

STUDIA METODOLOGICZNE

Piotr SULEWSKI

Ocena zdolności tablic dwudzielczych do wykrywania związku między uporządkowanymi cechami typu jakościowego

Tablice wielodzielcze są podstawowym i bardzo często stosowanym narzędziem służącym do badania związku między cechami. Tablicę, która powstaje w wyniku podziału danych według dwóch cech nazywamy dwudzielczą, trzech cech — trójdzielczą itd. Przedmiotem obecnej pracy są tablice dwudzielcze.

Procedura weryfikacji hipotezy podstawowej H_0 o braku stochastycznej zależności między badanymi cechami polega na ustaleniu poziomu istotności testu (najczęściej 5%, rzadziej 10%), a następnie na odnalezieniu w odpowiednich tablicach rozkładu statystyki testowej wartości krytycznej na tym poziomie istotności i porównaniu z wartością statystyki testowej. Gdy ta druga wartość jest choćby tylko nieco mniejsza od pierwszej, to H_0 zostaje przyjęta. Jednak należy zdawać sobie sprawę, jak duże jest prawdopodobieństwo, że przyjmując H_0 można przyjąć hipotezę fałszywą.

Przeprowadźmy zatem następujący eksperyment numeryczny. W pierwszym kroku jego realizacji ustalmy takie prawdopodobieństwa przynależności danej obserwacji do komórki leżącej w i -tym wierszu i j -tej kolumnie, by istniał związek między cechą „po kolumnach” i cechą „po wierszach”. Zatem hipoteza H_0 nie jest prawdziwa. Traktujemy numery wierszy i kolumn jako realizację dwuwymiarowej zmiennej losowej dyskretnej. Siłę związku liczbowo wyraża kowariancja zmiennych. W kroku drugim wygenerujmy wielokrotnie tablice dwudzielcze i przeprowadźmy na każdej z nich test weryfikujący hipotezę H_0 . Następnie policzmy, ile razy H_0 została przyjęta, chociaż słuszną nie jest. W ten

sposób uzyskamy krzywą operacyjno-charakterystyczną, którą można nazwać funkcją słabości testu. W praktyce posługujemy się funkcją mocy testu, będącą dopełnieniem funkcji słabości do jedności. Warto przy tym zauważyć, że nazwa „funkcja mocy testu” lepiej brzmi.

Stosując tablicę dwudzielczą powinniśmy zadać pytanie, jaką zdolność do wykrywania związku między cechami mają tablice dwudzielcze, tzn. jaka jest ich moc. Artykuł pokazuje, jak można odpowiedzieć na to pytanie.

Na podstawie przeprowadzonej analizy można poznać moc testu jedynie dla tablic dwudzielczych 2×2 , natomiast w przypadku tablic o większych wymiarach konieczne jest generowanie tablic dwudzielczych i określenie ich mocy za pomocą badań symulacyjnych. Przeprowadzone przez autora eksperymenty numeryczne pokazały, że wyznaczenie mocy testów na drodze analitycznej i porównanie uzyskanych wyników ze stosownymi wartościami empirycznymi jest możliwe — jak wspomniano wcześniej — jedynie dla tablicy dwudzielczej 2×2 . Przedstawione w pracy wyniki pozwalają czytelnikowi zorientować się, jak moc tablic dwudzielczych zależy od liczebności próby oraz od siły związku między cechami.

W literaturze statystycznej istnieje wiele miar statystycznych testowych związanych z tablicami dwudzielczymi. Spośród nich szczególnie często stosuje się statystykę χ^2_{XY} Pearsona. Istnieją jednak pewne ograniczenia w stosowaniu statystyki χ^2_{XY} Pearsona, która — co warto przypomnieć — ma asymptotyczny rozkład chi-kwadrat z $(w-1)(k-1)$ stopniami swobody. W celu zniesienia tych ograniczeń w pracy (Sulewski, 2015) proponuje się wyznaczanie wartości krytycznych za pomocą symulacji komputerowych metodą Monte Carlo.

Praca składa się z dwóch części. Część I ma charakter teoretyczny. Zdefiniowano w niej statystykę χ^2_{XY} Pearsona dla tablic dwudzielczych, centralny i niecentralny rozkład chi-kwadrat oraz kowariancję i współczynnik korelacji zmiennych. Przedstawiono także metodę generowania zawartości tablic dwudzielczych, sposób wyznaczania wartości krytycznych oraz metodę wyznaczania empirycznej i teoretycznej mocy testów. Część II pracy ma charakter eksperymentalny. Składają się na nią przykłady wyznaczania mocy tablic dwudzielczych wraz z ich implementacją komputerową napisaną w języku VBA (*Visual Basic for Applications*) arkusza kalkulacyjnego Excel. Kopię tej implementacji umieszczono w Internecie, co pomoże czytelnikom samodzielnie prowadzić badania statystyczne. Zamieszczono też wykaz użytych procedur i funkcji oraz liczne komentarze dotyczące kodów źródłowych, które są pomocne w poznaniu mechanizmu tworzenia tej implementacji.

