DIDACTICS
OF
MATHEMATICS

8(12)

The Publishing House
of Wrocław University of Economics
Wrocław 2011

Referee
Henryk Zawadzki
(University of Economics in Katowice)

Copy-editing
Dorota Pitulec

Proof-reading
Barbara Łopusiewicz

Typesetting
Elżbieta Szlachcic

Cover design
Robert Mazurczyk

Front cover painting: W. Tank, Sower
(private collection)

This publication is available at: www. journal.ue.wroc.pl and www.ibuk.pl.
Abstracts of published papers are available in the international database
The Central European Journal of Social Sciences and Humanities
http://cejsh.icm.edu.pl

Information on submitting and reviewing paper is available
on the Publishing House's website www.wydawnictwo.ue.wroc.pl

ISSN 1733-7941

The original version: printed
Printing: Printing House TOTEM
Print run: 200 copies

TABLE OF CONTENTS

D I D A C T I C S   O F   M A T H E M A T I C S

No. 8 (12) 2011

WIELKOŚĆ PRÓBY A ISTOTNOŚĆ WNIOSKOWANIA STATYSTYCZNEGO

Marek Kośny, Piotr Peternek

Abstract. Commonly used method of determining the quality of research results is the application of significance tests – both parametric and non-parametric. They enable researchers to eliminate low-quality results, when observed differences are the consequence of too small samples. Therefore relatively little attention – also in didactics – is paid to issues of big samples, when the number of observations amounts up to several thousands.

Whereas situations, when significance test is applied to big samples requires no less attention than the situations of too small samples. In this context the main aim of the paper is the analysis of results of the significance test for the correlation coefficient (parametric test) and chi-square test for independence (non-parametric test) for samples exceeding 1000 observations. For these tests it will be shown, how imprecise interpretation could lead to the significant distortion of obtained results.

Keywords: statistical inference, tests of significance, the correlation coefficient.

1. Wstęp

Jednym z podstawowych problemów napotykanych w badaniach empirycznych jest kwestia jakości uzyskanych wyników. W wypadku większości badań nie ma możliwości przebadania całej populacji, a badaniu poddawana jest jedynie pewna jej część – próba. W takiej sytuacji zagadnieniem o zasadniczym znaczeniu jest pytanie o to, na ile wyniki uzyskane dla analizowanej próby mogą być rozszerzone na całą populację. Inaczej mówiąc, jakie warunki muszą być spełnione, aby uogólnienie takie było uprawnione.

Wśród standardowo wymienianych warunków znajduje się losowość próby, która zakłada jej odpowiednią wielkość oraz odpowiednio skonstruowany operat losowania (szerzej na ten temat por. np. (Steczkowski, 1995)). W dalszej części artykułu skoncentrujemy się na pierwszym z wymienionych elementów – wielkości badanej próby. O ile bowiem problem reprezentatywności jest często trudny do jednoznacznego rozstrzygnięcia (oraz

Marek Kośny, Piotr Peternek
Department of Operations Research, Wrocław University of Economics, Komandorska Street 118/120, 53-345 Wrocław, Poland.
E-mail: marek.kosny@ue.wroc.pl, piotr.peternek@ue.wroc.pl

kontrolowania), o tyle wielkość próby jest elementem badania, na który badacz ma zazwyczaj bezpośredni wpływ. W praktyce, ze względu na koszty badania, najczęściej rozważanym w tym kontekście zagadnieniem jest problem minimalnej wielkości próby, niezbędnej do zagwarantowania możliwości uogólniania wyników z próby na całą populację. Jeśli bowiem założymy, że błąd jest wprawdzie nieunikniony, ale istnieje możliwość kontrolowania jego wielkości i prawdopodobieństwa jego wystąpienia, podanie uzyskanego wyniku oraz informacji o możliwych błędach w znacznym stopniu rozwiązuje problem wiarygodności uzyskanych wyników.

