

PIOTR SULEWSKI¹

MOC TESTÓW NIEZALEŻNOŚCI W TABLICY DWUDZIELCZEJ WIĘKSZEJ NIŻ 2×2

1. WPROWADZENIE

Moc testów niezależności to prawdopodobieństwo odrzucenia hipotezy zerowej H_0 , gdy nie jest ona prawdziwa. W każdym popularnym podręczniku ze statystyki znajdują się informacje o m.in. takich testach jak: niezależności, t Studenta, Kołmogorowa czy Behrensa-Fishera. Nie sposób nie zgodzić się zatem z tezą, że testy niezależności – obok wspomnianych wcześniej testów – są zapewne najczęściej stosowanymi narzędziami statystycznym. Dane do testów niezależności aranżuje się w postaci tablic dwudzielczych (TD).

W pracy porównano rodzinę sześciu tzw. „statystyk chi-kwadrat” – w skład której wchodzi powszechnie znana i często stosowana statystyka χ^2 Pearsona – ze statystyką modułową zaproponowaną w Sulewski (2013). Pearson w statystyce χ^2 użył kwadratu dlatego, aby móc skorzystać z rozkładu chi-kwadrat jako miary rozbieżności. Kwadrat użyty w liczniku statystyki χ^2 powoduje również, że duże rozbieżności między liczebnością empiryczną a oczekiwaną są jeszcze większe, a małe rozbieżności jeszcze mniejsze. Innym celem jego zastosowania było uniknięcie wzajemnego niwelowania się rozbieżności. Do tego jednak celu zamiast kwadratu odchyłeń można użyć także ich wartości bezwzględnej, stąd narodził się pomysł na statystykę modułową.

W Sulewski (2014) pokazano, że tzw. „statystyki chi-kwadrat”: χ^2 Pearsona, G^2 ilorazu wiarygodności, N Neymana, KL Kullbacka-Leiblera, FT Freemana-Tukeya oraz CR Cressiego-Reada, mają rozkład chi-kwadrat dla próbek liczących przynajmniej kilkaset elementów. Korzystanie ze „statystyk chi-kwadrat” – szczególnie dla małych próbek – wiąże się ze stosowaniem różnego rodzaju ograniczeń. Np. ze statystyki χ^2 Pearsona w tablicach dwudzielczych większych niż 2×2 można korzystać, gdy wszystkie liczebności oczekiwane są nie mniejsze od 1 oraz gdy nie więcej niż 20% tych wartości jest mniejsze niż 5 (Yates i inni, 1999; Shier, 2004). Zdaniem Cochran’a (1952) statystykę χ^2 Pearsona dla tablic większych niż 2×2 można stosować, gdy przynajmniej jedna z liczebności oczekiwanych jest większa niż 5. W celu zniesienia powyższych ograniczeń w niniejszej pracy zaproponowano wyznaczanie wartości krytycznych na drodze symulacji komputerowych metodą Monte Carlo. Wyznaczanie

¹ Akademia Pomorska w Słupsku, Instytut Matematyki, ul. Arciszewskiego 22, 76-200 Słupsk, Polska, e-mail: piotr.sulewski@apsl.edu.pl.

wartości krytycznych na podstawie np. 100.000 wartości statystyki testowej – w dobie wydajnych komputerów z procesorami wielordzeniowymi – nie stanowi problemu, bo trwa kilkadziesiąt sekund.

Wymienione powyżej testy porównano ze względu na ich moc – czyli zdolności TD $w \times k$ do odrzucenia H_0 mówiącej o tym, że nie ma związku między cechami X i Y . Do wyznaczenia mocy testów tablice dwudzielcze $w \times k$ generowano za pomocą metody słupkowej oraz zaproponowano miarę nieprawdliwości H_0 przyjmującą różne wartości dla danego schematu wyznaczania prawdopodobieństw p_{ij} ($i, j = 1, 2$), dla których H_0 jest niesłuszna.

Niniejsza praca jest kontynuacją rozważań podjętych w Sulewski (2016), gdzie omawianą tematykę zastosowano dla TD 2×2 . Pierwszym celem pracy jest przedstawienie teorii dotyczącej testów niezależności dla TD $w \times k$, drugim celem – wprowadzenie miary nieprawdliwości H_0 oraz porównanie jakości testów za pomocą ich mocy. Praca składa się z dwóch części. W części I realizowany jest cel pierwszy, zdefiniowano w niej TD $w \times k$ oraz siedem testów niezależności, opisano sposób generowania TD i wyznaczania wartości krytycznych. W części II realizowany jest cel drugi, zdefiniowano w niej miarę nieprawdliwości H_0 oraz wyznaczono moc analizowanych testów niezależności w celu wyłonienia najlepszego testu.

2. TABLICA DWUDZIELCZA $w \times k$

Tabela 1 przedstawia tablicę dwudzielczą $w \times k$, która składa się z wartości n_{ij} ($i = 1, 2, \dots, w; j = 1, 2, \dots, k$) rozkładu łącznego cech X i Y takich, że $\sum_{i=1}^w \sum_{j=1}^k n_{ij} = n$.

Tabela 1.

Tablica dwudzielcza $w \times k$ liczebności

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

Źródło: opracowanie własne.

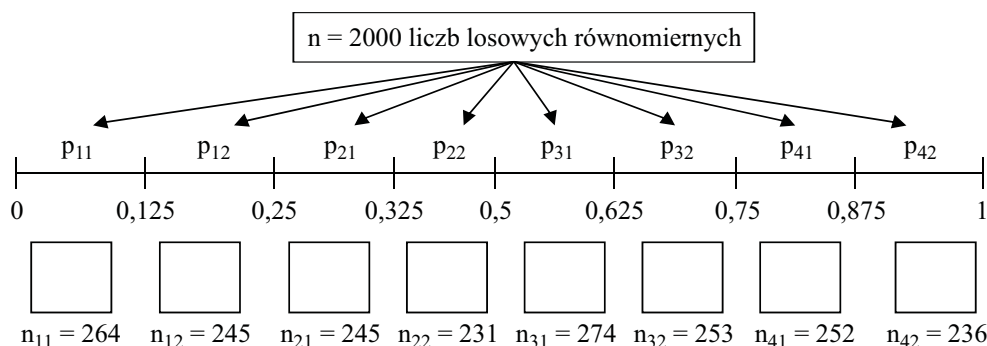
Liczebności oczekiwane i -tego wiersza i j -tej kolumny wyznacza się ze wzoru

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

Tablicę dwudzielczą $w \times k$ można także przedstawić wykorzystując prawdopodobieństwa p_{ij} ($i = 1, 2, \dots, w; j = 1, 2, \dots, k$) takie, że $\sum_{i=1}^w \sum_{j=1}^k p_{ij} = 1$. W odniesieniu do tabeli 1 wartości te wyznaczone są ze wzoru $p_{ij} = n_{ij} / n$ ($i = 1, 2, \dots, w; j = 1, 2, \dots, k$).

Generację TD $w \times k$ niezbędną do przeprowadzenia symulacji i wyznaczenia mocy testów przeprowadzono metodą słupkową. W tym celu 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 oczywiście warunek normalizacji $\sum_{i=1}^w \sum_{j=1}^k p_{ij} = 1$ i dobrano je w taki sposób, aby uzyskać żądaną wartość miary nieprawdźliwości H_0 .

