

Małgorzata Renigier

Selekcja zmiennych a priori przy nieklasycznych założeniach

Acta Scientiarum Polonorum. Administratio Locorum 3/2, 71-81

2004

Artykuł został zdigitalizowany i opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

SELEKCJA ZMIENNYCH A PRIORI PRZY NIEKLASYCZNYCH ZAŁOŻENIACH

Małgorzata Renigier

Katedra Gospodarki Nieruchomościami i Rozwoju Regionalnego,
Uniwersytet Warmińsko-Mazurski w Olsztynie

Streszczenie: Szeroko pojęta informatyzacja i powiększające się wciąż wartości informacji w dążeniu do tworzenia nowoczesnej „gospodarki informacyjnej” oraz zaawansowane techniki komputerowe pozwalają na coraz większe usystematyzowanie i unifikację procesów społeczno-gospodarczych często ujmowanych w postaci zaawansowanych modeli statystyczno-matematycznych.

Budowanie modeli statystycznych jest związane z wieloma etapami, a jednym z najważniejszych z nich jest dobór zmiennych do modelu. Artykuł ten jest propozycją przedstawienia metody wstępnej selekcji zmiennych przy nieklasycznych założeniach budowy modelu regresji związanych z nieliniowością zjawisk na rynku nieruchomości.

Słowa kluczowe: model regresji nieliniowy, selekcja zmiennych, analiza wariancji.

WSTĘP

Stworzenie jasno i precyzyjnie działającego schematu działań poszczególnych zjawisk w gospodarce kraju jest głównym przedmiotem działań wielu specjalistów w tej dziedzinie od wielu lat. Obecnie w dobie informatyzacji i zaawansowanych technik komputerowych usystematyzowanie i unifikacja procesów społeczno-gospodarczych często ujmowane jest w postaci zaawansowanych modeli statystycznych. Określenie wartości nieruchomości jako istotnego składnika przestrzeni może być ważnym instrumentem użytecznym w polityce gospodarczej.

Adres do korespondencji – Corresponding author: Małgorzata Renigier, Katedra Gospodarki Nieruchomościami i Rozwoju Regionalnego, Uniwersytet Warmińsko-Mazurski, ul. Prawocheńskiego 15, 10-720 Olsztyn.

Wartość nieruchomości określa się w procesie wyceny, która wykorzystuje między innymi metody statystyczne. Metody analizy statystycznej rynku polegają na wykorzystaniu do badań zjawisk rynkowych narzędzi statystycznych. Badania te polegają na sformułowaniu przy wykorzystaniu metod statystycznych diagnoz, wniosków, interpretacji struktury i zależności zjawisk zachodzących na rynku nieruchomości.

Analizy statystyczne są powszechnie stosowane w wycenie masowej ze względu na potrzebę przeanalizowania olbrzymich pojemnościowo zbiorów danych oraz wymogów uzyskania wyników o wysokiej precyzji, niezbędnych do prognozowania wartości, ustalania podatków od nieruchomości czy inwestycyjnego planowania przestrzeni.

Wzrogacenie świadomości struktur odpowiedzialnych za kształtowanie się i rozwój kraju poprzez modelowanie statystyczne jako wyraz zsyntetyzowanej informacji na temat czynników kształtujących wartość powinno wpłynąć na kształtowanie równowagi z wykorzystaniem spójnego systemu gospodarczego.

Najistotniejszą sprawą przy tworzeniu i interpretacji modeli do analizy rynku jest takie połączenie narzędzi matematyczno-statystycznych z wnikliwą merytoryczną analizą danych, by wynikiem tego było jak najwierniejsze odzwierciedlenie stanu rzeczywistego w postaci modelu.

DOBÓR ZMIENNYCH JAKO JEDEN Z NAJISTOTNIEJSZYCH ETAPÓW BUDOWY MODELU

Wytypowanie optymalnego zestawu zmiennych tworzących model kształtowania się badanego zjawiska ma uzasadnienie nie tylko ze względu na aspekt praktyczny, ale także na ich wymiar w interpretacji teoretycznej. Negatywny wpływ nieadekwatnych zmiennych może się przejawiać uzyskaniem błędnych wyników (zawyżonych lub zaniżonych), które jednocześnie zafałszują obraz rzeczywistych relacji. Za nieadekwatne zmienne można uznać obiekty, które są niekompatybilne z wybraną wcześniej metodą estymacji wartości parametrów modelu oraz rodzajem modelu. Istnieje wiele metod estymacji nieznanymi parametrów oraz rodzajów modeli statystycznych. Najpopularniejszymi i powszechnie stosowanymi modelami w analizach przestrzennych są modele regresji liniowej ze względu na prostotę wykonania i łatwość interpretacji wyników.