W rzeczywistości występuje jednak często znaczna rozbieżność między formalną interpretacją wyników wnioskowania statystycznego oraz ich percepcją przez odbiorców badania (a czasem także samych badaczy). Problem ten, zazwyczaj kojarzony jedynie z badaniami prowadzonymi na zbyt małych próbach[1], wydaje się jednak dotyczyć także prób dużych. Za taki stan rzeczy w znacznej mierze odpowiada – zdaniem autorów – sposób nauczania statystyki matematycznej. O ile bowiem problem małych prób jest dosyć szeroko rozważany w dydaktyce, o tyle niewiele uwagi poświęcane jest badaniom z wykorzystaniem dużych prób. W praktyce pozostawia to często wrażenie, że jedynym celem badania jest uzyskanie wyniku istotnego statystycznie. Pułapkom takiego rozumowania – gdy statystyczna istotność wyniku nie przesądza wcale o jego wartości – poświęcona jest dalsza część niniejszego artykułu. W kolejnych dwóch punktach przedstawione zostaną rozważania dotyczące, odpowiednio, testu na istotność współczynnika korelacji oraz testu niezależności χ^2, jako przykładów wnioskowania statystycznego.

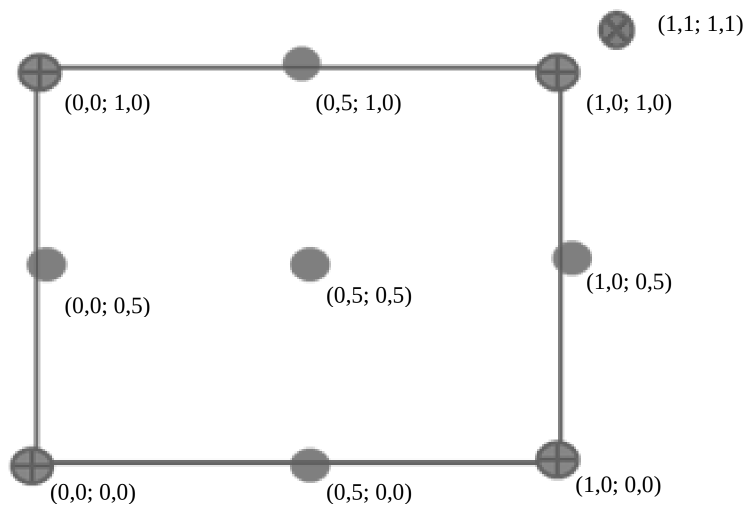
2. Istotność współczynnika korelacji

Pierwszy z zaprezentowanych przykładów będzie dotyczył testu na istotność współczynnika korelacji Pearsona. Współczynnik ten jest chyba najczęściej stosowaną w badaniach empirycznych miarą zależności między zmiennymi – w badaniach społecznych, ekonomicznych, ale także np. w medycynie.

Na rys. 1 zaprezentowany został przykładowy układ punktów tworzących próbę. Załóżmy, że zasadniczą część próby – zgodną z charakterem analizowanego zjawiska w populacji, który jest znany – stanowią punkty

[1] Dochodzi wtedy do nieuprawnionego wnioskowania na podstawie prób niegwarantujących faktycznie możliwości uogólniania wyników.

znajdujące się na krawędziach kwadratu oraz punkt w jego wnętrzu. W tym sensie punkt o współrzędnych (1,1; 1,1) będzie dalej traktowany jako obserwacja odstająca.



Rys. 1. Położenie analizowanych punktów

Źródło: opracowanie własne.

W próbie przedstawionej na rys. 1 (współrzędne oznaczają wartości obserwowanych zmiennych – odpowiednio X i Y) wartość współczynnika korelacji pomiędzy zmiennymi X i Y wyniosła 0,18.

Następnie, w analogiczny sposób, skonstruowano kolejne próby – w każdej z nich zwielokrotnione zostały obserwacje leżące na krawędziach kwadratu oraz w jego wnętrzu. Wartości współczynnika korelacji przy n-krotnym powtórzeniu w próbie wszystkich punktów z wyjątkiem obserwacji odstającej – punktu (1,1; 1,1) – przedstawione zostały w tab. 1.

Tabela 1. Zależność wartości współczynnika korelacji od liczby powtórzeń

Liczba powtórzeń	Wartość współczynnika korelacji
2	0,10
5	0,04
10	0,02
20	0,01

Źródło: opracowanie własne.

