Methods of Psychological Research Online 2003, Vol.8, No.1, pp. 43-79
Internet: http://www.mpr-online.de Department of Psychology

A Method for the Analysis of Hierarchical Dependencies between Items of a Questionnaire

Martin Schrepp[1]

This paper describes a method of explorative data analysis which allows to detect logical implications between items of a dichotomous questionnaire or test. These logical dependencies are organized to form a hierarchical structure (quasi-order) on the items. Our analysis method, which is called *Inductive Item Tree Analysis*, can be seen as a method of Boolean analysis. We discuss the relation of our method to other methods of Boolean analysis and to related methods of data analysis, like for example Guttman scaling and latent class analysis. The adequacy of our analysis method is tested in a simulation study. The results of this study show that the method is able to detect existing dependencies with high accuracy if enough data are available. We apply our method to some real data sets to demonstrate the advantages of an analysis of logical implications.

Keywords: *Item analysis statistical, test-items, mathematical modeling*

We describe a method which allows us to derive a hierarchical structure on the items of a questionnaire from observed response patterns. Assume that we have a questionnaire I with m items $\{i_1, \ldots, i_m\}$ and that a subject can agree (1) or disagree (0) with each of these items, i.e., the items are dichotomous. If n subjects respond to the questions in I this will result in a binary data matrix D with m columns and n rows[2].

Depending on the content of the items it is possible that the response of a subject to an item j determines her or his responses to other items. It is, for example, possible that each subject who agrees to item j will also agree to item i. In this case we say that *item j implies item i* and write shortly $j \rightarrow i$. The goal of our analysis method is to uncover such deterministic implications from the data set D.

---

[1] Address: e-mail: MartinSchrepp@aol.com

[2] We describe our method in the context of the analysis of dichotomous questionnaires. But it is clear that the same method can be used to analyze the logical dependencies between items from a test, where each item (problem) can be solved (1) or failed (0) by a subject.

Our method can be classified as a method of *Boolean analysis of questionnaires*. Boolean analysis was introduced by Flament (1976). The goal of a Boolean analysis is to detect deterministic dependencies between the items of a questionnaire in observed response patterns. These deterministic dependencies have the form of logical formulas connecting the items. Assume, for example, that a questionnaire contains items i, j, and k. Examples of such deterministic dependencies are then $i \to j$, $i \wedge j \to k$, and $i \vee j \to k$.

Since the basic work of Flament (1976) a number of different methods for Boolean analysis have been developed. See, for example, Buggenhaut and Degreef (1987), Duquenne (1987), Leeuwe (1974), Schrepp (1997, 1999), or Theuns (1994, 1998). These methods share the goal to derive deterministic dependencies between the items of a questionnaire from data, but differ in the algorithms to reach this goal.

Our method shares some features with *Item Tree Analysis* (Leeuwe, 1974). We call our method in the following *Inductive Item Tree Analysis* (short IITA). A comparison to Item Tree Analysis and an explanation for the chosen name will be given in the detailed description of our analysis method.

The investigation of deterministic dependencies has some tradition in educational psychology. The items represent in this area usually skills or cognitive abilities of subjects. Bart and Airasian (1974) use Boolean analysis to establish logical implications on a set of Piagetian tasks. Other examples in this tradition are the learning hierarchies of Gagné (1968) or the theory of structural learning of Scandura (1971).

Another example for the use of deterministic dependencies in psychology are approaches to formalize the diagnostic process of psychologists. The goal of this approach is to uncover the rules on which the decisions of diagnosticians are based. See Härtner, Mattes and Wottawa (1980) or Wottawa and Echterhoff (1982) for details.

A recent application of Boolean analysis can be found in Held and Korossy (1998) who analyze implications on a set of algebra problems. In this paper item tree analysis (Leeuwe, 1974) is used to extract logical implications from observed response patterns. The extracted implications are then compared to implications obtained by querying an expert.

Methods of Boolean analysis are used in a number of social science studies to get insight into the structure of dichotomous data. Bart and Krus (1973) use, for example, Boolean analysis to establish a hierarchical order on items that describe socially unaccepted behavior. Janssens (1999) used a method of Boolean analysis to investigate the integration process of minorities into the value system of the dominant culture.

Boolean analysis is an explorative method to detect deterministic dependencies between items. The detected dependencies must be confirmed in subsequent research. Methods to test such deterministic dependencies statistically are described, for example, in von Eye (1991).

Methods of Boolean analysis do not assume that the detected dependencies describe the data completely. There may be other probabilistic dependencies as well. Thus, a Boolean analysis tries to detect interesting deterministic structures in the data, but has not the goal to uncover all structural aspects in the data set. Therefore, it makes sense to use other methods, like for example latent class analysis, together with a Boolean analysis.

We explain this point with an example. Assume that we have 3 items i_1, i_2, and i_3. Assume further that we have detected the implication $i_1 \rightarrow i_2$, i.e., this implication is with the exception of a small number of counterexamples (which are attributed to random errors) true for all rows in the data set. Thus, we have an implication between i_1 and i_2, but the dependencies between i_1 and i_3 respectively i_2 and i_3 are not specified at all. Thus, it is for example possible that we have the probabilistic dependency *A subject who gives an answer in category 1 in item i_1 will answer with a probability of .7 to category 0 in item i_3*. Such dependencies are not in focus of a Boolean analysis, i.e., they are ignored by these analysis methods.

Deterministic dependencies, like for example $j \rightarrow i$, impose very strict conditions on the data. Thus, we can usually not assume to find too many of them in empirical data. It is also a reasonable outcome of a Boolean analysis that no deterministic dependencies at all are detected in a data set. This does not mean that the data set has no structure. Other methods, like for example latent class analysis or configural frequency analysis, may uncover in such cases structural aspects of a probabilistic nature.

Boolean analysis has some relations to other research areas. There is a close connection between Boolean analysis and *knowledge space theory*. The theory of knowledge spaces (see Doignon & Falmagne, 1998 or Albert & Lukas, 1999) provides a theoretical framework for the formal description of human knowledge. A knowledge domain is in this approach represented by a set Q of problems. The knowledge of a subject in the domain is then described by the subset of problems from Q he or she is able to solve. This set is called the *knowledge state* of the subject. Because of dependencies between the items (for example, if solving item j implies solving item i) not all elements of the power set of Q will, in general, be possible knowledge states.

The set of all possible knowledge states is called the *knowledge structure*. It can be used for the efficient diagnosis of knowledge (Doignon & Falmagne, 1985) or for the empirical evaluation of psychological models of problem solving behavior in the domain (see for example Albert, Schrepp & Held, 1994 or Schrepp, 1995).

Methods of Boolean analysis can be used to construct a knowledge structure from data (see Theuns, 1998 or Schrepp, 1999). The main difference between both research areas is that Boolean analysis concentrates on the extraction of structures from data while knowledge space theory focus on the structural properties of the relation between a knowledge structure and the logical formulas which describe it.

Closely related to knowledge space theory is *formal concept analysis*. See Ganter and Wille (1996) for an overview of this research area. Similar to knowledge space theory this approach concentrates on the formal description and visualization of existing dependencies. In contrast Boolean analysis offers a way to construct such dependencies from data.

Another related field is *data mining* (for an overview see Nakhaeizadeh, 1998). Data mining deals with the extraction of knowledge from large databases. Several algorithms (see for example Klementinnen, Mannila, Ronkainen, Toivonen & Verkamo, 1994 or Toivonen, 1996) are developed in this area which extract dependencies of the form $j \to i$ (called *association rules*) from the database.

The main difference between Boolean analysis and the extraction of association rules in data mining is the interpretation of the extracted implications. The goal of a Boolean analysis is to extract implications from the data which are (with the exception of random errors in the response behavior) true for all rows in the data set. For data mining applications it is sufficient to detect implications which fulfill a predefined level of accuracy. It is, for example in a marketing scenario, of interest to find implications which are true for more than x% of the rows in the data set. An online bookshop may be interested, for example, to search for implications of the form *If a customer orders book A he also orders book B* if they are fulfilled by more than 10% of the available customer data.

A method closely related to data mining techniques is the *GUHA method* (Hájek, Havel & Chytil, 1966 or Hájek & Havránek, 1977, 1978). The basic idea of this method is to use formal logic to generate all hypotheses which are of interest in a given research task and supported by the data. Statistical methods are used to evaluate these generated hypotheses.

We will discuss the connection of IITA to scaling techniques, latent class analysis, and configural frequency analysis in detail on the example of some data sets. The connections of IITA to the rule space approach (Tatsuoka, 1983), the unified model (DiBello, Stout & Roussos, 1995), and the competency-performance theory (Korossy, 1996) will be discussed in the detailed description of our analysis method.

IITA tries to detect implications, i.e., dependencies of the form $j \rightarrow i$, from the data set. What benefits can we gain from the study of such implications between items of a questionnaire?

An important possibility to apply the knowledge of existing implications between items is adaptive testing. Suppose we present the items on a computer screen and not in a paper-pencil form. Assume that item j implies item i. If the program presents item j to a subject and the subject agrees to the item, then we can conclude from $j \rightarrow i$ that the subject will also agree to item i. This item needs therefore not to be presented to the subject. Assume conversely that the program presents item i first and that the subject disagrees with that item. Then it follows from $j \rightarrow i$ that the subject will also disagree to item j. Thus, item j needs not to be presented to the subject. If we detect many implications between the items of a questionnaire this can lead to a severe reduction of the time a subject needs to complete the questionnaire.