Rozwój badań nad dynamiką i złożonością systemów doprowadził do popularyzacji pojęcia nieliniowości jako kategorii ujmowania zjawisk gospodarczych. Należy zwrócić uwagę na fakt, że każdy system gospodarczy jest tworem dynamicznym i nieustannie ewoluującym. Zobrazowanie go w postaci modeli nie jest proste i nie powinno być zbyt upraszczane. Efektem tego w najgorszym wypadku może być zniekształcenie badanego obszaru, dające mylne wnioski, a w najlepszym nie pozwalające na pogłębienie wiedzy na temat analizowanego zjawiska. W związku z tym wielu specjalistów uważa

(m.in. Hozer i Kuźmiński 2002, Peters 1997), że zastosowanie modeli liniowych regresji może nie dać zadowalających rezultatów.

W artykule przedstawiono procedurę selekcji zmiennych a priori, gdy nie znamy dokładnie zależności, w jakich występują wzajemne relacje między zmiennymi modelowymi lub gdy uznajemy za fakt powiązania nieliniowe.

OBSZAR BADAŃ ORAZ DOBÓR MIERNIKÓW SELEKCJI ZMIENNYCH

Wymienionej metodzie selekcji poddano 25 cech opisujących kształtowanie się cen na nieruchomości gruntowe nie zabudowane. Zmienne objaśniające wybrano po przeprowadzeniu badań literaturowych pod kątem wyselekcjonowania grupy zmiennych, które wyczerpująco opisują dany rodzaj nieruchomości.

Analizę przeprowadzono na podstawie transakcji zaistniałych na olsztyńskim rynku nieruchomości gruntowych, nie zabudowanych, przeznaczonych pod zabudowę mieszkaniową. W badaniach uwzględniono 821 transakcji przeprowadzonych w latach 1997-2003.

Istotną rzeczą przy wyborze zmiennych do modelu jest określenie, czy wywierają one istotny wpływ na zmienną objaśnianą, a jeśli tak, to z jaką siłą. Aby rozpoznać liczbowo siłę oraz kształt powiązań między zmiennymi objaśniającymi (X) i zmienną objaśnianą (Y) można zastosować następujące procedury statystyczne:

analiza wariancji, oparta głównie na pojedynczej metodzie analizy wariancji, wykorzystującej warunkowe statystyki opisowe, które za cel przyjmują określenie siły wpływu czynnika klasyfikacyjnego X na skalę zmienności zmiennej losowej Y .

analiza korelacji, oparta zazwyczaj, choć nie jedynie, na empirycznych tablicach korelacyjnych, wykorzystuje różne miary korelacji, których wynikiem jest liczbowe określenie siły (zawsze), kierunku (niekiedy) oraz kształtu (zawsze) zależności między zmiennymi X i Y .

Selekcję wybranych wcześniej zmiennych przeprowadzono na podstawie wymienionych procedur metod statystycznych.

OKREŚLENIE WPŁYWU ZMIENNYCH NIEZALEŻNYCH NA ZMIENNĄ ZALEŻNĄ Z WYKORZYSTANIEM ANALIZY WARIANCJI

Do przeprowadzonych badań wykorzystano wariant jednoczynnikowej analizy wariancji. Procedura sprowadza się do dekompozycji ogólnego stopnia zróżnicowania zmiennej objaśnianej (SST) według dwóch podstawowych i sumujących się źródeł jego powstawania: stopień zróżnicowania międzygrupowego (SSB) oraz stopień zróżnicowania wewnątrzgrupowego (SSE) (tzw. skład-

nika resztowego). Zatem schemat analizy wariancji wyrażonej jako rozłożenie ogólnej sumy kwadratów odchyłeń realizacji zmiennej objaśnianej wyraża się wzorem:

$$\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2 = \sum_{i=1}^k (\bar{y}_i - \bar{y})^2 n_i + \sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2 \quad [1]$$

gdzie:

$\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y})^2$ – to całkowita suma kwadratów odchyłeń od średniej ogólnej (SST),

$\sum_{i=1}^k (\bar{y}_i - \bar{y})^2 n_i$ – to suma kwadratów odchyłeń średnich grupowych od średniej ogólnej (SSB),

