Aplikace matematiky

Jiří Anděl

On interactions in contingency tables

Aplikace matematiky, Vol. 18 (1973), No. 2, 99–109

Persistent URL: `http://dml.cz/dmlcz/103456`

Terms of use:

© Institute of Mathematics AS CR, 1973

Institute of Mathematics of the Czech Academy of Sciences provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these *Terms of use*.

This document has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library* `http://dml.cz`

ON INTERACTIONS IN CONTINGENCY TABLES

JIŘÍ ANDĚL

(Received June 15, 1972)

The concept of a generalized logarithmic interaction is introduced. Šidák's method for multiple comparisons of the logarithmic interactions is considered. The proposed procedures are applied to Simpson's example. Finally, the logarithmic interactions are proposed for comparing 2 × 2 contingency tables.

1. INTRODUCTION

Statistical methods based on contingency tables are frequently used in practice. The classical χ^2-test and Fisher's factorial test for independence have been for a long time the main tools for detecting an association. In the present time the following methods are used besides them:

(a) The method based on maximal likelihood (see [8]).

(b) The method based on the information theory (see [6]).

(c) The Bayes method (see [7]).

(d) The method based on logarithmic models (which uses the analysis of variance).

(e) The method based on interactions.

The concept of the interaction was introduced by Bartlett [1] in 1935. Fisher's paper [3] initiated the investigation of the statistical properties of the interactions. Especially, Goodman's papers play an important role in the development of the theory of interactions during the last ten years. Some of them are the basis for the present paper.

We shall investigate contingency tables where no marginal total are fixed. Let us explain the basic concepts on a two-dimensional contingency table

(1)

$n_{11}\ n_{12} \dots n_{1c}$	$n_{1.}$
$n_{21}\ n_{22} \dots n_{2c}$	$n_{2.}$
$\dots\dots\dots$	$\dots$
$n_{r1}\ n_{r2} \dots n_{rc}$	$n_{r.}$
$n_{.1}\ n_{.2} \dots n_{.c}$	n

where

$$n_{i.} = \sum_{j=1}^{c} n_{ij}\,, \quad n_{.j} = \sum_{i=1}^{r} n_{ij}\,, \quad n = \sum_{i=1}^{r} n_{i.} = \sum_{j=1}^{c} n_{.j}\,.$$

In the sequel we shall assume that $n_{ij} > 0$ for all i, j.

For $r = c = 2$ we have a 2×2 table

(2)

n_{11}	n_{12}
n_{21}	n_{22}

.

The number

$$b = n_{11}n_{22}/n_{12}n_{21} \tag{3}$$

will be called the interaction in a 2×2 table (some authors call it "cross-product ratio"). Edwards [2] proved that under reasonable assumptions any measure of association in a 2×2 table must be a function of b. A good interpretation of the interaction b is the ratio of chances $(n_{11} : n_{21})/(n_{12} : n_{22})$.

Generally, let us have a matrix $\boldsymbol{\alpha} = (\alpha_{ij})_{i=1}^{r}\,_{j=1}^{c}$, where $\boldsymbol{\alpha} \neq \mathbf{0}$, $\sum_{i=1}^{r} \alpha_{ij} = 0$ for $j = 1, 2, \dots, c$, $\sum_{j=1}^{c} \alpha_{ij} = 0$ for $i = 1, 2, \dots, r$.

Define

$$d_\alpha = \sum_{i=1}^{r} \sum_{j=1}^{c} \alpha_{ij} \ln (n_{ij}/n) \tag{4}$$

and

$$b_\alpha = \exp \{d_\alpha\}\,. \tag{5}$$

Obviously, $d = \sum_{i=1}^{r} \sum_{j=1}^{c} \alpha_{ij} \ln n_{ij}$, which is more appropriate for calculations.

