Marian MATŁOKA

PRZEDZIAŁOWE I ROZMYTE MODELE REGRESJI

Wstęp

W języku ekonomicznym, w celu określenia siły i natężenia występowania pewnych zjawisk, często posługujemy się zwrotami „duży zysk", „niewielki obrót", „około 100 zł" itp. Opisanie tego typu wielkości za pomocą konwencjonalnego aparatu matematycznego jest niemożliwe. Często w takich przypadkach stosowana jest teoria zbiorów rozmytych i arytmetyki przedziałowej.

Rozmyta i przedziałowa analiza regresji jest techniką modelowania opisującą zależności między zmiennymi w wypadku, gdy związek ten jest rozmyty lub przedziałowy. Wiadomo, że w modelu ekonometrycznym pomiędzy wartościami rzeczywistymi a teoretycznymi zmiennej objaśnianej występują pewne odchylenia. Ujmowane są one w postaci składnika losowego. W modelach „rozmytym i przedziałowym rozmyte i przedziałowe parametry powodują rozmycie" wartości zmiennej objaśnianej, które uwzględniają już wszelkie czynniki losowe. Dzięki temu w rozmytych i przedziałowych modelach ekonometrycznych możemy pominąć składnik losowy.

W artykule dużo uwagi poświęcono modelom przedziałowym, które stanowią pewien szczególny przypadek modeli rozmytych i jednocześnie są punktem wyjścia do analizy regresji rozmytej. Jest to ogólnie występujący w literaturze kierunek badań[1] wynikający głównie z powodów obliczeniowych. Arytmetyka przedziałowa[2] wielokrotnie jest podstawą wykonania działań arytmetycznych na liczbach rozmytych.

[1] J. Kacprzyk, M. Fedrizzi, *Fuzzy regression analysis*, Omnitech Press, Warsow and Physica – Verlag, Heidelberg 1992; W. Łuczyński, M. Matłoka, *Fuzzy regression models and their applications*, The Journal of Fuzzy Mathematics, vol. 3, 1995, no. 3.

[2] G. Alefeld, J. Herzberger, *Introduction to interval computations*, Academic Press, New York 1983; R. E. Moore, *Methods and applications of interval analysis*, SIAM, Philadephia 1979.

Niniejszy artykuł stanowi pewne uogólnienie wyników otrzymanych w pracy M. Matłoki i W. Łuczyńskiego[3]. Prezentowane modele ilustrowane są przykładami liczbowymi w celu pokazania istoty problemu.

1. Pojęcia wstępne

W teorii zbiorów rozmytych liczbę rozmytą $\tilde{A}$ definiuje się między innymi jako funkcję $\tilde{A}: R \to \langle 0, 1\rangle$ spełniającą następujące warunki:

- istnieje $t_0 \in R$ takie, że $\tilde{A}(t_0) = 1$,
- istnieje zbiór zwarty $K \subset R$ taki, że

$$\tilde{A}^0 = \{t: \tilde{A}(t) > 0\} \subset K,$$

- dla dowolnego $\beta \in (0, 1\rangle$ zbiór $\tilde{A}^\beta = \{t: \tilde{A}(t) \geqslant \beta\}$ zwany β-obcięciem, jest zbiorem wypukłym,
- funkcja $\tilde{A}$ jest półciągła z góry, tzn. zbiór $\{t: \tilde{A}(t) < \beta\}$ jest otwarty dla każdego $\beta \in \langle 0, 1\rangle$.

