

**Małgorzata Szerszunowicz**

Uniwersytet Ekonomiczny w Katowicach

# **O ZASTOSOWANIU METOD PLANOWANIA EKSPERYMENTÓW W ANALIZIE DANYCH PRZESTRZENNYCH**

## **Wprowadzenie**

Obecnie w analizach statystycznych poszukuje się metod i narzędzi, które pozwolą na zrozumienie zachodzących przemian ekonomicznych i społecznych. Jednym ze sposobów odpowiedniego opisu charakteru badanych zjawisk jest uwzględnienie ich lokalizacji przestrzennej, co prowadzi do analizy danych o charakterze przestrzennym. Od połowy XX w. rozwija się metodologia analizy danych przestrzennych będąca swoistym połączeniem metod ekonometrii i statystyki przestrzennej.

Planowanie eksperymentów jest narzędziem statystycznej kontroli jakości szeroko wykorzystywanym w praktyce przedsiębiorstw produkcyjnych. Wdrożenie procedur planowania doświadczeń przed rozpoczęciem procesu produkcyjnego prowadzi do istotnego polepszenia rezultatów procesu i ma wpływ na jego aspekt ekonomiczny.

Analizy przestrzenne danych dotyczących np. rozwoju przemysłu czy stopy bezrobocia prowadzą do wyodrębnienia obszarów o niskim (województwa wschodnie) i wysokim poziomie rozwoju gospodarczego (województwa zachodnie). Przedmiotem niniejszego opracowania jest podjęcie problemu przestrzennego zróżnicowania poziomu wynagrodzeń w Polsce. W tym celu, w analizie danych przestrzennych, wykorzystane zostanie narzędzie planowania eksperymentów.

## **1. Podstawy analizy danych przestrzennych**

Pierwsze publikacje dotyczące danych o charakterze przestrzennym pojawiały się od lat 50. XX w., lecz za początek odrębnej dziedziny zajmującej się analizą danych przestrzennych uznaje się lata siedemdziesiąte XX w., kiedy to J.H.P. Paelinck wprowadził pojęcie ekonometrii przestrzennej (Paelinck, Klaassen, 1983).

Wraz ze wzrostem znaczenia zależności przestrzennych, metody ekonometrii przestrzennej znalazły zastosowanie nie tylko w badaniach regionalnych, geografii ekonomicznej czy analizach rynku nieruchomości, ale również w ekonomii oraz nowej ekonomii geograficznej. Następstwem intensywnego rozwoju metodologii ekonometrii przestrzennej są jej powiązania z innymi dziedzinami wiedzy, takimi jak ekonometria klasyczna, geografia, ekonomia, statystyka matematyczna, statystyka przestrzenna, eksploracyjna analiza danych przestrzennych oraz taksonomia (Zeliaś, 1991).

Jednym z początkowych etapów analizy danych przestrzennych jest ich klasyfikacja. Dane o charakterze przestrzennym można podzielić ze względu na ich źródło pochodzenia lub według typu informacji (Suchecki, 2010). W niniejszym opracowaniu wykorzystane zostaną dane statystyczne dostępne w Banku Danych Lokalnych Głównego Urzędu Statystycznego. Zgodnie z wymienionymi kryteriami klasyfikacji, ze względu na źródło pochodzenia, są to dane o charakterze ekonomiczno-społecznym. W związku z tym, że wartości wykorzystanych zmiennych są obserwowane dla pewnych fragmentów powierzchni, dane te są danymi obszarowymi.

Zazwyczaj w analizie danych przestrzennych ważna jest ich wizualizacja. Najczęściej w tym celu wykorzystuje się mapy uwzględniające podział administracyjny badanego obszaru lub mapy z naniesionymi izoliniami informującymi o natężeniu badanego zjawiska (Kopczewska, 2011). Ponadto wyróżnia się mapy kropkowe, kartodiagramy oraz trellisy (Suchecki, 2010).

