ISSN:1306-3111
e-Journal of New World Sciences Academy
2009, Volume: 4, Number: 3, Article Number: 3A0009

PHYSICAL SCIENCES
Received: November 2008
Accepted: June 2009
Series : 3A
ISSN : 1308-7304

Kamil Alakuş
Taner Tunç
Mehmet Gürcan
Ondokuz Mayis University
kamilal@omu.edu.tr
Samsun-Turkey

DOĞRUSAL REGRESYON ANALİZİNDE BAĞIMLI BULANIK DEĞİŞKENİNİN İNCELENMESİ VE MODEL SEÇİMİ

ÖZET

Bu çalışmamızda bağımlı bulanık değişkeni üzerinde durarak model tahmini yapmaya çalıştık. Model tahmini yaygın olarak bilinen yöntemlerin dışında gözlem değerlerinden elde edilen olasılıkları kullanarak hatanın beklenen değerini en küçük yapmaya dayanmaktadır. Bu yöntem ağırlıklı en küçük kareler yönteminin bir benzeridir.

Anahtar Kelimeler: Bağımlı bulanık değişkeni, Bağımsız Değişken Seçimi, Ağırlıklı En Küçük Kareler Yöntemi

INVESTIGATION OF THE FUZZY DEPENDENT VARIABLE IN THE LINEAR REGRESSION ANALYSIS AND MODEL SELECTION

ABSTRACT

In this study, we are estimated model with dependent fuzzy variable. Estimating the model may be minimized expected mean square error which is instead of the most commonly used method by using the probabilities of the observations. This method is also similar to the weighted least square method.

Keywords: Dependent Fuzzy Variable, Independent Variable Selection, Weighted Least Square Method

1. GİRİŞ (INTRODUCTION)

Regresyon analizinde bağımlı bulanık değişkenlerinin kullanılması son yıllarda oldukça ilgi çekici bir konu haline gelmiştir. Uygulamalarda genellikle bağımlı bulanık değişkenlerine ait örneklere az rastlanılmaktadır. Deneysel analizlerde elde edilen gözlemlerin sonuçları genellikle kesin sayısal veriler olmaktadır. Elde edilen bağımlı değişken değerinin bir aralıkta değer alabileceğinin söylenmesi ölçüm değerinin hatalı olabileceğinden veya ardı ardına yapılan aynı deneylerde aynı bağımsız değişken değerlerinin kullanılmasına karşılık elde edilen bağımlı değişken değerlerinin birbirlerinden çok az da olsa farklı gözlemlenmesinden kaynaklanmaktadır. Bu gibi durumlarda bağımlı değişken değerlerinin ortalamasının alınması yerine bağımlı değişkene bir bulanık değişkeni olarak bakılmasında yarar vardır. Ancak buradaki sorun bağımlı bulanık değişkeninin üyelik fonksiyonunun tanımlı olduğu aralıkta nasıl alınması gerektiğidir. Literatürde sıkça kullanılan üyelik fonksiyonları yerine yapılan deneylerden elde edilen bağımlı değişken gözlem değerlerinin olasılık fonksiyonunun kullanılması bu aşamada analize olumlu yönde bir katkı sağlamaktadır.

Tüm bunların yanı sıra seçilen model denkleminin uyumunu artırabilmek için de (yani; regresyon kareler toplamında bir azalış meydana getirebilmek için) bağımlı değişken değerlerinin gözlem değerleri civarında değiştirilebilmesi olanağını vermektedir. Aynı zamanda regresyon kareler toplamını oldukça az miktarda etkileyen ve modele katkısının çok az olduğu görülen değişkenlerin modelden atılabilmesine de olanak sağlamaktadır. Bu tip değişkenlerin modelden tamamen atılması yerine tüm değişkenlerle birlikte elde edilen bağımlı değişken değerlerinin, önemsiz olduğuna karar verilen değişkenler atıldıktan sonra da bir bulanık değişkeni gibi kullanılmasında analiz açısından yarar vardır. Aynı uygulama bağımsız değişkenlerin ilişkili olduğu durumda da kullanılabilmektedir. Birbiri ile ilişkili olan bağımsız değişkenleri modelden çıkartırken bağımlı değişkenin tüm değerlerini kullanabilmemizi sağlamaktadır. Sonuç itibari ile bağımlı bulanık değişkeninin kullanılması günümüzde birçok problemi ortadan kaldırdığından ve model denkleminin uyumunu artırdığından tüm regresyon analizinde vazgeçilmez bir hale gelmiştir.

