STATISTICS IN TRANSITION-new series, October 2009
Vol. 10, No. 2, pp. 277—299

STATISTICAL MODELS FOR SOCIAL STRESS ANALYSIS

Luisa Canal[1], Walenty Ostasiewicz[2]

ABSTRACT

This paper presents the critical overview of the statistical models that could be used in the social stress analysis. Such analysis is considered to consist of the identification of the social stressors, and of the measurement of their potency to destroy the social harmony. Four main groups of methods are discussed: item response models, factorial models, latent classification, and paired comparison.

1. Introduction

In the 1970s the world entered a period of transition, whose length is difficult to predict. Almost all social and economical arrangements are crumbling. Neo-classical liberalism seems to dominate the economic thought in all countries. The influence of multiple factors of the globalization is visible in all aspects of human life. Such transition has positive effects as well as enormous negative impacts on the social life. Values of citizenships are decaying, societies are being based on individualistic egotism. Above all, social ties that are necessary for social cohesion are diminishing.

In addition to the negative influence of the globalization processes, there are various internal phenomena which contribute negatively to the harmony of social life. Those phenomena distress people, irritate them, and in a consequence can even destroy the existing social order. In this paper, we assume that any phenomenon, event, or condition which may have a destructive impact on the social life can be called *social stressor*.

The literature concerning the stress analysis is quite rich. However, it is to be found mainly within the psychology, and above all in the medical psychology. Furthermore, it concentrates on psychological distress rather than on the social one. Problems of social distress are scarcely documented.

The basic premise of this paper is that the prevention of fraying of the social fabric is a more appropriate policy than its mending. In order to prevent the

[1] Trient University.
[2] Wroclaw University of Economics.

fraying of the fabric of our societies, the underlying destructive forces have to be identified and then evaluated. This paper discusses the available statistical methods which might be useful in analyzing the social stressors. Those methods include the identification, the measuring and the evaluation of perils and risks to social cohesion.

Social stressor is any phenomenon, event or condition of living which irritates people thus deteriorating the social atmosphere of common living. It has a negative impact on social cohesion and social harmony.

The adjective "social" in the term "social stressor" is used to emphasize the fact that the investigated phenomenon produces distress which could be common for a large group of people. It means that the existence of some *common sense* or common feature characterizing a whole group of people is assumed. As this characteristic is not observed directly, it is called a latent variable. It will be denoted by the symbol Z, and it is assumed that it "drives", commands or controls people's reaction to stressful phenomena. For the lack of established terminology, a latent variable Z will be called *susceptibility*, endurance, resistance or patience. To keep the discussion general enough, we admit a number of aspects of the susceptibility. Therefore, the trait Z is considered as a d-dimensional variable $Z = (Z_1, Z_2, ..., Z_d)$.

As different people are endowed with different amounts of susceptibility, we will interpret the trait Z as a random variable. The cumulative distribution of it is denoted by $H(z) = H(z_1, z_2, ..., z_d)$.

For illustrative purposes, here are a few examples of stressors.

1. Almost legalized political corruption,
2. Cynicism of politicians,
3. Brutality in TV movies,
4. Immoral behaviour of higher officials.

All phenomena of that kind will be denoted by symbols $Y_1, Y_2, ..., Y_p$. They will be also called *items*.

In order to assess how dangerous those phenomena are in destroying the social cohesion, the survey methodology will be applied. The measurement of the strength of a stressor can be done by "observing" people's reaction. By a reaction we mean an answer to a question concerning undesired phenomena. Two kinds of questions and two broad approaches to analysis of collected responses are being discussed: categorical responses and comparative responses.

To collect data, items $Y_1, Y_2, ..., Y_p$ representing appropriate stressors are prepared, and then they are administered to a group of n people (respondents), who are treated as experts or judges.

In the case of categorical responses, the task of the subject is to endorse or reject the administered item. If, for example, we would like to know whether or not a non-controlled immigration is a risk for social cohesion, we could form the

following item: "non-controlled immigration is dangerous for social cohesion", and then ask respondents to confirm or to reject this assertion.

All responses to *p* items $Y_1, Y_2, ..., Y_p$. given by *n* respondents are presented in the form of the following matrix

$$\begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & & & \\ y_{n1} & y_{n2} & \cdots & y_{np} \end{bmatrix}$$

Symbol y_{ij} has the following meaning

$$y_{ij} = \begin{cases} 1, \text{if item } Y \text{ is endorsed by ith respondent} \\ 0, \text{if item } Y \text{ is rejected by ith respondent} \end{cases}$$

In order to avoid the confusion, the *i-th* row of this matrix

$$y_i = (y_{i1}, y_{i2}, ..., y_{ip})$$

will be denoted by $y^i = (y_1^i, y_2^i, ..., y_p^i)$.

Reaction (or response) obtained from *i-th* individual forms the vector (the *i-th* row of data matrix):

$$y^i = (y_1^i, y_2^i, ..., y_p^i)$$

In the case of a generic individual, a simplified notation can be used:

$$y = (y_1, y_2, ..., y_p).$$

For typographical reasons, instead of y_j^i , a symbol y_{ij}. will be used.

In the case of comparative responses, the subject's task is to decide which item from a pair (Y_j, Y_k) is more dangerous for social cohesion. The results of this survey form the following matrix:

$$\begin{bmatrix} n_{11} & n_{12} & \ldots & n_{1p} \\ \vdots & \vdots & \vdots & \vdots \\ n_{p1} & n_{p2} & \cdots & n_{pp} \end{bmatrix}$$

Entry n_{jk} stands for the number of respondents who asserted that Y_j is at least as dangerous as Y_k. For convenience, we put $n_{jj} = n$, $j = 1,2,..., p.$

2. General model for categorical responses

There is a long tradition of investigating categorized data, particularly data obtained from the responses of people. Usually, the existence of some hypothetical variable "driving" these responses is assumed. The categorized responses are supposed to be determined by the individual's position in underlying space of latent characteristics, also called latent variables or latent traits. Usually, this space is a one-dimensional continuum.