Metodologia ekonometrii przestrzennej jest ściśle związana ze statystyką przestrzenną, która dostarcza właściwych metod pozwalających na opis złożonych struktur przestrzennych. Metody te pozwalają na wskazanie występujących prawidłowości przestrzennych lub określenie stopnia skorelowania zmiennych w danym obszarze. Ponadto dzięki statystyce przestrzennej możliwe jest rozpoznanie losowości badanego zjawiska w przestrzeni.

W ramach statystyki przestrzennej wyróżnia się trzy klasy metod stosowanych dla różnych typów danych:

- metody geostatyczne (dane powierzchniowe z atrybutami ilościowymi lub jakościowymi),
- metody analizy danych punktowych (obiektów dyskretnych),
- metody analizy danych obszarowych i punktowych atrybutowych.

Ze względu na rodzaj danych wykorzystywanych w niniejszym opracowaniu zostanie scharakteryzowana jedynie trzecia grupa metod statystyki przestrzennej.

Dane statystyczne określone dla obszarów lub punktów atrybutowych dotyczą zazwyczaj jednostek administracyjnych lub obiektów, które mają granice i mogą być opisane za pomocą współrzędnych geograficznych. Wizualizacja i analiza ilościowa danych obszarowych odbywa się zazwyczaj za pomocą technik eksploracyjnej analizy danych przestrzennych (ESDA – *Exploratory Spatial Data Analysis*), których celem jest m.in. prezentacja graficzna rozkładu przestrzennego zmiennych, identyfikacja obserwacji nietypowych oraz określenie reżimów przestrzennych. Przede wszystkim techniki eksploracyjnej analizy danych przestrzennych powinny umożliwić jasny opis danych i ich zrozumienie, ze szczególnym uwzględnieniem badania heterogeniczności i przestrzennej zależności. Pod pojęciem heterogeniczności przestrzennej rozumieć należy monotoniczne zmiany rozmieszczenia w przestrzeni wartości badanej zmiennej. W analizie danych przestrzennych jest również ważne określenie stopnia autokorelacji przestrzennej, czyli zjawiska grupowania się wysokich lub niskich wartości badanej zmiennej (autokorelacja dodatnia) bądź odwrotność tego zjawiska. W celu weryfikacji przestrzennej zależności badanej zmiennej wykorzystuje się statystyki globalne i lokalne. Statystyki globalne to jednoczłone wskaźniki informujące o stopniu ogólnego podobieństwa regionów, natomiast statystyki lokalne wyznaczane dla wszystkich uwzględnionych obszarów informują o skali podobieństwa bądź różnicy w porównaniu do obszarów sąsiadujących. Najczęściej stosowanymi statystykami globalnymi i lokalnymi są statystyki Morana oraz Geary’ego.

K. Lewandowska-Gwarda (2013) w swoim opracowaniu podkreśla rolę przestrzeni w różnych badaniach ekonomicznych. Natomiast w opublikowanym raporcie Głównego Urzędu Statystycznego (Główny Urząd Statystyczny, 2013) ukazały się dane dotyczące przeciętnego poziomu wynagrodzeń m.in. według podziału na regiony i województwa, na podstawie których można zauważyć istotne różnice przeciętnego wynagrodzenia pomiędzy poszczególnymi regionami. Można zatem rozważyć uwzględnienie czynnika przestrzennego w charakterystyce poziomu wynagrodzeń w Polsce.

## 2. Zarys teorii planowania eksperymentów

Metody planowania eksperymentu zostały po raz pierwszy zastosowane w doświadczalnictwie rolniczym przez R.A. Fishera, w latach 20. XX w. Skuteczność metod planowania doświadczeń sprawiła, że w późniejszych latach znalazły one zastosowanie w biologii, medycynie, psychologii, jak również w statystycznej kontroli jakości.