Jak wynika z danych przedstawionych w tab. 2, zawierającej typową interpretację wartości współczynnika korelacji, we wszystkich analizowanych wypadkach należy uznać, że nie występuje liniowa zależność pomiędzy cechami *X* i *Y*.

Tabela 2. Interpretacja wartości współczynnika korelacji

Wartość współczynnika korelacji	Interpretacja
Poniżej 0,2	Praktycznie brak związku liniowego między badanymi cechami
0,2 do 0,4	Zależność liniowa wyraźna, lecz niska
0,4 do 0,7	Zależność umiarkowana
0,7 do 0,9	Zależność znacząca
Powyżej 0,9	Zależność bardzo silna

Źródło: opracowanie własne na podstawie (Ostasiewicz, Rusnak, Siedlecka 2001, s. 311).

Niezależnie od interpretacji sformułowanej na podstawie przedziałów opisanych w tab. 2, statystyka matematyczna umożliwia formalną weryfikację, czy uzyskana wartość współczynnika korelacji istotnie różni się od zera. W tym celu formułowane są następujące hipotezy:

$$H_0\colon \rho = 0 \qquad\qquad H_1\colon \rho \neq 0,$$

gdzie ρ oznacza prawdziwą wartość współczynnika korelacji w populacji. Weryfikacja prawdziwości hipotezy H_0 dokonywana jest na podstawie statystyki:

$$T_{n-2} = \frac{R}{\sqrt{1-R^2}}\sqrt{n-2}\,, \tag{1}$$

która ma rozkład Studenta[2] o n–2 stopniach swobody (por. np. (Ostasiewicz, Rusnak, Siedlecka 2001, s. 282)). Jak wynika ze wzoru (1), wartość statystyki testowej zależy wyłącznie od wartości współczynnika korelacji w analizowanej próbie oraz od liczebności próby. Zamiast więc weryfikacji, czy uzyskany wynik jest istotny statystycznie, można sformułować pytanie, jak duża powinna być próba, by dla określonej wartości współczynnika korelacji oraz na zadanym poziomie istotności należało odrzucić hipotezę zerową na rzecz alternatywnej. Odpowiednie wartości określające minimal-

[2] Przy założeniu, że prawdziwa jest H_0 oraz że próba pochodzi ze zbiorowości o dwuwymiarowym rozkładzie normalnym.

ną wielkość próby można uzyskać metodami numerycznymi. W takiej sytuacji próba o zadanej liczebności (nie mniejszej niż wyznaczone minimum) zagwarantuje odrzucenie hipotezy zerowej, co oznacza, że otrzymana wartość współczynnika korelacji będzie istotnie różna od zera.

Przedstawiony sposób rozumowania nie jest wprawdzie zgodny z chronologią prowadzenia badania. Umożliwia jednak takie dostosowanie wielkości próby, aby otrzymany wynik był istotnie różny od zera bez względu na wartość tego współczynnika (wymaga to znajomości przynajmniej orientacyjnej wartości współczynnika korelacji w populacji – informacje takie mogą pochodzić np. z badania pilotażowego). Wyniki obliczeń dla poziomu istotności $\alpha = 0{,}05$ przedstawiono w tab. 3.

Tabela 3. Zależność minimalnej wielkości próby od wartości współczynnika korelacji

r	n_{min}	r	n_{min}
0,01	38 416	0,10	385
0,02	9 605	0,20	97
0,03	4 269	0,30	44
0,04	2 402	0,40	25
0,05	1 538	0,50	16
0,06	1 068	0,60	12
0,07	785	0,70	9
0,08	601	0,80	7
0,09	475	0,90	5

Źródło: opracowanie własne.

Wracając więc do przykładu opisanego w tab. 1, gdzie współczynnik korelacji przyjął wartość 0,01 dla próby 181-elementowej[3], można postawić pytanie o liczebność próby gwarantującą statystyczną istotność uzyskanego wyniku. Jeśli uwzględni się dane z tab. 3, minimalna wielkość próby to 38 416 – czyli 181-elementowa próba musiałaby zostać powielona 213 razy. Oznacza to, że w 38 416-elementowej próbie wystarczyłoby 213 (0,5%) wystąpień obserwacji (1,1; 1,1), aby wartość współczynnika korelacji, wynoszącą 0,01, uznać za istotnie różną od zera.