W załączonym w Internecie arkuszu kalkulacyjnym można także modelować wykresy funkcji gęstości $f(x)$ (arkusz wyk_r_f(x)) i dystrybuanty $F(x)$ (arkusz wyk_r_F(x)) rozkładu chi-kwadrat (centralnego) oraz funkcji gęstości $f_N(x)$ (arkusz wyk_r_fN(x)) i dystrybuanty $F_N(x)$ (arkusz wyk_r_FN(x)) niecentralnego rozkładu chi-kwadrat.

STATYSTYKA χ^2_{XY} PEARSONA

Tabl. 1 przedstawia tablicę dwudzielczą $w \times k$, która składa się z $w \cdot k$ wartości n_{ij} ($i = 1, \dots, w$; $j = 1, \dots, k$) rozkładu łącznego cech X i Y .

TABL. 1. TABLICA DWUDZIELCZA $w \times k$

Cecha X	Cecha Y				Razem
	Y_1	Y_2	...	Y_k	
X_1	n_{11}	n_{12}	...	n_{1k}	$n_{1\bullet}$
X_2	n_{21}	n_{22}	...	n_{2k}	$n_{2\bullet}$
...
X_w	n_{w1}	n_{w2}	...	n_{wk}	$n_{w\bullet}$
Razem	$n_{\bullet 1}$	$n_{\bullet 2}$...	$n_{\bullet k}$	n

Źródło: opracowanie własne.

Wartość n jest sumą wszystkich wartości n_{ij} tablicy dwudzielczej:

$$n = \sum_{i=1}^w n_{i\bullet} = \sum_{j=1}^k n_{\bullet j} = \sum_{i=1}^w \sum_{j=1}^k n_{ij} \quad (1)$$

Statystyka χ^2_{XY} Pearsona do badania niezależności cech w tablicy dwudzielczej $w \times k$ ma postać¹:

$$\chi^2_{XY} = \sum_{i=1}^w \sum_{j=1}^k \frac{(n_{ij} - e_{ij})^2}{e_{ij}} \quad (2)$$

gdzie:

n_{ij} — liczebność zaobserwowana w i -tym wierszu i j -tej kolumnie,

e_{ij} — liczebność oczekiwana i -tego wiersza i j -tej kolumny dana wzorem:

$$e_{ij} = \frac{n_{i\bullet} \cdot n_{\bullet j}}{n} \quad (i = 1, 2, \dots, w; \quad j = 1, 2, \dots, k) \quad (3)$$

W przedstawionym teście niezależności liczebność próby musi być znaczna, aby w każdej komórce tablicy znalazła się możliwie duża liczba obserwacji.

Dla tablicy dwudzielczej 2×2 wzór (2) można stosować, gdy $n \geq 40$ oraz $e_{ij} \geq 5$ dla każdego $i, j = 1, 2$. Jeżeli $1 \leq e_{ij} < 5$ dla pewnego $i, j = 1, 2$, to zda-

¹ Pearson (1900), s. 157—172.

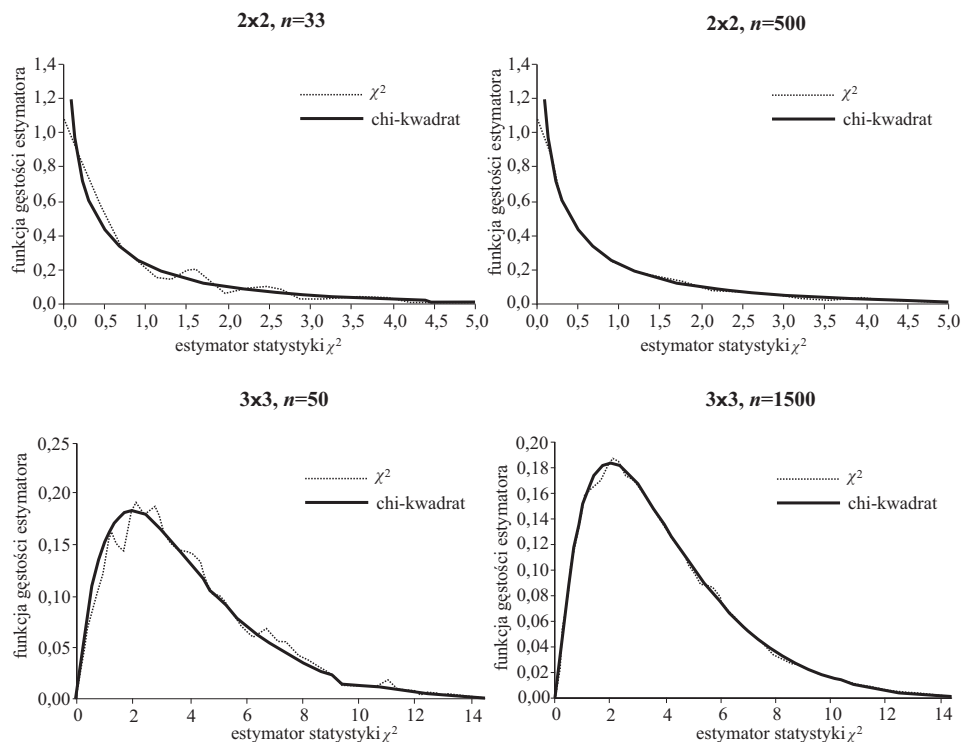
niem Campbella dla tabel 2×2 należy stosować test $n-1$ chi-kwadrat w postaci²:

$${}_{n-1}\chi_{XY}^2 = \frac{n-1}{n} \chi_{XY}^2 \quad (4)$$

Jeżeli $e_{ij} < 1$ dla pewnego $i, j = 1, 2$, to należy stosować test Fishera-Yatesa (Campbell, 2007).