Każda z n wygenerowanych liczb losowych równomiernych „wpada” do jednego z podprzedziałów i tym samym zostaje o jedną zwiększona liczba obiektów w odpowiadającej temu podprzedziałowi komórce TD. Wielkości n_{ij} spełniające równość $\sum_{i=1}^w \sum_{j=1}^k n_{ij} = n$ są liczebnością obiektów w poszczególnych komórkach TD. Rysunek 1 przedstawia schemat wypełniania komórek TD 4×2 dla liczebności próby $n = 2000$ i miary nieprawdźliwości H_0 równej 0, czyli $p_{ij} = 0,125$ dla każdego $i = 1, 2, \dots, 4; j = 1, 2$. Tabela 2 prezentuje odpowiadającą temu schematowi tablicę dwudzielczą.



Rysunek 1. Schemat wypełniania komórek TD 4×2

Źródło: opracowanie własne.

Tabela 2.

Tablica dwudzielcza $w \times k$ otrzymana metoda słupkową

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: opracowanie własne.

3. TESTY NIEZALEŻNOŚCI W TABLICY DWUDZIELCZEJ $w \times k$

Do badania niezależności cech X i Y w TD $w \times k$ skorzystano ze statystyki modułowej będącej modyfikacją statystyki χ^2 Pearsona (Sulewski, 2013):

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

oraz z rodziny statystyk, które dla TD $w \times k$ mają asymptotyczny rozkład chi-kwadrat z $(w-1)(k-1)$ stopniami swobody i dlatego są określane mianem „statystyk chi-kwadrat” (Cressie, Read, 1984). Są to:

- 1) Powszechnie znana i często stosowana statystyka χ^2 Pearsona (Pearson, 1900)

$$\chi^2 = \sum_{i=1}^w \sum_{j=1}^k \frac{(n_{ij} - e_{ij})^2}{e_{ij}}; \quad (3)$$

- 2) Statystyka G^2 ilorazu wiarygodności (Sokal, Rohlf, 2012)

$$G^2 = 2 \sum_{i=1}^w \sum_{j=1}^k n_{ij} \ln \left(\frac{n_{ij}}{e_{ij}} \right); \quad (4)$$

- 3) Statystyka N Neymana (Neyman, 1949)

$$N = \sum_{i=1}^w \sum_{j=1}^k \frac{(n_{ij} - e_{ij})^2}{n_{ij}}; \quad (5)$$

- 4) Statystyka KL Kullbacka-Leiblera (Kullback, 1959)

$$KL = 2 \sum_{i=1}^w \sum_{j=1}^k e_{ij} \ln \left(\frac{e_{ij}}{n_{ij}} \right); \quad (6)$$

- 5) Statystyka FT Freemana-Tukeya (Freeman, Tukey, 1950)

$$FT = 4 \sum_{i=1}^w \sum_{j=1}^k \left(\sqrt{n_{ij}} - \sqrt{e_{ij}} \right)^2; \quad (7)$$

- 6) Statystyka CR Cressiego-Reada (Cressie, Read, 1984)

$$CR = \frac{9}{5} \sum_{i=1}^w \sum_{j=1}^k n_{ij} \left[\left(\frac{n_{ij}}{e_{ij}} \right)^{2/3} - 1 \right]. \quad (8)$$

Charakterystykę statystyk „chi-kwadrat” dla tablic dwudzielczych, trójdzielczych i czterodzielczych wraz z implementacją komputerową można znaleźć w pracy Sulewskiego (2014).

W Sulewski, Motyka (2015a), korzystając z metody Monte Carlo, porównano jakość „statystyk chi-kwadrat” wyrażoną w funkcji mocy przy założeniu, że statystyki te – jaką twierdzą ich autorzy – podlegają rozkładowi chi-kwadrat. Badania wykazały, że jakość testu χ^2 Pearsona oraz CR Cressiego-Reada jest porównywalna i znacznie lepsza od pozostałych.

Liczebności n_{ij} ($i = 1, 2, \dots, w; j = 1, 2, \dots, k$) w TD $w \times k$ są zmiennymi losowymi. Godnym uwagi jest, że zmienne losowe będące odwrotnością lub ilorazem zmiennych losowych nie posiadają wartości oczekiwanej i w konsekwencji także niektórych momentów wyższych rzędów. Rozkład chi-kwadrat – jak powszechnie wiadomo – posiada wszystkie momenty.

Ponadto pożądanym jest, aby poszczególne składniki statystyki testowej – gdy H_0 o niezależności cech X i Y jest słuszna – były tego samego znaku i jak najbliższe 0. Składniki statystyk (4), (6) i (8) są różnego znaku i wzajemnie się rekompensują i tylko dla bardzo dużych prób podlegają rozkładowi chi-kwadrat (Sulewski, 2014). Aby to pokazać metodą słupkową opisaną powyżej wygenerowano tablicę dwudzielczą 3×3 o liczebności próby $n = 150$, gdy hipoteza H_0 o niezależności cech X i Y jest słuszna, czyli $p_{ij} = 1/9$ ($i, j = 1, 2, 3$). Tabela 3 przedstawia wartości poszczególnych składników „statystyk chi-kwadrat” (wartości ujemne pogrubiono).

Tabela 3.

Wartości składników „statystyk chi-kwadrat”

Składnik	χ^2	G^2	N	KL	FT	CR
$i=1, j=1$	0,029	1,362	0,028	-1,305	0,007	0,829
$i=1, j=2$	0,003	0,429	0,003	-0,424	0,001	0,259
$i=1, j=3$	0,060	-1,698	0,065	1,823	0,016	-0,995
$i=2, j=1$	0,381	-4,933	0,444	5,755	0,103	-2,813
$i=2, j=2$	0,616	-6,652	0,741	8,003	0,169	-3,755
$i=2, j=3$	2,547	14,901	1,816	-10,620	0,534	10,030
$i=3, j=1$	0,250	4,240	0,222	-3,769	0,059	2,647
$i=3, j=2$	0,638	7,481	0,538	-6,312	0,146	4,753
$i=3, j=3$	2,202	-8,301	3,699	13,945	0,702	-4,210
Razem	6,726	6,829	7,556	7,095	1,735	6,744

Źródło: opracowanie własne.

4. WYZNACZANIE WARTOŚCI KRYTYCZNYCH

W okresie ostatnich kilkudziesięciu lat zaproponowano różne testy niezależności dla tablic dwudzielczych oraz określono warunki ich stosowalności szczególnie dla małych prób, które najczęściej są przedmiotem badań statystycznych. W dobie coraz to szybszych komputerów z procesorami wielordzeniowymi oraz rozbudowaną pamięcią operacyjną, można za pomocą stosownego oprogramowania znieść te ograniczenia i drogą symulacyjną wyznaczyć wartości krytyczne.

Przy wyznaczaniu wartości krytycznych, gdy między cechami nie ma związku, zawartość tablic dwudzielczych $w \times k$ o liczebności próby n generowano za pomocą metody słupkowej przyjmując $p_{ij} = 1/(w \cdot k)$ ($i = 1, 2, \dots, w; j = 1, 2, \dots, k$). Dla każdej TD $w \times k$ obliczono $r = 10^5$ wartości każdej analizowanej statystyki testowej ϖ , które następnie uporządkowano w kolejności rosnącej. Wartość krytyczną na poziomie istotności $\alpha = 0,1$ wyznaczono ze wzoru:

$$CV_{\alpha} = \varpi_{(90000)}. \quad (9)$$

Tak duża liczba powtórzeń r przy wyznaczaniu wartości statystyki testowej zapewnia uzyskanie dokładnego wyniku.

5. MOC TESTU

Moc testu to prawdopodobieństwo odrzucenia hipotezy zerowej H_0 , gdy nie jest ona prawdziwa. Moc testu zależy od liczebności próby – im liczniejsza próba, tym większa moc. Moc testu zależy także od poziomu istotności testu – im niższy poziom istotności, tym mniejsza moc testu. Moc testu zależy również od siły związku między cechami – im większa siła związku, tym większa moc testu. Więcej informacji o mocy testu znaleźć można m.in. w Sulewski (2015b).

