The Korean Communications
in Statistics Vol. 7, No. 2, 2000
pp. 597-603

A Dimensionality Assessment for Polytomously Scored Items Using DETECT[1)]

Hae Rim Kim[2)]

Abstract

A versatile dimensionality assessment index DETECT has been developed for binary item response data by Kim (1994). The present paper extends the use of DETECT to the polytomously scored item data. A simulation study shows DETECT performs well in differentiating multidimensional data from unidimensional one by yielding a greater value of DETECT in the case of multidimensionality. An additional investigation is necessary for the dimensionally meaningful clustering methods, such as HAC for binary data, particularly sensitive to the polytomous data.

Keywords : DETECT, test dimensionality, polytomous, item response theory, item clustering

1. INTRODUCTION

The models for polytomously scored items have been developed for decades to meet several needs including extracting more information about examinees. Among many models of Samejima (1996), Masters (1982), and Muraki (1992) provide suitable frameworks. With such development of modeling for polytomously scored data, tools for assessing test dimensionality also need to be refined accordingly in case of polytomous data. For the binary data, Stout (1987) suggests a statistical test named DIMTEST to check the lack of test unidimensionality, while a versatile procedure DETECT has been developed by Kim (1994), and is useful in measuring the amount of multidimensionality, in counting the number of dimensions involved, and in clustering items into dimensionally homogeneous subgroups. However, both DIMTEST and DETECT use dichotomously scored data, hence the necessity comes out for polytomously

1) This paper is partially supported by Sangji University Research Grant in 1995.
2) Assistant Professor, Department of Applied Statistics, Sangji University, 660 Woosan-dong, Wonju, Kangwon-do, 220-702, Korea.
E-mail : hrkim@mail.sangji.ac.kr

scored data to develop procedures to assess the test dimensionality.

This paper investigates the possibility of extending DETECT's performance to the circumstances of polytomously scored data. The scheme of DETECT applies intact while it is used for the polytomously scored data. A primary simulation shows promising results in differentiating unidimensional data from multidimensional one when items are scored polytomously as well as scored dichotomously. A success in defining polytomous DETECT would provide practitioners with a tool to judge whether the test data are unidimensional or not in a variety of test circumstances including polytomous data cases. Moreover, with a suitable clustering procedure DETECT provides knowledge of how many abilities are involved in the test and what items are grouped together in terms of dimensionality.

This study, however, is aiming only to examine DETECT's extensibility to polytomous data using the given best partition of items. This partition can be found by other methods such as a statistical clustering procedure or the test maker's item specification in practice etc.. The next section briefly explains DETECT and suggests its possibility in case of polytomous data followed by a section showing results of a simulation study. The last section discusses relating work and future directions to fulfill the needs.

2. DETECT

It is assumed that a test is composed of disjoint clusters that are dimensionally distinctive from each other and each cluster is dimensionally homogeneous. Then it is conjectured that the conditional covariance estimate of an item pair given score of the remaining test is positive or negative subject to whether two items in the pair belong to the same cluster or not, respectively. The index DETECT by Kim (1994) combines non-zero second order conditional covariances of item pairs evidencing violation of unidimensionality by adding conditional covariances when two items come from the same cluster and subtracting conditional covariances when two items come from different clusters. That is, for an n item test and for an (i, j) item pair

$$DETECT = 100\ X \frac{1}{\frac{n(n-1)}{2}} \sum_{(i,j)} (-1)^{\delta_{ij}} (\widehat{Cov}_{ij} - \overline{Cov})$$

where summation extends over all item pairs and δ_{ij} is determined to be zero when two items belong to the same cluster or one otherwise. Note that the average $\overline{Cov}$ is the overall mean of covariances and the subtraction is introduced to extracting off the bias occurred possibly in the unidimensional case (Kim, 1994). Also note that 100 is multiplied for the ease of presentation and it is obvious that the maximum DETECT occurs if the correct cluster formation is utilized.

An objective of analysis employing DETECT involves finding cluster formation maximizing DETECT across formations. The number of clusters at the maximum DETECT corresponds to the number of dimensions present in the test, and each item is identifiable in terms of dimension to which it contributes. In addition, the amount of maximum DETECT is informative in indicating the degree of multidimensionality the test bears. An extensive choice of cluster formations, of course it is hoped that every formation is meaningful, is required in search for the maximum DETECT. Simulation studies show that DETECT finds extraordinarily the correct clusters for the tests having ''approximately'' simple structure. The underlying theory and details about DETECT can be found in Kim (1994) and Zhang and Stout (1999).