Then b_α is the interaction (corresponding to the matrix $\boldsymbol{\alpha}$) and d_α the logarithmic interaction. (We differ somewhat from the terminology used in [5].) The special case for a 2×2 table arises, when

$$\boldsymbol{\alpha} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

Because the contingency table (1) is a sample of the size n from the multinomial distribution with the probabilities

$$\begin{matrix} p_{11} & p_{12} & \cdots & p_{1c} \\ p_{21} & p_{22} & \cdots & p_{2c} \\ \cdots & \cdots & \cdots & \cdots \\ p_{r1} & p_{r2} & \cdots & p_{rc} \end{matrix}$$

d_α is an estimate for the theoretical logarithmic interaction

$$\delta_\alpha = \sum_{i=1}^{r} \sum_{j=1}^{c} \alpha_{ij} \ln p_{ij}$$

and b_α is an estimate for the theoretical interaction

$$\beta_\alpha = \exp \{\delta_\alpha\} .$$

2. GENERALIZED LOGARITHMIC INTERACTIONS

Let $(n_{i_1 \ldots i_h})$ be an h-dimensional contingency table with positive elements, corresponding to the probabilities $(p_{i_1 \ldots i_h})$, where $i_j = 1, 2, \ldots, k_j$; $j = 1, 2, \ldots, h$. Denote $n = \sum_{i_1} \ldots \sum_{i_h} n_{i_1 \ldots i_h}$. Let $\boldsymbol{\alpha} = (\alpha_{i_1 \ldots i_h})$ be such a system of real numbers that

$$\sum_{i_1} \cdots \sum_{i_h} \alpha_{i_1 \ldots i_h} = 0 , \quad \sum_{i_1} \cdots \sum_{i_h} |\alpha_{i_1 \ldots i_h}| > 0 .$$

Then

$$d_\alpha = \sum_{i_1} \cdots \sum_{i_h} \alpha_{i_1 \ldots i_h} \ln n_{i_1 \ldots i_h}$$

will be called the generalized logarithmic interaction and $b_\alpha = \exp \{d_\alpha\}$ the generalized interaction. Analogously

$$\delta_\alpha = \sum_{i_1} \cdots \sum_{i_h} \alpha_{i_1 \ldots i_h} \ln p_{i_1 \ldots i_h}$$

will be the generalized theoretical logarithmic interaction and $\beta_\alpha = \exp \{\delta_\alpha\}$ the generalized theoretical interaction. If, in addition, $\sum_{i_j} \alpha_{i_1 \ldots i_h} = 0$ for $j = 1, 2, \ldots, h$, then b_α, β_α are called (normal) interactions and d_α, δ_α (normal) logarithmic interactions without the adjective "generalized"; b_α and d_α are the sample values, β_α and δ_α the theoretical ones.

Theorem 1. *Let χ_t^2 be random variable having the chi-square distribution with t degrees of freedom. For $p \in (0, 1)$ define $\chi_t^2(p)$ by the formula*

$$P(\chi_t^2 > \chi_t^2(p)) = p .$$

Put

$$S_{d_\alpha}^2 = \sum_{i_1} \dots \sum_{i_h} (\alpha_{i_1 \dots i_h})^2 / n_{i_1 \dots i_h} .$$

Then the probability that

$$|d_\alpha - \delta_\alpha| / S_{d_\alpha} \leqq \chi_t(p) \tag{6}$$

holds for all logarithmic interactions (or generalized logarithmic interactions) simultaneously, converges to $1 - p$ *for* $n \to \infty$. *Here* $t = (k_1 - 1)(k_2 - 1) \dots \dots (k_h - 1)$ *for the logarithmic interactions and* $t = k_1 k_2 \dots k_h - 1$ *for the generalized ones.*

Proof. All possible tables $\boldsymbol{\alpha} = (\alpha_{i_1 \dots i_h})$ form a linear space of the dimension t (excluding the null-element). Take a basis

$$\boldsymbol{\alpha}_1 = (\alpha^1_{i_1 \dots i_h}), \dots, \quad \boldsymbol{\alpha}_t = (\alpha^t_{i_1 \dots i_h})$$

and consider the corresponding $\delta_1, \dots, \delta_t$ and $d_1, \dots, d_t$. Denote $\boldsymbol{\delta} = (\delta_1, \dots, \delta_t)'$, $\boldsymbol{d} = (d_1, \dots, d_t)'$. Put

