Folia Oeconomica
Acta Universitatis Lodziensis
ISSN 0208-6018 e-ISSN 2353-7663
www.czasopisma.uni.lodz.pl/foe/
4(330) 2017

DOI: http://dx.doi.org/10.18778/0208-6018.330.04

Piotr Sulewski
Pomeranian University in Słupsk, Faculty of Mathematics and Natural Sciences,
Institute of Mathematics, piotr.sulewski@apsl.edu.pl

A New Test for Independence in 2×2 Contingency Tables

Abstract: In statistical literature there exist many tests to reveal the independence of two qualitative variables in two-way contingency tables (CTs), in particular in 2×2 CTs. In this paper four independence tests were compared. These are: the chi-square test, being the most popular type of power divergence statistics; the modular test and the d-square test, which is a modification of the Pearson's test; the logarithmic minimum test which is a new proposal. Critical values for the tests listed above were determined with the Monte Carlo method. In order to compare the tests, the measure of untruthfulness of H_0 was proposed and the power of the tests was calculated.

Keywords: independence test, 2×2 contingency table, logarithmic minimum statistics, modular statistics, power divergence statistics, Monte Carlo method

JEL: C12, C14, C15, C46, C63

1. Introduction

The analysis of contingency tables (CTs) is one of the most common tasks performed by statisticians. CTs display the frequency distribution of two (two-way CTs) or more (multi-way CTs) categorical variables. Information presented as CTs features in a wide variety of areas such as the social sciences (Wickens, 1969), genetics (El Galta et al., 2008; Dickhaus et al., 2012), demography (Cung, 2013) and psychology (Iossifova et al., 2013).

Independence tests are probably one of the most commonly used statistical tools. Test data are arranged in the form of CTs, in particular 2×2 CTs. Well-known and commonly used are the (Pearson's) χ^2 test and the log likelihood ratio G^2 test. Garside and Mack (1976) compared the sizes of the χ^2 testand some of its corrected versions numerically. Authors noted that though the corrected versions are conservative in nature, the χ^2 test has the size closest to the nominal level α. For small CTs (not applicable 2×2) with small sample sizes, Lawal, Uptong, 1984) suggested a modification to the χ^2 test to make the size closer to the nominal level α. Numerous publications on CTs and the χ^2 test of independence were proposed, one can see e.g. (Meng, Chapman, 1966; Diaconis, Efron, 1985; Albert, 1990; Andrés et al., 1995), where the χ^2 test statistics was interpreted from various angles. Information about chi-square approximations of χ^2 and G^2 can be found in (Cochran, 1952; 1954; Koehler, Larntz, 1980; Cressie, Read, 1989). The χ^2 and G^2 tests provide consistent and asymptotically unbiased tests of independence (Haberman, 1981). These test statistics belong to the power divergence statistics (PDS) (Cressie, Read, 1984)

The Fisher exact test (Fisher, 1922) is also popular, independently developed by Irwin (1935) and also known as the Fisher-Irwin (FI) test. The FI test is most commonly applied to 2×2 CTs because it can be computationally time consuming for tables bigger than 2×2. Campbell (2007) recommended the use of the χ^2 test for large sample sizes and the FI test for small sample sizes. This test is also criticized for being too conservative and hence having lower power. Lydersen et al. (2009) recommended that the FI test should practically never be used.

CTs having very small or no cell counts are said to be sparse. Sparse CTs often containing cells with zero cell counts are of two types: sampling zero ($n^*_{ij} = 0$, $p_{ij} > 0$) and structural zero ($n^*_{ij} = 0$, $p_{ij} = 0$) . This article assumes, that cells count in CTs are positive, which means $n^*_{ij} = 0$, ($i, j = 1,2$).

Some researchers have investigated the test methods when there exists a natural ordering among the X_i values and Y_j values (Tab. 1). For details please consult e.g. (Agresti, 2002). In this paper, features X and Y are nominal in nature.

In hypothesis testing, the theory of bootstrap is well developed and provides a reasonably good answer in many parametric problems where a consensus is hard to reach (Hall, Wilson, 1991; Chang, Pal, 2008; Chang et al., 2011).

Statistical science has been enriched with many papers offering statistics related to testing independence. Haber (1987) compared the two-sided FI test with six nonrandomized unconditional exact tests with respect to their power. Zelterman (1987) proposed the D^2 test, which is another adjustment to the χ^2 test. Berry and Mielke (1988) used Monte Carlo methods to assess the relative fit of two asymptotic χ^2 tests, two asymptotic G^2 tests, and a recently developed nonasymptotic χ^2 test to the models specified by the null hypotheses of independence and homogeneity. Results of the study indicate that the nonasymptotic χ^2 test is superior in overall performance to the other four tests.

Lawal and Uptong (1990) compared the PDS with the modified χ^2 test statistics (Lawal, Uptong, 1984) by means of the statistical power. Cohen and Nee (1990) used the Monte Carlo methods and calculated the statistical power using the Rao F-test in CTs. Davis (1993) described a generalized chi-square approximation to the distribution of the χ^2 test statistics for testing independence in CTs. The new method consistently yields estimated p-value, which agrees closely with the exact result. In (Jeong et al., 2005), for the analysis of CTs having ordered row categories and ordered column categories, a bootstrap method was applied for the model-based G^2 test for independence. Taneichi and Sekiya (2007) considered a class of test statistics C_ϕ based on ϕ-divergence for the test of independence in CTs. The class of test statistics C_ϕ includes the test statistics R^a based on the PDS as a special case. The research shows that the transformed χ^2 test performs very well. Ceyhan (2010) compared the directional (i.e. one-sided) versions of the cell-specific nearest neighbor contingency tables (NNCT) tests with new directional NNCT tests for the two-class case using Monte Carlo simulations and statistical power. Yenigün et al. (2011) carried out a simulation study to see the empirical power performance of the maximal correlation test and compared it with χ^2 and G^2 tests of independence. When the underlying continuous variables are uncorrelated but dependent, authors pointed out some cases for which the maximal correlation test appears to be more powerful. Nandram et al. (2013) considered a G^2 test for quasi-independence in large CTs which are likely to have both structural and sampling zeros. A new procedure requires at least one sampling zero, and is an alternative to the commonly used ad hoc procedures of converting the zero cells to positive ones by adding a small constant (Clogg, Eliason, 1987; Beh, Farver, 2009). One drawback of the new procedure is that it is a conditional (on the set of positive cells) test but it is not conditional on the margins as in the FI test. It is also true that T^2 is an exact test and it does not rely on asymptotic theory. Yu (2014) allows the margins to be random and compares the power of the G^2 test, the Bayes factor test, and the FI test. Egozcue et al. (2015) examine the independence in CTs using simplicial geometry. Shan and Wilding (2015) extends the unconditional approach based on estimation and maximization to designs with the total sum fixed. The procedures based on the χ^2 test statistics, Yates's corrected and G^2 test statistics are evaluated with regard to actual type I error rates and powers. Lipsitz et al. (2015) propose Wald and score test statistics for independence based on weighted least squares estimating equations. In contrast to the Rao-Scott test statistics, the proposed Wald and score test statistics always exist. Comparing the Rao-Scott test statistics, the score and Wald statistics with respect to power, one can see that the Wald test statistics had the highest power. Lin et al. (2015) explore the accuracy of the χ^2 and G^2 tests through an extensive simulation study and then propose their bootstrap versions that appear to work better than the asymptotic tests. The bootstrap tests are useful even for small-cell frequencies as they maintain the nominal level quite accurately. Also, the proposed bootstrap tests are more convenient than the FI test which is often criticized for being too conservative. Vélez et al. (2016) propose and illustrate a new graphical method to perform diagnostic analyses in two-way CTs. In this method, one observation is added or removed from each cell at a time, whilst the other cells are held constant, and the change in a test statistics of interest is graphically represented. García and González-López (2016) present a new non-parametric independence test for sparse data, which is a generalization of the LIS test (Garcia, González-López, 2014) for the hypothesis of independence between two continuous random variables. The new test does not require the assumption of continuity for the random variables. This test is applied to two datasets and also compared with the χ^2 test. Lin et al. (2015) show, that a simple bootstrap version of the existing asymptotic tests can correct the size problems even for small sample sizes without going through the restrictive adjustment already reported in literature. In biomedical studies, ordered categorical variables are frequently encountered. Hui-Qiong et al. (2016) introduce a simple ordering test approach for the two-way r×c CTs with incomplete counts. The results show that the G^2 test statistics based on the bootstrap resampling methods performs satisfactorily from small to large sample sizes. The modular test has been compared with the family of PDS on 2×2 CTs (Sulewski, 2016a) and on CTs bigger than 2×2 (Sulewski, 2016b).

