ISAHP 1996, Vancouver, Canada, July 12-15, 1996

# THE AGREEMENT DEGREE OF ESTIMATIONS SET WITH REGARD FOR EXPERTS' COMPETENCE

V. G. Totsenko
Institute of Information Recording Problems
National Academy of Sciences
Kiev, Ukraine 252113

Abstract : The sets of expert estimates which are obtained directly or are results of processing of numerical equivalents of qualitative estimations are considered. The notions for an agreement coefficient for the expert estimation set are determined, as are detection thresholds defining the exactness for generalized estimations. The algorithms for calculation of agreement coefficient and detection and application thresholds with regard for experts' competence are suggested. An algorithm for extraction of a significant subset of expert estimations is developed.

## 1 Introduction

At the application of the method of analytical hierarchical processes (AHP) [1] is necessary to determine the expert estimations of the elements influence of hierarchy certain level onto the elements of neighboring higher levels. These estimates are received by the method of pair-wise comparisons with the subsequent processing of comparisons matrix. By results of this processing are the coefficients of influence which are a positive real numbers, not exceeding 1.

For rising the certainty of expert estimations are used some experts. But this goal can be achieved, if the estimations given by the different independent experts would be sufficiently agreed.

The determination of agreement of an expert estimations (EE) set precedes the stage of resulting estimation determination and seeks to determine the opportunity of using this set for obtaining such estimation and/or determination of some significant of its subset using which it is possible to determine the resulting estimation.

Without loss of generality we should consider that EE represent graduation numbers of some scale with $n$ graduations. If an estimation obtained from an expert directly or by transformations of EE is presented by a real number, then having been given the admissible error $\varepsilon$ it is not difficult to pass to the representation of this estimation as a scale graduation number with $n = 1/2\varepsilon$ graduations.

We should review the known methods from positions of their applicability to the agreement estimation of expert estimations of such type.

When solving the problem of agreement estimation it is necessary to solve at least two main problems. The first one consists in finding the way of answering the question: does an expert estimations set carry information or it is "information noise"? The second problem is in calculating the quantitative measure for agreement degree of an expert estimations set.

Among the methods for solution of the first problem let us note first of all those which are based on the application of Kendall's rank correlation coefficient [3-6], Spearman's coefficient [3, 7] or concordance coefficient [3]. The Kendall's rank correlation coefficient is bound up with another statistics: Mann-Whitney statistic [8], Pearson statistic $x^2$ for the case of linked ranks [3, 9], Kemeny-Snell distance [10].

The mentioned methods were developed for the agreement degree estimation of direct ranging results as well for the application of pair-wise comparisons. The admissible by these methods number of different relations does not exceed 3. In [1] is considered a problem for agreement degree estimation of pair-wise comparisons results in the 9-point numeric scale of preferences degrees. As an agreement index it is suggested to use the difference between the maximum eigenvalue of alternatives preferences matrix and the alternatives number. For estimating the presence of information it is proposed to compare the agreement index of comparisons results obtained by an expert with the agreement index calculated for a similar random numbers matrix.

The main difficulty of applying such methods is just bound up with the complexity for the magnitude justification of the degree threshold agreement value. Earlier the problem for calculation of the computability coefficient of pair-wise comparisons results by one expert defining the transitivity breakdown degree of objects preferences relations was considered in [3, 11]. The method is framed on the assumption that the results of comparisons are presented in the binary scale (1-exceeds, 0-concedes); the equivalence relation is not provided what does not permit to apply it at agreement estimation of expert estimations expressed in the multivalued scales.

For solving the second problem, i.e. the quantitative estimation of agreement degree there is developed a large number of methods.

In [2,3] it is suggested the opinion agreement of $m$ experts performing the pair-wise comparison and presenting the results of comparisons in the binary scale to estimate by the fit coefficient $u$ taking a value 1 at full concordance of opinions and at the absence of concordance it takes a value $u_{min} = 1/m - 1$ at an even $m$ and $u_{min} = 1/m$ at an odd one.