[3] Obejmującej 20-krotnie powtórzone obserwacje leżące na krawędziach kwadratu i w jego wnętrzu oraz jedną obserwację odstającą.

Opisane własności testu na istotność współczynnika korelacji są oczywiście powszechnie znane, nie zawsze jednak uświadamiane są w pełni ich konsekwencje – zwłaszcza gdy uwaga jest koncentrowana na małych próbach. Jako przykład poniżej zaprezentowany zostanie szkic zadania nr 4.18 (por. (Krysicki i in. 2000, s. 174)).

1. Pobrano próbę 25-elementową pochodzącą z dwuwymiarowego rozkładu normalnego.

2. Obliczono współczynnik korelacji $r = -0{,}1$ oraz przeprowadzono test istotności na poziomie istotności $\alpha = 0{,}05$.

Tabela 4. Dane zadania 4.18

	5	10	15	20	25
5			1		
6		1	3	2	
7	1	3	3	3	1
8		2	3	1	
9			1		

Źródło: opracowanie własne na podstawie (Krysicki i in. 2000, s. 174).

Wartość statystyki $t_e = -0{,}482$ nie znalazła się w zbiorze krytycznym $K = (-\infty; -2{,}069) \cup (2{,}069; \infty)$, co oznacza brak podstaw do odrzucenia hipotezy, że cechy X i Y są nieskorelowane.

W komentarzu do zadania 4.18 podana jest następująca informacja: „dla cech X i Y mających dwuwymiarowy rozkład normalny oznacza to ponadto, że brak podstaw do odrzucenia hipotezy, że cechy X i Y są niezależne” (por. (Krysicki i in. 2000, s. 175)).

Tabela 5. Dane zadania 4.18 po modyfikacji

	5	10	15	20	25
5			20		
6		20	60	40	
7	20	60	60	60	20
8		40	60	20	
9			20		

Źródło: opracowanie własne.

W nawiązaniu do wcześniejszych rozważań, powyższe zadanie zostało zmodyfikowane poprzez 20-krotną replikację wszystkich obserwacji, przez co otrzymana została próba zaprezentowana w tab. 5.

Dla tak skonstruowanej próby współczynnik korelacji wynosi w dalszym ciągu –0,1. Wartość statystyki testowej (ze względu na wielkość próby) wynosi już jednak $t_e = -2{,}243$, co oznacza, że znajduje się w zbiorze krytycznym $K = (-\infty; -1{,}965) \cup (1{,}965; \infty)$. W związku z tym hipotezę zerową o tym, że cechy X i Y są nieskorelowane, należy odrzucić. Pozostaje więc pytanie, czy dla cech X i Y, mających dwuwymiarowy rozkład normalny, oznacza to ponadto, że – w myśl komentarza do zadania 4.18 – należy odrzucić hipotezę, że cechy X i Y są niezależne?

3. Test niezależności χ2

Jako drugi przykład wpływu wielkości próby na interpretację wyników postanowiono wybrać test niezależności χ^2. Problem interpretacyjny pojawił się podczas badania prowadzonego przez autorów we współpracy z Izbą Skarbową we Wrocławiu, dotyczącego oceny funkcjonowania aparatu skarbowego na Dolnym Śląsku. Jednym z kluczowych elementów tego badania była weryfikacja zależności między charakterystyką respondenta i sposobem prowadzenia badania a uzyskanymi opiniami. Badaniem objętych zostało 12 urzędów skarbowych. Każdy z nich został oceniony przez respondentów, których liczba wahała się od 100 do 500. Liczba ankiet zgromadzonych w czasie całego badania przekroczyła 3000.

