

MAGDALENA OKUPNIAK

ZASTOSOWANIE ANALIZY BLOKOWEJ NA PRZYKŁADZIE BADANIA ZRÓŻNICOWANIA DOCHODÓW GOSPODARSTW DOMOWYCH

1. WSTĘP

Analiza blokowa należy do rodziny metod wielowymiarowej analizy danych statystycznych, często stosowanych w badaniach związanych z rolnictwem. Przeprowadza się ją głównie w celu oceny wpływu na wzrost i rozwój roślin różnorodnych czynników takich jak nasłonecznienie, wilgotność, nawozy mineralne, czy środki ochrony. Znalazła ona zastosowanie również w biologii, medycynie, inżynierii, fizyce, chemii, agrokulturze czy w eksperymentach przemysłowych. Zebrane dane jakie poddajemy badaniu podlegają podwójnej klasyfikacji. Z jednej strony jednostki podzielone są według poziomu czynnika (bądź rodzaju zastosowanego obiektu), z drugiej strony podzielone na możliwie jednorodne grupy (bloki). Stąd można wyróżnić analizę wewnątrz i między-blokową, które odpowiadają badaniu wpływu czynnika zastosowanego na jednostkach oraz badaniu wpływu podziału jednostek na grupy. Twórcą analizy blokowej jest brytyjski statystyk Ronald Aylmer Fisher, który w początkach XX wieku zdefiniował schemat układu bloków kompletnie zrandomizowanych. W tym szczególnym przypadku można również badać obiekty zastosowane na losowo wybranych jednostkach eksperymentalnych podzielonych na możliwie jednorodne grupy. Różnica w tym przypadku polega na tym, że grupy te są równoliczne (Oktaba, 1974).

2. CHARAKTERYSTYKA METODY ANALIZY BLOKOWEJ

ZAŁOŻENIA OGÓLNE

Zakładamy, że w doświadczeniu dysponujemy n jednostkami eksperymentalnymi, które podzielone są na b bloków (grup) tak, by możliwie najmniej różniły się między sobą w obrębie danego bloku. Ponadto każdy blok ma wielkość (pojemność) k_j dla $j = 1, \dots, b$. Na tak rozdysponowanych jednostkach stosujemy v obiektów, losowo przypisanych do danej jednostki eksperymentalnej. Tak by każdy z nich był użyty (replikowany) r_i razy ($i = 1, \dots, v$). (Raghavarao, Padget, 2005) Rolę obiektów mogą przyjąć również poziomy stosowanego czynnika. Ponadto warto zwrócić uwagę na to, iż w tej metodzie nie jest wymagane założenie o normalności rozkładu badanej cechy.

Model doświadczenia jest następujący:

$$Y = \mu \mathbb{1}_n + \Delta' \underline{\tau} + \mathbf{D}' \underline{\beta} + \varepsilon, \quad (1)$$

gdzie: Y – wartość badanej zmiennej,
 μ – średnia ogólna,
 $\underline{\tau}$ – wektor efektów obiektowych,
 $\underline{\beta}$ – wektor efektów blokowych,
 ε – wektor błędów losowych,
 $\Delta = (\delta_{ws})$, gdzie:

$$\delta_{ws} = \begin{cases} 0 & s\text{-ta obserwacja nie jest związana z obiektem } w, \\ 1 & s\text{-ta obserwacja dotyczy } w\text{-tego obiektu.} \end{cases}$$

$\mathbf{D} = (\gamma_{zs})$, gdzie:

$$\gamma_{zs} = \begin{cases} 0 & s\text{-ta obserwacja nie pochodzi z } z\text{-tego bloku,} \\ 1 & s\text{-ta obserwacja pochodzi z } z\text{-tego bloku.} \end{cases}$$

Warto zwrócić uwagę na to, że w doświadczeniu prowadzonym według metody układu bloków nie ma interakcji między efektem blokowym, a efektem obiektowym. Macierz Δ jest wymiaru $v \times n$, natomiast macierz \mathbf{D} wymiaru $b \times n$.

O błędach losowych ε zakładamy, że

- 1) ε jest wektorem nieskorelowanych błędów losowych,
- 2) $\varepsilon \sim N(0_n, \sigma^2 I_n)$.

