

PIOTR SULEWSKI¹

MOC TESTÓW NIEZALEŻNOŚCI W TABLICY TRÓJDZIELCZEJ $2 \times 2 \times 2$

1. WPROWADZENIE

Moc testów to prawdopodobieństwo odrzucenia hipotezy zerowej, gdy nie jest ona prawdziwa. Testy niezależności należą do najczęściej stosowanych narzędzi statystycznych, a dane do nich aranżuje się w postaci tablic wielodzielczych, a w szczególności tablic trójdzielczych (TT) $2 \times 2 \times 2$. Wydawać by się mogło, że o metodach analizy tablic wielodzielczych powiedziano już wszystko. Jednak w dobie szeroko rozwiniętej informatyzacji i możliwości przeprowadzania symulacji komputerowych, na wiele zagadnień można spojrzeć „nowym okiem”.

W (Cressie, Read, 1984) zaproponowano dla tablic dwudzielczych (TD) $w \times k$ rodzinę 6 statystyk, które mają asymptotyczny rozkład chi-kwadrat z $(w - 1)(k - 1)$ stopniami swobody. Do tzw. „statystyk chi-kwadrat” należą następujące statystyki: χ^2 Pearsona (Pearson, 1900), G^2 ilorazu wiarygodności (Sokal, Rohlf, 2012), N Neymana (Neyman, 1949), KL Kullbacka-Leiblera (Kullback, 1959), FT Freemana-Tukeya (Freeman, Tukey, 1950) oraz CR Cressiego-Reada (Cressie, Read, 1984). W pracy Sulewskiego (2014) pokazano, że „statystyki chi-kwadrat” są wystarczająco bliskie rozkładowi chi-kwadrat dla próbek liczących przynajmniej kilkaset elementów. Aby można było skorzystać z próbek mniej licznych wyznaczono wartości krytyczne na drodze symulacji komputerowych metodą Monte Carlo. Wyznaczanie wartości krytycznych na podstawie np. 100.000 wartości statystyki testowej – w dobie wydajnych komputerów z procesorami wielordzeniowymi – nie stanowi problemu.

W pracy Sulewski (2013) zaproponowano statystykę modułową będącą modyfikacją statystyki χ^2 Pearsona dla TD $w \times k$. Statystykę modułową porównano pod względem mocy testu ze „statystykami chi-kwadrat” dla TD 2×2 (Sulewski, 2016a) oraz dla wybranych rozmiarów TD większych niż 2×2 (Sulewski, 2016b).

Celem pracy jest przedstawienie teorii dotyczącej autorskiej statystyki modułowej mierzącej niezależność zmiennych dla TT $2 \times 2 \times 2$ i zbadanie jej własności w odniesieniu do znanych „statystyk chi-kwadrat”, opisanie procedury generowania zawartości tych tablic metodą słupkową, wprowadzenie nowej miary nieprawdziwości H_0 .

Praca składa się z siedmiu punktów. W punkcie drugim zdefiniowano TT $2 \times 2 \times 2$ oraz opisano metodę słupkową generacji zawartości tych tablic. W punkcie trzecim

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

rozszerzono postać „statystyk chi-kwadrat” oraz statystyki modułowej na TT $2 \times 2 \times 2$. W punkcie czwartym opisano sposób wyznaczania wartości krytycznych. W punkcie piątym zaproponowano miarę nieprawdopodobieństwa hipotezy H_0 zdefiniowaną dla różnych schematów wyznaczania prawdopodobieństw p_{ijt} ($i, j, t = 1, 2$), dla których H_0 jest niesłuszna. Wartości proponowanej miary siły związku porównano z wartościami współczynników: Graya-Williamsa (Gray, Williams, 1975), Marcotorchino (Marcotorchino, 1984), Lombardo (Lombardo, 2011). W punkcie szóstym testy niezależności wykorzystujące statystykę modułową oraz „statystyki chi-kwadrat” porównano ze względu na ich moc – czyli zdolność testu do odrzucenia hipotezy H_0 mówiącej o tym, że nie ma związku między cechami X, Y, Z w sytuacji, gdy w rzeczywistości jest ona fałszywa. Punkt ostatni poświęcono podsumowaniu.

2. TABLICA TRÓJDZIELCZA $2 \times 2 \times 2$

Tablicę, która gromadzi wynik podziału próby według trzech cech X, Y, Z nazywamy tablicą trójdzielczą (TT). Można sobie ją wyobrazić jako kostkę z „w” wierszami, „k” kolumnami i „p” płaszczyznami. TT można także zilustrować rozkładając płaszczyzny p obok siebie. TT ułatwia odczytywanie zależności między cechami tak ilościowymi, jak i jakościowymi. Najprostszą jej postacią jest tablica $2 \times 2 \times 2$ (tabela 1), która składa się z 8 liczebności n_{ijt} ($i, j, t = 1, 2$) rozkładu łącznego cech X, Y, Z .

Tabela 1.

Tablica trójdzielcza $2 \times 2 \times 2$

Z	Z ₁		Z ₂		Razem
	Y ₁	Y ₂	Y ₁	Y ₂	
X ₁	n_{111}	n_{121}	n_{112}	n_{122}	$n_{1..}$
X ₂	n_{211}	n_{221}	n_{212}	n_{222}	$n_{2..}$
Razem	$n_{.11}$	$n_{.21}$	$n_{.12}$	$n_{.22}$	n

Źródło: opracowanie własne.

