

Józef Czaja \*, Edward Preweda \*

## KILKA UWAG DO ANALIZY ROZKŁADU PRAWDOPODOBIENSTWA INFORMACJI RYNKOWYCH \*\*

### 1. Studium pojęć

W ostatnim okresie środowisko rzeczoznawców majątkowych często podejmuje dyskusje, dotyczące modelowania wielkości losowych w aspekcie wyceny nieruchomości. Wielu autorów w swoich publikacjach rozważa, jaki charakter mają informacje rynkowe oraz możliwości zastosowania do tych informacji modelowania statystycznego.

Na początku tych rozważań trzeba dać odpowiedź na pytanie, czy informacje dotyczące rynku nieruchomości, ujmujące ceny i atrybuty, mają charakter losowy. Odpowiedź na to pytanie jest jednoznaczna: każda informacja dotycząca zjawiska społecznego (formułowanego przez człowieka) ma charakter losowy, zatem każdy zbiór informacji (próbka losowa) o cenach i atrybutach nieruchomości, bez względu na jego liczebność, ma cechy zmiennej losowej wielowymiarowej. Jeżeli będziemy rozpatrywać zawężony rynek lokalny, na którym wszystkie nieruchomości z obrotu rynkowego posiadają identyczne atrybuty, to wtedy również mamy do czynienia ze zmienną losową, ale **jednowymiarową**.

Zasadniczą cechą zmiennej losowej opisującej rynek nieruchomości jest jej **rozkład prawdopodobieństwa**, czyli częstości pojawiania się ustalonych cen i wartości atrybutów nieruchomości na tym rynku. Na ogół informacje o poszczególnych nieruchomościach (analizowanych pojedynczo) mają identyczne prawdopodobieństwo, czyli **rozkład prawdopodobieństwa równomierny**. Oznacza to, że dla zbioru zawierającego

---

\* Akademia Górniczo-Hutnicza, Wydział Geodezji Górniczej i Inżynierii Środowiska

\*\* Pracę realizowano w ramach Badań Statutowych w Katedrze Informacji o Terenie

$n$  nieruchomości prawdopodobieństwo pojawienia się każdej nieruchomości w tym zbiorze (na rynku) jest jednakowe i wynosi

$$p = \frac{1}{n}.$$

Rzeczoznawca, dokonując analizy rynku w aspekcie zmiennej losowej **jednowymiarowej**, może grupować rozpatrywane nieruchomości według ustalonych przez siebie kryteriów, na przykład według przedziałów cenowych. W zależności od wyboru szerokości przedziałów cenowych w podgrupach rzeczoznawca może uzyskać różne częstości (prawdopodobieństwa) pojawienia się nieruchomości w tych przedziałach. Zestawienie tabelaryczne przedziałów cenowych i częstości występowania cen w tych przedziałach (ilorazów liczby nieruchomości wchodzących do przedziałów i ogólnej liczby nieruchomości analizowanego zbioru) nosi nazwę **funkcji rozkładu prawdopodobieństwa cen** na analizowanym rynku. Graficznym obrazem tej funkcji jest **histogram rozkładu prawdopodobieństwa (częstości) cen** nieruchomości na analizowanym rynku. Kształt takiego histogramu w dużym zakresie zależy od doboru przez rzeczoznawcę przedziałów cenowych dla rozważanych nieruchomości. Każda analiza rynku nieruchomości powinna być odniesiona do funkcji rozkładu prawdopodobieństwa cen i jej histogramu.

Wyrażenie „**rozkład prawdopodobieństwa zmiennej losowej**” może być rozpatrywane w różnych znaczeniach (pojęciach), które mają różne implikacje. Rzeczoznawca w procesie szacowania nieruchomości powinien wyraźnie określić, w jakim znaczeniu rozpatruje rozkład prawdopodobieństwa cen rynkowych.

W dalszych rozważaniach omówione będą, w formie syntetycznej, trzy różne znaczenia dotyczące **rozkładu prawdopodobieństwa cen nieruchomości** jako zmiennej jednowymiarowej:

- 1) rozkład prawdopodobieństwa wyróżnionej cechy nieruchomości,
- 2) rozkład prawdopodobieństwa wyników próby,
- 3) rozkład prawdopodobieństwa statystyki zdefiniowanej na próbie.