Bulanık regresyon analizi ile ilgili günümüzde yapılan çalışmalarda iki farklı yöntem kullanılmaktadır. Bunlardan birincisi bağımlı bulanık değişkeninin farklı değerler aldığı durumda en küçük kareler yöntemi kullanarak çok değişkenli regresyon analizinde olduğu gibi tahmin yapılmasıdır (Picrpaola ve Gastaldi, 2000). İkincisi ise doğrusal programlama yöntemi kullanılarak yapılan tahmindir. Bu yöntemde bağımsız değişken değerleri bulanık değişkeni olarak seçilmekte ve genellikle üyelik fonksiyonu olarak simetrik üçgen veya belirlenen aralıkta simetrik olacak şekildeki bir fonksiyon kullanılmaktadır. Bu tip örnekleri ilk olarak Tanaka ve Ucjima (1982) başlatmıştır. Bu araştırmacılardan sonra Chang ve Lee, 1994a, 1994b, Diamond, 1990; Diamond ve Kloeden, 1994; Chang ve Ayyuh, 2001 gibi araştırmacılar devam ettirmiştir.

2. ÇALIŞMANIN ÖNEMİ (RESEARCH SIGNIFICANCE)

Gerçekleştirdiğimiz bu çalışmada bağımlı bulanık değişkenli gözlem değerlerinden elde edilen olasılıkları yardımı ile ağırlıklı en küçük kareler yöntemini kullanarak model denklemini tahmin etmeye çalışacağız. Buna ek olarak modele katkısının az olduğuna karar verilen bağımsız değişkenlerin modelden çıkartılmasına karar verildiğinde bu işlemi yaparken bağımlı bulanık değişkeninin nasıl kullanıldığını sayısal örnek yardımıyla elde edilen bir veri kümesinde göstereceğiz.

3. BAĞIMLI BULANIK DEĞİŞKEN VE PARAMETRE TAHMİNİ (DEPENDENT FUZZY VARIABLE AND PARAMETER ESTIMATION)

Bağımlı bulanık değişkeninin değerleri $Y_1 = (y_{11},\dots,y_{1k_1})$, …, $Y_n = (y_{n1},\dots,y_{nk_n})$ ile gösterilsin. Burada görüldüğü gibi her bir bağımlı değişken değeri aynı sayıda bileşen içermek zorunda değildir. Olasılık değerlerini ise $P(Y_i = y_{ij}) = \alpha_{ij}$ ve $P(Y = y_{ij}) = \dfrac{\alpha_{ij}}{n}$ şeklinde gösterilsin. Bu olasılıklarla birlikte kullanılacak istatistik aşağıdaki (1) eşitliğinde verildiği gibidir.

$$E(Y-\hat{Y})^2 = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}(y_{ij}-\hat{Y}_i)^2 \,. \tag{1}$$

(1) eşitliğinde verilen istatistiğin tahmin denkleminin hangi parametreleri için en küçük olacağı en küçük kareler yöntemiyle bulunabilir. Şimdi model denklemini

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon \tag{2}$$

şeklinde olduğu düşünüldüğünde (1) istatistiğini en küçükleyen tahmin denkleminin parametreleri bulunabilir. Bunun için (1) istatistiğinin β_0, β_1 ve β_2 parametrelerine göre kısmi türevleri alınırsa (3) sistemindeki denklemler elde edilir.

$$\begin{aligned}
&\beta_0 n + \beta_1 \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}x_{1i} + \beta_2 \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}x_{2i} = \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}y_{ij}\\
&\beta_0 \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}x_{1i} + \beta_1 \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}x_{1i}^2 + \beta_2 \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}x_{1i}x_{2i} = \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}y_{ij}x_{1i}\\
&\beta_0 \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}x_{2i} + \beta_1 \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}x_{1i}x_{2i} + \beta_2 \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}x_{2i}^2 = \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}y_{ij}x_{2i}
\end{aligned} \tag{3}$$

Bu durumda $X'X$ ve $X'Y$ veri matrisleri sırası ile

$$X'X = \begin{bmatrix} n & \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}x_{1i} & \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}x_{2i}\\ \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}x_{1i} & \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}x_{1i}^2 & \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}x_{1i}x_{2i}\\ \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}x_{2i} & \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}x_{1i}x_{2i} & \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}x_{2i}^2 \end{bmatrix}, \tag{4}$$

$$X'Y = \begin{bmatrix} \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}y_{ij}\\ \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}y_{ij}x_{1i}\\ \sum_{i=1}^{n}\sum_{j=1}^{k_i}\alpha_{ij}y_{ij}x_{2i} \end{bmatrix} \tag{5}$$

verilebilir. Sonuç olarak parametre tahmin vektörü, $\hat{\beta} = (X'X)^{-1}X'Y$ eşitliğinden elde edilebilir. Burada yine $E(Y-\hat{Y}) = 0$ şartı sağlanmaktadır.