The significance of the latent traits, on which much of the psychological theory is based, has been precisely explained by Birnbaum: "nowhere is there any necessary implication that traits exist in any physical or psychological sense. It is sufficient that person behaves as if he were in the possession of a certain amount of each of a number of relevant traits and that he behave as if these amounts substantially determined his behaviour" (see [14]).

All responses y_{ij} are supposed to be random outcomes. The response $y_{ij} \in \{0,1\}$ given by the *i-th* individual to the *j-th* item is interpreted as a realization of a Bernoullian random variable Y_{ij} with a distribution

$$P(Y_{ij} = 1), P(Y_{ij} = 0) = 1 - P(Y_{ij} = 1).$$

In the case of a generic individual, a simpler notation will be used:

$$P(Y_j = 1), P(Y_j = 0) = 1 - P(Y_j = 1).$$

The probability distribution of the responses to all the items is denoted as

$$f(y) = \mathrm{P}(\mathrm{Y}_1 = y_1, ...Y_p = y_p).$$

This distribution is supposed to depend on a random variable Z, and the conditional distribution is denoted by

$$f(y|z) = P(Y_1 = y_1, ..., Y_p = y_p) \big| Z_1 = z_1, ..., Z_p = z_p).$$

Using the rule that $P(A) = P(A|B) \cdot P(B)$ we can infer the following representation:

$$f(y) = \int f(y|z) dH(z).$$

This is the fundamental representation of the probability distribution of the observed data. It forms the basis for all theories commonly called latent variables models (see [6]). Unfortunately, in such general settings, it has no use in practice, as it can be neither falsified nor verified.

To make the settings more restrictive, three fundamental hypotheses (assumptions) are introduced (see [15,24]).

The first hypothesis.

The first restriction requires that people with a greater level of the susceptibility are more inclined to react to the stressful phenomena. In other words, more susceptible people can be easier provoked by stressors beyond their endurance. Using more technical language, this assumption can be formulated as follows:

Probability of an affirmative response to an item representing a stressful phenomenon is a non-decreasing function of the respondent's susceptibility.

This assumption is called monotonicity (M), and formally it is expressed as follows:

(M) $P(Y=1|Z=z)$ is a coordinate-wise non-decreasing function in Z.

The second hypothesis.

The second assumption, known as conditional independence or latent independence (LI) requires that reactions of one subject to different stressors are independent.

Formally it takes the following form:

$$\text{(LI)} \qquad P(Y_1=y_1,...,Y_p=y_p|Z_1=z_1,...,Z_d=z_d)=\prod_{j=1}^{p}P(Y_j=y_j|Z=z)$$

Using notations

$$f(y|z)=P(Y_1=y_1,...,Y_p=y_p)|Z_1=z_1,...,Z_p=z_p)$$

$$f_j(y_j|z)=P(Y_j=y_j|Z_1=z_1,...,Z_d=z_d)$$

this condition can be written down as follows:

$$f(y|z)=\prod_{j=1}^{p}f_j(y_j|z)\,.$$

Hence, the basic representation will take the form:

$$f(y)=\int\prod_{j=1}^{p}f_j(y_j|z)dH(z)\,.$$

To be consistent with the main stream of publications the following notation is introduced:

$$\pi_{ij}(z)=P(Y_{ij}=1|Z=z)=P(Y_{ij}=1|Z_1=z_1,Z_2=z_2,...,Z_d=z_d).$$

More compact form of it will also be used:

$$\pi_{ij}(z) = P(Y_{ij} = 1 | Z = z).$$

Omitting the index of the individual, it will take the form:

$$\pi_j(z) = P(Y_j = 1 / Z = z)$$

Taking into account that:

$$P(Y_j = 0 / Z = z) = 1 - \pi_j(z)$$

and that $y_j \in \{0,1\}$ one arrives at the distribution

$$f_j(y_j / z) = P(Y_j = y_j / Z = z) = \pi_j(z)^{y_j} (1 - \pi_j(z))^{1-y_j}.$$

As a result, the following representation is obtained:

$$f(y) = \int \prod_{j=1}^{p} \pi_j(z)^{y_j} (1 - \pi_j(z))^{1-y_j} \, dF(z)$$

The probability distribution function for observed (manifest) data for the *i-th* subject will take the following form:

$$f(y^i) = \int \prod_{j=1}^{p} \pi_{ij}(z)^{y_{ij}} (1 - \pi_{ij}(z))^{1-y_{ij}} \, dF(z).$$

The third hypothesis.

The third assumption requires that the susceptibility is a one-dimensional continuum. It is called unidimensionality (U), and its formal expression is very simple:

(U) d=1

When Z is unidimensional, probability $\pi_{ij}(z)$ treated as a function of z is usually called item characteristic function, or item characteristic curve (ICC).

3. General implications

The basic representation

$$f(y) = \int f(y|z) dH(z)$$

with the conditions (LI), (M) and (U) defines reasonable non-parametric model for the observed data. The three conditions: (LI), (M), and (U) are minimal assumptions necessary to obtain a falsifiable model for the analysis of the perceived stressors.

It means that from this model one can infer a number of properties for the observed data. The most important implications are briefly summarized below.