W przypadku tablicy dwudzielczej większej niż 2×2 — zdaniem m.in. Yatesa, Moore'a i McCabe'a (1999) oraz Shiera (2004) — wzór (2) należy stosować, gdy $e_{ij} \geq 1$ dla każdego $i = 1, \dots, w; j = 1, \dots, k$, w sytuacji gdy nie więcej niż 20% wartości oczekiwanych e_{ij} jest mniejsze niż 5.

ROZKŁAD EMPIRYCZNY STATYSTYKI χ^2_{XY} PEARSONA I ROZKŁAD CHI-KWADRAT DLA TABELI $w \times k$ I LICZEBNOŚCI PRÓBY n



Źródło: opracowanie własne.

² Cochran (1952), s. 315—345; Campbell (2007), s. 3661—3675; Pearson (1947), s. 139—167.

Gdy hipoteza H_0 o niezależności cech X, Y jest słuszna, statystyka χ^2_{XY} Pearsona ma asymptotyczny rozkład chi-kwadrat z $df = (w-1)(k-1)$ stopniami swobody. Wykres przedstawia funkcję gęstości rozkładu chi-kwadrat z $df = 1$ i $df = 4$ stopniami swobody oraz empiryczną funkcję gęstości rozkładu statystyki χ^2_{XY} uzyskaną metodą Parzena dla tablicy 2×2 , 3×3 i liczebności próby n , gdy H_0 jest słuszna. Dokładny opis tej metody znaleźć można w pracy (Sulewski, 2013). Liczebność próby w przypadku tablicy 3×3 jest większa niż tablicy 2×2 , gdyż tym razem dysponujemy większą liczbą komórek. Liczebność próby dobrano w taki sposób, aby pokazać:

- że rozkład statystyki χ^2_{XY} Pearsona znacznie odbiega od rozkładu chi-kwadrat (2×2 , $n=33$; 3×3 , $n=50$);
- zgodność rozkładu statystyki χ^2_{XY} Pearsona z rozkładem chi-kwadrat (2×2 , $n=500$; 3×3 , $n=1500$).

KOWARIANCJA I WSPÓŁCZYNNIK KORELACJI

Współczynnik korelacji r Pearsona zależy nie tylko od siły związku między cechami, ale także od momentów centralnych rzędu II każdej zmiennej z osobna. Kowariancja natomiast wyraża „czystą” siłę związku między cechami. Zestawienie (1) przedstawia przykładowe wartości prawdopodobieństw p_{ij} dla tablic dwudzielczych 2×2 , przy ustalonej wartości współczynnika korelacji $r = 0$. Tablice z prawdopodobieństwami p_{ij} ($i = 1, \dots, w$; $j = 1, \dots, k$) wykorzystano w procesie generacji tablic dwudzielczych $w \times k$ opisaną poniżej metodą słupkową.

ZESTAWIENIE (1) WARTOŚCI PRAWDOPODOBIEŃSTW p_{ij}
W TABLICY DWUDZIELCZEJ 2×2 DLA $r = 0$

$p_{ij}, r = 0$			$p_{ij}, r = 0$		
ij	1	2	ij	1	2
1	0,250	0,250	1	0,229	0,250
2	0,250	0,250	2	0,249	0,272

$p_{ij}, r = 0$			$p_{ij}, r = 0$		
ij	1	2	ij	1	2
1	0,179	0,244	1	0,151	0,244
2	0,245	0,333	2	0,238	0,373

Źródło: jak przy tabl. 1.

Zbiór tablic dwudzielczych różniących się wartościami prawdopodobieństw p_{ij} , a jednocześnie o takiej samej wartości współczynnika korelacji r jest zapewne bardzo liczny. Aby radykalnie zmniejszyć licznosc tego zbioru współczynnik korelacji r zastapiono kowariancją cov_{XY} oraz wprowadzono dwa dodatkowe kryteria przynależności do zbioru — zerowe wartości momentów centralnych mieszanych rzędu trzeciego, mianowicie μ_{12} i μ_{21} . Kryteria te są pewnym unormalnieniem rozkładu. Jak wiadomo rozkład normalny dwuwymiarowy ma wszystkie momenty centralne mieszane rzędu wyższego niż dwa równe zero.