Z powyższej definicji wynika, że dla dowolnego $\beta \in \langle 0, 1\rangle$, $\tilde{A}^\beta$, jest zwartym i wypukłym przedziałem. Fakt ten jest bardzo ważny z punktu widzenia obliczeniowego. Każdy zwarty i wypukły przedział liczbowy A możemy utożsamić z jego środkiem a_c oraz promieniem a_w. Wtedy stosujemy zapis

$$A = \langle a_c, a_w\rangle = \{t \in R: a_c - a_w \leqslant t \leqslant a_c + a_w\}.$$

Mając dwa przedziały liczbowe $A = \langle a_c, a_w\rangle$ oraz $B = \langle b_c, b_w\rangle$, możemy zdefiniować:

1) dodawanie przedziałów:

$$A + B = \langle a_c + b_c, a_w + b_w\rangle,$$

2) mnożenie przedziału przez liczbę rzeczywistą:

$$k \cdot A = \langle k \cdot a_c, |k| \cdot a_w\rangle.$$

---

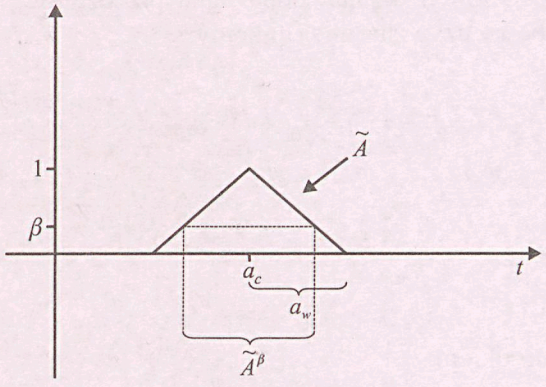
[3] W. Łuczyński, M. Matłoka, op. cit.

Operacje dodawania liczb rozmytych oraz mnożenia liczby rozmytej przez liczbę rzeczywistą wprowadza się na podstawie powyższych definicji, a mianowicie są one przeprowadzane na β-obcięciach liczb rozmytych.

Ze względów obliczeniowych bardzo wygodnymi liczbami rozmytymi są symetryczne, trójkątne liczby rozmyte, które określa się przez środek i rozrzut $\langle a_c, a_w \rangle_L$ oraz funkcję przynależności

$$\tilde{A}(t) = \max \left\{ 1 - \frac{|t - a_c|}{a_w}, 0 \right\}.$$

Przykładem takiej liczby rozmytej jest liczba przedstawiona na poniższym wykresie.



Nietrudno wykazać, że każde β-obcięcie tej symetrycznej, trójkątnej liczby rozmytej jest przedziałem liczbowym o środku a_c i promieniu $(1-\beta)a_w$. Zauważmy jednocześnie, że mając β-obcięcie pewnej symetrycznej, trójkątnej liczby rozmytej nietrudno tę liczbę znaleźć.

2. Regresja przedziałowa

Analiza modeli przedziałowych, czyli modeli regresji z przedziałowymi wartościami parametrów jest punktem wyjścia do analizy regresji rozmytej. Uzasadnieniem wprowadzenia parametrów przedziałowych jest niepełne odzwierciedlanie niejednoznaczności wpływu zmiennych objaśniających na kształtowanie się zmiennej objaśnianej.

Załóżmy, że:

- mamy m obserwacji (x_j, y_j), $i=1, 2, \ldots, m$;
- $A_i=\langle a_{ic}, a_{iw}\rangle$ są przedziałowymi wartościami parametrów modelu określonymi przez środek a_{ic} oraz promień a_{iw}, $i=1, 2, \ldots, n$.

Rozpatrzmy model przedziałowy[4] zapisany w następującej postaci

$$Y(x)=A_0+A_1\cdot x+A_2\cdot x^2+\ldots+A_n\cdot x^n.$$

Zauważmy, że dla dowolnego x_j mamy

$$Y(x_j)=\langle a_{0c}, a_{0w}\rangle+\langle a_{1c}, a_{1w}\rangle x_j+\ldots+\langle a_{nc}, a_{nw}\rangle x_j^n=\langle y_c(x_j), y_w(x_j)\rangle.$$

Na podstawie definicji działań dodawania przedziałów i mnożenia przedziału przez liczbę rzeczywistą otrzymujemy:

$$y_c(x_j)=a_{0c}+\sum_{i=1}^{n} a_{ic}\cdot x_j^i,$$

$$y_w(x_j)=a_{0w}+\sum_{i=1}^{n} a_{iw}\cdot |x_j^i|,$$

gdzie

- $y_c(x_j)$ jest środkiem,
- $y_w(x_j)$ jest promieniem

szacowanego przedziału $Y(x_j)$.

