

Ostermann, Rüdiger

Working Paper

Deviations from an Unidimensional Scale in the Unfolding Model

Volkswirtschaftliche Diskussionsbeiträge, No. 12-90

Provided in Cooperation with:

Fakultät III: Wirtschaftswissenschaften, Wirtschaftsinformatik und Wirtschaftsrecht, Universität Siegen

Suggested Citation: Ostermann, Rüdiger (1990) : Deviations from an Unidimensional Scale in the Unfolding Model, Volkswirtschaftliche Diskussionsbeiträge, No. 12-90, Universität Siegen, Fakultät III, Wirtschaftswissenschaften, Wirtschaftsinformatik und Wirtschaftsrecht, Siegen

This Version is available at:

<https://hdl.handle.net/10419/118710>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

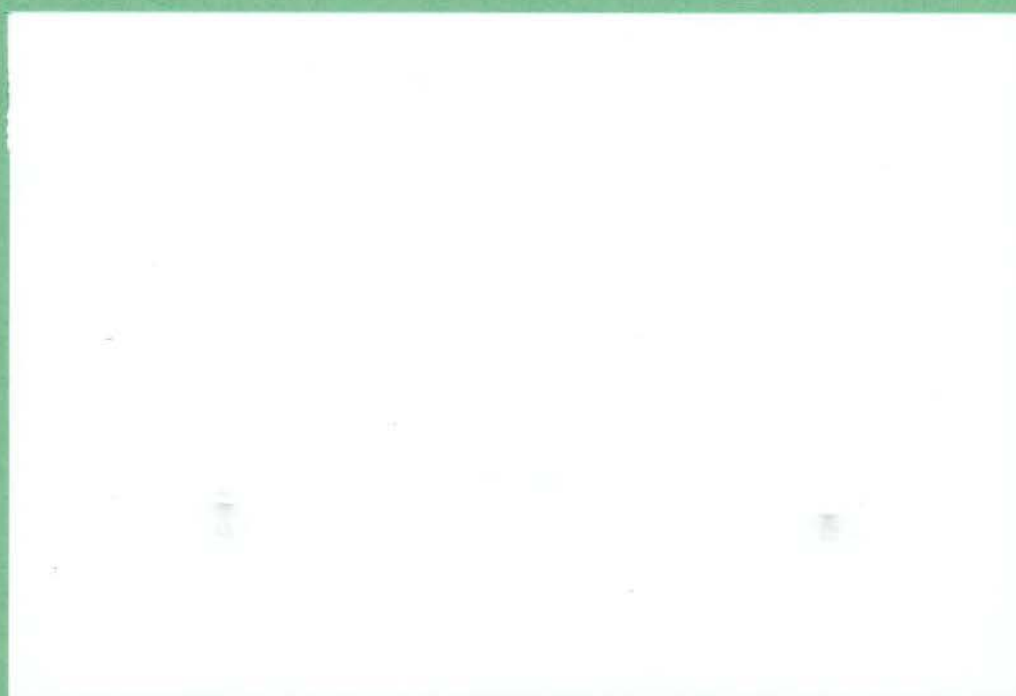
Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

**VOLKSWIRTSCHAFTLICHE
DISKUSSIONSBEITRÄGE**



**UNIVERSITÄT - GESAMTHOCHSCHULE - SIEGEN
FACHBEREICH WIRTSCHAFTSWISSENSCHAFTEN**

Deviations from
an Unidimensional Scale
in the Unfolding Model

by
Rüdiger Ostermann
University of Siegen

Diskussionsbeitrag Nr. 12-90

Deviations from an Unidimensional Scale in the Unfolding Model

Rüdiger Ostermann
University of Siegen
Computer Center
Hölderlinstr.3
D - 5900 Siegen

August 13, 1990

The unidimensional unfolding model proposed by Coombs (1964) is a tool for scaling preference data. Here we consider the special case where the sets of subjects and objects are identical. We get an unidimensional solution of our problem only if the data satisfy a well-defined structure. Therefore we propose a statistical error model including an algorithm which allows to construct an unidimensional solution for data with small differences from this well-defined structure, too. A new coefficient $\kappa_{d,n}$ (n is the number of subjects and objects, respectively) derived from well-known distance measures on permutations (Kendall's τ , Spearman's ρ and Spearman's *footrule* f) is used to describe this difference. The exact distribution of this new coefficient is given for small n (up to 9) as well as some asymptotical results ($n \rightarrow \infty$). Based on these distributions we are able to decide whether we use our new algorithm for a given dataset or not. We demonstrate our new approach by a well-known example of Coombs (1964).

Key words: unidimensional unfolding, preference data, permutations Kendall's τ , Spearman's ρ and *footrule*, dependent variables

1 Introduction

In this paper we consider a special case of the unidimensional unfolding model which has been proposed by Coombs in the early fifties, cf. Coombs (1964). In this special case the set of objects is identical to the set of subjects and Coombs has called it *nonmetric multidimensional psychological scaling*. We can describe our considered unidimensional unfolding problem in the following formalistic way:

Consider the situation where a set of n subjects O_1, \dots, O_n is given. Each of these subjects ranks the remaining $n - 1$ subjects in decreasing order of preference. Each subject produces a rank order $\pi_i = (\pi_i(1), \dots, \pi_i(n))$, where $\pi_i(j)$ is the rank of subject j in the judgement of subject i . Without loss of generality we assume $\pi_i(i) = 1$.

In the area of the unfolding model we usually work with I-scales (individual scales). The I-scale I_i of O_i is an individual preference ordering of the n objects with respect to itself. O_i is at the top of the I-scale I_i . The following formalistic relation combines these both approaches: If $\pi_i(j) = l$ the object O_j has the l -th position on the I-scale I_i .