Wykorzystując excelowski Solver wyznaczono takie prawdopodobieństwa p_{ij} , które minimalizują funkcję celu postaci:

$$Fc = [\text{cov}_{XY} - \text{cov} T_{XY}]^2 + \mu_{12}^2 + \mu_{21}^2 \quad (5)$$

W (5) $\text{cov} T_{XY}$ jest kowariancją teoretyczną, cov_{XY} — kowariancją empiryczną daną wzorem:

$$\text{cov}_{XY} = E[(X - E(X))(Y - E(Y))], \quad \text{cov}_{XY} \in \langle -s_X s_Y, s_X s_Y \rangle \quad (6)$$

natomiast μ_{12} i μ_{21} są momentami centralnymi rzędu III danymi wzorami:

$$\mu_{21} = E(X^2 Y) - E(X^2)E(Y) - 2E(X)\text{cov}_{XY} \quad (7)$$

$$\mu_{12} = E(XY^2) - E(X)E(Y^2) - 2E(Y)\text{cov}_{XY} \quad (8)$$

Zestawienie (2) przedstawia wartości prawdopodobieństw p_{ij} dla tablic dwudzielczych 2×2 , przy ustalonej wartości kowariancji cov_{XY} .

**ZESTAWIENIE (2) WARTOŚCI PRAWDOPODOBIENSTW p_{ij}
W TABLICY DWUDZIELCZEJ 2×2 DLA $\text{cov}_{XY} = \{0; 0,05; 0,10; 0,15\}$**

$p_{ij}, \text{cov}_{XY} = 0$			$p_{ij}, \text{cov}_{XY} = 0,05$		
ij	1	2	ij	1	2
1	0,250	0,250	1	0,306	0,203
2	0,250	0,250	2	0,197	0,294
$p_{ij}, \text{cov}_{XY} = 0,10$			$p_{ij}, \text{cov}_{XY} = 0,15$		
ij	1	2	ij	1	2
1	0,349	0,151	1	0,400	0,102
2	0,149	0,351	2	0,098	0,400

Źródło: jak przy tabl. 1.

GENEROWANIE ZAWARTOŚCI TABLIC DWUDZIELCZYCH

W celu generacji zawartości tablicy dwudzielczej $w \times k$ przedział $(0;1)$ podzielono na $w \cdot k$ podprzedziałów o szerokościach równych wartości prawdopodobieństw p_{ij} w taki sposób, że pierwszy podprzedział ma szerokość p_{11} , drugi — p_{12} , ..., k -ty — p_{1k} , ..., ostatni — p_{wk} . Wielkości p_{ij} spełniają warunek normalizacji:

$$\sum_{i=1}^w \sum_{j=1}^k p_{ij} = 1 \quad (9)$$

i dobrano je w taki sposób, aby uzyskać żadaną wartość kowariancji.

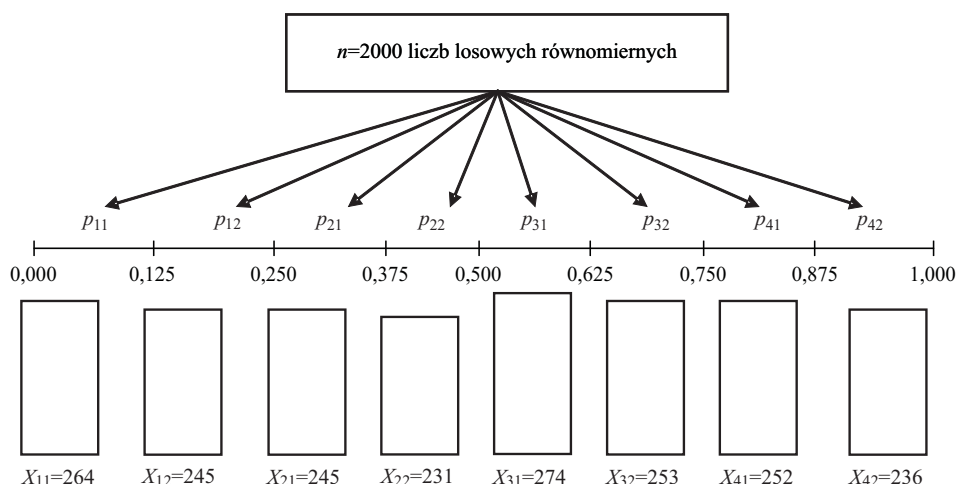
Każda z n wygenerowanych liczb losowych o rozkładzie równomiernym w przedziale $(0;1)$ wpada do jednego z $w \cdot k$ podprzedziałów i tym samym liczba obiektów w odpowiadającej temu podprzedziałowi komórce tablicy zostaje zwiększona o jedną. Wielkości n_{ij} spełniające równość:

$$\sum_{i=1}^w \sum_{j=1}^k n_{ij} = n \quad (10)$$

określają wielkość obiektów w poszczególnych komórkach tablicy.

Schemat przedstawia wypełnianie komórek tablicy dwudzielczej 4×2 dla liczebności próby $n = 2000$ i kowariancji $\text{cov}_{XY} = 0$.

SCHEMAT WYPEŁNIANIA KOMÓREK TABLICY 4x2



Ź r ó d ł o: opracowanie własne.

Tabl. 2 prezentuje odpowiadającą temu schematowi tablicę dwudzielczą.

TABL. 2. TABLICA DWUDZIELCZA 4×2 , GDY CECHY X I Y SĄ NIEZALEŻNE ($\text{cov}_{XY} = 0$)

Cecha X	Cecha Y		Razem
	Y_1	Y_2	
X_1	264	245	509
X_2	245	231	476
X_3	274	253	527
X_4	252	236	488
Razem	1035	965	2000

Źródło: jak przy tabl. 1.