W celu wyznaczenia mocy testu – czyli zdolności TD do odrzucenia hipotezy mówiącej o tym, że nie ma związku między cechami X i Y – gdy w istocie związek jest, niezbędna jest generacja tablic dwudzielczych. Dane podlegające opracowaniu muszą być danymi pochodzącymi z generatora liczb losowych, a nie danymi wziętymi z praktyki. Uwzględniając narzuconą siłę związku między cechami wyrażoną za pomocą miary nieprawdziwości H_0 danej wzorem

$$mn_{w \times k} = \sum_{i=1}^w \sum_{j=1}^k |p_{ij} - p_{i \cdot} \cdot p_{\cdot j}|, \quad (10)$$

do wypełnienia TD $w \times k$ skorzystano z metody „słupkowej” wykorzystującej prawdopodobieństwa p_{ij} ($i = 1, 2, \dots, w; j = 1, 2, \dots, k$).

Znając wartość statystyki testowej ϖ dla różnych liczebności próby n wyznaczono moc testu za pomocą wzoru $m = u/r$, gdzie u określa liczbę tych spośród $r = 10^5$

wszystkich możliwych przypadków, kiedy to wartość statystyki testowej ϖ jest nie mniejsza od wartości krytycznej cv_α .

Hipoteza zerowa H_0 – mówiąca o tym, że między cechami X i Y w TD $w \times k$ nie ma związku – jest słuszna, gdy $p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$ dla $i = 1, 2, \dots, w; j = 1, 2, \dots, k$. Zatem miara (10) przyjmuje wartość 0, gdy hipoteza H_0 jest słuszna. Im większe wartości mn , tym większa jest możliwość fałszywości H_0 .

Do wyznaczenia mocy testu dla TD $w \times k$ skorzystano z różnych schematów wyznaczania prawdopodobieństw p_{ij} ($i = 1, 2, \dots, w; j = 1, 2, \dots, k$), dla których hipoteza zerowa H_0 jest niesłuszna. Prawdopodobieństwa te – jak również przedziały zmienności miary mn dla omawianych schematów – przedstawiono w tabelach 5, 7, 9, 11, 13. Wynika z nich, że miara nieprawdziwości hipotezy H_0 zmienia się w różnych przedziałach. Minimalna liczebność próby dla danej TD i dla danego schematu została tak dobrana, aby liczebności oczekiwane były różne od zera. Maksymalną wartość próby ustalono tak, aby uzyskać maksymalną moc testu.

Dla statystyki modułowej $|\chi|$ oraz dla „statystyki chi-kwadrat” o największej mocy, uzyskane wyniki porównano za pomocą testu o równości prawdopodobieństw stawiając hipotezę zerową, że moce testów są równe, czyli różnice są statystycznie nieistotne.

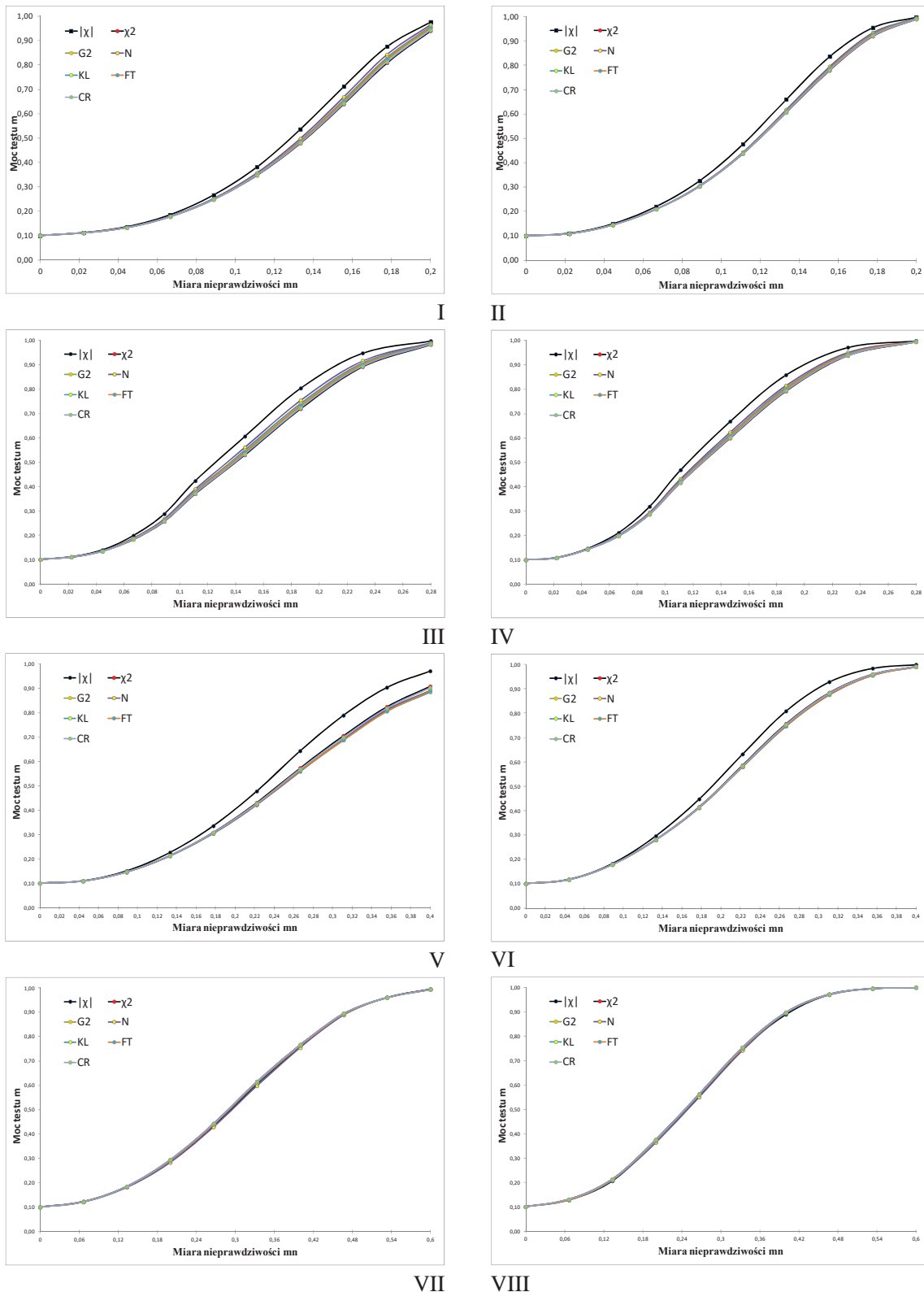
TABLICA DWUDZIELCZA 2×3

Tabela 4.

Schematy prawdopodobieństw ukazujące związek między cechami dla TD 2×3

Schemat A			
	Y_1	Y_2	Y_3
X_1	$p_o - k \Delta p$	p_o	p_o
X_2	p_o	p_o	$p_o + k \Delta p$
Schemat B			
	Y_1	Y_2	Y_3
X_1	$p_o - k \Delta p$	$p_o - k \Delta p$	p_o
X_2	p_o	$p_o + k \Delta p$	$p_o + k \Delta p$
Schemat C			
	Y_1	Y_2	Y_3
X_1	$p_o - k \Delta p$	$p_o - k \Delta p$	p_o
X_2	$p_o + k \Delta p$	$p_o + k \Delta p$	p_o
Schemat D			
	Y_1	Y_2	Y_3
X_1	$p_o - k \Delta p$	p_o	$p_o + k \Delta p$
X_2	$p_o + k \Delta p$	p_o	$p_o - k \Delta p$

Źródło: opracowanie własne.



