KM 37(1991)
pag 97 - 117

A WEIGHTED LOEVINGER H-COEFFICIENT EXTENDING MOKKEN SCALING TO MULTICATEGORY ITEMS

*Ivo W. Molenaar

ABSTRACT

A Mokken scale originally was a nonparametric item response model for measuring a not directly observable property (latent trait) from a person's answers to a set of dichotomously scored items. The concept was extended to items with more than two ordered answer categories by Molenaar. This extension and its implementation in the computer program MSP so far were based on a count of Guttman errors in the contingency table of each item pair.

The present paper argues that a weighted count of such errors is a more effective way to assess the closeness to the Guttman scalogram. It leads to a version of Loevinger's H-coefficient that is always equal to the ratio of the correlation between the two item scores and the maximum possible correlation given the marginal distributions per item.

After an introduction, the new coefficient is first explained by two examples. Then the new and the old coefficient are compared. Proofs are collected in the final section.

---

*Ivo W. Molenaar, Vakgroep Statistiek & Meettheorie, FPPSW, Rijksuniversiteit Groningen, Grote Kruisstraat 2/1, 9712 TS Groningen, tel. 050-636185

1. INTRODUCTION

Mokken (1971) has proposed a nonparametric item response model for dichotomous items. Statistical models for item responses are used to assess and improve the quality of a measurement procedure that infers a person's position on an unobservable latent trait (usually an attitude or an ability) from his/her answers to a set of items believed to be relevant to that trait. The core idea of item response theory is to express the probability of a positive answer of a person to an item as a function of the person's latent trait value and of the properties of the item, summarized in the socalled item characteristic curve that represents the probability of a positive answer as a function of the latent trait value.

Mokken's model is nonparametric, in the sense that neither the distribution of the latent trait across subjects, nor the item characteristic curves, are assumed to belong to a parametric family (like the normal or the logistic). As a result it only provides an ordering of the persons and of the items, as opposed to numerical estimates of person and item parameters. With the computer program MSP (Debets and Brouwer, 1989) one can test the assumptions of the model and obtain the resulting order for a given data matrix of persons by items. Moreover, if no satisfactory fit is found, a bottom up search procedure finds possible subsets of items that do form a Mokken scale.

A Mokken scale can also be viewed as a probabilistic extension of the Guttman scalogram. The major tool for assessing scalability according to this model is Loevinger's H-coefficient for two items $H = 1 - F/EF$, where F is the number of persons scoring positively on the most difficult and negatively on the easiest of the two items, and EF is the expected value of F under the null hypothesis that the item scores are independently distributed given the observed marginal distributions. Mokken (1971) derives from the asymptotic sampling distribution of H a one sided significance test whether the population H is zero or positive. He works out the search procedure mentioned above, in which H-values per item are only viewed as acceptable when they exceed a lower bound c to be provided by the user (default 0.3) .

As is explained in Molenaar (1982), Sijtsma (1988, p.32-34) and Debets & Brouwer (1989), the model, the scaling procedure and the use of Loevinger's H-coefficient can been extended to items with $M>2$ ordered categories. They will be scored in this paper as $1,2,\ldots,M$, but any other equidistant scoring system produces equivalent results. Indeed, Molenaar (1982) uses $m+1$ for the number of categories and scores them as $0,1,\ldots,m$. The new notation, however, is much easier for our proofs. If k items of this kind form a Mokken scale, they measure the same latent concept, and persons can be ordered by their Likert sum score across the k items as an indicator of their position on the latent trait.

The extension proceeds as follows. The score, denoted by X, obtained on an item, can be written as the sum of the dichotomous scores X_g on M ordered "item steps", defined by $X_g=1$ if $X\geq g$ and $X_g=0$ otherwise, for $g=1,2,..,M$. A person scoring $X=h$, say, has passed the first h item steps and failed the remaining M-h steps. Note that the very first item step $X\geq 1$ is trivially passed by everyone (and is omitted if scores $0,1,\ldots,m$ are used). Suppose, for example, that an ability item has $M=3$ ordered rating categories wrong, partially correct, wholly correct, scored as $X = 1$, 2 and 3 respectively. A competent person will probably earn three points on this item by passing the trivial step $X\geq 1$ but also both nontrivial item steps $X\geq 2$ (from wrong to partially correct) and $X\geq 3$ (from partially correct to wholly correct); thus $X_g=1$ for all three item step scores. A person with average competence will probably pass the first two item steps but fail the last one, and earn two points ($X=2$, $X_1=1$, $X_2=1$, $X_3=0$). An incompetent person will only pass the trivial first step and obtain the minimal score of $X=1$ with $X_1=1$, $X_2=0$, $X_3=0$.