**Ad 1) Rozkład prawdopodobieństwa wyróżnionej cechy** reprezentującej zmienną losową (np. cen ustalonego typu nieruchomości) **w całej populacji** (dla całego miasta) – w tym znaczeniu **rozkład populacji jest definiowany a priori za pomocą jego parametrów w formie hipotezy statystycznej**. Na przykład można założyć, że rozkład cen nieruchomości na analizowanym rynku ma parametry rozkładu normalnego czyli  $N(\mu, \sigma)$ . Tak sformułowane założenie nosi nazwę statystycznej hipotezy parametrycznej. Również można zakładać, że rozkład cen nieruchomości na analizowanym rynku ma cechy (postać) rozkładu chi-kwadrat. W takim przypadku sformułowane założenie nosi nazwę statystycznej hipotezy nieparametrycznej. W obu przypadkach należy zweryfikować postawione hipotezy statystyczne za pomocą odpowiednich testów parametrycznych lub za pomocą testów zgodności (nieparametrycznych).

W aspekcie wyceny nieruchomości nie jest konieczna znajomość rozkładu prawdopodobieństwa cen w całej populacji, a to oznacza, że rzeczoznawca nie musi stawiać hipotez statystycznych i nie musi dokonywać ich weryfikacji.

**Ad 2) Rozkład prawdopodobieństwa wyników w ramach dokonanej próby** (wyboru nieruchomości) – w tym znaczeniu **rozkład wyników ma charakter empiryczny** (doświadczalny), a każda nowa próba niesie nowe informacje dotyczące rozkładu prawdopodobieństwa. Najczęściej rzeczoznawca zainteresowany jest tym, w jakim stopniu analizowany rozkład wyników otrzymany w próbie może być zbliżony do rozkładu normalnego. Ogólnie można stwierdzić, że rozkład normalny jest dobrym modelem dla rozkładu zmiennej losowej mającej następujące cechy:

- występuje silna tendencja do przyjmowania wartości położonych blisko środka rozkładu (wartości przeciętnej);
- dodatnie i ujemne odchylenia od wartości środkowej są prawdopodobnie równe;
- liczebność odchyłeń gwałtownie spada w miarę oddalania się w obie strony od wartości środkowej.

Proste statystyki (parametry) opisowe, zdefiniowane za pomocą momentów absolutnych  $2r$  – rzędu ( $\mu_{2r}$ ) i odchylenia standardowego ( $\sigma$ ) wyliczonych z wyników próby, pozwalają analitycznie dowiedzieć czy wyniki otrzymane w próbie mają niektóre cechy rozkładu normalnego. Na przykład, jeśli parametr asymetrii

$$\gamma_1 = \mu_3 \sigma^{-3} \approx 0$$

jest bliski wartości zero, to rozkład wyników próby jest rozkładem symetrycznym, czyli zgodnym w tym zakresie z rozkładem normalnym. Gdy parametr spłaszczenia

$$\gamma_2 = \mu_4 \sigma^4 - 3 \approx 0$$

będzie bliski wartości zero, wtedy rozkład wyników próby posiada spłaszczenie podobne do spłaszczenia rozkładu normalnego, czyli równego zero. Dokładniejsze wnioskowanie o zgodności wyników próby z rozkładem normalnym można uzyskać na podstawie testów zgodności. Testy takie pozwalają uzyskać informacje jakie jest prawdopodobieństwo, że próba pochodzi z populacji o rozkładzie normalnym. Przykładem takich testów jest test Kołmogorowa–Smirnowa oraz test  $W$  Shapiro–Wilka. Jednak żaden z tych testów nie zastąpi wizualnej oceny rozkładu wyników przy wykorzystaniu histogramu rozkładu częstości rozpatrywanej zmiennej. Wykres histogramu ułatwi ocenę normalności, gdy na empiryczny histogram będzie dodatkowo nałożona wpasowana krzywa rozkładu normalnego. Badania Monte Carlo wskazują, że jeśli liczebność próby jest większa od 100, to założenie o normalności rozkładu populacji nie powinno być praktycznie brane pod uwagę.

