

## STUDIA METODOLOGICZNE

**Piotr SULEWSKI**

### Moc testów niezależności w tablicy dwudzielczej

---

**Streszczenie.** *W literaturze statystycznej istnieje wiele miar testowych do badania niezależności cech w tablicach dwudzielczych. W zaprezentowanej pracy do analizy statystycznej wybrano tzw. statystykę chi-kwadrat, w tym statystykę  $\chi^2$  Pearsona, a także przedstawiono propozycję Autora w postaci statystyki modułowej. W celu wyeliminowania ograniczeń dotyczących statystyki chi-kwadrat, wartości krytyczne dla całej analizowanej statystyki wyznaczono symulacyjnie metodami Monte Carlo. Do porównania testów zaproponowano miarę nieprawdziwości  $H_0$  oraz wyznaczono moc testów, czyli zdolność tablicy dwudzielczej  $2 \times 2$  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.

---

Moc testów to prawdopodobieństwo odrzucenia hipotezy zerowej  $H_0$ , gdy nie jest ona prawdziwa. Testy niezależności — obok m.in. testów  $t$ -studenta, Kołmogorowa czy Behrensa-Fishera — są zapewne najczęściej stosowanymi narzędziami statystycznymi. Dane do testów niezależności aranżuje się w postaci tablic dwudzielczych, a w szczególności tablic  $2 \times 2$ .

W używanej często frazie „statystyka testowa ma asymptotyczny rozkład chi-kwadrat” ukryta jest nieprzyjemna prawda, że dość często rozkład statystyki testowej z powodu małych próbek odbiega od rozkładu asymptotycznego, jakim jest rozkład chi-kwadrat.

W artykule porównano rodzinę tzw. „statystyki chi-kwadrat” — w skład której wchodzi powszechnie znana i często stosowana statystyka  $\chi^2$  Pearsona — ze statystyką modułową (Sulewski, 2013). W innej pracy (Sulewski, 2014) pokazano, że tzw. „statystyki chi-kwadrat” ( $\chi^2$  Pearsona,  $G^2$  ilorazu wiarygodności,  $N$  Neymana,  $KL$  Kullbacka-Leiblera,  $FT$  Freemana-Tukeya oraz  $CR$  Cressiego-Reada) jest wystarczająco bliska rozkładowi chi-kwadrat dla próbek liczących przynajmniej kilkaset elementów.

Korzystanie ze „statystyk chi-kwadrat” w tablicach dwudzielczych, szczególnie dla małych próbek, wiąże się ze stosowaniem różnego rodzaju ograniczeń. Na przykład ze statystyki  $\chi^2$  Pearsona można korzystać, gdy liczebność próby  $n \geq 1000$  oraz wielkości oczekiwane  $e_{ij} \geq 5$  dla każdego  $i, j = 1, 2$  (Cochran, 1952). Jeżeli  $1 \leq e_{ij} < 5$  dla pewnego  $i, j = 1, 2$ , to należy stosować test  $n-1$  chi-kwadrat (Pearson, 1947), natomiast w przypadku  $e_{ij} < 1$  dla pewnego  $i, j = 1, 2$  stosuje się test Fishera-Yatesa (Campbell, 2007).

Aby wyeliminować te ograniczenia zaproponowano wyznaczenie wartości krytycznych na drodze symulacji komputerowych metodą Monte Carlo, na podstawie np. 100000 wartości statystyki testowej, co w dobie wydajnych komputerów z procesorami wielordzeniowymi nie stanowi problemu, ponieważ trwa kilkadziesiąt sekund.

Wymienione testy porównano ze względu na ich moc, czyli zdolności testu do odrzucenia hipotezy  $H_0$  mówiącej o tym, że nie ma związku między cechami  $X$  i  $Y$  w sytuacji, gdy w rzeczywistości jest ona fałszywa.

Do wyznaczenia mocy testów tablice dwudzielcze  $2 \times 2$  generowano za pomocą metody słupkowej (Sulewski, 2015b) oraz zaproponowano miarę nieprawdziwości  $H_0$  przyjmującą różne wartości w przypadku danego schematu wyznaczania prawdopodobieństw  $p_{ij}$  ( $i, j = 1, 2$ ), dla których  $H_0$  jest niesłuszna.

Pierwszym celem pracy jest przedstawienie teorii dotyczącej testów niezależności dla tablicy dwudzielczej  $2 \times 2$ , wprowadzenie miary nieprawdziwości  $H_0$  oraz porównanie jakości testów za pomocą ich mocy. Drugi natomiast cel stanowi propozycja gotowego narzędzia do samodzielnego prowadzenia badań statystycznych w postaci pliku umieszczonego w Internecie, opisanego w językach polskim i angielskim.

Praca składa się z dwóch części. W części I zdefiniowano tablicę dwudzielczą  $2 \times 2$  oraz siedem testów niezależności, opisano sposób wyznaczania wartości krytycznych oraz porównano jakość testów niezależności na podstawie ich mocy.

Część II zawiera implementację zastosowanej teorii, na którą składa się oprogramowanie utworzone w edytorze języka VBA (*Visual Basic for Applications*)

arkusza kalkulacyjnego Microsoft Excel w formie pliku umieszczonego w Internecie<sup>1</sup>. Za pomocą tego pliku czytelnik może samodzielnie badać niezależność cech w tablicy dwudzielczej  $2 \times 2$  wykorzystując statystykę modułową.

*TABLICA DWUDZIELCZA  $2 \times 2$*

Tablica dwudzielcza  $2 \times 2$ , która składa się z wartości  $n_{ij}$  ( $i, j = 1, 2$ ) rozkładu łącznego cech  $X$  i  $Y$  takich, że  $\sum_{i=1}^2 \sum_{j=1}^2 n_{ij} = n$ .

**TABL. 1. TABLICA DWUDZIELCZA  $2 \times 2$  LICZEBNOŚCI**

Cecha $X$	Cecha $Y$		Razem
	$Y_1$	$Y_2$	
$X_1$ .....	$n_{11}$	$n_{12}$	$n_{1\bullet}$
$X_2$ .....	$n_{21}$	$n_{22}$	$n_{2\bullet}$
R a z e m .....	$n_{\bullet 1}$	$n_{\bullet 2}$	$n$

Ź r ó d ł o: opracowanie własne.