Another benefit from the study of implications is that they allow us to get information about the beliefs and cognitive processes which cause the answer behavior of subjects.

Assume again that we detect that item j implies item i. In some cases this will be a trivial fact because item j is just a stricter wording of the statement described in item i. Thus, the implication follows from the wording of the items. But if the implication does not result from the wording of the items we have to find out what are the common opinions (remember that $j \rightarrow i$ means that each subject who agrees to item j will also agree to item i) that cause this implication. Why do subjects connect the statements?

Thus, the study of implications between the items may lead to some insights about the reasons for the answer behavior and about the common opinions shared by all investigated subjects. We will discuss these points in more detail on the example of some data sets.

Description of Inductive Item Tree Analysis

Let $I = \{ i_1, ..., i_m\}$ be a set of m dichotomous items. Assume that $D = \{ d_1 , ..., d_n\}$ is a set of observed response patterns of n subjects who answered to the items in I. Every response pattern d can be considered as a mapping $d : I \to \{0,1\}$ which assigns the response $d(i)$ of a subject to each of the items in I.

IITA constructs a number of implications $j \to i$ from D. Since we interpret $j \to i$ as a logical implication the constructed implications should be reflexive ($i \to i$ for each item i) and transitive ($i \to j$ and $j \to k$ implies $i \to k$ for all items i,j,k). In other words the implications must form a *quasi-order*[3] on the item set I.

A response pattern d of a subject is called a *counterexample for an implication* $j \to i$ if the subject agrees to j and disagrees with i. A simple approach to detect all valid implications is to accept only those implications $j \to i$ for which no counterexample is observed in the data set D. Such an approach is clearly insufficient, since we have to consider random errors in the data. Such random errors can result from various sources.

A subject may, for example, be demotivated to work on the questionnaire and therefore answering some of the items inconsistent to her or his true opinion. Another possibility is that the subject misunderstands some items due to a lack of concentration. It is also possible that the subject marks by error the wrong answer category on the answer form, for example due to time pressure or lack in concentration.

We describe now our basic assumptions on the data. Therefore, we introduce the distinction between the cognitive state of a subject and her or his answers given in the questionnaire. The *response pattern* of the subject is a mapping d: $I \to \{0,1\}$ which assigns the observed response of the subject to each item. Some of these observed responses are caused by random influences, for example lack in concentration or decreasing motivation to answer the questions. Thus, the true opinion of the subject can differ from her or his response pattern. This true opinion can also be considered as a mapping s: $I \to \{0, 1\}$ and is called in the following the *latent pattern* of the subject.

[3] A quasi-order $\leq$ on I can be interpreted as a set $\{ (i,j) \mid i \leq j \}$ of item pairs. To denote that two items are connected by this quasi-order $\leq$ we use the notations $i \leq j$ and $(i,j) \in \leq$ simultaneously.

We assume that there is a quasi-order $\leq$ which describes all true implications between items and write $i \leq j$ for $j \rightarrow i$. This quasi-order determines the set S of all possible latent patterns by $S = \{s: I \rightarrow \{0,1\} \mid i \leq j \wedge s(j) = 1 \rightarrow s(i) = 1 \}$. This definition of S is equivalent with the definition of the quasi-ordinal knowledge space corresponding to $\leq$ in knowledge space theory (see Doignon & Falmagne, 1985). Interindividual differences are thus described in our approach by different latent patterns in S.

We introduced the distinction between latent pattern and response pattern to clarify our assumptions concerning random influences on the data. Please note that we do not make any assumptions on the structure of the underlying cognitive states, like for example the *rule space* approach (Tatsuoka, 1983 or Tatsuoka & Tatsuoka, 1987) or the *unified model* (DiBello et al., 1995). These theories establish a link between deterministic aspects of cognition and item response theory. They assume for a given set I of problems a list of k cognitive attributes or skills which are connected to the items by a k x n attribute-by-item matrix Q. Thus, the cognitive state of a subject can be represented by a k-tuple $a = (a_1,\ldots, a_k)$, where $a_i = 1$ if the subject has the cognitive attribute i and $a_i = 0$ otherwise. The response pattern of the subject is represented by a n-tuple $r = (r_1, \ldots, r_n)$, where $r_i = 1$ if the subject solved problem i and $r_i = 0$ otherwise. For a given cognitive state a it is possible to predict a response pattern r by the assumption that $r_i = 1$ if the cognitive state of the subject contains all the skills necessary to solve problem i, i.e., by the information contained in Q. With some additional assumptions such approaches allow to draw inferences from the observed response patterns of subjects to their cognitive states.

Another similar approach which we have to mention here is the *competency-performance theory* (Korossy, 1996). This theory describes the connection between the set of skills of a subject (the cognitive state of the subject) and the observed response pattern of the subject in the context of knowledge space theory.

It is necessary to make some assumptions on the nature of the cognitive states to apply these techniques to data analysis. In contrast our analysis method is purely explorative. It does not require any assumptions concerning underlying cognitive attributes.

We assume that the data set D is generated from S under the influence of random errors. We assume further that the probability of a random error does not depend on the item[4], i.e., is the same for all items. The goal of our analysis is to uncover $\leq$ from D.

[4] The assumption that the probability of a random error does not depend on the item is a necessary as-

Let $b_{ij} := \left|\{d \in D \mid d(j) = 1 \wedge d(i) = 0\}\right|$ be the number of observed counterexamples for the implication $j \rightarrow i$. We define a binary relation $\leq_0$ on I by:

$$i \leq_0 j \Leftrightarrow b_{ij} = 0\,. \tag{1}$$

This relation $\leq_0$ consists of all the implications which are not violated empirically in the data set D. Leeuwe (1974) showed that this relation is a quasi-order on the item set I.

IITA consists of two steps. In the first step, we construct a set $\{\leq_L \mid L = 0,1, 2, ..., n\}$ of quasi-orders on the item set I. In the second step, we calculate the fit of each constructed quasi-order $\leq_L$ to the data set D. We chose then the quasi-order which fits best to the data.

Construction of the Quasi-Orders $\leq_L$

We construct the relations $\leq_L$ inductively, starting with the quasi-order $\leq_0$.

Assume that we have constructed in step L of the construction process a quasi-order $\leq_L$. In step L+1 we add all item pairs to this quasi-order which have not more than L+1 counterexamples in the data set (i.e., $b_{ij} \leq L+1$) and do not cause an intransitivity[5] to item pairs already contained in $\leq_L$ or added in this step to $\leq_L$. The procedure to construct $\leq_{L+1}$ from $\leq_L$ consists of three steps.

First, we determine the set $A_{L+1} = \{\ (i,j) \mid b_{ij} \leq L+1 \wedge (i,j) \notin \leq_L\ \}$ of item pairs which have not more than L+1 counterexamples in the data set and are not contained in $\leq_L$.

Second, we repeat the two following operations until the set B_{L+1} is empty:

- We check for each element of A_{L+1} if it causes an intransitivity to other elements of $\leq_L \cup A_{L+1}$. Elements which cause such an intransitivity are marked. Let B_{L+1} be the set of all marked elements.
- We delete all marked elements from A_{L+1}.

---

sumption. If this assumption is heavily violated, then our analysis method should not be used. Assume, for example, that we have a set of mathematical problems and that some of these problems are open problems and some others are multiple choice problems with 4 answer alternatives. For the multiple choice problems the probability that a subject, who is not able to find the correct solution, marks by chance the correct answer alternative is around 25%. For the open problems the chance that a subject finds the correct solution by chance is close to 0%. Thus, our assumption on the error probabilities is heavily violated in this case and our analysis method should not be used for this set of items.

[5] It is not possible to define the relation $\leq_L$ simply by $i \leq_L j \Leftrightarrow b_{ij} \leq L$, since this relation is for $L > 0$ not always transitive (Leeuwe, 1974).

When this process ends none of the remaining implications in A_{L+1} causes an intransitivity to other elements in $\leq_L \cup A_{L+1}$.

Thus, in the third step we define $\leq_{L+1} = \leq_L \cup A_{L+1}$. The relation $\leq_{L+1}$ is by construction transitive.

This construction method results in a set $\{ \leq_L \mid L = 0,1, 2, ..., n \}$ of quasi-orders. It is clear from the construction method that $\leq_0 \subseteq \leq_1 \subseteq \leq_2 \subseteq ... \subseteq \leq_n$. Note that some of these quasi-orders may be identical. If, for example, $A_{L+1} = \emptyset$, then $\leq_L = \leq_{L+1}$. See Schrepp (1999) for a detailed description of the construction method.

A Method to Determine the Best Fitting Quasi-Order

Our inductive construction method results in a set of quasi-orders on *I*. Which of these quasi-orders describes the data best? To determine this optimal quasi-order we compare each $\leq_L$ to the data set *D*. Then we chose the quasi-order $\leq_L$ which fits the data relatively best among the constructed quasi-orders $\{ \leq_L \mid L = 0,1, 2, ..., n \}$.

Assume for a level *L* that $\leq_L$ is the correct quasi-order underlying the data set *D*. How many counterexamples for an implication $j \to i$ must we expect under this assumption?

Define for each item *i* in *I*:

$$p_i = \frac{\left|\{d \in D \mid d(i) = 1\}\right|}{n} \tag{2}$$

The number p_i is the relative frequency of subjects in the data set, who agreed to item *i*.