Kluczowym zagadnieniem, interesującym zarówno autorów, jak i przedstawicieli aparatu skarbowego, było określenie, czy uzyskane wyniki (odpowiedzi respondentów) zależą od cech respondenta oraz sposobu i miejsca ankietowania. Do zweryfikowania tej zależności zdecydowano się wykorzystać test niezależności χ^2, weryfikując następujące hipotezy:

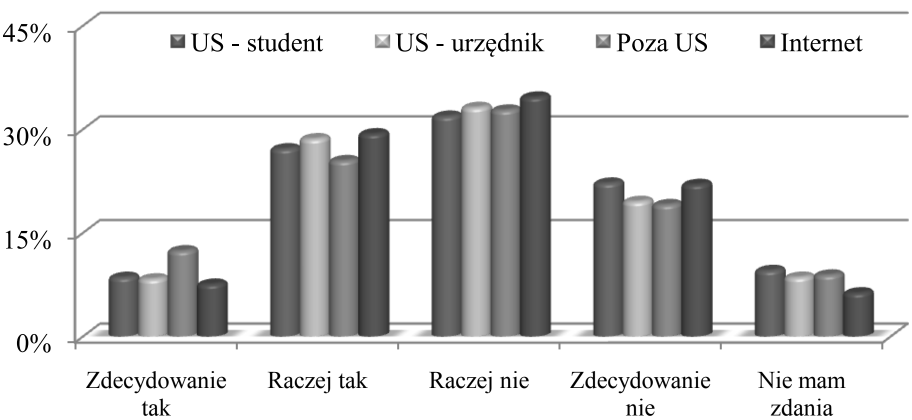
$$H_0: p_{ij} = p_{i\cdot} \cdot p_{\cdot j} \qquad H_1: p_{ij} \neq p_{i\cdot} \cdot p_{\cdot j},$$

gdzie p_{ij} oznacza łączny rozkład zmiennej (X, Y), a $p_{i\cdot}$ oraz $p_{\cdot j}$ rozkłady brzegowe zmiennej X i zmiennej Y. Weryfikacja odbywa się na podstawie wartości statystyki (por. (Ostasiewicz, Rusnak, Siedlecka 2001, s. 299)):

$$\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{s} \frac{\left(n_{ij} - n_{ij}^{'}\right)^2}{n_{ij}^{'}},$$

gdzie n_{ij} oznaczają liczebności empiryczne; n'_{ij} – liczebności teoretyczne; r – liczbę poziomów zmiennej X; s – liczbę poziomów zmiennej Y. Statystyka χ^2, przy założeniu prawdziwości H_0, ma asymptotyczny rozkład χ^2 o $k = (r – 1)(s – 1)$ stopniach swobody.

Testowanie przeprowadzono dla poszczególnych pytań zadanych w ankiecie. Na rys. 2 oraz 3 przedstawione zostały rozkłady odpowiedzi na dwa wybrane pytania.



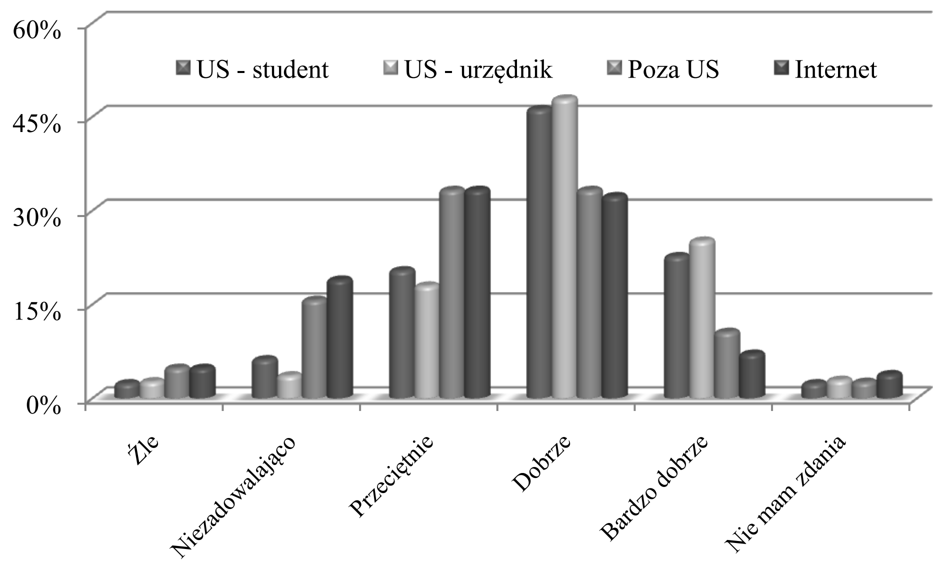
Rys. 2. Rozkłady odpowiedzi na pytanie „Czy uważa Pan(i), że płacenie podatków bywa czasem usprawiedliwione?”