First of all, it is worthy to note that no two of three assumptions define a restrictive model. If any of these three conditions is completely relaxed, then the resulting "model" will fit any distribution of binary data (see[19]). For example, it is well known that the assumption of unidimensionality and local independence are independent in the sense that neither, either one or both assumptions, can hold for the above representation (see [29,30]).

What follows from this result is the suggestion for testing of the hypothesis.

For example, the evidence that follows from condition (LI) and the lack of fit is that $d \neq 1$. Condition $d = 1$ and the lack of fit might be considered as the evidence of non-local independence.

The necessary conditions for the representation $f(y)$ were found by P. W. Holland in 1981 (see[16]). By weakening the condition of local independence ([16,18]) to the form of local non-negative dependence (LND) it was possible to establish the necessary as well as the sufficient conditions.

The (LI) and (M) conditions imply that vector $(Y_1, Y_2, ..., Y_p)$ is *associated*, and this means that for all monotone item summaries expressed by non-decreasing functions g_1 and g_2 holds the inequality: $Cov(g_1(Y), (g_2(Y)) \geq 0$. Furthermore, more powerful result was obtained (see[19,28]):

$$Cov(g_1(Y_A), (g_2(Y_A) \mid h(Y_B) = z) \geq 0$$

where g_1 and g_2 are non-decreasing functions, h is any function and $\{Y_A, Y_B\}$ is a partition of the set $\{Y_1, Y_2, ..., Y_p\}$.

The interpretation of this important theorem is given in [19,20].

To detect the violation of (LI) condition, one can use the conditional covariance functions $Cov(Y_i, Y_j / Z = z)$. If conditions (LI) and (U) hold, then conditional covariance functions are identically equal to zero:

$$Cov(Y_i, Y_j / Z = z) = 0 \text{ for all z, and all pairs } i \text{ and } j.$$

In practice, however, condition (LI) never holds exactly. For this reason, in order to provide tests and measures of unidimensionality, the concept of *essential independence* with respect to the latent characteristic Z was introduced (see [29,30]). This concept requires that the average value of $\left|Cov(Y_i, Y_j / Z = z)\right|$ over all item pairs Y_i and Y_j should be small for all values z, instead of being equal to zero.

Using this concept, the concept of *the essential unidimensionality* of a set of items $Y_1, Y_2, ..., Y_p$ has been defined as the minimal dimensionality necessary to satisfy the assumption of the essential independence. The statistical procedure for

testing the null hypothesis of the essential unidimensionality has been proposed by Stout in 1987 (see[29]). This rather complicated procedure requires the construction of an estimate of $Cov(Y_i, Y_j \mid Z = z)$ at each value of Z. One way to obtain this estimate is by kernel smoothing. Almost all tests of fit that have been already proposed for non-parametric models of item responses are rather complicated and usually their asymptotic properties cannot be determined mathematically (see[28]).

One can distinguish the following three important particular cases of the general settings (see[17]):

1. Non-parametrical models.

   Probabilities $\pi_j(z)$ and $H(z)$ are both allowed to vary over all non-parametric classes of functions.

2. Semi-parametric models.

   In this class $\pi_j(z)$ have parametric form, and $H(z)$ is allowed to vary over all non-parametric classes of functions.

3. Parametric models.

   In this class, $\pi_j(z)$ and $H(z)$ have parametric forms (see[17]):

$$\pi_j(z) = \pi_0(a_j(z - b_j), c_j)$$

$$H(z) = H_0(z - \frac{\mu}{\sigma}, v)$$

where π_0 and H_0 are specified functions, and z is a location parameter (see[17]).

4. Simple logistic model

4.1. Model derivation

Let us assume the following conditions:

1) each respondent is characterized by latent parameter z_i and to each stressor a real number α_j characterizing its "capability" to provoke distress can be assigned,
2) functions $\pi_j(z) = \pi_j(y_j|z)$ are continuous and strictly monotone increasing for all latent trait parameters z,
3) $\lim_{z\to-\infty} \pi_j(z) = 0$, $\lim_{z\to\infty} \pi_j(z) = 1$
4) the "principle of local stochastic independence" is satisfied for all responses,

5) the row score $T_i = \sum_{j=1}^{p} Y_{ij}$ is sufficient statistics for parameter z_i.

It has been proved (see for example [1,14]) that if these conditions are satisfied, then π_j has the following form:

$$\pi_j(z_i) = P(Y_j = 1/Z = z_i) = \frac{\exp(z_i - \alpha_j)}{1 + \exp((z_i - \alpha_j)}$$

This model is called a simple logistic model, or Rasch model (see [27]).

It depends on $n+p$ parameters:

$$\alpha_1, \alpha_2, ..., \alpha_p, z_1, z_2, ..., z_n .$$

Parameters determining susceptibility of the respondents $z_1, z_2, ..., z_n$ are treated as nuisance parameters.

When estimating these parameters one can follow one of three possible ways:

1) nuisance parameters can be estimated simultaneously with the parameters determining the intensity of the stressors,
2) nuisance parameters can be eliminated by the conditioning,
3) nuisance parameters can be integrated out.

In the first case, joint maximum likelihood (JML) method can be used, in the second case – conditional maximum likelihood (CML) method, and in the third one - marginal maximum likelihood (MML) method.