The aim of unidimensional unfolding is to construct an unidimensional, metric scale (J-scale, Joint-scale) where the rank order of the absolute distances of one object to every other corresponds to I_i ($i=1, \dots, n$). We want to demonstrate the unidimensional unfolding problem by an example (cf. Coombs (1964, p.454-456)):

Example 1:

457 college students of $n=6$ distinct religions were asked to make out an I-scale of the five other religions with respect to their own one. The mean I-scale of all members of each group yields the I-scale of a religion. These six I-scales are given below.

Table 1: Data of Coombs (1964)

| Religion | Number | Symbol | I-Scale |
|--------------|--------|--------|-------------|
| Catholic | 120 | A | A B C D E F |
| Episcopalian | 38 | B | B A C D E F |
| Lutheran | 57 | C | C D E B F A |
| Presbyterian | 100 | D | D E F C B A |
| Methodist | 116 | E | E D F C B A |
| Baptist | 26 | F | F E D C B A |

An example of a J-scale is given below. For all six religions the unfolding problem is solved in one dimension. The numerical representation of this J-scale is not unique, but the order of the objects on this metric scale is it. This J-scale yields e.g. for Lutheran (C) the following rank order which conforms with I_C : C D E B F A .

Fig.1: (A) J-Scale to the Data of Tab.1

| A | B | C | D | E | F |
|-----|-----|-----|------|------|------|
| 0.0 | 2.3 | 9.1 | 13.6 | 14.7 | 17.0 |

$$|D - C| = 4.5$$

$$|E - C| = 5.6$$

$$|B - C| = 6.8$$

$$|F - C| = 7.9$$

$$|A - C| = 9.1$$

Coombs (1964) has interpreted this J-scale as a *degree of prescribed rituals*. Elementary restrictions are given by the unidimensional unfolding problem:

- All I-scales have to end up with one of two objects.
- The I-scale of one of these two endobjects has to be exact reverse of that for the other object.

Each I-scale has only an ordinal structure. In our further analysis we do not assume that the I-scales are derived from an underlying quantitative scale. Nevertheless we want to regard a distance in an usual conception [$d(O_i, O_i)=0$, $d(O_i, O_j)=d(O_j, O_i)$ and $d(O_i, O_j) \leq d(O_i, O_k) + d(O_k, O_j)$]. In the example above the distance between A and B ($d(A, B)$) on the A-I-scale is not greater than the distance between A and C ($d(A, C)$). We use the qualitative nature of the I-scales to construct a qualitative J-scale, i.e. an order of the given subjects from the left to the right, or vice versa.

Then we can derive from this qualitative J-scale a quantitative one using

- the triangular analysis which yields a (partial) order of all pairwise distances and
- the Goode-Phillips- or the Chernikowa-algorithm to construct the desired metric scale. By means of one of these algorithms we solve a system of homogeneous linear inequalities which represent the (partial) order above.

For more details see e.g. Coombs (1964), Phillips (1971), Lehner & Noma (1980) or Hartung & Elpelt (1984). A discussion about the uniqueness of the metric representation can be found in these papers, too.

In this paper we consider the problem to find a qualitative J-scale which represents our given I-scales. Therefore we

- introduce a statistical model for this unfolding problem,
- propose an algorithm to find the 'best' qualitative J-scale,
- introduce three new coefficients which measure the deviation of the data from the well-defined unidimensional model structure,
- give the exact distribution of these coefficients for small n (up to 9) as well as some asymptotical results ($n \rightarrow \infty$).

2 A Statistical Error Model and Two Error Definitions

We understand the unfolding model as a deterministic one and so we do not assume that the I-scales are derived from an underlying quantitative scale. Therefore it is necessary to build an appropriate formalistic model. We can adapt many suggestions from the theory of the Linear Model but we have to modify these suggestions with regard to our special problem.

In the following discussion we assume that n I-scales are given. These n I-scales yield a datamatrix I . Starting from the datamatrix I we want to construct a J-scale. The use of the triangular analysis together with one of the algorithms (Goode-Phillips or Chernikowa) mentioned above is only possible if the given data are errorless. Therefore it is necessary to define (at least) two kinds of errors which may occur.

Def.1:

An I-scale I_i of O_i is incompatible with the (unknown) J-scale iff it is not compatible in the sense of the following definition. An I-scale I_i of O_i is compatible with the (unknown) J-scale iff the order of the objects on the J-scale between O_i and each of the endobjects is identical with the order of these objects on I_i . (see example 2).

Def.2:

An intransitive triple is given iff from (at least) three I-scales of the objects A, B and C the following inequality can be derived:

$$d(A, B) \leq d(B, C) \leq d(A, C) \leq d(A, B)$$

(see example 3).

These two kinds of errors are illustrated in the following examples.

Example 2:

Given the following four I-scales:

| | | | |
|---|---|---|---|
| A | B | C | D |
| B | D | C | A |
| C | D | B | A |
| D | C | B | A |

It is obvious that the unknown J-scale is ABCD or DCBA because the A-I-scale is the reverse one of the D-I-scale and the two other I-scales end up with the object A. But I_B is not compatible with these J-scales because on I_B D is followed by C.

Example 3:

An intransitive tripel is given by

| | | | |
|----------|----------|----------|----------|
| <i>A</i> | <i>B</i> | <i>C</i> | <i>D</i> |
| <i>B</i> | <i>C</i> | <i>D</i> | <i>A</i> |
| <i>C</i> | <i>B</i> | <i>A</i> | <i>D</i> |
| <i>D</i> | <i>C</i> | <i>B</i> | <i>A</i> |

By means of symmetry the following inequalities can be derived from these I-scales:

$$\begin{aligned}
 I_A : d(A, B) &\leq d(A, C) \\
 I_C : d(A, C) &\leq d(C, D) \\
 I_D : d(C, D) &\leq d(B, D)
 \end{aligned}$$

but

$$I_B : d(B, D) \leq d(A, B).$$

All I-scales from example 3 are compatible to ABCD and the datamatrix from example 2 contains no intransitive tripel. It is necessary to have only compatible I-scales to get an unidimensional J-scale. In the case of at least one incompatible I-scale one has to extent the unidimensional J-scale to a multidimensional one. Further it is important to have no intransitive triple to get a quantitative unidimensional J-scale. The sole existence of an intransitive triple (where all I-scales compatible) does not prevent the existence of a qualitative unidimensional J-scale. So we can derive Lemma 1 from Def.1 and 2 (without proof).