Our basic assumption on the data is that there is a (maybe empty) set of true logical implications between the items. These implications are true for all response patterns of subjects. A violation of such an implication is only possible by the influence of random errors. Let γ be the probability that a true implication is violated due to random errors. Under our assumption that $\leq_L$ is the correct quasi-order we are able to estimate the error probability γ from the observed b_{ij}-values. This can be done by:

$$\gamma = \frac{\sum \{b_{ij} / p_j * n \mid i \leq_L j \wedge i \neq j\}}{(\left|\leq_L\right| - m)} \tag{3}$$

In this formula $b_{ij}/p_j{*}n$ is the number of observed counterexamples to the implication $j \to i$ relative to the number of cases in which such a counterexample is possible, i.e., item j is answered with 1. The value $|\leq_L| - m$ is the number of non-reflexive implications in $\leq_L$.

Thus, γ represents the amount of random errors in the data under the assumption that $\leq_L$ is the correct quasi-order.

To determine how many counterexamples we have to expect for an implication $j \to i$ we have to distinguish two cases:

1. Assume $(i,j) \notin \leq_L$. Since there is no information available about the dependencies of i and j, we assume that these items are independent. Thus, the expected number t_{ij} of counterexamples to $j \to i$ (under the assumption that $\leq_L$ is the correct quasi-order) is equal to the expected value of the number of response patterns with $d(i) = 0$ and $d(j) = 1$. Therefore, $t_{ij} = (1\text{-}p_i) * p_j * n * (1\text{-}\gamma)$. Here $(1\text{-}p_i) * p_j * n$ is the expected number of response patterns with $d(i) = 0$ and $d(j) = 1$ and $(1\text{-}\gamma)$ is the probability that no random error has occurred.
2. Assume $(i,j) \in \leq_L$ and $i \neq j$. In this case (since we assume that $\leq_L$ is the correct quasi-order) all counterexamples to $j \to i$ must result from random errors. Thus, the expected number of counterexamples to $j \to i$ should be $t_{ij} = \gamma * p_j * n$. The value t_{ij} is the number of data patterns $p_j * n$ in which a subject agreed to item j (since only in this case a counterexample can occur) multiplied with the error probability γ.

We can now derive a measure for the fit between $\leq_L$ and the data set D by:

$$\mathit{diff}(\leq_L, D) = \frac{\sum_{i \neq j} (b_{ij} - t_{ij})^2}{(m^2 - m)} \tag{4}$$

The value $\mathit{diff}(\leq_L,D)$ is the mean quadratic difference between the observed counterexamples to $j \to i$ and the expected number of counterexamples under the assumption that $\leq_L$ is the correct quasi-order.

To determine the optimal tolerance level L we calculate the $\mathit{diff}(\leq_L,D)$ value for all levels $L = 0,1,\ldots,n$ and chose the level for which this value is minimal.

Note that our assumption concerning the stochastic independence of i and j in the case of $(i,j) \notin \leq_L$ may be wrong. But we use the t_{ij} only to compare the quasi-orders $\leq_L$. If the estimation of t_{ij} is wrong, this does influence all levels L in the same way. Thus, the quasi-order $\leq_L$ closest (we can compare two quasi-orders by their symmetric distance, see the description of our simulation study for a detailed definition) to the correct quasi-order should show the best *diff*-value.

The described method to compare the constructed quasi-orders $\{\leq_L \mid L = 0,1, 2, ..., n\}$ can also be used outside the context of explorative data analysis. In knowledge space theory quasi-orders on problem sets are sometimes derived from models of the underlying cognitive problem solving processes (see Albert, Schrepp & Held, 1994). In this case we can distinguish competing models of problem solving by their derived quasi-orders. It is then possible to use our method to determine which of these derived quasi-orders (and thus which of the investigated models of problem solving) fits best to the data.

Our algorithm shares some features with *Item tree analysis* (Leeuwe, 1974). A comparison of our algorithm to item tree analysis can be found in Schrepp (1997, 1999). In these papers both algorithms are compared concerning their ability to reconstruct the correct implications from simulated data. The results of these simulations (the simulation procedure in these papers is similar to the procedure we describe in the next section) show that IITA produces better results than ITA. Since the main advantage of the described algorithm compared to ITA is the inductive construction of the quasi-orders $\leq_L$ we have chosen the name *Inductive Item Tree Analysis.*

An Example

Assume that $I = \{a, b, c, d, e\}$ is a questionnaire with 5 items. Assume that the implications $a \to b$, $a \to c$, $a \to d$, $a \to e$, $b \to c$, $b \to d$, $b \to e$, $c \to e$, and $d \to e$ are true for this questionnaire. The following Hasse-diagram depicts these implications.

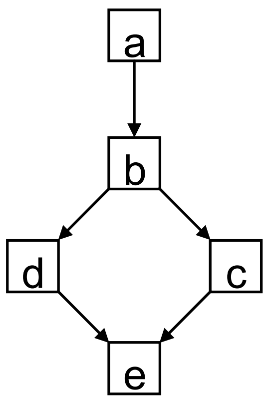

Figure 1. Hasse-diagram of the assumed implications.

The set S of all patterns consistent with these implications is given by:

$S = \{(1,1,1,1,1), (0,1,1,1,1), (0,0,1,1,1), (0,0,1,0,1), (0,0,0,1,1), (0,0,0,0,1), (0,0,0,0,0)\}$

If no errors are possible, then each subject must show a response pattern from S. We simulate the response behavior of 200 subjects to the items in I accordingly to our error model. Therefore, we randomly select 200 times a latent pattern s from S. For each item i we change $s(i)$ with probability .05 from 1 to 0 respectively from 0 to 1. The result is a simulated response pattern d. Thus, the simulated response patterns result from the latent patterns by applying an error rate of .05.

Table 1 shows the resulting simulated response patterns and their frequencies. Patterns with frequency 0 are not shown.

Table 1

Simulated Response Patterns and their Frequencies.

pattern	freq.	pattern (cont.)	freq. (cont.)	pattern (cont.)	freq. (cont.)
00101	30	10111	5	11110	2
00011	28	00100	4	01110	2
01111	27	11011	4	01000	1
11111	24	01101	3	01011	1
00001	24	10011	3	11101	1
00111	21	10001	2	00010	1
00000	14	01001	2	00110	1

We try now to reconstruct the quasi-order underlying the simulated data with the help of IITA. Table 2 shows the observed b_{ij}-values for the data set in Table 1.

Table 2

The Observed b_{ij}-Values for the Simulated Data Set from Table 1.

		j				
		a	b	c	d	e
	a	0	36	88	81	136
	b	10	0	61	59	113
i	c	9	8	0	37	64
	d	3	7	38	0	62
	e	2	5	9	6	0

We use our inductive method to construct the set $\{\leq_L \mid L = 0, ..., 200\}$ of quasi-orders on *I*. The resulting quasi-orders (we show only those quasi-orders which are not identical to a quasi-order with a higher tolerance level) are depicted in Figure 2 as Hasse-diagrams.

The value of $diff(\leq_L, D)$ is minimal for L=10. Thus, $\leq_{10}$ is the optimal solution for the simulated data set. Therefore, IITA detected the quasi-order underlying the data successfully. But this is of course not always possible.

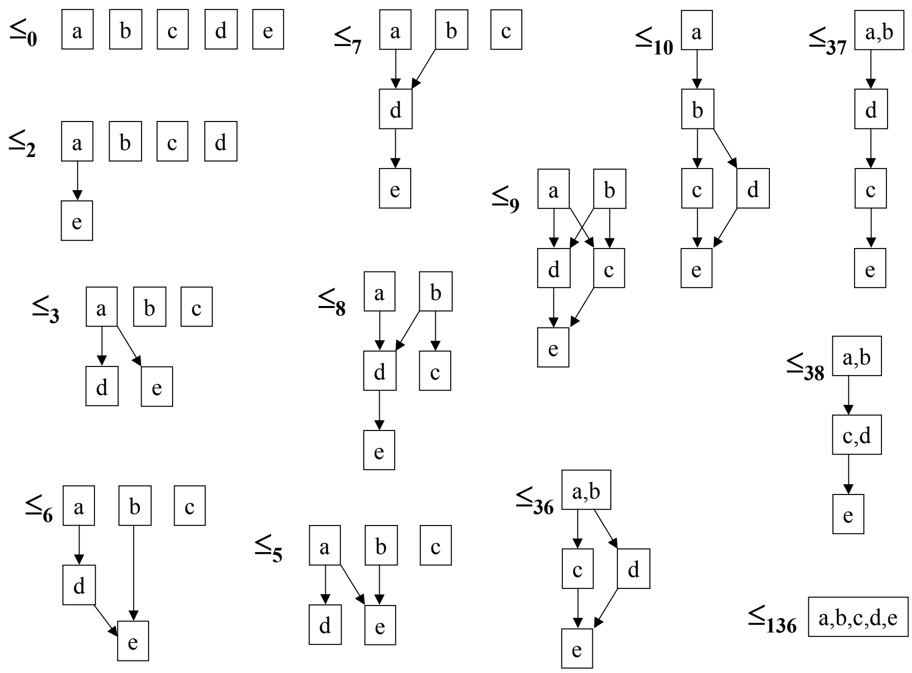

Figure 2. Hasse-diagrams of the constructed quasi-orders on I.