$\sum_{i=1}^k \sum_{j=1}^{n_i} (y_{ij} - \bar{y}_i)^2$ – to suma kwadratów odchyłeń wartości cechy od średniej grupowej (SSE)

przy założeniu, że: $i = 1, 2 \dots k$ (poziomy czynnika klasyfikacyjnego) oraz $j = 1, 2, \dots, n_j$ (numery jednostek obserwacji w i -tym poziomie czynnika klasyfikacyjnego)

Cechą charakterystyczną analizy wariancji jest to, że stanowi ona w zasadzie test dla hipotezy, iż wyróżniony czynnik nie wpływa na poziom zmiennej Y . Do oceny stosuje się statystykę F Fishera-Snedecora, która jest ilorazem średnich kwadratów odchyłeń międzygrupowych i wewnątrzgrupowych określoną wzorem:

$$F = \frac{MSB}{MSE} \quad [2]$$

gdzie:

$$MSB = \frac{SSB}{r - 1}$$

$$MSE = \frac{SSE}{n - r}$$

przy założeniu, że $r - 1$ i $n - r$ to liczba stopni swobody (df).

Wyniki przeprowadzonej analizy wariancji wpływu zmiennych X na zmienną Y znajdują się w tabeli 1. Wyrazem zakwalifikowania się zmiennej objaśniającej do grupy zmiennych wywierających wpływ na zmienną objaśnianą jest poziom p , który jest wynikiem relacji wysokości statystyki F obserwo-

Tabela 1. Wyniki analizy wpływu zmiennych niezależnych na zmienną zależną na podstawie analizy wariancji

Table 1. Results of variance analysis - influence of independent variables on the dependent variable

Pary zmiennych poddane analizie wariancji Pair of variables subjected to variance analysis	SSB	df	MSB	SSE	df	MSE	F	p
cena & pow	234051,8	6	39008,6	1923286	814	2362,7	16,51	0,000000
cena & data	760064,5	67	11344,2	1397273	753	1855,6	6,11	0,000000
cena & ilosc dz	20516,3	8	2564,5	2136822	812	2631,5	0,97	0,454445
cena & F wład	2602,3	1	2602,3	2154736	819	2630,9	0,99	0,320253
cena & front	95529,56	8	11941,19	2061808	812	2539,17	4,70	0,000012
cena & głęb	95529,56	8	11941,19	2061808	812	2539,17	4,70	0,000012
cena & ksztalt	67,8	1	67,8	2157270	819	2634,0	0,02	0,878914
cena & usyt	7790,1	1	7790,1	2149337	819	2627,5	2,96	0,087863
cena & E	245615,0	3	81871,6	1911723	817	2339,9	34,98	0,000000
cena & W	274234,7	3	91411,58	1883103	817	2304,9	39,65	0,000000
cena & G	638520,3	3	212840,1	1518818	817	1859,0	114,49	0,000000
cena & T	641534,7	3	213844,9	1515803	817	1855,3	115,25	0,000000
cena & K	641534,7	3	213844,9	1515803	817	1855,3	115,25	0,000000
cena & UT	97254,5	4	24313,6	2060083	816	2524,6	9,63	0,000000
cena & DD	90764,6	5	18152,9	2066573	815	2535,6	7,15	0,000001
cena & U d	455698,1	4	113924,5	1701640	816	2085,3	54,63	0,000000
cena & U k	10311,0	4	2577,7	2147027	816	2631,1	0,97	0,417738
cena & U I	16135,0	4	4033,75	2141203	816	2624,0	1,53	0,189404
cena & PZP	125874,6	2	62937,31	2031463	818	2483,4	25,34	0,000000
cena & F trans	2431,528	1	2431,5	2154906	819	2631,1	0,92	0,336676
cena & Atr woda	223912,8	5	44782,6	1933425	815	2372,3	18,9	0,000000
cena & Atr las	115053,3	5	23010,6	2042285	815	2505,9	9,18	0,000000
cena & Odl H	131125,6	6	21854,3	2026212	814	2489,2	8,77	0,000000
cena & Odl C	283612,1	5	56722,4	1873726	815	2299,1	24,67	0,000000
cena & Odl K	294218,3	5	58843,7	1863120	815	2286,0	25,74	0,000000

Źródło: Badania własne

Source: own studies.