As is worked out in Molenaar (1982), there is a logical relation between the item step scores X_h : if $X_h=1$ then $X_g=1$ for all $g<h$, and if $X_h=0$ then $X_g=0$ for all $g>h$. It would thus be incorrect to analyze the scalability of k items with M categories each by computing the Loevinger H coefficient for all pairs formed from the $k*(M-1)$ nontrivial dichotomous item step scores : the relation implies that step scores from the same item have perfect Guttman scalability.

Therefore an adequate analysis should be based directly on the value of H for each pair of multicategory items, with scores denoted by X and Y. It can be calculated from the M by M contingency table of X and Y in the following way. First order the estimated probabilities of the 2M-2 nontrivial item steps $X \geq g$ and $Y \geq h$ for $g,h = 2,3,\ldots,M$. Each person passing the easiest s of these nontrivial steps plus the two trivial ones and failing all other steps ($0 \leq s \leq 2M-2$) falls in one of the cells of the contingency table that are compatible with the cumulative Guttman idea. Frequencies in such conformal cells are underlined in the examples below and marked by stars in the MSP output. Persons falling in the other cells of the contingency table are in error with respect to the Guttman model.

In all publications up to now, except for a brief remark in Molenaar(1983), the H-coefficient for the scalability of the two items is then defined by $H' = 1 - F'/EF'$, where F' is the sum of the frequencies in the error cells and EF' is its expectation under the null model of global independence of the item scores X and Y with the given marginal distributions. Primes are added to F and H in order to distinguish them from the new version proposed in this paper.

The goal of this paper is to introduce an alternative definition of H, in which F' is replaced by a weighted sum of the frequencies in the error cells, to be denoted by F. As explained below, the weight per cell is the number of item step pairs for which the Guttman order is reversed in the answer to both items corresponding to that cell.

The next section gives two examples of the calculation of F' and F for two items with four categories each. Section 3 presents an evaluation of the differences between the old coefficient $H' = 1- F'/EF'$ and the new coefficient $H = 1 - F/EF$. Some favorable properties of H, announced there, will be proven in the last section.

2. TWO EXAMPLES

It is a relevant property of multicategory Mokken scaling that the order of the item steps and thus the separation into conformal and error cells differs for different item pairs. In the first example below it will be shown that the sequence of conformal cells (underlined) has first two vertical steps, then two horizontal ones, and next a vertical one followed by a horizontal one (VVHHVH); in the second example this order will be VHVHHV. Both examples deal with items scored 1,2,3,4 for the categories very often, often, rarely, never, taken from the GSLDT data set of Weijmar Schultz and Van der Wiel (1991, p.158-167).

	X=1	X=2	X=3	X=4	marg	step
Y=1	<u>3</u>	0	0	0	3	(178)
Y=2	<u>4</u>	7	3	0	14	175
Y=3	<u>10</u>	<u>22</u>	<u>34</u>	3	69	161
Y=4	9	17	<u>40</u>	<u>26</u>	92	92
marg	26	46	77	29	178	
step	(178)	152	106	29		

From the marginal frequencies, one derives the popularities per item step that are displayed in the last column and row. The number of persons passing the trivial first steps $X\geq1$ and $Y\geq1$ is of course 178. More interesting is that the most popular real item step is $Y\geq2$ passed by 175 persons. The next most popular is $Y\geq3$ passed by 161 persons, and then $X\geq2$ passed by 152 persons. Proceeding in this way one obtains that the cells 11, 21, 31, 32,... of the table are in agreement with the Guttman ordering of the item steps; these entries are underlined. In the table below the item steps are ordered from most to least popular (top to bottom). For each cell (column of the table) it is indicated whether the item step is passed (+) or failed (-). In a perfect Guttman pattern (underlined entry in the previous table) no minus sign in the corresponding column may occur above a plus sign.

	cell ij	11	12	13	14	21	22	23	24	31	32	33	34	41	42	43	44
item step	pop.																
Y≥2	175	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+
Y≥3	161	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+
X≥2	152	-	+	+	+	-	+	+	+	-	+	+	+	-	+	+	+
X≥3	106	-	-	+	+	-	-	+	+	-	-	+	+	-	-	+	+
Y≥4	92	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+
X≥4	29	-	-	-	+	-	-	-	+	-	-	-	+	-	-	-	+
error weight	w_{ij}	0	2	4	7	0	1	2	4	0	0	0	1	2	1	0	0
obs.frequency	n_{ij}	$\underline{3}$	0	0	0	$\underline{4}$	7	3	0	$\underline{10}$	$\underline{22}$	$\underline{34}$	3	9	17	$\underline{40}$	$\underline{26}$

The MSP output assigns error weight 1 to each of the error cells 12, 13, 14, 22, 23, 24, 34, 41, 42 and shows an error frequency F = 7+3+3+9+17 = 39. The error weight in the table above, however, is the number of Guttman violations between item steps : per column each combination of a plus sign with a minus sign above it contributes one violation. In cell 14, for example, the plus signs in rows 3 and 4 each have two minus signs above them, and the plus in the last row has three, leading together to a weight of 7.