WYZNACZANIE WARTOŚCI KRYTYCZNYCH

Wcześniej przedstawiono warunki, jakie muszą być spełnione, aby można było w testach niezależności dla tablic dwudzielczych stosować statystykę opisaną wzorem (2). W dobie coraz szybszych komputerów można za pomocą stosownego oprogramowania znieść te ograniczenia i na podstawie symulacji wyznaczyć wartości krytyczne.

Wartości krytyczne wyznaczono metodą Monte Carlo wyznaczając empiryczne kwantyle, którymi są wartości stosownej statystyki pozycyjnej (David, 1970). W celu wyznaczenia wartości krytycznych dla danej tablicy dwudzielczej oraz liczebności próby wyznaczono $R = 10^5$ wartości statystyki testowej (2) i za wartość krytyczną przyjęto ich $[(1 - \alpha)R]$ statystykę pozycyjną, gdzie α jest poziomem istotności oraz [...] jest częścią całkowitą liczby. Tak duża liczba powtórzeń w obliczeniach statystyki testowej zapewnia uzyskanie dokładnego wyniku. Przy wyznaczaniu wartości krytycznych, gdy między cechami nie ma związku, zawartość tablic dwudzielczych generowano za pomocą metody słupkowej. Generowanie zawartości tablic dwudzielczych innymi metodami, przy założeniu że między cechami nie ma związku, nie miało wpływu na uzyskane wielkości wartości krytycznych.

MOC TABLICY DWUDZIELCZEJ

W testowaniu hipotez statystycznych ich prawdziwość lub fałszywość ocenia się na podstawie wyników próby losowej. Przy takim postępowaniu istnieje możliwość popełnienia błędu. Wyniki z próby mogą przyczynić się do odrzucenia hipotezy H_0 , gdy w istocie jest prawdziwa. Mówi się wtedy o błędzie I rodzaju, a prawdopodobieństwo jego popełnienia oznacza się przez α . W sytuacji odwrotnej przyjmujemy H_0 , która jest fałszywa. Wtedy mamy do czynienia

nia z błędem II rodzaju, którego prawdopodobieństwo popełnienia oznacza się przez β .

Najlepszy byłby test, w którym prawdopodobieństwa α i β byłyby jak najmniejsze, jednak równoczesna ich minimalizacja nie jest możliwa. Dlatego testy buduje się w taki sposób, aby przy ustalonym z góry prawdopodobieństwie popełnienia błędu I rodzaju α zminimalizować prawdopodobieństwo popełnienia błędu II rodzaju β . Tak zbudowane testy nazywane są testami najmocniejszymi, odpowiada im bowiem przy ustalonym α największa moc M , czyli prawdopodobieństwo odrzucenia hipotezy zerowej H_0 , gdy nie jest ona prawdziwa. Moc testu zależy od liczebności próby — im liczniejsza próba, tym większa moc, a także od poziomu istotności testu — im niższy poziom istotności, tym mniejsza moc testu.

W literaturze statystycznej³ rekomendowanym poziomem mocy testu jest wartość większa lub równa 0,8. Oznacza to, że przy ustalonym poziomie istotności α wielkość błędu II rodzaju β nie może być większa niż 0,2.

Przystępując do testowania testów trzeba dysponować narzędziem do modelowania populacji generalnej, czyli do nadawania populacji generalnej określonej właściwości, jaką jest związek między cechami. W związku z tym dane podlegające opracowaniu muszą być danymi pochodzącymi z generatora liczb losowych, a nie danymi wziętymi z praktyki. Należy wiedzieć, jaki jest rzeczywisty związek między cechami, inaczej nie zbada się mocy testów.

W celu wyznaczenia empirycznej mocy testu — czyli zdolności tablicy dwudzielczej do odrzucenia hipotezy mówiącej o tym, że związku nie ma — niezbędne jest generowanie tablic dwudzielczych. Uwzględniając narzuconą siłę związku między cechami, do wypełnienia tablicy dwudzielczej skorzystano z metody słupkowej wykorzystującej rozkład równomierny.

Jeżeli dysponuje się tablicą dwudzielczą 2×2 z prawdopodobieństwami $TP_{ij} = 0,25$ ($i, j = 1, 2$) odpowiadającymi przypadkowi niezależności cech, a danej zależności cech odpowiada tablica dwudzielcza 2×2 z prawdopodobieństwami p_{ij} ($i, j = 1, 2$), to moc testu M_T dla statystyki χ^2 Pearsona na poziomie istotności α można wyznaczyć analitycznie korzystając z niecentralnego rozkładu chi-kwadrat (Zieliński, 1972):

$$M_T = 1 - F_N(\chi^2_{1-\alpha,1}; 1, \lambda) \quad (11)$$

gdzie:

$\chi^2_{1-\alpha,1}$ — wartość krytyczna testu chi-kwadrat,

³ Cohen (1965), s. 92—121.

