**ZESZYTY NAUKOWE AKADEMII MARYNARKI WOJENNEJ**
SCIENTIFIC JOURNAL OF POLISH NAVAL ACADEMY

2015 (LVI) 1 (200)

DOI: 10.5604/0860889X.1161260

**Piotr Sulewski[1], Ryszard Motyka[1]**

# POWER ANALYSIS OF INDEPENDENCE TESTING FOR CONTINGENCY TABLES

## ABSTRACT

Six tests for independence in a two-way contingency table namely chi-squared test, log likelihood ratio test, Neyman-modified chi-squared test, Kullback-Leibler test, Freeman-Tukey test, Cressie--Read test, were examined. It was accomplished with the Monte Carlo method. The Goodman--Kruskal $\tau$ index was used to fix dependence in two-way contingency table in Monte Carlo experiments. The examination consisted in determining power functions of the tests. Next, the power functions were compared to each other. It was revealed that differences in power are negligible.

Key words:
two-way contingency table, independence test, Monte Carlo method.

## BASIC OF CONTINGENCY TABLE

Let $X$ and $Y$ be two features of the same object having levels $X_1, X_2, ..., X_w$ and $Y_1, Y_2, ..., Y_k$. Testing for independency of these two features with appropriately arranged contingency table and $\chi^2$ statistics applied is probably one of the most common statisticians' tasks. Table 1 recalls a pattern of two dimensional contingency table. There are $n$ items classified with respect to $X$ and $Y$. This produces a table of a scheme shown below as table 1. Here $n_{ij}$ are the counts, being the numbers of objects classified to belong to the cell $(X_i, Y_j)$. The symbols $n_{1\bullet}, n_{2\bullet}, ..., n_{w\bullet}$ and $n_{\bullet 1}, n_{\bullet 2}, ..., n_{\bullet k}$ are marginal counts.

[1] The Pomeranian Academy, Institute of Mathematics, Arciszewskiego 22 Str., 76-200 Słupsk, Poland; e-mail: {piotr.sulewski; ryszard.motyka}@apsl.edu.pl

Table 1. A scheme of $w \times k$ contingency table

| Criterion X | Criterion Y | | | | | | Marginal counts |
|---|---|---|---|---|---|---|---|
| | $Y_1$ | $Y_2$ | ... | $Y_j$ | ... | $Y_k$ | |
| $X_1$ | $n_{11}$ | $n_{12}$ | ... | $n_{1j}$ | ... | $n_{1k}$ | $n_{1\bullet}$ |
| $X_2$ | $n_{21}$ | $n_{22}$ | ... | ... | ... | $n_{2k}$ | $n_{2\bullet}$ |
| ... | ... | ... | ... | ... | ... | ... | ... |
| $X_i$ | $n_{i1}$ | $n_{i2}$ | ... | $n_{ij}$ | ... | $n_{ik}$ | $n_{i\bullet}$ |
| ... | ... | ... | ... | ... | ... | ... | ... |
| $X_w$ | $n_{w1}$ | $n_{w2}$ | ... | $n_{wj}$ | ... | $n_{wk}$ | $n_{w\bullet}$ |
| Marginal counts | $n_{\bullet 1}$ | $n_{\bullet 2}$ | ... | $n_{\bullet j}$ . | ... | $n_{\bullet k}$ | $n$ |

The table of a form as above is a basis to test hypothesis commonly called the main and denoted $H_o$ that $X$ and $Y$ are independent. Pearson test statistics is of the form [11]:

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

where:

$$e_{ij} = \frac{n_{i\bullet}\cdot n_{\bullet j}}{n} \tag{1A}$$

are counts expected when $H_o$ is true. Statistics (1) have asymptotically a chi-square distribution on $(w-1)(k-1)$ degrees of freedom, when $H_o$ is true. Goodman and Kruskal [4] put forward the following measure of dependence in two dimensional contingency tables:

$$\tau = \frac{\sum_{i=1}^{w}\sum_{j=1}^{k} p_{ij}^2 / p_{\bullet j} - \sum_{i=1}^{w} p_{i\bullet}^2}{1-\sum_{i=1}^{w} p_{i\bullet}^2}, \quad p_* = \frac{n_*}{n}. \tag{1B}$$

**ALTERNATIVE TESTS OF INDEPENDENCE**

The opportunity to make something better that it currently is occurs everywhere if someone chooses to act. As the result statistical science has been enriched with five other statistics intended to test independency. The following competitive test statistics were developed after Pearson:

