Journal of Korean
Data & Information Science Society
2007, Vol. 18, No. 2, pp. 301 ~ 314

Separate Fuzzy Regression with Crisp Input and Fuzzy Output

Jinhee Yoon[1], Seunghoe Choi[2]

Abstract

The aim of this paper is to deal with a method to construct a separate fuzzy regression model with crisp input and fuzzy output data using a best response function for the center and the width of the predicted output. Also we introduce the crisp mean and variance of the predicted fuzzy value and also give some examples to compare a performance of the proposed fuzzy model with various other fuzzy regression model.

Keywords : Accuracy, Fuzzy Regression Model, Least Squares Method

1. 서론

회귀분석이란 한 개 혹은 여러 개 변수를 이용하여 다른 변수의 값을 설명하거나 예측하기 위해 변수들 사이의 함수 관계를 유추하거나 모형화(modeling)하여, 관측된 자료로부터 주어진 관계나 모형을 추정하는 통계적 분석방법이다. 이때 설명하는 변수를 독립변수 또는 설명변수라 하고, 설명변수에 종속되어 있거나 예측이 되는 변수를 종속변수 또는 반응변수라고 한다. 설명변수와 반응변수 사이의 관계를 모형화하는 과정에서 발생하는 문제는 불확실성(uncertainty)이다. 불확실성에는 시간이 흘러가거나 실험을 통하여 불확실성이 해결되는 확률적인 불확실성과 실험이나 시간과는 무관한 퍼지적 불확실성이 있다. 종속변수가 내포하고 있는 확률적인 불확실성에 대한 연구는 활발히 진행되어 많은 분야에서 응용되고 있다. Zadeh(1965)는 실험이나 시간과는 무관한 퍼지적 불확실성을 애매함(ambiguity)과 모호함(vagueness)으로 설명하고, 애매하고 모호한 문장이나 정보들을 처리하기 위하여 필요한 시스템을 구현하기 위하여 퍼지이론을 소개하였다.

퍼지적 불확실성을 포함하고 있는 회귀모형을 추정하기 위해 Tanaka 등(1980,

---

1) Instructor, Department of Mathematics, Yonsei University, Seoul, 120-749, Korea. E-mail : jin9135@yonsei.ac.kr.

2) Corresponding Author : Associate professor, Department of General Studies, Korea Aerospace University, Koyang, Kyungkido, 412-791, Korea. E-mail : shchoi@kau.ac.kr

1982)은 퍼지회귀모형(fuzzy regression model)

$$Y_i = A_0 + A_1 x_{i1} + \cdots + A_p x_{ip} \tag{1.1}$$

을 처음으로 소개하였다. 여기서 $x_{ij}(j=0,\cdots,p)$는 i번째 설명변수의 j번째 성분이고 $x_{i0}=1$이다. 또한 회귀계수 $A_j=(a_j,l_{a_j},r_{a_j})_{LR}$와 반응변수 $Y_i=(y_i,l_{y_i},r_{y_i})_{LR}$는 LR-퍼지수이다. Dubios와 Prade(1979)가 제안한 LR-퍼지수 $A=(a,l_a,r_a)_{LR}$의 소속함수(membership function) $\mu_A(x)$는

$$\mu_A(x) = \begin{cases} L_A\left(\dfrac{a-x}{l_a}\right) & \text{for} \quad 0 \le a-x \le l_a \\ R_A\left(\dfrac{x-a}{r_a}\right) & \text{for} \quad 0 \le x-a \le r_a \\ 0 & \text{for} \quad otherwise. \end{cases}$$

이고, 단조증가함수 L_A와 단조감소함수 R_A는

$$L_A(0) = R_A(0) = 1 \text{와 } L_A(1) = R_A(1) = 0$$

을 만족한다. 여기서 a를 LR-퍼지수 A의 중심이라 하고, l_a와 r_a를 각각 LR-퍼지수 A의 왼쪽과 오른쪽의 폭이라 한다. 만약 $L_A(x)=R_A(x)=1-x$이면 LR-퍼지수 A를 삼각퍼지수라고 부르고 $(a,l_a,r_a)_T$ 와 같이 표시하고, 특히 양 쪽 폭이 같으면 $(a,a_s)_T$와 같이 표시한다.

퍼지회귀모형 (1.1)에 주어진 퍼지회귀계수 A_j를 추정하기 위하여 수치해석적인 방법과 통계적인 방법이 여러 사람들에 의하여 제시되었다. 예측퍼지수의 폭을 최소화하는 수치해석적인 방법은 Savic과 Pedryzc(1991), Kao와 Chyu(2002), Sakawa와 Yano(1992), 그리고 Tanaka 등(1980, 1982, 1989)이 연구하였으며, Diamond(1988), Diamond와 Korner(1997), Kim과 Bishu(1998), Chang과 Ayyub(2001), 그리고 Kao와 Chyu(2003)등은 예측퍼지수와 관측퍼지수의 차를 최소화하는 통계적인 방법을 연구하였다. 그러나 설명변수가 실수인 퍼지회귀모형 (1.1)은

(1) 퍼지회귀계수와 반응변수의 소속함수 형태가 일치하는 경우
(2) 퍼지수인 종속변수의 폭에 대한 반응함수가 중심에 대한 반응함수와 일치하는 경우
(3) 종속변수의 폭과 중심에 대한 설명변수가 일치하는 경우

와 같은 경우에는 좋은 모형이 될 수 있으나 이와 같은 조건을 만족하지 않는 경우에는 추정된 퍼지회귀모형의 정확성이 떨어질 수 있다. 따라서 회귀계수와 반응변수의 소속함수에 독립적이고 종속변수에 대한 회귀모형이나 혹은 설명변수에 제약을 받지 않는 새로운 퍼지회귀모형을 연구할 필요가 있다.