The weighted sum of observed error frequencies $F = \sum_i \sum_j w_{ij} n_{ij}$ equals 51. If one replaces the observed frequencies by the null-expected frequencies $e_{ij} = n_{i+} n_{+j} / n_{++}$, not shown above but given in the MSP output, one obtains an expected error count of 62.97 in the unweighted case and 96.91 in the weighted case. Thus one obtains H′ = 1 - 39/62.97 = 0.3807 and H = 1 - 51/96.91 = 0.4737.

We now turn to the second example. Here the two least popular item steps happen to have an equal estimated probability of 29/178. In the table, X≥4 has been arbitrarily designed as the easiest one. This leads to F′= 57, EF′= 83.70, H′= 0.3190 for the unweighted and F = 86, EF = 147.80, H = 0.4181 for the weighted case.

	X=1	X=2	X=3	X=4	marg	step
Z=1	12	6	4	2	24	(178)
Z=2	8	15	18	5	46	154
Z=3	4	23	41	11	79	108
Z=4	2	2	14	11	29	29
marg	26	46	77	29	178	
step	(178)	152	106	29		

item step	pop.	cell ij: 11	12	13	14	21	22	23	24	31	32	33	34	41	42	43	44
Z≥2	154	-	-	-	-	+	+	+	+	+	+	+	+	+	+	+	+
X≥2	152	-	+	+	+	-	+	+	+	-	+	+	+	-	+	+	+
Z≥3	108	-	-	-	-	-	-	-	-	+	+	+	+	+	+	+	+
X≥3	106	-	-	+	+	-	-	+	+	-	-	+	+	-	-	+	+
X≥4	29	-	-	-	+	-	-	-	+	-	-	-	+	-	-	-	+
Z≥4	29	-	-	-	-	-	-	-	-	-	-	-	-	+	+	+	+
error weight	w_{ij}	0	1	3	5	0	0	1	2	1	0	0	0	4	2	1	0
obs.frequency	n_{ij}	12	6	4	2	8	15	18	5	4	23	41	11	2	2	14	11
alt.error weight		0	1	3	6	0	0	1	3	1	0	0	1	3	1	0	0

If Z≥4 is chosen as the easier step, the fifth and sixth line of plus and minus signs are interchanged. One easily verifies that this leads to the line of alternative error weights given at the end of the table. Now in the unweighted case one finds F′ = 54, EF′ = 84.03, and H′= 0.3574. For the weighted case one can verify that both F and its null-expected value remain unchanged, and H=0.4181 as before.

The MSP output prints a warning for the two equal cumulative marginals. The program takes the mean of the two F′ values (57+54)/2 = 55.5 and the mean of the two EF′ values $(83.70+84.03)/2 = 83.87$ and thus uses $H' = 1-55.5/83.87 = 0.3382$ for the analysis of the scale. Although this interpolation can be justified by the argument that there is equal probability that each of the $Z\geq4$ and $X\geq4$ steps is the easiest in the population, it is more convincing to use a coefficient that has the same value regardless of the choice. Later in this paper it will be shown that H always has this property.

3. COMPARISON OF H′ AND H

Note that in the dichotomous case, H′ and H coincide : there is only one error cell, and it gets weight one in both coefficients. In the case of more than two categories, the new coefficient H has the following advantages compared to the original coefficient H′:

1) In the case of equal item step probabilities, there is ambiguity which of two cells is an error cell. H has the same value for both choices, whereas H′ has not (the computer program MSP then gives both possible values and operates with a third value based on the means of F′ and EF′).
2) The interpretation as the observed correlation between the item scores divided by the maximum possible correlation given the marginals, holds for H′ in the dichotomous case only (H′= phi/phimax). For H this interpretation is valid regardless of the number of categories.
3) The null-expected frequency of errors EF equals the product of the sample size with the maximal possible covariance given the marginals, which makes it easier to obtain than EF′.
4) This maximal covariance is the covariance for the table with the given marginals in which all error cells have zero frequency.

The proofs of these properties are given in section 5.