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

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

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

**TABL. 2. TABLICA DWUDZIELCZA  $2 \times 2$  PRAWDOPODOBIEŃSTWA**

Cecha $X$	Cecha $Y$		Razem
	$Y_1$	$Y_2$	
$X_1$ .....	$p_{11}$	$p_{12}$	$p_{1\bullet}$
$X_2$ .....	$p_{21}$	$p_{22}$	$p_{2\bullet}$
R a z e m .....	$n_{\bullet 1}$	$p_{\bullet 2}$	1

Ź r ó d ł o: jak przy tabl. 1.

<sup>1</sup> <http://www.utogim.eu/power.xlsm>.

## TESTY NIEZALEŻNOŚCI W TABLICY DWUDZIELCZEJ $2 \times 2$

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

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

oraz z rodziny statystyki, która przy prawdziwości hipotezy  $H_0$  dla tablicy dwudzielczej  $2 \times 2$  ma asymptotyczny rozkład chi-kwadrat z 1 stopniem swobody, dlatego są określane mianem „statystyka chi-kwadrat” (Cressie, Read, 1984). Jest to statystyka:

1)  $\chi^2$  Pearsona (Pearson, 1900):

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

2)  $G^2$  ilorazu wiarygodności (Sokal, Rohlf, 2012):

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

3)  $N$  Neymana (Neyman, 1949):

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

4)  $KL$  Kullbacka-Leiblera (Kullback, 1959):

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

5)  $FT$  Freemana-Tukeya (Freeman, Tukey, 1950):

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

6) *CR* Cressiego-Reada (Cressie, Read, 1984):

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

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

Korzystając z metody Monte Carlo, w pracy Sulewskiego i Motyki (2015a) porównano jakość „statystyk chi-kwadrat” wyrażoną w funkcji mocy, przy założeniu, że asymptotycznym rozkładem statystyki testowej, przy prawdziwości hipotezy  $H_0$ , jest rozkład chi-kwadrat. Badania wykazały, że jakość testu  $\chi^2$  Pearsona oraz *CR* Cressiego-Reada jest znacznie lepsza od pozostałych.

Liczebności  $n_{ij}$  ( $i, j = 1, 2$ ) w tablicy dwudzielczej  $2 \times 2$  są zmiennymi losowymi. Godnym uwagi jest fakt, że zmienne losowe będące odwrotnościami lub ilorazami zmiennych losowych mogą nie mieć wartości oczekiwanej i w konsekwencji także niektórych momentów wyższych rzędów. Rozkład chi-kwadrat ma wszystkie takie momenty. Badania symulacyjne potwierdziły, że statystyki (5) i (6) dla małych prób nie podlegają rozkładowi chi-kwadrat (Sulewski, 2014).

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 statystyki (4), (6) i (8) są różnego znaku i wzajemnie się rekompensują i tylko w przypadku bardzo dużych prób podlegają rozkładowi chi-kwadrat (Sulewski, 2014). Tabl. 3 przedstawia wartości poszczególnych składników „statystyki chi-kwadrat” wyznaczonych symulacyjnie dla tablicy  $2 \times 2$  i liczebności próby  $n = 50$ , gdy  $H_0$  o niezależności cech  $X$  i  $Y$  jest słuszna. Tablicę dwudzielczą  $2 \times 2$  wygenerowano metodą słupkową opisaną w pracy Sulewskiego (2015b) dla  $p_{ij} = 0,25$  ( $i, j = 1, 2$ ).

**TABL. 3. WARTOŚCI SKŁADNIKÓW „STATYSTYK CHI-KWADRAT”**

Składniki	$\chi^2$	$G^2$	$N$	$KL$	$FT$	$CR$
$i=1, j=1$ .....	0,836	-3,287	1,076	4,233	0,236	-2,017
$i=1, j=2$ .....	1,363	4,351	0,999	-3,189	0,290	3,223
$i=2, j=1$ .....	0,981	4,192	0,777	-3,321	0,218	3,024
$i=2, j=2$ .....	1,600	-2,792	2,798	4,881	0,519	-1,554
<b>Suma</b> .....	<b>4,78</b>	<b>2,464</b>	<b>5,65</b>	<b>2,604</b>	<b>1,263</b>	<b>2,676</b>

Źródło: jak przy tabl. 1.

### WYZNACZANIE WARTOŚCI KRYTYCZNYCH

W ostatnich kilkudziesięciu latach zaproponowano różne testy niezależności dla tablic dwudzielczych oraz opisano warunki ich stosowania szczególnie dla

małych prób, które najczęściej są przedmiotem badań statystycznych. W dobie coraz szybszych komputerów oraz z rozbudowaną pamięcią operacyjną można za pomocą stosownego oprogramowania znieść te ograniczenia i symulacyjnie wyznaczyć wartości krytyczne. Przy ich wyznaczaniu, gdy między cechami nie ma związku, zawartość tablic dwudzielczych  $2 \times 2$  o wielkości próby  $n$  generowano za pomocą metody słupkowej przyjmując  $p_{ij} = 0,25$  ( $i, j = 1, 2$ ) (Sulewski, 2015b).

Niech  $\varpi$  będzie jedną z rozpatrywanych siedmiu statystyk. Dla każdej tablicy  $2 \times 2$  obliczono  $r = 10^5$  wartości statystyki testowej  $\varpi$ , które następnie uporządkowano w kolejności rosnącej. Wartość krytyczną na poziomie istotności  $\alpha = 0,1$  wyznaczono na podstawie dziewiątego decyła ze wzoru:

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

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

Generowanie zawartości tablic dwudzielczych innymi metodami, przy założeniu, że między cechami nie ma związku, nie miało wpływu na uzyskane wielkości wartości krytycznych.

### MOC TESTU

Moc testu to prawdopodobieństwo odrzucenia hipotezy zerowej  $H_0$ , gdy nie jest ona prawdziwa. Moc testu zależy od liczebności próby — im liczniejsza, tym większa moc. Zależy ona także od poziomu istotności testu — im niższy jego poziom, tym mniejsza moc testu oraz również od siły związku między cechami — im większa jest jego siła, tym większa jest moc testu. Więcej informacji o mocy testu można znaleźć w opracowaniu Sulewskiego (2015c).