Problem polega na znalezieniu takich wartości teoretycznych modelu (będących przedziałami), aby zawierały odpowiadające im wartości rzeczywiste i jednocześnie by przedziały te były jak najwęższe. Oznacza to, że należy rozwiązać następujące zagadnienie programowania matematycznego:

$$\sum_{j=1}^{m} y_w(x_j)\to\min$$

przy warunkach

$$y_j\in Y(x_j), \qquad j=1, 2, \ldots, m,$$

$$a_{iw}\geq 0, \qquad i=0, 1, 2, \ldots, n.$$

---

[4] W. Łuczyński, M. Matłoka, op. cit.

Biorąc pod uwagę postaci $y_w(x_j)$ oraz $Y(x_j)$, powyższe zagadnienie staje się następującym zadaniem programowania liniowego:

$$\sum_{j=1}^{m} (a_{0w} + a_{1w}|x_j| + a_{2w}|x_j^2| + \ldots + a_{nw}|x_j^n|) \to \min$$

przy warunkach

$$\sum_{i=0}^{n} a_{ic}x_j^i + \sum_{i=0}^{n} a_{iw}|x_j^i| \geqslant y_j \qquad j = 1, 2, \ldots, m,$$

$$\sum_{i=0}^{n} a_{ic}x_j^i - \sum_{i=0}^{n} a_{iw}|x_j^i| \geqslant y_j \qquad j = 1, 2, \ldots, m,$$

$$a_{iw} \geqslant 0, \qquad i = 0, 1, 2, \ldots, n.$$

Przykład 1

Rozpatrzmy dane zamieszczone w poniższej tabeli[5]

x_j	1	2	3	4	5	6	7	8
y_j	−37,2	−33,9	−23,5	−26,7	−15,9	−23,8	−12,3	−13,1

Rozwiązując przedstawione wcześniej zagadnienie programowania liniowego dla powyższych danych oraz $n = 1$, otrzymujemy następujący liniowy model przedziałowy:

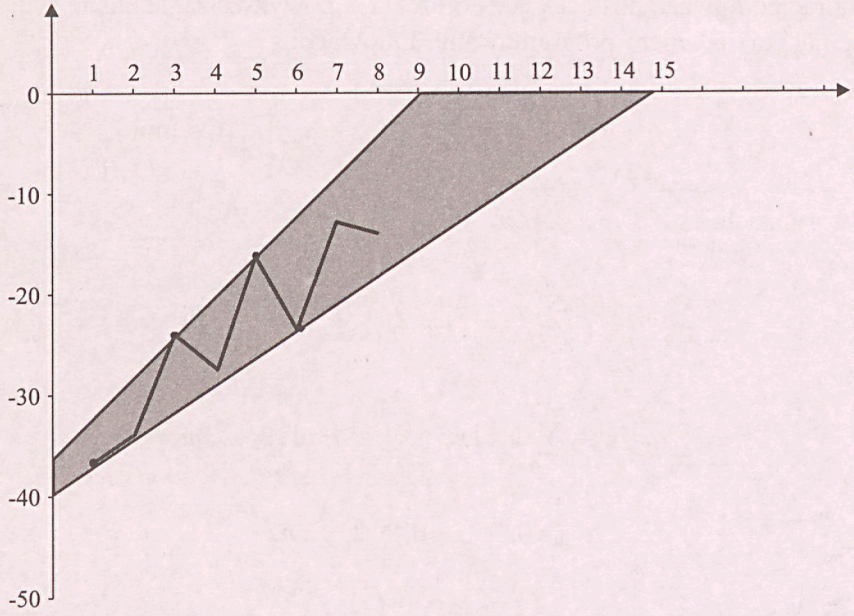
$$Y(x_j) = \langle -37{,}39;\ 2{,}49\rangle + \langle 3{,}24;\ 0{,}56\rangle \cdot x_j.$$

Jego ilustrację graficzną przedstawia poniższy wykres.