Note that F weighs deviations from the perfect Guttman pattern according to their severity, whereas F′ does not. In other words, passing a certain item step and failing another one which is much easier (in the sense of : there exist many other item steps in between) is punished more severely than passing the same step and failing the adjacent easier one. There is an analogy with the relation between kappa and weighted kappa, or between the Wilcoxon symmetry test and the sign test. The former weighs by the "number of observations in between" and the latter just uses the sign of the difference.

The extension of H for item pairs to a scalability coefficient per item or for the total scale proceeds exactly like for the dichotomous model: errors and expected errors are added across the item pairs containing the fixed item, and across all item pairs, respectively. In both cases the weighted coefficient counts the weighted number of inversions per person of the relevant item steps, whereas the unweighted one counts the number of relevant item pairs per person in which at least one step pair is reversed.

A possible drawback is that H, more than H', has an instable value when it is estimated from a small sample: having one person more or one person less in a heavily weighted cell makes more difference for calculating H than for H'. A similar instability occurs in the dichotomous case between a very popular and a very impopular item, because one person more or less in their error cell makes much difference when EF' is close to zero. This effect of small expected frequencies is discussed in detail in Molenaar (1982, section 3).

A pilot version of MSP using the new H rather than H' has been developed for the author by Debets. Running both versions on a dozen available datasets with three to seven ordered answer categories per item has led to the following tentative conclusions.

Pairwise H'-values are often close to pairwise H-values, but occasional absolute differences of 0.10 and even 0.30 occur, mostly with H larger than H'. The reverse, such as H'=0.35 and H =0.21, is more exceptional. It seems to occur for tables in which a relatively high frequency is found in a very unusual cell with many Guttman violations as counted by item steps in the wrong order. Since such a contingency table tends to produce a negative H' value even in the unweighted case, there is a general trend that pairwise H values lie further from zero than their unweighted counterparts. Exceptions to this trend do occur, however, notably for small samples of persons.

Item $H(i)$ values are much closer to item $H'(i)$ values than the pairwise coefficients; in general the new coefficients tend to be a bit higher, but occasionally they are lower than the old ones.

For the total scale H is almost always somewhat higher than H'. The median difference obtained lies between .05 and .07.

The differences between the weighted and unweighted cases tend to be more pronounced when the number of categories is larger.

It is not unusual to find some item pairs for which the old and the new pairwise coefficients have opposite signs, or in which an item coefficient $H'(i)$ falls just on the other side of the lower bound c for scalability than $H(i)$. Both phenomena may cause exclusion of an item in a search procedure with one coefficient while it will be included with the other.

4. DISCUSSION

The new coefficient H based on the weighted error count is superior in many respects to the older version based on the unweighted counts. The natural interpretation as the ratio of the correlation to the maximum possible correlation given the marginals is often easier than viewing H as the complement of the frequency ratio of the observed errors to the null-expected errors. In particular, taking the H value obtained from a sample of persons as an estimate of the corresponding value in the larger population of persons, is more natural in the former interpretation. Note that Mokken (1971) discusses the population H in both interpretations.

Attaching additional weight to errors which are more sharply in contrast with a score pattern expected under the Guttman scalogram assumption appears desirable in most applications of the Mokken model to multicategory items. The present author does not exclude, however, that the use of the unweighted version H' might be desirable in some cases, especially for a test consisting of a few items each with many categories.

Multicategory Mokken scaling may also be applied when the number of ordered categories varies between items. The reader can easily verify that the results in this paper remain valid for this case; the modifications are trivial but the notation becomes more cumbersome. This extension can already be handled in the present version of MSP which is based on the unweighted H'. It will also be included in the new version using the weighted H. If the same category set can be meaningfully used for all items, one is advised to do so : note that for example in the sum score for three two-category items and three four-category items the former set of items allows earning less points, although these items need not be less important for determining a person's position on the latent trait being measured. When some items do not allow a natural meaningful set of categories that corresponds to the categories used for the other items, it should be left to the investigator whether reduction to an equal number of categories per item is preferred to using unequal numbers of categories.

Mokken scaling has been successfully applied, both with dichotomous and with multicategory items, in many domains; for an overview see Sijtsma (1988, p.31). Incorporating the change proposed in the present paper hopefully will lead to the development of even more good Mokken scales in the future.

5. PROOFS

This section contains the proofs of the four properties mentioned in section 3. It will be convenient to prove them in reversed order, from 4) via some auxiliary results to 3) and finally 2) and 1). For the direct algebraic verification of the four properties, it is an obstacle that different cells play the role of error cells in different contingency tables, depending on the numerical values of the cumulative marginal proportions. In the proofs that follow below, this obstacle has been avoided by the use of general principles that hold for any contingency table, regardless of which cells are error cells.