Lemma 1:

- a. $\forall i = 1, \dots, n : \# (\text{compatible I-scales}) = \binom{n-1}{i-1}$
- b. \exists altogether 2^{n-1} compatible I-scales.
- c. \exists altogether $\frac{n!}{2}$ possible J-scales for n objects.

To solve the unidimensional unfolding problem by means of triangular analysis together with the Goode-Phillips- or Chernikowa-algorithm it is necessary

that no intransitive triple and no incompatible I-scale exist, but we want to scale (slightly) non-errorless data, too. Therefore we allow that some subjects make a small mistake at the construction of their I-scales. If the 'sum of such mistakes' for the whole set of subjects is not too big we suppose that the true J-scale is more likely an unidimensional than a multidimensional one. Now our problem is

- to find the 'best' J-scale and
- to determine whether the deviations from an unidimensional scale are too great.

Without loss of generality we can describe the objects on the J-scale by the order $1, \dots, N$. So we propose the following statistical error model.

Statistical error model:

$$I_i = \pi_i \circ \tilde{I}_i \quad i = 1, \dots, n$$

with

\tilde{I}_i the true I-scale (derived from the true, but unknown J-scale),

I_i the observed I-scale,

π_i permutation which transmits \tilde{I}_i into I_i ,

$P(I_i = \tilde{I}_i) = (n - 1)^{-1}$.

This model is very similar to the model by van Blokland-Vogelzand (1985) but she has studied the case where the set of subjects and the set of objects are non-identical. Therefore she has no problem with a dependence-structure between the different I-scales. We consider this dependence-structure later in detail.

The best (qualitative) J-scale is defined as that J-scale for which the observed I-scales \tilde{I}_i have minimal distance to the true I-scales $I_i (i = 1, \dots, n)$. For each J-scale and for each subject O_i we have a set $v_i = \{v_{i,1}, \dots, v_{i,l_i}\}$, $l_i = \binom{n-1}{i-1}$, of compatible I-scales and we choose that $v_{i,k}$ which solves

$$d(I_i, v_{i,k}) = \min_k .$$

d should be an appropriate distance measure on these I-scales or on the permutations of the numbers $1, \dots, n$, respectively. There exist several proposals to measure these distances. We consider the following three:

- Kendall's τ ,
- Spearman's ρ ,
- Spearman's footrule f .

It is possible to use or to consider also other distance measures like the Hamming, Ulam or Cayley distance. A brief survey about these distance measures and their properties can be found in Chritchlow (1985) or Diaconis(1988). If we understand an I-scale as a special presentation of a permutation of the numbers $1, \dots, n$ it is much easier to analyze our minimizing problem using the permutation notation mentioned in the introduction. If we consider two I-scales and denote them by π and σ we can write the three different distance measures as

$$\begin{aligned} \tau &= \# \{ \text{sign}\{\pi(r) - \pi(s)\} \neq \text{sign}\{\sigma(r) - \sigma(s)\} \} (r, s = 1, \dots, n) \\ &= \text{minimal numbers of transpositions to transform } \pi \text{ into } \sigma, \\ \rho &= \sum_{r=1}^n (\pi(r) - \sigma(r))^2, \\ f &= \sum_{r=1}^n |\pi(r) - \sigma(r)|. \end{aligned}$$

These measures are not normalized. Usually they are used in their normalized versions which we will denote by $\tilde{\tau}$, $\tilde{\rho}$ and \tilde{f} . The normalization of τ is done using the transformation

$$\tilde{\tau} = 1 - \frac{2\tau}{\binom{n}{2}}$$

because 0 is the minimal value of τ and $\binom{n}{2}$ is the maximal value and so $\tilde{\tau}$ lies in the interval $[-1, 1]$. For $\tilde{\rho}$ and \tilde{f} we use analogous transformations regarding that ρ has a range from 0 to $\frac{n^3-n}{3}$ and f a range from 0 to $\frac{n^2}{2}$ for even n and to $\frac{n^2-1}{2}$ for odd n .

Lemma 2:

The distances measures τ , f and $\rho' := \sqrt{\rho}$ are right-invariant metrics, i.e. they fulfill:

$$d(\pi, \sigma) > d(\pi, \pi) = 0,$$

$$\begin{aligned}
d(\pi, \sigma) &= d(\sigma, \pi), \\
d(\pi, \sigma) &\leq d(\pi, \kappa) + d(\kappa, \sigma), \\
d(\pi, \sigma) &= d(\pi \circ \kappa, \sigma \circ \kappa).
\end{aligned}$$

3 The 'best' J-Scale

To find the 'best' (qualitative) J-scale we have to solve the following steps:

- to find the I-scales which are compatible (to the J-scale) and which have minimal distances in the sum to the observed I-scales for all $i = 1, \dots, n$.
- to ensure that the datamatrix I contains no intransitive triple.

For the first step we have to solve the minimizing problem for each i mentioned above

$$d(I_i, v_{i,k}) = \min_k ,$$

where $v_i = \{v_{i,1}, \dots, v_{i,l_i}\}$, $l_i = \binom{n-1}{i-1}$, is the set of compatible I-scales. Then we have to calculate a coefficient

$$\kappa_{d,n} = \sum_{i=1}^n d(I_i, v_{i,k}) .$$

The 'best' (qualitative) J-scale is that J-scale where $\kappa_{d,n}$ is minimal. At last we have to proof that the datamatrix I contains no intransitive triple. If (at least) one intransitive triple exists exchange one of the involved I-scales with the second best compatible one and repeat the second step and calculate $\kappa_{d,n}$ once more.