W celu wyznaczenia mocy testu, czyli zdolności tablicy dwudzielczej do odrzucenia hipotezy mówiącej o tym, że nie ma związku między cechami  $X$  i  $Y$ , gdy w istocie związek jest, niezbędna staje się generacja tablic dwudzielczych. Dane podlegające opracowaniu muszą pochodzić z generatora liczb losowych i nie mogą być wzięte z praktyki. Uwzględniając narzuconą siłę związku między cechami, uzyskaną dzięki generatorowi liczb losowych i wyrażoną za pomocą miary nieprawdziwości  $H_0$  danej wzorem:

$$mn = \sum_{i=1}^2 \sum_{j=1}^2 |p_{ij} - p_{i\bullet} \cdot p_{\bullet j}| \quad (10)$$

do wypełnienia tablicy dwudzielczej  $2 \times 2$  skorzystano z metody „słupkowej” wykorzystującej prawdopodobieństwa  $p_{ij}$  ( $i, j = 1, 2$ ) (Sulewski, 2015b).

Znając wartość statystyki testowej  $\varpi$  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  $\varpi$  jest nie mniejsza od wartości krytycznej  $cv_\alpha$ .

Hipoteza zerowa  $H_0$  — mówiąca o tym, że między cechami  $X$  i  $Y$  w tablicy dwudzielczej  $2 \times 2$  nie ma związku — jest słuszna, gdy  $p_{ij} = p_{i\cdot} \cdot p_{\cdot j}$  dla  $i, j = 1, 2$ . Zatem miara (10) przyjmuje wartość 0, gdy hipoteza  $H_0$  jest słuszna. Im większe wartości  $mn$ , tym większa jest możliwość fałszywego  $H_0$ . Moc testu wyznaczono ze schematów A, B, C na podstawie prawdopodobieństwa  $p_{ij}$  ( $i, j = 1, 2$ ), dla którego hipoteza zerowa  $H_0$  jest niesłuszna (tabl. 4). Tabl. 5 przedstawia przedziały zmienności miary  $mn$  dla omawianych schematów prawdopodobieństwa.

**TABL. 4. PRAWDOPODOBIEŃSTWO UKAZUJĄCE ZWIĄZEK MIĘDZY CECHAMI X I Y WEDŁUG SCHEMATÓW A, B I C (przy założeniu  $p_o = 0,25$ ,  $\Delta p = 0,025$ ,  $k = 0,1, \dots, 9$ )**

Cecha X	Cecha Y	
	$Y_1$	$Y_2$
<b>Schemat A</b>		
$X_1$ .....	$p_o - k \cdot \Delta p$	$p_o$
$X_2$ .....	$p_o$	$p_o + k \cdot \Delta p$
<b>Schemat B</b>		
$X_1$ .....	$p_o - k \cdot \Delta p$	$p_o$
$X_2$ .....	$p_o + k \cdot \Delta p$	$p_o$
<b>Schemat C</b>		
$X_1$ .....	$p_o - k \cdot \Delta p$	$p_o + k \cdot \Delta p$
$X_2$ .....	$p_o + k \cdot \Delta p$	$p_o - k \cdot \Delta p$

Ź r ó d ł o: jak przy tabl. 1.

**TABL. 5. MIARA NIEPRAWDZIWOŚCI HIPOTEZY  $H_0$  DLA PRZEDSTAWIONYCH SCHEMATÓW**

$k$	Miara $mn$		
	schemat A	schemat B	schemat C
0 .....	0	0	0
1 .....	0,0025	0,05	0,1
2 .....	0,01	0,10	0,2
3 .....	0,0225	0,15	0,3
4 .....	0,04	0,20	0,4
5 .....	0,0625	0,25	0,5
6 .....	0,09	0,30	0,6

**TABL. 5. MIARA NIEPRAWDZIWOŚCI HIPOTEZY  $H_0$   
DLA PRZEDSTAWIONYCH SCHEMATÓW (dok.)**

$k$	Miara $mn$		
	schemat A	schemat B	schemat C
7 .....	0,1225	0,35	0,7
8 .....	0,16	0,40	0,8
9 .....	0,2025	0,45	0,9

Ź r ó d ł o: jak przy tabl. 1.

Z tabl. 5 wynika, że miara nieprawdziwości hipotezy  $H_0$  zmienia się w różnych przedziałach. Minimalną liczebność próby dla danego schematu tak dobrano, aby liczebności oczekiwane  $e_{ij}$  ( $i, j = 1, 2$ ) były różne od zera. Z kolei maksymalną wartość próby ustalono tak, aby uzyskać maksymalną moc testu.

Wykr. 1 przedstawia zależność mocy testów niezależności od miary nieprawdziwości  $H_0$  dla schematu prawdopodobieństwa A na poziomie istotności  $\alpha = 0,1$  i liczebności próby  $n = 50 \cdot i$  ( $i = 1, 2, \dots, 4$ ).

Wykr. 2 prezentuje natomiast zależność mocy testów niezależności od miary nieprawdziwości  $H_0$  dla schematu prawdopodobieństwa B na poziomie istotności  $\alpha = 0,1$  i liczebności próby  $n = 10 \cdot i + 15$  ( $i = 1, 2, \dots, 4$ ).

Wykr. 3 z kolei przedstawia zależność mocy testów niezależności od miary nieprawdziwości  $H_0$  dla schematu prawdopodobieństwa C na poziomie istotności  $\alpha = 0,1$  i liczebności próby  $n = 5 \cdot i + 15$  ( $i = 1, 2, \dots, 4$ ).

Dla statystyki modułowej  $|\chi|$  oraz statystyki  $\chi^2$  Pearsona jako reprezentanta „statystyki chi-kwadrat” uzyskane wyniki porównano za pomocą testu o równości prawdopodobieństwa stawiając hipotezę zerową, że moce testów są równe. Symbol „+” oznacza, że różnice między mocami testów są statystycznie istotne, natomiast symbol „-”, iż są statystycznie istotne (tabl. 6—8).