Rysunek 2. Moc testów niezależności dla TD 2×3 oraz liczebności próby n

I. Schemat A, $n = 125$

II. Schemat A, $n = 150$

III. Schemat B, $n = 125$

IV. Schemat B, $n = 150$

V. Schemat C, $n = 50$

VI. Schemat C, $n = 75$

VII. Schemat D, $n = 30$

VIII. Schemat D, $n = 40$

Źródło: opracowanie własne.

Tabela 5.

Miara nieprawdziwości H_0 dla TD 2×3 i $k = 0, 1, \dots, 9$

Schemat A	Schemat B	Schemat C	Schemat D
$mn_{2 \times 3} = k / 45$	$mn_{2 \times 3} = \begin{cases} k / 45 & k = 0, \dots, 5 \\ k^2 / 450 + k / 90 & k = 6, \dots, 9 \end{cases}$	$mn_{2 \times 3} = 2k / 45$	$mn_{2 \times 3} = k / 15$
$mn_{2 \times 3} \in \langle 0; 0,2 \rangle$	$mn_{2 \times 3} \in \langle 0; 0,28 \rangle$	$mn_{2 \times 3} \in \langle 0; 0,4 \rangle$	$mn_{2 \times 3} \in \langle 0; 0,6 \rangle$

Źródło: opracowanie własne.

Rysunek 2 przedstawia zależność mocy testów niezależności od miary mn dla TD 2×3 , na poziomie istotności $\alpha = 0,1$ i liczebności próby n .

Z rysunku 2 wynika, że „statystyki chi-kwadrat” dla TD 2×3 mają podobną moc, a występujące między nimi różnice są statystycznie nieistotne. Dla schematu A różnice między statystyką $|\chi|$, a pozostałymi analizowanymi statystykami są statystycznie istotne dla $mn \in \langle 0,133; 0,178 \rangle$. Dla schematu B oraz $n = 125$, $n = 150$ różnice te są statystycznie istotne odpowiednio dla $mn \in \langle 0,147; 0,280 \rangle$ i $mn \in \langle 0,111; 0,231 \rangle$. Dla schematu C statystyka $|\chi|$ ma istotnie większą moc od pozostałych statystyk, gdy $mn \in \langle 0,222; 0,4 \rangle$. Dla schematu D różnice między statystykami są statystycznie nieistotne.

TABLICA DWUDZIELCZA 2×4

Tabela 6.

Schematy prawdopodobieństw ukazujące związek między cechami dla TD 2×4

Schemat A				
	Y_1	Y_2	Y_3	Y_4
X_1	$p_o - k \Delta p$	p_o	p_o	p_o
X_2	p_o	p_o	p_o	$p_o + k \Delta p$
Schemat B				
	Y_1	Y_2	Y_3	Y_4
X_1	$p_o - k \Delta p$	$p_o - k \Delta p$	p_o	p_o
X_2	p_o	p_o	$p_o + k \Delta p$	$p_o + k \Delta p$
Schemat C				
	Y_1	Y_2	Y_3	Y_4
X_1	$p_o - k \Delta p$	$p_o - k \Delta p$	p_o	p_o
X_2	$p_o + k \Delta p$	$p_o + k \Delta p$	p_o	p_o

Tabela 6. (cd.)

Schemat D				
	Y_1	Y_2	Y_3	Y_4
X_1	$p_o - k \Delta p$	$p_o - k \Delta p$	$p_o + k \Delta p$	$p_o + k \Delta p$
X_2	$p_o + k \Delta p$	$p_o + k \Delta p$	$p_o - k \Delta p$	$p_o - k \Delta p$

Źródło: opracowanie własne.

Tabela 7.

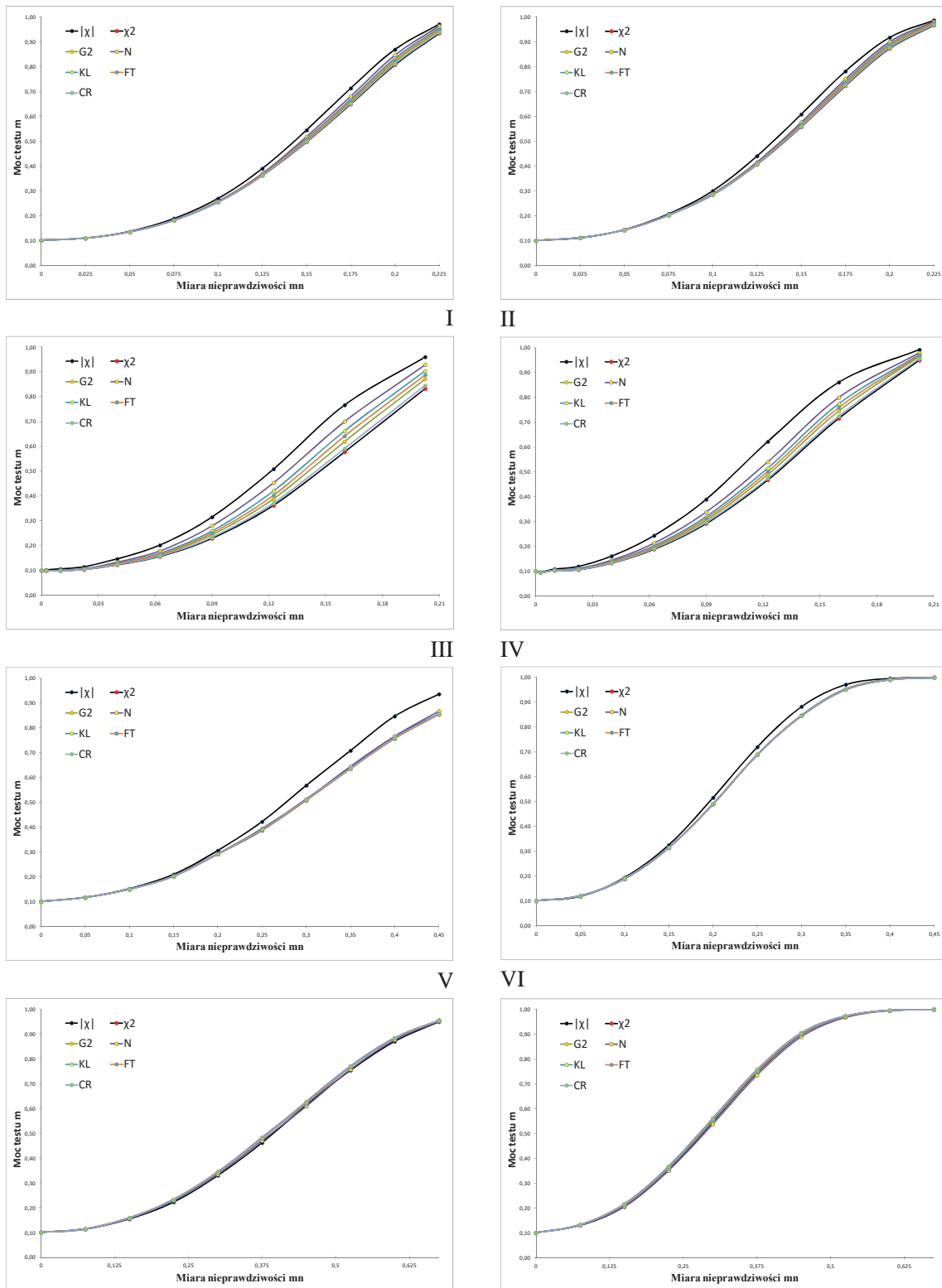
Miara nieprawdziwości H_0 dla TD 2×4 i $k = 0, 1, \dots, 9$

Schemat A	Schemat B	Schemat C	Schemat D
$mn_{2 \times 4} = k / 40$	$mn_{2 \times 4} = k^2 / 400$	$mn_{2 \times 4} = k / 20$	$mn_{2 \times 4} = 3k / 40$
$mn_{2 \times 4} \in \langle 0; 0,225 \rangle$	$mn_{2 \times 4} \in \langle 0; 0,2025 \rangle$	$mn_{2 \times 4} \in \langle 0; 0,45 \rangle$	$mn_{2 \times 4} \in \langle 0; 0,675 \rangle$

Źródło: opracowanie własne.