Źródło: opracowanie własne.

W obydwu zaprezentowanych pytaniach wyniki testu niezależności wskazały, że hipotezę zerową o niezależności wyników od miejsca ankietowania należy odrzucić, a *p-value* w obydwu wypadkach była niższa niż 0,01. Podczas opracowywania wyników badania oznaczało to, że istotna statystycznie zależność (czyli odrzucenie hipotezy H_0) powinna zostać orzeczona w wypadku prawie wszystkich zadanych pytań[4]. I o ile wynik ten wydaje się w pełni uzasadniony w kontekście właściwości zastosowanej metody, o tyle interpretacja wyników przez odbiorców badania może być niepoprawna. Pojęcie „statystycznej istotności” przestaje bowiem być intuicyjne w wypadku dużych prób – istotność uzyskanych różnic w odczuciu

[4] Problem ten występował, gdy analizie poddawana była cała badana próba jednocześnie. W wypadku analiz prowadzonych w odniesieniu do poszczególnych urzędów lub sposobów ankietowania wyniki były znacznie bardziej „intuicyjne”.

przeciętnego odbiorcy (porównującego np. graficzne prezentacje rozkładów na wykresie) nie pokrywa się z wynikami statystycznej analizy.



Rys. 3. Rozkłady odpowiedzi na pytanie „Jak ocenia Pan(i) precyzję i jasność wypowiedzi pracowników urzędu?"

Źródło: opracowanie własne.

W szczególnych wypadkach posługiwanie się pojęciem istotności statystycznej w kontekście badań prowadzonych na odpowiednio dużych próbach może prowadzić do niezamierzonej przez badacza interpretacji wyników badań. Między innymi z tego względu w raporcie z badania dolnośląskich urzędów skarbowych wyniki testowania hipotez o niezależności zostały dołączone w jednym z załączników jedynie jako informacja uzupełniająca.

4. Podsumowanie

Przedstawione w artykule rozważania dotyczyły wyłącznie dwóch przykładów z obszaru statystyki matematycznej. Problem wydaje się być jednak szerszy – wnioskowanie statystyczne, mimo popularyzacji technik komputerowych, w dużej mierze w dalszym ciągu nauczane jest w kontekście małych prób. Tymczasem, dzięki upowszechnieniu technik cyfrowych, zbiory danych o dużej i bardzo dużej liczebności są coraz powszechniej wykorzystywane w pracach badawczych oraz bieżącym funkcjonowaniu przedsiębiorstw i instytucji. Nie dotyczy to już wyłącznie danych finansowych o dużej częstotliwości. Systematycznie polepsza się bowiem dostępność danych z różnego typu badań, np. międzynarodowych badań porównawczych prowadzonych przez EUROSTAT oraz narodowe urzędy statystyczne.

I o ile w przeważającej części praktycznych zastosowań problemem w trakcie opracowywania wyników w dalszym ciągu pozostaje zbyt mała wielkość próby, o tyle jednak – jak starano się pokazać w niniejszym artykule – problemy interpretacyjne pojawiają się także w wypadkach, gdy próba jest odpowiednio duża. Wnioskowanie na podstawie dużych prób wymaga bowiem zwrócenia szczególnej uwagi na bardzo dużą wrażliwość testów na stosunkowo niewielkie różnice obserwowane w próbie.

Literatura

Krysicki W., Bartos J., Dyczka W., Królikowska K., Wasilewski M. (2000). *Rachunek prawdopodobieństwa i statystyka matematyczna w zadaniach*. Wydawnictwo Naukowe PWN. Warszawa.

Ostasiewicz S., Rusnak Z., Siedlecka U. (2001). *Statystyka. Elementy teorii i zadania*. Wydawnictwo Akademii Ekonomicznej we Wrocławiu. Wrocław.

Steczkowski J. (1995). *Metoda reprezentacyjna w badaniach zjawisk ekonomiczno-społecznych*. PWN. Warszawa-Kraków.