Doğrusal Olmayan Regresyon Modelinde Bulanık Tipli Bağımlı Değişkenlerin İncelenmesi

Analyzing of Fuzzy Type Dependent Variables in Nonlinear Regression Models

Mehmet GÜRCAN,[a]
Yüksel ÖNER,[b]
Cemil ÇOLAK[b]

[a]İstatistik Bölümü, Fırat Üniversitesi Fen Edebiyat Fakültesi, ELAZIĞ
[b]İstatistik Bölümü, Ondokuz Mayıs Üniversitesi Fen Edebiyat Fakültesi, SAMSUN

Geliş Tarihi/*Received:* 27.11.2008
Kabul Tarihi/*Accepted:* 22.12.2008

Yazışma Adresi/*Correspondence:*
Dr. Mehmet GÜRCAN
Fırat Üniversitesi Fen Edebiyat Fakültesi, İstatistik Bölümü,
23100, ELAZIĞ
mgurcan@firat.edu.tr

ÖZET Amaç: İncelememizde lineer olmayan regresyon modelinde bağımlı değişkenin ölçüm değerlerinin birden fazla olduğu durumların açıklanması amaçlanmıştır. **Yöntemler:** Bu çalışma özellikle ardışık olarak tekrarlanan deneylerin sonuçlarının incelenmesinde veya bağımlı değişkenin belli bir aralıkta değer aldığı durumlarda rahatlıkla uygulanabilmektedir. Bununla birlikte kullanılan yöntem, bağımlı değişken üzerine reel değerli bir dönüşüm uygulanarak sonuçların yorumlanması şeklinde de düşünülebilir. **Bulgular:** İncelenen örnekte toplam 20 adet gözlem yapılmış ve ardışık olarak deney dört kez uygulanmıştır. Örneğe ilişkin sonuçlar söz konusu yöntem kullanılarak elde edilmiştir. **Sonuç:** Çalışmanın devamında lineer olmayan bir model göz önüne alınarak ardışık dört kez tekrarlanan bir deneyin sonuçları gözlemlenerek bağımlı değişken değerleri incelenmiştir.

Anahtar Kelimeler: Doğrusal olmayan regresyon, bulanık değişken, bağımlı değişkenin fonksiyonu

ABSTRACT Objective: In this study, we aimed to explain nonlinear regression model of which dependent variable has more than one observed measurements for an explanatory variable. **Methods:** This study is easily applicable when carried out sequentially experiments or the case in which the dependent variable is taken the value on the closed interval. Together with the used method, the interpretation of results may be thought the real valued transformation on the dependent variable. **Results:** Totally, 20 measurements were carried out in the illustrated example, and the experiment was observed sequentially repeating four times. The related results of the example were obtained by using the explained method. **Conclusion:** Consequently, considering a nonlinear regression model, the values of dependent variable are investigated by observing sequentially repeating four times of an experiment.

Key Words: Nonlinear regression, fuzzy variable, function of dependent variable

Turkiye Klinikleri J Biostat 2009;1(1):16-9

Bulanık (Fuzzy) tipli değişkenler, belli bir aralıktaki değerleri belirlenen bir olasılıkla alan değişkenlerdir. Regresyon analizinde bulanık değişkenlerinin kullanılması öncelikle tahmin edilen regresyon denkleminin veri kümesini daha iyi açıklayabilmesini sağladığı gibi ölçüm değerleri net olarak bilinmeyen fakat aralık olarak tahmin edilebilen değişken tiplerinde de iyi sonuç verdiği için tercih edilebilmektedir. Bunun yanı sıra ardışık olarak tekrarlanan deneylerde bağımsız değişken ölçüleri değişmediği halde bağımlı değişken değerleri birbirlerinden farklılık gösterebilmektedir. Bunun nedeni modele katılmayan ve modeli cüzi miktarda etkileyebilecek göz ardı edilen değişkenlerin farklılaşması olarak kabul edilebilir. Bu gibi durumlarda da bağımlı değişken bir bulanık değişkeni gibi görülüp aldığı değerler bir aralık olarak belirlenerek bu aralıktaki olasılık değerleri deneysel olarak hesaplanabilir. Bağımlı değişkenin bulanık tipli olduğu durumda uygulanan regresyon yöntemini ilk olarak Picroporola ve Tommaso incelemiştir.[1] Fakat bu inceleme iki bağımlı değişken olduğu durumda uygulanan çok değişkenli regresyon modeliyle aynıdır. Literatürde temel olarak bulanık regresyon analizi iki yolla incelenmektedir. Bunlardan birincisi lineer programlama yöntemi ikincisi ise en küçük kareler yöntemidir. Her iki incelemede de bağımlı olan bulanık değişkeninin olasılıkları aralıkların merkezlerinde en büyük değerini alacak şekilde tasarlanmıştır. Bu yöntemlerden bazıları Chang ve Lee,[2,3] Chang ve Ayyub,[4] Diamond,[5] Diamond ve Kloeden[6] ve Tanaka ve ark.nın[7] yaptıkları çalışmalarda incelenmiştir