The $G^2$ statistics of Sokal and Rohlf [12]

$$G^2 = 2\sum_{i=1}^{r}\sum_{j=1}^{c} n_{ij} \ln\left(\frac{n_{ij}}{e_{ij}}\right); \tag{2}$$

The $N$ statistics of Neyman [10]

$$N = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{\left(n_{ij} - e_{ij}\right)^2}{n_{ij}}; \tag{3}$$

The $KL$ statistics of Kullback and Leibler [6]

$$KL = 2\sum_{i=1}^{r}\sum_{j=1}^{c} e_{ij} \ln\left(\frac{e_{ij}}{n_{ij}}\right) \tag{4}$$

The $FT$ statistics of Freeman and Tukey [3]

$$FT = 4\sum_{i=1}^{r}\sum_{j=1}^{k} \left(\sqrt{n_{ij}} - \sqrt{e_{ij}}\right)^2; \tag{5}$$

The $CR$ statistics of Cressie and Read [1]

$$CR = \frac{9}{5}\sum_{i=1}^{w}\sum_{j=1}^{k} n_{ij}\left[\left(\frac{n_{ij}}{e_{ij}}\right)^{2/3} - 1\right]. \tag{6}$$

Both Pearson and other originators claim that their statistics follow the chi--square distribution on $(w-1)(k-1)$ degrees of freedom provider $H_o$ is true. This claim seems irrelevant to $N$ and $KL$ statistics since they have $n_{ij}$ in denominators.

The counts $n_{ij}$ are random variables. Random variables that include reciprocals or ratios of random variables do not have expected value and frequently also some higher moments. In contrast chi-square distribution has all the moments. From the definition

$$D^{\lambda} = \frac{2}{\lambda(\lambda+1)}\sum_{i=1}^{w}\sum_{j=1}^{k} n_{ij}\left[\left(\frac{n_{ij}}{e_{ij}}\right)^{\lambda} - 1\right] \tag{7}$$

it can easily be seen that statistics (1) $(\lambda = 1)$, the log likelihood ratio statistics (2) $(\lambda = 0)$, the Neyman statistics (3) $(\lambda = -2)$, the modified log likelihood ratio statistics (4) $(\lambda = -1)$, the Freeman-Tukey statistics (5) $(\lambda = -0{,}5)$ and Cressie-Read statistics (6) $(\lambda = 2/3)$, are all special cases.

Statistics (6) emerges as an excellent and compromising alternative to the $D^0$ and $D^1$. Various properties and comparison of these so-called chi-squared tests can be found in [2, 5, 7, 9, 13].

## ANALYSIS OF THE POWER OF TESTS

### F i l l i n g  i n  c o n t i n g e n c y  t a b l e s

Phase 1. Preparation steps

Step P1. Setting parameters of simulation:

- dimension of contingency table $w \times k$, for instance $3 \times 3$;
- concrete value of $\tau$ that 'holds' in virtual general population we sample from, for instance $\tau_s = 0{,}5$;
- sample size $n$, for instance $n_s = 100$.

Step P2. Labeling cells. To cells of contingency table assign integers denoted $v$ that take values from $v = 1$ to $v_{\max} = w \times k$ starting from the leftmost upper corner. These are cell labels. For instance $p_5$ in label notation refers to $p_{23}$ in matrix notation.

Step P3. Determining a pattern of contingency table i.e. such a set of $p_{ij}^{s}$ values that substituted into (1B) yield $\tau_s = 0{,}5$.

Phase 2. The main steps

Repeat the following steps from M1 to M4 $n_s$ times:

Step M1. Generate random number $r_k$ uniformly distributed within $<0,1>$.

Step M2. Find such cell label $v^*$ that satisfies the following two sided inequality

$$\sum_{v=1}^{v^*-1} p_v < r_k \le \sum_{v=1}^{v^*} p_v \ . \tag{8}$$

Step M3. Having cell label $v^*$ determine cell index $ij$ .

Step M4. Increase cell count i.e. $n_{ij} = n_{ij} + 1$.

## Estimating power of test functions

Let $\tau$ be an appropriately defined measure of $XY$ dependency for instance of (1B) form. A standard course of action is that two competitive hypotheses are formed, namely:

- the null hypothesis, $H_0$ that says: $X$ and $Y$ are independent;
- alternative hypothesis, $H_1$ that says: $X$ and $Y$ are dependent.