In this paper the new logarithmic minimum statistics (LMS) is proposed, which was compared with three other statistics. The first one is the well-known and commonly used χ^2 test statistics (Pearson, 1904) representing the PDS. The second one is the D^2 test statistics (Zelterman, 1987), which is an adjustment to the χ^2 test statistics. The third one is the modular statistics (Sulewski, 2016a), which is another modification of the χ^2 test statistics. To use the smaller samples size, critical values were determined by means of the Monte Carlo simulation method. Lin et al. (2015) also determined the critical values of the χ^2 test using the Monte Carlo method. For the above test statistics, the power-of-the-test (PoT) function was determined. To calculate the power of the test, 2×2 CTs were generated by means of the bar method and the measure of untruthfulness of H_0 (MoU) for five probability scenarios were proposed.

2. Recalling 2×2 contingency table basics

Let X and Y be two features of the same object having levels X_1, X_2 and Y_1, Y_2. Testing for independence of these two features with appropriately arranged CT and χ^2 statistics applied is probably one of the most common tasks for statisticians. There are n items classified with respect to X and Y. This produces a table of a pattern shown below as Table 1 (EC Variant), where n^*_{ij} are counts of objects classified as belonging to the cell (X_i, Y_j) and $n^*_{1\bullet}$, $n^*_{2\bullet}$, $n^*_{\bullet 1}$, $n^*_{\bullet 2}$ are row and column marginal counts, respectively. Table 1 (EC Variant) is a basis for a test hypothesis that is commonly called the main and denoted H_0.

Table 1. Variants of presentation of 2×2 CTs

TP Variant	Y_1	Y_2	Total
X_1	p_{11}	p_{12}	$p_{1\bullet}$
X_2	p_{21}	p_{22}	$p_{2\bullet}$
Total	$p_{\bullet 1}$	$p_{\bullet 2}$	$p_{\bullet\bullet} = 1$

EC Variant	Y_1	Y_2	Total
X_1	n^*_{11}	n^*_{12}	$n^*_{1\bullet}$
X_2	n^*_{21}	n^*_{22}	$n^*_{2\bullet}$
Total	$n^*_{\bullet 1}$	$n^*_{\bullet 2}$	$n^*_{\bullet\bullet} = n$

TC Variant	Y_1	Y_2	Total
X_1	$n_{11} = np_{11}$	$n_{12} = np_{12}$	$n_{1\bullet}$
X_2	$n_{21} = np_{21}$	$n_{22} = np_{22}$	$n_{2\bullet}$
Total	$n_{\bullet 1}$	$n_{\bullet 2}$	$n_{\bullet\bullet} = n$

EP Variant	Y_1	Y_2	Total
X_1	$p^*_{11} = n^*_{11} / n$	$p^*_{12} = n^*_{12} / n$	$p^*_{1\bullet}$
X_2	$p^*_{21} = n^*_{21} / n$	$p^*_{22} = n^*_{22} / n$	$p^*_{2\bullet}$
Total	$p^*_{\bullet 1}$	$p^*_{\bullet 2}$	$p^*_{\bullet\bullet} = 1$

Source: own material

These are details of particular variants:

TP Variant (theoretical probabilities). Cells contain:

1) probabilities intrinsic to the phenomenon that is of our interest. Exact values of these probabilities are unknown to the investigator, or
2) probabilities of which values are arbitrarily set by the Monte Carlo experimenter.

EC Variant (experimental counts). Cells contain counts observed on a sample drawn from the general population being a subject of investigation.

TC Variant (theoretical counts). Cells contain the expected counts based on the Monte Carlo experiment.

EP Variant (empirical probabilities). Cells contain estimates of the unknown content of TP.

3. Scenarios of generation of 2×2 contingency tables

One can treat CTs as a mathematical expression of a certain phenomenon we cope with. This formulation suggests that there is some internal mechanism in this phenomenon that determines probabilities of particular XY combinations and ascribes these probabilities to cells of the table. Below is a “progenitor” of all the 2×2 CTs:

$$T_p = \begin{bmatrix} 0.25 & 0.25 \\ 0.25 & 0.25 \end{bmatrix}.$$

The variety of tables may be generated when portions of probability in quantity of a flow from “maternal” cells of T_p to other cells. In this paper five scenarios were developed that seem fundamental (Tab. 2).

Table 2. Scenarios of generation of 2×2CTs ($a = k \cdot 0.25 \cdot 10^{-3}$, $k = 0, 1, \ldots, 1000$)

Scenario I

	Y_1	Y_2
X_1	0.25	$0.25-a/2$
X_2	$0.25+a$	$0.25-a/2$

Scenario II

	Y_1	Y_2
X_1	$0.25-a$	0.25
X_2	0.25	$0.25+a$

Scenario III

	Y_1	Y_2
X_1	0.25	$0.25+a/2$
X_2	$0.25-a$	$0.25+a/2$

Scenario IV

	Y_1	Y_2
X_1	$0.25-a$	0.25
X_2	$0.25+a$	0.25

Scenario V

	Y_1	Y_2
X_1	$0.25-a$	$0.25+a$
X_2	$0.25+a$	$0.25-a$

Source: own material

In all above scenarios the flow portion $0 \leq a \leq 0.25$. Scenarios may be locally mutated by reversing rows or columns to better fit to the data analyzed. Surely, these are not unique. One may anticipate a variety of modifications as it is common in statistics. The others differ in degree of untruthfulness of H_0 which will be one of the subjects of the next section.

Each scenario has a multinomial distribution of its own. Particular formulas are easy to obtain by substituting probabilities in formula of the multinomial distribution with probabilities taken from relevant cells of CTs. For instance, the scenario V has the following distribution, obtained after simple transformations:

$$P_0(n_{11}, n_{12}, n_{21}, n_{22}) = \frac{n!}{4^n \cdot n_{11}! n_{12}! n_{21}! n_{22}!} \cdot (1-4a)^{n_{11}+n_{22}} \cdot (1+4a)^{n_{12}+n_{21}}. \quad (1)$$

4. The measure of untruthfulness of H_0

As it has already been stated in Section 2 particular classes of feature X are ascribed to rows and these of feature $Y$are ascribed to columns. Features X, Y are independent and H_0 is true, if

$$p_{ij} = p_{i\bullet} \cdot p_{\bullet j}. \tag{2}$$

When equality (2) is not fulfilled, H_0 is not true and an appropriate measure of untruthfulness of H_0 (MoU) is needed. In this paper the following MoU is put forward:

$$MoU = \sum_{i=1}^{2}\sum_{j=1}^{2}\left|p_{ij} - p_{i\bullet} \cdot p_{\bullet j}\right|. \tag{3}$$

It doubtlessly springs from essence of H_0 and seems to be of a very simple form.