$F_N(x; 1, \lambda)$ — dystrybuanta niecentralnego rozkładu chi-kwadrat dla 1 stopnia swobody i parametru niecentralności:

$$\lambda = n \sum_{i=1}^w \sum_{j=1}^k \frac{(p_{ij} - Tp_{ij})^2}{Tp_{ij}} \quad (12)$$

Znając wartość statystyki testowej χ_{XY}^2 danej wzorem (2) dla różnej liczebności próby n wyznaczono empiryczną moc testu za pomocą wzoru:

$$M_E = \frac{V - U}{V} \quad (13)$$

gdzie U określa liczbę tych spośród V wszystkich możliwych przypadków, kiedy to wartość statystyki χ_{XY}^2 jest mniejsza od wartości krytycznej $\chi_{1-\alpha, (w-1)(k-1)}^2$ na poziomie istotności α .

W celu porównania teoretycznej oraz empirycznej mocy testów przeprowadzono test równości prawdopodobieństw M_T i M_E wykorzystując statystykę:

$$z = \frac{M_E - M_T}{\sqrt{\frac{M_T \cdot (1 - M_T)}{n}}} \quad (14)$$

która ma asymptotycznie rozkład normalny. Hipotezę zerową $H_0 : M_E = M_T$ odrzuca się, jeżeli:

$$z \leq -u_{1-\alpha/2} \vee z \geq u_{1-\alpha/2} \quad (15)$$

Wartości krytyczne u_α można wyznaczyć za pomocą formuły arkuszowej — Rozkład.Normalny.S.Odw(α) albo za pomocą funkcji — Application. NormSInv(α).

Przykład 1

Zbadano zdolność tablicy dwudzielczej 2×2 do wykrywania związku między cechami X, Y na poziomie istotności $\alpha = 0,05$ wykorzystując statystykę χ_{XY}^2 dla $n \in \{50; 100\}$ i $\text{cov}_{XY} \in \{0; 0,05; 0,075; 0,1; 0,125; 0,15\}$. Empiryczne moce testów porównano z mocami teoretycznymi za pomocą testu równości prawdopodobieństw.

Zestawienie (3) zawiera wykaz procedur i funkcji wykorzystanych do badania mocy testu w tablicach dwudzielczych.

ZESTAWIENIE (3) PROCEDUR I FUNKCJI WYKORZYSTANYCH DO BADANIA MOCY TESTU

Wyszczególnienie	Typ	Realizowane zadanie
MocTestu	procedura	wyznaczanie mocy testu dla tablic dwudzielczych
ExpNij	funkcja	wyznaczanie liczebności oczekiwanej
SX2	funkcja	obliczanie sum brzegowych w wierszach
SY2	funkcja	obliczanie sum brzegowych w kolumnach
GenTab	funkcja	generowanie zawartości tablicy dwudzielczej
CzyscDane	procedura	czyszczenie komórek
Kwantyle	funkcja	wyznaczanie wartości krytycznej
Sort	funkcja	sortowanie
ChiKw1	funkcja	obliczanie wartości statystyki χ^2_{XY}
ChiKw2	funkcja	obliczanie wartości statystyki $n-1 \chi^2_{XY}$
Solver1	funkcja	przygotowanie Solvera
Solver2	funkcja	usuwanie warunków ograniczających z Solvera
DChiKw	funkcja	obliczanie dystrybuanty rozkładu chi-kwadrat (centralnego)
DNChiKw	funkcja	obliczanie dystrybuanty niecentralnego rozkładu chi-kwadrat
Frakcja	funkcja	przeprowadzanie testu równości prawdopodobieństw
Gamma	funkcja	wyznaczanie wartości funkcji gamma
GammaN	funkcja	wyznaczanie wartości niepełnej funkcji gamma

Źródło: jak przy tabl. 1.

Implementację komputerową wyznaczania mocy testu dla statystyki χ^2 Pearsona utworzoną w edytorze VBA arkusza kalkulacyjnego Excel przedstawiono w pliku MocChi, który umieszczono w Internecie⁴. Po wprowadzeniu wartości zmiennych, oznaczonych w zestawieniu (4) szarym tłem, uruchomiono procedurę MocTestu. Kody tej procedury uruchamiają excelowski Solver, który minimalizując funkcję celu (5) w arkuszu p -wa tworzy wartości prawdopodobieństw p_{ij} ($i = 1, \dots, w$; $j = 1, \dots, k$) wykorzystane w procesie generowania tablic dwudzielczych. Efektem końcowym działania procedury MocTestu jest teoretyczna i empiryczna moc testu oraz test równości prawdopodobieństw (zestawienie (4)).

ZESTAWIENIE (4) WYNIKÓW DZIAŁANIA PROCEDURY MOCY TESTU DLA $n = 50$

Wyszczególnienie	Wartości otrzymanych wyników
Liczba wierszy	$w=2$
Liczba kolumn	$k=2$
Poziom istotności	$\alpha=0,05$
Liczebność próby	$n=50$
Kowariancja	$cov_{XY}=0$
Liczba powtórzeń testu	$V=5000$
Teoretyczna moc testu	$M_T=0,051$
Empiryczna moc testu	$M_E=0,058$
Test równości	$H_0: M_T=M_E$ — nieodrzucona

Źródło: jak przy tabl. 1.

⁴ Pod adresem <http://www.utogim.eu/mocchi.xls>.