본 논문은 반응변수와 회귀계수가 LR-퍼지수이고, 설명변수가 실수인 퍼지회귀모형을 추정하기 위해 종속변수의 중심과 폭에 대한 회귀모형을 분리한 독립퍼지회귀모형

을 소개한다. 독립퍼지회귀모형을 추정하기 위해 최소제곱법을 이용하고 추정된 퍼지수의 평균과 분산을 유도한다. 본 논문에서 제시된 독립퍼지회귀모형의 정확성을 Tanaka 등(1989)이 추정한 퍼지회귀모형과 비교한다.

2. 독립퍼지회귀모형

본 절에서는 퍼지수인 종속변수의 중심과 폭에 대한 회귀모형을 분리한 독립퍼지회귀모형을 소개하고, 제안된 퍼지회귀모형을 추정하기 위하여 최소제곱법을 사용한다. LR-퍼지수 $A=(a,l_a,r_a)_{LR}$의 α-수준집합은

$$A(\alpha)\equiv\{x|\mu_A(x)\geq\alpha\}$$

와 같이 정의되고

$$A^*(\alpha)=\{x\in X:\mu_A(x)>\alpha\}$$

을 강 α-수준집합(strong α-level set)이라 한다. 여기서 $0<\alpha\leq 1$이고 $\alpha=0$인 경우 $A(0)$는 집합 $A^*(0)$의 폐포(closure)이다. LR-퍼지수 A의 α-수준집합 $A(\alpha)$는 중심이 a이고, 왼쪽과 오른쪽 폭이 각각 $l_aL_A^{-1}(\alpha)$와 $r_aR_A^{-1}(\alpha)$인 폐구간 $\left[a-l_aL_A^{-1}(\alpha),a+r_aR_A^{-1}(\alpha)\right]$이므로

$$A(\alpha)\cong\left(a,l_aL_A^{-1}(\alpha),r_aR_A^{-1}(\alpha)\right)$$

와 같이 표현할 수 있다. 또한 LR-퍼지수 A의 1-수준집합과 0-수준집합은 각각

$$A(1)\cong(a,0,0)\text{와 } A(0)\cong(a,l_a,r_a)$$

이다. 따라서 퍼지회귀모형 (1.1)에서 관측퍼지수 Y_i의 α-수준집합 $\left(y_i,l_{y_i}L_{Y_i}^{-1}(\alpha),r_{y_i}R_{Y_i}^{-1}(\alpha)\right)$는 예측퍼지수 $A_0+A_1x_{i1}+\cdots+A_px_{ip}$의 α-수준집합

$$\left(\sum_{k=0}^{p}a_kx_{ik},\ \sum_{k=0}^{p}l_{a_k}L_{A_k}^{-1}(\alpha)x_{ik},\ \sum_{k=0}^{p}r_{a_k}R_{A_k}^{-1}(\alpha)x_{ik}\right)$$

와 같다. 이것은 $\{(y_i,\boldsymbol{x_i})|i=1,\cdots,n\}$에 대한 함수 관계는 $\left\{(l_{y_i}L_{Y_i}^{-1}(\alpha),\boldsymbol{x_i})|i=1,\cdots,n\right\}$와 $\left\{(r_{y_i}R_{Y_i}^{-1}(\alpha),\boldsymbol{x_i})|i=1,\cdots,n\right\}$에 대한 관계와 같고, 반응변수의 값 y_i를 설명하는 독립변수의 개수가 α-수준집합 $Y_i(\alpha)$의 오른쪽과 왼쪽 폭인 $r_{y_i}R_{Y_i}^{-1}(\alpha)$와 $l_{y_i}L_{Y_i}^{-1}(\alpha)$ 에 대한 설명변수의 개수가 같음을 의미한다. 여기서 $\boldsymbol{x_i}=(1,x_{i1},\cdots,x_{ip})$이다. 그러나 관측퍼

지수 Y_i의 α-수준집합을 구성하는 중심과 왼쪽 끝점, 그리고 오른쪽 끝점에 대한 회귀모형이 반드시 일치하는 것은 아니다. 다음 예제는 퍼지회귀모형에 대한 α-수준집합의 중심과 양 끝점에 대한 회귀방정식이 반드시 일치하지 않음을 보여준다.

예제 1. Jennrich(1995)는 관절염 치료에 사용된 약품의 지속력을 이용하여 회귀분석을 설명하였다. <표 1>은 약품을 복용한 후 경과한 시간(x_i)과 LR-퍼지수로 변형한 약품의 지속력(Y_i)에 대한 Jennrich(1995)의 자료이다.