Replacing theoretical probabilities by empirical ones (Tab. 1 EP Variant) we get the sample MoU as

$$MoU_e = \frac{1}{n}\sum_{i=1}^{2}\sum_{j=1}^{2}\left|n_{ij}^{*} - \frac{n_{i\bullet}^{*} \cdot n_{\bullet j}^{*}}{n}\right| \tag{4}$$

Table 3 shows values of MoU at the points $k = 0, 1, \ldots, 1000$ (see Tab. 2), where the power of the test was determined. These values show that the MoU – for all probabilities scenarios – is changing with the constant step.

Table 3. Values of MoU under scenarios I–V

k	I	*k*	II	*k*	III	*k*	IV	*k*	V
0	0	0	0	0	0	0	0	0	0
51	0.0125	316	0.025	140	0.0375	100	0.05	100	0.1
106	0.025	448	0.05	265	0.075	200	0.1	200	0.2
163	0.0375	548	0.075	378	0.1125	300	0.15	300	0.3
226	0.05	633	0.1	483	0.15	400	0.2	400	0.4
293	0.0625	707	0.125	581	0.1875	500	0.25	500	0.5
368	0.075	775	0.15	673	0.225	600	0.3	600	0.6
452	0.0875	837	0.175	761	0.2625	700	0.35	700	0.7
553	0.1	895	0.2	844	0.3	800	0.4	800	0.8
684	0.1125	949	0.225	924	0.3375	900	0.45	900	0.9

Source: own material

5. Selected tests of independence

The opportunity to make something better than it currently is occurs everywhere if someone chooses to act. As a result statistical science has been enriched with many statistics intended to test independence.

In this paper the well-established and commonly used χ^2 test was selected. In relation to 2×2 CTs χ^2 test statistics is defined as

$$Q_1 = \chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(n_{ij}^* - e_{ij}^*\right)^2}{e_{ij}^*}, \tag{5}$$

where $e_{ij}^* = n_{i\bullet}^* \cdot n_{\bullet j}^* / n$ are expected counts. Statistics (5) asymptotically (i.e. sample size $n \to \infty$) follows the chi-square distribution with 1 degree of freedom, provided that H_0 is true. If not, and while n is large, test statistics follow the non-central chi-square distribution. For details please consult e.g. (Agresti, 2002).

Cressie and Read (1984) proposed the power divergence statistics (PDS). PDS for 2×2 CTs is given by

$$P^2 = \frac{2}{\lambda(\lambda+1)}\sum_{i=1}^{2}\sum_{j=1}^{2} n_{ij}^* \left\{\left(\frac{n_{ij}^*}{e_{ij}^*}\right)^{\lambda} - 1\right\} \quad -\infty < \lambda < \infty, \tag{6}$$

which is always positive and is defined as limits of P^2 at –1 and 0. This is a very rich class of test statistics and it contains many other test statistics. There are:

1) the χ^2 statistics ($\lambda = 1$) (formula (5)),
2) the G^2 statistics (the limit as λ goes to 0),
3) the Freeman-Tukey statistics ($\lambda = -0.5$),
4) the modified G^2 statistics (the limit as λ goes to –1),
5) the Neyman modified χ^2 statistics ($\lambda = -2$),
6) the Cressie-Read statistics ($\lambda = 2/3$).

Garcia-Perez and Nunez-Anton (2009) studied the whole family of the PDS. They found that χ^2 test statistics is the best up to a table density as low as 2. They showed that the G^2 statistics performs poorly and hence did not recommend it. In (Sulewski, 2016a), it was shown that the PDS- on probability scenarios II, IV, V (Tab. 2) – are of exactly/more or less the same power. Therefore, for further analysis in this paper, the well-known and commonly used χ^2 test statistics (1) was selected.

Another adjustment to the χ^2statistics is given by Zelterman (1987). The adjustment is the D^2 statistics, which in 2×2 CTs is given by

$$Q_2 = D^2 = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left(n_{ij}^* - e_{ij}^*\right)^2 - n_{ij}^*}{e_{ij}^*}. \tag{7}$$

D^2 test statistics is not a member of the PDS. Note that the contribution of $n^*_{ij} = 0$ to D^2 is the same as that of the χ^2 statistics.

In (Sulewski, 2013) the modular statistics was proposed, which is a modification of the χ^2 statistics for CTs and in 2×2 CTs is given by

$$Q_3 = |\chi| = \sum_{i=1}^{2}\sum_{j=1}^{2}\frac{\left|n^*_{ij} - e^*_{ij}\right|}{e^*_{ij}}. \tag{8}$$

The modular statistics was compared in terms of power with the PDS for the 2×2 CTs (Sulewski, 2016a). For scenarios II, IV (Tab. 2) it was shown that the modular statistics is more powerful than PDS.

In this paper the new logarithmic minimum statistics (LMS) was proposed, that in 2×2 CTs is defined as follows

$$Q_4 = LMS = -\sum_{i=1}^{2}\sum_{j=1}^{2}\ln\left[\frac{\min\left(n^*_{ij}, e^*_{ij}\right)}{\max\left(n^*_{ij}, e^*_{ij}\right)}\right]. \tag{9}$$

6. Determining of the power-of-the-test function

This section is devoted to *PoT* functions. An independent variable (an argument) is an appropriately defined measure of untruthfulness of H_0 (MoU). A dependent variable (a value of the function) is a probability to reject H_0 when it should be rejected as being untruth.

6.1. Generation of 2×2 contingency tables

In the simulation study, the generation of contingency tables is very important. The Markov Chain Monte Carlo (Diaconis, Efron, 1998; Cryan, Dyer, 2003; Chen et al., 2005; Cryan et al., 2006; Fishman, 2012), the Sequential Importance Sampling (Chen et al., 2005; Chen et al., 2006; Blitzstein, Diaconis, 2011; Yoshida et al., 2011), the probabilistic divide-and-conquer technique (Desalvo, Zhao, 2016), the Generalized Gamma Distribution (Sulewski, 2009), and the bar method (Sulewski et al., 2015) are all very popular approaches in literature about the generation of two-way contingency tables.

In this paper to generate 2×2 CTs the bar method was used. The algorithm of this method is as follows:

Phase 1

Step 1: Choose scenario I–V.
Step 2: Set MoU and value of k, a (Tab. 3).
Step 3: Calculate probabilities p_{ij} (i, j = 1, 2) (Tab. 2).
Step 4: Set label cells of 2×2 tables p_i (ij = 1, 2, 3) according to the rules

$$p_1 = p_{11}, p_2 = p_1 + p_{12}, p_3 = p_2 + p_{21}. \tag{10}$$

Step 5: Set sample size n.