Jak widać, parametry $A_i = \langle a_{ic}, a_{iw}\rangle$ modelu przedziałowego otrzymujemy poprzez znalezienie optymalnego rozwiązania powyższego zagadnienia programowania liniowego.

Omówiony model przedziałowy można uogólnić poprzez przyjęcie założenia, że mamy m obserwacji (x_j, Y_j), $j = 1, 2, \ldots, m$, gdzie obserwacja zmiennej objaśnianej jest przedziałem $Y_j = \langle y_{jc}, y_{jw}\rangle$ o środku y_{jc} oraz promieniu y_{jw}.

[5] Ibidem.



Przy tego typu założeniu w dalszym ciągu problem polega na wyznaczeniu przedziałowych wartości parametrów A_i modelu. Jednakże w tym przypadku możliwe są dwa podejścia.

Pierwsze z nich polega na przyjęciu założenia, że budowany przedział $Y(x_j)$ musi w całości pokrywać rzeczywiste wartości obserwacji Y_j, tzn. $Y_j \subseteq Y(x_j)$. Wtedy dla większej precyzji estymacji należałoby minimalizować ich szerokość. Oznacza to, że mamy do czynienia z następującym zagadnieniem programowania liniowego:

$$\sum_{j=1}^{m} y_w(x_j) \to \min$$

przy warunkach

$$Y_j \subseteq Y(x_j), \quad j=1, 2, \ldots, m,$$

$$a_{iw} \geqslant 0, \quad i=0, 1, 2, \ldots, n$$

lub

$$\sum_{j=1}^{m} (a_{0w} + a_{1w}|x_j| + a_{2w}|x_j^2| + \ldots + a_{nw}|x_j^n|) \to \min$$

przy warunkach

$$\sum_{i=0}^{n} a_{ic} x_j^i + \sum_{i=0}^{n} a_{iw} |x_j^i| \geqslant y_{jc} + y_{jw}, \quad j = 1, 2, \ldots, m,$$

$$\sum_{i=0}^{n} a_{ic} x_j^i - \sum_{i=0}^{n} a_{iw} |x_j^i| \leqslant y_{jc} - y_{jw}, \quad j = 1, 2, \ldots, m,$$

$$a_{iw} \geqslant 0, \quad i = 0, 1, 2, \ldots, n.$$

I w tym przypadku parametry A_i modelu przedziałowego są optymalnym rozwiązaniem powyższego zagadnienia programowania liniowego.

Przykład 2

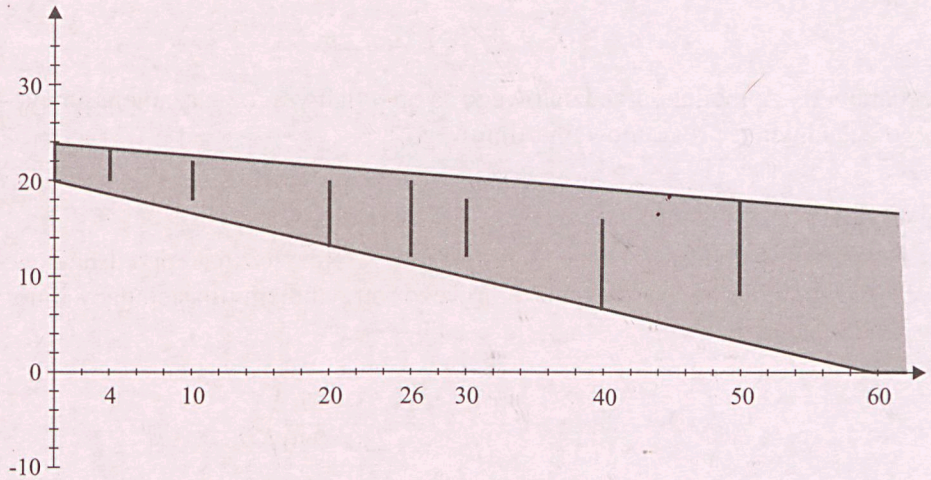
Rozpatrzmy dane zamieszczone w poniższej tabeli.

