

SŁAWOMIR DOROCCI

Uniwersytet Pedagogiczny w Krakowie, Polska • Pedagogical University of Cracow, Poland

BARTOSZ JENNER

Reckitt Benckiser, Hull, Wielka Brytania • United Kingdom

Recepta na nienormalność rozkładu i współzależność obserwacji z wykorzystaniem testów randomizacyjnych i testu Mantela – na przykładzie rozmieszczenia zasobów ludzkich w regionach Francji

Prescription for Abnormal Distribution and Interdependence of Observations Using Randomization Tests and Mantel Correlation Coefficient – the Example of Human Resources Distribution in the Regions of France

Streszczenie: Wykorzystując metody statystyczne w badaniach geograficznych, badacze często spotykają się z problemem niespełnienia przez posiadane dane założeń wymaganych przez liczne testy statystyczne. Dla przykładu wiele tzw. parametrycznych testów statystycznych opartych jest na założeniu normalności rozkładu danych. W praktyce jednak taka supozycja często nie jest zrealizowana, a rażące odchylenie rozkładu danych od rozkładu normalnego (np. rozkład J-kształtny) może prowadzić do wręcz absurdalnych wartości przedziałów ufności (np. przedział ufności dla średniej). Innym często spotykanym problemem z danymi pochodzącymi z badań geograficznych jest przestrzenna współzależność obserwacji. Wzajemne skorelowanie danych może przykładowo odzwierciedlać odległości między miejscami, gdzie te dane pozyskano. Dlatego wyniki pomiarów pochodzące z miejsc położonych w bliskiej odległości mogą być do siebie bardziej podobne niż obserwacje pochodzące z miejsc od siebie oddalonych. W takiej sytuacji oszacowanie wartości p-value dla testów badających zależności między zmiennymi (np. korelacja Pearsona) mogą być bardzo mylące, gdyż klasyczne metody statystyczne zakładają niezależność zmiennych. W powyższych sytuacjach badacz może wykorzystać tzw. testy randomizacyjne, które pozwalają obejść założenia normalności rozkładu czy niezależności obserwacji. Metody te, wraz z możliwością wykonywania złożonych obliczeń przy wykorzystaniu narzędzi informatycznych, stają się obecnie coraz bardziej popularne wśród badaczy. W artykule wyjaśniona została logika testów randomizacyjnych i dwa przykłady ich zastosowania: szacowanie przedziału dla średniej i obliczanie korelacji Mantela wraz z testowaniem jej istotności. W rezultacie przeprowadzonych obliczeń wykazano, że w obydwu przypadkach pominięcie założeń statystycznych prowadzi do otrzymania błędnych wyników. W tekście w celach zobrazowania powyższych metod statystycznych wykorzystano dane obrazujące wielkość i udział zasobów ludzkich dla nauki i techniki (HRST) w regionach Francji oraz tempo zmian tych wartości w czasie i przestrzeni. Obliczenia oparto na darmowym pakiecie statystycznym R oraz arkuszu kalkulacyjnym Excel.

Abstract: Using statistical methods in geographical research, researchers are often faced with the problem that their data do not comply with the assumptions required by a number of statistical tests. For example, many parametric statistical tests are based on the assumption of normal distribution of data. In practice, however, this supposition is often not met, and the serious deviation of data distribution from the normal distribution (e.g., J-shaped distribution) can lead to quite absurd values of confidence limits (for example, the

confidence interval). Another common problem with data from geographical research is the interdependence of observations. Mutual correlation of data can, for example, reflect the spatial distance between the places where that information was obtained. Therefore, the results of measurements from sites located in close proximity to each other may be more alike than observations from places more distant from each other. In such a situation, estimated p-values for tests investigating the relationships between variables (e.g. Pearson correlation) can be very confusing, because the "classical" statistical methods assume independence of the variables. In these situations, the researcher can use randomisation tests that allow him to "get around" the assumptions of normal distribution and independence of observations. These methods, along with the ability to perform complex calculations using computer tools, are becoming increasingly popular among researchers. The article explains the logic of randomisation tests and two examples of their use: estimating the interval for the mean and calculating Mantel correlation along with testing its significance. Calculations have shown that in both cases, disregard of the statistical assumptions leads to false results. In order to illustrate these statistical methods, data used in this paper show the volume and share of human resources in science and technology (HRST) in regions of France and the pace of changes in these values over time and space. Calculations were conducted using the free statistical software R Project and Excel spreadsheet.

Słowa kluczowe: Francja; kapitał ludzki; metody statystyczne; randomizacja; R-pakiet; test Mantela

Keywords: France; human capital; Mantel test; randomization; R-package; statistical methods

Otrzymano: 13 stycznia 2016

Received: 13 January 2016

Zaakceptowano: 1 czerwca 2016

Accepted: 1 June 2016

Sugerowana cytacja / Suggested citation:

Dorocki, S., Jenner, B. (2016). Recepta na nienormalność rozkładu i współzależność obserwacji z wykorzystaniem testów randomizacyjnych i testu Mantela – na przykładzie rozmieszczenia zasobów ludzkich w regionach Francji. *Prace Komisji Geografii Przemysłu Polskiego Towarzystwa Geograficznego*, 30(2), 186–197.