Phase 2

Repeat the following steps from 1 to $3n$ times:
Step 1: Set initial values of cells count i.e. $n_{ij} = 0$ (i, j = 1, 2).
Step 2: Generate random number rw uniformly distributed within (0, 1).
Step 3: Increase cells count of 2×2 tables according to the rule:

$$rw \leq p_1 \Rightarrow n_{11} = n_{11} + 1,\ \ p_1 < rw \leq p_2 \Rightarrow n_{12} = n_{12} + 1,$$

$$p_2 < rw \leq p_3 \Rightarrow n_{21} = n_{21} + 1,\ \ rw > p_3 \Rightarrow n_{22} = n_{22} + 1,$$

where p_i (i = 1, 2, 3) are given by (10).

6.2. Power of the test assessment

Let MoU be an appropriately defined measure of XY dependency (3). A standard course of action is that two competitive hypotheses are formed, namely:
1. The main hypothesis, H_0, that says: X and Y are independent.
2. An alternative hypothesis, H_1 that says: X and Y are dependent.

The PoT function takes the measure of dependency MoU as its argument and returns the probability of rejecting H_0 as dependency increases. Since there is no way to determine the function in an analytical way, we employ the Monte Carlo method which we can rely on in such situations.

Phase 1

Step 1: Set sample size n.
Step 2: Set significance level $\alpha = 0.1$.
Step 3: Repeat $m = 10^6$ times the following points (such a large number of repetitions guaranties avery accurate result):

– Generate of the 2×2 CT in accordance with the scenario as it was described above for $a = 0$ (Tab. 2).
– Calculate values of test statistics Q_{il} ($i = 1, \ldots, 4$; $l = 1, \ldots, m$).
Step 4: Present values of test statistics in increasing order.
Step 5: Calculate critical values of tests $cv1_i$ ($i = 1, \ldots, 5$) on the base of estimates of quantiles. Unknown values of quantiles are replaced by appropriate order statistics, namely

$$cv1_i = Q_{i(m[1-\alpha])} \quad (i = 1, \ldots, 4). \tag{11}$$

Step 6: Repeat $\omega = 100$ steps 3–5 and calculate finally critical values of tests cv_i ($i = 1, \ldots, 4$).

$$cv_i = \frac{1}{\omega}\sum_{t=1}^{\omega} cv1_{it} \quad (i = 1, \ldots 4). \tag{12}$$

Phase 2

Step 1: Choose scenario I–V
Step 2: Set a (Tab. 2).
Step 3: Set value of MoU (2) for which the test has to be carried-out (Tab. 3).
Step 4: Create a CQ_i ($i = 1, \ldots, 4$) that are counters of rejections of H_0 for Q_i statistics.
Step 5: Set initial values $CQ_i = 0$ ($i = 1, \ldots, 4$).
Step 6: Repeat $m = 10^6$ times the following points:
– Generate of the 2×2 CT in accordance with the scenario described above, taking into account a (Tab. 2).
– Calculate values of test statistics Q_i ($i = 1, \ldots, 4$).
– $CQ_i = CQ_i + 1$, when $Q_i > cv_i$ ($i = 1, \ldots, 4$).
Step 7: Calculate power of tests

$$PoT_i(MoU) = Q_{\ i}/m \; (i = 1, \ldots, 4).$$

6.3. The results

Different scenarios determine different intervals of achievable MoU values. Moreover, the sample size in particular scenarios differs. The reason is that the minimal sample size n for each particular scenario has to be chosen to guarantee that all the cell counts expected in a 2×2 contingency table will be nonzero. The maximal sample size n was, in turn, chosen to obtain tests of high power. Tables 4–8 show how MoU impacts PoT at significance level $\alpha = 0.05$ and for a given sample size *n*, Figures 1–5 – at significance level $\alpha = 0.01$. These tables and figures indicate that the MoU – for analyzed scenarios – changes in different intervals, and it takes the greatest value for the scenario V. Of course, *PoT* function is a strictly increasing function and its values are increasing together with the sample size *n*.

At the outset, test power functions of Q_i ($i = 1, 2, 3$) tests were visually compared. The most powerful test was selected out of them. It was definitely Q_3 then taken as a representative of the trio in question to compete against Q_4. Since the test power function has a meaning of a portion of untruthful hypotheses H_0 rejected, a widely known Z test for testing equality of two portions was employed. Obtained values significantly higher relative to Q_3 were marked in bold (see Tables 4–8).

Table 4. Sizes and power of four tests for scenario I and $\alpha = 0.05$

MoU	$n = 250$				$n = 500$				$n = 750$				$n = 1000$			
	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4
0	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.0125	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.07	0.07	0.07
0.025	0.07	0.07	0.07	0.07	0.09	0.09	0.09	0.09	0.10	0.10	0.10	0.10	0.12	0.12	0.12	0.12
0.0375	0.09	0.09	0.09	0.09	0.14	0.14	0.14	0.14	0.17	0.17	0.18	0.18	0.22	0.22	0.22	0.22
0.05	0.13	0.12	0.13	0.13	0.20	0.20	0.20	0.20	0.28	0.28	0.28	0.28	0.36	0.36	0.37	0.36
0.0625	0.17	0.17	0.18	0.17	0.30	0.30	0.31	0.31	0.41	0.41	0.42	0.42	0.52	0.52	0.53	0.53
0.075	0.23	0.22	0.24	0.23	0.40	0.40	0.41	0.41	0.55	0.55	0.57	0.57	0.67	0.67	0.69	0.69
0.0875	0.30	0.30	0.32	0.31	0.52	0.52	0.54	0.54	0.69	0.69	0.72	0.71	0.82	0.82	0.83	0.83
0.1	0.38	0.38	0.42	0.41	0.65	0.65	0.69	0.68	0.82	0.82	0.84	0.84	0.91	0.91	0.93	0.93
0.1125	0.48	0.48	0.54	0.53	0.78	0.77	0.82	0.81	0.91	0.91	0.93	0.93	0.97	0.97	0.98	0.98

Source: own material

Table 5. Sizes and power of four tests for scenario II and $\alpha = 0.05$

MoU	$n = 40$				$n = 60$				$n = 80$				$n = 100$			
	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4
0	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.025	0.06	0.06	0.07	0.07	0.06	0.06	0.06	0.07	0.06	0.06	0.06	0.07	0.06	0.06	0.06	0.07
0.05	0.06	0.06	0.08	0.09	0.07	0.07	0.08	0.09	0.07	0.07	0.08	0.09	0.08	0.08	0.09	0.10
0.075	0.08	0.08	0.11	0.13	0.09	0.09	0.11	0.14	0.11	0.11	0.14	0.15	0.12	0.12	0.15	0.17
0.1	0.09	0.09	0.14	**0.17**	0.12	0.13	0.17	**0.21**	0.15	0.16	0.21	0.24	0.19	0.19	0.25	0.28
0.125	0.10	0.11	0.18	**0.24**	0.17	0.18	0.25	**0.30**	0.22	0.24	0.32	**0.36**	0.28	0.29	0.37	**0.42**
0.15	0.12	0.13	0.22	**0.30**	0.22	0.24	0.35	**0.42**	0.32	0.35	0.46	**0.52**	0.41	0.43	0.54	**0.59**
0.175	0.14	0.16	0.28	**0.39**	0.30	0.32	0.48	**0.57**	0.45	0.49	0.63	**0.69**	0.57	0.59	0.72	**0.77**
0.2	0.15	0.17	0.34	**0.49**	0.39	0.41	0.62	**0.72**	0.59	0.63	0.80	**0.84**	0.74	0.76	0.88	**0.91**
0.225	0.16	0.19	0.38	**0.58**	0.47	0.50	0.76	**0.86**	0.74	0.77	0.92	**0.95**	0,88	0,90	0,97	**0,98**

