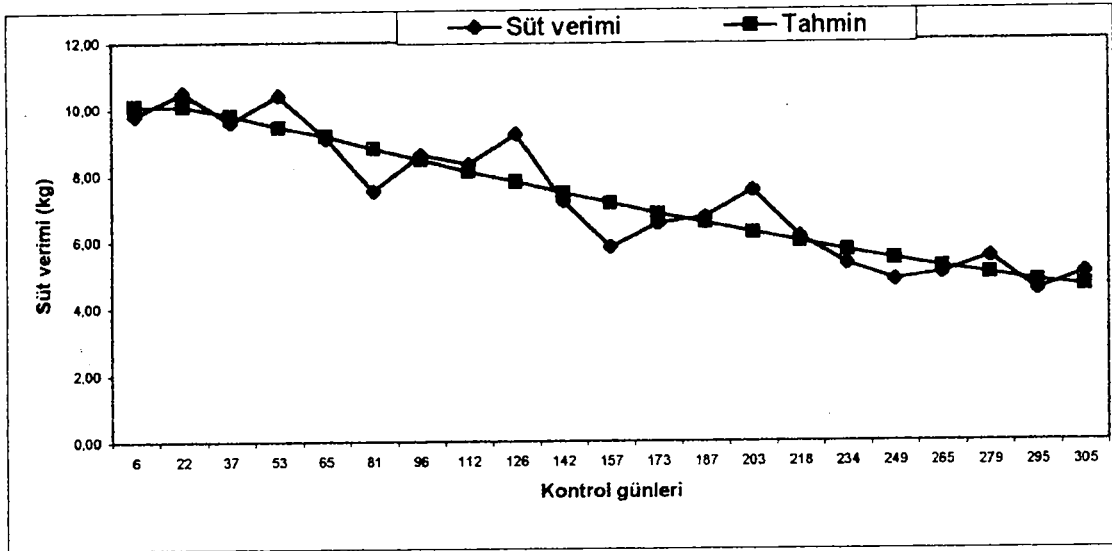
Laktasyon eğrisinde dönüşüm uygulamadan hesaplanan belirtme katsayısı değerleri 0.873 iken dönüşüm uygulandığında 0.872 olmuş ve böylece çok az bir fark oluşmuştur. Bu fark istatistik olarak önemli olmasa bile, uyumdaki %1'lik artış modelin açıklama düzeyini artırır. Hata kareler toplamı ve belirtme katsayısı sonuçları birlikte değerlendirildiğinde dönüşüm uygulamadan analiz yapmanın daha uygun olacağı söylenebilir (10).

Yumurta verimi için kullanılan modelde, hipotez kurularak yapılan analizde hata kareler toplamı 165.69

Tablo 1. Laktasyon süt verimi eğrisine ait analiz sonuçları.

Table 1. Result of nonlinear regression model of lactation curve.