<표 1> 퍼지화된 Jennrich의 자료

x_i	$Y_i=(y_i, s_{y_i})_T$
0	$(100, 1.07)_T$
2	$(86, 0.99)_T$
4	$(78, 1.1)_T$
14	$(60, 1.35)_T$
21	$(53, 1.74)_T$
42	$(47, 2.26)_T$
63	$(42, 3.1)_T$
97	$(40, 3.87)_T$
155	$(35, 5.51)_T$
217	$(33, 7.36)_T$

관절염 치료를 위해 사용된 약품의 지속력은 지수적으로 감소하므로 경과시간(x_i)과 지속력의 중심(y_i)에 대한 회귀모형은

$$y_i = 78.3\exp(-0.007x_i)$$

와 같은 비선형모형이고 지속력의 폭(s_{y_i})과 경과시간(x_i)에 대한 회귀방정식은

$$s_{y_i} = 1.037 + 0.029x_i$$

와 같다. 여기서 $\exp(x)$는 지수함수이다.

예제 1과 같이 종속변수의 중심과 폭에 대한 회귀모형과 설명변수가 일치하지 않는 경우 퍼지회귀모형 (1.1)을 사용하여 설명변수와 종속변수 사이의 관계를 설명하는 것은 추정된 퍼지회귀모형의 정확성을 감소시킬 수 있다. 따라서 이와 같은 경우에는 새로운 퍼지회귀모형을 연구할 필요가 있다.

반응변수와 회귀계수가 퍼지수인 퍼지회귀모형을 추정하기 위하여 본 논문에서는 퍼지회귀모형 (1.1)을 일반화한

$$Y_i = F(\boldsymbol{x_i}, \theta) \tag{2.1}$$

와 같은 독립퍼지회귀모형 (separate fuzzy regression model)을 생각한다. 여기서 $F(\boldsymbol{x_i},\theta) = \left(f(\boldsymbol{x_i^c}, C), g(\boldsymbol{x_i^l}, L), h(\boldsymbol{x_i^r}, R)\right)_{LR}$이고 함수 f와 g, 그리고 h는 실가함수이다. 또한, $\boldsymbol{x_i^c} = (x_{i0}^c, \cdots, x_{ip^c}^c)$, $\boldsymbol{x_i^l} = (x_{i0}^l, \cdots, x_{ip^l}^l)$, 그리고 $\boldsymbol{x_i^r} = (x_{i0}^r, \cdots, x_{ip^r}^r)$이고 c_i, l_i 그리고 r_i는 각각 실수로서 $C = (c_0, c_1, \cdots, c_{p^c})$, $L = \left(l_0, l_1, \cdots, l_{p^l}\right)$, 그리고 $R = \left(r_0, r_1, \cdots, r_{p^r}\right)$이다.

함수 f와 g, 그리고 h가 모두 선형이고 $p^c = p^l = p^r = p$와 $x_{ik}^c = x_{ik}^l = x_{ik}^r (k = 0, \cdots, p)$을 만족하면 독립퍼지회귀모형 (2.1)은

$$Y_i = A_0 + A_1 x_{i1} + \cdots + A_p x_{ip}$$

와 같이 변형되고, 또한 $l_{y_i} = r_{y_i} = 0$이고 $l_{a_k} = r_{a_k} = l_{e_i} = r_{e_i} = 0$이면 독립퍼지회귀모형 (2.1)은

$$(y_i, 0, 0)_{LR} = (a_0, 0, 0)_{LR} + (a_1, 0, 0)_{LR} x_{i1} + \cdots + (a_1, 0, 0)_{LR} x_{ip}$$

와 일치한다. 여기서 $A_k = \left(c_k, l_k, r_k\right)_{LR}$이다. 즉, 퍼지회귀모형 (1.1)과 회귀분석에서 사용하는 회귀모형은 독립퍼지회귀모형 (2.1)의 특별한 경우이다.

이제 독립퍼지회귀모형 (2.1)을 추정하는 방법에 대하여 알아보자. 관측퍼지수 Y_i와 예측퍼지수 $F(\boldsymbol{x_i},\theta)$의 소속함수 μ_Y와 μ_F를 알고 있는 경우에는 예측퍼지수 $F(\boldsymbol{x_i},\theta)$의 중심 $f(\boldsymbol{x_i^c}, C)$와 오른쪽 폭 $h(\boldsymbol{x_i^r}, R)$과 왼쪽 폭 $g(\boldsymbol{x_i^l}, L)$을 추정함으로써 $F(\boldsymbol{x_i},\theta)$의 α-수준집합을 추정할 수 있다. 두 퍼지수 Y_i와 $F(\boldsymbol{x_i},\theta)$의 0-수준집합은 각각 중심과 폭으로 구성되어 있다. 따라서 두 집합 $Y_i(0)$와 $F_0(\boldsymbol{x_i},\theta)$에 대한 차의 크기를 최소화하는 방법으로 독립퍼지회귀모형을 추정할 수 있다. 이를 위해 퍼지수의 중심과 폭을 분리하여 추정한 Savic과 Pedryzc(1991)의 2단계 방법과 퍼지수의 α-수준집합을 추정하기 위해 최소제곱법을 사용한 Kim과 Bishu(1998)의 방법을 이용하자.

독립퍼지회귀모형 (2.1)을 이용하여 관측퍼지수 Y_i에 대한 추정퍼지수 $\widehat{Y_i} = \left(\hat{y}_i, l_{\hat{y}_i}, r_{\hat{y}_i}\right)$를 구하기 위해 다음 2단계 방법을 정의하자.

