İstatistik Araştırma Dergisi Cilt 03 No 02 Özel Sayı Sayfa 89-95 Ağustos 2004

OLABİLİRLİK KURAMINA DAYALI REGRESYON ANALİZİ

Alper BAŞARAN* Süleyman GÜNAY*

ÖZET

Bulanık Doğrusal Regresyon literatürde çok çeşitli disiplinlerde modelleme çalışmalarında kullanılan bir yöntemdir. Bu çalışmada literatürdeki modelleme yöntemleri olan Tanaka yöntemi ve bulanık en küçük kareler yönemi ve onların çeşitli değiştirilmiş yönleri ele alınmıştır.

***Anahtar Kelimeler** : Bulanık Doğrusal Regresyon, Bulanık Küme, Bulanık Sayı.*

1. GİRİŞ

Klasik regresyonun varsayımlarının sağlanmaması durumunda elde edilen model ve bu modelden yapılan kestirimlerin çok sağlıklı olmayacağı açıktır. Bu gibi durumlarda literatürde olasılık kuramı baz alınarak geliştirilmiş bir çok yöntemle olumsuzluklar en aza indirgenmeye çalışılmıştır. Tanaka (1980) ilk olarak söz konusu varsayımların sağlanmaması durumunda olasılık kuramı değil olabilirlik (possibility) kuramına dayanarak yeni bir yöntem önermiştir. Bu yöntem klasik regresyondaki varsayımların sağlanmaması yanında eldeki veri kümesinin bağımlı değişkenle bağımsız değişken arasındaki ilişkiyi tam olarak belirlemesine yardımcı olmadığı durumlarda da kullanılmaktadır. Ayrıca uygulamada ölçümler sadece reel sayılar olmayabilir. Bunun yerine yaklaşık ifadeler ya da sözel ifadeler olabilir. Örneğin, 'yaklaşık 5 ya da orta, geniş ' gibi ifadelerle karşılaşabiliriz. Bulanık Doğrusal Regresyon (BDR) literatürde değişik disiplinlerde rastlanan ve geniş uygulama alanı elde etmiş bir yöntemdir. BDR Zadeh (1975) tarafından kuramı oluşturulan bulanık küme kuramına dayanmaktadır. BDR'ye geçmeden önce bulanık küme kuramı ile ilgili bazı tanımlar aşağıda verilmektedir.

2. BULANIK KÜME

Klasik küme kuramında bir elemanın bir kümeye ait olması karakteristik fonksiyon yardımıyla yapılır.

$$\Omega_A(a)=\begin{cases}1, a\in A\\0, a\notin A\end{cases} \tag{1}$$

Bu durum esas olarak bulanık kümenin özel bir durumudur. Zadeh (1975) tarafından ortaya konulan bulanık küme kuramı ise her elemanın kümeye belirli üyelik değerleri

* Hacettepe Üniversitesi, Fen Fakültesi, İstatistik Bölümü, Beytepe, Ankara, Türkiye

ile ait olduğunu ifade eder. Örneğin tam sayılar kümesi ele alınırsa ve tek sayılarla ilgileniliyorsa pozitif tek sayılar için karakteristik fonksiyon aşağıdaki gibi ifade edilir.

$$I_A = \begin{cases} 1, a = 2n+1 \\ 0, a = 2n \end{cases} \quad n \in N \tag{2}$$

Yukarıda ifade edilen 'yaklaşık 5 ya da orta, geniş gibi ifadeler bulanık ifadelerdir. Bu ifadeleri matematiksel olarak kullanışlı hale getirmek için üyelik fonksiyonlarından faydalanılır. Örneğin yaşa bağlı gençlik kavramını ele aldığımızda bu aşağıdaki matematiksel ifade ile gösterilir.

$$Y(x) = \begin{cases} 1, x < 40 \\ \frac{80-x}{40}, 40 \le x \le 60 \\ \frac{70-x}{20}, 60 < x \le 70 \\ 0, 70 < x \end{cases} \tag{3}$$

Bulanık küme ile ilgili bazı tanımlar aşağıdadır:

Tanım 1: Bir U kümesinin bulanık alt kümesi bu kümeden [0,1] aralığına bir fonksiyondur. $\mu_U(x): U \rightarrow [0,1]$. x burada U kümesinin genel terimini göstermektedir.