The selection of a compatible I-scale in the first step is not always unique. In these cases it is necessary to pursue all solutions using these different compatible I-scales. Also it is possible to exchange the two steps but some empirical studies have shown that it is better to execute the first step before the second one. If an I-scale is compatible it rarely belongs to an intransitive triple. We illustrate this comment by an example.

Example 4:

Given the following observed I-scale I_B and the corresponding J-scale (determined by the remaining three I-scales):

J-Scale A B C D
I-Scale B D A C .

In the case of $d = \tau$ the best compatible I-scales I_B — each of them with two transpositions — are B A C D and B C D A .

Frequency Distributions

We have introduced the statistical model to measure and to judge the deviation from an unidimensional scale. To rate the size of this deviation we need the exact distribution of $\kappa_{d,n}$. Therefore we have to calculate the exact frequency distribution of

$$\begin{aligned}\tau_{n,i}(I_i) &:= \min\{\tau(I_i, v_{i,1}), \dots, \tau(I_i, v_{r,i})\} \\ \rho_{n,i}(I_i) &:= \min\{\rho(I_i, v_{i,1}), \dots, \rho(I_i, v_{r,i})\} \\ f_{n,i}(I_i) &:= \min\{f(I_i, v_{i,1}), \dots, f(I_i, v_{r,i})\}\end{aligned}$$

which is the frequency distribution corresponding to our minimizing problem. Analogously we can define $\check{d}_{n,i}(I_i)$, which is the normalized version of $d_{n,i}(I_i)$ where d is τ, ρ or f . Here the minimum function converts to a maximum function:

$$\check{d}_{n,i}(I_i) := \max\{\check{d}(I_i, v_{i,1}), \dots, \check{d}(I_i, v_{r,i})\} .$$

Obviously exists for O_1 and O_n only one compatible I-scale. Here is the frequency distribution identical to the usual frequency distribution of τ, ρ or f . For more details see Kendall (1970). We have calculated these frequency distributions for all considered coefficients in the range of $n = 3, \dots, 9$ and $i = 1, \dots, \lfloor \frac{n+1}{2} \rfloor$. Because of symmetry the distributions of $d_{n, \lfloor \frac{n+1}{2} \rfloor + 1}, \dots, d_{n,n}$ can be ignored. In Fig.2-5 graphs of the frequency distributions of $\check{d}_{n,i}$ for distinct combinations of n and i are shown. We use the symbols t, r and f to mark the points of the different line graphs. Detailed tables for $\tau_{n,i}$ can be found in Ostermann(1987,1988).

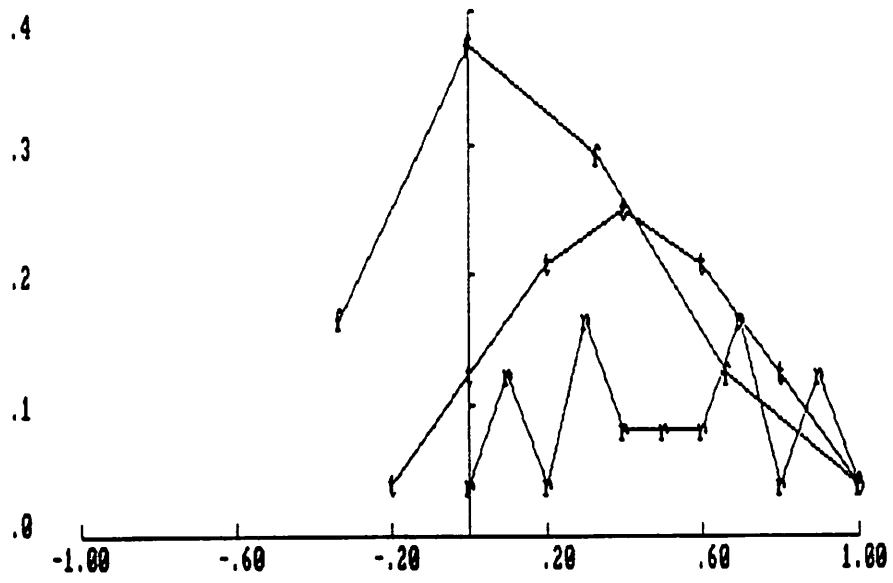


Fig.2: Frequency Distributions of $\hat{\tau}_{5,1}$, $\hat{\rho}_{5,1}$ and $\hat{f}_{5,1}$

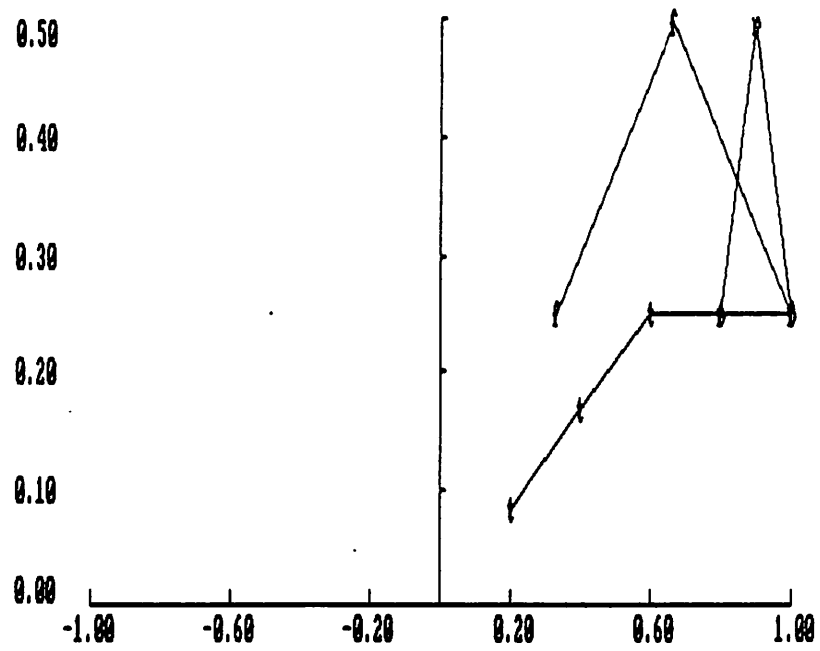


Fig.3: Frequency Distributions of $\hat{\tau}_{5,3}$, $\hat{\rho}_{5,3}$ and $\hat{f}_{5,3}$