The power of the test function has the measure of dependency $\tau$ as its argument and returns the probability of rejecting $H_0$ as dependency is growing. Since there is no way to determine the function in analytical way we employ the Monte Carlo method we can rely on in such situations.

Preparation phase

- set dimensions of the contingency table i.e. number of rows $w$ and number of columns $k$ ;
- set concrete value of $\tau = \tau_S$ for which the test has to be carried-out;
- set sample size $n = n_S$ ;
- set the confidence level $\alpha = \alpha_S$ ;
- calculate the number of degrees of freedom $d = (w-1)(k-1)$;
- determine $1-\alpha$ quantile of the chi-square distribution with $d$ degrees of freedom being the test critical value $\chi^2_{crit} = q_{1-\alpha,d}$ ;
- set number of repetitions $m_{rep}$ ;
- create a vector CoR of 6 rows that are counters of rejections of $H_0$; rows relate to particular test statistics from (1) to (6); set initial values to 0;
- determine the pattern of contingency table as it was described in section 3.

<u>The main phase</u>

Repeat $m_{rep}$ times what specified below:

- fill-in the contingency table in accordance with the pattern as it was described in section 4;
- calculate values of test statistics from (1) to (6);
- compare each of above values to critical value;
- increase by 1 the counter tied with this test statistics that exceeds critical value.

Calculate power of tests using statistics (1)–(6)

$$PoT_s(\tau_s) = CoR_i / m_{rep}\,; \quad s = 1,\ldots,6\,. \tag{9}$$

## The results

Let us choose $Pot_1(\tau)$ as the reference power function that is well grounded because the statistics of Pearson is both widely known and used. Power of chi-squared test is shown in figure 1. To compare other test functions define the relative difference as:

$$Rd_s(\tau) = \frac{Pot_s(\tau)}{Pot_1(\tau)} - 1; \quad s = 2,\ldots,6\,. \tag{10}$$

These relative differences are shown in figures from 2 to 5.