Ponadto należy zdefiniować pojęcie kontrastu:

Liniową funkcję parametryczną $\ell' \underline{\rho}$ wektora $\underline{\rho} = [\rho_1, \rho_2, \dots, \rho_d]$ nazywamy kontrastem, jeśli $\ell' \mathbb{1}_d = 0$. Kontrast nazywamy kontrastem elementarnym, jeśli ℓ ma tylko dwa niezerowe elementy i są one równe -1 i 1 (Raghavarao, Padget, 2005).

ANALIZA WEWNĄTRZBLOKOWA

Badanie prowadzone według analizy wewnątrzblokowej polega na tym, że eliminujemy efekt podziału jednostek eksperymentalnych na bloki i rozważamy problem, czy efekty pochodzące od zastosowanych obiektów mają wpływ na wyniki doświadczenia. Ze względu na nieestymowalność każdego parametru z osobna, analizujemy następujący układ hipotez:

$$\begin{aligned} H_0 : \tau_1 - \tau_v &= \tau_2 - \tau_v = \dots = \tau_{v-1} - \tau_v = 0 \\ H_1 : &\sim H_0 \end{aligned} \quad (2)$$

co możemy zrobić powołując się na następujące twierdzenie (Raghavarao, Padget, 2005):

Wszystkie elementarne kontrasty efektów obiektowych są estymowalne wtedy i tylko wtedy, gdy $\text{rz}(C) = v - 1$.

Powyższy układ hipotez możemy weryfikować za pomocą statystyki Walda (Raghavarao, Padget, 2005):

$$F = \frac{Q' C^- Q /_{v-1}}{R_0^2 /_{n-b-v+1}}, \quad (3)$$

gdzie:

$$\begin{aligned} C &= R - \mathbf{N} \mathbf{K}^{-1} \mathbf{N}', \\ C^- &\text{ - g-odwrotność macierzy } C, \\ Q &= T - \mathbf{N} \mathbf{K}^{-1} B, \\ R_0^2 &= SS_T - SS_B - SS_{T|B}. \end{aligned}$$

Macierz R jest macierzą diagonalną, $R = \text{diag}(r_1, r_2, \dots, r_v)$. Podobnie macierz K , gdzie $K = \text{diag}(k_1, k_2, \dots, k_b)$. Macierz incydencji \mathbf{N} jest wymiaru $v \times b$. Dalsze oznaczenia są następujące:

$$B = (B_1, \dots, B_b)', \text{ gdzie } B_j = \sum_{i=1}^v Y_{ij}$$

$$T = (T_1, \dots, T_v)', \text{ gdzie } T_i = \sum_{j=1}^b Y_{ij}$$

ANALIZA MIĘDZYBLOKOWA

Badanie prowadzone według analizy międzyblokowej polega na tym, że eliminujemy efekty pochodzące od zastosowanych obiektów i rozważamy problem, czy efekty wywodzące się z podziału jednostek eksperymentalnych na bloki mają wpływ na wyniki doświadczenia. Również w tym wypadku, ze względu na nieestymowalność każdego parametru z osobna, analizujemy poniższy układ hipotez:

$$\begin{aligned} H_0 &: \beta_1 - \beta_b = \beta_2 - \beta_b = \dots = \beta_{b-1} - \beta_b = 0 \\ H_1 &: \sim H_0 \end{aligned} \quad (4)$$

Wykorzystując poniższe twierdzenie (Raghavarao, Padget, 2005), które rozwiązuje problem nieestymowalności parametrów:

Wszystkie elementarne kontrasty efektów blokowych są estymowalne wtedy i tylko wtedy, gdy $\text{rz}(D^*) = b - 1$.