x_j	4	10	20	26	30	40	50
Y_j	(20; 23,52)	(18; 22)	$\left(13\frac{1}{3}; 20\right)$	(12; 20)	(12; 18)	$\left(6\frac{2}{3}; 16\right)$	(8; 18)

Rozwiązując przedstawione powyżej zagadnienie programowania liniowego dla powyższych danych oraz $n = 1$, otrzymujemy następujący liniowy model przedziałowy:

$$Y(x_j) = \langle 22; 2 \rangle + \left\langle -\frac{34}{150}; \frac{16}{150} \right\rangle \cdot x_j.$$

Jego ilustrację graficzną przedstawia poniższy wykres.



Zauważmy, że przedstawiony model przedziałowy zawiera wszystkie wyniki obserwacji zmiennej objaśnianej.

Drugie z podejść polega na przyjęciu założenia, że oszacowane przedziały nie wykraczają poza rzeczywiste wartości danych, tzn. $Y(x_j) \subseteq Y_j$. W tym przypadku należy maksymalizować ich szerokość tak, aby objęły największą część przedziałowych wartości obserwacji zmiennej objaśnianej. Oznacza to, że mamy do czynienia z następującym zagadnieniem programowania liniowego:

$$\sum_{j=1}^{m} y_w(x_j) \to \max$$

przy warunkach

$$Y(x_j) \subseteq Y_j, \quad j=1, 2, \ldots, m,$$

$$a_{iw} \geqslant 0, \quad i=0, 1, 2, \ldots, n$$

lub

$$\sum_{j=1}^{m} (a_{0w} + a_{1w}|x_j| + a_{2w}|x_j^2| + \ldots + a_{nw}|x_j^n|) \to \max$$

przy warunkach

$$\sum_{i=0}^{n} a_{ic} x_j^i + \sum_{i=0}^{n} a_{iw}|x_j^i| \leqslant y_{jc} + y_{jw}, \quad j=1, 2, \ldots, m,$$

$$\sum_{i=0}^{n} a_{ic} x_j^i - \sum_{i=0}^{n} a_{iw}|x_j^i| \geqslant y_{jc} - y_{jw}, \quad j=1, 2, \ldots, m,$$

$$a_{iw} \geqslant 0, \quad i=0, 1, 2, \ldots, n.$$

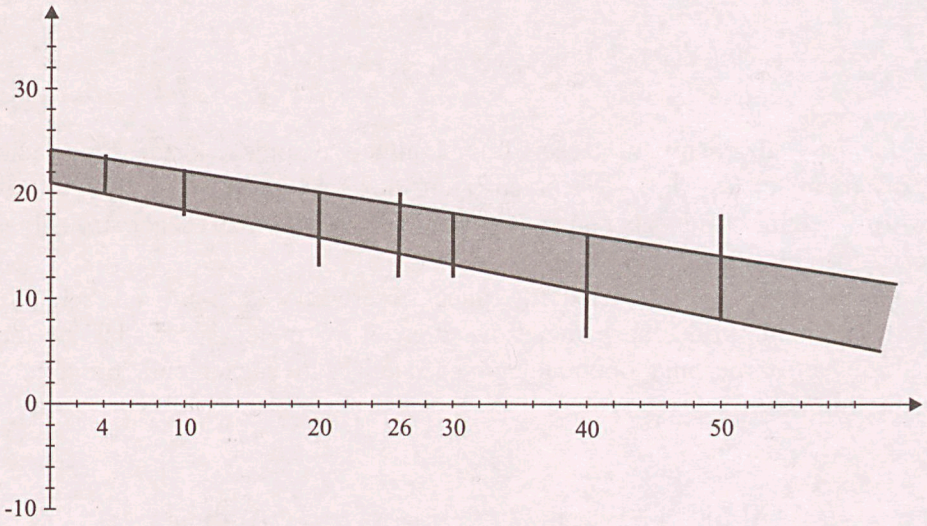
Parametry A_i modelu przedziałowego są optymalnym rozwiązaniem powyższego zagadnienia programowania liniowego.