Rysunek 3 przedstawia zależność mocy testów niezależności od miary mn dla TD 2×4 , na poziomie istotności $\alpha = 0,1$ i liczebności próby n .

Z rysunku 3 wynika, że dla schematów A, C, D „statystyki chi-kwadrat” dla TD 2×4 mają podobną moc, a występujące między nimi różnice są statystycznie nieistotne. Dla schematu A statystyka $|\chi|$ charakteryzuje się większą mocą niż pozostałe statystyki, jednak przewaga ta jest statystycznie istotna dla $n = 175$ i $mn = 0,175$. Podobna sytuacja jest dla schematu B, gdy $mn \in \langle 0,09; 0,203 \rangle$. Dla schematu C różnica istotna statystycznie na korzyść statystyki modułowej jest dla $n = 50$ i $mn \in \langle 0,3; 0,45 \rangle$ oraz $n = 100$ i $mn \in \langle 0,3; 0,35 \rangle$. Dla schematu D różnice między statystykami są nieistotne statystycznie.



Rysunek 3. Moc testów niezależności dla TD 2×4 oraz liczebności próby n

I. Schemat A, $n = 150$

V. Schemat C, $n = 50$

II. Schemat A, $n = 175$

VI. Schemat C, $n = 100$

III. Schemat B, $n = 150$

VII. Schemat D, $n = 30$

IV. Schemat B, $n = 200$

VIII. Schemat D, $n = 50$

Źródło: opracowanie własne.

TABLICA DWUDZIELCZA 3×3

Tabela 8.

Schematy prawdopodobieństw ukazujące związek między cechami dla TD 3×3

Schemat A			
	$p_0 = 1/9, \Delta p = p_0/10, k = 0,1,\dots,9$		
	Y_1	Y_2	Y_3
X_1	$p_0 - k \Delta p$	$p_0 - k \Delta p / 2$	p_0
X_2	$p_0 - k \Delta p / 2$	p_0	$p_0 + k \Delta p / 2$
X_3	p_0	$p_0 + k \Delta p / 2$	$p_0 + k \Delta p$
Schemat B			
	Y_1	Y_2	Y_3
X_1	$p_0 - k \Delta p$	$p_0 - k \Delta p$	p_0
X_2	p_0	p_0	p_0
X_3	$p_0 + k \Delta p$	$p_0 + k \Delta p$	p_0
Schemat C			
	Y_1	Y_2	Y_3
X_1	$p_0 - k \Delta p$	p_0	$p_0 + k \Delta p$
X_2	p_0	p_0	p_0
X_3	$p_0 + k \Delta p$	p_0	$p_0 - k \Delta p$

Źródło: opracowanie własne.

Tabela 9.

Miara nieprawdliwości H_0 dla TD 3×3 i $k = 0,1,\dots,9$

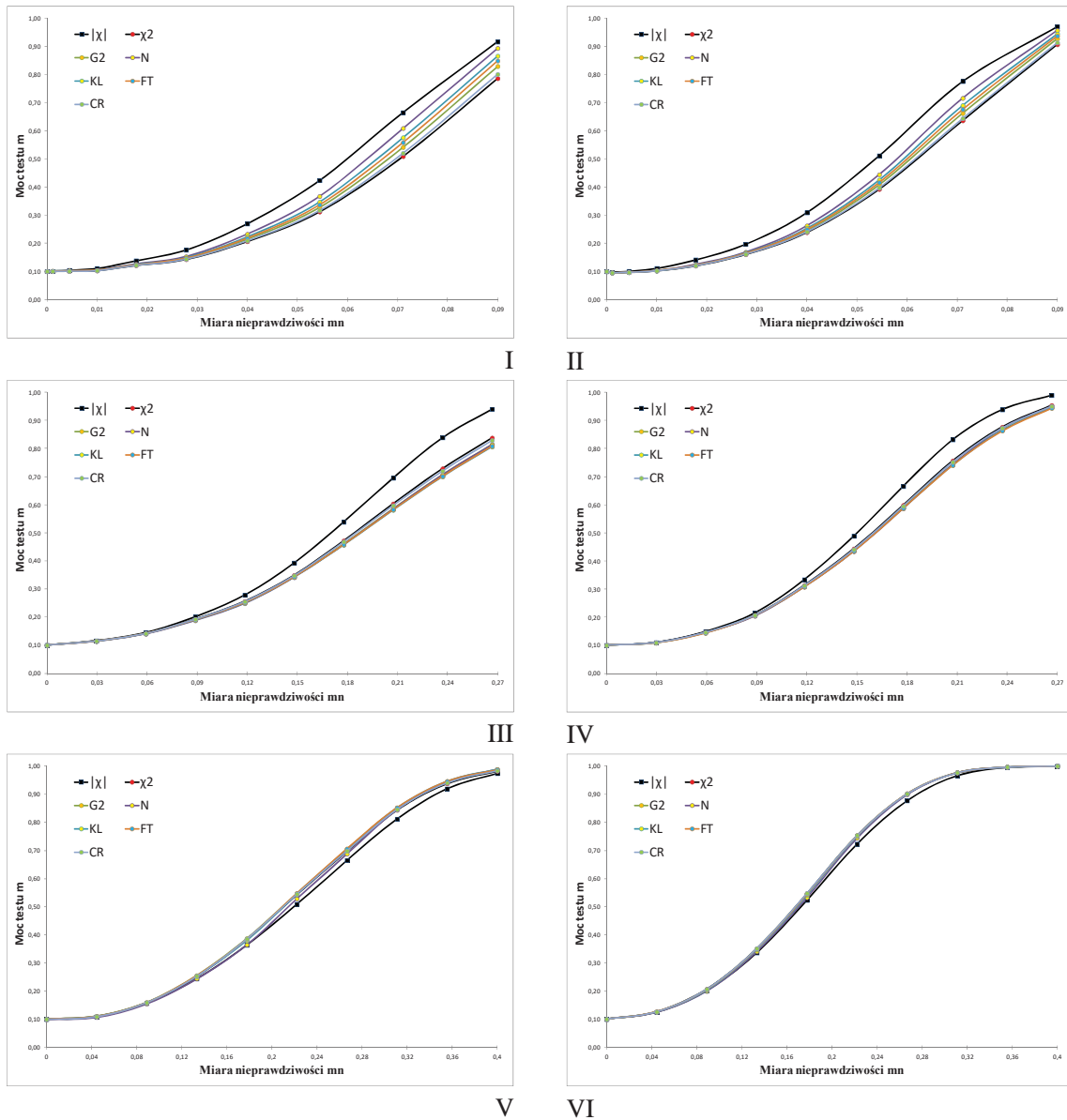
Schemat A	Schemat B	Schemat C
$mn_{3 \times 3} = k^2 / 900$	$mn_{3 \times 3} = 0,02963 \cdot k$	$mn_{3 \times 3} = 2k / 45$
$mn_{3 \times 3} \in \langle 0;0,09 \rangle$	$mn_{3 \times 3} \in \langle 0;0,2667 \rangle$	$mn_{3 \times 3} \in \langle 0;0,04 \rangle$

Źródło: opracowanie własne.

Rysunek 4 przedstawia zależność mocy testów niezależności od miary mn dla TD 3×3 , na poziomie istotności $\alpha = 0,1$ i liczebności próby n .

