Chapter 10
Formal Concept Analysis in Statistical Hypothesis Testing

Eraldo Nicotra and Andrea Spoto

1 Introduction

Hypothesis Testing consists of solving some kind of statistical test regarding the equality of group-referred parameters. Usually, this kind of procedure can be subdivided in three main steps: hypothesis definition, application of algebraic transformations over sampled parameters and conclusion. Solvers may present a wide variety of solution's patterns related to students' ability to solve the entire problem. The relation between the response patterns and the capability of each solver to master each part of the exercise could be mapped through a partial order and conveniently represented using the theoretical framework of Formal Concept Analysis (FCA) [1, 2] whose main concepts are introduced below.

The first basic notion of FCA is the *formal context* defined as a triple (G, M, I) where G is a set of *objects*, M is a set of *attributes* and I is a *binary relation* between the set of objects and the set of attributes. A formal context is usually represented by a Boolean matrix where each row is an object and each column is an attribute. Whenever a 1 is present in the entry (g, m), it means that the relation gIm for the specific g and m holds. Between objects and attributes of a formal context a *Galois connection* is defined. For all the sets $A \subseteq G$ and $B \subseteq M$, the following two transformations define the Galois connection:

$$A' := \{m \in M | gIm, \forall g \in A\} \tag{10.1}$$

$$B' := \{g \in G | gIm, \forall m \in B\} \tag{10.2}$$

E. Nicotra (✉)

Dipartimento di Pedagogia, Psicologia, Filosofia, University of Cagliari,
Via Is Mirrionis 1, Cagliari, Italy
e-mail: enicotra@unica.it

A. Spoto
Department of General Psychology, University of Padua, Via Venezia 8, Padua, Italy
e-mail: andrea.spoto@unipd.it

In words, A' is the collection of all the attributes that all the objects in A have in common. Dually B' is the collection of all the objects that possess all the attributes in B. It is now possible to introduce the fundamental notion of *formal concept*, that is a pair (A, B) that satisfies the following two conditions: $A = B'$ and $B = A'$. The *extent A* of the formal concept contains exactly those objects of G that have all the attributes in B; the *intent B* of the formal concept includes exactly those attributes satisfied by all the objects in A. A sub-concept super-concept relation is then defined in the following way:

$$(A_1, B_1) \le (A_2, B_2) \to A_1 \subseteq A_2 \tag{10.3}$$

or equivalently:

$$(A_1, B_1) \le (A_2, B_2) \to B_1 \supseteq B_2 \tag{10.4}$$

In words, a concept is of a lower level when it has a larger extent (or equivalently a smaller intent). The concepts of a context form a *complete lattice* [3] that is called the *concept lattice* of (G, M, I). The intents of a concept lattice are closed under intersection i.e. each intersection of sets of attributes is included in the lattice. In the present application the set of objects of our context consists of the response patterns to an exercise of "Hypothesis Testing for Paired Samples t-test", while the set of attributes consists of the subcomponents of the exercise that each pattern involves. In the Methods section we will go into more details about how the exercise has been divided in its subcomponents.

So far we introduced the deterministic part of the theoretical framework conceived for the present work. In evaluating knowledge some sort of probabilistic model has to be used in order to account for the variability of observed data. We decided to refer to the general framework of Item Response Theory and, more specifically, to the Rash Model [4].

The use of the *simple logistic model* was functional to obtain a probabilistic evaluation of the occurrence of each single response pattern observed to the task administered to the students. The application of the Rash simple logistic model, for dichotomous items allowed us to map each single deterministically detected items' structure obtained from the FCA solution into a probabilistic measurement system. This system, more realistically, conveys information about the probability to observe each specific response pattern, as the number of observations becomes sufficiently large. In details, the simple logistic model plays a fundamental role in discovering the measurement level of two parameters involved in items solution: the ability of a person to solve a problem and the difficulty of the asked items. We can now introduce some formalism to expose the mathematical structure of the Rasch's simple logistic model. The basic assumption of the model imposes that the probability to obtain a right solution of the proposed problem, for each respondent, can be related to the difficulty of the item and to the ability of the respondent. The more an individual is able the higher is the probability to obtain the correct solution to the problem, given the specific item difficulty. The mathematical relationship involved in the estimate of both the parameters of the Rasch model could be covered by the formula:

$$P(X_{ni} = 1|\beta_n, \delta_i) = \frac{e^{(\beta_n - \delta_i)}}{1 + e^{(\beta_n - \delta_i)}} \tag{10.5}$$

Where

- X_{ni} refers to response (X) elicited by the subject n to item i;
- β_n refers to the ability of subject n;
- δ_i refers to difficulty of item i;
- $X_{ni} = 1$ refers to a correct response of the item;
- e, indicates the base of the natural logarithm (i.e., $e = 2.718282$).