The Reproducibility Coefficient

As we have already mentioned a Boolean analysis has, in general, not the goal to explain a data set *D* completely by the generated deterministic dependencies, i.e., the derived implications in ≤. But to control the quality of the constructed quasi-order we need an estimation of how much of the structure in the data can be explained by these dependencies.

As we have seen the derived deterministic implications determine a set *S* of latent patterns which are compatible with these dependencies. A simple approach to measure the fit between *S* (respectively ≤) and *D* is to compute:

$$fit(D,S) = \frac{|\{d \in D \mid d \in S\}|}{|D|} \tag{5}$$

i.e., the relative frequency of observed data patterns compatible with the implications in $\leq$.

But this approach is not adequate. Assume, for example, that we have a linear order $i_1 \rightarrow i_2 \rightarrow \ldots \rightarrow i_{10}$ on a set of 10 items. Assume further a probability of .1 that the value $s(i)$ of an item in the latent pattern of a subject is not equal to the value $d(i)$ in the response pattern of the subject. The probability that an observed response pattern is an element of S is then approximately $0.9^{10} \approx .35$. Thus, even if the linear order fits perfect to the latent structure we can expect only 35% of the response patterns to be compatible with it.

If we have a linear order on a set of 20 items this value decreases to $.9^{20} \approx .12$. Thus, even a small value of $fit(D,S)$ does not indicate that the derived dependencies are not in accordance with the data.

A measure which is able to overcome the described problem is the *reproducibility coefficient* (Guttman, 1944). Let $dist(r,s) = |\{i \in I \mid r(i) \neq s(i)\}|$ be the distance between two response patterns. The value of $dist(r,s)$ is the number of items in which r and s differ. We calculate for every response pattern r the minimal distance $mdist(r,S) = \min\{dist(r,s) \mid s \in S\}$ to a pattern from S. We define now the reproducibility coefficient $repro(\leq, D)$ by:

$$repro(\leq, D) = 1 - \frac{\sum_{d \in D} mdist(d,S)}{|I| * n} \tag{6}$$

Here $|I| * n$ is the number of cells in the data matrix. Thus, the reproducibility coefficient can be interpreted as the percentage of cells in the data matrix which are compatible with S (and thus with $\leq$). We count for each response pattern d in D the minimal number of values $d(i)$ we have to change to transform d into a pattern in S. Thus, *repro* gives an insight about the minimal number of random errors that must have occurred under the assumption that all implications in $\leq$ are correct.

A Simulation Study

We test the ability of IITA to extract the true implications from the data in a simulation study. This simulation study should answer the following questions:

- Is the algorithm able to detect the valid implications with high accuracy if enough data are available?

- How much data do we need to ensure that the underlying quasi-order is reconstructed with high accuracy?

We use an item set I consisting of 9 hypothetical items for the simulation. The simulation procedure consists of the following steps:

1. A relation R on I is constructed randomly. Therefore, we loop over all possible pairs (i,j) of items with $i \neq j$ and include each pair (i,j) with probability δ in R. The reflexive pairs (i,i) are all included in R. The probability δ is changed randomly between the simulations to produce relations with different size.

2. The transitive closure $\leq$ of R is computed. Thus, $\leq$ is the smallest quasi-order which contains the relation R.

3. The set $S = \{s : I \rightarrow \{0,1\} \mid i \leq j \wedge s(j) = 1 \rightarrow s(i) = 1\}$ of all possible latent patterns consistent with $\leq$ is computed.

4. The set S is used to generate a simulated data set D with n data patterns. To generate this data set we repeat the following steps n times:

   a) An element $s \in S$ is chosen randomly.

   b) For each $i \in I$:

      i) if $s(i) = 1$, then $s(i)$ is set to 0 with probability τ

      ii) if $s(i) = 0$, then $s(i)$ is set to 1 with probability τ

5. We analyze D with IITA. This results in a quasi-order $\leq_{\text{opt}}$.

6. The quasi-orders $\leq$ and $\leq_{\text{opt}}$ are compared by:

$$\begin{aligned} \operatorname{dist}\left(\leq, \leq_{opt}\right) &= \left|\left(\leq \setminus \leq_{opt}\right) \cup \left(\leq_{opt} \setminus \leq\right)\right| \\ &= \left|\left\{(i,j) \mid \left((i,j) \in \leq \wedge (i,j) \notin \leq_{opt}\right) \vee \left((i,j) \notin \leq \wedge (i,j) \in \leq_{opt}\right)\right\}\right| \end{aligned} \tag{7}$$

The value of $\operatorname{dist}(\leq,\leq_{\text{opt}})$ is the number of implications in which $\leq$ and $\leq_{\text{opt}}$ differ. If $\operatorname{dist}(\leq,\leq_{\text{opt}}) = 0$ then the relation $\leq$ is perfectly reconstructed from the data.

The accuracy of IITA depends clearly on the number of available data patterns, i.e., the size of D, and the error probability τ. The number of available data patterns is varied as 50, 100, 200, 400, and 800. The error probability is varied as .03, .05, .08, and .1.

For each combination of values for n and the error probability τ we carry out 1000 simulations. The following table shows the mean value of dist($\leq$,$\leq_{\text{opt}}$) over these 1000 simulations. The values in parentheses indicate the *diff*-value divided by the number of non-reflexive item pairs (i,j).

Table 3

Mean Value of dist(≤,≤opt) for Different Error Probabilities and Different Values for n.

	n				
τ	50	100	200	400	800
.03	0.96 (.013)	0.66 (.009)	0.51 (.007)	0.42 (.006)	0.34 (.005)
.05	1.20 (.016)	0.93 (.013)	0.7 (.001)	0.65 (.009)	0.59 (.008)
.08	2.17 (.03)	1.83 (.025)	1.67 (.023)	1.62 (.023)	1.49 (.021)
.10	3.04 (.042)	2.47 (.034)	2.44 (.034)	2.35 (.033)	2.26 (.031)

In each simulation the algorithm had to decide for 72 item pairs (i,j) (since the 9 reflexive pairs (i,i) are always classified correctly) if $j \rightarrow i$ is true or not. As the simulation results show, this is done with high accuracy, even if the number of available response patterns is small. For example, for $\tau = .05$ and $n = 100$ in average less than 1 item pair is not classified correctly. Thus, the constructed quasi-order $\leq_{\text{opt}}$ differs in average in 1.3% of all non-reflexive item pairs from the true quasi-order $\leq$.

As we can see from Table 3 dist($\leq$,$\leq_{\text{opt}}$) increases with an increasing error level τ and decreases with an increasing number n of simulated response patterns.

In each simulation the not correctly classified pairs (i,j) can be divided into two subsets. The first subset consists of all implications $j \rightarrow i$ with $(i,j) \in \leq$ but $(i,j) \notin \leq_{\text{opt}}$. These are the correct implications which are not detected by IITA. The second subset consists of all implications $j \rightarrow i$ with $(i,j) \notin \leq$ but $(i,j) \in \leq_{\text{opt}}$. These are implications which are detected erroneously by IITA.

IITA is a method of explorative data analysis. Thus, the detected implications have to be investigated and confirmed in subsequent research. It is in general costly to investigate implications which then prove to be not true. Thus, the errors represented by the second subset should be small compared to the errors in the first subset.

The following table shows the mean value of the erroneously detected implications, i.e., the mean of $|(\leq_{\text{opt}} \setminus \leq)|$, over the 1000 simulations in each condition.

Table 4

Mean Value of $|(\leq_{\text{opt}} \setminus \leq)|$ *for Different Error Probabilities and Values for* n.

	n				
τ	50	100	200	400	800
.03	.64	.45	.33	.29	.24
.05	.44	.29	.16	.12	.12
.08	.33	.15	.09	.06	.04
.10	.21	.09	.05	.03	.02

The amount to which the erroneously detected implications contribute to the dist-value depends on the error probabilities. This is on the first sight a surprising result. The reason is that the algorithm uses the estimation of the error probability γ to chose the best fitting level.

If τ is small and the data set D contains only a small number of patterns, then the error constant γ will be overestimated in some simulations. In this case a higher tolerance level L will be accepted, i.e., the quasi-order $\leq_{\text{opt}}$ will contain erroneously detected implications.

If the number of data patterns is sufficiently large or if the error probabilities are relatively high, then most errors result from valid implications which are not detected by IITA.

Possible Extensions of the Analysis Method

IITA is designed to detect logical implications between items from observed response patterns. In this section we describe some extensions of this method.

Up to this point we restricted ourselves to direct implications. Depending on the content of the items it can be of interest to analyze more complex logical dependencies. It is, for example, possible that each subject who disagrees (0) to item j will agree (1) to item i. In this case we have to consider implications of the form $\neg j \rightarrow i$. Or it may be possible that each subject who agrees to item i and item j will agree to item k. In this case we have to consider implications of the form $i \wedge j \rightarrow k$.

We can use exactly the same method as described in the previous section to analyze such complex dependencies! Let X be a term consisting of variables for the items in I and the logical operators $\neg$, $\wedge$, and $\vee$. To analyze if $X \rightarrow i$ or $i \rightarrow X$ is true for some item i, we simply have to add a column X in the data matrix D. The value in this column is 1 if X is true for the item values in the corresponding row and 0 otherwise. Then we can simply analyze the extended data matrix by IITA.

If we want, for example, to consider the negations of the items, we have to add for each item i a item $\neg i$ to D. The value of this item $\neg i$ in each row is simply the complement of the value for i.