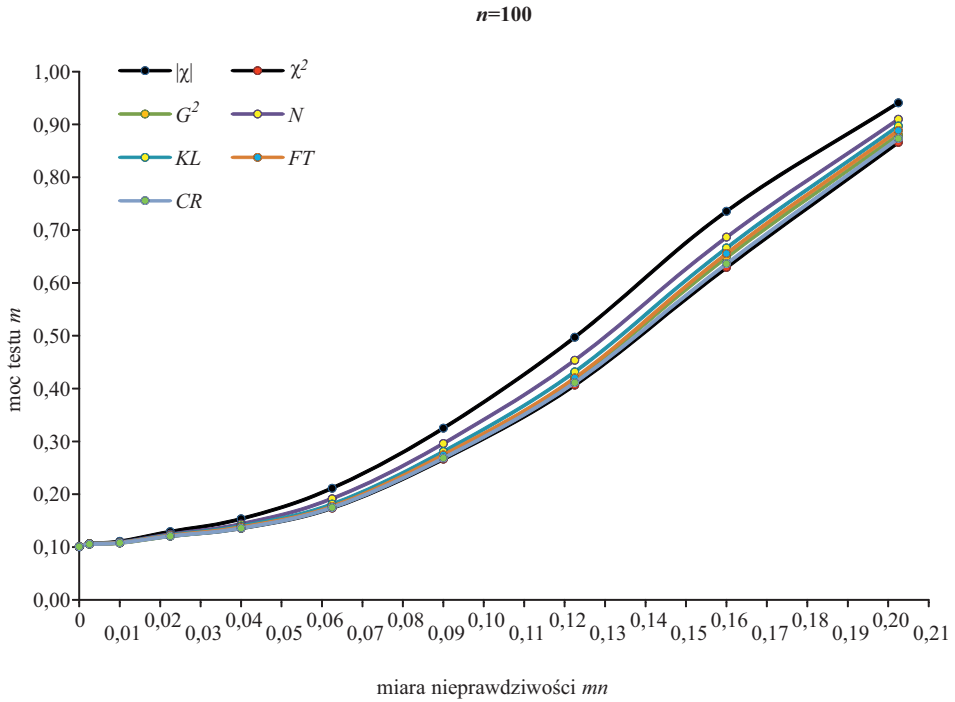
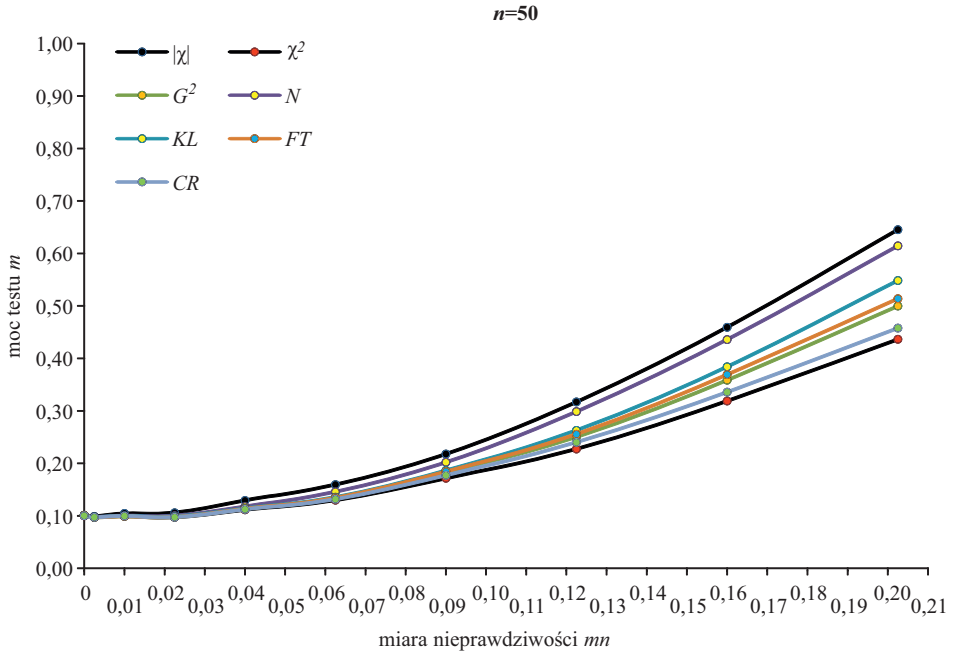
**TABL. 6. ISTOTNOŚĆ STATYSTYCZNA MOCY TESTÓW  $|\chi|$  I  $\chi^2$  DLA SCHEMATU A**

$mn$	$n=50$	$n=100$	$n=150$	$n=200$
0 .....	+	+	+	+
0,0025 .....	+	+	+	+
0,01 .....	+	+	+	+
0,0225 .....	+	+	+	+
0,04 .....	+	+	+	+
0,0625 .....	-	-	-	-
0,09 .....	-	-	-	-
0,1225 .....	-	-	-	-
0,16 .....	-	-	-	-
0,2025 .....	-	-	-	-

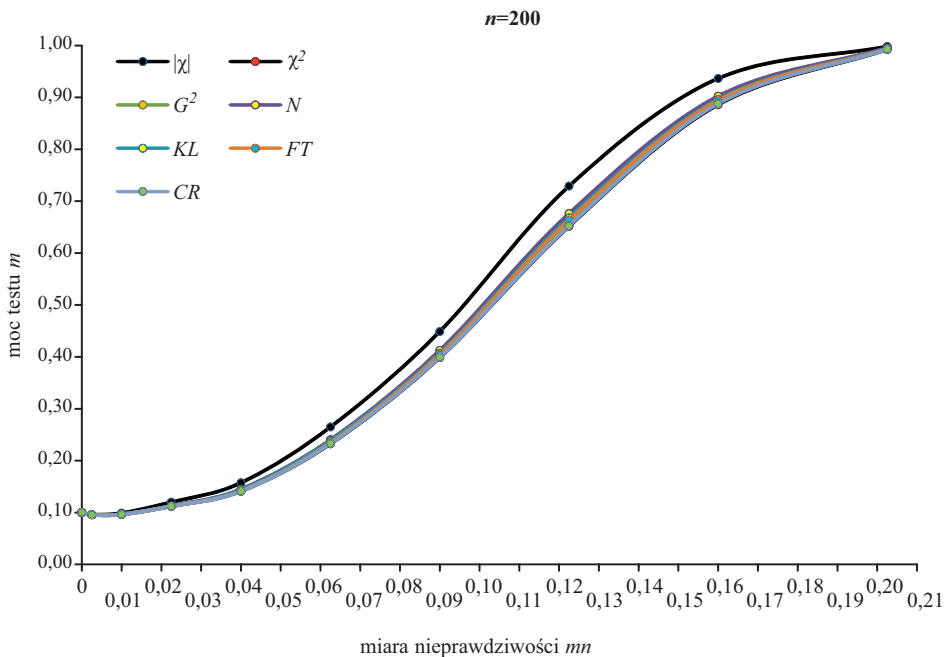
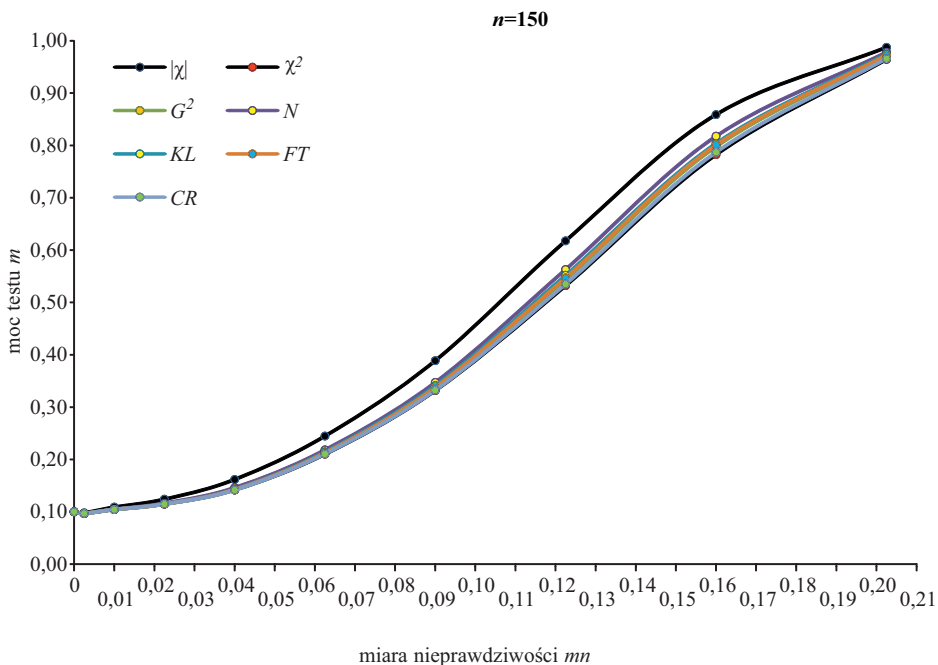
Ź r ó d ł o: jak przy tabl. 1.