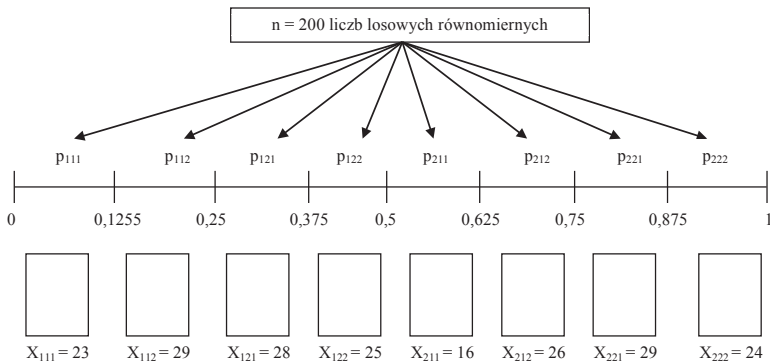
Liczebności oczekiwane dla TT $2 \times 2 \times 2$ – przy założeniu niezależności badanych zmiennych – wyznacza się ze wzoru

$$e_{ijt} = \frac{n_{i..} \cdot n_{.j.} \cdot n_{...t}}{n^2} \quad (i, j, t = 1, 2). \quad (1)$$

TT $2 \times 2 \times 2$ można także przedstawić za pomocą prawdopodobieństw p_{ijt} ($i, j, t = 1, 2$) takich, że $\sum_{i=1}^2 \sum_{j=1}^2 \sum_{t=1}^2 p_{ijt} = 1$. W odniesieniu do tabeli 1 wartości te wyznaczone są

ze wzoru $p^* = n^* / n$. Łatwo udowodnić, że sumy brzegowe liczebności oczekiwanych e_{ijt} są takie same jak sumy brzegowe liczebności n_{ijt} .

Generację zawartości TT $2 \times 2 \times 2$ niezbędną do przeprowadzenia symulacji i wyznaczenia mocy testów przeprowadzono metodą słupkową (Sulewski, 2014). W tym celu przedział $(0,1)$ podzielono na $2 \cdot 2 \cdot 2$ podprzedziałów o szerokościach równych wartości prawdopodobieństw p_{ijt} w taki sposób, że pierwszy podprzedział ma szerokość p_{111} , drugi – p_{112}, \dots , ostatni – p_{222} . Wielkości p_{ijt} spełniają oczywiście warunek normalizacji $\sum_{i=1}^2 \sum_{j=1}^2 \sum_{t=1}^2 p_{ijt} = 1$ i dobrano je w taki sposób, aby uzyskać żadaną wartość miary nieprawdziwości 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 TT. Wielkości n_{ijt} spełniające równość $\sum_{i=1}^2 \sum_{j=1}^2 \sum_{t=1}^2 n_{ijt} = n$ są liczebnością obiektów w poszczególnych komórkach TT. Rysunek 1 przedstawia schemat wypełnienia komórek TT $2 \times 2 \times 2$ dla liczebności próby $n = 200$, gdy hipoteza zerowa jest słuszna, czyli $p_{ijt} = 1/8$ dla każdego $i, j, t = 1, 2$. Tabela 2 prezentuje odpowiadającą temu schematowi TT $2 \times 2 \times 2$.



Rysunek 1. Schematyczne ujęcie metody słupkowej

Źródło: opracowanie własne.

Tabela 2.

TT $2 \times 2 \times 2$ uzyskana metodą słupkową

Z	Z ₁		Z ₂		Razem
	Y ₁	Y ₂	Y ₁	Y ₂	
X ₁	23	28	29	25	105
X ₂	16	29	26	24	95
Razem	39	57	55	49	200

Źródło: opracowanie własne.

3. TESTY NIEZALEŻNOŚCI W TT 2×2×2

W analizie niezależności cech X, Y, Z można badać: niezależność pełną, niezależność brzegową, niezależność częściową, niezależność łączną oraz niezależność warunkową. W pracy niniejszej badania ograniczone zostały tylko do niezależności pełnej, którą można wyrazić zapisem (XYZ, XY, XZ, YZ) .

Przy badaniu niezależności pełnej cech w tablicy 2×2×2 założono hipotezę zerową, że między cechami X, Y, Z związek nie występuje, tzn.

$$H_0 : p_{ijt} = p_{i..} \cdot p_{.j.} \cdot p_{..t} \quad (2)$$

dla każdego $i, j, t = 1, 2$.

Do badania niezależności pełnej cech X, Y, Z w tablicy 2×2×2 skorzystano ze statystyki modułowej, która ma postać

$$|\chi| = \sum_{i=1}^2 \sum_{j=1}^2 \sum_{t=1}^2 \frac{|n_{ijt} - e_{ijt}|}{e_{ijt}} \quad (3)$$

oraz z rodziny statystyk, które są rozszerzeniem „statystyk chi-kwadrat” dla TD (Cressie, Read, 1984). Zalicza się do nich:

1) statystykę χ^2 Pearsona

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \sum_{t=1}^2 \frac{(n_{ijt} - e_{ijt})^2}{e_{ijt}}, \quad (4)$$

2) statystykę G^2 ilorazu wiarygodności

$$G^2 = 2 \sum_{i=1}^2 \sum_{j=1}^2 \sum_{t=1}^2 n_{ijt} \ln \left(\frac{n_{ijt}}{e_{ijt}} \right), \quad (5)$$

3) statystykę N Neymana

$$N = \sum_{i=1}^2 \sum_{j=1}^2 \sum_{t=1}^2 \frac{(n_{ijt} - e_{ijt})^2}{n_{ijt}}, \quad (6)$$

4) statystykę KL Kullbacka-Leiblera

$$KL = 2 \sum_{i=1}^2 \sum_{j=1}^2 \sum_{t=1}^2 e_{ijt} \ln \left(\frac{e_{ijt}}{n_{ijt}} \right), \quad (7)$$

5) statystykę *FT* Freemana-Tukeya

$$FT = 4 \sum_{i=1}^2 \sum_{j=1}^2 \sum_{t=1}^2 \left(\sqrt{n_{ijt}} - \sqrt{e_{ijt}} \right)^2, \quad (8)$$

6) statystykę *CR* Cressiego-Reada

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

Powyższa rodzina statystyk dla TT $2 \times 2 \times 2$ przy prawdziwości hipotezy H_0 ma asymptotyczny rozkład chi-kwadrat z 4 stopniami swobody i dlatego jest określana mianem „statystyk chi-kwadrat”.