4.2. Elementary method

The brief review of the above mentioned methods starts with the presentation of the simplest method. In [22] it is called *elementaren Schätzungen.*

Suppose that the Rasch model is valid. This means that the equality

$$\pi_{ij} = \frac{\exp(z_i - \alpha_j)}{1 + \exp((z_i - \alpha_j)}$$

$$\text{holds for all } i = 1, 2, ..., n, \; j = 1, 2, ..., p .$$

The equivalent form of these equations is as follows:

$$Log(\pi_{ij}) = z_i - \alpha_j$$

Replacing π_{ij} by the observed frequency $\frac{n_{ij}}{n}$, parameters z_i and α_j can be found by the solution of the following system of $n \cdot p$ equations:

$$\log\left(\frac{n_{ij}}{n}\right) = z_i - \alpha_j$$

The solution is following [22]:

$$\hat{z}_i = \bar{y}_{i\cdot} - 0{,}5$$
$$\hat{\alpha}_j = \bar{y} - \bar{y}_{\cdot j}$$

where

$$\bar{y}_{i\cdot} = \frac{1}{p}\sum_{j=1}^{p} y_{ij} \quad , \quad \bar{y}_{ij} = \frac{1}{n}\sum_{i=1}^{n} y_{ij}, \qquad \bar{y} = \frac{1}{np}\sum_{i=1}^{n}\sum_{j=1}^{p} y_{ij},$$

Theoretically, this way of estimation is not satisfactory since one cannot make any inferences. Practically, it is justified by its simplicity and, as reported in [22], the results are very close to those obtained by the sound methods.

4.3.Joint maximum likelihood

In order to use the JML let us observe that

$$f(y;z,\alpha) = \prod_{j=1}^{p} \pi_{ij}^{\,y_{ij}} (1-\pi_{ij})^{1-y_{ij}} = \prod_{j=1}^{p} \left[\frac{\exp(z_i - \alpha_j)}{1+\exp((z_i - \alpha_j)}\right]^{y_{ij}} \left[\frac{1}{1+\exp((z_i - \alpha_j)}\right]^{1-y_{ij}}$$

Hence, the joint maximum likelihood function is the following:

$$L(z_1, z_2, ..., z_n, \alpha_1, \alpha_2, ..., \alpha_p) = \prod_{j=1}^{n}\prod_{i=1}^{p} \pi_{ij}^{\,y_{ij}} \left[1-\pi_{ij}\right]^{1-y_{ij}} =$$

$$= \exp\left(\sum_{i=1}^{n}\sum_{j=1}^{p} z_i y_{ij} - \sum_{i=1}^{n}\sum_{j=1}^{p} \alpha_i y_{ij}\right)\left(\prod_{j=1}^{n}\prod_{i=1}^{p}\left(1+e^{z_i-\alpha_j}\right)\right)^{-1} =$$

$$= \exp\left(\sum_{i=1}^{n} z_i t_i - \sum_{j=1}^{p} \alpha_j q_j\right)\left(\prod_{i=1}^{n}\prod_{j=1}^{p}\left(1+\exp\left(z_i-\alpha_j\right)\right)\right)^{-1}$$

where

$$t_i = \sum_{j=1}^{p} y_{ij} \quad q_j = \sum_{i=1}^{n} y_{ij}$$

Estimators of the parameters $\alpha_1, \alpha_2, ..., \alpha_p$ and $z_1, z_2, ..., z_n$ are to be obtained by solving the equations:

$$\frac{\partial \ln L(z_1, z_2, ..., z_n, \alpha_1, \alpha_2, ..., \alpha_p)}{\partial z_i} = 0$$

$$\frac{\partial \ln L(z_1, z_2, ..., z_n, \alpha_1, \alpha_2, ..., \alpha_p)}{\partial \alpha_j} = 0$$

Identifability requires that either

$$\sum_{i=1}^{n} z_i = 0 \quad \text{or} \qquad \sum_{j=1}^{p} \alpha_j = 0 .$$

After simplification:

$$t_i = \sum_{j=1}^{p} \frac{\exp(z_i - \alpha_j)}{1 + \exp(z_i - \alpha_j)}, \quad i = 1,2,...,n$$

$$q_j = \sum_{i=1}^{n} \frac{\exp(z_i - \alpha_j)}{1 + \exp(z_i - \alpha_j)}, \quad j = 1,2,..., p$$

The solution of these equations can be obtained only numerically.

This solution has however some theoretical drawbacks (see[1, 3]).

The number of nuisance parameters grows with the sample size *n*, and item parameters are not consistent (see[14]).

4.4. Conditional maximum likelihood

The drawbacks of JML method can be avoided by using theoretically satisfactory method known as conditional maximum likelihood (CML) method. This method forms two separate likelihoods: one for the item parameters, and the second for the respondent parameters (see [6]).

To obtain the estimating equations, let us observe that

$$f(y) = f(y_{i1}, y_{i2}, ..., y_{ip}|) = f(y|t_i) g(t_i|z_i)$$

Hence, the likelihood function has the form:

$$L(z_1, z_2, ..., z_n, \alpha_1, \alpha_2, ..., \alpha_p) = \prod_{i=1}^{n} f(y_i|t_i) g(t_i|z_i) = \prod_{i=1}^{n} f(y_i|t_i) \prod_{i=1}^{n} g(t_i|z_i) = L_C \cdot L_M$$

Part L_C is used for the estimation of α parameters, and L_M part is used for the estimation of z parameters.

The required conditional distribution

$$f(y|t_i) = P(Y = y|T_i = t_i) = f(y_{i1}, y_{i2},..., y_{ip}|T_i = t_i)$$

is obtained by dividing *f(y)* by $g(t_i|z_i)$.