Hipotezę zerową możemy weryfikować przy pomocy statystyki Walda (Raghavarao, Padget, 2005):

$$F = \frac{P(D^*)^- P /_{b-1}}{R_0^2 /_{n-b-v+1}}, \quad (5)$$

gdzie:

$$D^* = K - \mathbf{N}' R^{-1} \mathbf{N},$$

$$\begin{aligned} (D^*)^- &- \text{g-odwrotność macierzy } D^*, \\ P &= B - \mathbf{N}'R^{-1}T, \\ R_0^2 &= SS_T - SS_{Tr} - SS_{B|Tr}. \end{aligned}$$

3. SCHEMAT PRZEPROWADZONEGO BADANIA

Dochód jest jednym z zasadniczych czynników służących do określania poziomu zamożności i sposobu życia, pozwalający zaspokoić potrzeby gospodarstwa domowego. Duży wpływ na wysokość dochodu mają przede wszystkim liczebność gospodarstwa oraz jego skład demograficzny, który różnicuje poziom i strukturę potrzeb członków rodziny. Często w międzynarodowych analizach porównawczych zamożności, bazujących na dochodach gospodarstw domowych, stosuje się dochód ekwiwalentny na osobę w gospodarstwie, a nie powszechnie stosowany dochód na osobę. Skala ekwiwalentności koryguje dochód na osobę. Uwzględnia fakt, iż wraz ze wzrostem liczby osób w gospodarstwie domowym jego dochody nie rosną liniowo. Ponadto, bierze pod uwagę nie tylko liczbę osób wchodzących w skład gospodarstwa, ale także ich wiek. W badaniu posłużono się zmodyfikowaną skalą OECD liczoną według poniższego wzoru (GUS, 2011):

$$e = 1 + 0,5(LA - 1) + 0,3LD,$$

gdzie LA to liczba osób w wieku co najmniej 14 lat, a LD to liczba osób wchodzących w skład gospodarstwa w wieku co najwyżej 13 lat.

Celem analizy była weryfikacja hipotez o wpływie podziału gospodarstw domowych Polski na województwa i klasy miejscowości na wielkość średniego dochodu ekwiwalentnego i średniego dochodu na osobę w 2011 roku w oparciu o metodę analizy blokowej.

W badaniu wykorzystano dane pochodzące z Diagnozy Społecznej 2011 przeprowadzonej w Polsce przez Radę Monitoringu Społecznego. Badanie przyjmuje formę badania panelowego. Jest ono prowadzone od 2000 roku. W 2011 roku przeprowadzono je po raz szósty. Ankietyzacja respondentów miała miejsce w marcu, w celu uniknięcia efektu sezonowości. Opublikowany raport ma na celu charakterystykę polskiego społeczeństwa 22 lata po zmianie systemu, 7 lat po wstąpieniu Polski do Unii Europejskiej i 12 lat po pierwszym badaniu w ramach tego samego projektu. (Czapiński, Panek, 2012) Pytania zawarte w ankietach dotyczyły gospodarstw domowych oraz ich członków i odnosiły się między innymi do zachowań osób wchodzących w skład gospodarstwa domowego, ich postaw, a także dochodu, zasobności materialnej, kredytów, edukacji, leczenia. Baza danych zawiera informacje o 12387 gospodarstwach domowych.

Raport Diagnozy Społecznej 2011 (Czapiński, Panek, 2012) dostarcza dużo interesujących informacji dotyczących dochodów gospodarstw domowych dla poszczególnych województw. Na przykład potwierdza znany fakt, że w województwie mazowieckim obserwuje się najwyższe dochody netto na osobę zarówno w przeliczeniu na

jednostkę ekwiwalentną (czyli w przeliczeniu według zmodyfikowanej skali ekwiwalentności OECD), czy na gospodarstwo domowe. Natomiast najniższe dochody w województwie lubelskim. Interesujące wnioski można również wyciągnąć na temat dochodów netto w podziale na klasy miejscowości: wraz ze wzrostem liczby mieszkańców rosną dochody netto na gospodarstwo domowe, na osobę jak i na jednostkę ekwiwalentną.

Na potrzeby badania zostały przyjęte następujące definicje:

Obiekty – klasy miejscowości

W badaniach prowadzonych przez GUS, jak i w Diagnostyce Społecznej, określane w sześciu kategoriach. Warianty klas miejscowości:

1. miasta o liczbie mieszkańców 500 tys. i więcej,
2. miasta o liczbie mieszkańców 200-500 tys.,
3. miasta o liczbie mieszkańców 100-200 tys.,
4. miasta o liczbie mieszkańców 20-100 tys.,
5. miasta o liczbie mieszkańców poniżej 20 tys.,
6. wieś.