By applying the logistic model we were able to estimate the probability, for each subject, to partially or entirely solve the inferential problem and, conjointly, the probability associated to all the identified FCA solution's patterns. When the ability parameter β_n is higher than the related parameter δ_i the probability $P(X_{ni} = 1|\beta_n, \delta_i)$ of a correct response exceed 0.5; in the opposite case, the solution probability becomes smaller then 0.5. Both of the parameters estimated by the model identify a conjoint system of measurement of the collected dataset with respect to the psychological attitude of examinees and the complexity of the test to which they was exposed. The derived scale constitute a measure of a single latent trait.

It is now possible to introduce the specific application we carried out to test the applicability of the introduced elements.

2 Methods

2.1 Sample and Procedure

Participants were 256 students of the Psychometrics course 120 students were recruited at the University of Padua, while the remaining 136 students were recruited at the University of Cagliari. Students were asked to solve an "Hypothesis Testing for Paired Samples t-test" exercise. No time limit was imposed. Students were asked to solve the exercise in a paper and pencil way in order to allow the exact location of an error throughout the exercise. All participants attended the same theoretical and practice lessons on the topic during their Psychometrics course. Furthermore, both theoretical and practical lessons were carried out by the same teacher in both universities. Thus, given the fact that the sample size is almost the same for Padua and Cagliari, that the practical and theoretical lessons were exactly the same (same slides, same exercises, same text books) conducted by the same teacher, we can reasonably hypothesize the absence of substantial differences between the two groups.

The proposed exercise was initially subdivided into six main parts. After preliminary analysis through the Rash model, the following four parts were used to describe the whole exercise:

1. Hypothesis generation and formal expression;
2. Calculation of the test statistic t;
3. Identification of the critical value of test statistic t;
4. Assumption of the correct decision.

   Each part of the exercise was evaluated independently from the others, i.e., a "wrong" conclusion coherent with the obtained calculated value of the *t* statistics and with the critical value of the statistic was right scored. The scoring procedure assume that a *local independence* for each part of the problem exists.

2.2 Analysis

The first part of the analysis involved the construction of the formal context which had the four parts of the exercise as the attributes and the response patterns as objects. Figure 10.1 displays the implications existing among different patterns with respect to the attributes dimension of the FCA solution.
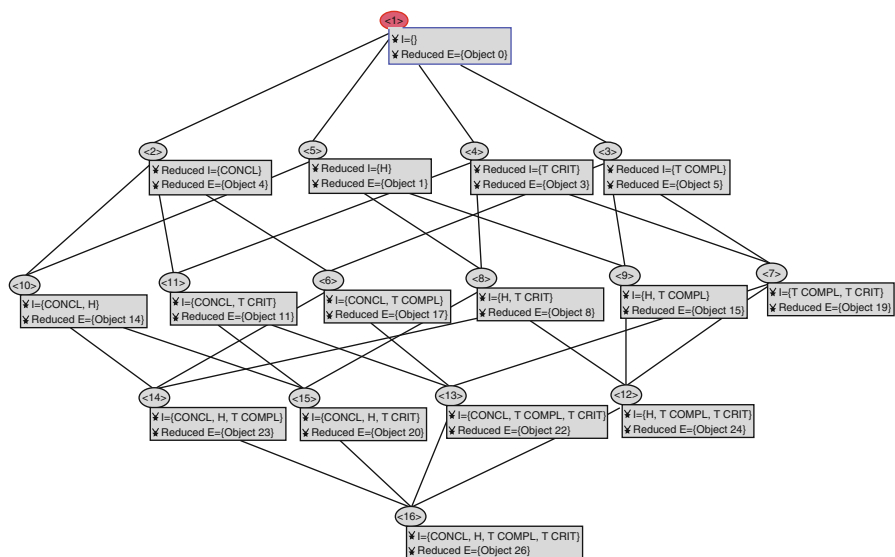