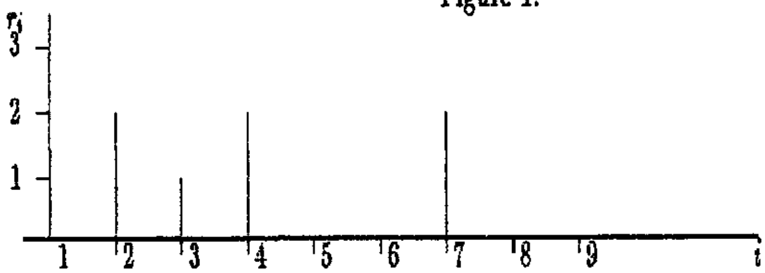
For the case when is determined the opinion agreement of $m$ experts each of which is ranging $n$ objects in [3] is suggested to use the fit coefficient expressed in terms of Kendall's rank correlation coefficients for expert's pairs. The generalization of this method for the case of the equivalence assumption is performed in [9]. It is possible similarly to frame expressions for determining the agreement coefficient on the base of Spearman fit coefficient.

In the case, when is applied the direct ranging of some alternatives by many experts, then for the agreement degree estimation of EE can be used a measure of answers variation proposed in [12].

There are promising the approaches to the agreement degree estimation based on the more detailed description of experts preferences structure based particularly on the entropy approach [14]. The method is developed for the case of direct ranging as well for pair-wise comparisons results of which are presented in the binary scale. The opportunities for assignment of equal ranks are not taken into account. Because of this the method is not applicable for solving the earlier mentioned problems.

In the paper at first "from the positions of common sekse" are formulated requirements to the agreement coefficient of expert estimations set and is synthesized an expression for its calculation. Further are validated approaches to determination of threshold values of agreement coefficient of expert estimations set and are suggested algorithms for calculation of these quantities. In conclusion is suggested the procedure of resulting estimation calculation.

Figure 1.




## 2 The agreement coefficient of expert estimations set

Let there be given an ordered by experts numbers set $V = \{v_j\}$, $j = (1, m)$ of expert estimations performed by $m$ experts. The estimations represent the numbers of scale graduations with $n$ graduations. It is necessary to estimate quantitatively the agreement degree of the set $V$.

Let us determine at first from purely qualitative reasonings the requirements to the coefficient $k(V)$ used as a quantitative estimation for agreement degree of the set $V$. It is convenient for this to represent the set $V$ by the $n$ -component vector $R = r_i$, $i = (1, n)$, where $r_i$ is the number of experts having shown as an estimation the $i$-th scale graduation.

For obviousness it is convenient to represent the vector $R$ as a diagram representing the axis $i$ on which in the points $\delta, \varphi, ..., \psi$ are constructed perpendiculars with a length of $r_\delta$, $r_\varphi$,..., $r_\psi$ (fig. 1-4).

Let be as an example: the number of experts is $m = 7$; the number of scale graduations is $n = 9$ ; $V = (2;2;3;4;4;7;7)$; Then: $R = (0;2;1;2;0;0;2;0;0;)$.

In the fig. 1 is shown a diagram corresponding to the vector $R$.

Intuitively it is clear that to the most agreement degree corresponds a case when the estimations of all experts is equal (because of that the only one component of vector $R$ is equal to $m$, and all the rest are equal to 0, fig. 2), and to the least agreement corresponds the case when the estimations of all $m$ experts are different and are uniformly distributed over the scale, i.e. each of $n$ scale graduations as an estimation called $m/n$ experts (fig. 3).

Without loss of generality it is possible to consider the least agreed the estimation set in which each scale graduation as an estimation showed exactly one expert (fig. 3 at $m = n$).

Besides such distribution it appears to consider mismatched to the utmost as well the distribution of EE in which $m$ experts divided into $k$ groups of experts who gave the same estimations, such that the minimum and maximum estimations gave for one group of experts and in addition the intervals between the estimations of adjacent groups are the same (in the fig. 4 is shown an

Figure 2.