Planowanie eksperymentów, jako jedno z narzędzi statystycznej kontroli jakości, pozwala na określenie czynników najsilniej oddziałujących na zmienną charakteryzującą badany proces, jak również umożliwia wskazanie wartości czynników, dla których zmienna wynikowa osiąga pożądaną wartość lub najmniejszą zmienność. W celu prawidłowego przeprowadzenia planowania doświadczeń należy m.in. szczegółowo sformułować problem, odpowiednio dobrać czynniki i zmienną charakteryzującą badany proces produkcyjny. Zasadniczym etapem eksperymentu jest jego realizacja oraz odpowiednia analiza otrzymanych wyników, na podstawie której zostaną sformułowane wnioski i zalecenia dla procesu produkcyjnego (Montgomery, 1997).

Ogólnie eksperyment należy rozumieć jako ciąg  $n$  kolejnych doświadczeń będących jednorazowym uzyskaniem wartości zmiennej objaśnianej  $Y$ , przy ustalonych wartościach czynników  $X_1, X_2, \dots, X_m$ , określonych odpowiednio na zbiorach  $\Xi_1, \Xi_2, \dots, \Xi_m$ . Obszarem eksperymentowania określa się zbiór punktów  $\mathbf{x} = (x_1, x_2, \dots, x_m)$ , gdzie  $x_i \in \Xi_i, i = 1, 2, \dots, m$ . Zbiór par  $P_n = \{\mathbf{x}_j, p_j\}_{j=1}^n$  określa się planem eksperymentu obejmującym  $n$  doświadczeń, gdzie  $\mathbf{x}_j = (x_{1j}, x_{2j}, \dots, x_{mj})$  oraz  $p_j = \frac{n_j}{n}$ , przy czym  $n_j$  oznacza liczbę doświadczeń w punkcie

$\mathbf{x}_j$  obszaru eksperymentowania, ponadto  $\sum_{j=1}^n n_j = n$  oraz  $\sum_{j=1}^n p_j = 1$  dla  $j = 1, 2, \dots, n$ .

Zależność pomiędzy zbiorem zmiennych objaśniających a ustaloną zmienną objaśnianą można przedstawić w postaci modelu statystycznego

$$Y(X_1, X_2, \dots, X_m) = y(X_1, X_2, \dots, X_m), \quad (1)$$

przy czym  $EY(X_1, X_2, \dots, X_m) = y(X_1, X_2, \dots, X_m)$ ,  $E\varepsilon = 0$  oraz  $V\varepsilon = \sigma^2$ , a  $\sigma^2$  jest wielkością stałą, niezależną od wartości poszczególnych czynników. Model ten przedstawić można w postaci ogólnego modelu liniowego (Wawrzynek, 2009) postaci  $\mathbf{Y} = \mathbf{F}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$  gdzie

$$\mathbf{Y}^T = (Y_1 \ Y_2 \ \dots \ Y_n) \quad (2)$$

$$\boldsymbol{\varepsilon}^T = (\varepsilon_1 \ \varepsilon_2 \ \dots \ \varepsilon_n) \quad (3)$$

$$\boldsymbol{\beta}^T = [\beta_1 \ \beta_2 \ \dots \ \beta_k] \quad (4)$$

$$\mathbf{f}^T(\mathbf{x}) = (f_1(\mathbf{x}) \ f_2(\mathbf{x}) \ \dots \ f_k(\mathbf{x})) \quad (5)$$

$$\mathbf{F} = \begin{bmatrix} f_1(\mathbf{x}_1) & \dots & f_k(\mathbf{x}_1) \\ \vdots & \ddots & \vdots \\ f_1(\mathbf{x}_n) & \dots & f_k(\mathbf{x}_n) \end{bmatrix} = [\mathbf{f}(\mathbf{x}_1) \ \mathbf{f}(\mathbf{x}_2) \ \dots \ \mathbf{f}(\mathbf{x}_n)]^T \quad (6)$$

oraz  $f_i(\mathbf{x}_j) \equiv x_{ij}$ , dla  $i = 1, 2, \dots, k, j = 1, 2, \dots, n$ .