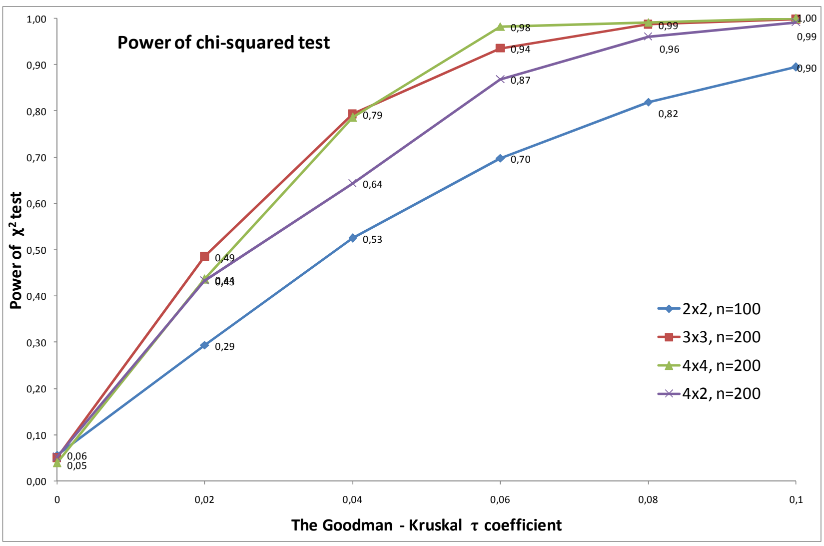


Fig. 1. Power of chi-squared test

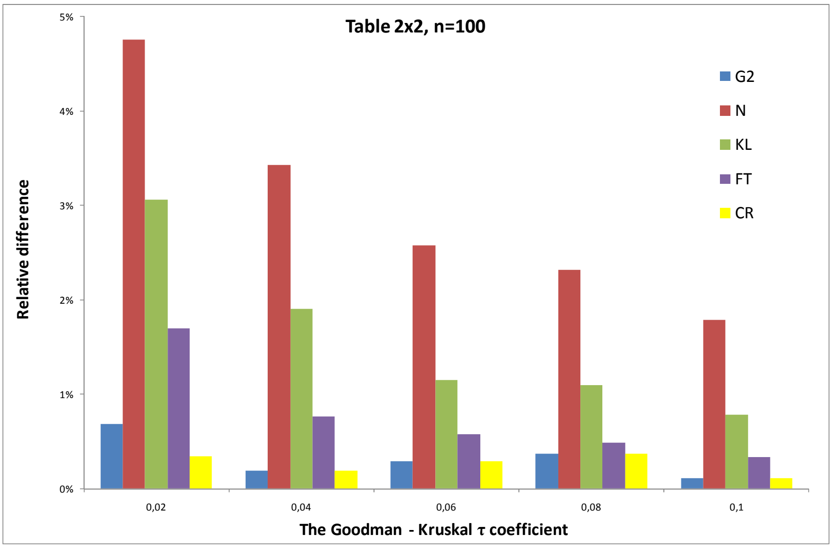


Fig. 2. Relative differences: table 2 x 2, $n$ = 100

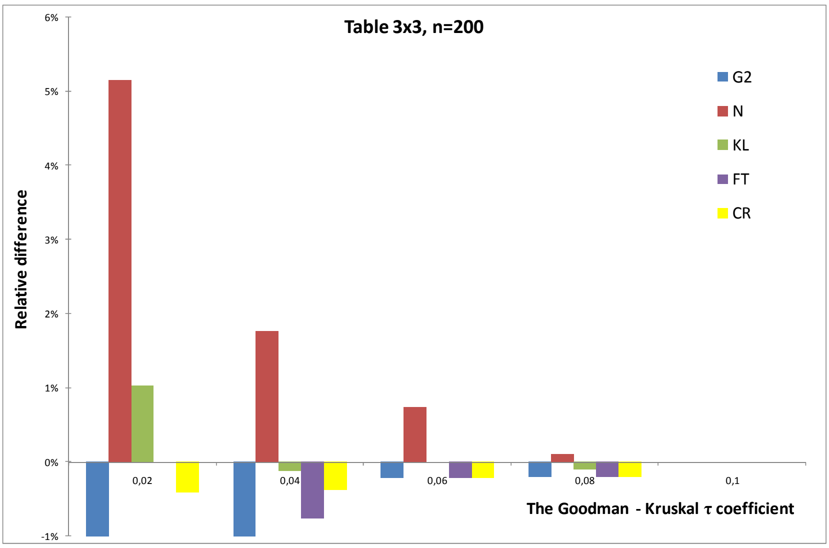


Fig. 3. Relative differences: table 3 x 3, $n$ = 200

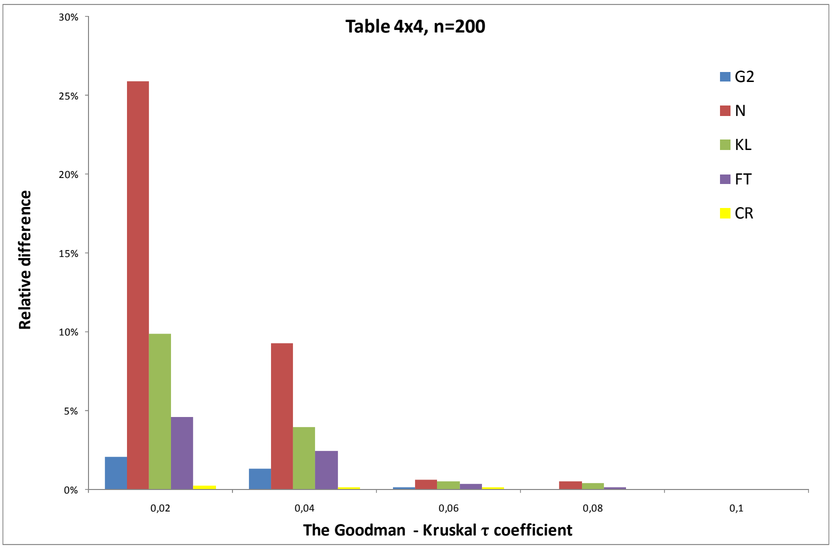


Fig. 4. Relative differences: table 4 x 4, $n$ = 200

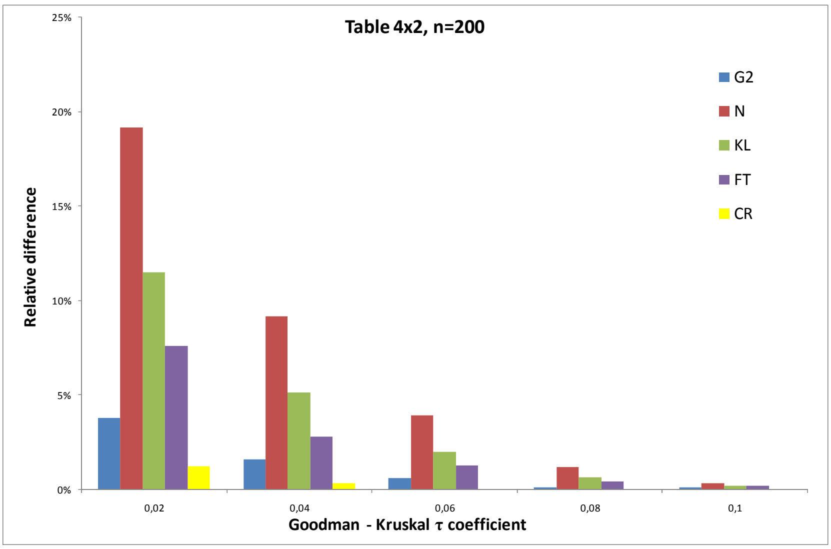


Fig. 5. Relative differences: table 4 x 2, $n$ = 200

## CONCLUSIONS

Differences (10) vary within interval of few percent. Some of them are even negative. Only test procedure that employs $N, KL$ statistics appears as outstanding one. An explanation is rather of this sort that $N, KL$ statistics simply, as it was mentioned in section 2, do not follow chi-square distribution. Concluding we may say that no noticeable progress was achieved in the domain of contingency tables by putting forward statistics (2)–(6).

## REFERENCES


[1] Cressie N., Read T., *Multinomial Goodness-of-Fit Tests*, 'Journal of the Royal Statistical Society', Series B (Methodological), 1984, Vol. 46, pp. 440–464.

[2] Fienberg S. E., *The use of chi-squared statistics for categorical data problems*, 'Journal of the Royal Statistical Society', Series B, 1979, Vol. 41, pp. 54–64.

[3] Freeman M. F., Tukey J. W., *Transformations related to the angular and the square root*, 'Annals of Mathematical Statistics', 1950, Vol. 21, pp. 607–611.

[4] Goodman L., Kruskal W., *Measures of Association for Cross classifications*, 'Journal of the American Statistical Association', 1954, Vol. 49, pp. 732–764.

[5] Horn S. D., *Goodness-of-fit tests for discrete data. A review and an applicatiob to a health impairment scale*, 'Biometrics', 1977, Vol. 33, pp. 237–248.