Our second extension deals with the handling of incomplete data. In many data sets we have to face the problem that some subjects answered only a part of the items in the questionnaire. To solve this problem we can simply exclude all incomplete response patterns from the analysis. This strategy is of course not optimal, since it throws away valuable information. We show in the following how the algorithm can be modified to handle also incomplete data.

Let d be the response pattern of a subject. If the subject does not answer item i we assign - to this item. Thus, d can be considered as a mapping d: $I \rightarrow \{0,1,-\}$.

Let o_i be the number of response patterns in which item i is not answered, i.e., $o_i = |\{ d \in D \mid d(i) = - \}|$ and q_{ij} be the number of subjects who agreed to item j and answered item i, i.e., $q_{ij} = | \{ d \in D \mid d(j) = 1 \wedge d(i) \neq - \}|$.

Now we are able to formulate the necessary modification of the algorithm. The only step we have to change is the method to determine the best fitting quasi-order. The first thing we have to change is the calculation of the t_{ij} values.

Assume $(i,j) \notin \leq_L$. As before t_{ij} should be the expected value of the number of response patterns with $d(i) = 0$ and $d(j) = 1$. Since we have to consider the case that an item is not answered we have to set

$$t_{ij} = (1 - (p_i + o_i)) * p_j * n * (1\text{-}\gamma). \tag{8}$$

Assume $(i,j) \in \leq_L$. The number of response patterns in which a counterexample to $i \to j$ can occur is given by q_{ij}. Thus,

$$t_{ij} = \gamma * q_{ij}. \tag{9}$$

In the calculation of γ we have to consider the cases in which an item is not answered. Thus,

$$\gamma = \frac{\sum \{b_{ij} \,/\, q_{ij} \mid i \leq_L j \wedge i \neq j\}}{(|\leq_L| - m)} \tag{10}$$

The rest of the algorithm stays the same.

Connections to Other Data Analytic Methods

There are some obvious connections between IITA (or any other method of Boolean analysis) and other well known methods of data analysis.

Guttman Scaling

IITA can be interpreted as a generalization of Guttman scaling. A Guttman scale defines a linear order $i_1 < i_2 < .... < i_m$ on a set $I = \{i_1, ..., i_m\}$ of items by simply setting $i_k < i_l$ if the number of subjects responding with a 1 to item i_k is lower than the number of subjects responding with a 1 to item i_l.

The assumption of a linear sequence of items poses heavy restrictions on the data. Thus, in practical applications a Guttman scale offers only in rare cases an adequate description of a data set. IITA is far less restricted, since we assume only a partial-order of logical implications. Thus, our method is able to produce a reasonable description of the logical implications between items for some data sets where no adequate Guttman scale exists.

As an example we analyze a data set from Suchman (1950). This data set is one of the classical examples for Guttman scaling. Suchman asked soldiers about the symptoms of fear they show under battle.

The following symptoms are asked:

1. Violent pounding of the heart
2. Sinking feeling in the stomach
3. Feeling sick in the stomach
4. Shaking or tremble all over
5. Feeling of stiffness
6. Feeling of weakness
7. Vomiting
8. Loosing control of the bowels
9. Urinating in pants

The following table shows the observed response patterns of 100 subjects where a 1 means that the symptom was observed.

Table 5

Observed Response Patterns and Their Frequencies in the Data from Suchman (1950).

Pattern	freq.	pattern (cont.)	freq. (cont.)	pattern (cont.)	freq. (cont.)	pattern (cont.)	freq. (cont.)
000000000	7	110100001	2	100001000	1	111001001	1
100000000	7	110101000	2	001100000	1	100101011	1
110000000	7	110101001	2	111000000	1	110110001	1
111111011	7	111101011	1	010110001	1	111101000	1
111101001	6	010000000	1	010101001	1	110100110	1
111111111	6	001000000	1	111001000	1	110110010	1
111111001	5	000010000	1	110110000	1	101111001	1
110001000	3	010001000	1	110010001	1	111011001	1
111100000	3	100000010	1	100001011	1	111111000	1
110001001	3	100010000	1	101110001	1	110111001	1
010000001	2	100000001	1	111110000	1	111110011	1
110100000	2	100100000	1	100111001	1	111101101	1
110111000	2	111101111	1				

The data can be very well described by a Guttman scale which is given by the left Hasse-diagram in Figure 3. This Guttman scale has a reproducibility coefficient of .93. Thus, 93% of the cells in the data matrix are in accordance with that linear order.

We analyze now the same data set with IITA. The following table shows the observed b_{ij}-values.

Table 6

The b_{ij}-Values for the Suchman Data.

		j								
		1	2	3	4	5	6	7	8	9
	1	0	6	2	3	2	2	0	0	4
	2	17	0	4	6	5	5	0	3	6
	3	44	35	0	20	11	19	1	5	12
	4	30	22	5	0	4	12	0	2	10
i	5	51	43	18	26	0	27	3	6	23
	6	34	26	9	17	10	0	1	4	10
	7	75	64	34	48	29	44	0	13	42
	8	63	55	26	38	20	35	1	0	32
	9	38	29	10	17	8	12	1	3	0

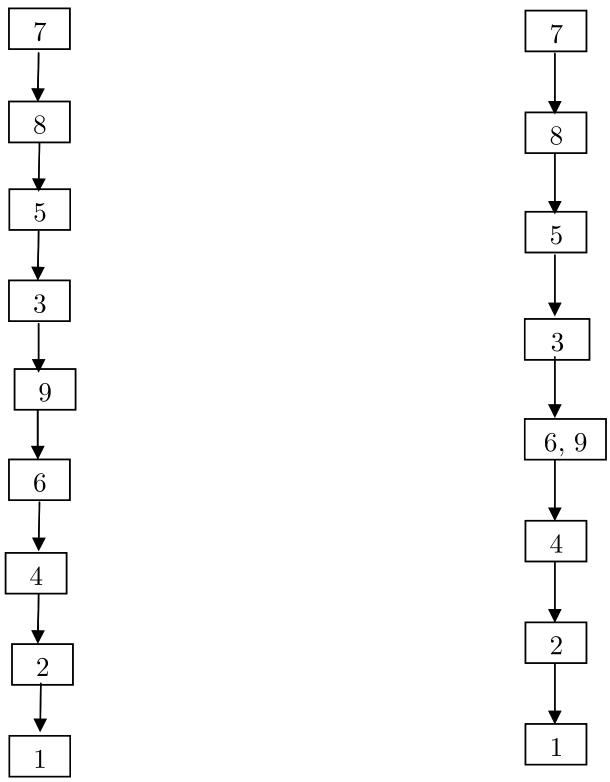

Figure 3. Guttman scale and best fitting quasi-order for the Suchman data.

The optimal fit is found at level 12. The resulting quasi-order is depicted as the right Hasse-diagram in Figure 3. The quasi-order has a reproducibility coefficient of .92. Thus, 92% of the cells in the data matrix can be explained by the quasi-order.

Figure 3 shows that our result is almost identical to the Guttman scale. The only exception is that we found the additional implication $6 \rightarrow 9$. This additional implication seems to be well supported by the data, since the reproducibility coefficient stays more or less the same when the additional implication is added to the set of implications derived from the Guttman scale.

The similarity of our analysis result to the Guttman scale is quite interesting, since IITA uses a totally different technique to derive an order from a data set.

Feature Pattern Analysis

Feature pattern analysis (Feger, 1994) can be seen as a generalization of Guttman scaling to higher dimensions. This method makes use of the information about the co-occurrence of the item values to built up a representation of the data by logical rules (for a geometric representation of FPA see, for example, Feger 2000).

Feature pattern analysis (FPA) is a stepwise procedure. First, it tries to find a representation of the data by analyzing the contingency tables of all item pairs. If no sufficient fitting representation is found the analysis operates on contingency tables of item triples, etc.

A solution of FPA in the first step is equivalent to a Guttman scale. FPA is a scaling method. The representation found by the method connects all items of the test or questionnaire.

We give now a rough description of FPA. For details see Feger (1994, 2000). Assume that we are in step n of the analysis. The analysis operates on all contingency tables resulting from the combination of n items. The researcher has to define a tolerance level L. In each contingency table *one* cell with a frequency $\leq L$ is defined to be a *zero cell*[6].

The interpretation of a zero cell is that the combination of item values defining this cell does not occur regular (with the exception of random errors) in the data. A zero cell can be represented by a number of logical dependencies. We show this by two examples.

Assume first that we have a contingency table of two items i and j defined by:

I	j	freq.	
0	0	20	
0	1	2	zero cell
1	0	18	
1	1	20	

[6] If in one of the contingency tables no cell with a frequency ≤ L exists, then FPA gives no solution for this step.

Thus, the pattern $i = 0$ and $j = 1$ does not occur regular in the data set. Therefore, we have the implications $j \rightarrow i$ and $\neg\, i \rightarrow \neg\, j$.

Assume now that we have a contingency table of three items i, j, and k and that the zero cell is given by $i = 1$, $j = 1$, and $k = 0$. Here we can derive the dependencies $i \wedge j \rightarrow k$, $i \wedge \neg\, k \rightarrow \neg\, j$, and $j \wedge \neg\, k \rightarrow \neg\, i$.