Bloki – Województwa

W analizie jako zmienne badane przyjęto:

1. średni miesięczny dochód ekwiwalentny netto gospodarstwa domowego
2. średni miesięczny dochód netto na osobę w gospodarstwie domowym

W tabelach 1 oraz 2 przedstawiono dane charakteryzujące przebadaną zbiorowość pod względem średniego dochodu ekwiwalentnego netto na osobę oraz średniego dochodu netto gospodarstwa domowego na osobę. Braki wartości oznaczają braki informacji, że w odniesieniu do danej kategorii nie udzielono ankietom odpowiedzi bądź też w obrębie danego województwa nie ma miejscowości należących do danej klasy.

Tabela 1.

Średni dochód ekwiwalentny netto na osobę w badanych gospodarstwach domowych według klasy miejscowości zamieszkania i województw

Średni dochód ekwiwalentny netto na osobę (zł)						
Województwo	Klasa miejscowości zamieszkania					
	miasta o liczbie mieszkańców (tys.)					wieś
	500 i więcej	200-500	100-200	20-100	poniżej 20	
Dolnośląskie	2329,07	.	1706,93	1766,58	1566,61	1456,73
Kujawsko-pomorskie	.	2139,47	1331,19	1304,97	1369,91	1358,93
Lubelskie	.	2235,49	.	1709,96	1635,82	1161,71
Lubuskie	.	.	2066,32	1765,27	1531,17	1379,55
Łódzkie	2010,86	.	.	1657,45	1506,43	1346,54
Małopolskie	2259,96	.	1883,38	1817,58	1660,55	1473,03
Mazowieckie	2952,28	1538,25	1985,45	1696,79	2100,87	1441,08
Opolskie	.	.	2099,01	1657,39	1558,12	1593,01

cd. Tabela 1.

Podkarpackie	.	.	1868,37	1755,93	1794,21	1201,60
Podlaskie	.	635,84	.	1669,73	1950,75	1289,40
Pomorskie	.	2177,79	1438,65	1896,71	1688,98	1553,60
Śląskie	.	2493,12	2015,51	1861,44	1666,24	1614,93
Świętokrzyskie	1300,00	1676,83	.	1578,74	1510,31	1279,96
Warmińsko-mazurskie	.	.	1687,19	1743,06	1427,50	1380,62
Wielkopolskie	2344,70	.	1568,84	1772,48	8,51	1473,92
Zachodnio-pomorskie	.	2265,28	2053,07	1664,83	1715,49	1223,96

Źródło: opracowanie własne na podstawie danych z badania Diagnoza Społeczna 2011

Tabela 2.

Średni dochód netto gospodarstwa domowego na osobę w badanych gospodarstwach domowych według klasy miejscowości zamieszkania i województw

Średni dochód netto gospodarstwa domowego na osobę (zł)						
Województwo	Klasa miejscowości zamieszkania					
	Miasta o liczbie mieszkańców (tys.)					wieś
	500 i więcej	200-500	100-200	20-100	poniżej 20	
Dolnośląskie	1734,17	.	1181,22	1344,15	1196,16	1029,65
Kujawsko-pomorskie	.	1583,93	931,66	980,80	1012,19	942,14
Lubelskie	.	1776,31	.	1229,77	1217,47	828,38
Lubuskie	.	.	1536,46	1343,17	1120,37	944,33
Łódzkie	1572,74	.	.	1231,04	1071,29	985,22
Małopolskie	1755,71	.	1198,33	1367,53	1180,89	1014,59
Mazowieckie	2250,20	1185,55	1492,99	1261,27	1458,72	1004,65
Opolskie	.	.	1492,95	1202,81	1213,74	1113,85
Podkarpackie	.	.	1504,76	1275,48	1184,66	804,24
Podlaskie	.	1242,11	.	1182,15	1366,78	880,61
Pomorskie	.	1587,28	1115,18	1439,30	1231,47	1085,55
Śląskie	.	1883,86	1576,45	1374,06	1248,17	1141,05
Świętokrzyskie	1300,00	1246,24	.	1129,28	1074,24	893,40
Warmińsko-mazurskie	.	.	1263,93	1304,95	974,25	965,61
Wielkopolskie	1814,90	.	1086,57	1310,09	1130,66	1006,39
Zachodnio-pomorskie	.	1703,05	1500,31	1228,13	1270,35	852,36