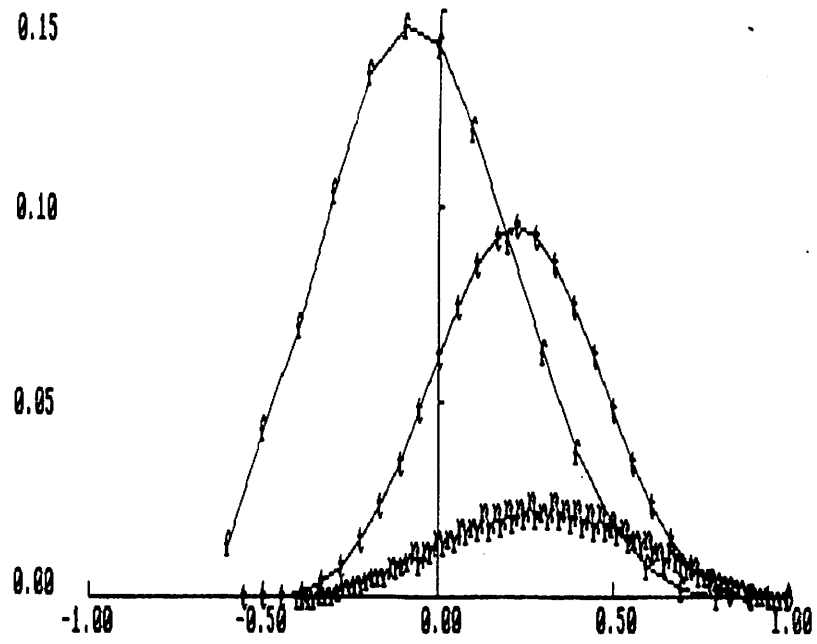


Fig.4: Frequency Distributions of $\tilde{\tau}_{9,1}$, $\hat{\rho}_{9,1}$ and $f_{9,1}$

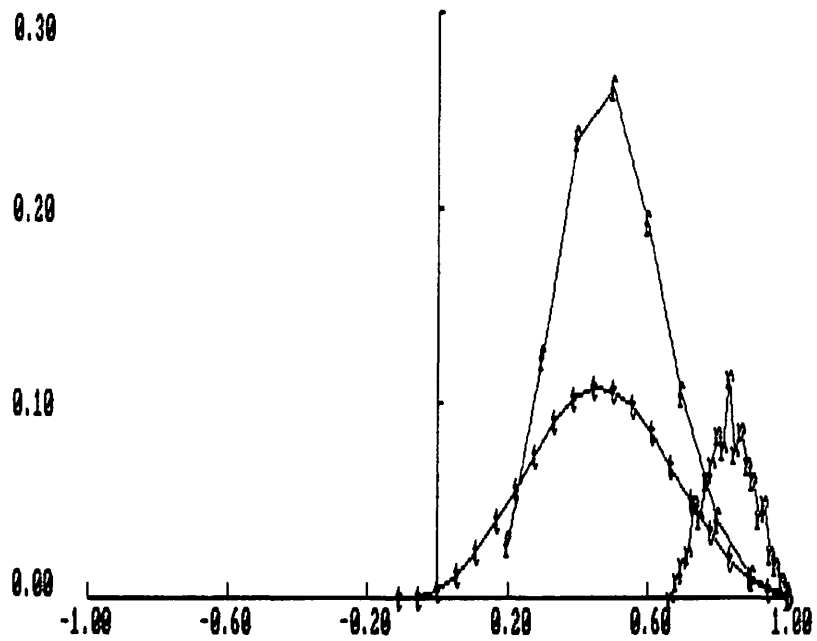


Fig.5: Frequency Distributions of $\tilde{\tau}_{9,5}$, $\hat{\rho}_{9,5}$ and $f_{9,5}$

Comment

The following remarks to the figures above seem necessary:

- We use the graphs of \tilde{d} instead of d because all graphs then lie in the interval $[-1, 1]$.
- All graphs are a little bit biased because we have not rescaled using an appropriate scale parameter. For the determination of such scale parameters see Kendall (1970) or Diaconis (1988).
- The graphs of $\tilde{\tau}$ tend more to a normal curve than the graphs of $\tilde{\rho}$, but both are symmetric.
- The graphs of f are non-symmetric and at -1 they do not tend to zero. The coefficient f has no well-defined properties. These defects are already described in Deuchler (1911).
- We get the impression that for fixed $i \geq 2$ and increasing n the graphs equal more and more the graph of $i = 1$. We have seen this phenomenon when preparing more of such graphs on our computer screen. In the discussion about asymptotical results we consolidate this impression.

In a second step we now have to calculate the frequency distribution of

$$\kappa_{d,n} = \sum_{i=1}^n d(I_i, v_{i,k}) .$$

Let $H(\kappa_{d,n})$ denote the absolute frequency distribution of $\kappa_{d,n}$. $\kappa_{d,n}(j)$ means that $\kappa_{d,n}$ has the value j . $d_{n,i}$ is defined analogously. It holds:

$$H(\kappa_{d,n}(j)) = \sum \prod_{i=1}^n H(d_{n,i}(j_i))$$

with

$$j = \sum_{i=1}^n j_i, \quad j_i = 0, \dots, \max(d_{n,i}) .$$

Obviously

$$\min \kappa_{d,n} = 0 \quad \text{and} \quad \max \kappa_{d,n} = \sum_{i=1}^n \max(d_{n,i})$$

is satisfied. In the tables 2-4 selected quantils for different n in the case of our three considered distance measures are shown. We restrict these tables to the 5%, 1% and 0.1% quantils.