To form a proper representation of the data the zero cells have to fulfill a consistency criterion. This criterion ensures that the derived dependencies are logically consistent. Again we show that with an example. Assume that we have three items i, j, and k. Assume further that we are in step 2 of a FPA. The zero cell in the contingency tables of i and j is $(i = 1,\ j = 0)$, the zero cell for j and k is $(j = 1,\ k = 0)$, and the zero cell for i and k is $(i = 1,\ k = 0)$. Thus, $i \rightarrow j$, $j \rightarrow k$, and $i \rightarrow k$, and these logical dependencies are consistent.

Assume now that the zero cells are given by $(i = 1,\ j = 0)$, $(j = 1,\ k = 0)$, and $(i = 0,\ k = 1)$. Thus, $i \rightarrow j$, $j \rightarrow k$, and $k \rightarrow i$. These logical dependencies are inconsistent.

FPA selects in step n one zero cell from each contingency table of the n items. Since it is possible that more than one cell in the contingency table shows a frequency less than the tolerance level L there can be multiple solutions for one step of FPA. In this case FPA computes for each solution the sum z of the frequencies in the zero cells. The solution with the lowest z-value is chosen as the solution for this step.

The z-value of a solution is also used as a measure for the adequacy of a solution, i.e., for the fit of the solution to the data.

We have to mention here one major difference between FPA and Guttman scaling on one side and methods of Boolean analysis on the other side. FPA and Guttman scaling try to find a representation of the data which connects *all* items. In contrast for a Boolean analysis it is a natural result that some of the items do not show a dependency to other items.

As an example we analyze a data set from Gloning, Lienert & Quatember (1972). This set contains the data of 162 aphasic patients of the Universitäts-Nervenklinik in Vienna. Each patient was tested with respect to 5 tasks.

These tasks are:

A point to an object on a picture (Example: Please show me the ship.)

B name an object on a picture (Example: Please tell me how this object is called.)

C repeat a sentence (Example: Please repeat exactly what I say.)

D name as fast as possible words beginning with a given letter (Please tell me as many word as possible starting with M.)

E the number of verbal and phonemic errors produced when the patient performs the tasks B, C, and D

The items were dichotomized at the median and recoded in a way that 1 indicates aphasic behavior and 0 indicates normal or almost normal behavior. The observed response patterns of the data set can be found in Table 9.

Feger (2000) used the aphasic data set as an example for his scaling method. His analysis supports a solution for dimension 2, which shows a z-value of 38. Thus, each of the 10 zero-cells is violated in average by only 3.8 of the 162 response patterns. This is a very good fit to the data. The zero cells of this solution and the corresponding logical dependencies are shown in Table 7. We restrict our analysis of logical dependencies in this example to logical formulas which do not contain negations of symptoms. Thus, Table 7 shows only one of the three logical dependencies, which follow from each zero-cell.

Table 7

Results of FPA for the Aphasic Data Set.

Cell	zero cell	violations	logical dependency
ABC	101	7	$A \wedge C \rightarrow B$
ABD	101	2	$A \wedge D \rightarrow B$
ABE	100	3	$A \rightarrow (B \vee E)$
ACD	101	1	$A \wedge D \rightarrow C$
ACE	100	2	$A \rightarrow (C \vee E)$
ADE	100	2	$A \rightarrow (D \vee E)$
BCD	101	6	$B \wedge D \rightarrow C$
BCE	100	7	$B \rightarrow (C \vee E)$
BDE	100	1	$B \rightarrow (D \vee E)$
CDE	011	7	$D \wedge E \rightarrow C$

Thus, each logical dependency has the form $X \rightarrow Y$ where X and Y are items or a conjunction or disjunction of two items.

We analyzed the aphasic data set with IITA. The analysis showed that the items A, B, and C are equivalent (i.e., A ↔ B ↔ C) and that item D implies the items A, B, and C (i.e., D → A, B, C).

To get a comparable result to FPA we analyze the aphasic data set again with the described extension of IITA to complex dependencies. We include for each conjunction $X \wedge Y$ or disjunction $X \vee Y$ of two items (with $X \neq Y$) a virtual item in the data matrix. This item has the value 1 if the conjunction or disjunction is true for the corresponding row in the data matrix and 0 otherwise. Thus, the total number of items is 25 (5 real items, 10 disjunctions, and 10 conjunctions). This analysis showed some additional and more complex dependencies.

A first interesting result is that all logical dependencies from Table 7 are also found by IITA. But our analysis uncovered also a number of new dependencies.

A Hasse-diagram or a listing of all the detected implications is for this 25 item data set not informative. Therefore we show the results in form of a minimal set (i.e., all detected implications can be derived logically from this set) of logical formulas:

$$D \rightarrow A,\ B,\ C \tag{11}$$

$$A \leftrightarrow B \leftrightarrow C \tag{12}$$

$$A \rightarrow D \vee E \tag{13}$$

$$B \rightarrow D \vee E \tag{14}$$

$$C \rightarrow D \vee E \tag{15}$$

The set S of all consistent patterns is given by:

$$S = \{\ (0,0,0,0,0),\ (0,0,0,0,1),\ (1,1,1,0,0),\ (1,1,1,1,0),\ (1,1,1,0,1)\ \}$$

The reproducibility coefficient for this solution is .9.

An interesting observation from our comparison to FPA and Guttman scaling is that if a good fitting scale exists this scale is (together with a number of additional dependencies) reproduced by an analysis with IITA.

Latent Class Analysis

There are also some similarities between Boolean analysis and latent class analysis.

The basic assumption of latent class analysis is that the set of subjects which produced the data set D can be split into several subsets which are called *latent classes*. Each latent class x is described by an n-tuple $(\pi_{x1}, ..., \pi_{xm})$ where $\pi_{xi} \in [0,1]$ for each i= 1, ..., m.

The interpretation of $(\pi_{x1}, ..., \pi_{xm})$ is *A subject in latent class x answers item i with probability* π_{xi} *positive*. Each latent class occurs with probability π_x in the population.

The result of a latent class analysis of a binary data set D is thus given by:

$$\begin{array}{c} \pi_1 \quad (\pi_{11}, ..., \pi_{1m}) \\ \ldots \\ \pi_t \quad (\pi_{t1}, ..., \pi_{tm}) \end{array} \tag{16}$$

where t is the number of latent classes.

For a more detailed introduction into latent class analysis see, for example, Clogg (1995) or Rost and Langheine (1997).

The result of an analysis with IITA is a partial quasi-order ≤ on the item set I. As we have seen this quasi-order corresponds to a set $S := \{s\text{: } I \rightarrow \{0, 1\} \mid i \leq j \wedge s(j) = 1 \rightarrow s(i) = 1\}$ of latent states. The discrepancies between S and the data set D are explained by random errors. Let τ be the probability of a random error.

Thus, latent class analysis and IITA have in common that they both produce a separation of the population into a number of latent classes or states. In latent class analysis these latent classes are n-tuples of probabilities. A latent class describes the answer probabilities of a subject in that class. In our method the latent states are n-tuples of 0s and 1s. A latent state describes the responses of a subject in that state.

The main difference of both analysis methods is the way the results are interpreted. Latent class analysis tries to fit a model to the data. Thus, the latent classes together with their frequencies try to explain the data completely. The fit of the model can be tested by statistical methods (for details see, for example, Rost & Langeheine, 1997).

IITA tries, in the tradition of Boolean analysis and data mining techniques, to uncover existing deterministic dependencies between items. It does not assume that these dependencies together with some error parameters explain the data set completely. The existence of other probabilistic dependencies is not explicitly modeled in IITA.

It is also possible to estimate the frequencies of the latent states and the error probability τ in IITA. The frequencies $f(s)$ of the latent states can be estimated in the following way.

For each response pattern $d \in D$ we calculate the minimal distance $mdist(d,S)$ to a latent state from S. Define the set M_d by $M_d = \{s \in S \mid dist(s, d) = mdist(d,S)\}$. M_d is the set of all latent states with a minimal distance to d. We define now $m_{s,d}$ by:

$$m_{s,d} = \begin{cases} 0 & s \notin M_d \\ 1/|M_d| & s \in M_d \end{cases} \tag{17}$$

The frequency $f(s)$ of a latent state s can then be estimated by:

$$f(s) = \sum_{d \in D} m_{s,d} / |D| \tag{18}$$

The probability τ can be estimated directly from the reproducibility coefficient as $\tau = 1 - repro(\leq, D)$.

As an example we analyze a data set from the *International Social Science Survey Programme* (ISSP) for the year 1995. The ISSP is a continuing annual program of cross-national collaboration on surveys covering important topics for social science research. The program conducts each year one survey with comparable questions in each of the participating nations.

The theme of the 1995 survey was national identity. We analyze the results for question 4 for the data set of Western Germany. The statement for question 4 was:

Some people say the following things are important for being truly German. Others say they are not important. How important do you think each of the following is …

1. to have been born in Germany
2. to have German citizenship
3. to have lived in Germany for most of one's life
4. to be able to speak German
5. to be a Christian
6. to respect Germany's political institutions
7. to feel German

The subjects had the response possibilities *Very important*, *Important*, *Not very important*, *Not important at all*, and *Can't choose to answer those statements*.

To apply IITA to this data set we have to change the answer categories[7] . The answers *Very important* and *Important* are coded as 1. The answers *Not very important* and *Not important at all* are coded as 0. Every response pattern containing the answer *Can't choose* to one of the items is removed from the data set. The resulting data set contains 1126 response patterns. The observed b_{ij}-values are shown in Table 8.