1단계. 독립퍼지회귀모형의 중심에 대한 최소제곱추정량 $\hat{C} = \left(\hat{c}_0, \hat{c}_1, \cdots, \hat{c}_{p^c}\right)$는

$$\sum_{i=1}^{n} \left(y_i - f(\boldsymbol{x_i^c}, C)\right)^2$$

을 최소화하는 방법으로 구하고, 최소제곱법을 이용하여 추정한 회귀선을 $\hat{y}_i = f(\boldsymbol{x}_i^c, \hat{C})$라 하자.

2단계. 예측퍼지수 $F(\boldsymbol{x}_i, \theta)$의 오른쪽과 왼쪽 폭에 대한 최소제곱추정량 $\hat{R} = \left(\hat{r}_0, \hat{r}_1, \cdots, \hat{r}_{p^r}\right)$와 $\hat{L} = \left(\hat{l}_0, \hat{l}_1, \cdots, \hat{l}_{p^l}\right)$는

$$\sum_{i=1}^{n}\left\{\left(l_{y_i} - g(\boldsymbol{x}_i^l, L)\right)^2 + \left(r_{y_i} - h(\boldsymbol{x}_i^r, R)\right)^2\right\} \tag{2.2}$$

를 최소화하는 방법으로 구하고, 주어진 퍼지회귀모형의 양 폭에 대한 최소제곱추정치는

$$l_{\hat{y}_i} = -\min\{0, k_i\} + g(\boldsymbol{x}_i^l, \hat{L}) \text{와 } r_{\hat{y}_i} = \max\{0, k_i\} + h(\boldsymbol{x}_i^r, \hat{R})$$

와 같이 정의하자. 그리고 모든 i에 대하여 $l_{y_i} = 0$이면 $l_{\hat{y}_i} = 0$ ($r_{y_i} = 0$이면 $r_{\hat{y}_i} = 0$)이고 $k_i = y_i - \hat{y}_i$라 정의하자.

참고: 표본 $\left\{(l_{y_i}, \boldsymbol{x}_i^l) | i = 1, \cdots, n\right\}$와 $\left\{(r_{y_i}, \boldsymbol{x}_i^r) | i = 1, \cdots, n\right\}$을 이용하여 추정한 퍼지수 $F(\boldsymbol{x}_i, \hat{\theta})$의 오른쪽 폭 $h(\boldsymbol{x}_i^r, \hat{R})$과 왼쪽 폭 $g(\boldsymbol{x}_i^l, \hat{L})$은 관측퍼지수 Y_i의 중심 y_i로부터의 거리이다. 그러므로 추정퍼지수 $\hat{Y}_i$의 폭을 두 퍼지수의 중심에 대한 오차 k_i만큼 평행이동함으로써 독립퍼지회귀모형의 정확성을 높일 수 있다. 또한 관찰된 퍼지수중 중심이 비정상적인 자료가 존재하면 추정된 퍼지수 $\hat{Y}_i$의 오른쪽 끝점이 추정된 퍼지수의 중심보다 왼쪽에 위치하거나 왼쪽 끝점이 추정된 퍼지수의 중심보다 오른쪽에 위치할 수 있다. 따라서 퍼지수의 왼쪽 폭 $l_{\hat{y}_i}$과 오른쪽 폭 $r_{\hat{y}_i}$을 각각 $\min\{0, k_i\}$과 $\max\{0, k_i\}$만큼 평행 이동함으로써 추정된 독립퍼지회귀모형의 정확성을 높이고 추정된 퍼지수의 중심과 끝 점이 서로 교차하는 문제를 해결할 수 있다.

함수 f와 g, 그리고 h가 모두 선형이고 $p^c = p^l = p^r = p$와 $x_{ik}^c = x_{ik}^l = x_{ik}^r (k = 0, \cdots, p)$을 만족하면 추정된 퍼지회귀모형은

$$\widehat{Y}_i = \widehat{A}_0 + \widehat{A}_1 x_{i1} + \cdots + \widehat{A}_p x_{ip}$$

이다. 여기서 $\widehat{A}_0 = \left(\hat{c}_0, -\min\{0, k_i\} + \hat{l}_0, \max\{0, k_i\} + \hat{r}_0\right)_{LR}$이고 $\widehat{A}_k = \left(\hat{c}_k, \hat{l}_k, \hat{r}_k\right)_{LR}$이다.

다음 예제는 <표 1>에서 제시된 퍼지자료에 대한 퍼지회귀모형을 추정한다.

예제 2. 설명변수가 실수이고 반응변수가 삼각퍼지수인 독립퍼지회귀모형을 추정하기 위해 $F(\boldsymbol{x}_i, \theta)$를 삼각퍼지수라 하자. <표 1>의 자료는 추정퍼지수 $F(\boldsymbol{x}_i, \theta)$의 양 쪽 폭에 대한 회귀모형은 선형임을 보여준다. 그러나 추정퍼지수 $F(\boldsymbol{x}_i, \theta)$의 중심에 대한

모형은 선형함수나 혹은 지수함수로 가정할 수 있다. 중심에 대한 반응함수를 선형으로 추정한 퍼지회귀모형은

$$\widehat{Y_i^l}=\left(72.155,-\min\left\{0,k_i^l\right\}+1.037,\max\left\{0,k_i^l\right\}+1.037\right)_T+(-0.24,0.029)_T x_i$$