WSTĘP

Wykorzystanie metod statystycznych w celu opisu geograficznego, jak zauważył pół wieku temu K. Dziewoński (1965), ma długą tradycję. Jednocześnie, co także podkreślił, wzrastające wyrafinowanie i złożoność metod statystycznych stwarza problemy w ich poprawnym wykorzystaniu i odpowiedniej interpretacji otrzymanych wyników. Dodatkowym utrudnieniem jest dostępność danych statystycznych, szczególnie w przypadku zagadnień społeczno-gospodarczych. Dotyczy ono zarówno kwestii porównywalności danych, jak i ich dostępności związanej z kwestiami finansowymi (Czapliński, 2008). Wielu geografów podczas swojej pracy spotyka się z problemem wykorzystania metod wnioskowania statystycznego w przypadku, gdy rozkład danych znacząco odbiega od rozkładu normalnego lub zebrane dane o współzależności zjawisk są ze sobą powiązane i nie spełniają wymogu niezależności (Wibig, 2013).

Poniższy artykuł ma na celu zilustrowanie metod wnioskowania statystycznego opartego na symulacjach komputerowych. Podejście to pozwala obejść założenia leżące u podstaw klasycznych metod statystyki matematycznej. Wykorzystanie tych narzędzi przez geografów wydaje się istotne głównie ze względu na częste przypadki analizy danych, które nie spełniają niektórych założeń stosowanych w statystyce tradycyjnej.

Jako przykłady przedstawione zostaną dwie sytuacje. Pierwsza, gdy chcemy oszacować przedział ufności dla średniej arytmetycznej, ale rozkład danych jest silnie

skośny i znacząco odbiega od rozkładu normalnego – czyli nie są spełnione założenia klasycznej metody estymacji dokładności oszacowania średniej.

W drugim przypadku chcemy zbadać związek między odległością pomiędzy danymi punktami a różnicami w natężeniu wybranego zjawiska. W przedstawionej sytuacji dane nie spełniają założenia niezależności wymaganego w klasycznej metodzie badania korelacji.

W poniższej egemplifikacji wykorzystano dane o wielkości i udziale zasobów ludzkich dla nauki i techniki (HRST – Human Resources for Science and Technology) w regionach Francji metropolitalnej (z pominięciem Korsyki) w latach 1995–2012, które zaczerpnięto z internetowej bazy danych Eurostatu. Wybór zagadnienia i tła zobrazowania wybranych metod statystycznych związany jest z zainteresowaniami autorów i prowadzonymi badaniami. Należy podkreślić, że przedstawione w artykule przykłady stanowią tylko egemplifikację omawianych metod statystycznych, natomiast nie mają na celu pokazania analiz przestrzennych. Autorzy zdają sobie sprawę z istnienia innych metod statystycznych, zaliczanych do tzw. metod klasycznych, które można zastosować do badania poruszanych w przykładach analiz przestrzennych, np. współczynnik autokorelacji przestrzennej Morana lub regresji geograficznie ważonej (Janc, 2006; Olejnik, 2013). Jednakże te klasyczne metody statystyczne nie nawiązują bezpośrednio do przedstawionych w artykule metod opartych na symulacjach komputerowych.

NIENORMALNOŚĆ ROZKŁADU A PRZEDZIAŁ UFNOŚCI DLA ŚREDNIEJ ARYTMETYCZNEJ

Pierwszym zagadnieniem poruszonym w artykule jest nienormalność rozkładu i kłopoty z określeniem przedziału ufności dla średniej arytmetycznej. Powyższy problem zostanie zobrazowany na podstawie liczby zasobów ludzkich dla nauki i techniki (HRST) w regionach Francji w 2000 roku. Jak widać z poniższego histogramu (ryc. 1), rozkład zasobów ludzkich dla nauki i techniki jest w regionach Francji nierównomierny. Ponad połowa z 21 analizowanych regionów Francji metropolitalnej znajduje się w najniższej klasie, ze średnio 80 tys. osób stanowiących zasoby dla gospodarki zaawansowanych technologii. Rozkład wartości jest zatem bardzo asymetryczny, co implikuje brak normalności rozkładu analizowanych danych. Asymetria ta wynika w głównej mierze z historycznej dominacji regionu stołecznego Île-de-France z metropolią paryską (Dorocki, Borowiec, 2011). Ogólnie dla Francji średnia wartość liczebności zasobów HRST dla regionów Francji wynosi 190,2 tys. osób, przy wartości odchylenia standardowego $s = 242,3$ tys. osób. Zachodzi zatem pytanie, jaki jest 95% przedział ufności dla otrzymanej średniej zasobów ludzkich dla nauki i techniki w regionach Francji metropolitalnej w 2000 roku.