Table 8

Observed b_{ij}-values for the data from question 4 of ISSP 1995.

		j						
		1	2	3	4	5	6	7
	1	0	335	232	450	107	487	342
	2	34	0	89	165	46	198	115
	3	93	251	0	313	81	360	227
i	4	19	35	21	0	26	82	37
	5	311	551	424	661	0	682	537
	6	24	36	36	50	15	0	32
	7	71	145	95	197	62	224	0

The optimal fit is found at level 95. The resulting quasi-order is depicted in Figure 4 as a Hasse-diagram. This quasi-order has a reproducibility coefficient of .94.

[7] Please note that for an analysis with latent class analysis a recoding of the data would be not necessary, since latent class analysis is able to handle polytomous data directly.

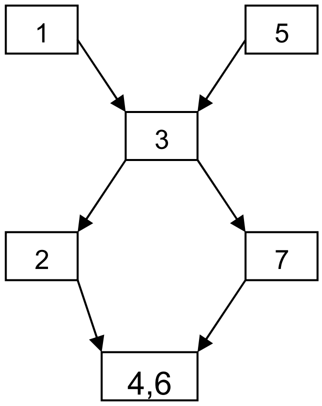

Figure 4. Best fitting quasi-order for the data from question 4 of ISSP 1995.

The set of latent states compatible with this quasi-order is given by:

$$S := \{ (0,0,0,0,0,0,0), (0,0,0,1,0,1,0), (0,1,0,1,0,1,0), (0,0,0,1,0,1,1), (0,1,0,1,0,1,1), (0,1,1,1,0,1,1), (1,1,1,1,0,1,1), (0,1,1,1,1,1,1), (1,1,1,1,1,1,1) \}.$$

If we estimate the error probability τ and the frequencies of the states we get $\tau = .06$ and the frequencies are given by:

State	Frequency
0000000	.046
0001010	.069
0101010	.078
0001011	.062
0101011	.111
0111011	.124
1111011	.226
0111111	.046
1111111	.239

We also applied the generalization of IITA to incomplete data to this data set. Here the answer *Can't choose* is coded as '-'. The full data set contains 1282 response patterns. The result of the analysis are exactly the same (naturally with other values for the b_{ij} and the optimal tolerance level) as described in Figure 4.

We analyze now the data for question 4 from ISSP 95 with latent class analysis. For the analysis we used the freeware program LEM from J. Vermunt[8].

As a criterion to determine the number of latent classes we used the Bayesean information criterion BIC[9], which gives an optimum of 3 latent classes for this data set.

The frequencies of the latent classes and the corresponding answer probabilities are given by:

.1111	(.0224, .1443, .0493, .3733, .3006, .7032, .1531)
.3927	(.9207, 1.000, .9516, 1.000, .6206, .9813, .9649)
.4959	(.3082, .7571, .5266, .9384, .1180, .9364, .7292)

Accordingly to this interpretation of the data there are three classes of subjects. Class 1 occurs with a frequency of 11% in the population and can be described by showing a high probability to find statement 6 to be important, a small probability to find statements 4 and 5 to be important, and a very small probability to find any of the remaining statements to be important. Class 2 occurs with a frequency of 39% in the data. A subject in this class finds statement 5 to be important with medium probability and all other items with a high probability to be important. Class 3 occurs with a frequency of 50% in the data. Subjects in this class find statements 4 and 6 with a very high probability to be important, statements 2 and 7 with a high probability to be important, and statement 3 with a medium probability to be important.

We summarize now the main differences and similarities between LCA and Boolean analysis. Both methods allow to split up the data into a number of latent states. For each latent state a frequency of occurrence in the population can be estimated.

The main difference between both methods lies in the nature of the latent states. In LCA the states are vectors of probabilities. In Boolean analysis the states describe vectors of 0s and 1s. Thus, if a subject is in a latent state in the sense of LCA we know with which probability he or she will answer any of the items positively. If a subject is in a latent state in the sense of Boolean analysis we know which answer he or she will give to each of the items (with the exception of a small fixed error probability).

[8] The program from J. Vermunt can be downloaded free of charge from the Web under the link www.kub.nl/faculteiten/fsw/organisatie/departementen/mto/software2.html.

[9] There exists a number of different approaches to determine the optimal number of latent classes. Since our goal is mainly to work out the differences of Boolean analysis and LCA we go not into detail here.

Thus, a latent class model with l classes shows that the population can be split into l subgroups. For each member of a subgroup we can determine the probability for his or her answers to each of the items.

The result of a Boolean analysis provides also a split of the population into subgroups, which are represented by the constructed deterministic states. But for each member of such a subgroup we know exactly how he or she will respond in principal (with the exception of some random errors in the response behavior) to each of the items.

Since the latent states in the sense of LCA are vectors of probabilities we can expect to find a sufficient explanation of the data with less states than in a comparable analysis by Boolean analysis, which represents the different possible response patterns by deterministic vectors of 0's and 1's.

Configural Frequency Analysis

Configural frequency analysis (Lienert, 1972) uses statistical techniques to search[10] for patterns in cross classification data.

Assume that we have a cross classification table of n (dichotomous) items. A configural frequency analysis (short CFA) starts from a base model which assigns an expected probability $P_{exp}(c)$ to each cell c in the cross classification table. Let $P_{obs}(c)$ be the observed frequency of a cell c in the cross classification table.

CFA makes use of statistical tests (for example Chi-square test or a binomial test) to find out if $P_{obs}(c) > P_{exp}(c)$ or $P_{exp}(c) > P_{obs}(c)$ is true for any cell c in the cross classification table. If $P_{obs}(c) > P_{exp}(c)$ is significant, then the cell c is called a *type*. If $P_{exp}(c) > P_{obs}(c)$ is significant, then the cell c is called an *antitype*.

CFA can be applied as a method for exploratory data analysis. Assume a set I of dichotomous items and a data set D of response patterns to that items. It is now possible to use CFA to analyze all cross classification tables of all item pairs, item triples, etc. The detected types respectively antitypes are then a description for the regularities in the data.

[10] Configural frequency analysis can also be used in a confirmatory context. See, for example, von Eye and Brandtstädter (1997).

Since a high number of tests is applied in this approach the α-level must be adjusted. This can be done, for example, by a Bonferroni adjustment, which replaces the overall α-level in each of the tests by α/t, where t is the number of tests. Another problem associated with this method is to choose the right base model. If there is no adequate information on the data available, then the only possible choice for the base model is to assume that all items are independent and to compute the expected frequencies from this assumption.

As an example for CFA we use again the aphasic data set from Gloning et al. (1972), which is already described in the section concerning FPA. We describe in the following the analysis of Lienert (1972) for the cross classification table of all 5 items. As base model it is assumed that all items are independent.

Remember that for each item i the value p_i is defined to be the relative frequency of rows in the data matrix D which show a value 1 for item i. The expected probability $P_{exp}(d)$ of a cell d in the cross classification table is (under the assumption that all items are independent) given by:

$$P_{exp}(d) = \prod_{\{i|d(i)=1\}} p_i \prod_{\{i|d(i)=0\}} (1 - p_i) \tag{19}$$

Table 9 (this table is taken from Lienert, 1972 p. 334) shows the cross classification table of all 5 items with the corresponding frequencies, expected probabilities and chi-square values.

The overall chi-square value is with 394.55 for 26 degrees of freedom highly significant. This shows that the assumption of item independence must be rejected. Lienert identified 4 types in the table. They show a highly significant difference $P_{obs}(c) > P_{exp}(c)$ and are indicated by an asterisk in the table.

We compare now these results with the results of an analysis of the aphasic data with IITA, which we have already described in the section concerning feature pattern analysis. Remember that the set S of all patterns consistent with the logical dependencies detected by IITA is given by:

$$S = \{ (0,0,0,0,0), (0,0,0,0,1), (1,1,1,0,0), (1,1,1,1,0), (1,1,1,0,1) \}$$

Thus, S contains all 4 patterns identified as types by CFA plus the additional pattern (1,1,1,0,0). Thus, the results of CFA are surprisingly similar to our analysis results.

This is a quite interesting result, since both methods represent very different approaches of data analysis.

Table 9

Cross Classification of the Aphasic Data set, from Lienert (1972).

Cell	Obs. Frequency	Exp. Frequency	Chi-Square
11111	5	3.549	0.594
11110	34	3.822	238.316*
11101	14	4.905	16.861*
11100	0	5.283	5.283
11011	0	3.637	3.637
11010	1	3.917	2.172
11001	7	5.028	0.773
11000	1	5.415	3.599
10111	0	3.637	3.637
10110	2	3.917	0.938
10101	4	5.028	0.210
10100	1	5.415	3.599
10011	0	3.728	3.728
10010	0	4.015	4.015
10001	3	5.154	0.900
10000	0	5.550	5.550
00111	1	4.436	2.661
01110	3	4.777	0.661
01101	2	6.132	2.784
01100	0	6.603	6.603
01011	0	4.547	4.547
01010	5	4.896	0.002
01001	7	6.285	0.081
01000	0	6.769	6.769
00110	2	4.896	1.713
00101	5	6.285	0.263
00100	7	6.769	0.008
00011	7	4.660	1.175
00010	8	5.019	1.771
00001	23	6.442	42.557*
00000	20	6.938	24.593*

Conclusions

We described an algorithm for exploratory data analysis which allows us to extract logical implications between items from observed response patterns. This algorithm, which we called IITA is based on a simple model of the data. We assume that there exists a number of valid logical implications, which form a quasi-order $\leq$ on the item set. Thus, without the influence of errors every subject should show responses which are compatible with $\leq$. We assume that each observed response pattern results from a pattern compatible with $\leq$ under the influence of random errors. This influence of random errors is described by an error parameter. A central assumption is that the error probability is independent from the items, i.e., the probability that an error occurs is the same for all items.