Ważnym etapem w planowaniu eksperymentów jest wyznaczenie postaci funkcji powierzchni odpowiedzi opisującej ogólny charakter zależności zmiennej objaśniającej od wartości czynników. Funkcję powierzchni odpowiedzi definiuje się za pomocą równania  $y = F\beta$ , gdzie macierz  $F$  i wektor  $\beta$  są określone za pomocą wzorów (4) i (6) odpowiednio. Realizacja klasycznych planów eksperymentów prowadzi zazwyczaj do estymacji parametrów liniowej funkcji powierzchni odpowiedzi postaci (Wawrzynek, 2009)

$$y(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_m x_m + \beta_{12} x_1 x_2 + \dots + \beta_{12\dots m} x_1 x_2 \dots x_m. \quad (7)$$

W literaturze rozważa się również funkcje powierzchni odpowiedzi będące wielomianami wyższych stopni (Wawrzynek, 1993). Estymacji parametrów funkcji powierzchni odpowiedzi dokonuje się za pomocą metody najmniejszych kwadratów.

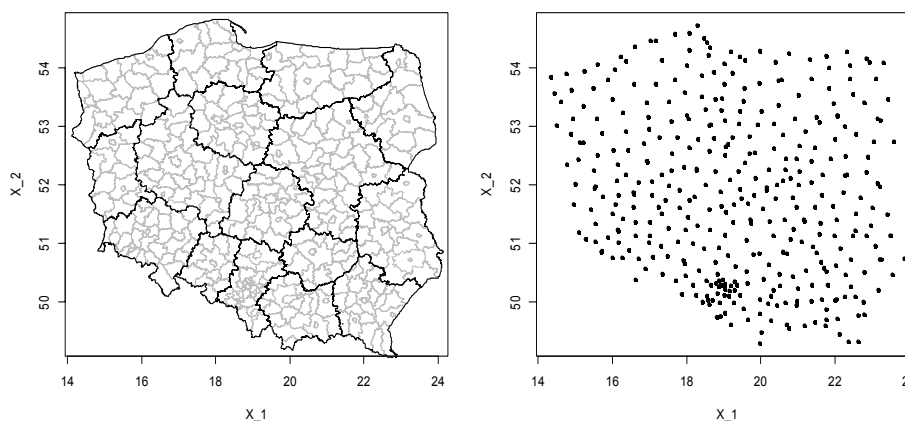
### 3. Wykorzystanie planowania eksperymentów w celu uzupełnienia analizy przestrzennej wynagrodzeń w Polsce

Wykorzystanie planowania eksperymentów w badaniach przestrzennych zostało zaproponowane m.in. przez W.G. Müllera (2010). Rozważa on wykorzystanie planowania eksperymentów w celu optymalnego rozmieszczenia czujników monitorujących stan zanieczyszczeń powietrza w pewnym regionie. Niniejszy rozdział stanowi propozycję implementacji elementów planowania eksperymentów w celu uzupełnienia analizy wynagrodzeń w Polsce.

Badaniu będą podlegać dane dotyczące wynagrodzeń w ujęciu brutto z uwzględnieniem powiatów Polski w roku 2012. W celu prezentacji problemu analizy przestrzennej wynagrodzeń w Polsce jako zagadnienia planowania doświadczeń wprowadzone zostały odpowiednie oznaczenia. Zmienną objaśnianą  $Y$  określono jako wartość wynagrodzenia brutto dla każdego z powiatów, zatem zmienna ta przyjmuje 380 wartości. W eksperymencie uwzględniono dwa czynniki  $X_1, X_2$  przyjmujące wartości ze zbiorów  $\Xi_1, \Xi_2$ , które zawierają odpowiednie wartości szerokości i długości geograficznej środków rozważanych obszarów. Wówczas obszar eksperymentowania utworzonego planu jest postaci

$$D = \{ (x_1, x_2) : x_1 \in \Xi_1, x_2 \in \Xi_2 \} \quad (8)$$

i składa się z 380 punktów odpowiadających środkom powiatów. Graficzną prezentację obszaru eksperymentowania przedstawiono na rysunku 1.