**Wykr. 1. MOC TESTÓW NIEZALEŻNOŚCI DLA SCHEMATU A WEDŁUG LICZEBNOŚCI PRÓBY**

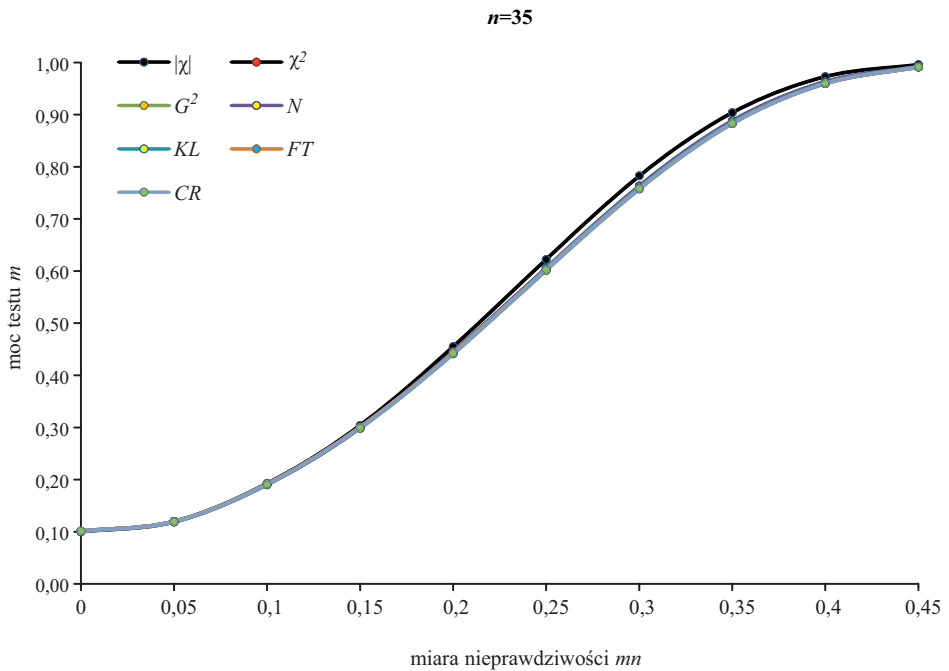
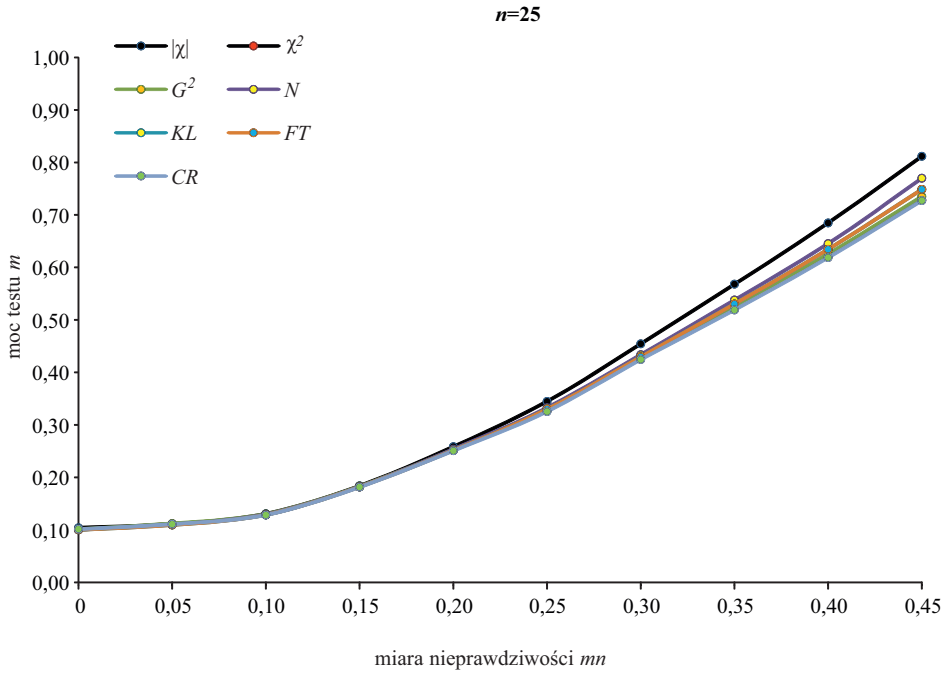