Source: own material

Table 6. Sizes and power of four tests for scenario III and α = 0.05

MoU	$n = 25$				$n = 40$				$n = 55$				$n = 70$			
	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4
0	0.05	0.05	0.05	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.0375	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06
0.075	0.06	0.06	0.07	0.07	0.07	0.08	0.08	0.08	0.08	0.08	0.09	0.09	0.10	0.10	0.10	0.10
0.1125	0.08	0.08	0.08	0.09	0.11	0.11	0.12	0.12	0.13	0.13	0.14	0.15	0.16	0.16	0.18	0.18
0.15	0.10	0.10	0.11	0.12	0.17	0.17	0.19	0.19	0.20	0.20	0.22	0.24	0.26	0.26	0.29	0.30
0.1875	0.12	0.12	0.14	0.16	0.23	0.23	0.27	0.28	0.29	0.29	0.33	0.36	0.39	0.39	0.43	0.45
0.225	0.15	0.15	0.19	0.21	0.31	0.32	0.38	0.40	0.42	0.42	0.48	0.51	0.54	0.54	0.60	0.63
0.2625	0.17	0.17	0.23	**0.27**	0.41	0.42	0.51	0.54	0.56	0.57	0.65	**0.69**	0.71	0.71	0.78	0.80
0.3	0.20	0.20	0.28	**0.34**	0.50	0.52	0.64	**0.69**	0.71	0.72	0.81	**0.84**	0.85	0.85	0.91	0.93
0.3375	0.22	0.23	0.34	**0.41**	0.60	0.62	0.78	**0.84**	0.84	0.85	0.93	**0.95**	0.95	0.95	0.98	0.99

Source: own material

Table 7. Sizes and power of four tests for scenario IV and α = 0.05

MoU	$n = 25$				$n = 35$				$n = 45$				$n = 55$			
	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4
0	0.05	0.05	0.05	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.05	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.06	0.07	0.06	0.07	0.07	0.07	0.07
0.1	0.08	0.07	0.08	0.08	0.09	0.09	0.09	0.09	0.11	0.10	0.11	0.11	0.11	0.11	0.12	0.12
0.15	0.11	0.11	0.11	0.12	0.15	0.15	0.16	0.15	0.17	0.17	0.18	0.18	0.20	0.20	0.21	0.21
0.2	0.16	0.16	0.17	0.17	0.22	0.22	0.24	0.23	0.28	0.28	0.29	0.29	0.32	0.32	0.34	0.34
0.25	0.23	0.22	0.24	0.24	0.33	0.33	0.36	0.35	0.41	0.41	0.44	0.44	0.48	0.48	0.51	0.51
0.3	0.31	0.30	0.34	0.34	0.45	0.45	0.50	0.49	0.57	0.57	0.61	0.61	0.65	0.65	0.69	0.69
0.35	0.40	0.38	0.45	0.46	0.59	0.59	0.66	0.65	0.72	0.73	0.77	0.77	0.81	0.80	0.84	0.85
0.4	0.51	0.48	0.57	0.59	0.73	0.73	0.80	0.80	0.86	0.86	0.90	0.91	0.92	0.92	0.95	0.95
0.45	0.61	0.57	0.70	0.73	0.85	0.85	0.92	0.92	0.95	0.95	0.97	0.98	0.98	0.98	0.99	0.99

Source: own material

Table 8. Sizes and power of four tests for scenario V and α = 0.05

MoU	$n = 25$				$n = 35$				$n = 45$				$n = 55$			
	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4	Q_1	Q_2	Q_3	Q_4
0	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.06	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.1	0.07	0.07	0.07	0.07	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.10	0.10	0.10	0.10
0.2	0.15	0.15	0.14	0.14	0.16	0.16	0.16	0.16	0.18	0.18	0.18	0.18	0.22	0.22	0.22	0.22
0.3	0.25	0.26	0.25	0.25	0.31	0.31	0.31	0.31	0.37	0.37	0.37	0.36	0.42	0.42	0.43	0.42
0.4	0.42	0.42	0.41	0.41	0.51	0.50	0.51	0.51	0.59	0.59	0.60	0.58	0.67	0.66	0.67	0.67
0.5	0.60	0.60	0.60	0.59	0.71	0.71	0.71	0.71	0.80	0.80	0.80	0.79	0.86	0.86	0.87	0.86
0.6	0.77	0.78	0.77	0.77	0.87	0.87	0.88	0.87	0.93	0.93	0.93	0.93	0.96	0.96	0.97	0.96
0.7	0.90	0.90	0.90	0.90	0.97	0.96	0.97	0.96	0.99	0.99	0.99	0.99	1.00	1.00	1.00	1.00
0.8	0.98	0.98	0.98	0.98	0.99	0.99	1.00	0.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
0.9	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Source: own material

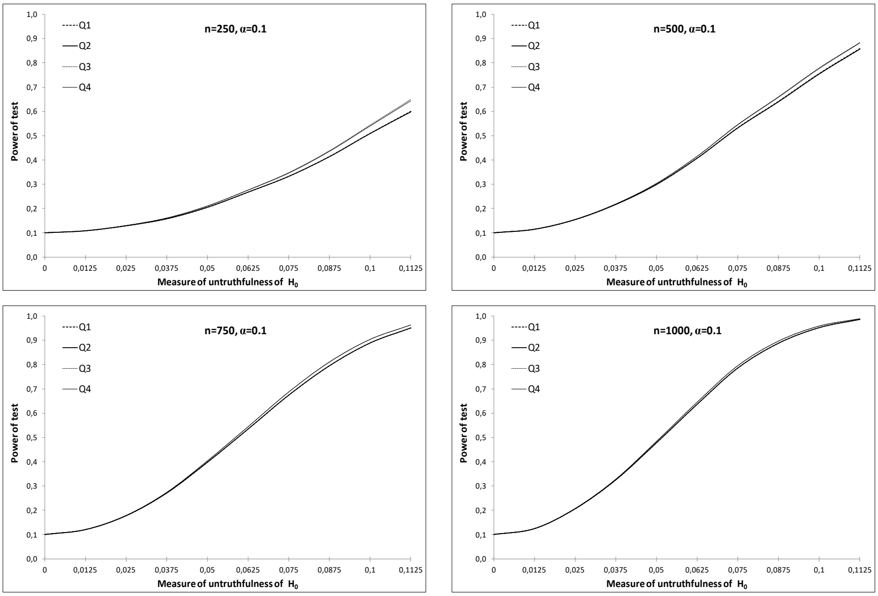

Figure 1. The power-of-test function for $n = 250 \cdot i$ ($i = 1, \ldots, 4$). Scenario I