Table 2: Selected Quantils of $\kappa_{\tau,n}$

| n | $\alpha = 0.05$ | $\alpha = 0.01$ | $\alpha = 0.001$ |
|-----|-----------------|-----------------|------------------|
| 3 | 0 | 0 | 0 |
| 4 | 1 | 0 | 0 |
| 5 | 5 | 4 | 3 |
| 6 | 15 | 12 | 9 |
| 7 | 30 | 25 | 21 |
| 8 | 52 | 46 | 40 |
| 9 | 82 | 75 | 67 |

Table 3: Selected Quantils of $\kappa_{\rho,n}$

| n | $\alpha = 0.05$ | $\alpha = 0.01$ | $\alpha = 0.001$ |
|-----|-----------------|-----------------|------------------|
| 3 | 0 | 0 | 0 |
| 4 | 2 | 0 | 0 |
| 5 | 12 | 8 | 4 |
| 6 | 46 | 30 | 20 |
| 7 | 92 | 74 | 56 |
| 8 | * | * | * |
| 9 | * | * | * |

Table 4: Selected Quantils of $\kappa_{f,n}$

| n | $\alpha = 0.05$ | $\alpha = 0.01$ | $\alpha = 0.001$ |
|---|-----------------|-----------------|------------------|
| 3 | 0 | 0 | 0 |
| 4 | 2 | 0 | 0 |
| 5 | 8 | 6 | 2 |
| 6 | 22 | 18 | 14 |
| 7 | 42 | 38 | 32 |
| 8 | 70 | 64 | 56 |
| 9 | * | * | * |

4 Examples

By means of the algorithm mentioned above is it always possible to solve the unidimensional unfolding problem. But does it makes sense for all data situations ? We want to discuss this question by a first example.

Example

Given four objects and their I-scales which yield the datamatrix I:

| | | | |
|----------|----------|----------|----------|
| <i>A</i> | <i>B</i> | <i>C</i> | <i>D</i> |
| <i>B</i> | <i>D</i> | <i>A</i> | <i>C</i> |
| <i>C</i> | <i>A</i> | <i>D</i> | <i>B</i> |
| <i>D</i> | <i>C</i> | <i>B</i> | <i>A</i> |

Applying our algorithm in the case of $d = \tau$ we get the results of Tab.5.

Tab.5: Number of used transpositions of A, B, C and D
for all possible J-scales

| J-scale | $\tau_{4,A}$ | $\tau_{4,B}$ | $\tau_{4,C}$ | $\tau_{4,D}$ | $\kappa_{\tau,4}$ |
|---------|--------------|--------------|--------------|--------------|-------------------|
| ABCD | 0 | 2 | 2 | 0 | 4 |
| ABDC | 1 | 0 | 2 | 0 | 3 |
| ACBD | 1 | 1 | 1 | 1 | 4 |
| ACDB | 2 | 1 | 0 | 0 | 3 |
| ADBC | 2 | 0 | 3 | 1 | 6 |
| ADCB | 3 | 2 | 1 | 0 | 6 |
| BACD | 0 | 2 | 0 | 1 | 3 |
| BADC | 1 | 1 | 1 | 1 | 4 |
| BCAD | 1 | 3 | 0 | 2 | 6 |
| BDAC | 2 | 0 | 0 | 2 | 4 |
| CABD | 0 | 0 | 1 | 2 | 3 |
| CBAD | 0 | 1 | 2 | 3 | 6 |

We need at least three transpositions to get a datamatrix where all I-scales are compatible to one J-scale and where we have no intransitive triplet. How good does our new datamatrix represent our empirical matrix I? If we compare our result $\kappa_{\tau,4} = 3$ to the corresponding value of table 2 we see that $\kappa_{\tau,4} = 3$ is greater than the 5%-quantil which is 1. In the case of 4 objects using τ as distance measure it is only acceptable to use unidimensional unfolding if we need one transposition at most. For the data of this example a multidimensional approach is more appropriate.

As a second example for demonstrating our new algorithm we use the data of Coombs (1964, p.456-462).

Example

Given the following six I-scales which yield the datamatrix I_0 :

| | | | | | |
|----------|----------|----------|----------|----------|----------|
| <i>A</i> | <i>C</i> | <i>D</i> | <i>B</i> | <i>F</i> | <i>E</i> |
| <i>B</i> | <i>D</i> | <i>E</i> | <i>A</i> | <i>F</i> | <i>C</i> |
| <i>C</i> | <i>A</i> | <i>F</i> | <i>D</i> | <i>B</i> | <i>E</i> |
| <i>D</i> | <i>B</i> | <i>F</i> | <i>E</i> | <i>A</i> | <i>C</i> |
| <i>E</i> | <i>D</i> | <i>B</i> | <i>F</i> | <i>A</i> | <i>C</i> |
| <i>F</i> | <i>D</i> | <i>C</i> | <i>E</i> | <i>B</i> | <i>A</i> |

This matrix I_0 violates the two elementary restrictions:

- I_F does not end up with C or E.
- The I-scale I_E is not the reverse one of I_C .

Using our new algorithm we get the results of the tables 6a and 6b for the potential J-scales CAFBDE and CAFDBE, respectively.