W związku z zaistniałą sytuacją silnej asymetrii analizowanego zjawiska aproksymacja takiego J-kształtnego rozkładu danych za pomocą rozkładu normalnego może prowadzić do absurdalnych wniosków. Gdyby przyjąć w powyższym przypadku, że rozkład jest normalny, to analiza średniej i odchylenia standardowego sugeruje, że co piąty region (ok. 21%) ma ujemną liczbę osób stanowiących zasoby ludzkie w sektorach wysokich technologii. Wobec tego wiarygodność oszacowania przedziału ufności

dla średniej na podstawie założenia normalności rozkładu może być bardzo nierzetelna lub wręcz błędna.

Jak zatem policzyć miarodajny przedział ufności dla średniej arytmetycznej, gdy dane są rozkładem silnie skośnym? Jedną z metod jest tzw. bootstrap (bootstrapping), który pozwala na oszacowanie rozkładu próby prawie każdej dowolnej statystyki za pomocą bardzo prostych metod (Davison, Hinkley, 2008). Metoda bootstrap jest stosowana w celu określenia średniej μ z populacji o funkcji F , na podstawie danych pozyskanych przez losowe pobranie próbek z tej populacji:

$$\mu = \int x dF(x) \quad [1]$$

średnia z próbki ma analogiczną funkcję empiryczną rozkładu, czyli

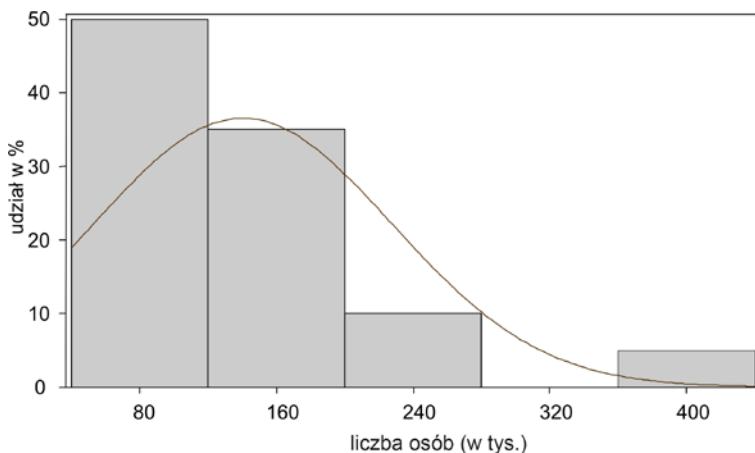
$$\hat{F}(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x) \quad [2]$$

gdzie $X_1; \dots; X_n$ oznaczają dane. Zatem prognozy średniej bootstrapowanej populacji μ równe są średniej próbki

$$\bar{X} = \int x dF(x) = \frac{1}{n} \sum_{i=1}^n X_i \quad [3]$$

Również prognozy z estymacji bootstrapowej z wariancji populacji są równe odchyleniom wartości próby.

Ryc. 1. Udział regionów według zasobów ludzkich dla nauki i techniki w regionach Francji metropolitalnej w 2000 roku



Źródło: opracowanie własne na podstawie danych Eurostatu

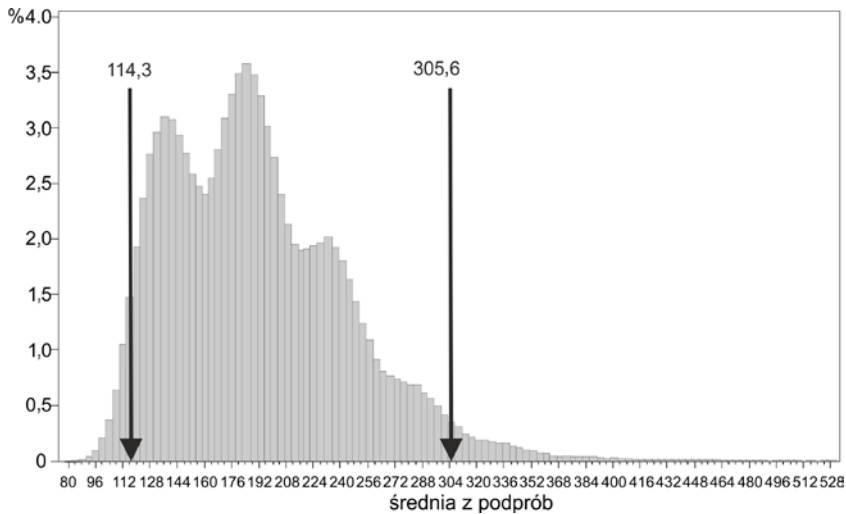
Procedura postępowania mająca na celu zastosowanie metody bootstrapowej w dużej mierze bazuje na wykorzystaniu obliczeniowych technik komputerowych. Pierwszym krokiem jest wylosowanie ze zwracaniem n -elementów z zebranej n -elementowej próby. W trakcie losowania niektóre wartości w związku ze zwracaniem

zostaną powtórzone w powstałej próbie losowej. Następnie obliczamy wartość średnią dla wylosowanych danych. Przedstawioną procedurę wykonujemy wielokrotnie. Im większa liczba losowań, tym otrzymany wynik będzie bardziej wiarygodny. W przedstawionym przypadku dokonano 500 tys. losowań podprób i otrzymano tyleż samo średnich. Do losowania można wykorzystać funkcję losowania w darmowym programie R. Losowanie następuje ze stworzonego zbioru zawierającego interesujące nas dane liczebności HRST w regionach, nazwanego *regions*, poprzez wprowadzenie polecenia:

```
replicate(500000,mean(sample(regions,21,replace=TRUE))).
```