Tanım 2: Bir A bulanık kümesinin $\alpha(h)$ kesit kümesi üyeliği en az $\alpha(h)$ olan elemanların kümesidir. $[A]_h = \{a \mid \mu_A(a) \ge h\} \forall h \in [0,1]$

Tanım 3: Eğer en az bir eleman bu kümede üyelik derecesi olarak 1 değerini veriyorsa bu bulanık küme normaldir.($\exists x, \mu_A(x) = 1$).

Tanım 4: Bir bulanık alt küme konvekstir. $\forall \lambda \in [0,1]$

$$\mu_A(\lambda a_1 + (1-\lambda)a_2) \ge Min(\mu_A(a_1) \wedge \mu_A(a_2))$$

Yukarıda verilen genel U kümesi yerine reel sayılar $\Re$ kümesi kullanılırsa ve bu kümenin bulanık alt kümelerinde işlemler yapilirsa o zaman bulanık kümenin özel bir türü olan bulanık sayılara ulaşılır. Dubois ve Prade (1980) bulanık sayıların detaylı çalışmasını yapmışlardır. Bu sayılar genel biçimiyle L-R sayıları olarak ifade edilir ve aşağıdaki gibi gösterilir. M bulanık alt küme olsun.

$$\mu_M(x) = \begin{cases} L(\frac{m-x}{\alpha}), x \le m \\ R(\frac{x-m}{\beta}), x \ge m \end{cases} \tag{4}$$

Burada L bulanık sayısının sol taraf biçimini R de sağ taraf biçimini göstermektedir. L-R sayılarının özel bir durumu üçgensel bulanık sayılarıdır. Üçgensel bulanık sayılar simetrik ya da simetrik olmayan olmak üzere ikiye ayrılırlar. Ayrıca yamuk(trapezoidal) bulanık sayıları BDR da kullanılmaktadırlar. Bulanık üçgensel simetrik sayılar parametrik olarak aşağıdaki gibi ifade edilir.

$$A_i = (\alpha_i, c_i),\ i = 1.....m$$
$$Y_i = (y_i, e_i)\ ,\ i = 1......m$$
$$\hat{Y}_i = (\hat{y}_i, \hat{e}_i)\ ,\ i = 1......m$$

Eldeki verileri (Y_i, x_i) şeklinde gösteriyoruz.

$$A_i = [\alpha_i - c_i, \alpha_i + c_i], i = 1.....m \text{ ve } Y_i = [\hat{y}_i - e_i, \hat{y}_i + e_i], i = 1...m$$

Burada α_i' ler çekirdek değerler c_i' ler yayılım değerleridir. $\hat{y}_i$ ler ve e_i ler de sırasıyla çekirdek ve yayılım değerlerini göstermektedirler.

3. BULANIK DOĞRUSAL REGRESYON

3.1 Tanaka Yöntemi

BDR bulanık küme kuramı baz alınarak Tanaka (1980) tarafından önerilmiştir. BDR çözüm yöntemi olarak iki ana yönde ilerlemektedir. Birincisi doğrusal programlama yöntemine dayanılarak regresyon denklemindeki parametrelerin tahminidir (Tanaka, 1980). Diğeri ise Diamond (1980) tarafından önerilen bulanık en küçük kareler yöntemidir. Birinci yöntem araştırmacılar tarafından eleştiriler almış ve zaman içinde geliştirilmiştir. Bu yöntemin sakıncaları uç değerlere karşı aşırı hassastır. Ne kadar çok gözlem işleme dahil edilirse tahmin edilen aralık genişler buda iyi tahmin yapmayı zorlaştırır. İkinci yöntem ise klasik en küçük kareler yönteminin bir uzantısı olup daha iyi sonuçlar vermektedir. BDR geçmeden önce literatürde karşılaşılan veri türleri aşağıdaki tabloda gösterilir.