**Wykr. 1. MOC TESTÓW NIEZALEŻNOŚCI DLA SCHEMATU A WEDŁUG LICZEBNOŚCI PRÓBY (dok.)**

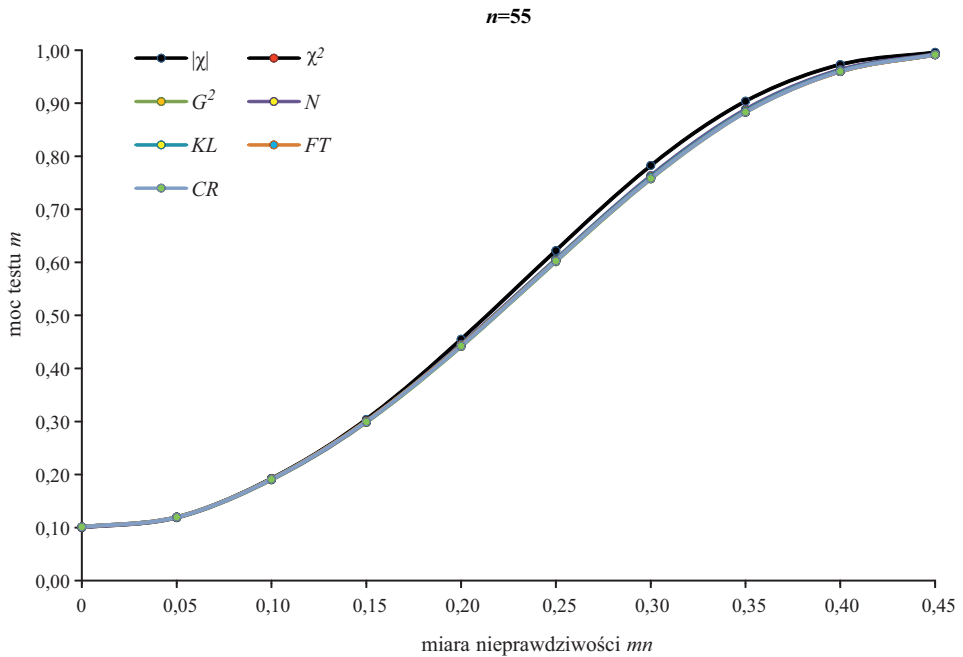
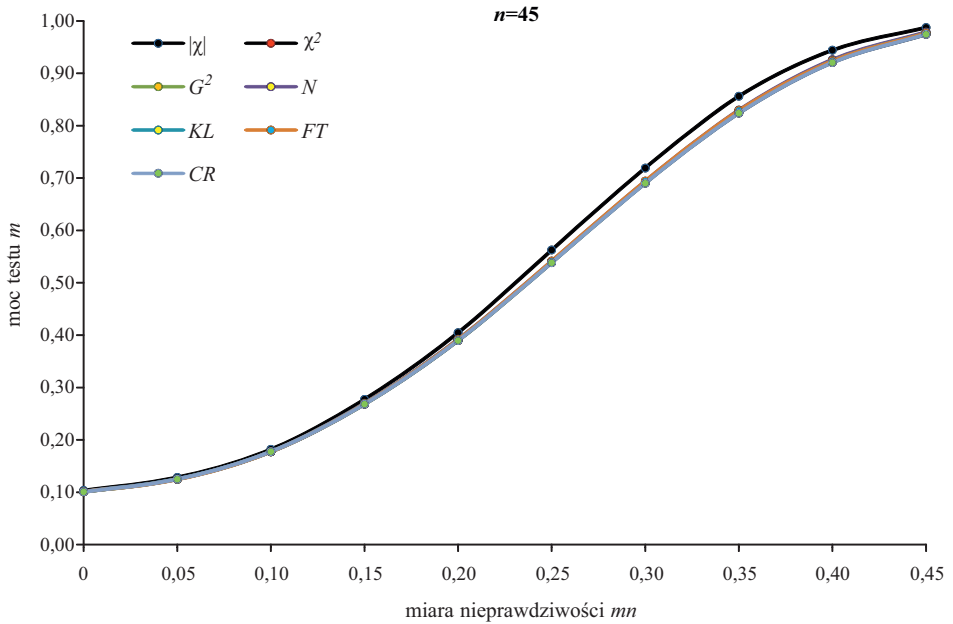


Ź r ó d ł o: jak przy tabl. 1.

**Wykr. 2. MOC TESTÓW NIEZALEŻNOŚCI DLA SCHEMATU B WEDŁUG LICZEBNOŚCI PRÓBY**

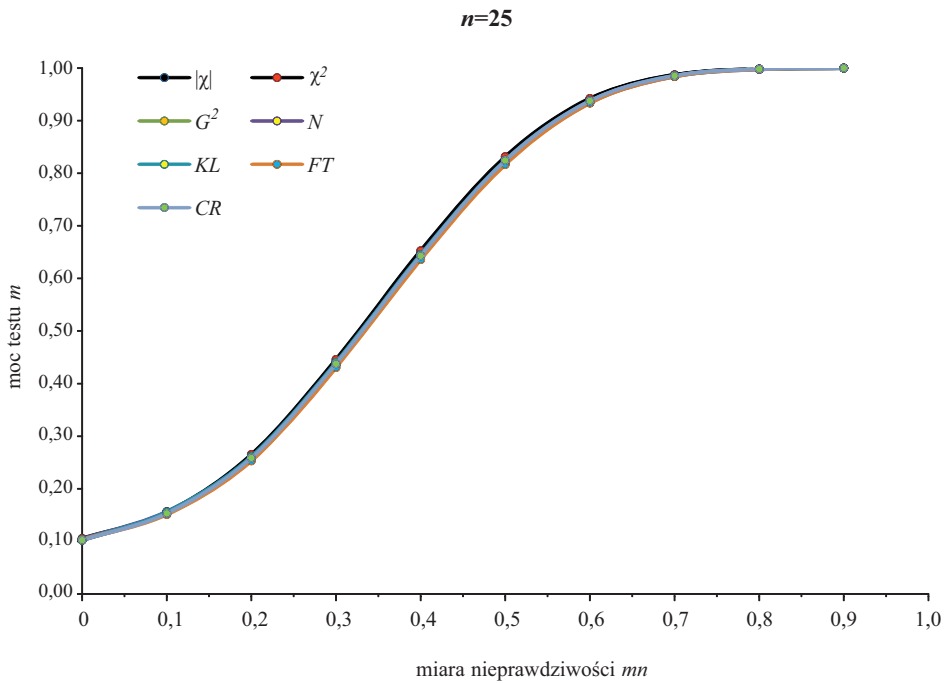
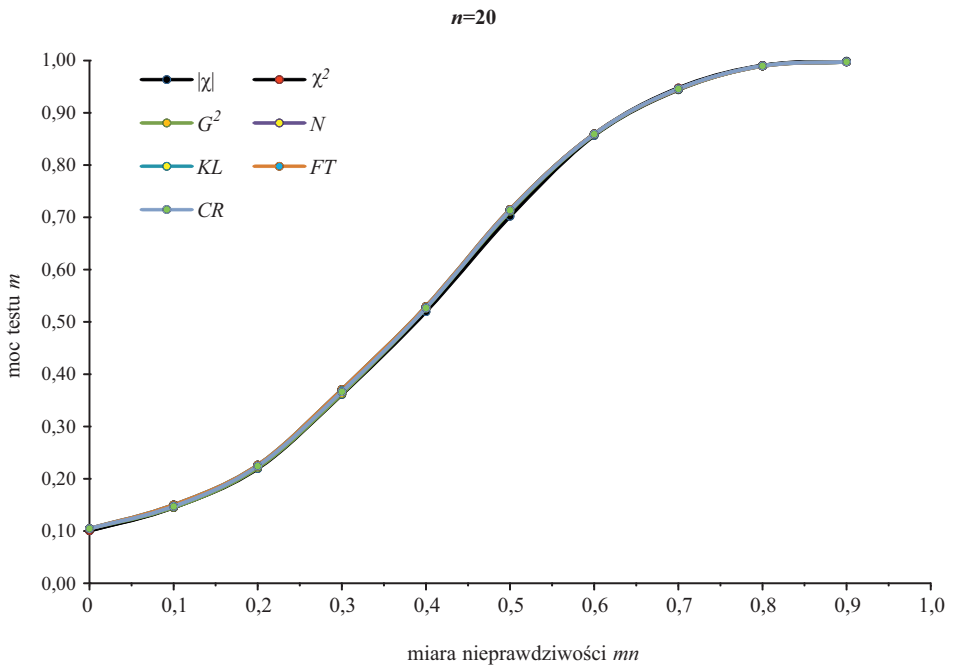