Rozkład wartości średnich z uzyskanych losowych podprób stanowi przybliżenie rozkładu statystyki z próby. Dla otrzymanego rozkładu obliczamy odpowiednie percentyle. W naszym przypadku w celu oszacowania górnego i dolnego krańca 95% przedziału ufności obliczono percentyle 2,5% i 97,5%. Obliczenia można dokonać, zliczając podpróby z szeregu prostego lub wykorzystując funkcję z pakietu R: `quantile(x,.025)` i `quantile(x,.975)`.

Ryc. 2. Rozkład wartości średnich (w tys.) z bootstrapowanej próby i obliczanie 95% przedziału ufności dla średniej arytmetycznej



Źródło: opracowanie własne na podstawie danych Eurostatu

Obliczając przedział ufności dla średniej arytmetycznej wielkości zasobów ludzkich dla nauki i techniki w regionach Francji w 2000 roku, posłużono się zarówno metodą klasyczną, jak i bootstrapową w celu porównania otrzymanych wyników. Do obliczeń przedziału ufności dla średniej posłużono się wzorem:

$$CI = \bar{x} \pm \frac{t \cdot s}{\sqrt{n}} \quad [4]$$

gdzie t to 2,5% i 97,5% percentyl.

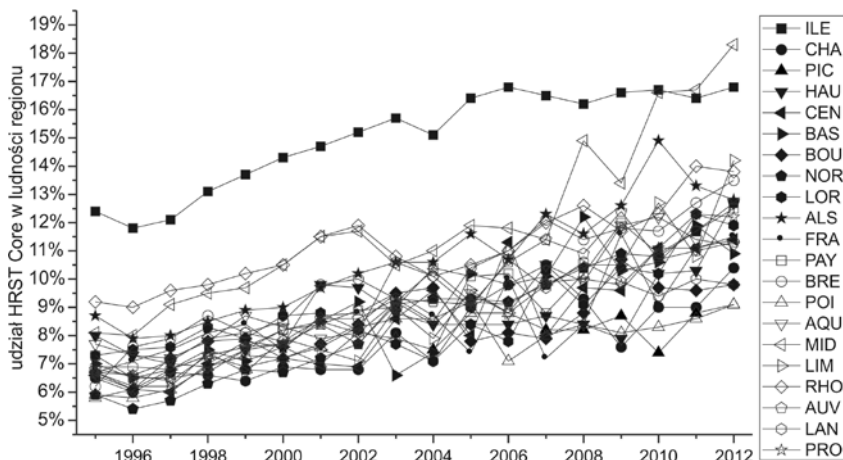
Przyjmując zatem normalność rozkładu analizowanych danych, otrzymujemy przedziały ufności dla średniej arytmetycznej liczebności zasobów ludzkich w regionach Francji na poziomie 95%CI = [79,7; 299,9], podczas gdy dokonując obliczeń na podstawie danych otrzymanych metodą bootstrappingu: 95%CI = [114,3; 305,6]. Tak więc dla analizowanych danych metoda klasyczna nie doszacowuje wartości dolnego krańca przedziału ufności w ponad 40%.

Podany przykład wydaje się w przypadku badań geograficznych ważny, ponieważ w wielu opracowaniach odnaleziono zastosowanie metody określenia wartości granicznych przedziału ufności z pominięciem określenia rozkładu analizowanych danych.

BADANIE WSPÓLZALEŻNOŚCI MIĘDZY OBSERWACJAMI ZALEŻNYMI

W drugiej części artykułu podjęto zagadnienie analizy współzależności (korelacji) między obserwacjami zależnymi. Autorzy podjęli próbę znalezienia odpowiedzi na pytanie: „Czy na podobieństwo intensywności zjawiska, jakim jest wzrost udziału personelu badawczego (HRST Core) w populacji regionów Francji w latach 1995–2012, ma wpływ położenie geograficzne w aspekcie fizycznej odległości między tymi regionami?”. Czy sąsiadujące ze sobą regiony mają podobne tempo wzrostu kadry naukowej?

Ryc. 3. Zmiana udziału zasobów ludzkich dla nauki i techniki (HRST Core) w populacji regionów Francji w latach 1995–2012