Źródło: opracowanie własne na podstawie danych z badania Diagnoza Społeczna 2011

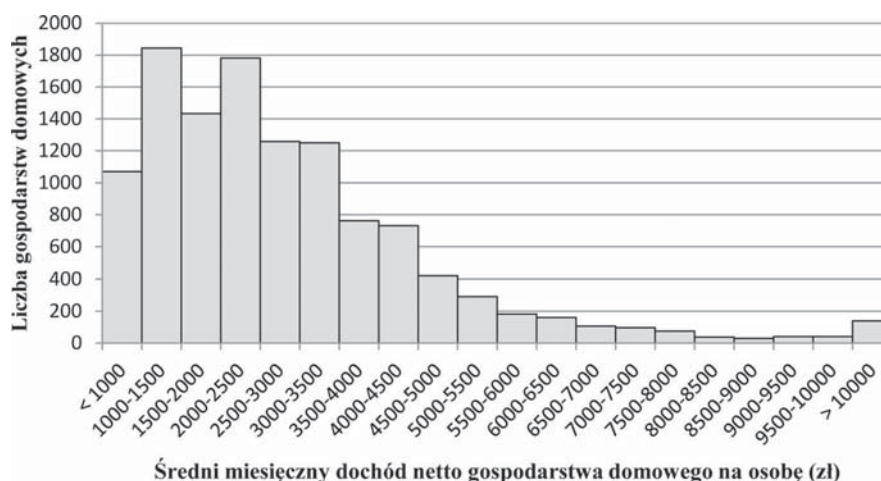
Przed przeprowadzeniem badania sprawdzono jak kształtuje się rozkład gospodarstw domowych ze względu na średni miesięczny dochód netto na osobę. Na podstawie podstawowych charakterystyk (tab. 3) jak i histogramu (rys. 1) można wnioskować

o silnej prawostronnej asymetrii rozkładu, a więc zdecydowanej przewadze gospodarstw domowych o zarobkach do 3613 zł. Największą różnicę w średnim dochodzie netto na osobę między największą i najmniejszą klasą miejscowości zaobserwować można w województwie mazowieckim, gdzie wynosi ona ponad 1200 zł. Natomiast najmniejszą w województwie świętokrzyskim, równą blisko 400 zł.

Tabela 3. Charakterystyki rozkładu średniego miesięcznego dochodu netto gospodarstw domowych na osobę (zł)

Statystyki opisowe	
Dominanta	1326,71
Błąd standardowy	23,22
Mediana	2428,27
Kwartyl 1	1507,94
Kwartyl 3	3613,02
Odchylenie ćwiartkowe	1052,54
Pozycyjny wsp. zmienności	43%
Pozycyjny wsp. asymetrii	1,12
Zakres	52220
Minimum	100
Maksimum	52320

Źródło: opracowanie własne na podstawie danych z badania Diagnoza Społeczna 2011



Rysunek 1. Rozkład gospodarstw domowych ze względu na średni miesięczny dochód netto na osobę (zł)

Źródło: opracowanie własne na podstawie danych z badania Diagnoza Społeczna 2011

Tabela 4.

Wyniki analizy wewnątrzblokowej dla średniego miesięcznego dochodu ekwiwalentnego netto oraz średniego miesięcznego dochodu netto gospodarstwa domowego na osobę

	Zmienna i postać hipotezy zerowej	Wyniki analizy
Analiza wewnątrz-blokowa	Średni miesięczny dochód ekwiwalentny netto na osobę $H_0 : \alpha_1 - \alpha_6 = \alpha_2 - \alpha_6 = \dots = 0$	F = 15,1458 p-value = 0,0000 df1 = 5 df2 = 53
	Średni miesięczny dochód netto gospodarstwa domowego na osobę $H_0 : \alpha_1 - \alpha_6 = \alpha_2 - \alpha_6 = \dots = 0$	F = 29,1003 p-value = 0,00000 df1 = 5 df2 = 53

Źródło: opracowanie własne na podstawie danych z badania Diagnoza Społeczna 2011

Opierając się na przedstawionych wyżej wynikach można powiedzieć, że podział gospodarstw domowych pod względem klas miejscowości ma wpływ zarówno na wysokość średniego miesięcznego dochodu ekwiwalentnego netto na osobę jak i na wysokość średniego miesięcznego dochodu netto gospodarstwa domowego na osobę.