$$a_{qs} = \sum_{i_1} \dots \sum_{i_h} \alpha^q_{i_1 \dots i_h} \alpha^s_{i_1 \dots i_h} / p_{i_1 \dots i_h} ,$$

$$a^n_{qs} = n \sum_{i_1} \dots \sum_{i_h} \alpha^q_{i_1 \dots i_h} \alpha^s_{i_1 \dots i_h} / n_{i_1 \dots i_h} ,$$

$$\boldsymbol{A} = (a_{qs})^t_{q=1}{}^t_{s=1} , \quad \boldsymbol{A}_n = (a^n_{qs})^t_{q=1}{}^t_{s=1} .$$

From the asymptotic normality of the multinomial distribution and theorem 6a.2 (III) in Rao's book [8] it follows that the random vector $n^{1/2}(\boldsymbol{d} - \boldsymbol{\delta})$ has asymptotically the normal distribution $N_t(\boldsymbol{0}, \boldsymbol{A})$. Denote by F_n the distribution function of $n^{1/2}(\boldsymbol{d} - \boldsymbol{\delta})$ and by H_n that of the normal distribution $N_t(\boldsymbol{0}, \boldsymbol{A}_n)$. From the fact that

$$\lim_{n \to \infty} \sup_{x_1 \dots x_t} |F_n - H_n| = 0 \tag{7}$$

(see (6a.2.11) in [8]) we conclude that H_n converges to the distribution function of $N_t(\boldsymbol{0}, \boldsymbol{A})$ and $n(\boldsymbol{d} - \boldsymbol{\delta})' \boldsymbol{A}_n^{-1} (\boldsymbol{d} - \boldsymbol{\delta})$ has asymptotically the chi-square distribution with t degrees of freedom provided $\boldsymbol{A}_n$ is regular. Now take $\boldsymbol{v} = (v_1, \dots, v_t)' \neq \boldsymbol{0}$ and denote

$$\alpha_{i_1 \dots i_h} = \sum_{q=1}^{t} v_q \alpha^q_{i_1 \dots i_h} , \quad \boldsymbol{\alpha}_0 = (\alpha_{i_1 \dots i_h}) .$$

Because $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_t$ is a basis and $\boldsymbol{v} \neq \boldsymbol{0}$, we see that $\boldsymbol{\alpha}_0 \neq \boldsymbol{0}$. After elementary computation we get

$$\boldsymbol{v}' \boldsymbol{A}_n \boldsymbol{v} = n \sum_{i_1} \dots \sum_{i_h} (\alpha_{i_1 \dots i_h})^2 / n_{i_1 \dots i_h} > 0 . \tag{8}$$

Thus the matrix $\boldsymbol{A}_n$ is positive definite and, therefore, regular.

If $\boldsymbol{v}$ varies over the t-dimensional space (excluding the null-element) then $\boldsymbol{v}'\boldsymbol{d}$ gives all logarithmic interactions (normal or generalized).

It is well known that the events

$$(9) \qquad n(\boldsymbol{d} - \boldsymbol{\delta})' \boldsymbol{A}_n^{-1}(\boldsymbol{d} - \boldsymbol{\delta})/\chi_t^2(p) \leqq 1$$

and

$$(10) \qquad [\boldsymbol{v}'(\boldsymbol{d} - \boldsymbol{\delta})]^2 \leqq \boldsymbol{v}'\boldsymbol{A}_n\boldsymbol{v}\, \chi_t^2(p)/n \quad \text{for all} \quad \boldsymbol{v}$$

are equivalent. A geometric proof is in [9], Appendix III, but the equivalence can be proved very easily and shortly using the Schwarz inequality. The probability of (9) converges to $1 - p$ for $n \to \infty$ according to our previous consideration. As for (10), $d_0 = \boldsymbol{v}'\boldsymbol{d}$ is a logarithmic interaction (normal or generalized) belonging to $\boldsymbol{\alpha}_0$, $\delta_0 = \boldsymbol{v}'\boldsymbol{\delta}$ its theoretical value in the population and $\boldsymbol{v}'\boldsymbol{A}_n\boldsymbol{v}/n = S_{d_0}^2$ according to (8). The proof is finished.