Rys. 1. Graficzna prezentacja obszaru eksperymentowania  $D$

Zgodnie z przedstawioną w poprzednim rozdziale terminologią, plan tak skonstruowanego eksperymentu obejmuje 380 doświadczeń.

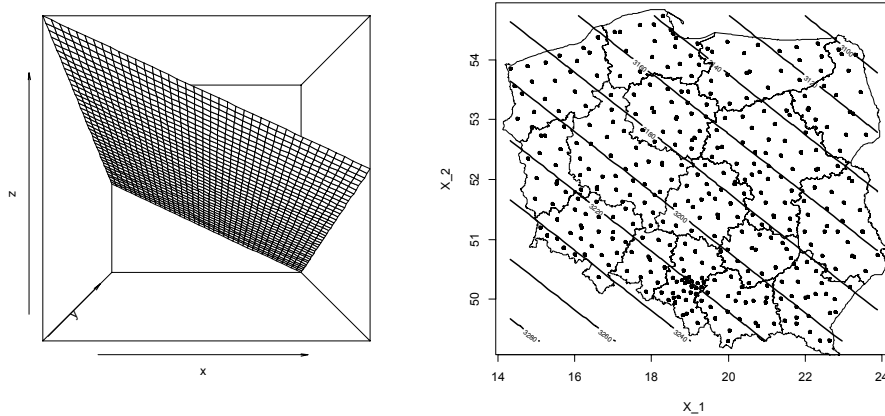
Niech przestrzenna zależność wynagrodzeń określona będzie za pomocą liniowej funkcji powierzchni odpowiedzi postaci:

$$y(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2. \quad (9)$$

Realizacja przedstawionego planu eksperymentu prowadzi do estymacji parametrów funkcji powierzchni odpowiedzi postaci (9). W wyniku przeprowadzonej estymacji otrzymano oszacowania wartości parametrów  $\beta_0$ ,  $\beta_1$ ,  $\beta_2$ , wówczas funkcję powierzchni odpowiedzi zapisać można za pomocą równania:

$$y(x) = 4429,24 - 10,16x_1 - 20,20x_2. \quad (10)$$

Wyznaczoną funkcję prezentującą przestrzenne rozmieszczenie wynagrodzeń w Polsce prezentuje rysunek 2. Zobrazowano również obszar eksperymentowania wraz z naniesionymi izoliniami określającymi zmienność przestrzenną wynagrodzeń w Polsce.



Rys. 2. Funkcja powierzchni odpowiedzi charakteryzująca rozmieszczenie przestrzenne wynagrodzeń w Polsce oraz obszar eksperymentowania z naniesionymi izoliniami

Analogiczne rozważania przeprowadzono dla funkcji powierzchni odpowiedzi postaci:

$$y(x) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2 + \beta_4 x_1^2 + \beta_5 x_2^2 \quad (11)$$

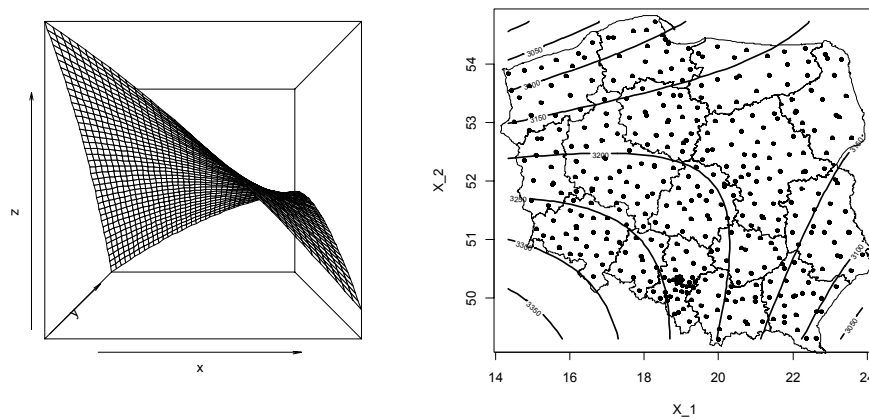
oraz

$$y(x) = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_4 x_2^2 + \beta_5 x_1^3 + \beta_6 x_2^3. \quad (12)$$