VERİ YAPISI VE PARAMETRE TAHMİNİ

Ardışık olarak tekrarlanan deneylerde elde edilen bağımlı değişken değerleri farklılık gösterdiğinde bağımlı değişken değerlerini bir aralıkta sınırlayarak deneysel olarak elde edilen olasılıkları üyelik fonksiyonu şeklinde yorumlayabiliriz. Bu durumda bağımsız değişkenlerin vektörü $X=(x_1,\ldots,x_r)$ ve parametre vektörü $\Theta=(\theta_1,\ldots,\theta_s)$ olmak üzere model denklemi $Y=f(X,\Theta)+\varepsilon$ olup her bir, $X_i=(x_{i1},\ldots,x_{ir})$, $1\le i\le n$ için l_i tane tekrarlanan deneylerden elde edilmiş olan $y_{i1},\ldots,y_{il_i}$ gibi bağımlı değişken değerleri gözlemlenmiş olur. y_i bağımlı değişkeninin deneysel olarak hesaplanmış y_{ij}, $1\le j\le l_i$ değerini alma olasılığını μ_{ij} ile gösterelim. Bu olasılığı y_i bulanık değişkeninin üyelik değeri olarak göreceğiz. Böylelikle gözlemlerimiz bağımsız değişkenler için $x_{ij}, 1\le i\le n$, $1\le j\le r$ şeklinde ve bağımlı $f(X,\Theta)$ kenler için $(y_{ij},\mu_{ij}), 1\le i\le n, 1\le j\le l_i$ şeklinde olacaktır. Burada (y_{ij},μ_{ij}) *çifti* deneyde bağımlı değişkenin aldığı değeri ve bu değeri alma olasılığını göstermektedir.

Şimdi $f(X,\Theta)$ fonksiyonunu tahmin edilen Θ_0 başlangıç noktası civarında birinci türevleri içerecek edecek şekilde Taylor serisine açalım.

$$f(X,\Theta)=f(X,\Theta_0)+\nabla f(X,\Theta_0)(\Theta-\Theta_0)+\varepsilon$$
$$f(X,\Theta)-f(X,\Theta_0)=\nabla f(X,\Theta_0)(\Theta-\Theta_0)+\varepsilon$$

Buradan kullanacağımız istatistiği aşağıdaki şekilde hesaplayabiliriz,

$$S^2=E[f(X,\Theta)-f(X,\Theta_0)-\nabla f(X,\Theta_0)(\Theta-\Theta_0)]^2$$

Bu formül, hata karelerinin beklenen değerinin minimum olması için kullanılacaktır. Yukarıdaki Taylor açılımı sadece birinci türevleri içermektedir. Kalan kısmın hatasını minimuma indirebilmek için S^2 istatistiği kullanılabilir.