[6] Kullback S., *Information Theory and Statistics*, Wiley, New York 1959.

[7] Lancaster H. O., *The chi-squared distribution*, Wiley, New York 1969.

[8] Light R. J., Margolin B. H., *An Analysis of Variance for Categorical Data*, 'Journal of the American Statistical Association', 1971, Vol. 66, No. 335, pp. 534–544.

[9] Moore D. S., *Recent developments in chi-square tests for goodness-of-fit*, Mimeograph series 459, Department of Statistics, Purdue University, 1976.

[10] Neyman J., *Contribution to the Theory of the $x^2$ Test*, Proc. (First) Berkeley Symp. on Math. Statist. and Prob. (Univ. of Calif. Press), 1949, pp. 239–273.

[11] Pearson K., *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', 1900, Vol. 50, pp. 157–172.

[12] Sokal R. R., Rohlf F. J., *Biometry: the principles and practice of statistics in biological research*, Freeman, New York 2012.

[13] Watson G. S., *Some recent results in chi-square goodness-of-fit tests*, 'Biometrics', 1959, Vol. 15, pp. 440–468.

# ANALIZA PORÓWNAWCZA TESTU NIEZALEŻNOŚCI DLA TABLIC DWUDZIELCZYCH

## STRESZCZENIE

W artykule zbadano sześć testów niezależności dla tablic dwudzielczych, do których należą: test chi-kwadrat Pearsona, test największej wiarygodności, test Neymana, test Kullbacka-Leiblera, test Freemana-Tukeya, test Cressiego-Reada. Dokonano tego metodą Monte Carlo. Wyniki są do siebie podobne. Na wyróżnienie zasługują test chi-kwadrat Pearsona oraz test Cressiego-Reada. Można uznać, że jakość tych dwóch testów (wyrażona w funkcji mocy) jest porównywalna, ale lepsza od pozostałych. Index $\tau$ Goodmana-Kruskala wykorzystano do badania zależności w tablicach dwudzielczych za pomocą eksperymentów Monte Carlo.

Słowa kluczowe:
tablice dwudzielcze, test niezależności, metoda Monte Carlo.