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

Liczebności n_{ijt} ($i, j, t = 1, 2$) w TT $2 \times 2 \times 2$ są zmiennymi losowymi. Godnym uwagi jest fakt, że zmienne losowe będące odwrotnościami lub ilorazami zmiennych losowych (patrz statystyki (6) i (7)) mogą nie posiadać 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 (5), (7) i (9) są różnego znaku i wzajemnie się rekompensują i tylko dla prób liczących przynajmniej kilkaset elementów podlegają rozkładowi chi-kwadrat (Sulewski, 2014). Tabela 3 przedstawia wartości poszczególnych składników „statystyk chi-kwadrat” wyznaczonych symulacyjnie dla TT $2 \times 2 \times 2$ i liczebności próby $n = 200$, gdy H_0 o niezależności cech X, Y, Z jest słuszna. Tabelę tę wygenerowano metodą słupkową opisaną powyżej dla $p_{ijt} = 0,125$ ($i, j, t = 1, 2$). Wartości ujemne pogrubiono.

Tabela 3.

Wartości składników „statystyk chi-kwadrat”

Składnik	χ^2	G^2	N	KL	FT	CR
$i = 1, j = 1, t = 1$	0,667	-7,297	0,801	-8,758	0,730	-4,123
$i = 1, j = 1, t = 2$	0,260	-5,197	0,235	-4,702	0,247	-3,224
$i = 1, j = 2, t = 1$	0,010	-1,036	0,010	-1,015	0,010	-0,625
$i = 1, j = 2, t = 2$	0,041	-2,076	0,040	-1,994	0,041	-1,262
$i = 2, j = 1, t = 1$	2,573	-18,370	1,948	-13,907	2,228	12,111

Tabela 3. (cd.)

Składnik	χ^2	G^2	N	KL	FT	CR
$i = 2, j = 1, t = 2$	1,719	-11,081	2,338	-15,074	1,993	-6,011
$i = 2, j = 2, t = 1$	0,773	-8,234	0,932	-9,929	0,847	-4,645
$i = 2, j = 2, t = 2$	0,346	-6,328	0,310	-5,674	0,327	-3,938
Razem	6,389	-6,394	6,614	-6,468	6,422	-6,383

Źródło: opracowanie własne.

4. WYZNACZANIE WARTOŚCI KRYTYCZNYCH

W okresie ostatnich kilkudziesięciu lat zaproponowano różne testy niezależności dla tablic kontyngencji oraz określono warunki ich stosowalności szczególnie dla małych prób, które najczęściej są przedmiotem badań statystycznych. Warunki te głównie dotyczą bardzo popularnego i powszechnie stosowanego testu niezależności χ^2 Pearsona. W dobie powszechnie stosowanych i ciągle rozwijających się technologii komputerowych 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 TT $2 \times 2 \times 2$ o liczebności próby n wygenerowano za pomocą metody słupkowej przyjmując $p_{ijt} = 0,125$ ($i, j, t = 1, 2$).

Niech ϖ będzie jedną z rozpatrywanych siedmiu statystyk. Dla każdej TT $2 \times 2 \times 2$ obliczono $r = 10^5$ razy wartości statystyki testowej ϖ , które następnie uporządkowano w kolejności rosnącej. Wartość krytyczną na poziomie istotności $\alpha = 0,1$ wyznaczono na podstawie dziewiątego decyla ze wzoru $cv_\alpha = \varpi_{(90000)}$. Tak duża liczba powtórzeń r przy wyznaczaniu wartości statystyki testowej zapewnia uzyskanie dokładnego wyniku.

5. MIARY SIŁY ZWIĄZKU DLA TT $2 \times 2 \times 2$

Przeprowadzając badanie populacji generalnej istotna jest nie tylko zależność istniejąca między cechami, ale także jej siła. Teoria poświęcona miarom siły związku dla TT nie jest tak bogata jak dla TD, gdzie na szczególną uwagę zasługuje współczynnik τ Goodmana–Kruskala (Goodman, Kruskal, 1979). Numerycznymi rozszerzeniami tego współczynnika dla TT są: współczynnik Graya–Williamsa (Gray, Williams, 1975), współczynnik Marcotorchino (Marcotorchino, 1984) oraz współczynnik Lombardo (Lombardo, 2011). Informacje o innych mniej popularnych miarach można znaleźć w pracach Tuckera (1963), Harshmana (1970), Beha, Davy (1998) i Lombardo, Beha (2010).

Jeżeli dwie cechy (np. Y, Z) są cechami objaśniającymi lub predyktorami wzajemnie zależnymi (ang. *dependent predictor variables*), a trzecia cecha (np. X) jest

cechą objaśnianą lub odpowiedzi (ang. *response variable*), to współczynnik Graya–Williamsa ma postać:

$$\tau_{GW} = \left[\sum_{i=1}^2 \sum_{j=1}^2 \sum_{t=1}^2 p_{\bullet jt} (p_{ijt} \cdot p_{\bullet jt}^{-1} - p_{i\bullet\bullet})^2 \right] \cdot \left[1 - \sum_{i=1}^2 p_{i\bullet\bullet}^2 \right]^{-1}, \quad (10)$$

gdzie $p_* = n_* / n$. Współczynnik ten przyjmuje wartości z przedziału $\langle 0, 1 \rangle$. Jeżeli cechy Y i Z nie dostarczają informacji o cesze X , wtedy $\tau_{GW} = 0$. Wraz ze wzrostem informacji dostarczanej przez cechy Y i Z o cesze X wartość tego współczynnika zbliża się do jedności.