Przykład 3

Rozpatrzmy ponownie dane z przykładu 2. Rozwiązując przedstawione wyżej zagadnienie programowania liniowego, otrzymujemy następujący liniowy model przedziałowy:

$$Y(x_j) = \left\langle 22\frac{7}{13}; 1\frac{6}{13} \right\rangle + \left\langle -\frac{53}{230}; \frac{7}{230} \right\rangle x_j.$$

Jego ilustrację graficzną przedstawia poniższy wykres.



Wykres pokazuje, że wyniki obserwacji zmiennej objaśnianej zawierają wartości teoretyczne liniowego modelu przedziałowego. Zatem przyjęte założenia są spełnione.

3. Regresja rozmyta

Model regresji rozmytej przyjmuje postać

$$\tilde{Y}(x)=\tilde{A}_0+\tilde{A}_1 x+\ldots+\tilde{A}_n x^n,$$

gdzie $\tilde{A}_i$, $i=0, 1, \ldots, n$, są liczbami rozmytymi.

Ze względów obliczeniowych wygodnie jest przyjąć, że $\tilde{A}_i$ są symetrycznymi, trójkątnymi liczbami rozmytymi, określonymi przez ich środki i tzw. rozrzuty $(a_{ic}, a_{iw})_L$ o następującej funkcji przynależności

$$\tilde{A}(t)=\max\{1-|t-a_{ic}|/a_{iw};\ 0\}.$$

Stosując zasadę rozszerzania Zadeha, model rozmyty możemy zapisać w postaci:

$$\tilde{Y}(x_j)=\left(\sum_{i=0}^{n} a_{ic}\, x_j^i,\ \sum_{i=0}^{n} a_{iw}\,|x_j^i|\right)_L,$$

co oznacza, że $\tilde{Y}(x_j)$ jest również trójkątną, symetryczną liczbą rozmytą.

Ponadto nietrudno wykazać, że dla dowolnego $\beta \in (0, 1\rangle$ model rozmyty można utożsamić z modelem przedziałowym

$$\tilde{Y}(x_j)^{\beta} = \tilde{A}_0^{\beta} + \tilde{A}_1^{\beta} x_j + \tilde{A}_2^{\beta} x_j^2 + \ldots + \tilde{A}_n^{\beta} x_j^n.$$

Jak powiedzieliśmy w części drugiej niniejszej pracy, każde β-obcięcie liczby rozmytej $(a_{ic}, a_{iw})_L$ jest przedziałem liczbowym $\langle a_{ic}; (1-\beta)\, a_{iw}\rangle$. Zauważmy jednocześnie, że mając β-obcięcie pewnej symetrycznej, trójkątnej liczby rozmytej, nietrudno tę liczbę znaleźć.

Oznacza to, że gdy będziemy mieć obserwacje (x_j, y_j), $j = 1, 2, \ldots, m$, wtedy problem szukania parametrów regresji rozmytej sprowadzi się do znalezienia rozwiązania optymalnego następującego zagadnienia programowania liniowego:

$$\sum_{j=0}^{m} (a_{0w} + a_{1w}|x_j| + a_{2w}|x_j^2| + \ldots + a_{nw}|x_j^n|) \to \min$$

przy warunkach

$$\sum_{i=0}^{n} a_{ic} x_j^i + (1-\beta) \sum_{i=0}^{n} a_{iw} |x_j^i| \geqslant y_j, \qquad j = 1, 2, \ldots, m,$$

$$\sum_{i=0}^{n} a_{ic} x_j^i - (1-\beta) \sum_{i=0}^{n} a_{iw} |x_j^i| \leqslant y_j, \qquad j = 1, 2, \ldots, m,$$

$$a_{iw} \geqslant 0, \qquad i = 0, 1, 2, \ldots, n.$$

Gdy będziemy dysponować danymi obserwacji (x_j, Y_j), gdzie Y_j jest przedziałem, wtedy niezależnie od podejść rozpatrywanych w części trzeciej niniejszego artykułu problem rozwiązuje się analogicznie.