Source: own material

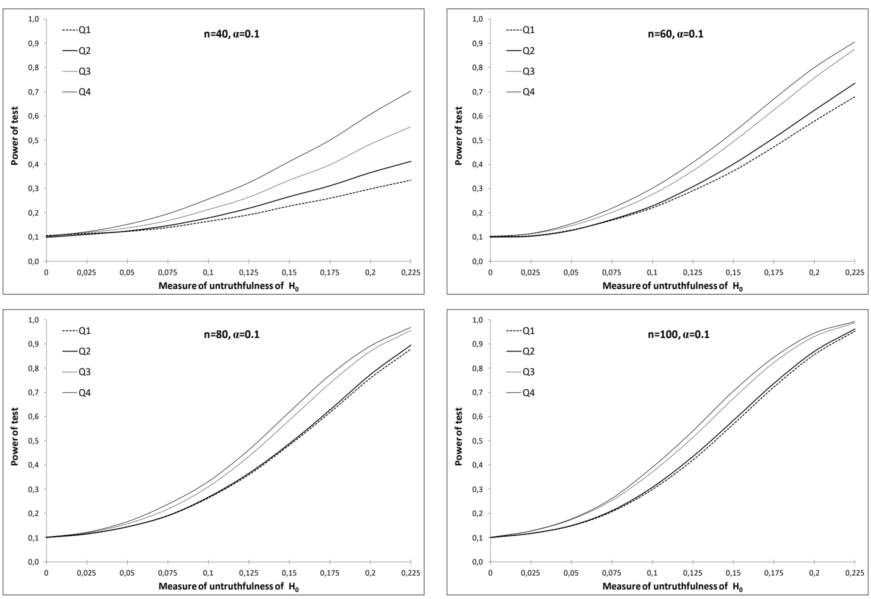

Figure 2. The power-of-test function for $n = 20 \cdot i + 20$ ($i = 1, \ldots, 4$). Scenario II

Source: own material

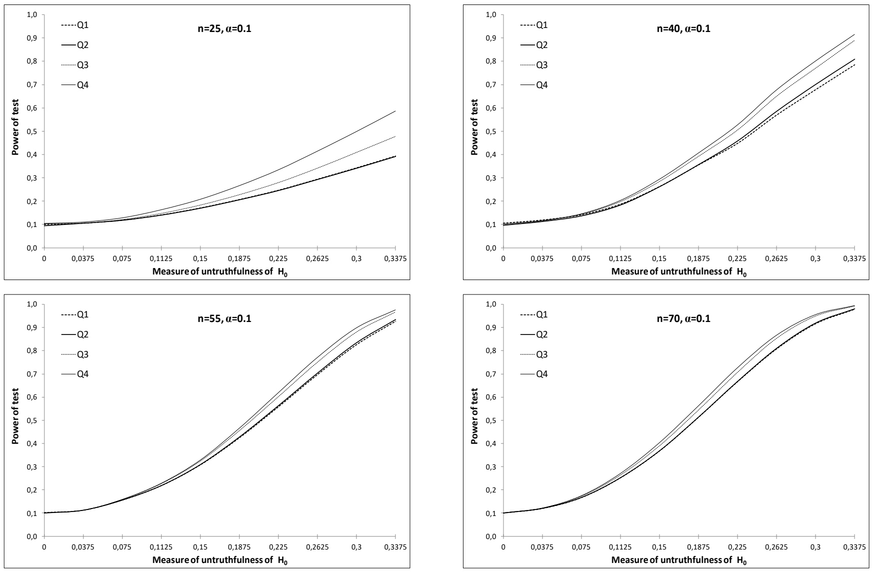

Figure 3. The power-of-test function for $n = 15 \cdot i + 10$ ($i = 1, \ldots, 4$). Scenario III

Source: own material

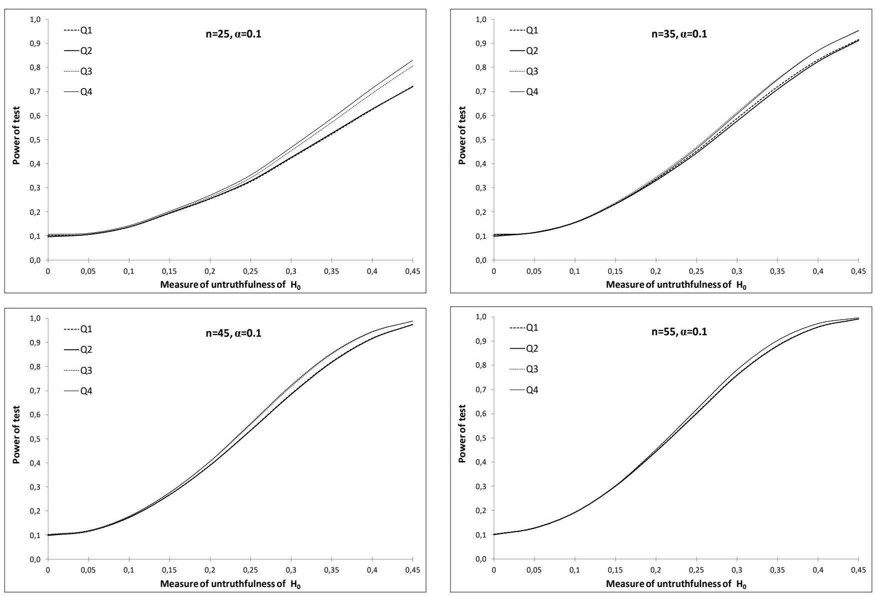

Figure 4. The power-of-test function for $n = 10 \cdot i + 15$ ($i = 1, \ldots, 4$). Scenario IV

Source: own material

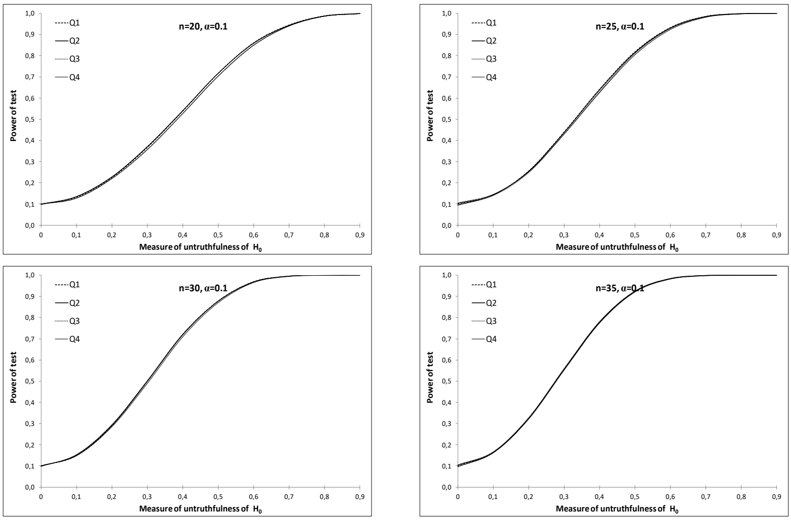

Figure 5. The power-of-test function for $n = 5 \cdot i + 15$ ($i = 1, \ldots, 4$). Scenario V

Source: own material

For weak dependence in scenarios I–IV and in scenario V, the power functions of these tests are very similar. Under scenario I, for strong dependence and for the smaller samples, the highest power function corresponds to Q_3 and Q_4 statistics. Related to Scenarios II and III the highest power function corresponds to Q_4 statistics, this is particularly visible for the smaller samples. Under scenario IV, Q_3 and Q_4 statistics have the highest power and Q_4 statistics is more powerful for sample size $n = 25$.

7. Examples

In a general population, a testing for independence of two features was carried out. The obtained results are presented in Table 9. Table 10 gives the average of 100 simulated critical values using $m = 10^6$ of $Q_i^\alpha \ (i = 1,\ldots,4)$ for $n = 40$ and $\alpha \in \{0.05, 0.1\}$. This table summarizes other findings too.

Table 9. Empirical data in 2×2 CTs

X	***Y***		**Total**
	Y_1	Y_2	
X_1	1	10	11
X_2	10	19	29
Total	11	29	40

Source: own material

Table 10. Critical values of four tests

Test	MoU	Value of statistics	Critical value	
			α = 0.05	α = 0.01
Q_1	0.2025	2.579	3.861	2.640
Q_2		–37.421	–0.129	–1.355
Q_3		1.274	1.264	1.067
Q_4		1.661	1.321	1.101

Source: own material

The preferred tests are Q_3, Q_4 according to which null hypothesis is rejected, whereas the Q_1, Q_2 tests retain the null.