It is vital to employ the vast set of meaningful cluster formations of the test with which DETECTs are calculated during the search of maximum DETECT. In order to consider all possible clusterings an enormous combinational search would be required. Hence necessity of an efficient and powerful method comes out to cover a large range of possible clusterings. One choice in the present paper has been the Hierarchical Agglomerative Clustering (HAC) by Roussos (see Roussos, Stout, and Marden (1994) for details). It is shown that results from HAC are supportive in confirming hypothetical structure of the tests (Kim, 1994). The proximity used in HAC is the covariance between items given score of the remaining test.

In dichotomously scored data study by Kim (1994) DETECT changes its value from near zero for the unidimensional case to 2.84 for 2-dimensional data with 0.3 correlation coefficient when 40 items are divided into 20/20 simple structured groups. In 3-dimensional 13/13/14 item formation DETECT turns out to be 2.00 with 0.3 correlation coefficient while in 4-dimensional 10/10/10/10 it is 1.51. With gradually increasing correlation coefficient such as 0.5, 0.7, 0.8, and 0.9 the value of DETECT decreases gradually from 1.77, 0.91, 0.82, and 0.44 in 2-dimensional cases, implying that multidimensionality decreases as traits are collapsing together. Very similar facts are found in 3- and 4-dimensional cases the bigger correlation coefficient traits have the smaller multidimensionality is detected manifesting the smaller value of DETECT. One astonishing finding in the binary data study is that the maximum DETECT occurs with the known correct item formation for the simulated data and with the most suspicious item formation for the real data. It is important enough to mention that with the item formation yielding the maximum DETECT one might obtain knowledge about which items together are assessing certain ability, etc.. For more details, see Kim (1994).

DETECT uses conditional covariances of an item pair as building blocks. With polytomously scored items these covariances are computed in exactly the same way, but the magnitude of covariances grows much subject to the scoring function of items involved. In other words, even with polytomously scored data the notion and method of DETECT to detect the dimensionality of the latent test space remains same as with binarily scored data. But the bigger range of scoring function yields relatively bigger DETECTs.

3. A SIMULATION STUDY

3.1 Polytomous Item Response Model

Let θ be the trait to be measured. In the generalized partial credit model by Muraki (1992) the probability of choosing the h_{th} category for an item with m categories is written as

$$P_h(\theta) = \frac{\exp[\sum_{v=1}^{h} Z_v(\theta)]}{\sum_{c=1}^{m} \exp[\sum_{v=1}^{c} Z_v(\theta)]}$$

and

$$Z_h(\theta) = Da(\theta - b_h) = Da(\theta - b + d_h),$$

where D is a scaling constant that puts the θ ability scale in the same metric as the normal ogive model $(D=1.7)$, a is a slope parameter, b_h is an item-category parameter, b is an item-location parameter, and d_h is a category parameter.

Here $P_h(\theta)$ is called the item category response function. By letting T_h be the scoring function one can define the item response function $T(\theta)$ for the polytomously scored item to be the expected value of the scoring function T over categories;

$$T(\theta) = \sum_{h=1}^{m} T_h P_h(\theta).$$

If the scoring function is increasing with h, the model is for the ordered categorical responses.

3.2 The Method of Simulation

The simple structured data are simulated using the generalized partial credit model for 1-, 2-, 3-, and 4-dimensional cases. The length of each test is set to be 40 and 3000 examinees are generated from the standard normal distribution. For 2-dimensional test items are divided into two dimensionally homogeneous groups of the same length 20 each (20/20). In order to investigate the influence of proximity between traits the correlation coefficient of traits is varied from 0.3 to 0.9. In the case of 3-dimension 13/13/14 item group formation is used while in the case of 4-dimension 10/10/10/10 item group formation is used. For 3- and 4-dimensional cases correlation coefficients stay the same among different trait pairs, resulting in the balanced trait structure.

In each item the number of categories is set to be 4, i.e., $m=4$. For the category parameters d, several sets of parameters are used in this simulation. See Muraki (1992) for details. The below is a summary of the item parameters employed in the simulation.