Jeżeli dwie cechy (np. Y, Z) są cechami objaśniającymi lub predyktorami wzajemnie niezależnymi (ang. *independent predictor variables*), a trzecia cecha (np. X) jest cechą objaśnianą lub odpowiedzi (*response variable*), to współczynnik Marcotorchino ma postać:

$$\tau_M = \left[\sum_{i=1}^2 \sum_{j=1}^2 \sum_{t=1}^2 p_{\bullet j\bullet} \cdot p_{\bullet\bullet t} (p_{ijt} \cdot p_{\bullet\bullet t}^{-1} \cdot p_{\bullet j\bullet}^{-1} - p_{i\bullet\bullet})^2 \right] \cdot \left[1 - \sum_{i=1}^2 p_{i\bullet\bullet}^2 \right]^{-1}, \quad (11)$$

gdzie $p_* = n_* / n$. Jeżeli $p_{ijt} = p_{i\bullet\bullet} p_{\bullet j\bullet} p_{\bullet\bullet t}$, to $\tau_M = 0$, natomiast gdy $p_{ijt} = p_{i\bullet\bullet} p_{\bullet\bullet t}$, to $\tau_M = 1$.

Jeżeli dwie cechy (np. X, Y) są cechami objaśnianymi lub odpowiedzi (ang. *independent response variables*) a trzecia cecha (np. Z) jest cechą objaśniającą lub predyktorem, to współczynnik Lombardo – zwany także współczynnikiem delta – ma postać

$$\tau_L = \left[\sum_{i=1}^2 \sum_{j=1}^2 \sum_{t=1}^2 p_{\bullet\bullet t} (p_{ijt} \cdot p_{\bullet\bullet t}^{-1} - p_{i\bullet\bullet} \cdot p_{\bullet j\bullet})^2 \right] \cdot \left[1 - \sum_{j=1}^2 p_{i\bullet\bullet}^2 \cdot \sum_{i=1}^2 p_{\bullet j\bullet}^2 \right]^{-1}, \quad (12)$$

gdzie $p_* = n_* / n$. Jeżeli $p_{ijt} = p_{i\bullet\bullet} p_{\bullet j\bullet} p_{\bullet\bullet t}$, to $\tau_L = 0$. Wartość maksymalna $\tau_L = 1$, gdy $p_{1jt} = 0$ dla pewnych $(j, t = 1, 2)$ i $p_{2bc} = 0,5$ dla pewnych $(b, c = 1, 2; b \neq j \wedge c \neq t)$.

Więcej informacji na temat trzech powyższych współczynników – w tym także ich implementację komputerową w języku VBA – można znaleźć w pracach Sulewski (2014, 2015).

Bazując na klasycznej definicji niezależności cech X, Y, Z proponuje się miarę nieprawdziwości H_0 w postaci:

$$mn = \sum_{i=1}^2 \sum_{j=1}^2 \sum_{t=1}^2 |p_{ijt} - p_{i\bullet\bullet} \cdot p_{\bullet j\bullet} \cdot p_{\bullet\bullet t}|. \quad (13)$$

Hipoteza zerowa H_0 – mówiąca o tym, że między cechami X, Y, Z w TT $2 \times 2 \times 2$ nie ma związku – jest słuszna, gdy $p_{ijt} = p_{i\bullet\bullet} \cdot p_{\bullet j\bullet} \cdot p_{\bullet\bullet t}$ dla $i, j, t = 1, 2$. Zatem miara (13) przyjmuje wartość 0, gdy hipoteza H_0 jest słuszna. Im większe wartości mn , tym większa jest możliwość fałszywości H_0 .

Miarę mn jako naturalną – wynikającą z definicji niezależności cech – wykorzystano w procesie wyznaczania mocy testów.

6. MOC TESTU

W celu wyznaczenia mocy testu – czyli zdolności TT do odrzucenia hipotezy zerowej H_0 mówiącej o tym, że nie ma związku między cechami X, Y, Z – gdy w istocie związek jest, niezbędna staje się generacja zawartości TT. Uwzględniając narzuconą siłę związku między cechami uzyskaną dzięki generatorowi liczb losowych i wyrażoną za pomocą miary nieprawdźliwości mn , do wypełnienia TT $2 \times 2 \times 2$ skorzystano z metody „słupkowej” wykorzystującej prawdopodobieństwa p_{ijt} ($i, j, t = 1, 2$).

Znając wartość statystyki testowej ϖ dla różnych liczebności próby n wyznaczono moc testu za pomocą wzoru $m = q / r$, gdzie q określa liczbę tych spośród $r = 10^5$ wszystkich możliwych przypadków, kiedy to wartość statystyki testowej ϖ jest nie mniejsza od wartości krytycznej cv_{α} .

Do wyznaczenia mocy testu skorzystano ze schematów A, B, C, AB, AC, BC wyznaczania prawdopodobieństw p_{ijt} ($i, j, t = 1, 2$) (tabela 4).

Tabela 4.

Schematy prawdopodobieństw ukazujące związek między cechami X, Y, Z

Schemat A

Z	Z_1		Z_2	
$X \setminus Y$	Y_1	Y_2	Y_1	Y_2
X_1	$p_o - k \cdot \Delta p$	p_o	$p_o - k \cdot \Delta p$	p_o
X_2	p_o	$p_o + k \cdot \Delta p$	p_o	$p_o + k \cdot \Delta p$

Schemat B

Z	Z_1		Z_2	
$X \setminus Y$	Y_1	Y_2	Y_1	Y_2
X_1	$p_o - k \cdot \Delta p$	p_o	$p_o - k \cdot \Delta p$	p_o
X_2	$p_o + k \cdot \Delta p$	p_o	$p_o + k \cdot \Delta p$	p_o

Schemat C

Z	Z_1		Z_2	
$X \setminus Y$	Y_1	Y_2	Y_1	Y_2
X_1	$p_o - k \cdot \Delta p$	$p_o + k \cdot \Delta p$	$p_o - k \cdot \Delta p$	$p_o + k \cdot \Delta p$
X_2	$p_o + k \cdot \Delta p$	$p_o - k \cdot \Delta p$	$p_o + k \cdot \Delta p$	$p_o - k \cdot \Delta p$