Tabela 5.

Wyniki analizy międzyblokowej dla średniego miesięcznego dochodu ekwiwalentnego netto oraz średniego miesięcznego dochodu netto gospodarstwa domowego na osobę

	Zmienna i postać hipotezy zerowej	Wyniki analizy
Analiza między-blokowa	Średni miesięczny dochód ekwiwalentny netto na osobę $H_0 : \beta_1 - \beta_{16} = \beta_2 - \beta_{16} = \dots = 0$	F = 1,68843 p-value = 0,08210 df1 = 15 df2 = 53
	Średni miesięczny dochód netto gospodarstwa domowego na osobę $H_0 : \beta_1 - \beta_{16} = \beta_2 - \beta_{16} = \dots = 0$	F = 1,47680 p-value = 0,14810 df1 = 15 df2 = 53

Źródło: opracowanie własne na podstawie danych z badania Diagnoza Społeczna 2011

Na podstawie powyższych wyników można stwierdzić, iż podział gospodarstw domowych pod względem województw nie ma wpływu na wysokość średniego miesięcznego dochodu ekwiwalentnego netto na osobę. Podobne wnioski można sformułować dla średniego miesięcznego dochodu netto gospodarstwa domowego na osobę, na którego wysokość nie ma wpływu podział gospodarstw domowych pod względem województw. (p-value oznacza poziom istotności, df1 oraz df2 liczbę stopni swobody dla statystyki F)

Dla porównania otrzymanych wyników przeprowadzono test jednoczynnikowej analizy wariancji (ANOVA) (por. tabele 6 i 7).

W rezultacie można stwierdzić, że zarówno wyniki analizy blokowej jak i jednoczynnikowej analizy wariancji pokrywają się. Wartości statystyk empirycznych, graniczny poziom istotności jak i stopnie swobody są takie same dla analizy wewnątrzblo-

Tabela 6.

Wyniki testu ANOVA dla średniego dochodu netto gospodarstwa domowego na osobę

Źródło	Suma kwadratów	Stopnie swobody	Średnia suma kwadratów	F	Istotność
Model skorygowany	3851910,084	20	192595,504	8,524	0,000
Stała	105112197,197	1	105112197,19	4652,02	0,000
Województwo	500524,599	15	33368,307	1,477	0,148
Klasa miejscowości	3287599,430	5	657519,886	29,100	0,000
błąd	1197532,814	53	22594,959		
Ogółem	118475919,276	74			
Ogółem skorygowane	5049442,898	73			

Źródło: opracowanie własne na podstawie danych z badania Diagnoza Społeczna 2011

Tabela 7.

Wyniki testu ANOVA dla średniego dochodu ekwiwalentnego netto na osobę

Źródło	Suma kwadratów	Stopnie swobody	Średnia suma kwadratów	F	Istotność
Model skorygowany	5374872,916	20	268743,646	5,093	0,000
Stała	197338923,890	1	197338923,89	3739,64	0,000
Województwo	1336459,968	15	89097,331	1,688	0,082
Klasa miejscowości	3996168,106	5	799233,621	15,146	0,000
błąd	2796779,892	53	52769,432		
Ogółem	225828476,852	74			
Ogółem skorygowane	8171652,808	73			

Źródło: opracowanie własne na podstawie danych z badania Diagnoza Społeczna 2011

kowej (zaznaczone kursywą) oraz analizy międzyblokowej. Należałoby zatem sprawdzić, czy w ogólności obie postacie statystyk prowadzą do tych samych wyników.

4. WNIOSKI

- Podział gospodarstw domowych według województw nie ma wpływu na średni dochód netto gospodarstwa domowego na osobę oraz średni miesięczny dochód ekwiwalentny netto na osobę.
- W odniesieniu do podziału jednostek według klas miejscowości, podział ten ma wpływ na wartości badanych zmiennych.
- Wyniki przeprowadzonych badań analizy blokowej i jednoczynnikowej analizy wariancji prowadzą do takich samych wartości statystyk testowych.