이고 $k_i^l=y_i-72.155+0.24x_i$이다. 같은 방법으로 지수함수를 이용하여 추정한 퍼지회귀모형 $\widehat{Y_i^e}=\left(\widehat{y_i}^e,l_{\widehat{y_i^e}},r_{\widehat{y_i^e}}\right)_T$ 은 다음과 같다.

$$\widehat{y_i}^e=78.3\exp(-0.007x_i)\ ,\ k_i^e=y_i-\widehat{y_i}^e,$$

$$l_{\widehat{y_i^e}}=-\min\left\{0,k_i^e\right\}+1.037+0.029x_i,\ r_{\widehat{y_i^e}}=\max\left\{0,k_i^e\right\}+1.037+0.029x_i.$$

<표 2>는 선형함수와 지수함수를 이용하여 추정한 퍼지수 $\widehat{Y_i^e}$와 $\widehat{Y_i^l}$을 보여준다. 추정된 두 퍼지수 $\widehat{Y_i^e}$와 $\widehat{Y_i^l}$의 폭은 지수함수와 선형함수로 추정된 중심의 영향으로 일치하지 않는다.

<표 2> Jennrich의 자료에 대한 추정량

x_i	$\widehat{Y_i^e}=(\widehat{y_i}^e,l_{\widehat{y_i^e}},r_{\widehat{y_i^e}})_T$	$\widehat{Y_i^l}=(\widehat{y_i}^l,l_{\widehat{y_i^e}},r_{\widehat{y_i^e}})_T$
0	$(78.3,1.04,22.74)_T$	$(72.16,1.04,28.88)_T$
2	$(77.21,1.09,9.89)_T$	$(71.68,1.09,15.42)_T$
4	$(76.14,1.15,3.01)_T$	$(71.2,1.15,7.95)_T$
14	$(70.99,12.43,1.44)_T$	$(68.8,10.24,1.44)_T$
21	$(67.6,16.25,1.65)_T$	$(67.12,15.77,1.65)_T$
42	$(58.36,13.61,2.26)_T$	$(62.08,17.33,2.26)_T$
63	$(50.38,11.24,2.86)_T$	$(57.04,17.9,2.86)_T$
97	$(39.71,3.85,4.14)_T$	$(48.88,12.73,3.85)_T$
155	$(26.46,5.53,14.07)_T$	$(34.96,5.53,5.57)_T$
217	$(17.14,7.33,23.19)_T$	$(20.08,7.33,20.25)_T$

<표 2>는 퍼지회귀모형의 중심과 폭에 대한 회귀모형에 따라 추정된 퍼지회귀모형의 정확성이 변하는 것을 보여준다. 다음 절에서는 추정된 회귀모형의 정확성에 대하여 소개한다.

3. 추정된 퍼지수의 평균과 분산

본 절에서는 독립퍼지회귀모형을 이용하여 추정된 퍼지수에 대한 평균과 분산을 소개하고, 관측퍼지수와 추정퍼지수의 차에 대한 평균과 분산를 구한다.

Carlsson과 Fuller(2001)가 소개한 퍼지수에 대한 평균과 분산을 이용하여 추정된 퍼지수를 순서화나 혹은 비퍼지화할 수 있다. Carlsson과 Fuller(2001)가 제시한 퍼지수에 대한 평균과 분산을 이용하면 LR-퍼지수 $A=(a,l_a,r_a)_{LR}$의 평균과 분산은 각각

$$E(A)=a+2\int_0^1 \alpha\left(r_a R_A^{-1}(\alpha)-l_a L_A^{-1}(\alpha)\right)d\alpha$$

와

$$Var(A)=\frac{1}{2}\int_0^1 \alpha\left(r_a R_A^{-1}(\alpha)+l_a L_A^{-1}(\alpha)\right)^2 d\alpha$$

이다. 두 삼각 퍼지수 $A_1=(a_1,l_1,r_1)_T$와 $A_2=(a_2,l_2,r_2)_T$ 에 대한 평균의 차는

$$E(A_1)-E(A_2)=(a_1-a_2)+\frac{(l_2-l_1)-(r_2-r_1)}{6}$$

이고, 두 퍼지수 차에 대한 분산은

$$Var(A_1-A_2)=\frac{(r_1+l_1+r_2+l_2)^2}{24}$$

이다. 따라서 관측퍼지수 Y_i의 소속함수 μ_Y가 중심에 대하여 대칭이면 즉, $L_Y(x)=R_Y(x)$을 만족하면 추정퍼지수 $\widehat{Y}_i$에 대하여 다음이 성립한다.