Schemat AB

Z	Z_1		Z_2	
$X \setminus Y$	Y_1	Y_2	Y_1	Y_2
X_1	$p_o - k \cdot \Delta p$	p_o	$p_o - k \cdot \Delta p$	p_o
X_2	p_o	$p_o + k \cdot \Delta p$	$p_o + k \cdot \Delta p$	p_o

Schemat AC

Z	Z ₁		Z ₂	
X \ Y	Y ₁	Y ₂	Y ₁	Y ₂
X ₁	$p_o - k \cdot \Delta p$	p_o	$p_o - k \cdot \Delta p$	$p_o + k \cdot \Delta p$
X ₂	p_o	$p_o + k \cdot \Delta p$	$p_o + k \cdot \Delta p$	$p_o - k \cdot \Delta p$

Schemat BC

Z	Z ₁		Z ₂	
X \ Y	Y ₁	Y ₂	Y ₁	Y ₂
X ₁	$p_o - k \cdot \Delta p$	p_o	$p_o - k \cdot \Delta p$	$p_o + k \cdot \Delta p$
X ₂	$p_o + k \cdot \Delta p$	p_o	$p_o + k \cdot \Delta p$	$p_o - k \cdot \Delta p$

Gdzie $p_o = 0,125$, $\Delta p = 0,125 \times 10^{-3}$, $k = 0, \dots, 1000$.

Źródło: opracowanie własne.

Wartość minimalna miar siły związku (13) wynosi zero, dlatego w tabeli 5 przedstawiono tylko wartości maksymalne czterech miar siły związku dla różnych schematów prawdopodobieństw. Moc testu została wyznaczona dla wartości miary mn w punktach $u = 0, 1, \dots, 9$. Miara ta dla wszystkich schematów prawdopodobieństw zmienia się ze stałym krokiem. W celu uzyskania takiej sytuacji należało odpowiednio dobrać wartości indeksu nieprawdziwości k (tabela 6).

Tabela 5.

Wartości maksymalne miar siły związku

Miara	A	B	C	AB	AC	BC
τ_{GW}	0,111	0,333	1	0,222	0,644	0,733
τ_M	0,111	0,333	1	0,378	0,716	0,733
τ_L	0,026	0,091	0,333	0,099	0,241	0,234
mn	0,25	0,5	1	0,44	0,67	0,75

Źródło: opracowanie własne.

Tabela 6.

Wartość indeksu nieprawdziwości k w punkcie u

u	A	B, C, BC	AB	AC
0	0	0	0	0
1	316	100	115	107
2	448	200	226	211
3	548	300	334	314
4	633	400	438	416

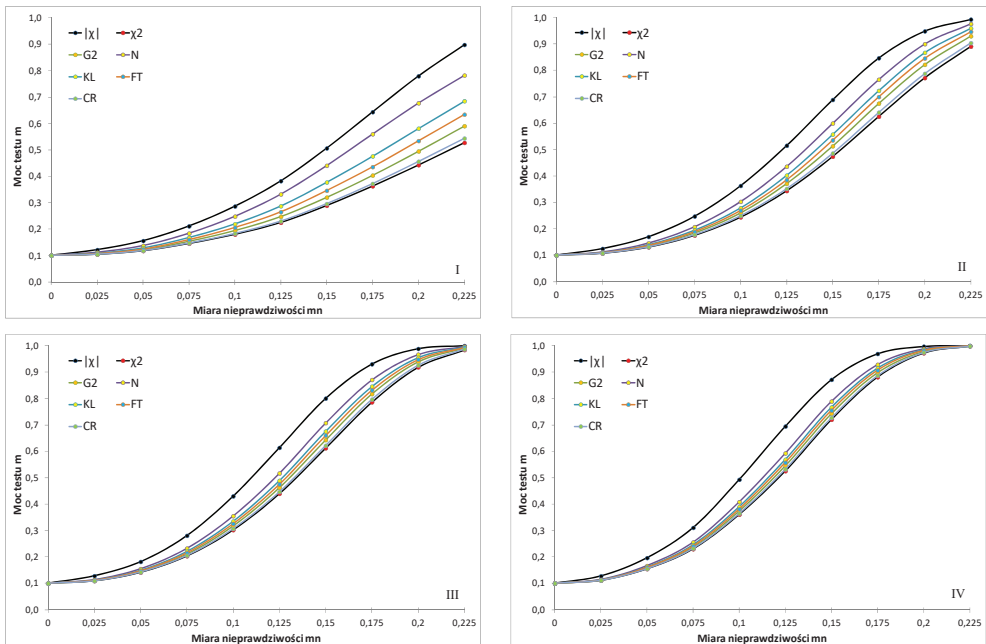
Tabela 6. (cd.)

u	A	B, C, BC	AB	AC
5	707	500	539	516
6	775	600	636	615
7	837	700	733	713
8	895	800	826	813
9	949	900	917	904
10	1000	1000	1000	1000

Źródło: opracowanie własne.