Z rysunku 4 wynika, że dla schematów B, C „statystyki chi-kwadrat” dla TD 3×3 mają podobną moc, a występujące między nimi różnice są statystycznie nieistotne. Dla schematu A statystyka $|\chi|$ charakteryzuje się większą mocą niż statystyka N, jednak przewaga ta jest statystycznie istotna dla $n = 300$ i $mn \in \langle 0,04;0,09 \rangle$ oraz $n = 400$ i $mn \in \langle 0,028;0,09 \rangle$. Statystyka $|\chi|$ charakteryzuje się także większą mocą

od pozostałych statystyk dla schematu B i przewaga ta jest statystycznie istotna dla $mn \in \langle 0,148; 0,267 \rangle$. Dla schematu C „statystyki chi-kwadrat” mają większą moc od statystyki modułowej, jednak przewaga ta jest statystycznie istotna tylko dla $n = 50$ i $mn = 0,311$.



Rysunek 4. Moc testów niezależności dla TD 3×3 oraz liczebności próby n

I. Schemat A, $n = 300$

IV. Schemat B, $n = 100$

II. Schemat A, $n = 400$

V. Schemat C, $n = 50$

III. Schemat B, $n = 75$

VI. Schemat C, $n = 75$

Źródło: opracowanie własne.

TABLICA DWUDZIELCZA 3×4

Tabela 10.

Schematy prawdopodobieństw ukazujące związek między cechami dla TD 3×4

Schemat A		$p_0 = 1/12, \Delta p = p_0/10, k = 0,1,\dots,9$			
	Y_1	Y_2	Y_3	Y_4	
X_1	$p_0 - k \Delta p$	$p_0 - k \Delta p / 2$	$p_0 - k \Delta p / 3$	p_0	
X_2	$p_0 - k \Delta p / 2$	p_0	p_0	$p_0 + k \Delta p / 2$	
X_3	p_0	$p_0 + k \Delta p / 3$	$p_0 + k \Delta p / 2$	$p_0 + k \Delta p$	
Schemat B					
	Y_1	Y_2	Y_3	Y_4	
X_1	$p_0 - k \Delta p$	$p_0 - k \Delta p$	p_0	p_0	
X_2	p_0	p_0	p_0	p_0	
X_3	$p_0 + k \Delta p$	$p_0 + k \Delta p$	p_0	p_0	
Schemat C					
	Y_1	Y_2	Y_3	Y_4	
X_1	$p_0 - k \Delta p$	$p_0 - k \Delta p / 2$	$p_0 + k \Delta p / 2$	$p_0 + k \Delta p$	
X_2	p_0	p_0	p_0	p_0	
X_3	$p_0 + k \Delta p$	$p_0 + k \Delta p / 2$	$p_0 - k \Delta p / 2$	$p_0 - k \Delta p$	

Źródło: opracowanie własne.

Tabela 11.

Miara nieprawdziwości H_0 dla TD 3×4 i $k = 0,1,\dots,9$

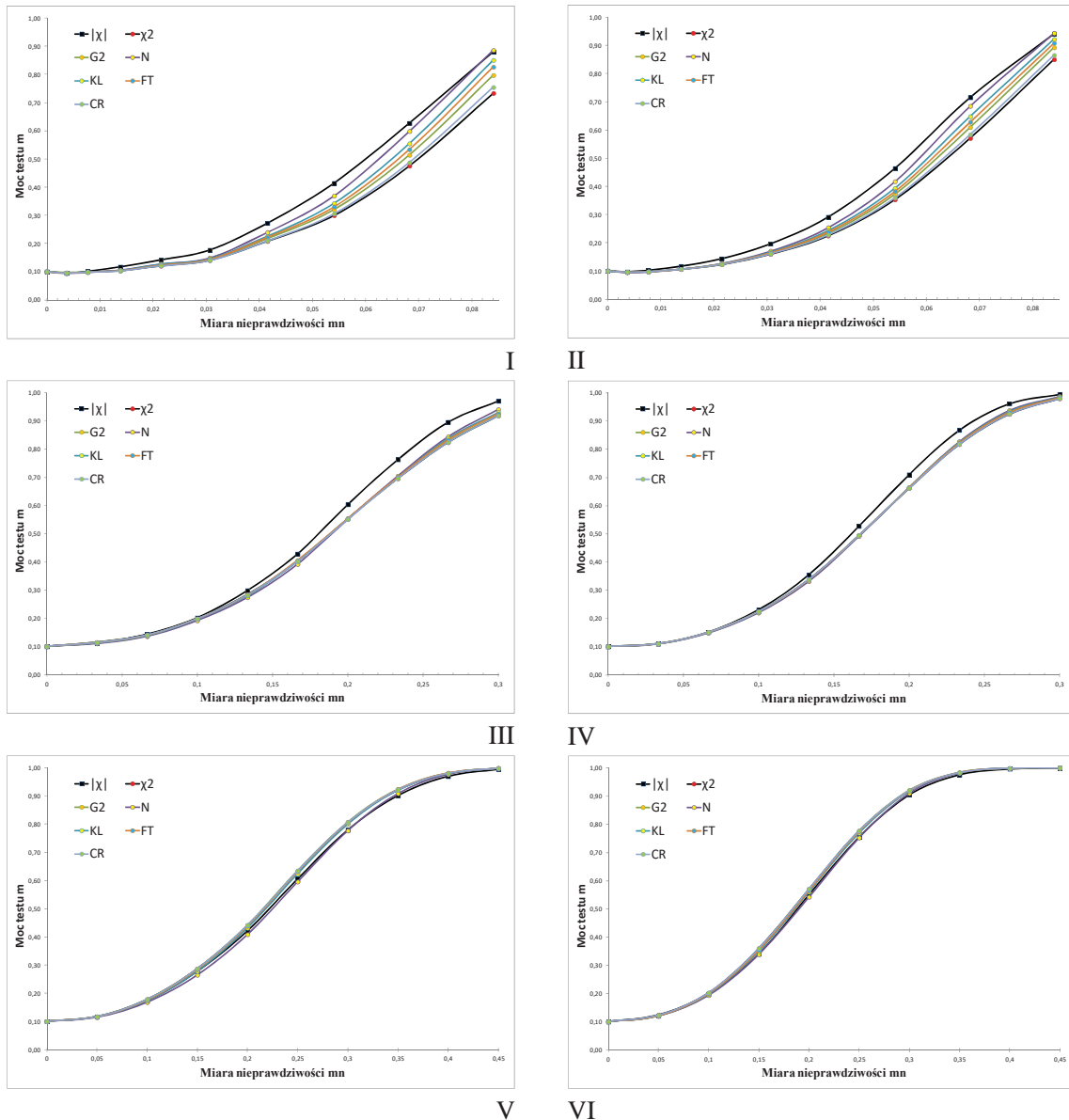
Schemat A	Schemat B	Schemat C
$mn_{3 \times 4} = k^2 / 1250 + k / 500$	$mn_{3 \times 4} = k / 30$	$mn_{3 \times 4} = k / 20$
$mn_{3 \times 4} \in \langle 0; 0,084 \rangle$	$mn_{3 \times 4} \in \langle 0; 0,03 \rangle$	$mn_{3 \times 4} \in \langle 0; 0,45 \rangle$

Źródło: opracowanie własne.

Rysunek 5 przedstawia zależność mocy testów niezależności od miary mn dla TD 3×4 , na poziomie istotności $\alpha = 0,1$ i liczebności próby n .