Jak wykazano wyżej, wyniki otrzymywane w próbach mogą być podstawą do weryfikowania hipotez statystycznych dotyczących rozkładu prawdopodobieństwa w populacji.

Rzeczoznawca majątkowy dokonując wyboru nieruchomości do porównania, formułuje statystyczną próbę prostą, w której – dla zmiennych losowych reprezentowanych

przez nieruchomości – zaobserwowana została w sposób ilościowy wybrana ich cecha, np. cena. Na podstawie wyników statystycznej próby rzeczoznawca, poprzez procedurę wyceny, formułuje nową funkcję na zmiennych losowych, która prowadzi do prognozy wartości rynkowej nieruchomości. Funkcja taka stanowi nową zmienną losową i nosi nazwę odpowiedniej **statystyki**. Sformułowane przez rzeczoznawcę statystyki powinny być przedmiotem wnioskowania statystycznego, a także powinny być estymowane punktowo i przedziałowo.

**Ad 3) Rozkład prawdopodobieństwa sformułowanej statystyki** na podstawie wyników z próby (nazywany często w skrócie „**rozkładem z próby**”) – w tym znaczeniu rozkład prawdopodobieństwa wybranej **statystyki** jako nowej zmiennej losowej wynika z **matematycznego rozumowania**, stąd **rodzaj rozkładu z próby nie zależy od wyników otrzymywanych w próbach**, gdyż zawsze jest rozkładem teoretycznym. Natomiast parametry tego rozkładu można określać na podstawie wyników otrzymywanych w próbach losowych. Rzeczoznawca w procesie wyceny nieruchomości najczęściej korzysta z niżej opisanych **rozkładów z próby**.

- Można wykazać, co udowodnił Fisher, że dla prób losowych o liczebności większej od 30 ( $n > 30$ ), bez względu na to, jaką ma postać rozkład cen nieruchomości w populacji, rozkład wartości przeciętnej z próby będzie zawsze posiadał cechy rozkładu normalnego o parametrach

$$N\left(\mu, \frac{\sigma}{\sqrt{n}}\right).$$

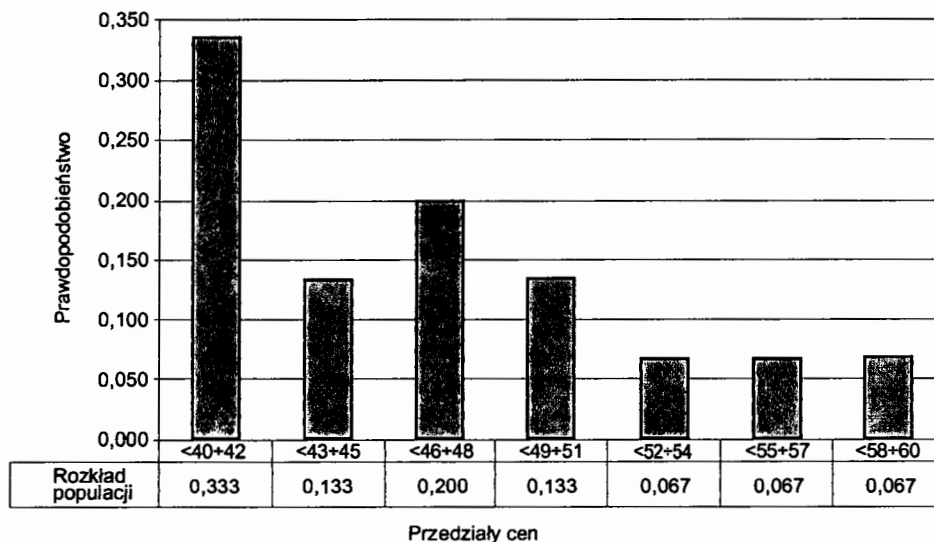
- Można wykazać, co udowodnił Snedecor, że dla prób losowych o liczebności większej od 30 ( $n > 30$ ), niezależnie od postaci rozkładu cen nieruchomości w populacji, rozkład wariancji z próby będzie zawsze posiadał cechy rozkładu chi-kwadrat.
- Do estymacji przedziałowej wykorzystuje się **statystykę**, która stanowi **iloraz wartości przeciętnej i odchylenia standardowego**. Można wykazać, co udowodnił Goset, że dla prób losowych o liczebności większej od 30 ( $n > 30$ ), niezależnie od postaci rozkładu cen nieruchomości w populacji, statystyka ta ma zawsze rozkład *t*-Studenta.