4. SAYISAL ÖRNEK (NUMERICAL EXAMPLE)

Laboratuar ortamında yapılan bir deneyde 4 adet bağımsız değişken seçilmiş olsun. Deney tesadüfî olarak tekrarlandığında deney ortamında göz ardı edilen veya ölçülemeyen etkilerin değişimiyle her bir deneyde elde edilen gözlem değerleri çok az da olsa farklılık gösterecektir. Bu yöntemin kullanılmasının amacı hem modele katkısı az olan bağımsız değişkenleri modelden çıkartabilmek ama bu işlemi yaparken etkilerini göz ardı etmemek ve hem de deney düzeneğinde kontrol altına alınamayan değişkenlerin etkilerini modele yansıtmak içindir. Deney düzeneğinde kontrol altına alınamayan etkilere örnek olarak havadaki oksijen miktarı, deneyi yapan kişinin vücut ısısı, ölçüm aletlerinin kullanım şeklinden dolayı deformasyonu gibi değişkenler verilebilir. Bunlar az da olsa gözlem değerleri üzerinde etki yapabilmektedir. Örneğin biyokimya laboratuarlarında aynı hastanın kan değerleri art arda alınan örneklerde çok az da olsa faklı çıkabilmektedir. Şimdi böylesi bir çalışmada Tablo 1'deki veriler gözlemlenmiş olsun.

Tablo 1. $Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\varepsilon$ Modeline Ait Veriler.
(Table 1. Data for the model $Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\varepsilon$).

Bağımsız Değişkenler					Bağımlı Değişken
Sıra No	X_1	X_2	X_3	X_4	Y
1	0.5	0.1	2	5	1.65
2	0.8	0.9	3.5	6.3	4.65
3	1.2	1.3	8	11	6.65
4	1.6	2	7.01	10.005	9.55
5	2.1	2.4	6.66	12.987	11.75
6	2.5	3.1	54.05	12.232	14.65
7	2.8	4	13	4	17.95
8	3.1	4.8	32	1	20.95
9	3.4	5.7	42.6	0.66	24.25
10	3.9	6.3	14.5	15	27.05
11	4.2	7.2	76	8.09	30.35
12	4.5	8	77	7.3	33.35
13	4.9	8.6	0.43	32	35.95
14	5.3	9.2	30.11	21.04	38.55
15	5.9	10	12	45	42.15
16	6.5	10.8	37	0.87	45.75
17	7.0	11.5	14	28	48.85
18	7.4	12.1	25	22.2	51.45
19	7.9	13	54.7	42.9	55.15
20	8.6	13.7	32	65.7	58.65
21	9.1	14.7	41.08	35	62.65
22	9.8	15.6	13	41.3	66.75
23	10.5	16.3	11.01	21	70.25
24	11.0	17.4	89	32.5	74.55
25	11.6	18.9	33	12.76	80.25
26	12.1	19.3	5.98	21	82.45
27	12.8	20.6	8	5	87.75
28	13.4	21.2	43	9.5	90.75
29	14.1	22	21	24.5	94.55
30	15	22.8	91	23.4	98.75

Tablo 2. $Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\varepsilon$ Modeline Ait Sonuçlar.
(Table 2. Results for the model $Y=\beta_0+\beta_1X_1+\beta_2X_2+\beta_3X_3+\beta_4X_4+\varepsilon$).

Parametreler	Tahmin Değeri	Standart Hata	Z-testi	α - olasılığı
β_0	0,3750	0,0343	8,9558	0,0000
β_1	2,0016	0,0487	41,0865	0,0000
β_2	3,0003	0,0305	98,2759	0,0000
β_3	-0,0002	0,0006	-0,2665	0,7920*
β_4	0,0010	0,0011	0,8594	0,3983*