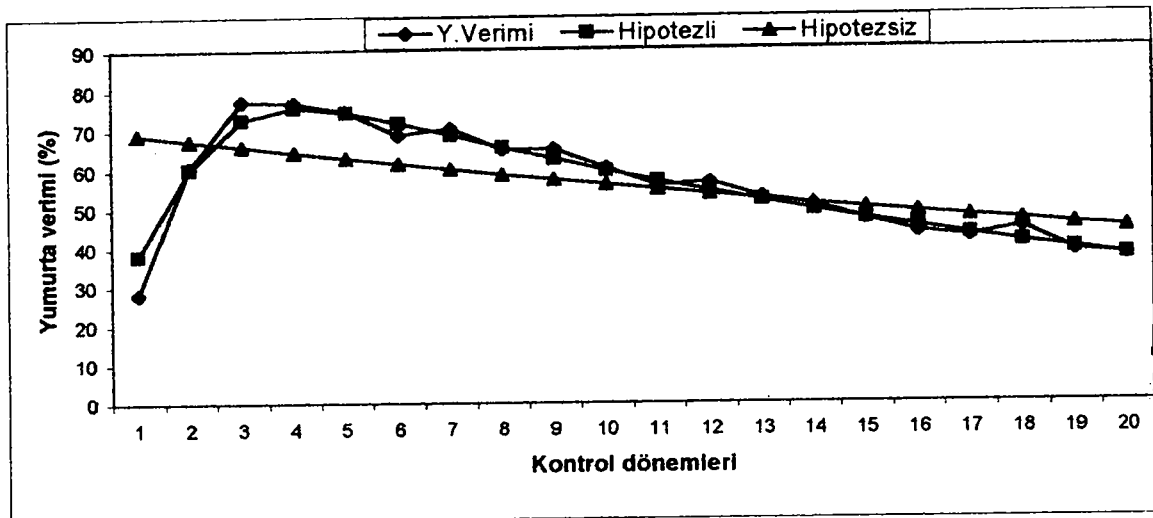
Kontrol günü	Süt verimi (kg)	Dönüşüm kullanılan		Dönüşüm kullanılmayan	
		Artıklar	Tahmin	Artıklar	Tahmin
6	9.8	-0.30	10.10	-0.22	10.02
22	10.5	0.42	10.08	0.38	10.12
37	9.6	-0.20	9.80	-0.29	9.89
53	10.4	1.25	9.45	.115	9.55
65	9.1	-0.07	9.17	-0.17	9.27
81	7.5	-1.30	8.80	-1.40	8.90
96	8.6	0.15	8.45	0.06	8.54
112	8.3	0.21	8.09	0.13	8.17
126	9.2	1.42	7.78	1.35	7.85
142	7.2	-0.23	7.43	-0.30	7.50
157	5.8	-1.32	7.12	-1.37	7.17
173	6.5	-0.30	6.80	-0.34	6.84
187	6.7	0.17	6.53	0.14	6.56
203	7.5	1.27	6.23	1.25	6.25
218	6.1	0.14	5.96	0.12	5.98
234	5.3	-0.39	5.69	-0.39	5.69
249	4.8	-0.64	5.44	-0.64	5.44
265	5.0	-0.19	5.19	-0.18	5.18
279	5.5	0.52	4.98	0.54	4.96
295	4.5	-0.25	4.75	-0.22	4.72
305	5.0	0.39	4.61	0.42	4.58



Şekil 1. Bir Jersey inçe ait laktasyon eğrisi grafiği.
Figure 1. Lactation curve of a Jersey cow.

Tablo 2. On dört günlük dönemlere ait yumurta verimi değerleri ve model tahmin değerleri.
Table 2. Model estimate values of egg production (%) at 2 weeks periods.

Dönemler (Hafta)	Yumurta verimi	Hipotezle		Hipotezsiz	
		Tahmin	Artıklar	Tahmin	Artıklar
25-26	28.04	38.07	-10.03	68.91	-40.87
27-28	60.67	60.18	0.49	67.33	-6.66
29-30	77.24	72.62	4.62	65.78	11.46
31-32	76.80	75.67	1.13	64.27	12.53
33-34	74.43	74.37	0.06	62.79	11.64
35-36	68.61	71.64	-3.03	61.34	7.27
37-38	70.17	68.55	1.62	59.93	10.24
39-40	65.05	65.46	-0.41	58.55	6.50
41-42	64.97	62.47	2.50	57.21	7.76
43-44	60.10	59.59	0.51	55.89	4.21
45-46	55.66	56.85	-1.19	54.61	1.05
47-48	56.35	54.23	2.12	53.35	3.00
49-50	52.69	51.74	0.95	52.12	0.57
51-52	50.53	49.35	1.18	50.93	-0.40
53-54	46.85	47.08	-0.23	49.76	-2.91
55-56	43.46	44.91	-1.45	48.61	-5.15
57-58	42.13	42.84	-0.71	47.49	-5.36
59-60	44.41	40.87	3.54	46.40	-1.99
61-62	38.39	38.99	-0.60	45.33	-6.94
63-64	37.19	37.19	0.00	44.29	-7.10