Jeżeli rozkład prawdopodobieństwa cen w populacji będzie układał się w przybliżeniu według rozkładu normalnego, to sformułowane wyżej **rozkłady z próby** będą spełnione już dla prób powyżej 7 nieruchomości.

Rzeczoznawca majątkowy, stosując statystyczne procedury wyceny, zawsze szacuje bezpośrednio lub pośrednio wartość przeciętną z cen nieruchomości wybranych do porównania (ze statystycznej próby). Szacowana wartość rynkowa nieruchomości, jako nowa zmienna losowa (**statystyka**) powinna mieć określone odchylenie standardowe, gdyż wtedy rzeczoznawca może w sposób jednoznaczny określić dla tej statystyki parametry rozkładu z próby, a następnie może sformułować przedziały ufności oraz weryfikować dowolne hipotezy statystyczne.

## 2. Praktyczna ilustracja twierdzenia o rozkładzie prawdopodobieństwa wartości przeciętnej z próby<sup>1)</sup>

Z rynku nieruchomości gruntowych wybrano 15 nieruchomości podobnych, które reprezentują populację ustalonego typu nieruchomości. Ceny nieruchomości zestawione w tabeli 1 mają charakter losowy, a rozkład ich prawdopodobieństwa nie przypomina w najmniejszym stopniu rozkładu normalnego, co ilustruje histogram przedstawiony na rysunku 1.



Rys. 1. Histogram częstości cen nieruchomości z bazy reprezentującej populację

Tabela 1. Ceny 15 nieruchomości podobnych reprezentujących lokalny rynek

Numer	1	2	3	4	5	6	7	8
Cena [zł/m <sup>2</sup> ]	40,00	40,00	57,00	44,00	45,00	53,00	46,00	43,00
Numer	9	10	11	12	13	14	15	
Cena [zł/m <sup>2</sup> ]	47,00	48,00	41,00	49,00	41,00	56,00	41,00	

Liczebność populacji – 15,  
 Cena minimalna – 40,00 zł/m<sup>2</sup>,  
 Cena średnia – 46,07 zł/m<sup>2</sup>,  
 Cena maksymalna – 57,00 zł/m<sup>2</sup>,  
 Odchylenie standardowe w bazie – 5,64 zł/m<sup>2</sup>.

<sup>1)</sup> Obliczenia do przykładu wykonał Marek Kulczycki.

Przeanalizujemy rozkład prawdopodobieństwa średniej arytmetycznej z próby. Z bazy 15-elementowej ( $n = 15$ ) wybieramy 7-elementowe podzbiory ( $k = 7$ ) i liczymy dla każdego z nich wartość średniej arytmetycznej (ze względu na specyfikę wyboru nieruchomości do wyceny rozpatrujemy tylko te podzbiory, które różnią się składem nieruchomości). Liczebność tak zdefiniowanej zmiennej losowej wynosi

$$N = \binom{n}{k} = \frac{n!}{k! \cdot (n-k)!} = \frac{15!}{7! \cdot 8!} = 6435.$$

Należy tutaj podkreślić, że rozpatrywana zmienna losowa (wartość przeciętna) o liczebności 6435 posiada **rozkład prawdopodobieństwa z próby**. Dla ustalonych przedziałów cenowych średniej arytmetycznej obliczono liczbę prób przynależnych poszczególnym przedziałom, a następnie stosując wzór

$$p_i = \frac{n_i}{N},$$

obliczono prawdopodobieństwo. Wyniki tych obliczeń zamieszczono w tabeli 2.