Załóżmy teraz, że mamy m obserwacji $(x_j, \tilde{Y}_j)$, gdzie obserwacja zmiennej objaśnianej jest trójkątną, symetryczną liczbą rozmytą. W zależności od podejścia, a więc czy zakładamy, że $\tilde{Y}_j \subseteq \tilde{Y}(x_j)$ lub $\tilde{Y}_j(x_j) \subseteq \tilde{Y}_j$, będziemy dokonywać minimalizacji lub maksymalizacji. W warunkach ograniczających z kolei należy uwzględnić β-obcięcia liczb $\tilde{Y}_j$.

Przykład 4

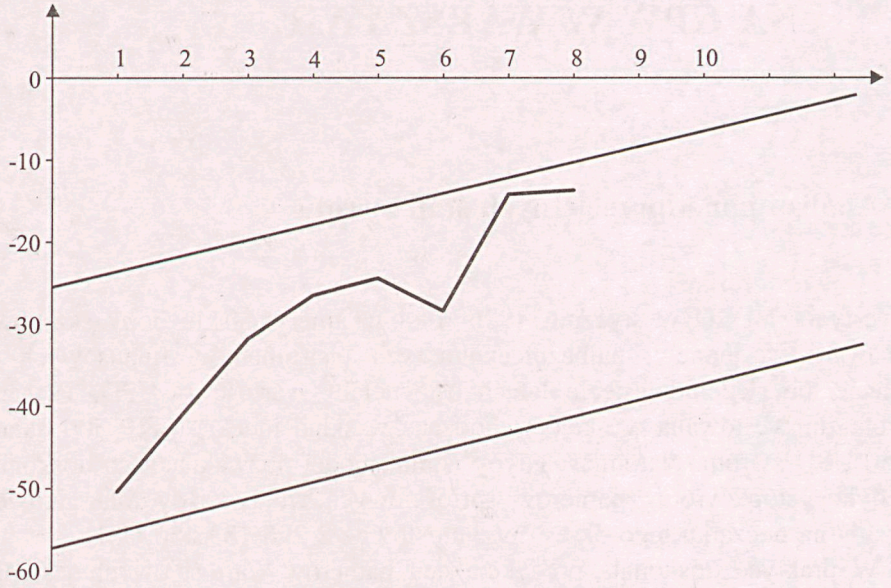
Rozpatrzmy dane zamieszczone w poniższej tabeli[6]

x_j	1	2	3	4	5	6	7	8
y_j	−50,4	−41	−31,8	−26,5	−24,4	−28,4	−14,2	−13,7

Dla $n=3$ oraz $\beta=0{,}5$ otrzymujemy następujący model przedziałowy:

$$\tilde{Y}(x_j)^{0,5} = \langle -44{,}142312;\ 16{,}46374\rangle + \langle 1{,}996443;\ 0\rangle x_j + \langle 0;\ 0{,}044521\rangle x_j^3.$$

Jak widać na poniższym wykresie wszystkie dane zawarte są w 0,5-obcięciu modelu rozmytego.



Literatura

Alefeld G., Herzberger J., *Introduction to interval computations*, Academic Press, New York 1983

Kacprzyk J., Fedrizzi M., *Fuzzy regression analysis*, Omnitech Press, Warsow and Physica – Verlag, Heidelberg 1992

Łuczyński W., Matłoka M., *Fuzzy regression models and their applications*, The Journal of Fuzzy Mathematics, vol. 3, 1995, no. 3

Moore R. E., *Methods and applications of interval analysis*, SIAM, Philadephia 1979

[6] Ibidem.