Przeprowadzając powyższe symulacje dla pozostałych wartości kowariancji otrzymano następujące wyniki:

ZESTAWIENIE (5) TEORETYCZNEJ I EMPIRYCZNEJ MOCY TESTU DLA $n = 50$

cov_{XY}	M_T	M_E	$H_0: M_T=M_E$
0,000	0,051	0,058	nieodrzucona
0,050	0,293	0,304	nieodrzucona
0,075	0,564	0,591	nieodrzucona
0,100	0,807	0,836	nieodrzucona
0,125	0,942	0,960	nieodrzucona
0,150	0,989	0,998	nieodrzucona

Źródło: jak przy tabl. 1.

Uruchamiając procedurę MocTestu dla liczebności próby $n = 100$ wyznaczono teoretyczną i empiryczną moc testu oraz przeprowadzono test równości prawdopodobieństw (zestawienie (6)).

ZESTAWIENIE (6) WYNIKÓW DZIAŁANIA PROCEDURY MOCY TESTU DLA $n = 100$

Wyszczególnienie	Wartości otrzymanych wyników
Liczba wierszy	$w=2$
Liczba kolumn	$k=2$
Poziom istotności	$\alpha=0,05$
Liczebność próby	$n=100$
Kowariancja	$cov_{XY}=0$
Liczba powtórzeń testu	$V=5000$
Teoretyczna moc testu	$M_T=0,053$
Empiryczna moc testu	$M_E=0,049$
Test równości	$H_0: M_T=M_E$ — nieodrzucona

Źródło: jak przy tabl. 1.

Przeprowadzając powyższe symulacje dla pozostałych wartości kowariancji otrzymano następujące wyniki:

ZESTAWIENIE (7) TEORETYCZNEJ I EMPIRYCZNEJ MOCY TESTU DLA $n = 100$

cov_{XY}	M_T	M_E	$H_0: M_T=M_E$
0,000	0,053	0,049	nieodrzucona
0,050	0,516	0,535	nieodrzucona
0,075	0,851	0,868	nieodrzucona
0,100	0,979	0,983	nieodrzucona
0,125	0,999	1,000	nieodrzucona
0,150	1,000	1,000	nieodrzucona

Źródło: jak przy tabl. 1.

Dla $cov_{XY} = 0$ (zestawienie (5) i (7)), gdy hipoteza zerowa H_0 o braku związku między cechami X i Y jest słuszna, moc testu oscyluje wokół założonego poziomu istotności α , co potwierdza prawidłowość przeprowadzonych symu-

lacji. Wraz ze wzrostem cov_{XY} oraz n moc testu rośnie i największe wartości przyjmuje dla $\text{cov}_{XY} = 0,15$, gdy niespełnienie H_0 jest największe. We wszystkich przypadkach hipoteza zerowa $H_0 : M_T = M_E$ o równości mocy testów nie została odrzucona. Fakt ten także potwierdza prawidłowość przeprowadzonych symulacji komputerowych metodą Monte Carlo dla tablic dwudzielczych 2×2 i można za pomocą proponowanego narzędzia wyznaczać empiryczną moc testów także dla tablic większych niż 2×2 (przykład 2).

Przykład 2

Zbadano zdolność tablicy dwudzielczej 4×2 do wykrywania związku między cechami X, Y na poziomie istotności $\alpha = 0,05$, wykorzystując statystykę χ^2_{XY} dla $n \in \{100, 150\}$ i $\text{cov}_{XY} \in \{0; 0,05; 0,075; 0,1; 0,125; 0,15; 0,175; 0,2\}$.

Zestawienie (3) zawiera wykaz procedur i funkcji wykorzystanych do badania mocy testu w tablicach dwudzielczych. Po wprowadzeniu wartości zmiennych oznaczonych szarym wypełnieniem (zestawienie (8)) uruchomiono procedurę MocTestu, w wyniku działania której wyznaczono empiryczną moc testu.

ZESTAWIENIE (8) WYNIKÓW DZIAŁANIA PROCEDURY MOCY TESTU DLA $n = 100$

Wyszczególnienie	Wartości otrzymanych wyników
Liczba wierszy	$w=4$
Liczba kolumn	$k=2$
Poziom istotności	$\alpha=0,05$
Liczebność próby	$n=100$
Kowariancja	$\text{cov}_{XY}=0$
Liczba powtórzeń testu	$V=5000$
Empiryczna moc testu	$M_E=0,053$

Źródło: jak przy tabl. 1.

Przeprowadzając powyższe symulacje dla pozostałych wartości kowariancji oraz dla $n = 150$ otrzymano następujące wyniki:

ZESTAWIENIE (9) EMPIRYCZNEJ MOCY TESTU DLA TABLICY 4×2 i $n \in \{100, 150\}$

cov_{XY}	$n=100$	$n=150$
0,000	0,053	0,054
0,050	0,126	0,183
0,075	0,257	0,355
0,100	0,417	0,600
0,125	0,609	0,811
0,150	0,801	0,942
0,175	0,922	0,989
0,200	0,966	0,999

Źródło: jak przy tabl. 1.

Dla $\text{cov}_{XY} = 0$ (zestawienie (9)), gdy hipoteza zerowa H_0 o braku związku między cechami X i Y jest słuszna, moc testu oscyluje wokół założonego poziomu istotności α , co potwierdza prawidłowość przeprowadzonych symulacji. Wraz ze wzrostem cov_{XY} oraz n moc testu rośnie i największe wartości przyjmuje dla $\text{cov}_{XY} = 0,2$, gdy niespełnienie H_0 jest największe.