W wyniku przeprowadzonej estymacji funkcji powierzchni odpowiedzi (11) na podstawie wyników eksperymentu otrzymano funkcję wyrażoną równaniem:

$$y(x) = -520,36x_1 + 342,77x_2 - 1,15x_1 x_2 - 5,49x_1^2 + 10,66x_2^2. \quad (13)$$

Natomiast graficzną prezentację funkcji powierzchni odpowiedzi i jej odpowiednika w postaci mapy izochromatycznej prezentuje rysunek 3.

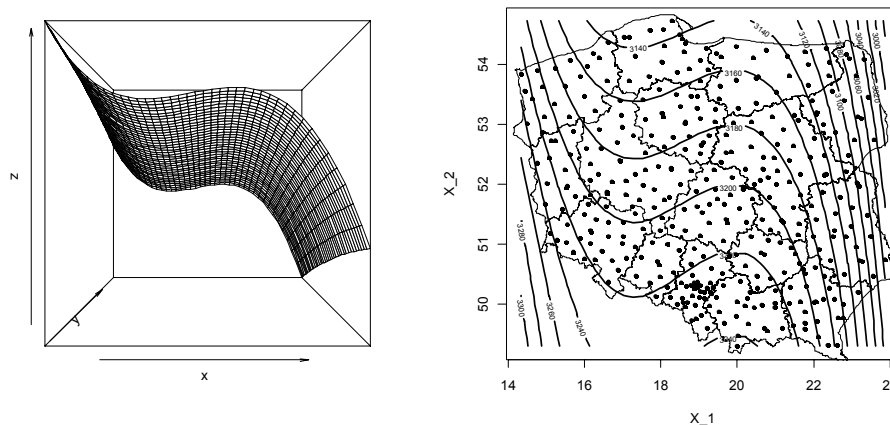


Rys. 3. Graficzna prezentacja funkcji powierzchni odpowiedzi (13) z odpowiadającą jej mapą izochromatyczną

Uzyskane oszacowanie powierzchni odpowiedzi (12) dane jest równaniem:

$$y(\mathbf{x}) = -1377,07x_1 + 653,03x_2 + 74,44x_1^2 - 11,83x_2^2 - 1,33x_1^3 + 0,07x_2^3, \quad (14)$$

które definiuje powierzchnię odpowiedzi przedstawioną na rysunku 4.



Rys. 4. Graficzna prezentacja funkcji powierzchni odpowiedzi (14) z odpowiadającą jej mapą izochromatyczną

Powyższe rysunki świadczą o tym, że istotne jest poszukiwanie takiej postaci funkcji powierzchni odpowiedzi, która będzie się charakteryzować najlepszym dopasowaniem do danych empirycznych, czyli będzie najlepiej odwzorowywać przestrzenne zróżnicowanie wynagrodzeń w Polsce. Ponadto graficzna prezentacja wyznaczonych funkcji powierzchni odpowiedzi postaci (10), (13) i (14) w prosty sposób ukazuje nierównomierności poziomu wynagrodzeń w Polsce. W szczególności powierzchnia odpowiedzi postaci (14) najlepiej z wyznaczonych funkcji przedstawia rozmieszczenie poziomu wynagrodzeń wskazując wyraźną różnicę pomiędzy wysokością wynagrodzeń między wschodnią a centralną i zachodnią częścią Polski.