Tablo 2'den model denkleminin tahmini

$$\hat{Y}=0.3075+2.0016X_1+3{,}0003X_2-0.0002X_3+0.0010X_4 \qquad (6)$$

olarak elde edilir. Z-hesap değerinden de görüldüğü gibi X_3 ve X_4 değişkenlerinin modele katkısının olmadığı söylenebilir. Doğal olarak tahmin denkleminde bu değişkenlerin kullanılmaması tahmin değerlerini göz ardı edilemez şekilde etkilemeyecektir. Diğer bir deyişle, tahmin denkleminde bu değişkenlerin kullanılmaması hata kareler toplamında kayda değer bir artışa neden olmaz. Tahmin denklemini daha az bağımsız değişkenle oluşturabilmek için bu değişkenleri modelden çıkartmak mümkündür. Bu durumda bu değişkenleri yok sayarak doğrudan tahmin denkleminden çıkartmak yerine tüm değişkenlerin bulunduğu tahmin denkleminden yararlanarak modele katkısı önemli olan X_1 ve X_2 değişkenleri sabit kalmak suretiyle X_3 ve X_4 değişkenlerinin değerleri her seferinde farklılaştırılarak birkaç bağımlı değişken değeri elde edilebilir. Bu işlem deney uygulanırken değişken değerleri sabit tutulduğunda kontrol edilemeyen fakat etkisinin göz ardı edilebileceği faktörlerden dolayı gözlem sonucu farklı çıkabilecek ardı ardına yapılan deneylerin gözlem sonuçlarının elde edilmesiyle aynı özelliğe sahip olacaktır. Yapılan her iki işlemde de elde edilen bağımlı değişken değerleri bir bağımlı bulanık değişken değeri olup bu özelliğine dayanılarak incelenebilir. Burada modelden çıkartılmak istenen değişkenlerin modele katkısı çok az olarak görülse de tahmin denkleminden çıkartıldığında bağımlı bulanık değişken değerlerinde etkisi olacağından tahmin değerlerinde bu değişkenler, tahmin denkleminde olmamasına karşılık, etkili olabileceklerdir. Sonuç olarak çıkartılmak istenen değişkenlerin etkileri tahmin denkleminde bulunmamalarına karşılık tahmin değerlerinde karşımıza çıkabilecektir.

Tablo 1'deki verilerle yapılan incelemede kullanılmak istenilmeyen X_3 ve X_4 değişkenlerinin değerleri değiştirilerek yeni bağımlı bulanık değişken değerleri Tablo 3'deki şekilde elde edilmiştir. Bu işlem yapılırken her bir gözlem sırasındaki X_1 ve X_2 değişken değerleri sabit tutulmak zorundadır. Tablo 3'deki bağımsız değişken değerlerinin farklı sayıda olması araştırmacıya isteğine göre değişmektedir. Özellikle kimyacılar laboratuar deneylerinde elde etmek istedikleri sonucu bulabilmek için veya deney düzeneğinde ideal ortamı yakalayabilmek için deneyi tekrarlarlar. Bu durumda modelleme yapabilmek için yapılan tüm deney sonuçlarını kullanmayarak istedikleri ideal sonucu kullanırlar. Bu durum diğer yapılan deneylerin önemini ortadan kaldırmaktadır. Buna göre yapılan diğer deneylerin sonuçlarını da göz önüne alabilmek için bu yöntem kullanılır. Burada atılmak istenen değişkenlerin model üzerindeki etkisi az olacağından bağımlı değişken değerleri birbirlerine yakın çıkacaktır. Burada da yine Tablo 3'deki gibi veriler elde edilir. Model kurulduğunda modelde istenen değişkenler bulunmakla birlikte atılan değişkenlerin de bağımlı değişken değerleri üzerinde etkisi olacaktır. Çünkü boyut indirgemede normalde atılan değişkenler yok sayılarak bağımlı değişken tahmini yapılmaktadır. Bu ise modele katkısı ne kadar az olursa olsun bir bağımsız değişkenin göz ardı edilmesi demektir. Bizim elde ettiğimiz modelde atılan değişkenler gözükmese bile bağımlı değişken tahminleri üzerinde etkileri hala bulunmaktadır. Tablo 3'deki yeni değişkenlerle işlem yapıldığında tahmin denkleminde göz ardı edilen X_3 ve X_4 değişken değerleri de tahmin değerlerinde bir katkısı olacaktır. Tablo 3'deki veriler (4) ve (5) matrisleri kullanıldığında X_1 ve X_2 değişkenlerinden meydana gelen yeni tahmin modeli elde edilebilecektir. Buradan elde edilen sonuçlar ise Tablo 4'de verilmiştir.