8. Conclusion

Under the considered scenarios MoU changes in different intervals. It is because different scenarios determine different intervals of achievable MoU values. Moreover, sample sizes in particular scenarios are different. The reason is that the minimal sample sizes for each particular scenario have to be chosen to guarantee counts expected for each $i, j = 1, 2$ to be nonzero. The maximal sample size was, in turn, chosen to obtain high power of tests.

Under scenarios I–IV, especially for smaller samples and for strong dependence, the new test (LMS) is of higher power than other tests analyzed in this paper. Under scenario V, where MoU takes the greatest values, the power functions of these tests are very similar. An identical situation can be observed in scenarios I–IV for weak dependence.

Therefore it can be concluded, that the LMS is not less effective than the Q_i ($i = 1, 2, 3$) tests, and it may be an alternative to them. Under some scenarios, LMS is of higher power, that is understood in the sense of the proposed measure MoU. With the increase of the variability range of MoU, the power function of LMS is more similar to other tests.

References

Agresti A. (2002), *Categorical Data Analysis*, Wiley, New Jersey.

Albert J.H. (1990), *A Bayesian test for a two-way contingency table using independence*, "Prior. Canadian Journal of Statistics", vol. 18, no. 4, pp. 347–363.

Andrés A.M., Tejedor I.H., Mato A.S. (1995), *The Wilcoxon, Spearman, Fisher, χ^2, Student and Pearson Tests and 2x2 Tables*, "The Statistician", pp. 441–450.

Beh E.J., Farver T.B. (2009), *An evaluation of non-iterative methods for estimating the linear-by-linear parameter of ordinal log-linear models*, "Australian & New Zealand Journal of Statistics", vol. 51, no. 3, pp. 335–352.

Berry K.J., Mielke P.W. (1988), *Monte Carlo comparisons of the asymptotic chi-square and likelihood-ratio tests with the no asymptotic chi-square tests for sparse r×c tables*, "Psychological Bulletin", vol. 103, no. 2, p. 256.

Blitzstein J., Diaconis P. (2011), *A sequential importance sampling algorithm for generating random graphs with prescribed degrees*, "Internet Mathematics", vol. 6, pp. 489–522.

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, pp. 3661–3675.

Ceyhan E. (2010), *Directional clustering tests based on nearest neighbor contingency tables*, "Journal of nonparametric Statistics", vol. 22, no. 5, pp. 599–616.

Chang C.H., Lin J.J., Pal N. (2011), *Testing the equality of several gamma means: a parametric bootstrap method with applications*, "Computational Statistics", vol. 26, no. 1, pp. 55–76.

Chang C.H., Pal N. (2008), *A revisit to the Behrens–Fisher problem: comparison of five test methods*, "Communications in Statistics – Simulation and Computation", vol. 37, no. 6, pp. 1064–1085.

Chen Y., Diaconis P., Holmes S.P., Liu J.S. (2005), *Sequential Monte Carlo methods for statistical analysis of tables*, "Journal of the American Statistical Association", vol. 100, pp. 109–120.

Chen Y., Dinwoodie I.H., Sullivant S. (2006), *Sequential importance sampling for multiway tables*, "The Annals of Statistics", pp. 523–545.

Clogg C.C., Eliason S.R. (1987), *Some common problems in log-linear analysis*, "Sociological Methods & Research", vol. 16, no. 1, pp. 8–44.

Cochran W.G. (1952), *The χ^2 test of goodness of fit*, "The Annals of Mathematical Statistics", pp. 315–345.

Cochran W.G. (1954), *Some methods for strengthening the common χ^2 tests*, "Biometrics", vol. 10, no. 4, pp. 417–451.

Cohen J., Nee J.C. (1990), *Robustness of Type I Error and Power in Set Correlation Analysis of Contingency Tables*, "Multivariate Behav. Res.", vol. 25, no. 3, pp. 341–350.

Cressie N., Read T. (1984), *Multinomial Goodness-of-Fit Tests*, "J. R. Stat. Soc. Ser. B. Stat. Methodol.", vol. 46, pp. 440–464.

Cressie N., Read T.R. (1989), *Pearson's χ^2 and the log likelihood ratio statistics G2: a comparative review*, "International Statistical Review/Revue Internationale de Statistique", pp. 19–43.

Cryan M., Dyer M. (2003), *A polynomial-time algorithm to approximately count contingency tables when the number of rows is constant*, "Journal of Computer and System Sciences", vol. 67, pp. 291–310.

Cryan M., Dyer M., Goldberg L.A., Jerrum M., Martin R. (2006), *Rapidly mixing Markov chains for sampling contingency tables with a constant number of rows*, "SIAM Journal on Computing", vol. 36, pp. 247–278.

Cung C. (2013), *Crime and Demographics: An Analysis of LAPD Crime Data*, "M. sc. Thesis", UCLA, Department of Statistics, Los Angeles.

Davis C.S. (1993), *A new approximation to the distribution of Pearson's chi-square*, "Statistica-Sinica", pp. 189–196.

Desalvo S., Zhao J.Y. (2016), *Random Sampling of Contingency Tables via Probabilistic Divide-and-Conquer*, "ArXiv preprint", ArXiv, 1507.00070v4.

Diaconis P., Efron B. (1985), *Testing for independence in a two-way table: new interpretations of the chi-square statistics*, "The Annals of Statistics", pp. 845–874.

Diaconis P., Sturmfels B. (1998), *Algebraic algorithms for sampling from conditional distributions*, "The Annals of Statistics", vol. 26, pp. 363–397.

Dickhaus T., Straßburger K., Schunk D., Morcillo-Suarez C., Illig T., Navarro A. (2012), *How to analyze many contingency tables simultaneously in genetic association studies*, "Statistical Applications in Genetics and Molecular Biology", vol. 11, no. 4, pp. 1544–6115.

Egozcue J.J., Pawlowsky-Glahn V., Templ M., Hron K. (2015), *Independence in contingency tables using simplicial geometry*, "Communications in Statistics-Theory and Methods", vol. 44, no. 18, pp. 3978–3996.

El Galta R., Stijnen T., Houwing-Duistermaat J.J. (2008), *Testing for genetic association: a powerful score test*, "Stat Med.", vol. 27, no. 22, pp. 4596–4609.

Fisher R.A. (1922), *On the interpretation of* χ^2 *from contingency tables, and the calculation of P*, "Journal of the Royal Statistical Society", vol. 85, no. 1, pp. 87–94.

Fishman G.S. (2012), *Counting contingency tables via multistage Markov chain Monte Carlo*, "Journal of Computational and Graphical Statistics", vol. 21, pp. 713–738.

García J.E., González-López V.A. (2014), *Independence tests for continuous random variables based on the longest increasing subsequence*, "Journal of Multivariate Analysis", vol. 127, pp. 126–146.

García J.E., González-López V.A. (2016), *Independence test for sparse data*, "International Conference of Numerical Analysis and Applied Mathematics 2015", AIP Publishing, vol. 1738, no. 1, p. 140002.

Garcia-Perez M.A., Nunez-Anton V. (2009), *Accuracy of the power-divergence statistics for testing independence and homogeneity in two-way contingency tables*, "Commun. Stat. – Simul. Comput.", vol. 38, pp. 503–512.