It happens that we consider a few logarithmic interactions only. Then the simultaneous bounds for their theoretical values given in (6) could be rather wide. For such a case Goodman derived the following Theorem 2.

Theorem 2. *Let Φ denote the distribution function of $N(0, 1)$ and u its inverse. Then the probability that the intervals*

$$(11) \qquad (d_i - u(1 - p/2w)\, S_{d_i}, \quad d_i + u(1 - p/2w)\, S_{d_i})$$

cover δ_i for all $i = 1, 2, \ldots, w$ simultaneously, is asymptotically at least $1 - p$.

Proof. See [5]. The proof is based on Tukey's theorem employing Bonferroni inequality (see [11], 10.5(a)).

We give another theorem concerning this problem.

Theorem 3. *Put*

$$(12) \qquad c = u(\tfrac{1}{2} + \tfrac{1}{2}(1 - p)^{1/w})\,.$$

Then the probability that the intervals

$$(13) \qquad (d_i - cS_{d_i}, d_i + cS_{d_i})$$

cover δ_i for all $i = 1, 2, \ldots, w$ simultaneously, is asymptotically at least $1 - p$.

Proof. Denote $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_w)'$, $\boldsymbol{d} = (d_1, \ldots, d_w)'$. Let d_k and δ_k correspond to $\boldsymbol{\alpha}_k = (\alpha^k_{i_1 \ldots i_h})$, $k = 1, 2, \ldots, w$. Put

$$a^n_{qs} = n \sum_{i_1} \ldots \sum_{i_h} \alpha^q_{i_1 \ldots i_h} \alpha^s_{i_1 \ldots i_h} / n_{i_1 \ldots i_h} \,,$$

$$\mathbf{A}^*_n = (a^n_{qs})^w_{q=1}{}^w_{s=1} \,.$$

Denote by F^*_n the distribution function of $n^{1/2}(\mathbf{d} - \boldsymbol{\delta})$ and by H^*_n that of $N_w(\mathbf{0}, \mathbf{A}^*_n)$. Similarly as (7) in the proof of Theorem 1 we obtain that

$$\lim_{n \to \infty} \sup_{x_1, \ldots, x_w} |F^*_n - H^*_n| = 0 \,. \tag{14}$$

Let $\mathbf{Y} = (Y_1, \ldots, Y_w)'$ have $N_w(\mathbf{0}, \mathbf{A}^*_n)$. Then, according to Šidák's theorem [10]

$$P(|Y_1| < c_1, \ldots, |Y_w| < c_w) \geqq P(|Y_1| < c_1) \ldots P(|Y_w| < c_w)$$

holds. Denote $\operatorname{var} Y_i = \sigma^2_i$, $i = 1, 2, \ldots, w$. (Obviously $\sigma^2_i = nS^2_{d_i}$.) Then

$$P(|Y_1| < c\sigma_1, \ldots, |Y_w| < c\sigma_w) \geqq P(|Y_1| < c\sigma_1) \ldots P(|Y_w| < c\sigma_w) = 1 - p \,.$$

But (14) implies that

$$P(|d_1 - \delta_1| < cS_{d_1}, \ldots, |d_w - \delta_w| < cS_{d_w}) - P(|Y_1| < c\sigma_1, \ldots, |Y_w| < c\sigma_w)$$

converges to zero for $n \to \infty$. This concludes the proof.

It is easy to see that Theorem 2 and Theorem 3 hold for generalized logarithmic interactions, too. Moreover, both of them are valid even in the case when some logarithmic interactions are evaluated from marginal contingency tables. It suffices to notice that such a vector of logarithmic interactions has asymptotically a simultaneous normal distribution.

It is evident that both of the methods described in Theorem 2 and 3 hold for $w = 1$, too.