Tablo 3. $Y_{ij} = \beta_0 + \beta_1 X_{1j} + \beta_2 X_{2j} + \varepsilon_j$ modeline ait veriler

(Table 3. Results for the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon$)

Sıra No	Bağımsız Değişkenler		Bağımlı Değişken
	X_1	X_2	Y
1	0.5	0.1	1.65; 1.65; 1.65; 1.63; 1.32
2	0.8	0.9	4.65; 4.67; 4.67; 4.62; 4.67
3	1.2	1.3	6.65; 6.61; 6.65; 6.61; 6.64
4	1.6	2	9.55; 9.55; 9.55
5	2.1	2.4	11.75; 11.77; 11.076; 11.77
6	2.5	3.1	14.65; 14.65; 14.654
7	2.8	4	17.95; 17.99; 17.95
8	3.1	4.8	20.95; 20.95; 20.95; 20.95; 20.92
9	3.4	5.7	24.25; 24.34; 25; 25.7; 25.001
10	3.9	6.3	27.05; 27.05; 27.055
11	4.2	7.2	30.35
12	4.5	8	33.35; 33.36; 33.35; 33.36
13	4.9	8.6	35.95; 35.003; 35.98; 35.43; 35.95
14	5.3	9.2	38.55; 38.55; 38.506
15	5.9	10	42.15; 41.003
16	6.5	10.8	45.75; 45.75; 45.76
17	7.0	11.5	48.85; 48.87; 48.85; 48.85
18	7.4	12.1	51.45; 51.55; 51.45; 51.45
19	7.9	13	55.15
20	8.6	13.7	58.65; 58.60; 58.65; 58.64; 58.72
21	9.1	14.7	62.65; 62.65
22	9.8	15.6	66.75; 66.753; 66.75
23	10.5	16.3	70.25; 70.22; 70.22
24	11.0	17.4	74.55; 74.55; 74.55; 74.55
25	11.6	18.9	80.25; 82.001
26	12.1	19.3	82.45; 82.64; 82.45
27	12.8	20.6	87.75; 87.76; 87.75; 87.75; 87.75
28	13.4	21.2	90.75; 90.75; 90.754
29	14.1	22	94.55; 94.58; 94.55
30	15	22.8	98.75; 98.75; 98.75; 98.77

Tablo 4. $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$ Modeline Ait Sonuçlar.
(Table 4. Results for the model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$).

Parametreler	Tahmin Değeri	Standart Hata	Z-testi	α-olasılığı
β_0	0,3700	0,0487	7,5931	0,0000
β_1	1,9503	0,0829	23,5388	0,0000
β_2	3,0324	0,0515	58,8972	0,0000
Hata Kareler ortalaması: σ_e^2=0,2786; s.d.=27				

Tablo 4 ile elde edilen yeni tahmin denklemi (6) modelinden farklı olarak

$$\hat{Y} = 0.37 + 1.9503 X_1 + 3{,}0324 X_2 + \varepsilon \qquad (7)$$

şeklinde olup bu tahmin modelindeki üç parametrenin de önemli olduğu görülmektedir.

5. SONUÇ VE TARTIŞMA(CONCLUSION AND REMARKS)

Çalışmada yer alan uygulamadan da görüleceği gibi kullanılan yöntem regresyon analizinde modele katkısı az olan değişkenlerin modelden çıkartılmasında kolaylık sağlamaktadır. Bununla birlikte önemsiz değişkenler modelden çıkartılırken regresyon kareler toplamında herhangi bir değişime neden olmamaktadır. Ayrıca kullanılan yöntem aralarında ilişki bulunan bağımsız değişkenlerin düzenlenebilmesinde de rahatlıkla kullanılabilir.

KAYNAKLAR (REFERENCES)

- Chang, P.T. and Lee, E.S., (1994a). Fuzzy Linear Regression With Spreads Unrestricted In Sign. Comput. Math. Apply. 28, 61-70.
- Chang, P.T., and Lee, E.S., (1994b). Fuzzy Least Absolute Deviations Regression Based On The Ranking Of Fuzzy Numbers. IEEE Conference On Fuzzy Systems At 1994 World Congress On Computational Intelligence. Orland, FL, June 28, 26-29.
- Chang, Y.-H. O. and Ayyuh, B.M., (2001). Fuzzy Regression Methods-A Comparative Assessment. Fuzzy Sets And Systems 119, 187-203.
- Diamond, P., (1990). Higher Level Fuzzy Sets Arising In Linear Regression. Fuzzy Sets And Systems 96, 265-275.
- Diamond, P. and Kloenden, P., (1994). Metric Spaces Of Fuzzy Sets. Theory And Applications. Singapore: World Scientific.
- Picrpaola, D. and Gastaldi, T., (2000). A Least-Squares Approach To Fuzzy Linear Regression Analysis. Comp. Stat. And Data Annuli. 34, 427-440.
- Tanaka, H. and Ucjima, K.A., (1982). Linear Regression Analysis With Fuzzy Model. IEEE, Systems, Trans. Systems Man. Cybernetic. SMC-2, 903-907.