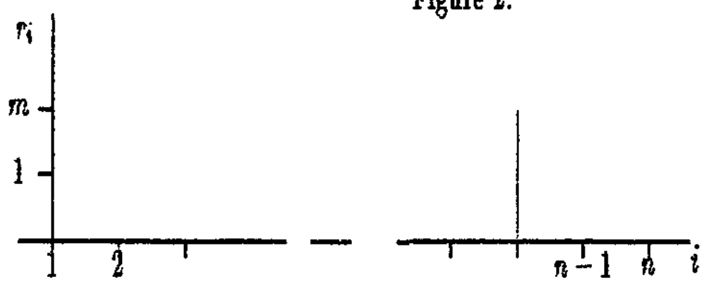
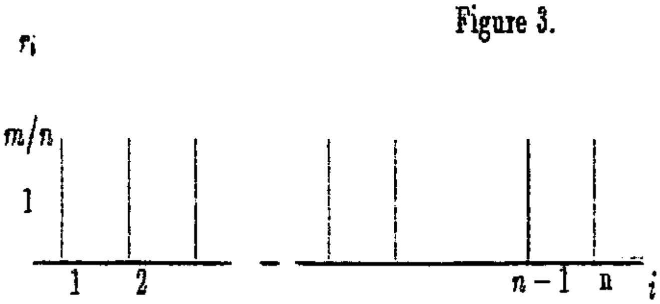



Figure 3.




example of such distribution).

This statement follows from the fact that such distributions are transformed evidently into the mismatched to the utmost distribution of the first type, if to transform the original scale with $n$ graduations into the scale with $m$ graduations having eliminated from the initial scale those graduations which were not called as an estimation by no one expert.

Let us seek an expression for $k(V)$ a composition of some components. Let us require that the first component $\varphi(V)$ at the mismatched to the utmost distribution would take on the maximum value and at the matched to the utmost distribution it would take on the value of $\varphi(V) = 0$.

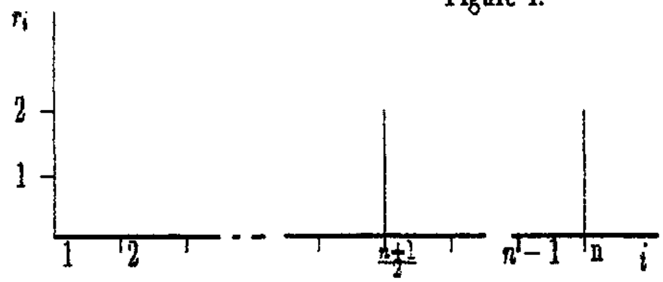
One of functions satisfying these requirements is the entropy function

$$\varphi(V) = \sum_{j=1}^{n} r_j/m \cdot log_2 r_j/m; (1)$$

The function $\varphi(V)$ "feels" the change of group number of coincident estimations in each group. But the differences of estimation values do not change it. At the same time it is clear intuitively that the increase of these differences (at the same number of groups) is taken as a decrease of agreement degree for a set $V$ (for, example, the vector $R$ shown in the fig. 5 is taken as a more agreed one than the vector $R$ shown in the fig. 6).

In addition it is necessary to provide the invariance of the agreement function value at the simultaneous and the same change of all estimations by the same value $\eta$ (it is not difficult to see

Figure 4.




that in this case on the distribution diagram all its components will shift by the value of $\eta$ ). This property can be provided by introducing into the general function of agreement degree estimation a component representing a sum of even powers of weighed differences for estimations and the mean estimation $a$:

$$\chi(V)=\sum_{i=1}^{n} r_i|i-a|;(2)$$

(it is supposed that the scale is uniform and each of its graduations can be chosen as an estimation).

The expression for calculation of mean estimation is determined by the scale type. For scale of intervals, relations, differences or absolute by calculation mean estimation is used an expression for arithmetic mean. For order, hyperorder, nominal and number scales the arithmetic mean finding operation is not correct. In these cases the median is to be chosen as a mean estimation [15].