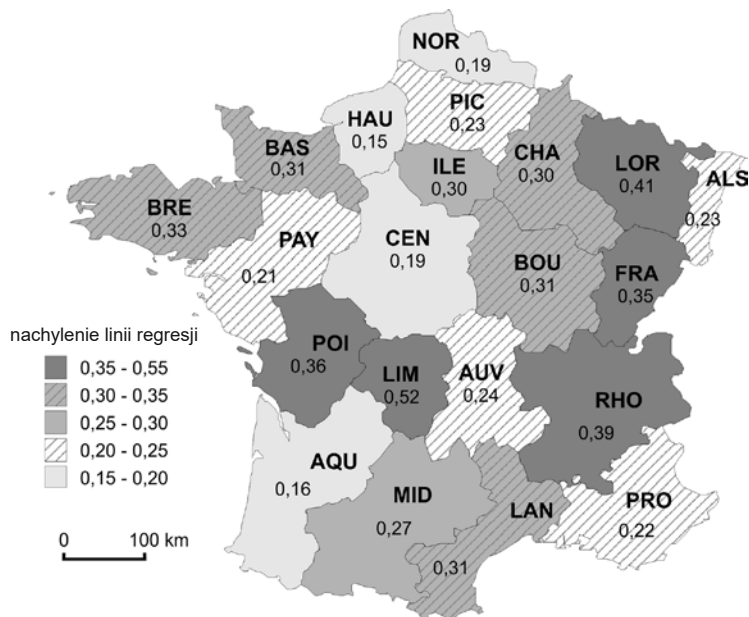


Źródło: opracowanie własne na podstawie danych Eurostatu

Analiza graficznego obrazu zmiany udziału zasobów ludzkich dla nauki i techniki (HRST Core) w populacji regionów Francji (ryc. 3) sugeruje, że zarówno udział, jak i tempo wzrostu kadry naukowej w regionach Francji w latach 1995–2012 są różne. Największy udział zasobów ludzkich dla nauki i techniki w analizowanym okresie posiadał region Île-de-France (ILE), jednakże dzięki realizowanej od 2006 roku regionalnej strategii wspierania innowacji (Midi-Pyrénées Innovation – MPI) największe wartości w 2012 roku odnotowano w regionie Midi-Pyrénées (MID). Również w pozostałych

regionach udział kadry naukowej w populacji regionów był zróżnicowany, oscylując w przedziale od 9% do 14%, i wykazywał odmienne tempo wzrostu. Analizując regionalne zróżnicowanie tempa zmian wyrażonego w nachyleniu krzywej regresji (ryc. 4), największy wzrost odnotowały regiony wschodnie Francji oraz zachodnie (Limousin i Poitou-Charentes). Dla regionów Francji współczynnik zmienności tempa wzrostu udziału ludności HRST Core wynosi $CV = 31\%$ ($s = 0,09$; $= 0,29$).

Ryc. 4. Tempo wzrostu udziału zasobów ludzkich dla nauki i techniki (HRST Core) w ogólnej populacji regionów Francji w latach 1995–2012



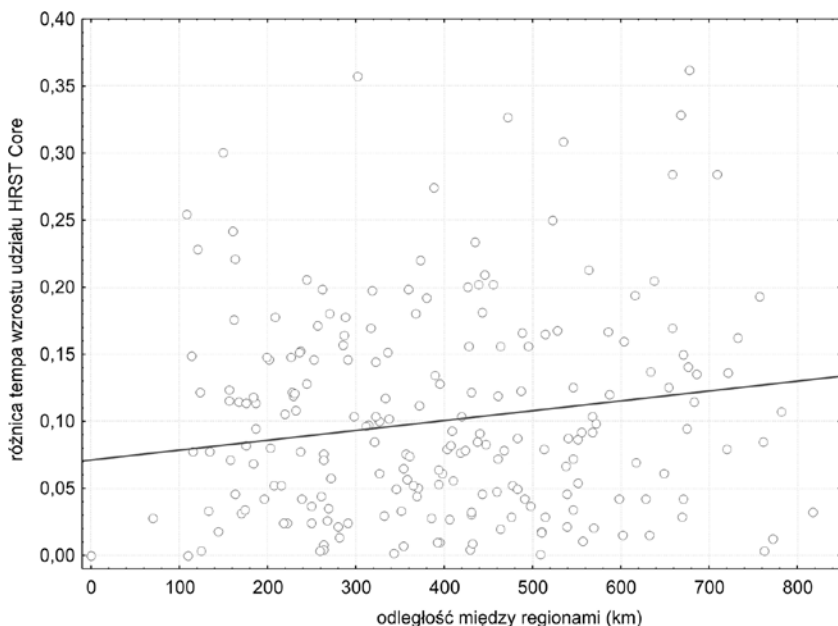
Źródło: opracowanie własne na podstawie danych Eurostatu

Podczas analizowania rozkładu tempa zmiany udziału kadry naukowej w regionach Francji rodzi się pytanie: „Czy regiony położone blisko siebie cechują się podobnym tempem wzrostu udziału zasobów dla HRST Core?”. Podobnie jak w poprzednim przykładzie, również do rozpatrywania istotności korelacji w celu dokonania porównania otrzymanych wyników wykorzystano zarówno metodę klasyczną, jak i symulacji komputerowej.