The problem of finding a joint distribution of two variables, with given marginal distributions, that maximizes their correlation or covariance, has a long history. The treatment by Whitt (1976), for example, refers to Fréchet (1951); see also Naddeo (1987, p.98). Although some of our theorems can be derived from these earlier results, the present note presents a complete set of proofs, in a notation consistent with our problem of H coefficients.

In each contingency table of two M-category items, there exists a sequence of 2M-1 conformal cells, in which the s easiest item steps for $2 \le s \le 2M$ have been passed and the 2M-s more difficult ones have been failed. The first such cell (s=2) is the 11 cell, in which only the two trivial item steps have been passed; the last cell (s=2M) is the MM cell. For each conformal cell, the next conformal cell lies either immediately to its right or immediately below it. If one joins consecutive conformal cells by line segments, cell 11 is joined to cell MM by segments that move downward or to the right, never upward or to the left. All cells outside this trajectory are error cells.

It is very important for the proofs that we only consider contingency tables with the same marginals as the observed one; they form the socalled isomarginal family. In the proof of theorem 1 this allows us to fill the remainder of certain rows or columns with zeroes; the goal of finding the table Z with zero entries in all error cells is achieved stepwise by putting in each cell the maximum frequency permitted given the assignment to previous cells and the constraint of equal marginals.

In proving theorems 2 and 5 the property is used that different members of the isomarginal family can always be obtained from one another by a sequence of elementary shift operations of the following type. If the new table has a frequency of one unit less in a certain cell than the previous table, it has one unit more in some cell of the same row (fixed row sum). Moreover, it has another row in which the column that was just decreased by one must be increased by one, and in which the column that was just increased by one must be decreased by one, in order to keep the marginal values of these two columns fixed.

<u>Theorem 1</u>. For each M by M contingency table of two items X and Y with entries n_{ij}, there exists a unique contingency table with entries z_{ij} which has the same marginals and zero entries in all error cells.

Proof

A numerical illustration of the procedure follows after the proof. Denote by n_{ij} the observed frequency of X=i, Y=j and by n the grand total. Let n_{i+} and n_{+j} be the individual marginal frequencies of X=i and Y=j, respectively. Finally let N_{i+} and N_{+j} be the item step freqencies observed for X≥i and Y≥j respectively; they were denoted by "step" in the examples of section 2. Note that they differ from ordinary cumulative frequencies by counting "i or more" rather than "i or less", and similarly for j.

The first conformal item cell is cell 1,1.

If g,h is the last cell established to be conformal, then the next conformal cell lies either to its right or below it. The choice is trivial when there is no cell to the right (h=M) or no cell below (g=M). Otherwise the choice depends on which of the corresponding step frequencies is largest. The following rule establishes (a) the frequency assigned in the Z-table to the currently considered cell and to the remaining cells in the same row/column; (b) the identity of the next conformal cell :

When either h=M or $N_{g+1,+} \geq N_{+,h+1}$, then g+1,h is the next conformal cell and one puts

$$z_{gh} = n_{g+} - \sum_{j=1}^{h-1} z_{gj} \qquad \text{and if } h<M \text{ also} \qquad z_{gj} = 0 \text{ for all } j>h.$$

When either g=M or $N_{g+1,+} < N_{+,h+1}$ holds, then g,h+1 is the next conformal cell and one puts

$$z_{gh} = n_{+h} - \sum_{i=1}^{g-1} z_{ih} \qquad \text{and if } g<M \text{ also} \qquad z_{ih} = 0 \text{ for all } i>g.$$

If g=M and h=M, one only applies any of the two formulas for z_{gh} for g=h=M; they both assign all the remaining frequency to the last cell.

This algorithm uniquely determines the conformal cells of the given table. It assigns zero frequency to each error cell, because each such cell either has all conformal cells of the same row to its left (in which case it has zero z-frequency by the first rule) or it has all conformal cells of the same column above it (in which case the second rule assigns zero z-frequency to it). Moreover, the rules clearly ensure that the marginals are preserved; once this holds for the first M-1 rows and columns it must also hold for the last one.

Let us illustrate now in some detail how the perfect Guttman table Z with entries z_{ij} can be obtained from a given table N with entries n_{ij} by a sequence of elementary shift operations in which four entries forming a rectangle are increased and decreased by one unit. In this algorithm, used in the proofs of theorems 2 and 5, the table currently under consideration forms a frequency matrix T with entries t_{ij}. Next a different matrix D is defined, with entries $d_{ij} = t_{ij} - z_{ij}$. Because T and Z must have the same marginals, the matrix D is doubly centered (row and column sums are always zero).