Let us observe that

$$f(y) = f(y_{i1}, y_{i2},..., y_{ip}) = \frac{\exp(z_i t_i - \sum_{j=1}^{p} \alpha_j y_{ij})}{\prod_{j=1}^{p}(1 + \exp((z_i - \alpha_j))},$$

then, by summation over all possible vectors $(y_{i1}, y_{i2},..., y_{ip})$ with the total $t_i = \sum y_{ij}$
one gets

$$g(t_i|z_i) = P(T_i = t_i) = P(\sum_{j=1}^{p} Y_{ij} = t_i) = \frac{\sum_{=t_i} \exp(z_i t_i - \sum_{j=1}^{p} \alpha_j y_{ij})}{\prod_{j=1}^{p}(1 + \exp((z_i - \alpha_j))}$$

where symbol $\sum_{=t_i}$ means that the summation is extended to all observations $(y_{i1}, y_{i2},..., y_{ip})$ such that $y_{i1} + y_{i2} + ... + y_{ip} = t_i$.

Hence, the conditional distribution $f(y|t)$ has the following form (see[1,5])

$$f(y|t_i) = f(y_{i1}, y_{i2},..., y_{ip}|T_i = t_i) = \frac{\exp(-\sum_{j=1}^{p} \alpha_j y_{ij})}{\gamma_{t_i}}$$

where $\gamma_{t_i} = \sum_{=t_i} \exp(-\sum_{j=1}^{p} \alpha_j y_{ij})$

The conditional likelihood L_C is independent of z_i (as it should be), and the estimates are determined by solving the equations (see [1,5]):

$$\frac{\partial L_c}{\partial \alpha_j} = 0.$$

For the estimation of z_i, the L_M part of the likelihood function is used.

Let us write down the function $g(t_i|z_i)$ in the equivalent form as follows (see [4]):

$$f(t_i|z_i) = \frac{\exp(z_i t_i)}{\prod_{j=1}^{p}(1+\exp((z_i-\alpha_j))} \cdot \sum_{=t_i} \exp(-\sum_{j=1}^{p}\alpha_j y_{ij})$$

Estimates of z_i are obtained by the maximization of the following likelihood function, known as marginal likelihood (see[4]):

$$L_M = L_M(\alpha_1, \alpha_2, ..., \alpha_p, z_1, z_2, ..., z_n) = \prod_{i=1}^{n} g(t_i|z_i).$$

System $\frac{\partial L_M}{\partial z_i} = 0$ is equivalent to (see [4])

$$t_i = \sum_{j=1}^{p} \frac{\exp(z_i - \alpha_j)}{1+\exp(z_i-\alpha_j)}$$

which can be solved for z_i replacing parameters α_j with their estimates $\hat{\alpha}_j$ obtained by conditional likelihood method.

4.5. Population likelihood

Suppose now that people's susceptibility to stressful phenomena is interpreted as a real valued random variable Z with a density distribution $h(z)$. Respondents are treated now as a random sample, and each z_i , $i = 1,2,...,n$ is interpreted as the realization of random variable Z_i, with distribution given by $h(z)$ or $H(z)$.

Furthermore, assume that $Z \sim N(\mu, \sigma^2)$. The problem is in the estimation of μ and σ^2. Parameters μ and σ^2 are estimated by the means of the so-called population likelihood function L_p. To derive this function let us observe that the probability density function of observed data can be expressed as follows (see [1,2,4]):

$$f(y) = \int f(y|t_i) \cdot g(t_i|z) dH(z) = \int f(y|t_i) \cdot g(t_i|z) h(z, \mu, \sigma^2) dz$$

hence, the likelihood function will have the form:

$$L(\alpha_1,\alpha_2,...,\alpha_p,\mu,\sigma^2)=\prod_{i=1}^{n} f(y_i|t_i)\prod_{i=1}^{n}\int g(t_i|z)h(z,\mu,\sigma^2)dz = L_C \cdot L_P$$

We see that it is a product of two functions L_C and L_P. For the estimation of parameters μ and σ^2 two stage procedure is applied.

First, parameters α_j , $j=1,2,...,p$ are estimated using the conditional likelihood function L_C . Then, they are treated as known, and they are used in the population likelihood function L_P for the estimation of μ and σ^2 .

It has been proven (see for example [3]) that observed score $T_i = Y_{i1} + Y_{i2} + ... + Y_{ip}$ is sufficient for (μ,σ^2). Statistics T_i takes on values $t=0,1,...,\mathrm{n}$. Let n_t be a number of response vectors with score t, i.e. n_t is a number of individuals with score t.

As respondents are sampled randomly, vector $(n_0,n_1,...,n_p)$ is to be interpreted as a realization of random vector $(N_0,N_1,...,N_p)$ which follows a multinomial distribution

$$M(n,g_0,g_1,...,g_p)$$

i.e.

$$P(N_0=n_0,N_1=n_1,...,N_p=n_p)=\frac{n!}{n_0!n_1!...n_p!}g_0^{n_0}g_1^{n_1}...g_p^{n_p}$$

Probabilities g_t are the following (see [3,4]):

$$g_t = g_t(\mu,\sigma^2)=\sum_{=t}\exp(-\sum_{j=1}^{p}\alpha_j y_{ij})\int \exp(z\cdot t)h(z,\mu,\sigma^2)\prod_{j=1}^{p}(1+\exp(z-\alpha_j))^{-1}dz$$

Population likelihood function is the following (see [1]):

$$L_p(\alpha_1,\alpha_2,...,\alpha_p,\mu,\sigma^2)=\prod_{t=0}^{p} g_t^{n_t}$$

Computational procedures for maximization of this function with the respect to μ and σ^2 are rather complicated (see [4]).

Statistics for testing goodness of fit is the following (see [3]):

$$z = 2\sum_{t=0}^{p} n_t[\ln n_t - \ln(ng_t(\hat{\mu},\hat{\sigma}^2))]$$

which is approximately χ^2 -distributed with *p-2* degrees of freedom for large *n*.