**Tabela 2.** Wyniki uzyskane z analizy częstości średniej arytmetycznej z prób

Przedział	Liczba prób o średniej z przedziału $n_i$	Prawdopodobieństwo $p_i$
<40,00÷41,00	0	0,000
<41,00÷42,00	6	0,001
<42,00÷43,00	119	0,018
<43,00÷44,00	446	0,069
<44,00÷45,00	970	0,151
<45,00÷46,00	1474	0,229
<46,00÷47,00	1513	0,235
<47,00÷48,00	1131	0,176
<48,00÷49,00	558	0,087
<49,00÷50,00	190	0,030
<50,00÷51,00	28	0,004
<51,00÷52,00	0	0,000
	<b><math>N = \sum n_i = 6435</math></b>	<b><math>\sum p_i = 1</math></b>

Empiryczna zmienna losowa reprezentowana przez średnią arytmetyczną z podzbiorów 7-elementowych posiada następujące parametry:

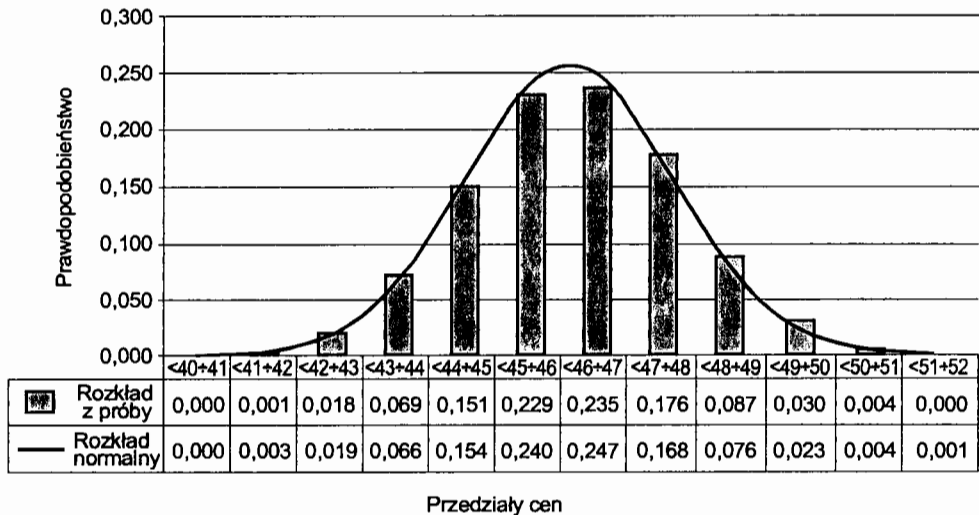
$$\hat{\mu} = 46,07 \text{ zł/m}^2 - \text{wartość przeciętna średnich cen z prób,}$$

$$\hat{\sigma} = 1,56 \text{ zł/m}^2 - \text{odchylenie standardowe,}$$

stąd funkcja gęstości prawdopodobieństwa rozkładu normalnego przyjmuje postać

$$f(X) = \frac{1}{1,56 \cdot \sqrt{2\pi}} \cdot e^{-\frac{(X-46,07)^2}{2 \cdot 1,56^2}} \quad (1)$$

Wyniki próby losowej zawarte w tabeli 2 zostały zilustrowane na rysunku 2. W histogram rozkładu prawdopodobieństwa została wpasowana funkcja gęstości prawdopodobieństwa rozkładu normalnego o równaniu (1).



**Rys. 2.** Rozkład prawdopodobieństwa średniej arytmetycznej z prób 7-elementowych losowanych z bazy 15-elementowej

Dodatkowo, dla zadanego poziomu istotności można określić przedział ufności. Ponieważ liczba elementów próby (wybieranych do porównania nieruchomości) jest mniejsza od 21, należy stosować kwantyle rozkładu Studenta dla poziomu  $1 - \alpha = 0,95$  i  $k - 1 = 6$  stopni swobody

$$t(1 - \alpha; k - 1) = t(0,95; 6) = 2,45,$$

dlatego przedział ufności przyjmuje postać

$$\mu_{1-\alpha} = \mu \pm t(1 - \alpha; k - 1) \cdot \hat{\sigma} \Rightarrow \mu_{0,95} = 46,07 \pm 2,45 \cdot 1,56$$

i ostatecznie

$$\mu_{0,95} = 46,07 \text{ zł/m}^2 \pm 3,82 \text{ zł/m}^2,$$

co oznacza, że wykorzystując do porównania siedem nieruchomości, losowo wybranych z bazy zawierającej piętnaście nieruchomości, 95% rzeczoznawców poda wartość nieruchomości mieszczącą się w tym przedziale, zaś 5% rzeczoznawców może podać wartość spoza przedziału.