Tab.6a: Results for the data of Coombs

| J-scale: CAFDBE | | | | |
|---------------------|------------------------------------|--------------|--------------|-----------------------|
| Observed I-scale | Selected, compatible I-scale | $\tau_{6,i}$ | $\rho_{6,i}$ | Value of $f_{6,i}$ |
| ACDBFE | ACFDDBE | 2 | 8 | 4 |
| BDEAFC | BDEFAC | 1 | 2 | 2 |
| CAFDBE | CAFDBE | 0 | 0 | 0 |
| DBFEAC | DBFEAC | 0 | 0 | 0 |
| EBDFAC | EBDFAC | 1 | 2 | 2 |
| FDCEBA | FDACBE | 4 | 14 | 6 |
| | FDDBEAC | 4 | 14 | 6 |
| | FDBACE | 4 | 16 | 8 |

Tab.6b: Results for the data of Coombs

| Observed I-scale | J-scale: CAFBDE | | | |
|---------------------|------------------------------------|--------------|--------------|-----------------------|
| | Selected, compatible I-scale | $\tau_{6,i}$ | $\rho_{6,i}$ | Value of $f_{6,i}$ |
| ACDBFE | ACFBDE | 3 | 8 | 4 |
| BDEAFC | BDEFAC | 1 | 2 | 2 |
| CAFDBE | CAFBDE | 1 | 2 | 2 |
| DBFEAC | DBFEAC | 0 | 0 | 0 |
| EBDFAC | EDBFAC | 0 | 0 | 0 |
| FDCEBA | FBDACE | 5 | 20 | 8 |

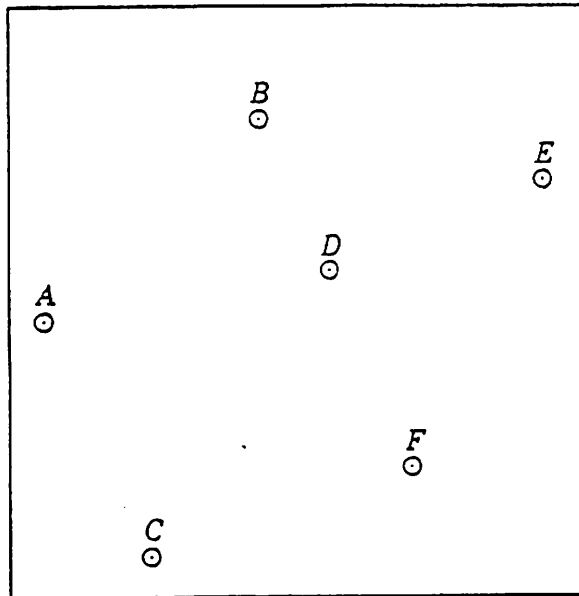
The selection of a compatible I-scale I_F at the J-scale CAFDBE is not unique. Therefore we have to pursue three solutions. The values of $\kappa_{d,6}$ are $\kappa_{\tau,6} = 8, \kappa_{\rho,6} = 26$ or 28 and $\kappa_{f,6} = 14$ or 16 in the case of the J-scale CAFDBE. In the case of CAFBDE they are $\kappa_{\tau,6} = 10, \kappa_{\rho,6} = 32$ and $\kappa_{f,6} = 16$. So we decide for CAFDBE independent of our distance measure. The values of $\kappa_{d,6}$ correspond to a quantil lower than 0.01.

Using the triangular analysis and the Goode-Phillips-algorithm we get the following three unidimensional solutions (Fig.6) which correspond to our three different compatible I-scales I_F . We have to compare these three different unidimensional solutions to a twodimensional solution of Coombs (1964, p.462) in Fig.7. If we project the twodimensional solution onto its \overline{CE} -line we get a unidimensional one which is very similar to our three unidimensional ones. So it makes no difference which representation we use — an one- or a two-dimensional representation of the given dataset. We prefer the one-dimensional representation which has a simpler interpretation.

Fig.6: Three Unidimensional Solutions

| Solution with $I_F = \text{FDACBE}$ | | | | | |
|-------------------------------------|-------|-------|-------|-------|-------|
| C | A | F | D | B | E |
| 0.0 | 4.76 | 47.62 | 69.05 | 80.95 | 100.0 |
| Solution with $I_F = \text{FDBEAC}$ | | | | | |
| C | A | F | D | B | E |
| 0.0 | 20.75 | 62.26 | 77.35 | 84.90 | 100.0 |
| Solution with $I_F = \text{FDBACE}$ | | | | | |
| C | A | F | D | B | E |
| 0.0 | 11.86 | 35.58 | 57.61 | 72.86 | 100.0 |

Fig.7: Twodimensional Solution of Coombs
(Copy of Coombs (1964))



5 Asymptotical Results

In Tab.7 we have listed the first central moments of the frequency distributions of $d_{n,1}$ (in dependence of n), which can be found in Diaconis (1988).

Tab.7: The First Two Central Moments

| d | $E(d_{n,1})$ | $Var(d_{n,1}) = \sigma^2(d_{n,1})$ |
|--------|--------------------------|------------------------------------|
| τ | $\frac{\binom{n}{2}}{2}$ | $\frac{n(n-1)(2n+5)}{72}$ |
| ρ | $\frac{n^3-n}{6}$ | $\frac{n^2(n-1)(n+1)^2}{36}$ |
| f | $\frac{n^2-1}{3}$ | $\frac{(n+1)(2n^2+7)}{45}$ |

For $n \rightarrow \infty$ the result of the usual normalization

$$\frac{d - E(d)}{\sigma(d)}$$

can be treated as an unit normal random variable. (Here a possible continuity correction is not considered.) Now we can formulate the following theorem.

Theorem 1:

For fixed $i \in \{1, \dots, n\}$ holds:

$$\lim_{n \rightarrow \infty} E(\tilde{\tau}_{n,i}) = 0.$$

Proof:

$d=\tau$ Let i be fixed and $v_i = \{v_{i,1}, \dots, v_{i,l_i}\}$ be the set of compatible I-scales.
 $\forall k, l \in \{1, \dots, l_i\}$ holds

$$\tau(v_{i,k}, v_{i,l}) \in O(n)$$

and so (using the attributes of right-invariant metrics)

$$E(\tau_{v_i,k}) \geq E(\tau_{n,i}) \geq E(\tau_{v_i,k}) - E(O(n)) .$$

Therefore

$$0 = E(\tilde{\tau}_{v_i,k}) \leq E(\tilde{\tau}_{n,i}) \leq E(\tilde{\tau}_{v_i,k}) + E(O(n))$$

because $\sigma^2(\tau) \in O(n^3)$.

$d=\rho$ It is $\sigma^2(\rho) \in O(n^5)$ and the maximal value of ρ between two compatibles I-scales is an element of $O(n^2)$.