Lineer olmayan regresyon modellerinde parametre hesaplamaları için kullanılan yöntem genellikle Gauss Newton algoritmasıdır. Şimdi bu algoritmanın genelleştirilmiş bir durumu olan yöntemi kullanalım.[8] ile lineer olmayan regresyon modellerindeki ardışık hesaplama sürecini gösterelim. Bu durumda parametre tahmin değeri *k*-inci adımda,

$$T(\Theta_k)=Z_k^{-1}(X)Z_k(Y)+\Theta_k$$

eşitliği ile hesaplanabilir. Hesaplamalarda hangi adımda durulacağı, algoritma devam ederken iki ardışık hesaplama arasındaki bulunan parametre tahmin değerlerinin farkına bağlıdır. Araştırmacı bu farkı yeterince küçük gördüğünde hesaplamayı bırakabilir. Yukarıdaki formülde

$$Z(X)=[z_{mn}]\ ,\ 0\le m,n\le r$$
$$z_{mn}=\sum\nolimits_{i,j}\mu_{ij}\mu_i u_{im}^{[m]}u_{in}^{[n]}$$

ve ayrıca,

$$Z(Y)=[z_m]$$
$$z_m=\sum\nolimits_{i,j}\mu_{ij}\mu_i y_{ij}u_{im}^{[m]}$$

olarak tanımlanır. $[m]$ ifadesi $m=0$ için $[m]=0$ ve $m\neq 0$ için $[m]=1$ olarak tanımlanmıştır. Ayrıca u_{ij} değerleri $\nabla f(X,\Theta_0)$ vektörünün bileşenlerinin değerlerinden hesaplanmaktadır. Klasik uygulamalarda rahatlıkla $\mu_i=1/n$ olarak alınabilir.

TABLO 1: Ardışık olarak tekrarlanan deneye ilişkin veriler.

X1	X2	X3	Y			
0.1	0.3	0.5	1.644	1.7	1.71	1.54
1.1	1.5	1.2	1.96	2	1.888	1.988
3	2	1	2.016	1.95	1.95	1.95
3.002	2.001	1.09	2.09	2	2.67	2.5
2.21	3	3.5	4.111	3.867	2.98	4.001
5	5.6	1	3.759	3.759	3.759	3.759
0.001	3	2.125	2.446	3.001	3	3
11.54	6	2	7.79	8	7.01	7.05
5.5	15	7.5	256.27	255.01	260.001	258.31
21.89	4.87	3.008	9.022	8.99	9.09	9.09
15	7	5.04	18.81	18	18	17.95
19	7.05	7.05	32.952	31.888	33	33
2.05	12	3	22.723	22.76	22	22
2.05	24	2	7862.01	7863	7862.09	7862.09
22	12	13	538	538	538	538
25	0.44	0.1	2.24	2	2.987	2.999
35	0.55	2.46	3.74	3.743	3.74	3.743
38	8	4	50.71	50.727	50.48	50.712
37.5	10	9.05	170.88	169	168.99	169
7.05	15	15	1204	1203	1200	1209

Örnek.

İnceleyeceğimiz bu örnekte toplam 20 adet gözlem yapılmış ve ardışık olarak deney dört kez uygulanmış olup, veriler elde edilmiştir. İlgili veriler Tablo 1'de verilmiştir.

Kullanılacak olan $f(X,\Theta)$ fonksiyonu aşağıdaki formda olup,

$$f(X,\Theta) = \theta_0 + \theta_1 \exp(\theta_2 x_2)\sin x_1 + \theta_3 \exp(\theta_4 x_3)$$

bu fonksiyon yardımıyla başlangıç vektörü $\Theta_0 = (0.5\ ,1\ ,0.2\ ,0.25\ ,0.24)^{/}$ olarak hesaplanmıştır. Bunu takiben $Z_0(X)$ ve $Z_0(Y)$ matrisleri aşağıdaki gibi hesaplanmıştır,

$$Z_0(X) = \begin{bmatrix} 1 & 0.14 & 0.057 & 0.614 & 0.285 \\ 0.14 & 0.126 & 0.0169 & 0.094 & 0.046 \\ 0.057 & 0.614 & 0.0068 & 0.038 & 0.019 \\ 0.614 & 0.094 & 0.038 & 0.404 & 0.193 \\ 0.285 & 0.046 & 0.019 & 0.193 & 0.093 \end{bmatrix}$$

$$Z_0(Y) = (0.619, 0.274, 0.112, 1.163, 0.545)^{/}$$

Sonuç olarak algoritma tekrar edildiğinde parametre tahmin vektörü

$$\hat{\Theta} = (0.901\ ,0.953\ ,0.501\ ,0.987\ ,0.511)^{/}$$

olarak hesaplanacaktır.