ALGORITHM. For any given matrix N, create the matrix Z as described in Theorem 1. Put initially T = N.

(1) Inspect matrix D = T - Z. If D is the null matrix we are ready. If not,

g = number of the first row in which D has a nonzero entry;

p = number of the first column in which D has a positive entry in row g;

q = number of the first column in which D has a negative entry in row g;

h = number of the first row for which D has a positive entry in column q.

Clearly g exists because D is non-null, and p,q,h exist because D is doubly centered. By definition h>g. In the proof of theorem 2 it will be shown that also p>q.

Now use the elementary shift, adding one to t_{gq} and t_{hp}, while diminishing t_{gp} and t_{hq} by one. As D = T-Z, the same changes occur in D. Repeat this shift until one of the two diminished d-values equals zero. Then go back to (1) because now one or more of the four values g,h,p,q have been changed.

This completes the algorithm. For later proofs, we define the raw cross product sum

$$CP = t_{++} * E\,XY = \sum_i \sum_j i*j*t_{ij}. \qquad (1)$$

Each shift increases CP by an amount CPC = gq-gp-hq+hp = (h-g)(p-q).
The algorithm is demonstrated for the first example of section 2. Conformal cells are underlined. For later use, we also give the weights w_{ij} and the weighted number of Guttman errors F (see theorem 5).

	table N				marg	step	table Z				weights W			
	3	0	0	0	3	(178)	3	0	0	0	0	2	4	7
	4	7	3	0	14	175	14	0	0	0	0	1	2	4
	10	22	34	3	69	161	9	46	14	0	0	0	0	1
	9	17	40	26	92	92	0	0	63	29	2	1	0	0
marg	26	46	77	29	178									
step	(178)152	106	29											

Comment : Z is obtained from N by the steps in the proof of Theorem 1.
The "step" entries N_{i+} and N_{+j} determine the next conformal cell.
Now we determine g,q,p,h for each step of the algorithm. The four cells of matrix D involved in the shift operation are printed in bold.

table T				table D=T-Z				situation
3	0	0	0	0	0	0	0	CP = 1629 F = 51
4	7	3	0	**-10**	**7**	3	0	g=2 q=1 p=2 h=3 CPC=1 applied once
10	22	34	3	**1**	**-24**	20	3	
9	17	40	26	9	17	-23	-3	
3	0	0	0	0	0	0	0	CP = 1630 F = 50
5	6	3	0	**-9**	**6**	3	0	g=2 q=1 p=2 h=4 CPC=2 applied 6 times
9	23	34	3	0	-23	20	3	
9	17	40	26	**9**	**17**	-23	-3	
3	0	0	0	0	0	0	0	CP = 1642 F = 38
11	0	3	0	**-3**	0	**3**	0	g=2 q=1 p=3 h=4 CPC=4 applied 3 times
9	23	34	3	0	-23	20	3	
3	23	40	26	**3**	23	**-23**	-3	
3	0	0	0	0	0	0	0	CP = 1654 F = 26
14	0	0	0	0	0	0	0	g=3 q=2 p=3 h=4 CPC=1 applied 20 times
9	23	34	3	0	**-23**	**20**	3	
0	23	43	26	0	**23**	**-20**	-3	
3	0	0	0	0	0	0	0	CP = 1674 F = 6
14	0	0	0	0	0	0	0	g=3 q=2 p=4 h=4 CPC=2 applied 3 times
9	43	14	3	0	**-3**	0	**3**	
0	3	63	26	0	**3**	0	**-3**	
3	0	0	0	0	0	0	0	CP = 1680 F = 0
14	0	0	0	0	0	0	0	D=0 ready
9	46	14	0	0	0	0	0	
0	0	63	29	0	0	0	0	

<u>Theorem 2</u>. The covariance and the correlation of two item scores with fixed marginals attain their maximum when all error cells have frequency zero.

<u>Proof</u>

Denote the two item scores by X and Y. Note that the marginal distributions of X and Y are fixed by assumption. Thus their means and variances are fixed as well. Maximizing the correlation between X and Y is thus equivalent to maximizing their cross product sum CP given by (1) across all sets of cell frequencies t_{ij} which give the correct marginals.

Let T denote an arbitrary member of the isomarginal family determined by the marginals of X and Y, and let Z denote the table from this family in which all error cells have frequency zero. Unless already T=Z, we transform T into Z using our algorithm. At each shift operation, CP changes by $(h-g)(p-q)$. If this is always positive, then Z must have a larger CP than T and our proof is complete.