Fig. 10.1: The graphical representation of the implications among the parts of the exercise. In each node of the lattice a formal concept including a set of objects (response patterns) and a set of attributes (parts of the exercise) is represented. The four parts of the exercise are named as follows: H = hypothesis formulation and formal expression; T COMPL = calculation of the test statistic t; T CRIT = identification of the critical value of the test statistic t; CONCL = assumption of the correct conclusion

We then applied the Rasch simple logistic model to the context by means of both the RUMM 2020 and R statistical software. Through these analysis two main issues were obtained: on the one hand, it was possible to evaluate each part of the exercise in terms of its own difficulty (i.e. a location on the difficulty continuum was obtained for each item); on the other hand, through the evaluation of the residuals of the model, the ideal learning path moving from knowing nothing to knowing everything was found among different alternatives. The residual of a pattern indicates the discrepancy between the observed pattern and the expectations given by the model. The higher the residual, the greater is the discrepancy between the pattern and the model. Through this information it can be identified a path that minimizes the residuals, i.e. a set of steps that best represent the ideal learning process of a student.

3 Results

Table 10.1 displays both the location of the four parts of the exercise in terms of difficulty, and the residual analysis on the observed response patterns.

Table 10.1: The items' locations and the residual values of each response pattern

Items difficulty	Exercise part				Location
	I1: Hypothesis generation and formal expression				−0.422
	I2: Calculation of the test statistic t				0.402
	I3: Identification of the critical value of test statistic t				−0.223
	I4: Assumption of the correct decision				0.062
	Pattern/Item	I1	I2	I3	I4	Residual
	1	0	0	0	0	0.903
	2	0	0	0	1	−2.179
	3	0	0	1	0	−0.944
	4	0	1	0	0	−1.714
	5	1	0	0	0	−0.609
	6	0	0	1	1	0.594
	7	0	1	0	1	0.195
Residual analysis	8	1	0	0	1	−1.276
	9	0	1	1	0	1.254
	10	1	0	1	0	0.298
	11	1	1	0	0	−1.587
	12	0	1	1	1	0.774
	13	1	1	0	1	4.016
	14	1	0	1	1	0.086
	15	1	1	1	0	1.205
	16	1	1	1	1	−0.972

The higher the location, the more difficult is the item. The patterns are presented in a Boolean form in which a 1 in an entry means that the pattern includes the item

From the table it can be argued that the simplest part of the proposed exercise is the generation of the hypothesis and its formal expression, while the more difficult part is the calculation of the t statistics. It is noteworthy that the identification of the critical value of the test statistics appears to be less difficult than the assumption of the correct decision. With respect to the residual analysis, it can be highlighted that an ideal learning path includes the steps going from the empty set to the complete exercise through the mastering of item 1, then items 1 and 3, then items 1, 3 and 4. This path is included in the formal context and it is the one that both minimizes the residual of the model, and follows the difficulty order displayed by items locations.
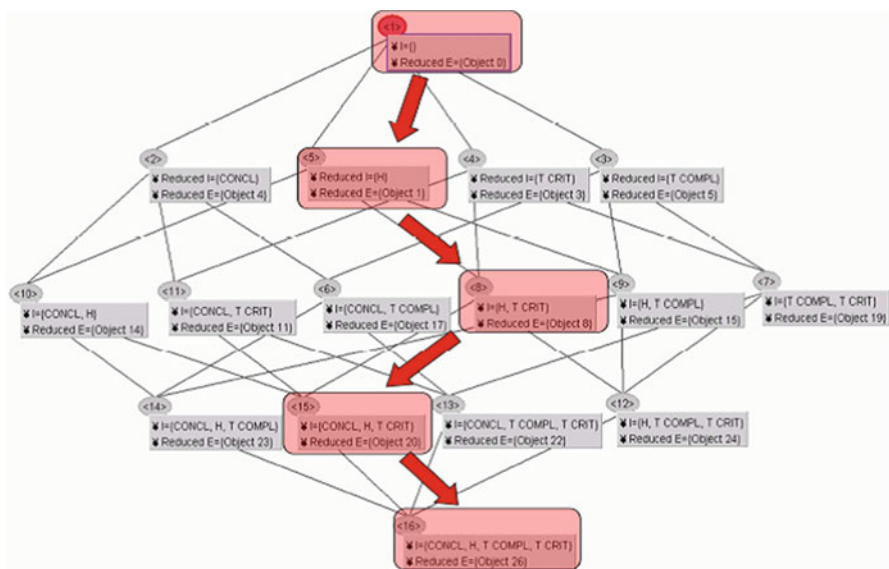