Stosując metodę tradycyjną, dokonujemy badania istotności statystycznej za pomocą prostej korelacji między zmienną niezależną, którą w naszym przypadku są odległości pomiędzy środkami geometrycznymi regionów pozyskane z publikacji INSEE, a zmienną zależną: różnice w tempie wzrostu udziału kadry naukowej w populacji regionów. Trudność tej tradycyjnej metody polega na braku wiarygodności obliczonej wartości prawdopodobieństwa testowego p-value, ponieważ test korelacji Pearsona zakłada niezależność obserwacji, a także normalność rozkładu zmiennych. Natomiast część powiązań dotyczy tych samych regionów, zatem dane te są zależne. Pomijając

jednak ten fakt i przyjmując niezależność obserwacji, korelacja, choć bardzo słaba, 0,18, jest statystycznie istotna na poziomie $\alpha = 0,05$. Nie ma zatem podstaw do odrzucenia tezy o istnieniu zależności pomiędzy odległością fizyczną między regionami Francji a różnicami w tempie wzrostu udziału zasobów ludzkich dla nauki i techniki w populacji regionów w badanym okresie.

Ryc. 5. Zależność liniowa między odległościami pomiędzy regionami a różnice w tempie wzrostu udziału HRST Core w populacji regionów Francji w latach 1995–2012



Źródło: opracowanie własne na podstawie danych Eurostatu

W przypadku korelacji Mantela algorytm postępowania i logika testu wyglądają następująco. W komórkach kwadratowej macierzy wyliczamy odległości między geometrycznymi centrami regionów. Otrzymujemy zatem wartość odległości pomiędzy centrum każdego regionu od centrów wszystkich pozostałych regionów. Następnie budujemy analogiczną macierz różnic między tempami wzrostu udziału zasobów ludzkich dla nauki i techniki w populacji regionów Francji (tab. 1).

Następnym krokiem jest obliczenie sumy iloczynów odpowiadających sobie komórek: macierzy odległości i macierzy różnic tempa wzrostu:

$$Z = \sum_{i=1}^n \sum_{j=1}^n A_{ij} B_{ij} \tag{5}$$

W naszym przypadku obliczenia wyglądałyby następująco:

$$Z = 157 * 0,07 + 114 * 0,15 + 156 * 0,12 + 124 * 0,00 + 263 * 0,01 + \dots + 156 * 0,12 + 264 * 0,00 + \dots$$

W praktyce można obliczyć standaryzowaną sumę iloczynów r , tak aby wartość Z zawierała się między -1 a $+1$. Znormalizowania dokonujemy za pomocą standaryzowanej transformacji normalnej, gdzie średnia macierzy jest odejmowana od każdego elementu, a następnie otrzymane różnice są dzielone przez odchylenie standardowe macierzy.

$$r = \frac{\sum \sum std A_{ij} B_{ij}}{n-1} \quad [6]$$

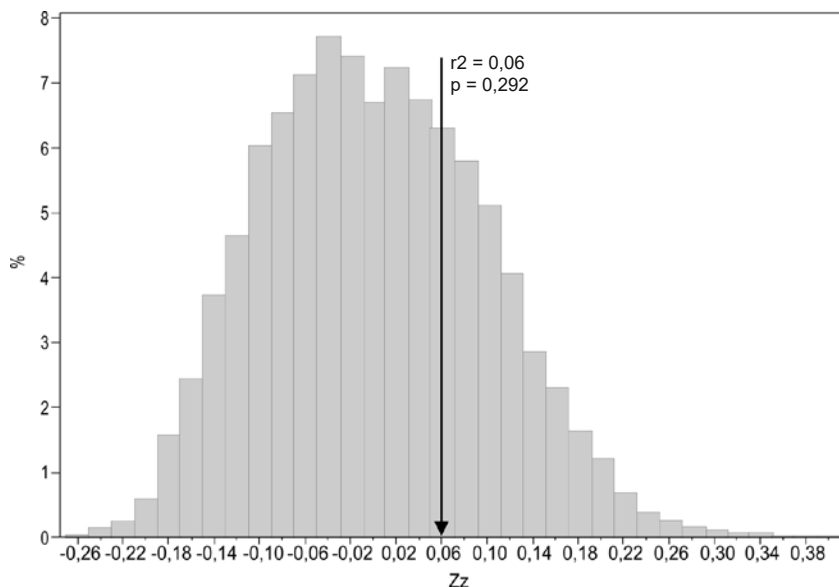
Im bardziej odpowiadające sobie w macierzach wartości komórki są ze sobą korelujące, tym otrzymana suma iloczynów jest większa.

Tab. 1. Procedura testowania istotności korelacji uwzględniającej współzależność obserwacji

Macierz odległości (lewa)								Macierz różnic tempa wzrostu (prawa)							
Region	ILE	CHA	PIC	HAU	CEN	BAS	...	Region	ILE	CHA	PIC	HAU	CEN	BAS	...
ILE	0	157	114	156	124	263		ILE	0.00	0.07	0.15	0.12	0.00	0.01	
CHA	157	0	133	261	264	402		CHA	0.07	0.00	0.08	0.04	0.08	0.08	
PIC	114	133	0	133	236	285		PIC	0.15	0.08	0.00	0.03	0.15	0.16	
HAU	156	261	133	0	229	156		HAU	0.12	0.04	0.03	0.00	0.12	0.12	
CEN	124	264	236	229	0	264		CEN	0.00	0.08	0.15	0.12	0.00	0.00	
BAS	263	402	285	156	264	0		BAS	0.01	0.08	0.16	0.12	0.00	0.00	
...								...							