(1) $E(\widehat{Y}_i)=f(\boldsymbol{x}_i^c,\widehat{C})+\left(k_i+h(\boldsymbol{x}_i^r,\widehat{R})-g(\boldsymbol{x}_i^l,\widehat{L})\right)\int_0^1 \alpha L_Y^{-1}(\alpha)d\alpha.$

(2) $Var(\widehat{Y}_i)=\frac{1}{2}\left(|k_i|+h(\boldsymbol{x}_i^r,\widehat{R})+g(\boldsymbol{x}_i^l,\widehat{L})\right)^2\int_0^1 \alpha\left(L_Y^{-1}(\alpha)\right)^2 d\alpha.$

(3) $\sum_{i=1}^n E\left(Y_i-\widehat{Y}_i\right)=\sum_{i=1}^n k_i+\sum_{i=1}^n\left(k_i^r-k_i^l-k_i\right)\int_0^1 \alpha L_Y^{-1}(\alpha)d\alpha.$

(4) $\sum_{i=1}^n Var(Y_i-\widehat{Y}_i)=\frac{1}{2}\sum_{i=1}^n\left(2l_{y_i}+2r_{y_i}+|k_i|-k_i^l-k_i^r\right)^2\int_0^1 \alpha\left(L_Y^{-1}(\alpha)\right)^2 d\alpha.$

여기서 $k_i^l=l_{y_i}-g(\boldsymbol{x}_i^l,\widehat{L})$이고 $k_i^r=r_{y_i}-h(\boldsymbol{x}_i^r,\widehat{R})$이다. 따라서 만약 예측퍼지수 $F(\boldsymbol{x}_i,\theta)$을 구성하는 함수 f와 g, 그리고 h가 상수항을 가지면 독립회귀모형을 이용하여 추정한 퍼지수와 관측퍼지수의 차에 대한 평균의 합은 0임을 알 수 있다.

4. 퍼지회귀모형의 정확성

본 절에서는 최소제곱법을 이용하여 추정된 독립퍼지회귀모형의 정확성을 조사하기 위하여 Kim과 Bishu(1998)가 제시한 척도를 사용하고, 예측퍼지수의 중심과 폭에 대한 회귀모형을 구분하지 않고 추정한 퍼지회귀모형과 본 논문에서 추정된 독립퍼지회귀모형에 대한 정확성을 비교한다.

여러 가지 방법으로 추정된 퍼지회귀모형의 정확성을 비교하기 위하여 Kim과 Bishu (1998)는 퍼지수의 소속함수에 대한 적분을 사용하였다. 그들이 제안한 정의를 사용하여 두 퍼지수 $A_1=(a_1,l_1,r_1)_{LR}$와 $A_2=(a_2,l_2,r_2)_{LR}$의 차를 소속함수에 대한 적분과 0-수준집합 사이의 거리를 이용하여 다음과 같이 정의하자.

$$D(A_1,A_2)= inf\{|a_1-a_2| : a_i \in A_i(0)\} + \int_{-\infty}^{\infty} |\mu_{A_1}(x)-\mu_{A_2}(x)| dx. \tag{4.1}$$

식 (4.1)을 이용하여 추정된 퍼지회귀모형에 대한 정확성을 조사하기 위하여 관찰된 퍼지수 Y_i와 추정된 퍼지수 $\widehat{Y}_i$의 차에 대한 척도

$$M(Y,\widehat{Y}) = \sum_{i=1}^{n} m\left(Y_i, \widehat{Y}_i\right) \tag{4.2}$$

를 사용하자. 여기서

$$m\left(Y_i,\widehat{Y}_i\right)=\begin{cases} D\left(Y_i,\widehat{Y}_i\right) & \text{for } \int_{-\infty}^{\infty}\mu_{Y_i}(x)dx<1 \\ \dfrac{D\left(Y_i,\widehat{Y}_i\right)}{\int_{-\infty}^{\infty}\mu_{Y_i}(x)dx} & \text{for } \int_{-\infty}^{\infty}\mu_{Y_i}(x)dx \ge 1 \end{cases}$$

이다. 척도 $M(Y,\widehat{Y})$는 관찰된 퍼지수와 추정된 퍼지수의 소속함수와 두 퍼지수 사이의 거리에 영향을 받는다. 척도 (4.2)의 값이 0에 가까울수록 추정된 퍼지회귀모형의 정확성은 높다고 할 수 있다. 다음 예제는 Jennrich(1995)의 자료에 대한 퍼지회귀모형을 선형함수와 지수함수를 이용하여 추정한 결과의 정확성을 설명한다.

예제 3. <표 1>은 관절염 치료에 사용된 약품에 대한 자료이다. 예제 2에서 설명변수가 실수이고 반응변수가 퍼지수인 자료에 대한 퍼지회귀모형을 추정하였다. 예측퍼지수의 폭에 대한 모형은 선형함수를 사용하였고, 중심에 대한 회귀모형은 지수함수와 선형함수를 이용하였다.

<표 3> 추정된 퍼지모형의 오차와 정확성

x_i	중심에 대한 오차		정확성 척도	
	$\lvert y_i - \hat{y}_i^e \rvert$	$\lvert y_i - \hat{y}_i^l \rvert$	$m\left(Y_i, \hat{Y}_i^e\right)$	$m\left(Y_i, \hat{Y}_i^l\right)$
0	21.7	27.84	11.937	14.842
2	8.79	14.32	6.08	8.98
4	1.86	6.8	1.772	4.628
14	10.99	8.8	5.719	4.829
21	14.6	14.12	5.776	5.628
42	11.36	15.08	3.944	4.875
63	8.38	15.04	2.475	3.803
97	0.29	8.88	0.11	2.215
155	8.54	0.04	1.649	0.012
217	15.86	12.92	2.113	1.811
Total	102.37	123.84	41.575	51.623

지수함수와 선형함수를 사용하여 추정한 값과 관측퍼지수의 중심에 대한 오차가 <표 3>의 왼쪽에 제시되어 있다. 지수함수에 대한 오차 $\lvert y_i - \hat{y}_i^e \rvert$의 합 102.37는 선형함수에 대한 오차 $\lvert y_i - \hat{y}_i^l \rvert$의 합 123.83보다 작다. 따라서 독립퍼지회귀모형의 중심에 대한 회귀모형은 선형함수보다는 지수함수를 사용하는 것이 더 좋을 수 있다.