5. Feelings of morale classification

Some phenomena, particularly those of the political nature, might provoke drastically different reaction in different groups of people.

In the simplest case, the society under the investigation (respondents) could be divided into two classes. These classes could be called, for example, "content" and "malcontent", or "sensible" and "insensible". In such a dichotomized situation one can assume that the latent trait Z is a binary random variable with distribution $\eta = P(Z = 1) = p$ (respondent is content), $1-\eta = P(Z = 0) = p$ (respondent is malcontent).

Belonging to either of classes can be determined based on binary responses to items $Y_1, Y_2, ..., Y_p$.

Introducing notations

$$\pi_{j1} = P(Y_j = 1 \mid Z = 1) \quad , \qquad \pi_{j0} = P(Y_j = 1 \mid Z = 0) \quad ,$$

the basic representation for observed data

$$f(y) = \int g(y \mid z) dH(z)$$

will take the following form

$$f(y) = \eta \prod_{j=1}^{p} \pi_{j1}^{y_j} (1-\pi_{j1})^{1-y_j} + (1-\eta) \prod_{j=1}^{p} \pi_{j0}^{y_j} (1-\pi_{j0})^{1-y_j} \ .$$

In more general case we can assume that respondents belong to c classes indicated by numerals $1,2,...,c$.

Let us introduce notations:

$$w_k = P(W = k) \quad , \quad k = 1,2,...,c$$
$$\pi_{jk} = P(Y_j = 1 \mid Z = k)$$
$$g_j(y_j \mid z_k) = P(Y = y_j \mid Z = z_k) = \pi_{jk}^{y_j} (1-\pi_{jk})^{1-y_j} ,$$

then

$$f(y) = \sum_{k=1}^{c} w_k \prod_{j=1}^{p} g_j(y_j) = \sum_{k=1}^{c} w_k \cdot \prod_{j=1}^{p} \pi_{jk}^{y_j} (1-\pi_{jk})^{1-y_j} .$$

This probability distribution function is a finite mixture density and it depends on $pc + c$ parameters:

$$w_1, w_2, ..., w_c, \pi_{11}, \pi_{12}, ..., \pi_{1c}, ..., \pi_{p1}, \pi_{p2}, ..., \pi_{pc}.$$

The log-likelihood function is as follows (see [6,12,13]):

$$L = \sum_{i=1}^{n} \ln(\sum_{k=1}^{c} w_k \prod_{j=1}^{p} g_j(y_{ij} | z_k)),$$

where z_k stands for the event "$Z = k$."

Remembering that $\sum_{k=1}^{c} w_k = 1$, the estimates are derived from the equations (see [6,12]):

$$\hat{w}_k = \frac{1}{n} \sum_{i=1}^{n} \hat{P}(k | y_k^i) \quad , \quad k = 1,2,...,c$$

$$\hat{\pi}_{jk} = \frac{1}{n \cdot \hat{w}_k} \sum_{i=1}^{n} y_{ij} \cdot \hat{P}(k | y^i) \quad , \quad j = 1,2,..., p, k = 1,2,..., c$$

Symbol $\hat{P}(k | y^i)$ is the estimate of probability that the individual with response vector $y^i = (y_1^i, y_2^i, ..., y_p^i) = (y_{i1,} y_{i2}, ..., y_{ip})$ belongs to the class *k*.

Although these equations have a simple form, posterior probability has a rather complicated form:

$$\hat{P}(k | y^i) = \frac{\hat{w}_k \prod_{j=1}^{p} \hat{\pi}_{jk}^{y_j^i} (1 - \hat{\pi}_{jk})^{1-y_j^i}}{\sum_{k=1}^{c} w_k \prod_{j=1}^{p} \hat{\pi}_{jk}^{y_j^i} (1 - \hat{\pi}_{jk})^{1-y_j^i}}.$$

For the solution the E-M algorithm has to be used (see [6]).

6. Factor analysis

Let us assume now that the reaction (response) to stressor Y_j is proportional to susceptibility Z and it is additively disturbed by a random error U_j.

Denoting the proportionality coefficient by α_j, we have $Y_j = \alpha_j Z + U_j$

If $Z \sim N(0,1)$ and $U_j \sim N(0, \sigma_j)$, and they are independent, then by introducing variable ε_j with a standard normal distribution, we can write

$$Y_j = \alpha_j Z + \sigma_j \varepsilon_j$$

This means that the response is a linear combination of susceptibility and random disturbances.

Without loosing generality, one can assume that Y_j is standardized by its standard deviation, i.e. $Var(Y_j) = 1$.

From this we have

$$Var(Y_j) = Var(\alpha_j Z) + Var(\sigma_j \varepsilon_j) = \alpha_j^2 + \sigma_j^2 = 1.$$

which implies that $\sigma_j^2 = 1 - \alpha_j^2$.

Hence, we have the following representation (see also [7,8]):

$$Y_j = \alpha_j Z + \sqrt{1 - \alpha_j^2} \cdot \varepsilon_j$$

The next crucial assumption is that we are not able to observe reaction to the stressor Y_j as it escalates. We are only able to note when the intensity of a stressor reaches a level when the respondent bursts out in anger (irritation).