Let us represent the agreement coefficient $k(V)$ as:

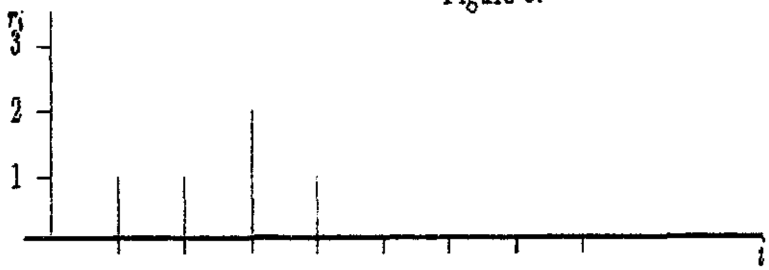
$$k(V)=1-C[\varphi(V),\chi(V)]$$

Let us determine requirements to the function $C$. Obviously, it should increase /decrease monotonically with growth/ decrease of $\varphi(V)$, $\chi(V)$. The simplest functions satisfying this requirement are the sum and the product. It is not difficult to see that these functions are easily transformed one into the other (e.g.by change of the product by the sum of logarithms of arguments). Therefore we assume:

$$C[\varphi(V),\chi(V)]=\varphi(V)+\chi(V),(3)$$

As follows from expressions (1)-(2) the ranges of values for functions $\varphi(V)$, $\chi(V)$ depend substantially not only on character of estimations distribution, but also on the scale graduations number and on the experts number.

Figure 5.




Therefore as an agreement measure we shall use the agreement coefficient equal to:

$$k(V) = 1 - \frac{C(R)}{C(R_0)} \quad (4)$$

where $R_0$ is the estimations vector corresponding intuitively to the worst set $V_0$ of estimations in which each estimation gave exactly one expert; $R$ is the estimations vector corresponding to the set $V$ of expert estimations under investigation.

As follows from (2), the value of function $\chi(V)$ depends substantially on the number $m$ of experts. Because of that to two distributions taken as equally matched, but having different values of $m$, will correspond different values of $\chi(V)$ while the value of $\chi(V_0)$ will remain unchanged in consequence of accepted agreement relative to the structure of $V_0$ and invariability of the number $n$ of scale graduations. This causes the necessity of introducing in (2) the correcting multiplier $1/m$.

For providing the fulfillment of condition $k(V_0) = 0$ let us transform the expression (4) to the form:

$$k(V) = \left(1 - \frac{C(R)}{C(R_0)}\right) Z \quad (5)$$

With regard to (1)-(5) we obtain:

$$k(V) = \left(1 - \frac{\frac{1}{m}\sum_{i=1}^{n} r_i \left| i - \frac{1}{m}\sum_{i=1}^{n} i \cdot r_i \right| - \sum_{i=1}^{n} \frac{r_i}{m} \ln \frac{r_i}{m}}{G \sum_{i=1}^{n} \left| i - \frac{n+1}{2} \right| + \ln n}\right) Z, \quad (6)$$

where $n$ is the number of scale graduations; $m$ is the number of experts; $r_i$ is the number of experts which showed as an estimation the $i$-th scale graduation;

$$G = \frac{m}{\ln(m)\, n \ln(n)}$$ - is the scale coefficient;

$$z = \begin{cases} 1, \ if \ z* = TRUE; \\ 0, \ if \ z^* = FALSE; \end{cases}$$

$$z^* = \overline{(r_1 = r_n = m/k)} \vee \bigvee_{d=1}^{k-1} \overline{(r_{i_d} = r_{i_{d+1}})} \vee \bigvee_{d=1}^{k-1} \overline{(i_d - i_{d+i})} = const \quad (7)$$

$i_d$ is the number of scale gradation choosen as an expert estimation, $d = (1, k)$; $k$ is the number of expert groups who gave the same estimations.