Garside G.R., Mack C. (1976), *Actual type 1 error probabilities for various tests in the homogeneity case of the 2×2 contingency table*, "The American Statistician", vol. 30, no. 1, pp. 18–21.

Haber M. (1987), *A comparison of some conditional and unconditional exact tests for 2x2 contingency tables: A comparison of some conditional and unconditional exact tests*, "Communications in Statistics – Simulation and Computation", vol. 16, no. 4, pp. 999–1013.

Haberman S.J. (1981), *Tests for independence in two-way contingency tables based on canonical correlation and on linear-by-linear interaction*, "The Annals of Statistics", vol. 9, no. 6, pp. 1178–1186.

Hall P., Wilson S.R. (1991), *Two guidelines for bootstrap hypothesis testing*, "Biometrics", pp. 757–762.

Hui-Qiong L., Guo-Liang T., Xue-Jun J., Nian-Sheng T. (2016), *Testing hypothesis for a simple ordering in incomplete contingency tables*, "Computational Statistics & Data Analysis", vol. 99, pp. 25–37.

Iossifova R., Marmolejo-Ramos F. (2013), *When the body is time: spatial and temporal deixis in children with visual impairments and sighted children*, "Research in Developmental Disabilities", vol. 34, no. 7, pp. 2173–2184.

Irwin J.O. (1935), *Tests of significance for differences between percentages based on small numbers*, "Metron", vol. 12, no. 2, pp. 84–94.

Jeong H.C., Jhun M., Kim D. (2005), *Bootstrap tests for independence in two-way ordinal contingency tables*, "Computational Statistics & Data Analysis", vol. 48, no. 3, pp. 623–631.

Koehler K.J., Larntz K. (1980), *An empirical investigation of goodness-of-fit statistics for sparse multinomials*, "Journal of the American Statistical Association", vol. 75, no. 370, pp. 336–344.

Lawal H.B., Uptong G.J.G. (1984), *On the use of* χ^2 *as a test of independence in contingency tables with small cell expectations*, "Australian Journal of Statistics", vol. 26, pp. 75–85.

Lawal H.B., Uptong G.J.G. (1990), *Comparisons of Some Chi-squared Tests for the Test of Independence in Sparse Two-Way Contingency Tables*, "Biometrical Journal", vol. 32, no. 1, pp. 59–72.

Lin J.J., Chang C.H., Pal N. (2015), *A revisit to contingency table and tests of independence: bootstrap is preferred to Chi-Square approximations as well as Fisher's exact test*, "Journal of Biopharmaceutical Statistics", vol. 25, no. 3, pp. 438–458.

Lipsitz S.R., Fitzmaurice G.M., Sinha D., Hevelone N., Giovannucci E., Hu J.C. (2015), *Testing for independence in J×K contingency tables with complex sample survey data*, "Biometrics", vol. 71, no. 3, pp. 832–840.

Lydersen S., Fagerland M.W., Laake P. (2009), *Recommended tests for association in 2×2 tables*, "Statistics in Medicine", vol. 28, no. 7, pp. 1159–1175.

Meng R.C., Chapman D.G. (1966), *The power of chi square tests for contingency tables*, "Journal of the American Statistical Association", vol. 61, no. 316, pp. 965–975.

Nandram B., Bhatta D., Bhadra D. (2013), *A likelihood ratio test of quasi-independence for sparse two-way contingency tables*, "Journal of Statistical Computation and Simulation", vol. 85, no. 2, pp. 284–304.

Pearson K. (1904), *On the theory of contingency and its relation to association and normal correlation*, "K. Pearson, Early Papers".

Shan G., Wilding G. (2015), *Unconditional tests for association in 2×2 contingency tables in the total sum fixed design*, "Statistica Neerlandica", vol. 69, no. 1, pp. 67–83.

Sulewski P. (2009), *Two-by-two Contingency Table as a Goodness-of-Fit Test*, "Computational Methods in Science and Technology", vol. 15, no. 2, pp. 203–211.

Sulewski P. (2013), *Modyfikacja testu niezależności [Modification of the independence test]*, "Statistical News – Central Statistical Office", vol. 10, pp. 1–19.

Sulewski P. (2014), *Statystyczne badanie współzależności cech typu dyskretne kategorie [The statistical study of features interdependence of discrete categories type]*, Pomeranian University, Slupsk.

Sulewski P., Motyka R. (2015), *Power analysis of independence testing for contingency tables*, "Scientific Journal of Polish Naval Academy", vol. 56, no. 1, pp. 37–46.

Sulewski P. (2016a), *Moc testów niezależności w tablicy dwudzielczej [Power of independence test in the contingency tables]*, "Statistical News – Central Statistical Office", vol. 8, pp. 1–17.

Sulewski P. (2016b), *Moc testów niezależności w tablicy dwudzielczej większej niż 2x2 [Power of independence test in the 2´2 contingency tables bigger than 2x2]*, "Statistical Review", vol. 63, no. 2, pp. 191–209.

Taneichi N., Sekiya Y. (2007), *Improved transformed statistics for the test of independence in r×s contingency tables*, "Journal of Multivariate Analysis", vol. 98, no. 8, pp. 1630–1657.

Vélez J.I., Marmolejo-Ramos F., Correa J.C. (2016), *A Graphical Diagnostic Test for Two-Way Contingency Tables*, "RevistaColombiana de Estadística", vol. 39, no. 1, pp. 97–108.

Wickens T.D. (1969), *Multiway Contingency Tables Analysis for the Social Sciences*, Psychology Press, United States.

Yenigün C.D., Székely G.J., Rizzo M.L. (2011). *A Test of Independence in Two-Way Contingency Tables Based on Maximal Correlation*, "Communications in Statistics-Theory and Methods", vol. 40, no. 12, pp. 2225–2242.

Yoshida R., Xi J., Wei S., Zhou F., Haws D. (2011), *Semigroups and sequential importance sampling for multiway tables*, "ArXiv preprint", ArXiv, 1111, p. 6518.

Yu Y. (2014), *Tests of independence in a single 2×2 contingency table with random margins*, Doctoral dissertation, Worcester Polytechnic Institute, Worcester.

Zelterman D. (1987). *Goodness-of-fit tests for large sparse multinomial distributions*, "J. Am. Stat. Assoc.", vol. 82, pp. 624–629.

Nowy test niezależności dla tablic dwudzielczych 2×2

Streszczenie: W literaturze statystycznej istnieje wiele miar do ujawniania niezależności dwóch zmiennych jakościowych w tabelach kontyngencji, w szczególności w tabelach dwudzielczych 2×2. W niniejszym artykule porównano cztery testy niezależności. Są to: test chi-kwadrat, jako najbardziej znany przedstawiciel statystyk power divergence, test modułowy oraz test d-kwadrat, jako modyfikacje testu Pearsona, test logarytmiczno-minimalny, będący nową propozycją. Wartości krytyczne dla wyżej wymienionych testów zostały wyznaczone metodami Monte Carlo. W celu porównania testów zaproponowano miarę nieprawdziwości H_0 i wyznaczono ich moc.

Słowa kluczowe: test niezależności, tablica dwudzielcza 2×2, statystyka logarytmiczno-minimalna, statystyka modułowa, statystyki power divergence, metoda Monte Carlo

JEL: C12, C14, C15, C46, C63

Received: 2016-12-17; verified: 2017-03-23. Accepted: 2017-09-07