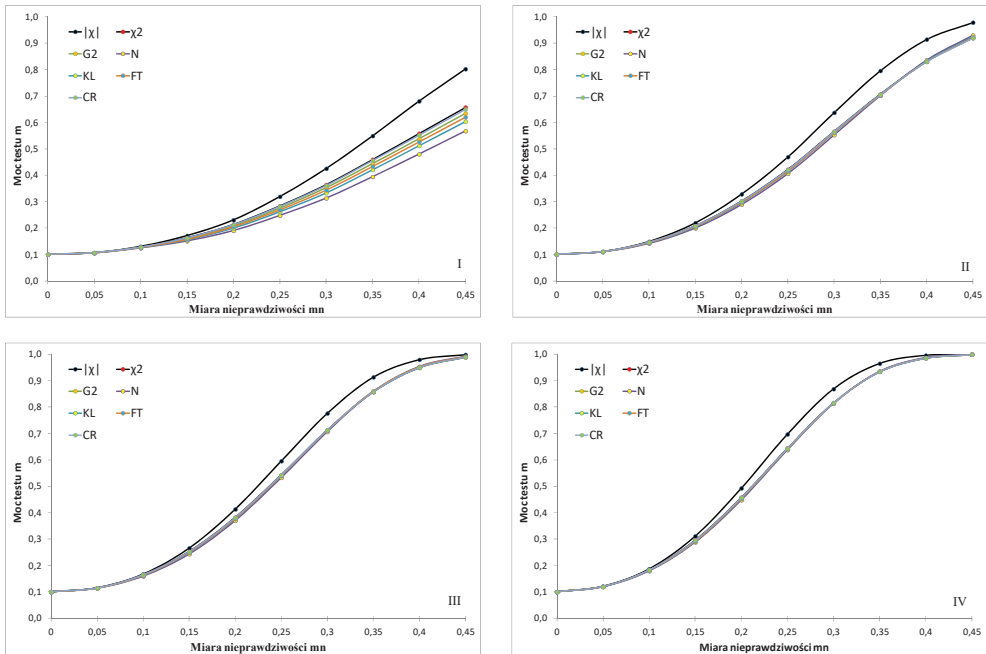
Aby można było porównać analizowane testy niezależności, liczebności komórek TT muszą być niezerowe, czyli $n_{ijt} \neq 0$ dla każdego $i, j, t = 1, 2$, w sytuacji przeciwnej nie da się obliczyć wartości statystyk G^2 , N , KL . W związku z powyższym moc testów niezależności zostanie obliczona dla $u = 0, 1, \dots, 9$. Minimalną liczebność próby n dla danego schematu dobrano tak, aby $n_{ijt} \neq 0$ dla każdego $i, j, t = 1, 2$. Maksymalną liczebność próby ustalono tak, aby uzyskać maksymalną moc testu.

Rysunki 2–7 przedstawiają zależność mocy testów niezależności od miary nieprawdliwości H_0 dla różnych schematów prawdopodobieństw, na poziomie istotności $\alpha = 0,1$ i liczebności próby n .



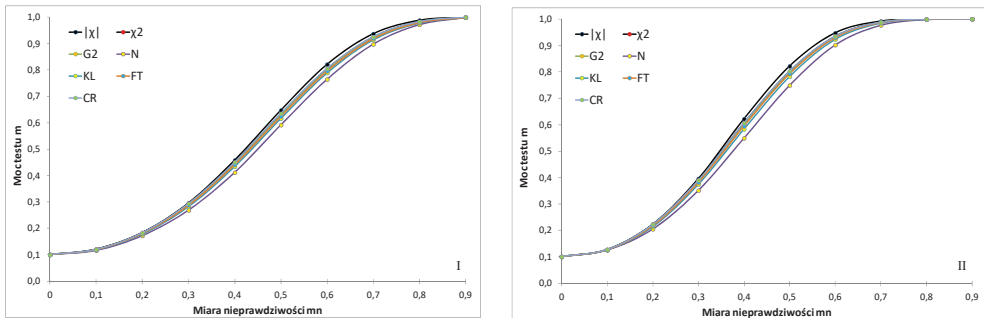
Rysunek 2. Moc testów niezależności dla schematu A prawdopodobieństw i liczebności próby $n = 40$ (I), $n = 60$ (II), $n = 80$ (III), $n = 100$ (IV)

Źródło: opracowanie własne.



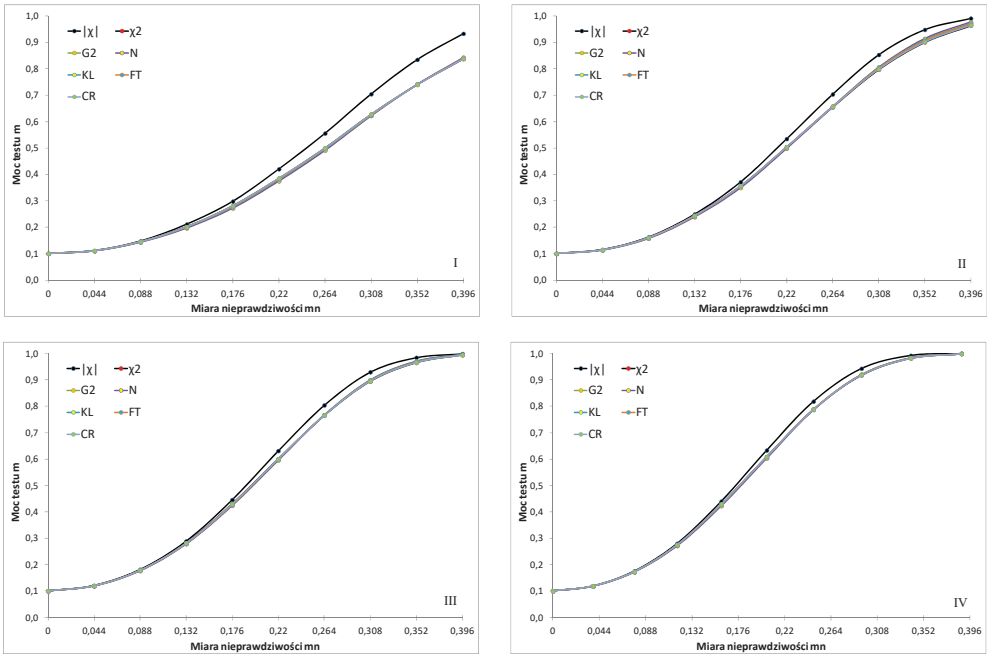
Rysunek 3. Moc testów niezależności dla schematu B prawdopodobieństw i liczebności próby $n = 40$ (I), $n = 60$ (II), $n = 80$ (III), $n = 100$ (IV)

Źródło: opracowanie własne.