Z rysunku 5 (schemat A) wynika, że „statystyki chi-kwadrat” charakteryzują się różną mocą. Przy $n = 400$ różnice między statystyką $|\chi|$ a statystyką N jako reprezentanta „statystyk chi-kwadrat” o największej mocy, są statystycznie istotne dla $mn \in \langle 0,031; 0,054 \rangle$, z kolei przy $n = 500$ – dla $mn \in \langle 0,042; 0,054 \rangle$. Dla schematu B moc testu dla „statystyk chi-kwadrat” jest bardzo podobna. Różnice między staty-

styką $|\chi|$ a pozostałymi statystykami są statystycznie istotne dla $mn \in \langle 0,2;0,3 \rangle$ z korzyścią dla statystyki modułowej. Dla schematu C i próby $n = 75$ funkcje mocy testów dla statystyki $|\chi|$ i N są bardzo podobne. Dla $n = 100$ funkcje mocy analizowanych testów mają podobne przebiegi.



Rysunek 5. Moc testów niezależności dla TD 3×4 oraz liczebności próby n

I. Schemat A, $n = 400$

IV. Schemat B, $n = 125$

II. Schemat A, $n = 500$

V. Schemat C, $n = 75$

III. Schemat B, $n = 100$

VI. Schemat C, $n = 100$

Źródło: opracowanie własne.

TABLICA DWUDZIELCZA 4×4

Tabela 12.

Schematy prawdopodobieństw ukazujące związek między cechami dla TD 4×4

Schemat A		$p_0 = 1/16, \Delta p = p_0/10, k = 0,1,\dots,9$			
	Y_1	Y_2	Y_3	Y_4	
X_1	$p_0 - k \Delta p$	$p_0 - k \Delta p / 2$	$p_0 - k \Delta p / 3$	p_0	
X_2	$p_0 - k \Delta p / 2$	$p_0 - k \Delta p / 3$	p_0	$p_0 + k \Delta p / 3$	
X_3	$p_0 - k \Delta p / 3$	p_0	$p_0 + k \Delta p / 3$	$p_0 + k \Delta p / 2$	
X_4	p_0	$p_0 + k \Delta p / 3$	$p_0 + k \Delta p / 2$	$p_0 + k \Delta p$	
Schemat B					
	Y_1	Y_2	Y_3	Y_4	
X_1	$p_0 - k \Delta p$	$p_0 - k \Delta p$	p_0	p_0	
X_2	$p_0 - k \Delta p / 2$	$p_0 - k \Delta p / 2$	p_0	p_0	
X_3	$p_0 + k \Delta p / 2$	$p_0 + k \Delta p / 2$	p_0	p_0	
X_4	$p_0 + k \Delta p$	$p_0 + k \Delta p$	p_0	p_0	
Schemat C					
	Y_1	Y_2	Y_3	Y_4	
X_1	$p_0 - k \Delta p$	$p_0 - k \Delta p / 2$	$p_0 - k \Delta p / 3$	p_0	
X_2	$p_0 - k \Delta p / 2$	$p_0 - k \Delta p / 3$	p_0	$p_0 + k \Delta p / 3$	
X_3	$p_0 - k \Delta p / 3$	p_0	$p_0 + k \Delta p / 3$	$p_0 + k \Delta p / 2$	
X_4	p_0	$p_0 + k \Delta p / 3$	$p_0 + k \Delta p / 2$	$p_0 + k \Delta p$	

Źródło: opracowanie własne.

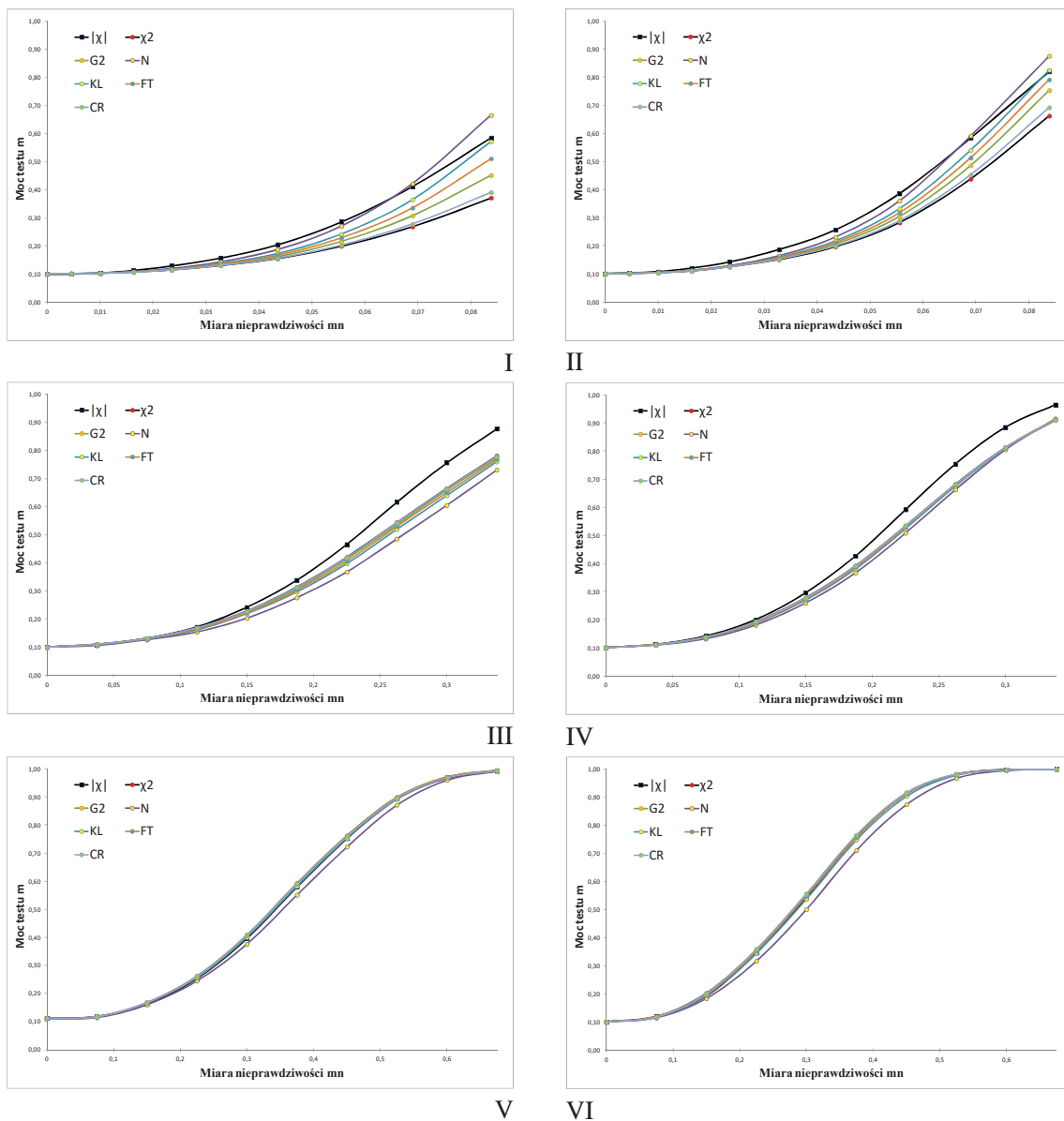
Tabela 13.

Miara nieprawdziwości H_0 dla TD 4×4 i $k = 0,1,\dots,9$

Schemat A	Schemat B	Schemat C
$mn_{4 \times 4} \in \left\{ \begin{array}{l} 0, 0,005, 0,01, 0,016, 0,024, 0,033 \\ 0,044, 0,056, 0,069, 0,084 \end{array} \right\}$	$mn_{4 \times 4} = 3k / 80$	$mn_{4 \times 4} = 3k / 40$
$mn_{4 \times 4} \in \langle 0; 0,084 \rangle$	$mn_{4 \times 4} \in \langle 0; 0,338 \rangle$	$mn_{4 \times 4} \in \langle 0; 0,675 \rangle$

Źródło: opracowanie własne.