The property $h>g$ is obvious from their definitions. The property $p>q$, on the other hand, does not hold for any pair of matrices, but here it can be derived from the procedure by which Z has been constructed. In order to show this, consider any fixed stage of the algorithm in which the current matrix is denoted by T and the difference T-Z by D. Again g denotes the first row of D that has a non-zero entry; on this row let j denote the column in which the first non-zero entry occurs. Our claim is that $d_{gj} < 0$.

The reader may verify from the proof of theorem 1 that the rules for creating Z imply that any element of Z, in particular our current gj-cell, always obtains the maximum possible frequency given the frequenties assigned to all cells to its left and to all cells on higher rows. For all such cells the current d-value is already zero. Therefore T and Z both belong to the subfamily of the isomarginal family with the same fixed values in all earlier cells. Given the constraints that this implies, z_{gj} assumes the maximum possible value and t_{gj} is free. Thus $d_{gj} = t_{gj} - z_{gj} \leq 0$, but it is unequal to zero by the definition of j and thus it must be negative. As a consequence, the first negative element of row g always stands to the left of the first positive element, by which $q<p$ is proven.

Theorem 3. Consider two cells $i,j-1$ and i,j of row i of the contingency table of two M-category items X and Y. Let w_{ij} denote the number of pairs of item steps for which one step is passed and an easier step is failed by a person with X=i and Y=j.

(a) There exists a unique integer g $(1-i \leq g \leq M-i)$ such that in row $i+g$ the cells in columns $j-1$ and j are both conformal.

(b) $w_{ij}-w_{i,j-1}= g$.

(c) The difference in w-value decreases by 1 for each subsequent row in the table :

$$w_{i+1,j} - w_{i+1,j-1} = w_{i,j} - w_{i,j-1} -1.$$

Proof

a) is correct because the sequence of line segments joining cell 11 to cell MM passes exactly once from column $j-1$ to column j.

For the case $g = 0$, $w_{ij}= w_{i-1,j}= 0$ as all conformal cells have no errors, and b) follows directly.

If $g > 0$, the g steps down from row i to row $i+g$ have been failed in cell $i,j-1$. Going from there to cell ij, the step $Y \geq j$ which was failed in cell $i,j-1$ is now passed, and nothing changes for any other item step. Thus going from $i,j-1$ to ij increases w by an amount g.

The $g < 0$ case follows by an analogous reasoning, so b) holds for all three situations.

Statement c) follows immediately from b), because going from row i to row $i+1$ the unique number g decreases by one.

Note that theorem 3 allows us to calculate the weight matrix without inspecting columns of plus and minus signs like in section 2 : first establish which cells are conformal and give them weight zero, next fill one row using b) and finally apply c) to find the other rows. In the example used before, we have

$$N = \begin{matrix} \underline{3} & 0 & 0 & 0 \\ \underline{4} & 7 & 3 & 0 \\ \underline{10} & \underline{22} & \underline{34} & 3 \\ 9 & 17 & \underline{40} & \underline{26} \end{matrix} \quad \text{so we already know that } W = \begin{matrix} 0 & . & . & . \\ 0 & . & . & . \\ 0 & 0 & 0 & . \\ . & . & 0 & 0 \end{matrix}$$

By result b) $w_{34} - w_{33} = 1$, because we have to go one row down to find conformal cells in the third and fourth column.

So our third row is 0 0 0 1

with differences of 0 0 1, respectively. By c) the fourth row has differences of -1 -1 0 so $w_{42} = 1$ and $w_{41} = 2$. The second row has differences of 1 1 2 so $w_{22} = 1$, $w_{23} = 2$, $w_{24} = 4$. The first row has differences of 2 2 3 so $w_{12} = 2$, $w_{13} = 4$, $w_{14} = 7$.

Theorem 4. For any four cells situated in rows i and i+g and columns j and j+h of a contingency table, the second order difference of w and the analogous difference of the row and column numbers used in the cross product i*j matrix are equal to -g*h and g*h respectively :

$w_{ij} - w_{i,j+h} - w_{i+g,j} + w_{i+g,j+h} = -g*h$ and

$i*j - i(j+h) - (i+g)j + (i+g)(j+h) = g*h.$

Proof

The second order difference considered is the sum of such differences for adjacent columns. For the w matrix the desired result then follows from repeated application of statement b) in the preceding theorem. For the cross product matrix it is very simple algebra.

Theorem 5. The weighted number of errors F in the observed table is equal to the difference between the maximum possible cross product sum CP_{max} and the observed sum CP, and thus to the sample size n times the difference between the maximum possible covariance and the observed covariance.