symbole zmiennych niezależnych zgodnie z oznaczeniami:

symbols of independent variables symbols according to designation:

cena – cena;

cena – value;
 pow – powierzchnia;
 pow – area;
 data – data transakcji;
 data – date of transaction
 liczba dz – liczba działek będąca przedmiotem jednej transakcji;
 liczba dz – number of plots
 F wład – forma władania nieruchomością;
 F wład – form of possession of a real estate
 front – front działki;
 front – front of plot
 głęb – głębokość działki;
 głęb – depth of plot
 kształt – kształt działki;
 kształt – shape of plot
 usyt – usytuowanie działki;
 usyt – location plot
 E – wyposażenie w sieć elektroenergetyczną;
 E – electric energy supply system
 W – wyposażenie w sieć wodociągową;
 W – water-supply system
 G – wyposażenie w sieć gazową;
 G – gas supply system
 T – wyposażenie w sieć telekomunikacyjną;
 T – communications network
 K – wyposażenie w sieć kanalizacyjną;
 K – sewage system
 UT – ukształtowanie terenu ;
 UT – topographic features
 DD – dojazd i dostęp do działki ;
 DD – approach and access to plot
 Ud – uciążliwość sąsiedztwa związana z ruchem kołowym ;
 Ud – burdensome neighborhood – vicinity of heavy-traffic roads
 Uk – uciążliwość sąsiedztwa związane z liniami kolejowymi;
 Uk – burdensome neighborhood – vicinity of rail lines
 Ui – inne uciążliwości sąsiedztwa związane np. z zakładem przemysłowym;
 Ui – burdensome neighborhood – vicinity of industrial plants and other
 PZP – przeznaczenie w planie;
 PZP – land function in a local plan
 F trans – forma transakcji;
 F trans – form of transaction
 Atr woda – atrakcyjność położenia w stosunku do obszarów leśnych i parków;
 Atr woda – attractiveness of location in relation to woodland areas and parks
 Atr las – atrakcyjność położenia w stosunku do obszarów zbiorników wodnych;
 Atr las – attractiveness of location in relation to water reservoirs
 Odl H – lokalizacja związana z odległością od najważniejszego ośrodka handlowo-usługowego;
 Odl H – location connected with the distance to the main shopping and service center
 Odl C – lokalizacja związana z odległością od centrum miasta;
 Odl C – location connected with the distance to the city center
 Odl K – lokalizacja związana z odległością od komunikacji ogólnej;
 Odl K – location connected with the distance to means of transport

wanej i wartości dopuszczalnej, przy założonym poziomie istotności na wysokości 0,05. Oznacza to, iż w 95% istnieje zaufanie do wyników testów istotności, a wysokość tę przyjęto z uwagi na 5-procentową możliwość popełnienia błędów w pomiarze zmiennych lub sposobie kodowania zmiennych.

OKREŚLENIE SIŁY WPLYWU ZMIENNYCH NIEZALEŻNYCH NA ZMIENNĄ ZALEŻNĄ PRZY WYKORZYSTANIU ANALIZY KORELACJI

Zmienne wyselekcjonowano na podstawie różnych rodzajów rachunków korelacyjnych, które odzwierciedlają kwantyfikację wzajemnych powiązań między zmiennymi X i Y .

Do określenia związków korelacyjnych między wybranymi danymi zastosowano trzy miary korelacyjne, tj współczynniki: gamma, współczynnik korelacji liniowej, Pearsona, a także wskaźnik korelacyjny Pearsona. Charakterystyka wymienionych miar korelacyjnych przedstawia się następująco:

- **statystyka gamma** jest nieparametryczną miarą korelacji i jest zalecana, jeśli dane zawierają wiele powiązanych obserwacji; współczynnik gamma opiera się na prawdopodobieństwie; liczy się go jako różnicę między prawdopodobieństwem, że uporządkowanie dwóch zmiennych jest zgodne, a prawdopodobieństwem, że jest niezgodne, podzieloną przez 1 minus prawdopodobieństwo występowania obserwacji powiązanych (StatSoft, 2001);
- **współczynnik korelacji liniowej Pearsona** jest estymatorem korelacyjnym nieznanego parametru, którego konstrukcja wiąże się z teoretyczną macierzą kowariancji i wariancji [Luszniewicz, Słaby 2001];
- **wskaźnik korelacyjny Pearsona**, którego podstawą konstrukcji jest jednoczynnikowa analiza wariancji; miara ta nie podlega zniekształceniu efektem krzywoliniowości wzajemnych relacji między badanymi zmiennymi i w pomiarze siły korelacji są uwolnione od tego zjawiska; wskazuje jednak na kierunek korelacji, ponieważ zawsze jest dodatni [Luszniewicz, Słaby 2001].