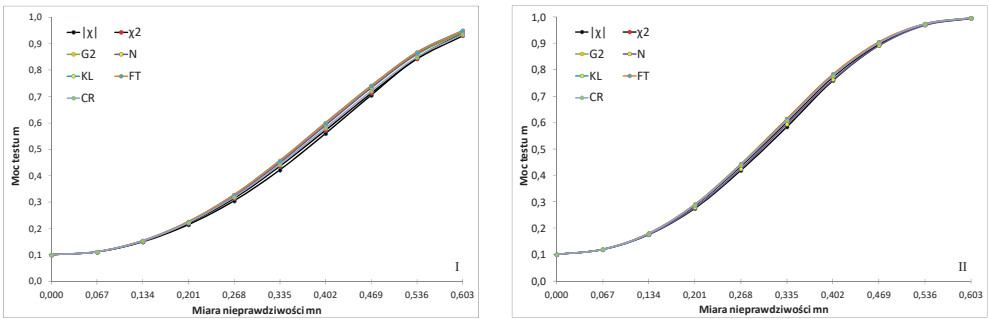
Rysunek 4. Moc testów niezależności dla schematu C prawdopodobieństw i liczebności próby $n = 30$ (I), $n = 40$ (II)

Źródło: opracowanie własne.



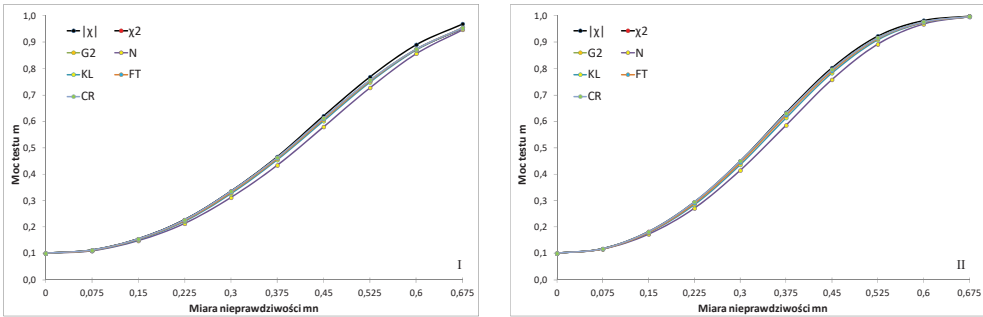
Rysunek 5. Moc testów niezależności dla schematu AB prawdopodobieństw i liczebności próby $n = 60$ (I), $n = 80$ (II), $n = 100$ (III), $n = 120$ (IV)

Źródło: opracowanie własne.



Rysunek 6. Moc testów niezależności dla schematu AC prawdopodobieństw i liczebności próby $n = 30$ (I), $n = 40$ (II)

Źródło: opracowanie własne.



Rysunek 7. Moc testów niezależności dla schematu BC prawdopodobieństw i liczebności próby $n = 30$ (I), $n = 40$ (II)

Źródło: opracowanie własne.

Dla statystyki modułowej $|\chi|$ oraz statystyki χ^2 Pearsona jako reprezentanta „statystyk chi-kwadrat”, uzyskane wyniki porównano za pomocą testu o równości prawdopodobieństw stawiając hipotezę zerową, że moce testów są równe. Symbol „+” oznacza, że różnice między mocami testów są statystycznie istotne, symbol „-” – są statystycznie istotne (tabele 7–9).

Tabela 7.

Istotność statystyczna mocy testów $|\chi|$ i χ^2 dla schematów A i B

Schemat A

Schemat B

mn	$n = 100$	$n = 150$	$n = 200$	$n = 250$
0	+	+	+	+
0,025	+	+	+	+
0,05	-	-	-	-
0,075	-	-	-	-
0,1	-	-	-	-
0,125	-	-	-	-
0,15	-	-	-	-
0,175	-	-	-	-
0,2	-	-	-	-
0,225	+	-	+	+

mn	$n = 40$	$n = 60$	$n = 80$	$n = 100$
0	+	+	+	+
0,05	+	+	+	+
0,1	+	+	+	+
0,15	+	+	+	+
0,2	+	+	+	+
0,25	-	-	-	-
0,3	-	-	-	-
0,35	-	-	-	-
0,4	-	-	-	-
0,45	-	-	-	+

Źródło: opracowanie własne.

Tabela 8.

Istotność statystyczna mocy testów $|\chi|$ i χ^2 dla schematów C i AB

Schemat C			Schemat AB				
mn	$n = 30$	$n = 40$	mn	$n = 60$	$n = 80$	$n = 100$	$n = 120$
0	+	+	0	+	+	+	+
0,1	+	+	0,044	+	+	+	+
0,2	+	+	0,088	+	+	+	+
0,3	+	+	0,132	+	+	+	+
0,4	+	+	0,176	+	+	+	+
0,5	+	+	0,22	+	+	+	+
0,6	+	+	0,264	-	-	-	-
0,7	+	+	0,308	-	-	-	-
0,8	+	+	0,352	-	-	-	-
0,9	+	+	0,396	+	+	+	+

Źródło: opracowanie własne.

Tabela 9.

Istotność statystyczna mocy testów $|\chi|$ i χ^2 dla schematów AC i BC

Schemat AC			Schemat BC		
mn	$n = 30$	$n = 40$	mn	$n = 30$	$n = 40$
0	+	+	0	+	+
0,067	+	+	0,075	+	+
0,134	+	+	0,15	+	+
0,201	+	+	0,225	+	+
0,268	+	+	0,3	+	+
0,335	+	+	0,375	+	+
0,402	+	+	0,45	+	+
0,469	+	+	0,525	+	+
0,536	+	+	0,6	+	+
0,603	+	+	0,675	+	+

Źródło: opracowanie własne.