In view of

$$u(1 - p/2w) \geqq u(\tfrac{1}{2} + \tfrac{1}{2}(1 - p)^{1/w}) \,, \quad 0 < p < 1 \,, \quad w \geqq 1 \,, \tag{15}$$

we see that Theorem 3 always gives better results compared to Theorem 2 (although the differences are small in practical cases). The inequality (15) can be proved either by a direct calculation or from the fact that Theorem 2 is essentially based on the Bonferroni inequality

$$P(|Y_1| < c_1, \ldots, |Y_w| < c_w) \geqq 1 - \sum_{i=1}^{w} P(|Y_i| \geqq c_i) \,,$$

but clearly the same inequality holds also for $P(|Y_1| < c_1) \ldots P(|Y_w| < c_w)$ so that this last product of probabilities yields a closer lower bound for $P(|Y_1| > c_1, \ldots \ldots, |Y_w| < c_w)$.

3. EXAMPLE

Let us analyse the following tables:

Male

	Treated	Untreated
Alive	800	400
Dead	500	300

Female

	Treated	Untreated
Alive	1200	200
Dead	1500	300

It is the classical Simpson's example (see [2]), only the frequencies were multiplied by 100. We have to do with $2 \times 2 \times 2$ contingency table (suppose the case where no marginal totals are fixed). The "natural" interaction (800.300)/(400.500) evaluated from the first subtable using (3) appears to be the generalized interaction considered in the whole table corresponding to

$$\boldsymbol{\alpha} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

It is the main reason why the generalized interactions were introduced. They make it possible to consider all the interesting subtables separately. (Another application of the generalized interactions consists in the possibility to test the hypothesis that the probability that a treated male is living equals to 0.75, say.) Let us test the following hypotheses on the usual 5% level of significance:

(1) The survival of males does not depend on the treatment.

(2) The survival of females does not depend on the treatment.

(3) The survival of treated subjects does not depend on the sex.

(4) The survival of untreated subjects does not depend on the sex.

(5) The treatment of living subjects does not depend on the sex.

(6) The treatment of dead subjects does not depend on the sex.

Obviously, the answers may be based on the generalized logarithmic interactions. We have

$$\chi_7(0.05) = 3.755\ ; \quad u(1 - 0.05/12) = 2.639\ ;$$

$$u(\tfrac{1}{2} + \tfrac{1}{2} \,.\, 0.95^{1/6}) = 2.633\ .$$

Thus, the method described in Theorem 3 is the most favourable and will be used.

For each hypothesis we shall write $\boldsymbol{\alpha}$, d, confidence interval (c.i.) for δ based on (13), and χ^2 calculated for the corresponding 2×2 table by the usual way. The values of χ^2 are mentioned for comparison only, they can't be used for simultaneous testing.

(1) $$\boldsymbol{\alpha} = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \quad d = 0.1823,$$

$$\text{c.i.} \quad (-0.0686, 0.4332), \quad \chi^2 = 3.663.$$

(2) $$\boldsymbol{\alpha} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \quad d = 0.1823,$$

$$\text{c.i.} \quad (-0.0788, 0.4434), \quad \chi^2 = 3.386.$$

(3) $$\boldsymbol{\alpha} = \begin{pmatrix} 1 & 0 \\ -1 & 0 \end{pmatrix}, \begin{pmatrix} -1 & 0 \\ 1 & 0 \end{pmatrix}, \quad d = 0.6931,$$

$$\text{c.i.} \quad (0.5116, 0.8746), \quad \chi^2 = 102.56.$$

(4) $$\boldsymbol{\alpha} = \begin{pmatrix} 0 & 1 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 0 & -1 \\ 0 & 1 \end{pmatrix}, \quad d = 0.6931,$$

$$\text{c.i.} \quad (0.3797, 1.0065), \quad \chi^2 = 34.29.$$

(5) $$\boldsymbol{\alpha} = \begin{pmatrix} 1 & -1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} -1 & 1 \\ 0 & 0 \end{pmatrix}, \quad d = -1.0986,$$

$$\text{c.i.} \quad (-1.3564, -0.8408), \quad \chi^2 = 132.06.$$

(6) $$\boldsymbol{\alpha} = \begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ -1 & 1 \end{pmatrix}, \quad d = -1.0986,$$

$$\text{c.i.} \quad (-1.3530, -0.8442), \quad \chi^2 = 135.42.$$

Our generalized logarithmic interactions are chosen in the present example in such a way that their theoretical values under independence are zeros. This always holds for non-generalized logarithmic interactions but it is not the general property of the generalized ones.