Tablo 1. Veri türleri

Y_i	X_i
Reel	Reel
Bulanık	Reel
Bulanık	Bulanık

Bir BDR aşağıdaki şekilde ifade edilir:

$$\hat{Y}_i = A_0 + A_1 X_{i1} + A_2 X_{i2} + ... + A_n X_{in} \quad (5)$$

Bu denkelemde (Y_i, X_i) bağımlı ve bağımsız değişkenler olmak üzere eldeki veri kümesini göstermektedir. A_i ler ise bulanık sayıları göstermektedirler. BDR formülasyonunda tahmin edilen değerler ile gözlemlenen değerler arasındaki karşılıklı ilşkiye göre üç farklı formülasyon geliştirilmiştir. Bu formülasyonlara geçmeden önce kısaca bu ilişkiler fonksiyonel olarak şöyle yazılabilir.

$$Y_i \subseteq_h \hat{Y}_i = A_1 X_{i1} + ... + A_n X_{in} \qquad (6)$$

$$Y_i \supseteq_h \hat{Y}_i = A_1 X_{i1} + ... + A_n X_{in} \qquad (7)$$

$$[Y_i]_h \cap [\hat{Y}_i = A_i X_{i1} + ... + A_n X_{in}]_h \neq \varnothing \qquad (8)$$

Birinci denklem gözlemlerin bulanık sayı tahminleri tarafından kapsandığını, ikinci denklem gözlemlerin bulanık sayıları kapsadığını ve son denklemde bulanık sayı tahminleri ile gözlemlerin kesişiminin boş küme olamadığını ifade eder. İlk yöntem olarak ifade edilen ve doğrusal programlama yöntemi ile parametrelerin tahmininin yapılmasında kullanılan üç formül aşağıdaki gibidir (Tanaka, 1980).

$$Min \quad J(c) = \sum_n c|x_i| \qquad (9)$$

$$y_i + \frac{1}{|L(h)|} e_i \leq \alpha x_i + \frac{1}{|L(h)|} c|x_i|$$

$$y_i - \frac{1}{|L(h)|} e_i \geq \alpha x_i - \frac{1}{|L(h)|} c|x_i|$$

$$c_i \geq 0, i = 1.....n$$

$$Max \quad J(c) = \sum_n c|x_i| \qquad (10)$$

$$y_i + \frac{1}{|L(h)|} e_i \geq \alpha x_i + \frac{1}{|L(h)|} c|x_i|$$

$$y_i - \frac{1}{|L(h)|} e_i \leq \alpha x_i - \frac{1}{|L(h)|} c|x_i|$$

$$c_i \geq 0, i = 1.....n$$

$$Min \quad J(c) = \sum_n c|x_i| \qquad (11)$$

$$y_i + \frac{1}{|L(h)|} e_i \geq \alpha x_i - \frac{1}{|L(h)|} c|x_i|$$

$$y_i - \frac{1}{|L(h)|} e_i \leq \alpha x_i + \frac{1}{|L(h)|} c|x_i|$$

$$c_i \geq 0, i = 1.....n$$

Burada c_i değerleri bulanık sayıların yayılım değerleridir. $L(h) = 1 - h$ olup, bulanık regresyonda uyum iyiliğini göstermektedir.