Podsumowanie

W badaniu niezależności cech w tablicy $w \times k$ najważniejsza jest zaproponowana przez Pearsona statystyka χ^2_{XY} . W celu zniesienia ograniczeń w zakresie stosowania statystyki χ^2_{XY} posiadającej rozkład chi-kwadrat z $(w-1)(k-1)$ stopniami swobody wartości krytyczne wyznaczono na podstawie symulacji.

Artykuł udziela odpowiedzi na pytanie, jaka jest zdolność tablic dwudzielczych do wykrywania związku między cechami, inaczej mówiąc — jaka jest ich moc. W tym celu generowano zawartość tablic dwudzielczych i wyznaczono empiryczną moc testów. Dla tablicy 2×2 obliczono także teoretyczną moc testów, którą porównano z wartością empiryczną. Brak podstaw do odrzucenia hipotezy o równości tychże prawdopodobieństw potwierdza prawidłowość przeprowadzonych symulacji metodą Monte Carlo.

Dzięki gotowej implementacji komputerowej i dokładnie opisanym przykładom czytelnik może samodzielnie badać moc testów z wykorzystaniem statystyki χ^2_{XY} Pearsona.

Analiza mocy testu staje się bardzo ważnym narzędziem w pracy badacza, szczególnie na etapie planowania doświadczeń, jak również po zakończeniu badań, jako element skuteczności podjętych działań.

dr Piotr Sulewski — Akademia Pomorska w Słupsku

LITERATURA

- Campbell I. (2007), *Chi-squared and Fisher-Irwin tests of two-by-two tables with small sample recommendations*, „Statistics in Medicine”, Vol. 26, No. 19
- Cochran W. G. (1952), *The χ^2 test of goodness of fit*, „Annals of Mathematical Statistics”, Vol. 23, No. 3
- Cohen J. (1965), *Some statistical issues in psychological research*, [w:] B. B. Wolman (red.), *Handbook of clinical psychology*, New York: Academic Press
- David H. A. (1970), *Order statistics*, Wiley, New York
- Pearson E. (1947), *The choice of statistical tests illustrated on the interpretation of data classed in a 2×2 table*, „Biometrika”, Vol. 34
- Pearson K. (1900), *On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling*, „Philosophy Magazine”, Series (5), Vol. 50

- Shier R. (2004), *The Chi-squared test for two-way tables*, Mathematics Learning Support Centre
- Sulewski P. (2013), *Modyfikacja testu niezależności*, „Wiadomości Statystyczne”, nr 10, GUS
- Sulewski P. (2015), *Wyznaczanie obszaru krytycznego przy testowaniu niezależności w tablicach wielodzielczych*, „Wiadomości Statystyczne”, nr 3, GUS
- Yates D., Moore D., McCabe G. (1999), *The Practice of Statistics* (1st Ed.), New York, W. H. Freeman
- Zieliński R. (1972), *Tablice statystyczne*, PWN, Warszawa

SUMMARY

Proposed by Pearson in 1900 χ^2_{XY} formula is still the most important measure to study the characteristics independence, especially since it has its extension for three variable and higher tables. The question is, what is the ability of two variable tables to detect relationship between features, what is their power. It is difficult to answer this question on the basis of the analysis. The best way seems to be generating two variable tables and determine power through simulation studies. For the 2x2 two variable table is it also possible to designate test power on the analytical way as well as comparison of obtained analytical results with empirical values.

The work results will allow the reader to get an idea of the extent to which power of two variable tables depends on the sample size and the strength of the association between features. Aim of this study is to provide a ready computer implementation to test power of two variable tables stated as a set on the Internet. Presented theory and some examples will help readers to explore the test power using Pearson's χ^2 statistics and model the course of the density function and cumulative distribution central and non-central chi-square distribution.

РЕЗЮМЕ

Предложенная Пирсоном в 1900 г. статистика χ^2_{XY} , является все еще самым важным измерителем для обследования независимости характеристик, тем более что она имеет свое расширение для трехразделительных таблиц и выше.

Тем не менее возникает вопрос, какой является способность двухразделительных таблиц для обнаружения связи между характеристиками, то есть какой является их мощность. Трудно ответить на этот вопрос на основе анализа, поэтому наилучшим способом кажется быть разработка двухразделительных таблиц и определение мощности с использованием моделированных обследований. Для двухразделительной таблицы 2x2 возможным является также определение мощности критерия с использованием анализа и сопоставления полученных результатов с эмпирическими

значениями. Представленные в статье результаты позволяют заинтересованным читателям выяснить, в какой степени мощность двухразделительных таблиц зависит от численности выборки и силы связи между характеристиками.

Целью статьи является предоставление готовой компьютерной имплементации для обследования мощности критериев двухразделительных таблиц в виде файла в Интернете. Представленные теория и примеры позволяют анализировать мощность критериев с использованием статистики χ^2 Пирсона, а также моделировать ход функции плотности и функции распределения центрального и нецентрального распределения хи-квадрат.