Z rysunków 2–7 wynika, że testy niezależności wykorzystujące „statystyki chi-kwadrat” we wszystkich schematach prawdopodobieństw charakteryzują się podobną mocą dla danej liczebności próby n i miary mn . Wyjątkiem jest tylko schemat A dla którego miara mn przyjmuje najmniejsze wartości. Test wykorzystujący statystykę modułową ma podobną moc jak pozostałe testy niezależności dla schematów C, AC, BC. Dla schematu A test wykorzystujący statystykę modułową ma większą moc niż testy związane ze „statystykami chi-kwadrat” i ta przewaga jest statystycznie istotna dla $mn > 0,025$. Podobna sytuacja jest także dla schematu B, gdy $mn > 0,02$ oraz dla schematu AB, gdy $mn > 0,022$.

7. 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.

Artykuł zwraca uwagę na fakt, że testy niezależności wykorzystujące „statystyki chi-kwadrat” we wszystkich schematach prawdopodobieństw charakteryzują się podobną mocą dla danej liczebności próby n i miary mn . Nieznacznym wyjątkiem od tej reguły jest schemat A, w którym miara nieprawdziwości mn przyjmuje najmniejsze wartości. Statystyka modułowa charakteryzuje się podobną mocą jak „statystyki chi-kwadrat” dla schematów C, AC, BC oraz większą mocą dla schematów A, B, AB. Przewaga na korzyść nowej statystyki szczególnie jest widoczna dla mniejszych liczebności próby w ramach danego schematu.

Badając niezależność trzech cech w pewnych schematach prawdopodobieństwa dla statystyki modułowej uzyskano wyższą moc wykrywania zależności, rozumianej w sensie zaproponowanej miary mn . Na korzyść proponowanej statystyki przemawia także fakt, że można z niej skorzystać w sytuacji, gdy komórki TT są puste. Statystyki G^2 ilorazu wiarygodności, N Neymana i KL Kullbacka-Leiblera nie spełniają tego założenia.

Zatem można wnioskować, że statystyka modułowa jest nie mniej skuteczna niż „statystyki chi-kwadrat” i może stanowić alternatywę, zwłaszcza uwzględniając jej inne zalety.

LITERATURA

- Beh E. J., Davy P. J., (1998), Partitioning Pearson's Chi-squared Statistic for a Completely Ordered Three-way Contingency Table, *The Australian and New Zealand Journal of Statistics*, 40, 465–477.
- 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.

- Goodman L., Kruskal W., (1954), Measures of Association for Cross Classifications, *Journal of the American Statistical Association*, 49, 732–764.
- Gray L. N., Williams J. S., (1975), Goodman and Kruskal's Tau b: Multiple and Partial Analogs, in: *Proceedings of the Social Statistics Section, American Statistical Association*.
- Harshman R. A., (1970), Foundations of the PARAFAC Procedure: Models and Conditions for an Explanatory Multi-modal Factor Analysis, *UCLA Working Papers in Phonetics*, 16, 1–84.
- Kullback S., (1959), *Information Theory and Statistics*, Wiley, New York.
- Lombardo R., (2011), Three-way Association Measure Decompositions: The Delta index, *Journal of Statistical Planning and Inference*, 141, 1789–1799.
- Lombardo R., Beh E. J., (2010), Simple and Multiple Correspondence Analysis for Ordinal-scale Variables Using Orthogonal Polynomials, *Journal of Applied Statistics*, 37 (12), 2101–2116.
- Marcotorchino F., (1984), *Utilisation des Comparaisons par Paires en Statistique des Contingences*, Partie (I), *Etude IBM F069*, France.
- Neyman J., (1949), Contribution to the Theory of the χ^2 Test, *Proceedings of the [First] Berkeley Symposium on Mathematical Statistics and Probability*, (Univ. of Calif. 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.
- 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., (2015), Miary związku między cechami w tablicy trójdzielczej, *Wiadomości Statystyczne*, GUS, 1, 13–27.
- Sulewski P., (2016a), Moc testów niezależności w tablicy dwudzielczej, *Wiadomości Statystyczne*, GUS, 8, 1–17.
- Sulewski P., (2016b), Moc testów niezależności w tablicy dwudzielczej większej niż 2×2 , *Przegląd Statystyczny*, 63 (2), 191–209.
- Tucker L. R., (1963), Implications of Factor Analysis of Three-way Matrices for Measurement of Change, w: Harris C. W. (red.), *In Problems in Measuring Change*, 122–137.

MOC TESTÓW NIEZALEŻNOŚCI W TABLICY TRÓJDZIELCZEJ $2 \times 2 \times 2$

Streszczenie

Pierwszym celem pracy jest przedstawienie teorii dotyczącej autorskiej statystyki modułowej mierzącej niezależność zmiennych dla tablic trójdzielczych $2 \times 2 \times 2$ i zbadanie jej własności w odniesieniu do znanych „statystyk chi-kwadrat”. Drugim celem jest opisanie procedury generowania zawartości tych tablic metodą słupkową. Trzecim celem jest zaproponowanie miary nieprawdziwości hipotezy zerowej, a także porównanie jakości testów niezależności za pomocą ich mocy. Wartości krytyczne dla testów niezależności wyznaczono symulacyjnie metodami Monte Carlo.

Słowa kluczowe: statystyka modułowa, tablica trójdzielcza, test niezależności, metoda słupkowa, Monte Carlo

POWER ANALYSIS OF INDEPENDENCE TESTING
FOR THREE-WAY CONTINGENCY TABLE $2 \times 2 \times 2$

A b s t r a c t

The first aim of this paper is to present the theory of the proposal of the author in the form of modular statistics for three-way contingency table $2 \times 2 \times 2$ and examine its properties in relation to known “chi-squared statistics”. The second aim is to describe the procedure of generating the content of these tables using the bar method. The third aim is to propose the measure of untruthfulness of null hypothesis as well as to compare the quality of independence tests using their power. Critical values for all analyzed statistics were determined by simulation methods of Monte Carlo.

Keywords: three-way contingency tables, modular statistics, independence test, bar method, Monte Carlo method