In other words the agreement coefficient is equal to 0 when each from the following conditions is fulfilled: - the number of experts who gave the same estimation is the same and equal to $m/k$; - the minimum and maximum estimations gave $m/k$ experts; - the differences between two any estimations are the same.

From (6)-( 7) follows that the values of agreement coefficient are within the limits of [0,1], and a completely agreed set of expert estimations corresponds the agreement coefficient equal to 1.

# 3 The accounting of experts competence

Numerically the competence of the $j$-th expert we shall estimate by the relative competence coefficient $\zeta_j$ in the group. These coefficients should meet the following conditions:

$$\forall j \, [0 < \zeta_j < 1];$$
$$\sum_{j=1}^{m} \zeta_j = 1.$$

The relative competence coefficients can be determined using the method of pair-wise comparisons [1] with subsequent normalization of obtained values. Thus in the case of the equal experts competence of the group we have:

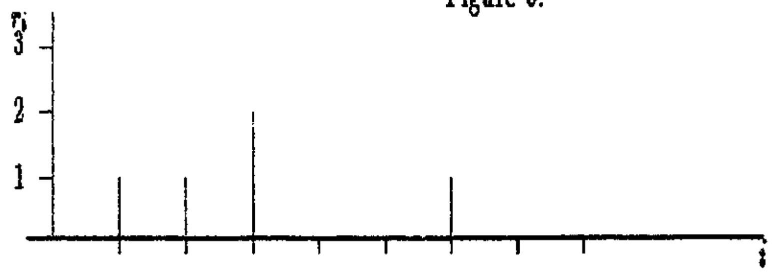
$$\forall j \, [\zeta_j = 1/m].$$

Referring to [6] we notice that in this case the multiplier $r_i/m$ can be interpreted as a sum of competence indices of equally competent experts who gave as an estimation the $i$-th scale graduation. From here follows a rule of expression transformation (6),(7) for the general case of unequal experts competence: the multipliers $r_i$ are to be replaced by $\sigma_i m$, where $\sigma_i$ is the sum of experts competence coefficients who gave as an estimation the $i$-th scale graduation.

# 4 The threshold values of agreement coefficient

At determination of the threshold value $T_0$ of agreement coefficient we shall proceed from the following considerations.

Figure 6.




Let us notice above all that the notion and the procedure for determination of the quantitative value for threshold is introduced with the purpose creation an instrument for selection of significant subsets of expert estimations, i.e. of estimations subsets using which it is possible to calculate consistent, having sense, generalized agreed estimations. In this connection the method for determination of the agreement koefficient threshold is based on the construction procedure of significant subsets $(SS)$ of expert estimations. When constructing the significant subset $V_\varphi$ of some set $V \supseteq V_\varphi$ of $EE$ set $V$ it is necessary to solve two problems.

The first one consists in solving a question: does the set $V$ contain information or this set of $EE$ is "information noise"? If $V$ is information noise, than it is apparent that for extraction of $V_\varphi$ is required the additional information which can be obtained only from experts in the course of repeated examination.

The solution of this problem is carried out by calculation of agreement coefficient $k_c(V)$ of the set $V$ and by comparison its value with the detection threshold $T_0$. If $k_c(V) \geq T_0$,i.e. it is established that $V$ contains information, then to this set may be applied the procedure for extraction of significant subset $V_\varphi$. It consists in successive correction or/and elimination from the set $V$ of "extreme", i.e. the most different from the mean estimations, and in verification of admissibility of the agreement coefficient value $k_c(V_\varphi)$ of the obtained in such way subset.

This verification consists in the comparison of $k_c(V_\varphi)$ with the application threshold $T_a$. If $k_c(V_\varphi) \geq T_a$,then the subset $V_\varphi$ considered to be significant and is used fordetermination of the generalized expert estimation.

The essence of an approach to the determination of detection threshold consists in the fact that the vector $R_r$ of expert estimations carrying the minimum admissible amount of information is constructed and then according to (6) for this vector is determined the value of agreement coefficient which is taken as a detection threshold.