Przedstawione zastosowanie narzędzia planowania eksperymentów może zostać szerzej wykorzystane w analizie struktury wynagrodzeń w Polsce, m.in. przez wzbogacenie o dane opracowane na podstawie Badania Aktywności Ekonomicznej Ludności. Wówczas można podjąć próbę scharakteryzowania terytorialnego zróżnicowania wysokości wynagrodzeń ze względu na wiek, poziom wykształcenia, płeć lub sekcję według Polskiej Klasyfikacji Działalności.



## Podsumowanie

Dzięki rozwijającej się metodologii ekonometrii i statystyki przestrzennej coraz więcej uwagi poświęca się uwzględnieniu charakteru przestrzennego analizowanych danych statystycznych.

W niniejszym opracowaniu zaproponowano wykorzystanie elementów procedury planowania eksperymentów w analizie danych przestrzennych. Uzyskane rezultaty stanowią swoiste uzupełnienie eksploracyjnej analizy danych przestrzennych. Ponadto wskazano istotny związek teorii planowania eksperymentów z analizą danych przestrzennych, co wzbogaca metodologię analiz przestrzennych i umożliwia dalsze wykorzystanie metod eksperymentowania w tej dziedzinie.

Przedstawiona analiza danych dotyczących przeciętnego wynagrodzenia wskazuje na słuszność wykorzystania czynnika przestrzennego w charakterystyce wynagrodzeń w Polsce. Uzyskane rezultaty prowadzą do wyodrębnienia obszarów o wysokim i niskim poziomie wynagrodzeń, co stanowi potwierdzenie danych dotyczących poziomu rozwoju gospodarki w Polsce.

## Literatura

- Główny Urząd Statystyczny (2013), Zatrudnienie i wynagrodzenia w gospodarce narodowej w 2012 r., [http://www.stat.gov.pl/gus/5840\\_1890\\_PLK/HTML.htm](http://www.stat.gov.pl/gus/5840_1890_PLK/HTML.htm).
- Kończak G. (2007), Metody statystyczne w sterowaniu jakością produkcji, Wydawnictwo Akademii Ekonomicznej, Katowice.
- Kopcewska K. (2011), Ekonometria i statystyka przestrzenna z wykorzystaniem programu R CRAN, Wydawnictwo CeDeWu, Warszawa.
- Lewandowska-Gwarda K. (2013), Rola przestrzeni w badaniach ekonomicznych, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń, s. 145-158.
- Montgomery D.C. (1997), Introduction to Statistical Quality Control, John Wiley & Sons, New York.
- Müller W.G. (2010), Collecting Spatial Data, Springer Verlag, Heidelberg.
- Paelinck J.H.P, Klaassen L.H.K. (1983), Ekonometria przestrzenna, Państwowe Wydawnictwo Naukowe, Warszawa.
- Sucheckie, B. (2010), Ekonometria przestrzenna, Wydawnictwo C.H. Beck, Warszawa.
- Wawrzynek J. (1993), Statystyczne planowanie eksperymentów w zagadnieniach regresji małej próby, Wydawnictwo Akademii Ekonomicznej, Wrocław.
- Wawrzynek J. (2009), Planowanie eksperymentów zorientowane na doskonalenie jakości produktu, Wydawnictwo Uniwersytetu Ekonomicznego, Wrocław.
- Zeliaś A. (1991), Ekonometria przestrzenna, PWE, Warszawa.

---

## ON THE USE OF DESIGN OF EXPERIMENTS METHODS IN THE ANALYSIS OF SPATIAL DATA

### Summary

Nowadays, in statistical analysis is the demand of methods which allows for understanding ongoing economic and social changes. One way of the corresponding description of analyzed phenomena is to consider their spatial location, what leads to an analysis of spatial data.

Design of experiments is a tool of statistical quality control, which is used in practice of manufacturing companies. Methods of design of experiments are used to improve the results of the production process and have an influence on his economic aspect.

The spatial analysis of data, such as the development of industry or the unemployment rate, leads to identifying areas of high and low level of economic development. The aim of this paper is to consider the problem of spatial differences of wages in Poland. For this purpose, in the analysis of spatial data, will be used the design of experiments methods.