As for the given example, we see that the hypotheses (3), (4), (5) and (6) should be rejected.

Somebody could prefer the following complex of hypotheses:

$$(1') = (1), \quad (2') = (2), \quad (3') = (3), \quad (4') = 4,$$

(5′) The treatment does not depend on the sex regardless to living.

(6′) The degree of the association (measured by interaction) between treatment and living is the same for both male and female.

We use Theorem 3 again. The above results (1)–(4) hold for (1′)–(4′), too. The test of (5′) is based on the marginal contingency table for which $d = -1.07$ and the confidence interval (13) is $(-1.25, -0.89)$. Under independence the theoretical value equals to zero. Thus the hypothesis (5′) is rejected.

	Male	Female
Treated	1300	2700
Untreated	700	500

The test of (6′) is based on the ratio of interactions the subtables

$$\frac{800.300}{500.400} \Big/ \frac{1200.300}{1500.200} \tag{16}$$

or on its logarithm. But (16) corresponds to

$$\alpha = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}, \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$$

and we get $d = 0$; the confidence interval is $(-0.3621, 0.3621)$.

4. ON A METHOD FOR COMPARING 2 × 2 CONTINGENCY TABLES

Some methods for comparing contingency tables were considered in [4]. We propose a method based on logarithmic interactions.

Let us have j contingency tables

$$\begin{array}{|cc|}\hline a_1 & b_1 \\ c_1 & d_1 \\ \hline\end{array} \quad \cdots \quad \begin{array}{|cc|}\hline a_k & b_k \\ c_k & d_k \\ \hline\end{array}$$

which can be considered as independent samples from k populations. We want to test the hypothesis that all the populations have the same interaction which (a) is supposed to be known and equal to δ, (b) is not known. Denote d_i, and δ_i, the logarithmic interaction, and the theoretical logarithmic interaction in the i-th table, respectively. The value $S^2_{d_i}$ is given in this special case by the formula $S^2_{d_i} = 1/a_i + 1/b_i + 1/c_i + 1/d_i$.

(a) Let $k \geqq 1$. From the asymptotic normality of $d_1, \ldots, d_k$ (when $a_i + b_i + c_i + d_i \to \infty$ for $i = 1, 2, \ldots, k$) and their independence it follows that

$$T = \sum_{i=1}^{k} (d_i - \delta)^2 / S_{d_i}^2$$

has asymptotically the chi-square distribution with k degrees of freedom. Therefore, if $T > \chi_k^2(p)$, we reject the hypothesis. (The case $\delta = 0$ can often occur.)

(b) When δ is not known, the test procedure is based on the following Theorem 4.

Theorem 4. *Put*

$$\hat{d} = (\sum_{i=1}^{k} 1/S_{d_i}^2)^{-1} \sum_{i=1}^{k} d_i/S_{d_i}^2 .$$

If $\delta_1 = \ldots = \delta_k$ and $a_i + b_i + c_i + d_i \to \infty$ *for* $i = 1, 2, \ldots, k$, *then the random variable*

$$V = \sum_{i=1}^{k} (d_i - \hat{d})^2 / S_{d_i}^2$$

has asymptotically the chi-square distribution with $k - 1$ *degrees of freedom.*

Proof. Let $X_1, \ldots, X_k$ be independent random variables, where X_i has $N(\mu, \sigma_i^2)$, $i = 1, 2, \ldots, k$. We assume that $\sigma_1^2, \ldots, \sigma_k^2$ are known positive constants and μ is an unknown parameter. Remark that

$$\hat{\mu} = (\sum_{i=1}^{k} 1/\sigma_i^2)^{-1} \sum_{i=1}^{k} X_i/\sigma_i^2$$

is the best linear unbiassed estimator for μ. Applying theorem 3b.4(II) in [8] we can prove that the random variable

$$\sum_{i=1}^{k} (X_i - \hat{\mu})^2 / \sigma_i^2$$

has the chi-square distribution with $k - 1$ degrees of freedom. The rest of the proof follows from the asymptotic normality of the random variables $d_1, \ldots, d_k$.