Our simulation study shows that IITA is able to reconstruct the true logical implications with high accuracy from observed data patterns.

We compared IITA with other well known methods of data analysis. A comparison to scaling methods like Guttman scaling or feature pattern analysis shows that if a good fitting scale exists this scale is reproduced by IITA. Since our analysis method is not restricted to a special structure it is no surprise that in these cases some additional dependencies are found, which can not be derived from the scale.

In our comparison to configural frequency analysis we found that the detected types of such an analysis are with one exception identical to the latent states which can be derived from the logical dependencies detected by IITA.

In these three comparisons IITA produced results which are very similar to the results of the methods it is compared to. Thus, a natural question is what the benefits of an analysis with IITA are. Please note that the data sets used in these comparisons are data sets which are often used to demonstrate the described analysis methods. Thus, there is some agreement that these methods produce very good results on these data sets. It is therefore a promising result that IITA produced very similar results in these examples. It is also clear that IITA can be applied in areas where some of the other methods will not produce reasonable results. For example, the scaling methods Guttman scaling and FPA poses heavy restrictions on the data (for example uni-dimensionality in Guttman scaling) and can thus only be applied to data sets which fulfill these requirements.

We compared IITA also with latent class analysis. Both analysis methods compute quite different forms of output. IITA searches for logical dependencies between items which can be used to compute the set of latent states compatible with them. If a subject is in a latent state, we know which answer he or she will give to each of the items (with the exception of a small fixed error probability). The result of a latent class analysis is a set of probability vectors, which are called latent classes. Thus, if a subject is in such a latent class we know with which probability he or she will answer any of the items positively.

References

Albert, D., Schrepp, M., & Held, T. (1994). Construction of knowledge spaces for problem solving in chess. In G. Fischer & D. Lamming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 123-135). New York: Springer.

Albert, D., & Lukas, J. (1999). *Knowledge spaces: theories, empirical research and applications*. Mahwah, NJ: Erlbaum.

Bart, W. M., & Krus, D. J. (1973). An ordering-theoretic method to determine hierarchies among items. *Educational and psychological measurement*, *33*, 291-300.

Bart, W. A., & Airasian P. W. (1974). Determination of the ordering among seven Piagetian tasks by an ordering-theoretic method. *Journal of Educational Psychology*, *66*(2), 277-284.

Buggenhaut, J., & Degreef, E. (1987). On dichotomization methods in Boolean analysis of questionnaires. In E. E. Roskam & R. Suck (Eds.), *Mathematical psychology in progress* (pp. 447-453). Amsterdam, NY: North Holland.

Clogg, C. C. (1995). Latent Class Models. In G. Arminger, C. C. Clogg & M. E. Sobel (Eds.), *Handbook of statistical modeling for the social and behavioral sciences* (pp. 311-359).

DiBello, L. V., Stout, W. F., & Roussos, L. A. (1995). Unified cognitive/psychometric assessment likelihood-based classification techniques. In P. D. Nichols, S. F. Chipman & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 361-389). Hillsdale, NJ: Erlbaum.

Doignon, J. P., & Falmagne, J. C. (1985). Spaces for the assessment of knowledge. *International Journal of Man-Machine Studies*, *23*, 175-196.

Doignon, J. P., & Falmagne, J. C. (1998). *Knowledge Spaces*. Berlin: Springer.

Duquenne, V. (1987). Conceptual implications between attributes and some representation properties for finite lattices. In B. Ganter, R. Wille & K. E. Wolff (Eds.), Beiträge zur Begriffsanalyse: Vorträge der Arbeitstagung Begriffsanalyse, Darmstadt 1986 (pp. 213-239). Mannheim: BI Wissenschafts-Verlag.

Eye, A. von (1991). *Prädiktionsanalyse*. Weinheim: Psychologische Verlagsunion.

Eye, A. von, & Brandtstädter, J. (1997). Configural frequency analysis as a searching device for possible causal relationships. *Methods of Psychological Research Online*, *2*(2), 7-23.

Feger, H. (1994). *Structure analysis of co-occurence data*. Aachen: Shaker.

Feger, H. (2000). Configural frequency analysis and feature pattern analysis: Some comparative observations. *Psychologische Beiträge*, *42*(3), 448-468.

Flament, C. (1976). *L'analyse booleenne de questionnaire*. Paris: Mouton.

Gagné, R. M. (1968). Learning hierarchies. *Educational Psychology*, *6*, 1-9.

Ganter, B., & Wille, R. (1996). *Formale Begriffsanalyse: Mathematische Grundlagen*. Berlin: Springer.

Gloning, K., Lienert, G.A., & Quatember, R. (1972). Konfigurationsfrequenzanalyse aphasiespezifischer Testleistungen. *Zeitschrift für klinische Psychologie und Psychotherapie*, *20*, 115-122.

Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review*, *9*, 139-150.

Hájek, P., Havel, I., & Chytil, M. (1966). The GUHA-method of automatic hypotheses determination. *Computing*, *1*, 293-308.

Hájek, P., & Havránek, T. (1977). On generation of inductive hypotheses. *International Journal of Man-Machine Studies*, *9*, 415-438.

Hájek, P., & Havránek, T. (1978). *Mechanizing hypothesis formation: mathematical foundations for a general theory*. Berlin: Springer.

Härtner, R., Mattes, K., & Wottawa, H. (1980). Computerunterstützte Hypothesenagglutination zur Erfassung komplexer Zusammenhänge. *EDV in Medizin und Biologie*, *2*, 53-59.

Held, T., & Korossy, K. (1998). Data-analysis as heuristic for establishing theoretically founded item structures. *Zeitschrift für Psychologie*, *206*, 169-188.

Janssens, R. (1999). A Boolean approach to the measurement of group processes and attitudes. The concept of integration as an example. *Mathematical Social Sciences*, *38*, 275-293.

Klemettinen, M., Mannila, H., Ronkainen, P., Toivonen, H., & Verkamo, I. (1994). Finding interesting rules from large sets of discovered association rules. In N. R Adam,.B. K. Bhargava & Y. Yesha (Eds.), *CIKM94: proceedings of the Third International Conference on Information and Knowledge Management*, November 29-December 2, 1994, National Institute of Standards and Technology (NIST), Gaithersburg, Maryland (pp. 401-407). New York: ACM Press.

Korossy, K. (1996). Kompetenz und Performanz beim Lösen von Geometrie-Aufgaben. *Zeitschrift für Experimentelle Psychologie*, *43*, 279-318.

Leeuwe, J. F. J. van (1974). Item tree analysis. *Nederlands Tijdschrift voor de Psychologie*, *29*, 475-484.

Lienert, G. A. (1972). Die Konfigurationsfrequenzanalyse. Assoziationsstruktur klinischer Skalen und Symptome. *Zeitschrift für klinische Psychologie und Psychotherapie*, 20, 231-248.

Nakhaeizadeh, G. (1998). *Data mining*. Heidelberg: Physika Verlag.

Rost, J., & Langeheine, R. (1997). *Applications of latent trait and latent class models in the social sciences*. New York: Waxmann.

Scandura J. M. (1971). Deterministic theorizing in structural learning: Three levels of empiricism. *Journal of Structural Learning, 3*, 21-53.

Schrepp, M. (1995). Modeling interindividual differences in solving letter series completion problems. *Zeitschrift für Psychologie*, *203*, 173-188.

Schrepp, M. (1997). An algorithm for the analysis of hierarchical relations between bi-valued test items. In W. Bandilla & F. Faulbaum (Eds.), *SoftStat'97 Advances in Statistical Software 6* (pp. 115-122). Stuttgart: Lucius & Lucius.

Schrepp, M. (1999). On the empirical construction of implications on bi-valued test items. *Journal of Mathematical Social Sciences*, *38*(3), 361–375.

Suchman, E. A. (1950). The utility of scalogram analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star & J. A. Clausen (Eds.), *Measurement and prediction* (pp. 122-171). Princeton, NJ: Princeton University Press.

Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, *20*(4), 345-354.

Tatsuoka, M. M., & Tatsuoka, K. K. (1987). Rule space. In S. Kotz & N. Johnson (Eds.), *Encyclopedia of Statistical Sciences* (pp. 217–221). New York: Wiley.

Theuns, P. (1994). A dichotomization method for Boolean analysis of quantifiable co-occurence data. In G. Fischer & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics and methodology* (pp. 309-402). Springer: New York.

Theuns, P (1998). Building a knowledge space via Boolean analysis of co-occurence data. In C. E. Dowling, F. S. Roberts, and P. Theuns (Eds.), *Recent Progress in Mathematical Psychology* (pp. 173-194). Hillsdale, NJ: Erlbaum.

Toivonen, H. (1996). Sampling large databases for association rules. *Proceedings of the 22th International Conference on Very Large Databases (VLDB96)*, Mumbay (India) September 1996 (pp. 134-145).

Wottawa, H., & Echterhoff, K. (1982). Formalisierung der diagnostischen Urteilsfindung: Ein Vergleich von linearen und auf Psychologenaussagen gestützte konfigurale Ansätze. *Zeitschrift für Differentielle und Diagnostische Psychologie*, *3*(4), 301-309.