Denoting this level by symbol γ_j we can define the new variable Y_j^* as follows (see [10]):

$$Y_j^* = \begin{cases} 1 & if \quad Y_j \geq \gamma_j \\ 0 & if \quad Y_j < \gamma_j \end{cases}$$

The probability that a subject with the susceptibility *z* bursts out, i.e. this subject endorses item Y_j is the following

$$P(Y_j^* = 1 \Big| z) = \frac{1}{\sigma\sqrt{2\pi}} \int_{\gamma_j}^{\infty} \exp(-(x - \alpha_j z)^2 / 2\sigma^2) dx$$

Let

$$x_j = -\frac{\gamma_j - \alpha_j \cdot z}{\sigma}$$

then using Φ for cumulative distribution function of the standard normal variable, we have $P(Y_j^* = y_j \big| z) = \begin{cases} \Phi(x_j) & , \quad when \quad y_j = 1 \\ 1 - \Phi(x_j) & , \quad when \quad y_j = 0 \end{cases}$

The probability of the response vector $y = (y_1, y_2, ..., y_p)$ is given by

$$\prod_{j=1}^{p} P(Y_j^* = y_j | z)$$

Let us observe that all response vectors are divided into 2^p mutually exclusive categories (patterns). These categories are conveniently indexed by the decimal numbers:

$$k = 1 + \textit{binary number } y_{k1}y_{k2}...y_{kp} \quad , \quad k = 1,2,...,2^p$$

If P_k denotes the probability of *k-th* category, then we have (see [8])

$$P_k = \int_{-\infty}^{\infty} \prod_{j=1}^{p} P(Y_1^* = y_{k1}, Y_2^* = y_{k2}, ..., Y_p^* = y_{kp} | z) \varphi(z) dz$$

The definite integral in this formula cannot be expressed in the closed form, in [7] it is evaluated by Gauss-Hermite quadrature.

Remember that P_k determines the probability of the k-th category of response vector. This probability depends on threshold values $\gamma_1, \gamma_2, ...\gamma_p$ and correlations $\alpha_1, \alpha_2, ...\alpha_p$, which in turn are interpreted as the strength of stressors $Y_1, Y_2, ...Y_p$. The numbers N_k are multinomially distributed with the parameters n and P_k.

Therefore, the likelihood function

$$L = \frac{n!}{\prod_k n_k!} \prod_{k=1}^{2^p} P_k^{n_k}$$

is maximized with respect to α_j and γ_j using Newton-Raphson method.

Christofferson proposed simpler method of the generalized least squares using only first-order and second-order marginal proportions.

7. Stressor scaling by paired comparisons

Definition of models based on paired comparisons requires different assumptions (see[11]). First of all, one has to assume that each stressor can be located on a sensation scale. To determine this location, stressors are presented to individuals for judgement. They are presented in pairs, so that an individual can evaluate which stressor is more intensive.

Let us assume that stressors $Y_1, Y_2, ..., Y_p$ have intrinsic force to cause distress in the population, or to irritate people. As a result, a disruption of social

ties can occur. The strength of the negative impact of stressor Y_j on social cohesion is denoted as α_j, $j = 1,2,...,p$.

It is assumed that there is a probability space on which following random variables Y_{ij} are defined:

$$Y_{ij} = Y_i - Y_j .$$

and interpreted as the amount of predominance of Y_i over Y_j.

Quantities α_j are then treated as positional parameters of Y_j.

Assuming that ε_{ij} is a random variable having density $f(u)$ symmetric about 0, we arrive at the model

$$Y_{ij} = \alpha_i - \alpha_j + \varepsilon_{ij} .$$

Let π_{ij} denote the probability of the predominance of Y_i over Y_j:

$$\pi_{ij} = P(Y_i \prec Y_j) = P(Y_i - Y_j > 0) .$$

Because of symmetry of $f(u)$ the above formula will have the following form:

$$\pi_{ij} = P(Y_{ij} > 0) = P(\alpha_i - \alpha_j + \varepsilon_{ij} > 0)$$

$$= 1 - P(\alpha_i - \alpha_j + \varepsilon_{ij} \le 0) = F(\alpha_i - \alpha_j) = F(\delta_{ij})$$

where $F(u)$ is the cumulative distribution function (see[25,26]).

In the simplest case, the function of the uniform distribution over the interval $(-0{,}5 \;\; 0{,}5)$ can be used:

$$H(u) = \frac{1}{2} + u, \quad -0.5 \le u \le 0.5 .$$

However, either normal or logistic model is usually considered.

In the first case we have

$$\pi_{ij} = \frac{1}{\sqrt{2\pi}} \int_{-(\alpha_i - \alpha_j)}^{\infty} e^{-\frac{1}{2}z^2} dz$$

and in the second case we have (see[6,9]):

$$\pi_{ij} = \int_{-(\alpha_i - \alpha_j)}^{\infty} \frac{e^z}{(1+e^z)^2} dz = \frac{1}{1+e^{\alpha_j - \alpha_i}}$$

This model is known as Bradley-Terry model.

Positional parameters $\alpha_1, \alpha_2, ..., \alpha_p$ determine the scale values for stressors $Y_1, Y_2, ..., Y_p$.

With no loss of generality it may be assumed that

$$\sum_{i=1}^{p} \alpha_i = 0.$$

This equality guaranties the uniqueness of estimates $\hat{\alpha}_j$, $j = 1,2,..., p$.

For the estimation of these parameters paired comparisons are used.

If by n_{ij} we denote the number of individuals (respondents, experts, judges) who considered stressor Y_i more stressful than Y_j, then we can conclude that n_{ij} is a binomial variable

$$n_{ij} \sim B(n, \pi_{ij})$$

where n is the number of experts, and π_{ij} is the probability of "success".