$d=f$ It is $\sigma^2(f) \in O(n^3)$ and the maximal value of f between two compatibles I-scales is an element of $O(n)$.

Using the same arguments we formulate Theorem 2.

Theorem 2:

a)

$$\lim_{n \rightarrow \infty} \text{Range}(\tilde{d}_{n,i}) = 2 \quad \text{for fixed } i.$$

b)

$$\begin{aligned} \lim_{n \rightarrow \infty} F_{d_{n,i}}^* &= \Phi(x) \quad \text{for fixed } i \text{ and} \\ \tau_{n,i}^* &= \tilde{\tau}_{n,i} \cdot \left(\frac{9}{2} \frac{n(n-1)}{2n+5} \right)^{\frac{1}{2}} , \\ \rho_{n,i}^* &= \tilde{\rho}_{n,i} \cdot (n-1)^{\frac{1}{2}} , \\ f_{n,i}^* &= \tilde{f}_{n,i} \cdot \left(\frac{45}{16} \frac{n^4}{(n+1)(2n^2+7)} \right)^{\frac{1}{2}} . \end{aligned}$$

The normalization of $d_{n,i}^*$ is derived from the results of Tab.7 . So we treat the function $d_{n,i}^*$ as an unit normal random variable.

Finally we transfer the results of the theorems 1 and 2 to the coefficient $\kappa_{d,n}$.

Theorem 3:

Let

$$\kappa_{d,n}^* = \frac{1}{n} \left(1 - \frac{2\kappa_{d,n}}{\max(d_{n,1})} \right) \cdot \text{scale factor},$$

where the scale factor is the corresponding factor of theorem 2. Then it holds

a)

$$\lim_{n \rightarrow \infty} E(\kappa_{n,i}^*) = 0 ,$$

b)

$$\lim_{n \rightarrow \infty} F_{\kappa_{n,i}^*} = \Phi(x) .$$

6 Conclusion

In this paper we have proposed a statistical error model including an algorithm for solving the unidimensional unfolding problem for slightly non-errorless data, too. By means of the exact distribution of a new coefficient it is possible to check whether it makes sense to use this algorithm because now the unidimensional unfolding problem can be solved for all data. These check shall prevent misuse of our new algorithm.

Some asymptotical results are given. They support that our model is embedded in a solid theory because the asymptotic distribution of our new coefficient tends to a normal curve. In an example we have shown that now it is possible to scale data in one dimension where Coombs (1964) had to scale these data in two dimensions.

References

- Chritchlow, D.E. (1985), *Metric Methods for Analyzing Partially Ranked Data*, Springer, Berlin Heidelberg
- Coombs, C.H. (1964), *A Theory of Data*, Wiley, New York
- Deuchler, G. (1914), Über die Methoden der Korrelationsrechnung in der Pädagogik und Psychologie, *Zeitschrift für Pädagogische Psychologie und Experimentielle Pädagogik*, 15, 114-242
- Diaconis, P. (1988), *Group Representations in Probability and Statistics*, Lecture Notes - Monograph Series, Institute of Mathematical Statistics, Volume 11
- Hartung, J. & Elpelt, B. (1984), *Multivariate Statistik*, Oldenbourg, München
- Kendall, M.G. (1970), *Rank Correlations Methods*, Fourth Edition, Griffin, London
- Lehner, P.E. & Noma, E. (1980), A New Solution to the Problem of Finding All Numerical Solutions to Ordered Metric Structures, *Psychometrika*, 45, 135-137
- Ostermann, R. (1987), *Eine Erweiterung der Unfolding-Technik bei fehlerbehafteten Daten*, University of Dortmund, Doctoral Thesis
- Ostermann, R. (1988), A Statistical Model for Unidimensional Unfolding, in: Jansen, M.G.H. & van Schuur, W.H. (1988), *The Many Faces of Multivariate Analysis*, Proceedings of the SMABS-88 conference, Groningen, 270-281
- Phillips, J.P.N. (1971), A Note on the Representation of Ordered Metric Scaling, *British Journal of Mathematical and Statistical Psychology*, 24, 239-250
- Van Blokland-Vogelesang (1985), Unidimensional Unfolding Complemented by Feigin & Cohen's Error model and by Davison's Test for Independence, *Computational Statistics Quarterly*, 2, 43-52
- Van Blokland-Vogelesang (1988), Unfolding and Consensus Rankings: A Prestige Ladder for Technical Occupations, Paper distributed at the SMABS-88 conference, Groningen

Seit 1989 erschienene Diskussionsbeiträge:
Discussion papers released as of 1989/1990:

- 1-89 **Klaus Schöler**, Zollwirkungen in einem räumlichen Oligopol
- 2-89 **Rüdiger Pethig**, Trinkwasser und gewässergüte. Ein Plädoyer für das Nutzerprinzip in der Wasserwirtschaft
- 3-89 **Rüdiger Pethig**, Calculus of Consent: A Game-theoretic Perspective. Comment
- 4-89 **Rüdiger Pethig**, Problems of Irreversibility in the Control of Persistent Pollutants
- 5-90 **Klaus Schöler**, On Credit Supply of PLS-Banks
- 6-90 **Rüdiger Pethig**, Optimal Pollution Control, Irreversibilities, and the Value of Future Information
- 7-90 **Klaus Schöler**, A Note on "Price Variation in Spatial Markets: The Case of Perfectly Inelastic Demand"
- 8-90 **Jürgen Eichberger and Rüdiger Pethig**, Constitutional Choice of Rules
- 9-90 **Axel A. Weber**, European Economic and Monetary Union and Asymmetries and Adjustment Problems in the European Monetary System: Some Empirical Evidence
- 10-90 **Axel A. Weber**, The Credibility of Monetary Target Announcement: An Empirical Evaluation
- 11-90 **Axel A. Weber**, Credibility, Reputation and the Conduct of Economic Policies Within the European Monetary System
- 12-90 **Rüdiger Ostermann**, Deviations from an Unidimensional Scale in the Unfolding Model