- Metodę analizy blokowej można stosować w przypadku, gdy nie jest spełnione założenie o rozkładzie normalnym badanej zmiennej wymagane dla jednoczynnikowej analizy wariancji.
- Różnica między metodą układu bloków, a jednoczynnikową analizą wariancji tkwi w tym, że w analizie blokowej jedna obserwacja może dotyczyć kilku jednostek. Oznacza to otrzymanie macierzyowej postaci wyjściowej ogólnego modelu doświadczenia. Parametry nie są estymowalne z osobna, zatem zmienia się postać hipotezy zerowej, w której analizujemy kontrasty (różnice parametrów).

Uniwersytet Ekonomiczny w Poznaniu

LITERATURA

- [1] Czapiński J., Panek T. (red.), (2012), *Diagnoza społeczna 2011*, Warszawa, <http://www.diagnoza.com> [dostęp: 15.02.2012].
- [2] GUS, (2011), *Budżety gospodarstw domowych w 2010r.*, Warszawa, s. 19.
- [3] Krzyśko M., (2000), *Wielowymiarowa analiza statystyczna*, UAM, Poznań.
- [4] Oktaba W., (1974), *Elementy statystyki matematycznej i metodyka doświadczalnictwa*, Państwowe Wydawnictwo Naukowe PWN, Warszawa.
- [5] Raghavarao D., Padgett L.V., (2005), *Block designs*, World Scientific, Singapore.
- [6] Rao R.C., (1982), *Modele liniowe statystyki matematycznej*, Państwowe Wydawnictwo Naukowe PWN, Warszawa.

ZASTOSOWANIE ANALIZY BLOKOWEJ NA PRZYKŁADZIE BADANIA ZRÓŻNICOWANIA DOCHODÓW GOSPODARSTW DOMOWYCH

Streszczenie

Analiza blokowa to rodzaj wielowymiarowej analizy danych często stosowany w badaniach związanych z rolnictwem. Badania te zwykle przeprowadza się w celu oceny wpływu na wzrost i rozwój roślin różnych czynników takich jak nawozy mineralne, czy środki ochrony roślin. Głównym jej prekursorem był brytyjski statystyk R. A. Fisher, który na początku XX wieku sformułował schemat układu bloków kompletnie zrandomizowanych. Jest to szczególnie przypadek układu bloków, w którym badamy wpływ danego obiektu zastosowanego na losowo wybranych jednostkach eksperymentalnych podzielonych na równoliczne i możliwie jednorodnie grupy.

Celem pracy jest zaprezentowanie wyników zastosowania analizy blokowej w badaniu dochodu ekwiwalentnego gospodarstw domowych. Jako źródło informacji wykorzystano dane zgromadzone przez Radę Monitoringu Społecznego w ramach Diagnozy Społecznej 2011. Baza zawierała informacje o 12387 gospodarstwach domowych. Każde z nich zostało opisane za pomocą 259 zmiennych. W przeprowadzonym badaniu, w którym dokonano analizy wewnątrz- i międzyblokowej, uwzględnione zostały zmienne takie jak: położenie geograficzne gospodarstwa domowego, liczba osób wchodzących w skład gospodarstwa oraz klasa miejscowości.

Słowa kluczowe: analiza wewnątrzblokowa, analiza międzyblokowa, dochód ekwiwalentny

APPLICATION OF THE BLOCK ANALYSIS TO THE INVESTIGATION OF HOUSEHOLD
INCOME DIVERSITY

A b s t r a c t

Analysis of the block designs is one of multivariate data analysis methods, which is often used in a agricultural. This type of research is usually carried out to assess the impact of various factors on the growth and development of plants, such as fertilizers pesticides and others. The main precursor of block designs was a British statistician R. A. Fisher, who in early twentieth century formulated a randomized block designs. It is a particular kind of the block designs, in which we examine the impact of an object used to randomly selected experimental units divided into equinumerous and possibly homogenous groups.

The aim of this study is to present results of applying the block analysis in a equivalent household income research. Data was obtained from the Council for Social Monitoring – The Social Diagnosis 2011. The base contained of 12387 households, which were described by 259 variables. The variables which were used in a intra-and inter-block analysis research were: administrative location of household, number of people included into the household and class of the village.

Key words: intrablock analysis, interblock analysis, equivalent household