TARTIŞMA

İncelenen model, çok değişkenli regresyon modellerine benzemekle birlikte bağımlı değişken değerlerinin aldığı değer sayısı her bir gözlemde aynı olmak zorunda değildir. Ayrıca bağımlı değişken değerlerinin aldığı değerlerin olasılıkları farklı seçildiğinde olasılığı yüksek olan bağımlı değişken değerinin parametre hesaplamalarındaki etkisi daha yüksek olacaktır. Tüm bunlarla birlikte lineer olmayan regresyon modellerinde parametre hesaplanırken Gauss Newton algoritması uygulanmaktadır. Ancak bu algoritmanın yapısında döngüsel olarak işleyen çok değişkenli bir regresyon modelinin adımsal işlemleri mevcut değildir. Bu bakımdan çalışmanın yapısı lineer olmayan regresyon modellerine de farklı bir bakış açısı getirmektedir.

Gauss Newton algoritması lineer olmayan regresyon modellerinde parametre tahmini için temel bir algoritma yönteminde olduğu gibi, lineer olmayan regresyon modellerinde tanımlanan genel bir ardışık parametre tahmin sürecinden alınarak, sürecin özelliklerine bağlı kalınmak suretiyle uygun seçilen tanımların sonucu altında elde edilmiştir.[8] Ele alınan ardışık hesaplama süreci genel bir yapıya sahip olup, araştırmanın metoduna uygun olarak yapılacak olan kısıtlamalar sayesinde çeşitli yöntemlere uyarlanabilmektedir.

Bildiğimiz kadarıyla, yapılan inceleme ile ilgili veya benzer bir çalışma tekniği literatürde gözükmediği için çalışmada yararlanılan kaynakların dışında konu ile ilgili başka bir kaynak incelenememiştir.

SONUÇ

Sonuç olarak, yapılan çalışmada bağımlı değişkenin bulanık tipli olduğu durum incelenmiş olup, bu inceleme bulanık tipli olmadığı durumda da kullanılabilmektedir. Şöyle ki, deney bir kez yapıldığında bağımlı değişken değerleri tek olacağından yöntem Gauss-Newton algoritmasına dönüşmektedir.

KAYNAKLAR

1. D'Urso P, Gastaldi T. A least-squares approach to fuzzy linear regression analysis. Comp. Stat. and Data Anal. 2000; 34: 427-440.
2. Chang PT, Lee ES. Fuzzy linear regression with spreads unrestricted in sign. Comput. Math. Appl. 1994; 28: 61-70.
3. Chang PT, and Lee ES. (b). Fuzzy least absolute deviations regression based on the ranking of fuzzy numbers. IEEE Conference on Fuzzy Systems at 1994 World Congress on Computational Intelligence. Orlando, FL, June 1994; 28: 26-29.
4. Chang YHO, Ayyub BM. Fuzzy regression methods-a comparative assessment. Fuzzy Sets and Systems 2001; 119: 187-203.
5. Diamond P. Higher level fuzzy sets arising in linear regression. Fuzzy Sets and Systems 1990; 96: 265-275.
6. Diamond P, Kloeden P. Metric Spaces of Fuzzy Sets: Theory and Applications. World Scientific, Singapore; 1994.
7. Tanaka H, Uejima KA, Assai K. Linear regression analysis with fuzzy model. IEEE, Systems, Trans. Systems Man. Cybernet 1982: SMC-12; 903-907.
8. Öner Y, Gürcan M. The basic algorithm for parameter estimation in nonlinear models. OMÜ, Fen-Edebiyat Fak. Fen Dergisi 1997:8(1); 119-25.