Źródło: opracowanie własne na podstawie danych Eurostatu

Ryc. 6. Standaryzowany rozkład przepermutowanych wartości Z_z odnoszący się do odległości geograficznej regionów Francji i tempa zmian udziału zasobów HRST Core w regionach



Źródło: opracowanie własne na podstawie danych Eurostatu

W dalszej kolejności komórki jednej z macierzy przedstawiamy w sposób losowy (ale symetryczny), a następnie obliczamy wartość Z dla nowego rozkładu. Krok ten wykonujemy wiele razy, minimum 1000, czyli otrzymamy 1000 nowych Z_z . W ten sposób uzyskujemy dane do oszacowania rozkładu Z , jaki wynikałby z losowego powiązania danych lub braku powiązania. Jeśli oryginalna (przed permutacją) wartość Z była wyjątkowo duża lub nadzwyczaj mała, czyli sytuowała się w skrzydle rozkładu, to sugeruje to, że pierwotne ułożenie komórek miało pewną strukturę. Rozkład danych nie był losowy.

Skorygowaną wartość p -value możemy wyliczyć jako proporcję przepermutowanych wartości Z_z różnej od wartości Z sprzed permutacji. W teście Mantela hipoteza zerowa zakłada brak zależności pomiędzy analizowanymi macierzami odległości i danej zmiennej. Jeżeli wartość otrzymanej statystyki jest wyższa niż $\alpha = 0,05$, nie ma podstaw do odrzucenia hipotezy zerowej.

W naszej sytuacji w 29,2% przypadkach wartość przepermutowanych wartości Z_z była większa od wartości Z sprzed permutacji.

$$Z_z\% > Z\% = 29,2\%; p\text{-value} = 0,292$$

Stąd p -value jest równe 0,292 (istotność 0,05). Tym samym korelacja nie jest istotna. Na poziomie istotności 0,05 nie mamy zatem podstaw do odrzucenia hipotezy zerowej o braku zależności pomiędzy odległością fizyczną między regionami Francji a różnicami w tempie wzrostu udziału HRST Core w ich populacji. Bliskość geograficzna regionów nie wpływa na podobieństwo wzrostu udziału zasobów ludzkich dla nauki i techniki core w społeczeństwie regionów Francji.

Test Mantela można wykonać za pomocą pakietu R i funkcji *mantel.rtest* (Bruin, 2006). Poniżej przedstawiono kod testu Mantela, w którym dokonano 10 tys. interakcji.

```
> hrst<-read.table(" hrst.csv", sep=";", header=T)
> region.dists <- dist(cbind(hrst$lon, hrst$lat))
> hrst.dists <- dist(hrst$hrst)
> as.matrix(region.dists)[1:21, 1:21]
> as.matrix(hrst.dists)[1:21, 1:21]
> mantel.rtest(region.dists, hrst.dists, nrepet = 10000)
```

Monte-Carlo test

Observation: 0.051952364

Call: mantel.rtest(m1 = region.dists, m2 = hrst.dists, nrepet = 1e+05)

Based on 10000 replicates

Simulated p-value: 0.2704761

Jak można zauważyć, zarówno otrzymany wynik korelacji 0,051, jak i poziomu ufności 0,27, w tym przypadku jest inny niż otrzymany w wyniku poprzednich obliczeń, co związane jest z losową permutacją, dlatego wyniki za każdym razem będą różne.

PODSUMOWANIE

Wyniki analiz statystycznych są wiarygodne tylko wówczas, gdy spełnione zostaną założenia stosowanych metod statystycznych. Jeśli założenia te nie są uwzględniane przez badaczy, to nie można oczekiwać wiarygodnych wyników analizy statystycznej (*rubbish in, rubbish out*). Jak wykazano na przedstawionych przykładach, dobranie złych metod do posiadanych danych prowadzi do błędnych wyników. Przedstawione w tekście intensywne metody komputerowe, chociaż są mniej eleganckie z powodu braku precyzyjnej powtarzalności wyników, pozwalają jednak na uzyskanie rzetelnych ocen parametrów i wiarygodnych testów statystycznych. Podkreślić również należy, że wykorzystane w tekście przykłady miały na celu głównie zobrazowanie wykorzystania metod statystycznych, a nie analizę problemu rozmieszczenia zasobów ludzkich we Francji.