<표 3>의 오른쪽은 선형함수와 지수함수를 이용하여 추정한 독립퍼지회귀모형의 정확성을 보여준다. 예측퍼지수의 중심에 대한 모형으로 선형함수를 사용하여 추정한 퍼지회귀모형의 척도 $m\left(Y_i, \hat{Y}_i^l\right)$의 합은 51.623이고 지수함수를 이용하여 추정한 모형의 척도 $m\left(Y_i, \hat{Y}_i^e\right)$의 합은 41.575이다. 따라서 예측퍼지수에 대한 중심이나 폭에 대한 적합한 회귀모형을 선택하여 추정된 퍼지회귀모형의 정확성을 높일 수 있다.

추정된 독립퍼지회귀모형의 정확성을 높이기 위해서는 예제 3과 같이 예측퍼지수의 중심과 폭에 대한 적당한 회귀모형을 선택한 후 회귀모형을 설명하는 독립변수를 선택하는 것도 중요하다. 반응변수에 크게 영향을 주지 못하는 변수는 설명변수에서 제거함으로써 추정된 퍼지회귀모형의 정확성을 높일 수 있다. 다음 예제는 추정된 퍼지회귀모형의 정확성은 설명변수의 선택에 따라 영향을 받는 것을 보여준다.

예제 4. Tanaka 등(1982)은 퍼지회귀모형을 설명하기 위해 일본의 집값(Y_i)에 대한 자료를 제시하였다.

<표 4> Tanaka의 자료

$Y_i = \left(y_i, s_{y_i}\right)_T$	x_{i1}	x_{i2}	x_{i3}	x_{i4}	x_{i5}
$(6060,550)_T$	1	38.09	36.43	5	1
$(7100,50)_T$	1	62.10	26.50	6	1
$(8080,400)_T$	1	63.76	44.71	7	1
$(8260,150)_T$	1	74.52	38.09	8	1
$(8650,750)_T$	1	75.38	41.40	7	2
$(8520,450)_T$	2	52.99	26.49	4	2
$(9170,700)_T$	2	62.93	26.49	5	2
$(10310,200)_T$	2	72.04	33.12	6	3
$(10920,600)_T$	2	76.12	43.06	6	2
$(12030,100)_T$	2	90.26	42.64	7	2
$(13940,350)_T$	3	85.70	31.33	7	3
$(14200,250)_T$	3	95.27	27.64	6	3
$(16010,300)_T$	3	105.98	27.64	6	3
$(16320,500)_T$	3	79.25	66.81	6	3
$(16990,650)_T$	3	120.5	32.25	6	3

반응변수 Y_i에 영향을 주는 설명변수는 재료(x_{i1}), 일 층과 이 층의 넓이(x_{i2}, x_{i3}), 그리고 일반 방과 일본의 전통적인 방의 수(x_{i4}, x_{i5})를 사용하였다.

Chang(2001), Kim과 Bishu(1998), 그리고 Tanaka 등(1989)은 <표 4>에 제시된 집값에 대한 퍼지회귀모형을 추정하였다. Chang(2001)과 Kim과 Bishu(1998)이 추정한 퍼지회귀모형은 음수인 폭을 가진다. 모든 설명변수를 사용하여 추정한 Tanaka 등(1989)이 제시한 퍼지회귀모형은

$$\begin{aligned}\widehat{Y}_i = &(-329.75, 429.22)_T + (2268.60, 114.50)_T x_{i1} + (116.20, 2.34)_T x_{i2} \\ &+ (109.38, 0)_T x_{i3} + (-732.80, 0)_T x_{i4} + (-723.43, 18.78)_T x_{i5}\end{aligned}$$

이다. 위 식에서 두 설명변수 x_{i3}와 x_{i4}는 집값의 폭에는 영향을 주지 못함을 암시한다. 따라서 퍼지회귀모형의 폭에 대한 설명변수를 선택함으로써 추정된 모형을 정확성을 높일 수 있다. 회귀분석에서 사용하는 변수선택법과 상관분석을 이용하여 <표

4>에 제시된 자료에 대한 중심과 폭에 대한 회귀모형은

$$f(\boldsymbol{x_i^c}, C) = c_0 + c_1 x_{i1} + c_2 x_{i2} + c_3 x_{i3} + c_4 x_{i4} \text{와 } h(\boldsymbol{x_i^r}, R) = g(\boldsymbol{x_i^l}, L) = l_0 + l_1 x_{i3} \tag{4.3}$$

이다. 최소제곱법을 이용하여 회귀모형 (4.3)을 추정한 결과는 다음과 같다.

$$f(\boldsymbol{x_i^c}, \widehat{C}) = -1237.1 + 2140.6 x_{i1} + 94.2 x_{i2} + 81.4 x_{i3} - 343.1 x_{i4}$$

$$.h(\boldsymbol{x_i^r}, \widehat{R}) = g(\boldsymbol{x_i^l}, \widehat{L}) = 292.1 + 2.99 x_{i3}.$$