It is not difficult to infer that

$$\alpha_j = \frac{1}{n} \sum_{k=1}^{n} \delta_{kj}$$

This equality suggests the method of estimation:

$$\hat{\alpha}_j = \frac{1}{n} \sum_{k=1}^{n} d_{kj}$$

where values of d_{kj} are to be obtained from the system of equations:

$$F(d_{ij}) = \frac{n_{ij}}{n}, \quad i,j = 1,2,...,p$$

The ratio $\frac{n_{ij}}{n}$ is the maximum likelihood estimate $\hat{\pi}_{ij}$ of π_{ij}.

If $n_{ij} = 0,$ than it is suggested to put $\hat{p}_{ij} = \frac{1}{2n}$, and in the case when $n_{ij} = 1$, to put $\hat{p}_{ij} = 1 - \frac{1}{2n}$.

Mosteller proved that estimates $\hat{\alpha}_j$ are least squares estimates (see[25,26]).

For testing the hypothesis that there is no difference between all stressors

$$\alpha_1 = \alpha_2 = ... = \alpha_p$$

one can use the following statistics (see[26]):

$$A_f = 4npf^2(0)\sum_{j=1}^{p}\hat{\alpha}_j^2$$

where the subscript *f* indicates that $\hat{\alpha}_j$ have been computed using the function $f(u)$.

Under the null hypothesis, A_f has an asymptotic chi square distribution with *p-1* degrees of freedom as number of sampled respondents tends to infinity.

8. Conclusions

From the discussion presented in this paper it follows that the statistical methods developed in different fields of psychology, education and bioassay can be easily adopted for modelling of the social phenomena. Particularly, the methods of item response theory can be directly used for social stressors analysis. Merely little changes in the interpretation of parameters are needed.

REFERENCES

1. ANDERSEN E.B. (1980): Discrete Statistical Models with Social Science Applications, North-Holland, Amsterdam.
2. ANDERSEN E.B., Latent trait models , Journal of econometrics, 22, 1983, 215–227
3. ANDERSEN E.B., Comparing latent distributions, Psychometrika, 45, 1980, 121–134
4. ANDERSEN E.B., MADSEN M., Estimating the parameters of the latent population distribution, *Psychometrika* 42, 1977, 357–374

5. ANDRICH D., Rasch models for measurement, SAGE University Paper, 1988
6. BARTHOLOMEW D.J., KNOTT M., Latent variable models and factor analysis, ARNOLD, London, 1999
7. BOCK R.D., AITKIN M. (1981): Marginal maximum likelihood estimation of item parameters: application of an EM algorithm, *Psychometrika* 46, pp. 443–459
8. BOCK R.D., LIEBERMAN M.(1970): Fitting a response model for n dichotomously scored items, *Psychometrika* 435(2), pp. 179–197.
9. BRUNK H. D. (1960), Mathematical models for ranking from paired comparisons, American Statistical Association Journal, 9, 503–21
10. CHRISTOFFERSON A. (1975): Factor analysis of dichotomised variables, *Psychometrika* 40(1), pp. 5–31.
11. DAVID H.A., The method of paired comparisons, Griffin, London, 1969
12. EVERITT B.S.(1984): An introduction to latent variable models, Chapman& Hall, London.
13. EVERITT B.S., HAND D.J.(1981): Finite mixture distributions, Chapman& Hall, London.
14. FISCHER G.H., MOLENAAR I.W. eds (1995): Rasch models: foundations, recent developments and applications, Springer-Verlag, New York.
15. HAMBLETON R.K, SWAMINATHAN H., ROGERS H.J. (1991): Fundamentals of itemresponse theory, Sage Publications, Newbury Park, CA.
16. HOLLAND P.W. (1981): When are item response models consistent with observe data?, *Psychometrika* 46, pp. 79–92.
17. HOLLAND P.W. (1990): On the sampling theory foundations of item response theory models, *Psychometrika 55,* pp. 577–601.
18. HOLLAND P.W., ROSENBAUM P.R. (1986): Conditional association and unidimensionality in monotone latent variable models, Annals of Statistics, vol. 14, pp. 1523–1543
19. JUNKER B.W. (1993): Conditional association, essential independence and monotone unidimensional item response models, *Annals of Statistics* 21, pp. 1359–1378.
20. JUNKER B.W., ELLIS J.L. (1997): A characterization of monotone unidimensional latent variable models, Annals of Statistics 25, pp. 1327–1343.

21. JUNKER B.W., SIJTSMA K., Nonparametric Item Response Theory in action, Applied Psychological Measurement,25, 2001, 211–220

22. KRAUTH J., Testkonstruktion und Testtheorie, BELTZ, 1995

23. LAZARSFELD P.F., HENRY N.W. (1968): Latent structure Analysis, Houghton-Mifflin, New York.

24. MOKKEN R.J. (1997): Nonparametric models for dichotomous responses, in W. van der Linden and R.K. Hambleton, eds, Handbook of Modern Item Response Theory, Springer-Verlag, New York pp. 351–367.

25. MOSTELLER F. (1951), Remarks on the method of paired comparisons, I, *Psychometrika* 16, pp. 3–9

26. NOETHER G., (1960), Remarks about a paired comparison model, *Psychometrika 25*, pp. 357–367

27. RASCH G. (1960): Probabilistic models for some intelligence and attainment tests, Pœdagogiske Institut, Copenhagen.

28. ROSENBAUM P. R.(1984), Testing the conditional independence and monotonicity assumptions of item response theory, *Psychometrika* , 49, 425–435.

29. STOUT W. (1987): A nonparametric approach for assessing latent trait unidimensionality, *Psychometrika* 52, pp. 589–617.

30. STOUT W. (1990): A new item response theory modelling approach with applications to unidimensionality assessment and ability estimation, *Psychometrika* 55, pp. 293–325.