Literatura

References

- Bruin, J. (2006) (2015, 11 stycznia). Newtest: command to compute new test, UCLA: Statistical Consulting Group. Pozyskano z: <http://www.ats.ucla.edu/stat/stata/ado/analysis/>
- Czapliński, P. (2008). Problematyka badawcza przemysłu w geografii na tle nauk ekonomicznych, *Prace Komisji Geografii Przemysłu Polskiego Towarzystwa Geograficznego*, 11, 46–52.
- Davison, A.C., Hinkley, D.V. (2008). *Bootstrap Methods and their Application. Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge: University Press.
- Dorocki, S., Borowiec, M. (2011). Rola kapitału ludzkiego w procesie kształtowania innowacyjności układów regionalnych Francji. W: B. Namyślak (red.). *Przekształcenia regionalnych struktur funkcjonalno-prze-strzennych*. Wrocław: Instytut Geografii i Rozwoju Regionalnego. Uniwersytet Wrocławski, 215–230.
- Dziwowski, K. (1965). Zagadnienia integracji analizy kartograficznej i statystycznej w badaniach geograficznych. *Przegląd Geograficzny*, 37(4), 585–597.
- Fortin, M., Gurevitch, J. (2001). Mantel Tests: Spatial Structure in Field Experiments. W: S.M. Scheiner, J. Gurevitch (red.). *Design and Analysis of Ecological Experiments*. Nowy Jork: Chapman & Hall, 308–315.
- Janc, K. (2006). Zjawisko autokorelacji przestrzennej na przykładzie statystyki I Morana oraz lokalnych wskaźników zależności przestrzennej (LISA) – wybrane zagadnienia metodyczne. W: T. Komornicki, Z. Podgórski (red.). *Idee i praktyczny uniwersalizm geografii, Dokumentacja Geograficzna*, 33, 76–83.
- Olejnik, A. (2013). Wybrane metody testowania modeli regresji przestrzennej. *Przegląd Statystyczny*, 60(3), 381–393.
- Wibig, J. (2013). O współczynniku korelacji liniowej raz jeszcze. W: K. Jarzyna (red.). *Zastosowanie metod statystycznych w geografii*. Kielce: Instytut Geografii Uniwersytetu Jana Kochanowskiego w Kielcach, 47–61.

Sławomir Dorocki, doktor nauk humanistycznych w dyscyplinie historia (Instytut Europeistyki, Uniwersytet Jagielloński), absolwent studiów z zakresu geografii społeczno-ekonomicznej Uniwersytetu Pedagogicznego w Krakowie. Adiunkt w Instytucie Geografii Uniwersytetu Pedagogicznego w Krakowie. Zainteresowania badawcze skupiają się wokół problematyki regionów i procesów regionalizacji społeczno-gospodarczej, ze szczególnym uwzględnieniem zróżnicowania przestrzeni europejskiej oraz procesów integracji europejskiej i uwarunkowań historycznych, a także zastosowania metod komputerowych i statystycznych w badaniach nad zróżnicowaniem przestrzeni.

Sławomir Dorocki, Ph.D., Department of Entrepreneurship and Spatial Management, Institute of Geography, Pedagogical University of Cracow. Sławomir Dorocki, graduated from the Pedagogical University of Cracow, MA degree in geography, Ph.D. in history (Institute of European Studies of the Jagiellonian University). Associate Professor at the Pedagogical University of Cracow, Institute of Geography. His research interests are tied with regional problems and processes of socio-economic regionalization, with particular emphasis on

the diversity of Europe, processes of European integration, historical conditions and application computer and statistical methods in the study of diversity of space.

Adres/address:

Uniwersytet Pedagogiczny w Krakowie
Instytut Geografii
Zakład Przedsiębiorczości i Gospodarki Przestrzennej
ul. Podchorążych 2, 30-084 Kraków, Polska
e-mail: sdorocki@up.krakow.pl

Bartosz Jenner, absolwent Wydziału Biologii i Nauk o Ziemi Uniwersytetu Jagiellońskiego w Krakowie (2001), okresowo studiował też biologię na uniwersytecie w Wolverhampton w Wlk. Brytanii. Aktualnie pracuje jako statystyk w dziale Global Medical Affairs firmy Reckitt Benckiser w Wlk. Brytanii. Poprzednio pracował jako epidemiolog w Klinice Toksykologii na Wydziale Lekarskim UJ oraz przebył roczny staż post-doc w Max Planck Institute for Demographic Research w Niemczech. Posiada doświadczenie w prowadzeniu rozmaitych badań naukowych, badań klinicznych nad lekami prowadzonych według przepisów GCP oraz badań skuteczności i bezpieczeństwa kosmetyków. Jest współautorem około 20 publikacji. Członek Polskiej Grupy Narodowej Międzynarodowego Towarzystwa Biostatystyki Klinicznej (ISCB) oraz Statisticians in the Pharmaceutical Industry (PSI).

Bartosz Jenner, graduated from the Department of Biology and Earth Sciences of the Jagiellonian University in Krakow (2001) – he also temporarily studied biology at the University of Wolverhampton – UK. He currently works as a statistician at the Global Medical Affairs of Reckitt Benckiser (UK). In the past he worked as epidemiologist at the Department of Toxicology of the Jagiellonian University Medical College and completed 1 year post-doc fellowship in Max Planck Institute for Demographic Research in Germany. He has acquired experience in various research studies, clinical trials and GCP procedures. He is co-author of 20 papers. He is a member of International Society of Clinical Biostatistics (ISCB) and Statisticians in the Pharmaceutical Industry (PSI).

Adres/address:

Reckitt Benckiser
5a Hawthorne Av.
HU10 6JG Willerby, United Kingdom
e-mail: Bartosz.Jenner@rb.com