Proof

In order to establish the relation F = CPmax - CP, we shall change the observed table N into the table Z defined in Theorem 1. This will be done by our algorithm, which consists of a suitable sequence of elementary shift operations that preserve the marginals. For each such operation (which may take place between non-adjacent rows and columns) we have already seen that CP is increased by an amount (h-g)*(p-q).

Adding one to t_{gq} and t_{hp} while subtracting one from t_{gp} and t_{hq} implies that $F = \sum_{i,j} t_{ij} * w_{ij}$ changes by an amount $w_{gq} - w_{gp} - w_{hq} + w_{hp}$. By theorem 4, this means that F decreases by (h-g)*(p-q).

This implies that the total change in F during the algorithm is minus the total change in CP, as is also illustrated in the example after theorem 1. Now CP_{max} is the cross product sum for matrix Z (theorem 2), for which all error cells are zero and thus $F_Z = 0$. Therefore

$F = -(F_Z - F)$ = minus total F-change = total CP-change = CP_{max} - CP.

Because both tables have the same marginals, they have the same means, and the statement on the difference between the covariances follows from that on the difference between the CP values, using the definition of a covariance.

Theorem 6. The expected value of the weighted error count for an item pair X,Y under the null hypothesis of independence and given marginals equals the maximal covariance given the marginals multiplied by the sample size, or in formula

$$E\ F = \Sigma\ \Sigma\ e_{ij} w_{ij} = n\ cov_{max}(X,Y).$$

where $e_{ij} = n_{i+} n_{+j} / n$.

Proof

It is easy to modify the proof of theorem 5 and the algorithm for transforming N into Z for the situation where the given frequencies n_{ij} are replaced by the null-expected frequencies e_{ij}, which have the same marginals as N and Z. Now at each step a suitable fractional number can be added to the gq and hp cells and subtracted from the gp and hq cells, rather than an integer number. If we use the index e to denote values obtained from e_{ij}, this new version of theorem 5 yields that $E\ F = CP_{max} - CP_e$. Next one obtains CP_e from $0 = cov_e(X,Y) = CP_e/n - EX\ EY$. As a result one has

$$E\ F = CP_{max} - n\ EX\ EY = n\ cov_{max}(X,Y).$$

<u>Theorem 7</u>. The weighted coefficient H equals both $cov(X,Y)/cov_{max}(X,Y)$ and $r(X,Y)/r_{max}(X,Y)$.

<u>Proof</u>

The identities $F = n[cov_{max}(X,Y)-cov(X,Y)]$ and $EF = n*cov_{max}(X,Y)$ can be obtained from theorems 5 and 6. Substitution into $H = 1 - F/EF$ gives the first result.

The ratio of the correlations equals that of the covariances, because the standard deviations are fixed within the isomarginal family.

<u>Theorem 8</u>. If an observed frequency table has two item steps $X\geq g$ and $Y\geq h$ with equal observed frequencies, the values of F, EF and H are not changed by choosing one or the other of the two equally popular item steps as the easiest one.

<u>Proof</u>

The three quantities have been interpreted in terms of covariances, and any covariance is a symmetric function of its arguments.

Acknowledgement

The author is grateful to Debets for programming the weighted H, to Mokken, Sijtsma and an unknown reviewer for comments on an earlier version, and to Verhelst for a simplification of the proof of theorem 2.

References

Debets, P. and Brouwer, E. (1989). User's Manual MSP. Groningen : iec ProGAMMA.

Frechet, M. (1951). Sur les tableaux de correlation dont les marges sont donnees. Annales de l'universite de Lyon Sect.A 14, 53-77.

Mokken, R.J. (1971). A theory and procedure of scale analysis. New York, Berlin : De Gruyter (Mouton).

Molenaar, I.W. (1982). Mokken scaling revisited. Kwantitatieve Methoden 3, No 8, 145-164.

Molenaar, I.W. (1983). Measuring with Mokken scales, statistical aspects (in Dutch). Bulletin Vereniging voor Statistiek 16 nr 4, 16-18.

Naddeo, A.(ed.). Italian contributions to the methodology of statistics. Padova : Cleup, Societa Italiana di Statistica.

Sijtsma, K. (1988) Contributions to Mokken's nonparametric item response theory. Ph. D. Thesis R.U. Groningen. Amsterdam: Free University Press.

Weijmar Schultz, W.C.M., and Van der Wiel, H.B.M. (1991) Sexual functioning after gynaecological cancer treatment. Ph.D. thesis R.U.Groningen.

Whitt, W. (1976). Bivariate distributions with given marginals. Annals of Statistics 4, 1280-1289.

Ontvangen: 19-12-1990
Geaccepteerd: 28-03-1991