<표 5> 추정된 퍼지모형의 정확성

i	정확성 척도	
	$m(Y_i, \widehat{Y_i^t})$	$m(Y_i, \widehat{Y_i})$
1	0.34	0.723
2	13.27	10.312
3	1.32	0.258
4	3.94	0.484
5	0.01	0.514
6	1.39	0.821
7	0.37	0.488
8	4.65	1.54
9	0.81	0.809
10	9.47	6.657
11	2.83	0.721
12	3.28	1.012
13	3.68	2.214
14	1.06	0.739
15	1.20	0.373
Total	47.62	27.665

<표 5>는 Tanaka 등(1989)이 추정한 퍼지회귀모형과 독립퍼지회귀모형에 대한 척도 $m(Y_i, \widehat{Y_i^t})$와 $m(Y_i, \widehat{Y_i})$의 합을 보여준다. 예측퍼지수의 중심과 폭에 대한 설명변수를 선택하여 추정한 독립퍼지회귀모형에 대한 척도 $m(Y_i, \widehat{Y_i})$의 합은 27.665로 Tanaka 등(1989)이 추정한 퍼지회귀모형에 대한 척도 $m(Y_i, \widehat{Y_i^t})$의 합 47.62보다 작다. 이것은 예측퍼지수의 중심과 폭에 대한 설명변수를 적당한 방법으로 선택하는 것이 추정된 퍼지회귀모형의 정확성을 높일 수 있음을 설명한다.

5. 결론

본 논문에서는 회귀계수와 반응변수가 퍼지수인 퍼지회귀모형에서 관찰된 퍼지수의 중심과 폭에 대한 반응함수가 일치하지 않을 수 있음을 확인하고, 추정된 퍼지수의 중심과 폭에 대한 반응함수를 서로 구분한 독립퍼지회귀모형을 소개하였다. 최소제곱법을 이용하여 추정된 독립퍼지회귀모형의 정확성은 예측퍼지수의 중심과 폭에 대한 반응함수의 선택과 반응함수에 대한 설명변수의 선택에 의존함을 확인하였다. 또한 관측퍼지수와 추정된 퍼지수의 차에 대한 평균과 분산을 조사하였다. 그리고 본 논문에서 제시된 독립퍼지회귀모형은 Tanaka 등(1989)이 추정한 퍼지회귀모형보다 더 정확할 수 있음을 예제를 통하여 확인하였다.

앞으로 일반적인 퍼지회귀모형에 대한 정확성을 유도하기 위해 독립퍼지회귀모형에 대한 통계적인 성질을 연구할 필요가 있다.

참고 문헌

1. Carlsson, C. and Fuller, R. (2001). On possibilistic mean value and varience of fuzzy number, *Fuzzy Sets and Systems*, Vol. 122, 315-326.
2. Chang, Y. (2001). Hybrid fuzzy least-squares regression analysis and its reliability measures, *Fuzzy Sets and Systems*, Vol. 119, 225-246.
3. Chang, Y. O. and Ayyub, B. M. (2001). Fuzzy regression methods-a comparative assessment. *Fuzzy Sets and Systems*, Vol. 119, 187-203.
4. Diamond, P. (1988). Fuzzy least squares. *Information Sciences*, Vol 46, 141-157.
5. Diamond, P. and Korner, R. K. (1997). Extended fuzzy linear models and least-squares estimates. *Journal of Computational and Applied Mathematics*, Vol. 9, 15-32.
6. Dubios, D. and Prade, H. (1979). Fuzzy real algebra and some results, *Fuzzy Sets and Systems*, Vol. 2, 327-348.
7. Jennrich, R. I. (1995), *An introduction to computational statistics - regression analysis*, Prentice-Hall International, Inc.
8. Kao, C. and Chyu, C. (2002). A fuzzy linear regression model with better explanatory power, *Fuzzy Sets and Systems*, Vol. 126, 401-409.
9. Kao, C. and Chyu, C. (2003). Least Squares estimates in fuzzy regression analysis, *European Journal of Operational Research*, Vol. 148, 420-435.
10. Kim, B. and Bishu, R. R. (1998). Evaluation of fuzzy linear regression models by comparing membership functions. *Fuzzy Sets and Systems*, Vol. 100, 343-352.
11. Sakawa, M. and Yano, M. (1992). Multiobjective fuzzy linear regression analysis for fuzzy input-output data. *Fuzzy Sets and Systems*, Vol. 47, 173-181.
12. Savic, D. and Pedryzc, W. (1991). Evaluation of fuzzy linear regression

models. *Fuzzy Sets and Systems*, Vol. 39, 51-63.
13. Tanaka, H., Hayashi, I. and Watada, J. (1989). Possibilistic linear regression analysis for fuzzy data. *European Journal of Operational Research*, Vol. 40, 389-396.
14. Tanaka, H., Uejima, S. and Asai, K. (1980). Fuzzy linear regression model. *International Congress Applied Systems and Cybernetics*, Vol 4, 2933-2938.
15. Tanaka, H., Uejima, S. and Asai, K. (1982). Linear regression analysis with fuzzy model. *IEEE Transaction on Systems*, Man, and Cybernetics, Vol. 12, 903-907.
16. Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, Vol 8, 338-353.

[2007년 2월 접수, 2007년 3월 채택]