References

[1] *Bartlett M. S.* (1935): Contingency table interactions. J. Roy. Stat. Soc. Supp. 2, 248—252.

[2] *Edwards A. W. F.* (1963): The measure of association in a 2 × 2 table. J. Roy. Stat. Soc. ser. A 126, 109—114.

[3] *Fisher R. A.* (1962): Confidence limits for a cross-product ratio. Austral. J. Statist. 4, 41.

[4] *Goodman L. A.* (1963): On methods for comparing contingency tables. J. Roy. Stat. Soc. ser. A 126, 94—108.

[5] *Goodman L. A.* (1964): Simultaneous confidence limits for cross-product ratios in contingency tables. J. Roy. Stat. Soc. ser. B 26, 86—102.

[6] *Kullback S.* (1959): Information theory and statistics. Wiley, New York.
[7] *Lindley D. V.* (1964): The Bayesian analysis of contingency tables. Ann. Math. Statist. 35, 1622—1643.
[8] *Rao C. R.* (1965): Linear statistical inference and its applications. Wiley, New York.
[9] *Scheffé H.* (1959): The analysis of variance. Wiley, New York.
[10] *Šidák Z.* (1967): Rectangular confidence regions for the means of multivariate normal distributions. J. Amer. Stat. Assoc. 62, 626—633.
[11] *Wilks S. S.* (1962): Mathematical statistics. Wiley, New York.

Souhrn

O INTERAKCÍCH V KONTINGENČNÍCH TABULKÁCH

JIŘÍ ANDĚL

Budiž $(n_{i_1 \ldots i_h})$ h-rozměrná kontingenční tabulka s kladnými četnostmi a $\boldsymbol{\alpha} = (\alpha_{i_1 \ldots i_h})$ h-rozměrná tabulka reálných čísel taková, že platí

$$\sum_{i_1} \cdots \sum_{i_h} \alpha_{i_1 \ldots i_h} = 0 \,, \quad \sum_{i_1} \cdots \sum_{i_h} |\alpha_{i_1 \ldots i_h}| > 0 \,.$$

Nechť daná kontingenční tabulka nemá žádné pevné marginální četnosti, takže ji lze považovat za výběr z multinomického rozdělení s kladnými pravděpodobnostmi $(p_{i_1 \ldots i_h})$. Položme

$$\delta_\alpha = \sum_{i_1} \cdots \sum_{i_h} \alpha_{i_1 \ldots i_h} \ln p_{i_1 \ldots i_h} \,, \quad d_\alpha = \sum_{i_1} \cdots \sum_{i_h} \alpha_{i_1 \ldots i_h} \ln n_{i_1 \ldots i_h} \,,$$

$$\beta_\alpha = \exp\{\delta_\alpha\} \,, \quad b_\alpha = \exp\{d_\alpha\} \,.$$

Pak b_α a d_α se nazývají zobecněná interakce, resp. zobecněná logaritmická interakce; β_α a δ_α jsou teoretická zobecněná interakce, resp. teoretická zobecněná logaritmická interakce. Simultánní intervaly spolehlivosti pro zobecněné teoretické logaritmické interakce lze konstruovat pomocí metod uvedených ve větách 1, 2 a 3. Je uveden Simpsonův příklad týkající se tabulky $2 \times 2 \times 2$, který je navrhovanými metodami podrobně vyhodnocen. Poslední část práce je věnována testu hypotézy, že k nezávisle pořízených čtyřpolních kontingenčních tabulek pochází ze základních souborů se stejnou interakcí. Přitom se rozlišují dva případy. První nastává tehdy, je-li tato společná interakce součástí hypotézy (je-li tedy dána), druhý případ se týká situace, kdy tato hodnota není dána. Navržený test je v obou případech založen na logaritmických interakcích.

Author's address: RNDr. *Jiří Anděl,* CSc., Matematicko-fyzikální fakulta Karlovy university, Sokolovská 83, 186 00 Praha 8 - Karlín.