Table 1. The mean and standard deviation of the generated item parameters

	slope parameter a	item-location parameter b
mean	1.00	0.00
std. dev.	0.50	0.80
minimum	0.35	-2.50
maximum	2.50	2.00

3.3 Results of A Simulation Study

This study only calculates the DETECT with the known item formation instead of searching the maximum value over possible item formations because the most dimensionally sensitive and efficient clustering procedure is yet to be developed. The performance of DETECT with polytomously scored data is promising and consistent with results found in the binary case. When unidimensional data are employed DETECT turns out to be near zero, about -0.04, implying no departure from unidimensional case. Table 2 shows that DETECT ranges from 17.29 to 2.69 in 2-dimensional case as correlation coefficient increases from 0.3 to 0.9, meaning as two traits are collapsing in terms of correlation coefficient the amount of multidimensionality decreases. Similar results can be found in 3- and 4-dimensional cases, checking DETECT ranging from 15.38 to 2.40 in 3-dimension and from 12.50 to 1.94 in 4-dimension, respectively. We notice that among the same correlation coefficient cases across various dimensions DETECTs decrease as more dimensions are introduced.

DETECT is calculated with known item formation 20/20, 13/13/14, or 10/10/10/10 instead of searching for the maximum over possible dimensionally homogeneous item formation found by an external clustering process. However, by introducing a percentage

$$r = \frac{DETECT}{abs(DETECT)} X\ 100$$

where $abs(DETECT) = \frac{1}{n(n-1)/2} \sum_{(i,j)} (-1)^{\delta_{ij}} |\widehat{Cov}_{ij} - \overline{Cov}|$, we see that DETECTs in Table 2 are the maxima of the cases. That is, $r = 100$ means that all entries of DETECT from the same cluster are added up while the other from different clusters are subtracted, resulting in DETECT as same as $abs(DETECT)$.

Table 2. DETECT and r for polytomous data*

correlation coefficient	2D	3D	4D
0.3	17.29 (100.0)	15.38 (100.0)	12.50 (100.0)
0.5	13.69 (100.0)	10.81 (100.0)	8.54 (100.0)
0.7	7.57 (100.0)	6.02 (100.0)	5.98 (100.0)
0.8	4.88 (100.0)	4.70 (99.9)	3.50 (99.8)
0.9	2.69 (99.9)	2.40 (99.7)	1.94 (97.3)

* The percentage r is in the parenthesis.

We recall that the conditional covariance estimate of an item pair given score of the remaining test is positive or negative subject to whether two items in the pair belong to the same cluster or not, respectively. That is, item formation information used here is the very one yielding maximum DETECT. Table 2 shows only one r, 97.3 deviates from 100 significantly when very high correlation coefficient 0.9 is used in 4-dimensional case. It is obvious that the range of DETECT for polytomous data is much wider than the counterpart for dichotomous data because of the bigger range of the raw item response, which is 0-4 for polytomous data and 0-1 for dichotomous data, respectively.

4. CLOSING

DETECT performs extraordinarily well when more than two categories are used in an item to check the amount of multidimensionality existing in the data. The amount of multidimensionality is manifesting by resulting in a huge DETECT such as 17.29 in 2-dimensional 0.3 correlation coefficient case. Also the unidimensional case differs clearly from the other by having near zero DETECT. The results of this study can be used as a benchmark for the future study. One related ongoing work is to investigate the cases of mixed structure and unbalanced cases as well as the case of having items with different number of categories. Instead of finding possibly best partition of items in terms of dimensionality this study uses known item formation. Thus more work needs to be done developing a successful process which is sensitive enough to item dimensionality for the polytomously scored data. A study employing real data also needs to be performed as a relevant future work.

References

[1] Kim, H. R. (1994). New Techniques for the Dimensionality Assessment of Standardized Test Data. A Doctoral Dissertation, University of Illinois at Urbana-Champaign.

[2] Masters, G. N. (1982). A Rasch Model for Partial Credit Scoring. *Psychometrika* 47, 149-174.

[3] Muraki, E. (1992). A Generalized Partial Credit Model: Application of an EM Algorithm. *Applied Psychological Measurement* 16, 159-176.

[4] Roussos, L., Stout, W., and Marden, J. (1994). Analysis of the Multidimensional Simple Structure of Standardized Tests using DIMTEST with Hierarchical Cluster Analysis. Unpublished Manuscript. Department of Statistics, University of Illinois at Urbana-Champaign.

[5] Samejima, F. (1996). Graded Response Model. In W. Linden and R. Hambleton (Eds.), *Handbook of Modern Item response Theory*. New York, NY; Springer.

[6] Stout, W. F. (1987). A Nonparametric Approach for assessing Latent Trait Unidimensionality. *Psychometrika* 52, 589-617.

[7] Zhang, J. and Stout, W. (1999). The Theoretical DETECT Index of Dimensionality and Its Application to Approximate Simple Structure. *Psychometrika* 64, 213-249.