Fig. 10.2: The graphical representation of the results obtained through the analysis of the residuals. The highlighted nodes of the lattice represent the ideal learning path from the empty set (at the *top* of the figure), to the whole exercise (at the *bottom*)

Figure 10.2 displays the learning path depicted by the analysis of residuals. It can be seen how the path follows the increasing difficulty rate of the four parts of the exercise. By following the highlighted path the ideal sequence of concepts is learned. On the other hand, the present representation could be used as a reference point for calibrating an adaptive knowledge assessment tool. In fact, both the lattice and the path could be intended as the set representation of the implications existing among the parts of the exercise. Thus, if a student cannot formulate the correct hypothesis about the exercise, it would be almost certain that he/she will fail in computing the test statistic. The represented learning path is a total order. Through the FCA approach each partial order among items can be represented and implemented into an algorithm for the adaptive assessment of knowledge.

4 Discussion

The present work confirms the possibility of describing the solution process of a statistical exercise in terms of a number of steps that conduct from the total incapability to solve the exercise to a complete mastery of it. This process could be conveniently represented through the FCA and adequately described, from a probabilistic point of view, by the Rasch model. The use of FCA could play an important role in both algorithmically describe the solution process, and in planning specific teaching strategies accounting for the increasing difficulty of the proposed arguments and the logical sequence of exercise solution [5]. From the results it is possible to conclude that a reasonable solution to improve the ability of students to solve the proposed exercise may involve two main steps: the first one referring to the formal and logical part of the inferential process (hypothesis formulation, identification of the critical value of the test statistics, decision); while a second step of the learning process could be devoted to the more operational part of the exercise (i.e., the calculation of the value of the t statistics). The second part should follow the first one in order to maximize the probability of observing a correct answer to the exercise.

The presented approach could be fruitfully applied in adaptive knowledge assessment of students. More specifically, the implications obtained either a priori from theoretical consideration, or from a set of observed data, could be used to calibrate an algorithm that moves within the formal context by asking each time the more informative item. The procedure could allow to both improve assessment efficiency, and to precisely identify what exactly a student knows and what are his/her critical learning steps. Some software are already available for completing this task, but none of them refer to the FCA for detecting the specific relations among items [6–8].

Future works could investigate in more detail both the applicability of the method to other problem types [9] and to consider more sophisticated probabilistic models such as the partial credit [4].

References

1. Ganter, B., & Wille, R. (1999). *Formal concept analysis: Mathematical foundations*. Berlin: Springer.
2. Duquenne, V. (1999). Latticial structures in data analysis. *Theoretical Computer Science, 217*, 407–436.
3. Birkoff, G. (1967). *Lattice theory*. Providence, RI: American Mathematical Society Colloquium Publication.
4. Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Mahwah, NJ: Erlbaum.
5. Belohlavek, R., De Baets, B., & Konecny, J. (2014). Granularity of attributes in formal concept analysis. *Information Sciences, 260*, 149–170.
6. Schrepp, M. (1999). On the empirical construction of implications between bi-valued test items. *Mathematical Social Sciences, 38*, 361–375.

7. Schrepp, M. (2003). A method for the analysis of hierarchical dependencies between items of a questionnaire. *Methods of Psychological Research Online, 8*(1), 43–79.
8. Ünlü, A., & Malik, W. A. (2008). Psychometric data analysis: A size/fit trade-off evaluation procedure for knowledge structures. *Journal of Data Science, 6*, 491–514.
9. Xu, W., Pang, J. Z., & Luo, S. (2014). A novel cognitive system model and approach to transformation of information granules. *International Journal of Approximate Reasoning, 55*, 853–866.