Let us construct the vector $R_r$ on the basis of the vector $R_0$ which corresponds to the intuitively understandable maximum mismatching of experts opinions and therefore does not carry any information. It is also obvious that as $R_\tau$ it is to be taken a vector to which corresponds the mean estimation which differs the least from the estimation calculated for the vector $R_0$. Let us denote this minimal distinguishable difference of estimations with $\delta$.

If the resulting estimation represents an integer ( the number of scale graduation) then at its determination is carried out inevitably the rounding off to the nearest integer. Taking this into account as $R_\tau$ is to be taken the distribution of estimations for a such number of experts with which the generalized estimation $a_\tau$ differs from the generalized estimation $a_0$ of the distribution $R_0$ by the least noticeable value, i.e. by one scale graduation. Such change of the generalized estimation with regard for rounding off will take place if $\delta = 0.5$.

From the distribution $R_0$ we shall construct the distribution $R_\tau$ in the following way: let us eliminate the estimation of one expert being in the vector $R_0$ on the graduation $\eta$ and let us locate it additionally on the graduation $\lambda$ so that owing to this the resulting estimation would be equal to $a + \delta$. For providing the possibility for construction of $R_\tau$ with any scale sizes the magnitude of $\eta$ should be chosen the least admissible, i.e. $\eta = 1$.

Since the vector $R_0$ and $R_\tau$ differ only by the numbers of expert which gave as an estimation the first and the $\lambda$-th scale graduation, then using as a generalized estimation of the arithmetic mean follows:

$\lambda = [\delta n + 1]$

Such method for determination of detection threshold is applicable by using the scales of intervals, relations, differences and absolute scales. By using order, hyperorder, nominal and number scales the vector $R_\tau$ is constructed starting from the fact that for such scales the median is the generalized estimation.

Let us consider the procedure for determination of the applicability threshold.

DEFINITION.

The applicability threshold $T_u$ is called the agreement coefficient value of the set $V_2$ for estimations of two experts in which the experts estimations differ by extremely admissible according to user's opinion value.

The mentioned definition does not impose restrictions on the number of experts in the estimations set under investigation. As significant can be the estimations subset of any experts number the agreement coefficient of which is not lower of $T_u$. Into the set $V_2$ are included two experts starting from the ease of notion formulation "the admissible difference of estimations" (e.g. as admissible is considered the difference of estimations no more than by $k (k = 1, 2)$ scale graduations).

# 5 The determination of the significant subset of expert estimations and the calculation of generalized expert estimation

The ultimate goal of the expert estimations set $V$ processing is the determination of a generalized expert estimation $a$.

DEFINITION

The significant subset $V_s$ of the expert estimations set $V$ in the $n$-valued order scale is called a subset $V_s \subseteq V$, for which is $k(V_s) \geq T_u$.

The generalized expert estimation $a$ of the expert estimations set $V$ is determined for its significant subset. It is obvious that any set of expert estimations contains some set of significant subsets. This follows from the fact that any subset containing a single expert estimation has as follows from (6,7) an agreement coefficient equal to 1 and therefore is formally significant. However, actually, such set has scarcely any sense to be used, since in this case are ignored opinions of the rest of experts. The mentioned statement only shows that by eliminating from the initial set the estimations of a part of experts estimations it is possible to increase the agreement coefficient. At the same time in order that the agreement improvement process would not contradict "common sense", it is evident that one must,firstly, apply the estimations elimination procedure of a part of estimations only with reference to a set carrying useful information, i.e. having an agreement coefficient which is more that the detection threshold and, secondly, set restrictions on the extremely admissible minimum of experts estimations number in the significant set. Following the same "common sense" this minimum number of experts estimations is expedient to restrict by 3.