**Wykr. 2. MOC TESTÓW NIEZALEŻNOŚCI DLA SCHEMATU B WEDŁUG LICZEBNOŚCI PRÓBY (dok.)**

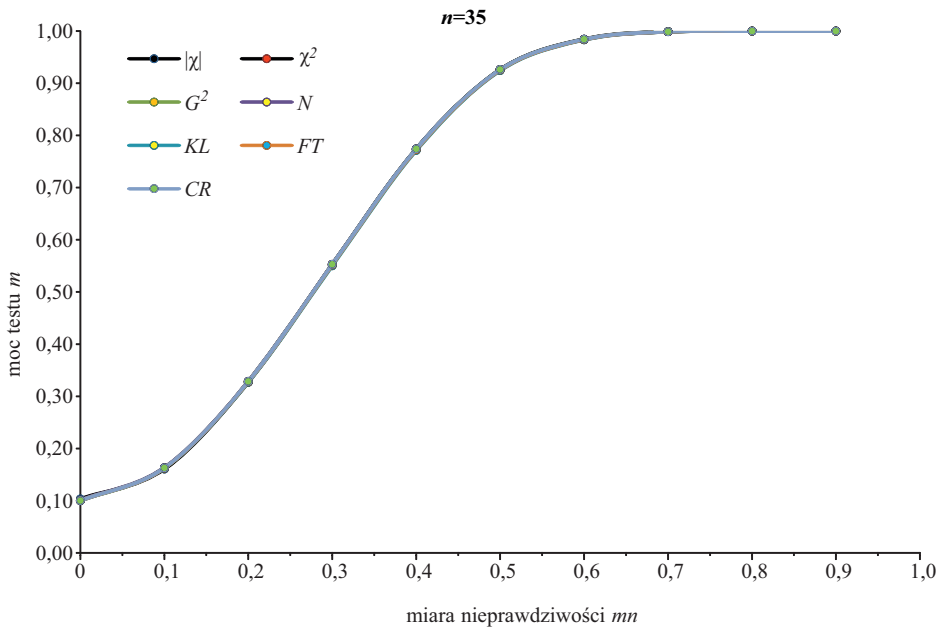
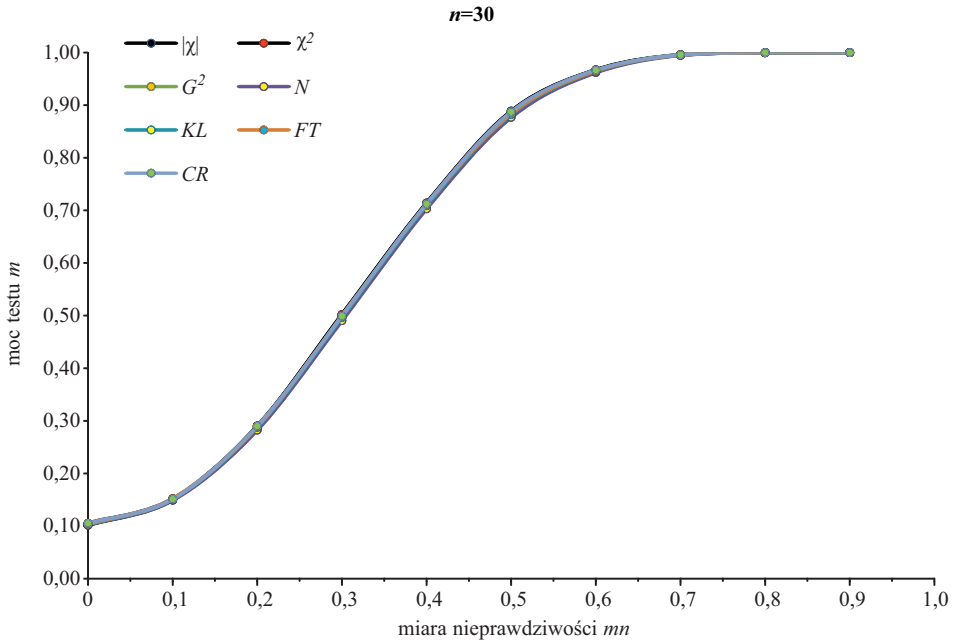


Źródło: jak przy tabl. 1.

**Wykr. 3. MOC TESTÓW NIEZALEŻNOŚCI DLA SCHEMATU C WEDŁUG LICZEBNOŚCI PRÓBY**



Wykr. 3. MOC TESTÓW NIEZALEŻNOŚCI DLA SCHEMATU C  
WEDŁUG LICZEBNOŚCI PRÓBY (dok.)