Genelde Max problemi ve Birleştirme problemleri için optimal çözüm bulunamadığı ve yorumlamada bazı zorluklarla karşılaşıldığından dolayı Min problemleri literatürde en sık kullanılan problem türüdür.

$$\hat{Y}_i = (\alpha_1, c_1)_L x_1 + \dots\dots + (\alpha_n, c_n)_L x_n \tag{12}$$

yazılır. Burada Y'nin üyelik fonksiyonunun açık biçimi,

$$u(y_i) = \begin{cases} 1 - \dfrac{|y_i - X'\alpha|}{c^t|X|}, X \neq 0 \\ 1, X = 0, y \neq 0, \forall i = 1 \dots m \\ 0, X = 0, y = 0 \end{cases} \tag{13}$$

Buradan amaç, $1 - \dfrac{|y_i - X^t\alpha|}{c^t|X|} \geq h, \forall i = 1 \dots m$ kısıtı altında yayılımın minimize edilmesidir. En son olarak Min problemi aşağıdaki doğrusal programlama problemi olarak ifade edilir:

$$\begin{aligned} & MIN \sum_{j=0}^{n} (c_j \sum_{i=1}^{n} |x_{ij}| \\ & \sum_{j=0}^{n} \alpha_j x_{ij} + (1-h) \sum_{j=}^{n} c_j |x_{ij}| \geq y_i + (1-h) e_i \\ & \sum_{j=0}^{n} \alpha_j x_{ij} - (1-h) \sum_{j=0}^{n} c_j |x_{ij}| \leq y_i - (1-h) e_i \\ & c_j \geq 0, a \in R, x_{i0} = 1, (0 \leq h \leq 1; \forall \dot{I} = 1, \dots, n) \end{aligned} \tag{14}$$

3.2 Çok Amaçlı Programlama İle Bulanık Doğrusal Regresyon

Tanaka (1980) tarafından önerilen yöntemin en büyük sakıncası uç değerler karşı duyarlı olmasıdır. Bu yöntemin uç değerlere karşı duyarlı olması (Redden, 1996; Woodall, 1996) tarafından belirtilmiştir. Bu yöntemin daha güvenli sonuçlar vermesi ve bu sakıncalardan etkilenmesini aza indirmek için çok amaçlı programlama yöntemine dayanan yeni bir formülasyon önerilmiştir (Özelkan, 2000, Duckstein, 2000). Bu yöntem sayesinde uç değerlerin model üzereindeki etkileyici gücü azalmış ve tahminler daha iyi olmuştur. Bu yötem aşağıdaki gibi ifade edilebilir.

$$\begin{aligned} & \text{Min-Dominate } \{V, E^p\} \\ & (\hat{y}_i - (1-h) + e_i) - y_i \leq \varepsilon_{L,i} \\ & y_i - (\hat{y}_i - (1-h) + e_i \leq \varepsilon_{R,i} \end{aligned} \tag{15}$$

$$\varepsilon_{L,i} \ , \varepsilon_{R,i} \geq 0 \ \forall n$$
$$E^p = \sum_n (\varepsilon_{L,i}{}^p + \varepsilon_{R,i}{}^p) \qquad 1 \leq p \leq \infty$$
$$V = \sum_{i=1}^{n} \{P_R(\alpha) - P_L(\alpha)\}$$

Buradan elde edilen uzlaşma çözümünde kararı karar verici belirler.

3.3 Bulanık Doğrusal Regresyonda Bulanık En Küçük Kareler Yöntemi

İkinci yöntem olarak bulanık en küçük kareler yöntemi önerilmiştir (Diamond, 1980). Bu yöntem en küçük kareler yönteminin genelleştirilmiş hali olup birinci yönteme göre daha iyi sonuçlar vermektedir. Bu problem için formülasyon aşağıdaki gibidir.

$$r(A_0 + A_1 X_{i1}, Y_i) = (\alpha_0 + c_0 + \alpha_1 X_{i1} + c_1 X_{i1} - \hat{y}_i - e_i)^2 + (\alpha_0 + \alpha_1 X_1 - \hat{y}_i - e_i)^2 \quad (16)$$
$$+ (\alpha_0 - c_0 + \alpha_1 X_{i1} - c_1 X_{i1} - \hat{y}_i + e_i)^2$$