Rysunek 6 przedstawia zależność mocy testów niezależności od miary mn dla TD 4×4 , na poziomie istotności $\alpha = 0,1$ i liczebności próby n .

Rysunek 6. Moc testów niezależności dla TD 4×4 oraz liczebności próby n I. Schemat A, $n = 300$ IV. Schemat B, $n = 125$ II. Schemat A, $n = 500$ V. Schemat C, $n = 60$ III. Schemat B, $n = 100$ VI. Schemat C, $n = 80$

Źródło: opracowanie własne.

Z rysunku 6 (schemat A) wynika, że różnice między statystyką $|\chi|$ a statystyką N jako reprezentanta „statystyk chi-kwadrat” o największej mocy, są statystycznie nieistotne z wyjątkiem $mn = 0,084$. Dla schematu B różnice między statystyką $|\chi|$ a pozostałymi statystykami są statystycznie istotne dla $mn > 0,188$ z korzyścią dla statystyki modułowej. Dla schematu C analizowane statystyki mają podobną moc z wyjątkiem statystyki N , która dla $n = 80$ i $mn \in \langle 0,3; 0,525 \rangle$ ma istotnie mniejszą moc niż pozostałe statystyki.

6. PODSUMOWANIE

Kwadrat użyty w liczniku statystyki χ^2 powoduje, że duże rozbieżności między liczebnością empiryczną a teoretyczną są jeszcze większe, a małe rozbieżności jeszcze mniejsze. Innym celem zastosowania kwadratu było uniknięcie wzajemnego niwelowania się rozbieżności. Do tego jednak celu zamiast kwadratu odchyień można użyć także ich wartości bezwzględnej, stąd pomysł na statystykę modułową.

Zaproponowana miara mn dla analizowanych schematów prawdopodobieństw zmienia się w różnych przedziałach. Minimalna liczebność próby dla danej TD i dla danego schematu została tak ustalona, by liczebności oczekiwane były niezerowe. Maksymalną wartość próby dobrano tak, by uzyskać maksymalną moc testu.

Artykuł zwraca uwagę na fakt, że testy niezależności wykorzystujące „statystyki chi-kwadrat” tylko dla schematów prawdopodobieństw o najmniejszym zakresie zmienności miary mn charakteryzują się różną mocą dla danej liczebności próby n i miary mn . Największą moc wśród nich ma statystyka N i ona jest bezpośrednio porównywana ze statystyką modułową. Jeżeli widoczna jest przewaga statystyki modułowej nad „statystykami chi-kwadrat”, to głównie dla nie początkowych wartości miary mn w ramach danego przedziału zmienności tej miary. Dla ostatniego analizowanego w ramach danej TD schematu nieprawdziwości H_0 przedstawiającego największy zakres zmienności miary mn , test wykorzystujący statystykę modułową ma podobną moc jak pozostałe testy niezależności, z wyjątkiem TD 3×3 , liczebności próby $n = 50$ i miary $mn = 0,311$, kiedy to statystyka $|\chi|$ ma istotnie mniejszą moc od pozostałych statystyk.

Reasumując statystyka modułowa $|\chi|$ charakteryzuje się większą mocą niż „statystyki chi-kwadrat”, szczególnie jest to widoczne dla schematów prawdopodobieństw o mniejszych zakresach zmienności miary mn . Dla schematów C lub D o największym zakresie zmienności miary mn statystyka modułowa nie odstępuje pod względem mocy od pozostałych analizowanych statystyk. Na korzyść statystyki modułowej przemawia także fakt, że można z niej skorzystać w sytuacji, gdy jedna komórka lub dwie leżące na przekątnej są puste. Statystyk G^2 , N i KL nie można stosować w tej sytuacji.

LITERATURA

- Cochran W. G., (1954), Some Methods for Strengthening the Common χ^2 Tests, *Biometrics*, 10, 417–451.
- Cressie N., Read T., (1984), Multinomial Goodness-of-Fit Tests, *Journal of the Royal Statistical Society, Series B (Methodological)*, 46 (3), 440–464.
- Freeman M. F., Tukey J. W., (1950), Transformations Related to the Angular and the Square root, *Annals of Mathematical Statistics*, 21, 607–611.
- Kullback S., (1959), *Information Theory and Statistics*, Wiley, New York.
- Neyman J., (1949), Contribution to the Theory of the χ^2 Test, *Proceedings of the (First) Berkeley Symposium on Mathematical Statistics and Probability*, University of California Press, 239–273.

- 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 (50), 157–172.
- Shier R., (2004), The Chi-squared Test for Two-Way Tables, *Mathematics Learning Support Centre*.
- Sokal R. R., Rohlf F. J., (2012), *Biometry: The Principles and Practice of Statistics in Biological Research*, Freeman, New York.
- Sulewski P., (2013), Modyfikacja testu niezależności, *Wiadomości Statystyczne*, GUS, 10, 1–19.
- Sulewski P., (2014), *Statystyczne badanie współzależności cech typu dyskretne kategorie*, Akademia Pomorska, Słupsk.
- Sulewski P., Motyka R., (2015a), Independence Test. A Comparative Analysis of Its Six Variants, *ZN AMW*, Gdynia, LVI,1, 37–46.
- Sulewski P., (2015b), Ocena zdolności tablic dwudzielczych do wykrywania związku między uporządkowanymi cechami typu jakościowego, *Wiadomości Statystyczne*, GUS, 5, 1–16.
- Sulewski P., (2016), Moc testów niezależności w tablicy dwudzielczej 2×2 , *Wiadomości Statystyczne*, GUS, 8.
- Yates D., Moore D., McCabe G., (1999), *The Practice of Statistics*, 1st Ed., New York, W. H. Freeman.

MOC TESTÓW NIEZALEŻNOŚCI W TABLICY DWUDZIELCZEJ WIĘKSZEJ NIŻ 2×2

Streszczenie

W literaturze statystycznej istnieje wiele miar testowych do badania niezależności cech w tablicach dwudzielczych. Do analiz statystycznych wybrano rodzinę sześciu tzw. „statystyk chi-kwadrat” – w tym statystykę χ^2 Pearsona – oraz propozycję autora w postaci statystyki modułowej. W celu uwolnienia się od ograniczeń stosowalności „statystyk chi-kwadrat”, wartości krytyczne dla wszystkich analizowanych statystyk wyznaczono symulacyjnie metodami Monte Carlo. W celu porównania testów zaproponowano miarę nieprawdliwości H_0 oraz wyznaczono moc testów, czyli zdolność tablicy dwudzielczej $w \times k$ do odrzucenia H_0 mówiącej o tym, że między cechami X i Y nie ma związku.

Słowa kluczowe: tablica dwudzielcza, test niezależności, wartości krytyczne, Monte Carlo

POWER ANALYSIS OF INDEPENDENCE TESTING FOR TWO-WAY CONTINGENCY TABLES BIGGER THAN 2×2

Abstract

In the statistical literature there are many test measures to study the independence features in the two-way contingency tables. For statistical analysis, the family of six so-called “chi-squared statistic” was selected – including Pearson’s χ^2 statistics – and the proposal of the author in the form of modular statistics. In order to free themselves from the limitations of the applicability of the “chi-squared statistic”, critical values for all analyzed statistics were determined by simulation methods of Monte Carlo. In order to compare the tests, the measure of untruthfulness of H_0 was proposed and calculated the power of the tests which is the ability of two-way contingency tables to reject null hypothesis which says that between features X and Y there is no relation.

Keywords: two-way contingency tables, independence test, critical values, Monte Carlo study