Źródło: jak przy tabl. 1.

**TABL. 7. ISTOTNOŚĆ STATYSTYCZNA MOCY TESTÓW  $|\chi|$  I  $\chi^2$  DLA SCHEMATU B**

<i>mn</i>	<i>n</i> =25	<i>n</i> =35	<i>n</i> =45	<i>n</i> =55
0 .....	+	+	+	+
0,05 .....	+	+	+	+
0,10 .....	+	+	+	+
0,15 .....	+	+	+	+
0,20 .....	+	+	+	+
0,25 .....	+	+	+	+
0,30 .....	+	+	+	+
0,35 .....	-	-	-	+
0,40 .....	-	-	-	+
0,45 .....	-	-	-	+

Ź r ó d ł o: jak przy tabl. 1.

**TABL. 8. ISTOTNOŚĆ STATYSTYCZNA MOCY TESTÓW  $|\chi|$  I  $\chi^2$  DLA SCHEMATU C**

<i>mn</i>	<i>n</i> =20	<i>n</i> =25	<i>n</i> =30	<i>n</i> =35
0 .....	+	+	+	+
0,1 .....	+	+	+	+
0,2 .....	+	+	+	+
0,3 .....	+	+	+	+
0,4 .....	+	+	+	+
0,5 .....	+	+	+	+
0,6 .....	+	+	+	+
0,7 .....	+	+	+	+
0,8 .....	+	+	+	+
0,9 .....	+	+	+	+

Ź r ó d ł o: jak przy tabl. 1.

Z wyk. 1—3 wynika, że testy niezależności wykorzystujące „statystyki chi-kwadrat” we wszystkich schematach prawdopodobieństwa charakteryzują się podobną mocą dla danej liczebności próby *n* i miary *mn*. Dla schematu C test wykorzystujący statystykę modułową ma podobną moc, jak pozostałe testy niezależności. Dla schematów A i B test wykorzystujący statystykę modułową ma większą moc niż testy związane ze „statystykami chi-kwadrat”. Szczególnie jest to widoczne w sytuacji, gdy miara nieprawdziwości  $H_0$  zmienia się według schematu A, czyli przyjmuje najmniejsze wartości.

## Podsumowanie

Artykuł pokazuje, że testy niezależności wykorzystujące „statystyki chi-kwadrat” we wszystkich schematach prawdopodobieństwa charakteryzują się podobną mocą dla danej liczebności próby *n* i miary *mn*. Gdy ma miejsce schemat C i miara nieprawdziwości  $H_0$   $MN = 0,1 \cdot i$  ( $i = 0,1, \dots, 9$ ), test wykorzystujący statystykę modułową ma podobną moc, jak pozostałe testy niezależności.

Dla schematów A i B test wykorzystujący statystykę modułową ma większą moc niż testy związane ze „statystyką chi-kwadrat”. Szczególnie jest to widoczne wtedy, gdy miara nieprawdopodobieństwa  $H_0$  zmienia się według schematu A.

Dzięki gotowej implementacji komputerowej czytelnik może samodzielnie badać niezależność cech w tablicy dwudzielczej  $2 \times 2$  za pomocą statystyki modułowej, która spośród omawianych testów charakteryzuje się największą mocą. Na korzyść tej statystyki przemawia także fakt, że można z niej skorzystać w sytuacji, gdy jedna komórka lub dwie leżące na przekątnej są puste. Statystyka  $G^2$  ilorazu wiarygodności,  $N$  Neymana i  $KL$  Kullbacka-Leiblera nie spełniają tego założenia.

---

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, s. 3661—3675.
- Cochran W. G. (1952), *The  $\chi^2$  test of goodness of fit*, „*Annals of Mathematical Statistics*”, Vol. 25, s. 315—345.
- Cressie N., Read T. (1984), *Multinomial Goodness-of-Fit Tests*, „*Journal of the Royal Statistical Society*”, Series B (Methodological), Vol. 46, No. 3, s. 440—464.
- Freeman M. F., Tukey J. W. (1950), *Transformations related to the angular and the square root*, „*Annals of Mathematical Statistics*”, Vol. 21, s. 607—611.
- Kullback S. (1959), *Information Theory and Statistics*, Wiley, New York.
- Neyman J. (1949), *Contributions to the theory of the  $\chi^2$  test*. In: Neyman, J. (Ed.), *Proceedings of the First Berkeley Symposium on Mathematical Statistics and Probability*. University of California Press, Berkeley, s. 239—273.
- Pearson E. (1947), *The choice of statistical tests illustrated on the interpretation of data classed in a  $2 \times 2$  table*, „*Biometrika*”, Vol. 34, s. 139—167.
- 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*”, Series 5, Vol. 50, s. 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*”, nr 10, s. 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*, „*Zeszyty Naukowe Akademii Marynarki Wojennej*”, nr 1, s. 37—46.
- Sulewski P. (2015b), *Wyznaczenie obszaru krytycznego przy testowaniu niezależności w tablicach wielodzielczych*, „*Wiadomości Statystyczne*”, nr 3, s. 1—18.
- Sulewski P. (2015c), *Ocena zdolności tablic dwudzielczych do wykrywania związku między uporządkowanymi cechami typu jakościowego*, „*Wiadomości Statystyczne*”, nr 5, s. 1—16.



**Summary.** *In the statistical literature there are many test measures to study the independence of features in the two-way contingency tables. For statistical analysis, the family of six so-called "chi-squared statistic" was selected — including Pearson's  $\chi^2$  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 Monte Carlo methods. 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 method.

**Резюме.** *В статистической литературе существует много испытательных тестов для обследования независимости показателей в таблицах сопряженности 2x2. В представляемой разработке для статистического анализа были выбраны так называемая «статистика хи-квадрат», в том числе статистика  $\chi^2$  Пирсона, а также модульная статистика по выбору автора. Для того, чтобы освободиться от ограничений касающихся «статистики хи-квадрат», критические значения для всей проанализированной статистики были определены с помощью методов моделирования Монте-Карло. Для сравнения тестов была предложена мера неправильности  $H_0$ , а также установлено мощность критерия, то есть способность таблицы сопряженности 2x2 отвергнуть нулевую гипотезу  $H_0$  говорящую о том, что между признаками  $X$  и  $Y$  нет связи.*

**Ключевые слова:** таблица сопряженности 2x2, тест независимости, критические значения, метод Монте-Карло.