3.4 H Kesit Yöntemi İle Min Problem Çözümü

Tablo 1'de görüldüğü üzere hem Y hem de X değişkeninin bulanık olduğu durumlarda yukarıda ifade edilen formülasyonlar problemi çözmede başarılı olmamaktadırlar. Bunun sebebi iki bulanık sayının çarpımı tam olarak ifade edilen üçgensel ya da yamuk bulanık sayıyı vermemektedir ve Zadeh (1975) tarafından önerilen genişletme prensibiyle Y değişkenin üyelik fonksiyonu yazılamamaktadır. Sakawa ve Yano (1992) Tanaka'nın (1980) de önerdiği Min problemine karşılık olarak aşağıdaki formülasyonu önerdiler.

$$[A \otimes X_i]_h = [Y^{L*}{}_{ih}, Y^{R*}{}_{ih}] \quad (17)$$
$$Y^{L*}{}_{ih} = \sum_{j \in J_1} (\alpha_j - L^{-1}(h)c_j)(\tilde{x}_{ij} - L^{-1}(h)d_{ij}) + \sum_{j \in J_2 \cup J_3} (\alpha_j - L^{-1}(h)c_j)(\tilde{x}_{ij} + L^{-1}(h)d_{ij}) \quad (18)$$
$$Y^{R*}{}_{ih} = \sum_{j \in J_1 \cup J_2} (\alpha_j + L^{-1}(h)c_j)(\tilde{x}_{ij} + L^{-1}(h)d_{ij}) + \sum_{j \in J_3} (\alpha_j + L^{-1}(h)c_j)(\tilde{x}_{ij} - L^{-1}(h)d_{ij})$$

Kapalı biçimde ifade edilemeyen üyelik fonksiyonu h kesiti şeklinde yukarıdaki gibi ifade edilir. Buradan Tanaka'nın (1980) Min problemindeki gibi yayılım değerlerini minimize edecek şekilde problem aşağıdaki gibi ifade edilir.

$$Min \qquad \sum_{i=1}^{n} (Y^{R*}{}_{ih} - Y^{L*}{}_{ih}) \quad (19)$$
$$Y^{R*}{}_{ih} \geq \bar{y}_i + L^{-1}(h)e_i$$
$$-Y^{L*}{}_{ih} \geq -\bar{y}_i + L^{-1}(h)e_i \ , \ i = 1 \ldots n$$

4. SONUÇ

Literatürde başka varsayımlar altında BDR konusunda çeşitli sayıda makale mevcuttur. Burada kısaca BDR deki değişkenlerin türlerine göre önerilen ve uygulamada ve daha sonra çıkan çeşitli problemlere göre yeniden düzenlenen formülasyonlardan bahsedilmektedir.

KAYNAKLAR

DUBOIS, D VE PRADE, H. (1980), *Fuzzy sets and Systems: Theory and Apllications*, New york, Academic Press.

ÖZELKAN,Ö VE DUCKTEIN, R. (2000), *Multiobjective Fuzzy Regression*, European Journal of Operation Research, 126,637-650.

REDDEN, D ve WOODALL,H. (1996), *Further Examination of Fuzzy Regression*, Fuzzy Sets and Systems,79,203-211.

SAKAWA, M, ve YANO, H. (1992), *Multiobjective Fuzzy Linear Regression Analysis for Fuzzy Input-Output Data*, Fuzzy Sets and Systems, 47, 173,181.

TANAKA, H. (1987), *Fuzzy Data Analysis By Possibilistic Linear Modells*, Fuzzy Sets and Systems,24,363-375.

ZADEH, L (1975), *The Concept of a linguistic Variable and Its Application to Approximate Reasoning*, Information Sciences,8,199-249

REGRESSION ANALYSIS BASED ON POSSIBILITY THEORY

ABSTRACT

In this study, regression analysis based on possibility theory has been rewieved.This modelling tool helps researchers gain insigth about their subjects under the circumstances that classical statistical modelling tools does not help much.This modelling technique is widely used in many areas.This review is restricted to the cases that consist of only Tanaka's method and fuzzy least squares method and their modified versions based on defined variables.

Key Words : *Fuzzy Linear Regression, , Fuzzy Number, Fuzzy Set.*