### 3. Wnioski metodyczne

W oparciu o powyższą analizę można wysnuć następujący wniosek. Gdy liczebność bazy nieruchomości do wyboru będzie większa od 15 (optymalnie  $n > 30$ ), bez względu na rozkład prawdopodobieństwa cen nieruchomości rynkowych tworzących bazę, szacowanie wartości przeciętnej z różnych podzbiorów (prób) nieruchomości wybieranych do porównania o liczbie  $k \geq 7$  będzie odbywało się według rozkładu normalnego. Stąd wszystkie implikacje płynące z rozkładu normalnego wartości przeciętnej powinny być wykorzystane w wycenie nieruchomości.

W ogólnym przypadku analiza statystyczna rynku nieruchomości dotyczy wielowymiarowej zmiennej losowej. Szacowanie wartości nieruchomości może być wykonywane przy różnych założeniach dotyczących rozkładu prawdopodobieństwa poszczególnych zmiennych losowych. W uproszczonym ujęciu analizę statystyczną rynku można prowadzić według następujących modeli:

- ceny nieruchomości mają rozkład zmiennej losowej jednowymiarowej, zaś atrybuty porównywanych nieruchomości są identyczne, więc stanowią stałe parametry analizowanego rynku;
- ceny nieruchomości z poszczególnymi ich atrybutami tworzące pary (cena – atrybut) stanowią zmienne losowe dwuwymiarowe;
- ceny nieruchomości i ich atrybuty określają zmienną losową wielowymiarową, gdzie ceny stanowią zmienną objaśnianą zaś atrybuty stanowią zmienne objaśniające.

W pierwszym przypadku parametrami rozkładu cen nieruchomości są **wartość przeciętna i odchylenie standardowe z próby oraz odchylenie standardowe wartości przeciętnej z próby**. Na podstawie tych parametrów można określać wagi dokładności według zasady Markowa, a następnie wartość nieruchomości jako średnią ważoną z prognozowanych cen dla poszczególnych atrybutów. Na podstawie **rozkładu z próby** statystyki  $t$ -Studenta można formułować przedziały ufności i weryfikować hipotezy statystyczne, dotyczące między innymi minimalnej liczby nieruchomości do porównania. Liczba ta będzie zależeć przede wszystkim od przyjętego poziomu istotności, co wyraźnie omówiono w [9].

W drugim przypadku parametrami rozkładu zmiennych losowych dwuwymiarowych (cena – atrybut) są wartości przeciętne i odchylenia standardowe w rozkładach brzegowych oraz współczynnik korelacji (Pearsona). Na podstawie tych parametrów można bezpośrednio wyznaczyć prostą linię regresji, która reprezentuje liniowy model zmienności cen nieruchomości względem wybranego atrybutu. Zbiór informacji rynkowych rozpatrywany w modelu regresji nie musi posiadać cech rozkładu normalnego, natomiast cechy tego rozkładu powinien posiadać zbiór odchyłek losowych (resztowych) tego modelu. Wykorzystując metodę najmniejszych kwadratów można również estymować nieliniowe modele dwuwymiarowej regresji. Ustalony dla każdego atrybutu model regresji może być wykorzystany do prognozowania wartości nieruchomości i jej wagi.



Na podstawie tych parametrów można określić wartość przeciętną i jej odchylenie standardowe, które posiadają **rozkład z próby**. Rozkład z próby statystyki *t*-Studenta pozwala formułować przedziały ufności i weryfikować hipotezy statystyczne.