Şekil 2. Yumurta verimine ait grafik.
Figure 2. Egg production curve.

iken, hipotez kurulmadan yapılan analizde 2593.42 olmuştur. Uygun hipotez kurularak yapılan hesaplamalarda hata kareler toplamı değerinin azaldığı görülmüştür. İstatistik analizlerde hata payının çok az olması beklenir. Çünkü bu durum da hipotezlerin sonuçları geçerli ve güvenilir olabilir. Belirtme katsayısı "R²" hipotez kurulan modelde 95.68 olurken hipotez kurulmayan modelde 32.40 olarak hesaplanmıştır. Modelin uyumunun bir göstergesi olan belirtme katsayısı hipotezler altında %95.68'e çıkmıştır. Sonuçta, daha elverişli olan ve veri yapısını daha iyi açıklayan bir model hipotezlerle desteklenmiştir. Hipotez kurularak daha uygun analiz sonuçları elde edilebildiği gibi, uygun parametre tanım aralığı da hipotezlerle belirlenebilmektedir.

Doğrusal olmayan regresyon analizi, geçmişte uygulama zorluğu nedeniyle tercih edilmemiş, ancak günümüzde bilgisayarlar sayesinde kolayca ve kısa sürede uygulanabilmektedir. Eldeki herhangi bir veri yapısı eğer doğrusal olmayan bir durum gösteriyorsa, bunu dönüşümlerle veya basit modellerle açıklamaya çalışmak yerine doğrusal olmayan modellerle açıklamak daha uygun olmaktadır. Bu tip veri yapısına fizik, biyoloji, sağlık, hayvancılık alanlarında sıkça rastlanılmaktadır. Örneğin, büyüme eğrileri, yumurta verimleri, süt verimi, gibi. Sonuç olarak, veri yapısı seçilecek modelin tipini belirlemede önemlidir.

Kaynaklar

1. **Draper NR, Smith H** (1966): *Applied regression analysis*. John Wiley and Sons Inc, New York.

2. **Cason JA, Ware GO** (1990): *Analysis of flock egg production curves using generalized growth function*. Poultry Sci, **69**, 1064-1069.
3. **Gori, E** (1985): *Asymptotic properties of nonlinear least squares estimators when the parameters are subject to equality constraints*. Statistica, **44**, 699-709.
4. **Jenrich RI** (1969): *Asymptotic properties of non-linear least squares estimators*. Ann Math Statist, **40**, 633-643.
5. **Lutkepohl H** (1983): *Nonlinear least squares estimation under nonlinear equality constraints*. Econ Lett, **13**, 191-196.
6. **Malinvaud E** (1970): *The consistency of nonlinear regression*. Ann Math Statist, **41**, 959-969.
7. **Marquardt WD** (1963): *An algorithm for least-squares estimation of nonlinear parameters*. J Soc Indust Appl Math, **11**, 431-441.
8. **Nazlıgül A** (1992): *Yılın Değişik Dönemlerinde Kuluçkadan Çıkan Broiler Anaçların Bazı Verim Özellikleri Yönünden Karşılaştırılması*. Doktora Tezi. Ankara Üniversitesi, Sağlık Bilimleri Enstitüsü, Ankara.
9. **Orman MN** (1991): *Constrained Nonlinear Least Squares Estimation: A Milk Production Study*. Master Thesis. Middle East Technical University, Ankara.
10. **Wood PDP** (1969): *Factors affecting the shape of the lactation curve in cattle*. Anim Prod, **11**, 307-316.
11. **Wu CF** (1981): *Asymptotic theory of nonlinear least squares estimation*. Ann Statist, **9**, 501-513.

Geliş tarihi: 27.2.2001 / Kabul tarihi : 18.4.2001

Yazışma adresi:

Yrd.Doç.Dr. Mehmet N. Orman
Ankara Üniversitesi Veteriner Fakültesi
Biyometri Anabilim Dalı
06110 Dışkapı, Ankara
E-posta: orman@veterinary.ankara.edu.tr