Determination of significant subset of expert estimates and the computation of integrated expert estimation is carried out in accordance with next procedure. In the beginning it is necessary to compute the agreement coefficient $k_c(V)$ of $V$ expert estimates set and also meanings of the detection threshold $T_0$ and the application threshold $T_u$. If $k_c(V) < T_0$, (8) that to suggest experts all right competence coefficient increasing to reconsider its estimates. After the change of every estimate to determine $k_c(V)$ to check (8). If $T_0 \leq k_c(V) < T_u$, (9) that to compute arithmetic mean $a$ of estimates for set $V$.

Further it is necessary to propose experts all right competence coefficient increase to reconsider its estimates with the aim of the reduction of their difference from $a$. Availability rather experts with equal to competence in the first place is suggested to reconsider estimate to that expert, at whom estimate greatest differs from $a$. After the change of any estimate is verified the accomplishment of condition (9). If upon certain step was performed the condition $k_c(V) > T_u$ (10) that the set $V$ is significant and $a$ computed for it, is accepted for agreed integrated estimate.

If after suggestions all experts to reconsider its estimates failed attain the execution of condition

(10) that, as from least competent expert, reject their estimates. If we have rather the experts of similar competence than to reject at the out set estimate , the most distinguished from $a$. After rejecting each estimate is verified the accomplishment of condition (9). If it is not carried out that process continues, if the number of remained experts more minimally allowable (usually admit ($n_{\min} = 3$). Otherwise is infer about what brigade of experts can't give co-ordinated estimate and its necessary to replace.

CONCLUSION.

In the paper is substantiated an analytical method for determination the numerical index of agreement degree of estimations set given by different experts. The algorithms for calculation of detection threshold, allowing to determine the presence of useful information in a set of expert estimations, and application threshold describing the user's requirements to the of exactness if resulting estimation determination are suggested. The procedures of obtaining certain generalized expert estimations are developed.

The work is supported by the Fund of Fundamental Investigations at the State Committee on Science and Technologies of Ukraine.

REFERENCES

1.Saaty T.L.Vzaimodistvija v ierarhicheskih sistemah.- " Izv. AN SSSR. Tehnicheskaja kibernetika", 1979,N1, s.68-84 (Russian).

2.David H.A.The method of paired comparisons(University of North Carolina,Charles Griffin. Company Limited,London,1969)

3.Kendall M.Rangovie korreljatsii,-Statistika,1975,214s.(Russian)

4.Hollander M.,Wolfe A.D. Nonparametric statistical methods.N.Y., Wiley,1973 p.500

5.Panneton M.,Robillard P.Algorithm AS Kendall's S frequency distribution.-"Appl.Statist.",1972 348.

6.Best D.J. Extended tables for Kendall's tau. -"Biometrika", 1973, v.60, N2, p.429-430.

7.De Jonge C.,Van Montfort M.A.J. The null distribution of Spearman's S when K=12.-"Statistica Neerlandica", 1972, v.26, N1, p.15-17.

8.Gelberg I.G. Sootnoshenie mezdy statistikoj Manna-Whytny i koefficientom korrelacii Kendalla.- "Teoria veroiatnostej i ee primenenie", 1974, t.XIX, N1, s.211-384 (Russian).

9.Kendall M.,Stuart A.,Statisticheskie vivodu i svijazi.M., Nauka , 1973, s.899 (Russian).

10.Mirkin B.G. Problemu gruppovogo vibora. -M., Nauka.-1974.,s.254, (Russian)

11.Kendall M.G.,Smitth B.B. On the method of paired comparisons.-"Biometrika",1940,v.31,p.324 345.

12.Becker G.M.,De Groort M.M.,MarschakF. Probabilities of choice among very similar objects.- "Behavioral Sci.", 1963, v8, p.306-311.

13.Kendall M.Rank correlation methods.L.,1955.

14.Nikolaev V.I.,Temnov V.N. Ob odnom metode formirovanija ekspertnoj otsenki.-"Izv.AN SSSR.Technicheskaja kibernetika",1973, N 5,s.32-36,(Russian).

15.Litvak B.G. Ekspertnaja informacija.Metodi poluchenija i analiza.M., Radio i svjaz, 1982, p.183 (Russian).