W trzecim przypadku ceny nieruchomości można modelować za pomocą liniowych modeli regresji wielorakiej (wielokrotnej), w których ceny reprezentują zmienną zależną, zaś atrybuty stanowią zmienne niezależne. Parametry takich modeli wyznacza się zawsze za pomocą metody najmniejszych kwadratów, przy wykorzystaniu rachunku macierzowego. Miarą dopasowania modelu regresji do cen rynkowych jest kwadrat współczynnika wielorakiej regresji. Estymowane modele wielorakiej regresji można wykorzystać do prognozowania wartości rynkowej oraz do oceny dokładności tej prognozy. Można wykazać, co udowodnił Fisher, że dla takiego modelu rozkład prawdopodobieństwa cen w populacji może być dowolny. Wnioskowaniu statystycznemu są poddawane odchyłki losowe do modelu.

Zmienną losową wielowymiarową można również opisywać za pomocą nieliniowych modeli regresji wielorakiej, w których atrybuty są wyrażane za pomocą nieliniowych funkcji różnej postaci. Modele takie występują pod ogólną nazwą stochastycznych modeli hybrydowych. Aby określić parametry takiego modelu, funkcje opisujące nieliniowe modele regresji muszą być rozwijane w szereg Taylora. Na podstawie tego rozwinięcia zestawia się układ równań aproksymujących, który można rozwiązać według zasady najmniejszych kwadratów, za pomocą rachunku macierzowego.

Zamieszczone pojęcia, terminy i opisy modeli statystycznych są przedstawione w formie lapidarnej i nie wyczerpują treści zagadnień związanych z zawodem rzeczoznawcy majątkowego. Wiele praktycznych przykładów w zakresie zastosowania modeli statystycznych do analizy rynku i wyceny nieruchomości można znaleźć w [8].

### Literatura

- [1] Cegielski P.: *Przykład alternatywnej techniki analizy statystycznej rynku – część 2*. Warszawa, Rzeczoznawca Majątkowy, nr 2, 1999
- [2] Czaja J.: *Estymacja uogólnionych modeli liniowych*. Warszawa, PAN, Geodezja i Kartografia, t. XLIII, z. 3, 1994
- [3] Czaja J. i in.: *Wycena nieruchomości majątkowych metodą cenowo-porównawczą*. Warszawa, Przegląd Geodezyjny, 12, 1996
- [4] Czaja J.: *Zamiast techniki porównywania parami – technika interpolacyjna – wyceny nieruchomości*. Katowice, Kwartalnik Nieruchomość, nr 3, 1996
- [5] Czaja J.: *Modele statystyczne w informacji o terenie*. Kraków, Wydawnictwa AGH 1997
- [6] Czaja J.: *Interval Estimation of Generalized Linear Models*. Warszawa, PAN, Geodezja i Kartografia, t. XLVI, z. 1, 1997
- [7] Czaja J.: *Praktyczna weryfikacja różnych technik wyceny w podejściu porównawczym*. Kraków, Rzeczoznawca Małopolski, nr 2, 1997
- [8] Czaja J.: *Metody i systemy szacowania nieruchomości*. Kraków, podręcznik dla rzeczoznawców majątkowych, 1999

- 
- [9] Czaja J.: *Podejście porównawcze wyceny nieruchomości w aspekcie standardów*. Warszawa, Rzeczoznawca Majątkowy, nr 4, 1999
- [10] Czaja J.: *Metody i systemy określania wartości nieruchomości*. Kraków, UWND AGH 1999
- [11] Mazurkiewicz E.: *Uwagi do procedur określania wartości rynkowej nieruchomości przy zastosowaniu analizy statystycznej rynku*. Warszawa, Rzeczoznawca Majątkowy, nr 2, 1999
- [12] Prystupa M.: *Wycena nieruchomości metodą cenowo-porównawczą*. Warszawa, Biblioteczka rzeczoznawcy majątkowego 1997
- [13] Prystupa M.: *Jaki standard dla podejścia porównawczego?* Warszawa, Rzeczoznawca Majątkowy, nr 4, 1999
- [14] Rao C.R.: *Modele liniowe statystyki matematycznej*. Warszawa, PWN 1982

Recenzent

prof. dr hab. inż. Stanisław Latoś