Wyniki badań związanych z analizami związków korelacyjnych między zmiennymi przedstawia tabela 2. Dane istotne pogrubiono.

Tabela 2. Analiza korelacyjna zmiennych
Table 2. Correlation analysis of variables

Pary zmiennych pomiedzy którymi pomierzono związki korelacyjne Correlations between pairs of variables	Miary korelacyjne Correlation measures				
	Współczyn- nik Gamma Gamma coefficient	Współczyn- nik korelacji liniowej Pearsona Pearson correlation coefficient	Wskaźnik korelacji Pearsona Pearson correlation coefficient	F	p
cena & pow	-0,25	-0,22	0,33	4,05	0,00000
cena & data	0,37	0,42	0,59	17,71	0,00000
cena & liczba dz	0,01	-0,01	0,09	0,27	0,99984
cena & F wład	-0,03	-0,03	0,03	0,03	1,00000
cena & front	-0,08	-0,04	0,21	1,53	0,05001
cena & głęb	-0,26	-0,22	0,21	1,53	0,05001
cena & kształt	0,05	0,01	0,01	0,00	1,00000
cena & usyt	0,03	0,06	0,06	0,11	1,00000
cena & E	0,17	0,27	0,34	4,33	0,00000
cena & W	0,20	0,26	0,36	4,94	0,00000
cena & G	0,56	0,53	0,54	13,65	0,00000
cena & T	0,59	0,53	0,54	13,65	0,00000
cena & K	0,43	0,42	0,54	13,65	0,00000
cena & UT	0,12	0,15	0,21	1,53	0,05001
cena & DD	-0,05	0,00	0,21	1,53	0,05001
cena & U d	0,59	0,43	0,46	8,90	0,00000
cena & U k	-0,08	-0,04	0,07	0,16	0,99999
cena & U I	-0,02	-0,03	0,09	0,27	0,99984
cena & PZP	0,25	0,24	0,24	2,02	0,00274
cena & F trans	0,02	0,11	0,03	0,03	1,00000
cena & Atr woda	0,27	0,31	0,33	4,05	0,00000
cena & Atr las	0,05	0,07	0,23	1,85	0,00805
cena & Odl H	0,09	0,00	0,25	3,49	0,00000
cena & Odl C	-0,02	-0,14	0,36	4,94	0,00000
cena & Odl K	0,02	-0,10	0,37	5,26	0,00000

Źródło: Badania własne

Source: own studies

Za istotne należy uznać, że dwie pierwsze miary korelacyjne, tj. współczynnik gamma oraz współczynnik korelacji Pearsona, mają ograniczenia liniowe. Oznacza to, że parametry te efektywnie mierzą siłę i prawidłowo wskazują kierunek korelacji, kiedy zależności między zmiennymi X i Y mają przebieg liniowy lub zbliżony do liniowego. W związku z tym, że nie jest znany do tej pory fakt, z jakiego rodzaju wzajemnym skorelowaniem zmiennych mamy do czynienia (liniowym lub nieliniowym) za decydującą miarę uznano wskaźnik korelacyjny Pearsona, który jako jedyny z wymienionych jest odporny na zależności krzywoliniowe. Dwie pierwsze miary zastosowano jedynie do oceny różnorodności wyników mierników korelacji.

Tabela 3. Wynik wstępnej selekcji zmiennych na podstawie analizy wariancji i wskaźnika korelacji Pearsona
Table 3. Results of pre-selection of variables on the basis of variance analysis and the Pearson correlation coefficient

Pow	UT
Data	DD
liczba dz	Ud
F wład	U k
Front	Ui
Głęb	PZP
Kształt	F trans
Usyt	atr- woda
E	atr las
W	odl H
G	odl C
T	odl K
K	

– zmienne odrzucone zaznaczono pogrubioną czcionką

– reject variables are printed in bold

– oznaczenia jak w tabeli 1

– symbols as in Table 1

Źródło: Badania własne

Source: own studies

WYNIK WSTĘPNEJ SELEKCJI ZMIENNYCH

Po zastosowaniu wyżej wymienionych procedur selekcji zmiennych, których miarami były analizy jednoczynnikowej wariancji oraz siły korelacyjne do dalszych badań wyselekcjonowano 17 zmiennych. Wynik znajduje się w tabeli 3.

PODSUMOWANIE

Przystępując do analizy zjawisk gospodarczych, należy najpierw sprecyzować ich strukturę działania i rozwoju. W naturze człowieka często leży jak największe uproszczenie i uporządkowanie spraw go otaczających. Jednak bez względu na to, jak bardzo będziemy się starać, świat nie jest prostym tworem z łatwymi rozwiązaniami.

Najczęściej stosowaną, a tym samym najbardziej popularną, jak dotąd, postacią funkcyjną modelu przedstawiającego zależności opisujące rynek nieruchomości jest model liniowy, ze względu na prostotę wykonania i łatwość interpretacji wyników. Należy jednak zwrócić uwagę na istotny fakt, że każdy system gospodarczy jest tworem dynamicznym i nieustannie ewoluującym. Zobrazowanie go w postaci modeli nie jest proste i nie powinno być zbyt upraszczane. Efektem tego, w najgorszym wypadku, może być zniekształcenie badanego zjawiska pozwalające na mylne wnioski, a w najlepszym nie pozwalające na pogłębienie wiedzy na temat analizowanej dziedziny.

Jednym z największych problemów związanych z wyceną nieruchomości z wykorzystaniem analiz statystycznych jest koszt uzyskania informacji o danych oraz czas niezbędny do ich obliczeń. Stworzenie optymalnego zbioru danych nie wiąże się jedynie z minimalizacją wymienionych czynników (skądinąd niezwykle istotnych), ale także z wyborem zmiennych w taki sposób, aby najlepiej odzwierciedlały kształt badanego zjawiska, który często nie jest liniowy.

Próba poznania mechanizmu działania realnego świata w postaci modeli statystycznych jest cennym źródłem wiedzy, pod warunkiem jednak, że model ten w miarę możliwości oddaje dynamikę oraz heterogeniczność badanych zjawisk.

Istnieje wiele metod i narzędzi doboru zmiennych do modelu, których jednak podstawowym warunkiem użyteczności jest spełnienie założenia liniowości między analizowanymi cechami. Z uwagi na ten fakt zaproponowano procedurę selekcji zmiennych a priori z zastosowaniem mierników doboru zmiennych, których cechą szczególną jest ich odporność na zależności krzywoliniowe, a w związku z tym – większe „pole manewru” przy realizacji następnym etapów kreowania modelu mającego być jak najwierniejszym odzwierciedleniem realnych zależności analizowanych zjawisk.

PIŚMIENNICTWO

- Adamczewski Z. 2002. Nieliniowe i nieklasyczne algorytmy w geodezji. Oficyna wydawnicza PWN. Warszawa.
- Chow G.C., 1995. Ekonometria, Wydawnictwo Naukowe PWN, Warszawa.
- Czaja J., Marczevska B., Świątoniowska D., Żak M. 1994. Określenie modelu regresji i współczynników korelacji w powszechnej wycenie. *Przeg. Geod.* 7.
- Czaja J. 1997. Modele statystyczne w informacji o terenie. AGH. Kraków.
- Czaja J. 1999. Metody i systemy określania wartości nieruchomości. AGH uczelniane wydawnictwa naukowo-dydaktyczne Kraków.
- Gruszczyński M., Podgórska M. 2000. Ekonometria. Oficyna Wydawnicza Szkoły Głównej Handlowej. Warszawa.
- Hozer J., Kuźmiński W. 2002. Metody analizy statystycznej rynku w wycenie nieruchomości. PFSRM Warszawa.
- Józwiak J., Podgórski J. 1997. Statystyka od podstaw. Polskie Wydawnictwo Ekonomiczne. Warszawa.
- Luszniewicz A., Słaby T. 2001. Statystyka, teoria i zastosowania. Wydawnictwo C.H. Beck. Warszawa.
- Welfe A. 1998. Ekonometria. Metody i ich zastosowanie. Polskie Wydawnictwo Ekonomiczne. Warszawa.
- Peters E. 1997. Teoria chaosu a rynki kapitałowe. Wig-Press Warszawa.

A PRIORI SELECTION OF VARIABLES WITH NON-CLASSICAL ASSUMPTIONS

Abstract. Widely understood computerization and the continually growing value of information in the aspiration for creating a modern "informative economy", as well as advanced computer techniques, enable to systemize and unify the economic and social processes often presented in the form of developed statistical and mathematical models.

Building statistical models involves many stages, and one of the most important among them is the selection of model variables. This paper presents a preliminary method for variable selection with non-classical assumptions while developing a regression model, connected with non-linearity of the phenomena taking place on the real estate market.

Key words: nonlinear regression model, selection of variables, analysis of variance.

Zaakceptowano do druku 2004.10.28

Accepted for print