第40卷 第5期
2016年9月
江西师范大学学报(自然科学版)
Journal of Jiangxi Normal University(Natural Science)
Vol. 40 No. 5
Sep. 2016

文章编号: 1000-5862(2016)05-0456-09

A Simulation Approach to Selecting the Appropriate Method of Determining the Number of Factors to Retain in Factor Analysis

Cui Ying

(Department of Education Psychology, Faculty of Education, University of Alberta, Edmonton T6G25G, Canada)

Abstract: The appropriate use of unidimensional and multidimensional IRT requires the knowledge of the correct number of constructs assessed by an educational measurement. Factor analysis may be used to determine the number of constructs (or factors) that a set of test items really measures by providing a description of the underlying structure of the test. Several rules have been suggested by researchers about how to decide the number of factors to retain in factor analysis. It should be noted that there are a number of variables that contribute to the inaccuracy and inconsistency of the decision rules. The study suggests a simulation method for selecting the appropriate rule to decide the number of factors for a specific data set.

Key words: construct validity; item response theory; factor analysis; number of constructs
中图分类号: B 841.7 **文献标志码:** A **DOI:** 10.16357/j.cnki.issn1000-5862.2016.05.02

0 Introduction

Unidimensional Item Response Theory models are helpful when only one construct is measured by the test. When the unidimensionality assumption holds, each item on the test only measure one ability trait or a specific combination of traits. However, psychological processes have consistently been found to be more complex than they first appear, and an increasing number of educational measurements assess an examinee on more than one trait factor[1]. Multidimensional IRT is a methodology that shows promise for dealing with this form of complexity in psychological assessments[2]. If the dimensionality of response data is established, the appropriate unidimensional or multidimensional IRT model is applied to estimate item parameters. The necessity arises, then, to determine the number of trait factors assessed by an educational measurement. Factor analysis may be used to determine the number of constructs (or factors) that a set of test items really measures by providing a description of the underlying structure of the test[3]. The purpose of this paper is to show how factor analysis can help assess the dimensionality of an educational measurement and a simulation approach is introduced to select the correct method to determine the number of factors to retain.

This paper is structured as follows. First, factor analysis is briefly reviewed. Of particular interest is the discussion of the factor analysis for items scored in two categories. Then, a review of different rules for determining the number of factors to retain will be undertaken. Next, using the factor analysis to assess the dimensionality of a real test will be conducted. Due to the inconsistency of the different decision rules and the uncertainty of the correct number of factors and also considering the fact that the impact of various conditions (for example, the number of variables and component saturation etc.) on the different rule's accuracy is unlikely same, a simulation study will be performed to select an relatively accurate decision rule to determine the number of factors in the particular condition of this real test. Finally, the future research possibilities will be discussed.

0.1 Factor Analysis

Factor analysis is a data reduction statistical tech-

收稿日期: 2016-07-10
基金项目: Insight Development Grant from the Social Sciences Humanities Research Council of Canada(430-2011-0101).
作者简介: 崔 迎(1978-),女,辽宁鞍山人,副教授,博士,主要从事教育测量学、个人拟和分析、认知诊断测量和应用统计在社会科学的运用研究.

nique, attempting to explain the pattern of correlations of a set of observed variables by identifying a smaller set of latent variables, or factors. This is obtained by factoring the Pearson Product Moment correlation matrix of the observed variables into a pair of matrices: the correlation matrix between factors and the factor by variable weight matrix. There are two basic models in factor analysis: component factor model and common factor model. In component factor model, there is no error term appeared in the model. It is assumed that the factors of the model can perfectly explain the observed variables and also any observed error should be explained as the unfitness of the model to the data. In the common factor model, the factors are divided into two groups: common factors and unique factors. Common factors are those contributing two or more variables in the model. It is assumed each variable can't be explained by common factors perfectly. Each variable of the model has the unique variance independent of other variables. That is, the total variance of each variable is explained by both common factors and unique factors.

However, the traditional factor analysis procedures produce meaningful and interpretable results only if the variables being analyzed are truly continuous and also multivariate normally distributed. Dichotomous items are obviously problematic. I. G. Bernstein and G. Teng [4] argue that applying the traditional factor procedures to the polytomous-scored often produces just as bad results as the dichotomous items, if not worse than. One of the recommendations is to conduct a traditional factor analysis on the tetrachoric correlations matrix in the case of dichotomous data or the polychoric correlations matrix in the case of polytomous data rather than on the matrix of pearson correlations. The tetrachoric and polychoric correlations assume that the latent variables underlying the observed dichotomous and polytomous variables are continuous and each latent variable has threshold values that discretize the latent variable to produce dichotomous and polytomous scores. Factor analysis of tetrachoric or polytomous correlation matrices are essentially factor analyses of the relations among latent response variables that are assumed to underlie the data and that are assumed to be continuous and normally distributed [5].

Another approach is to conduct a full information factor analysis, a factor analytic technique based on the latent trait model. This approaches bypasses the problems associated with correlations matrices entirely, uses all of the information contained in the response category pattern frequencies [6].

0.2 Rules for Determining the Number of Factors

It is critical to extract the correct number of factors. Both the over-and under-factoring will distort the definition of major factors [7]. There are a number of rules having been suggested by researchers to determine the number of factors in factor analysis. However, the decisions by different rules are often not consistent with each other. Thus, it is essential to select the appropriate rule to decide the number of factors. In this paper, four widely-used rules are discussed. These rules are: Kaiser-Guttman rule; the scree test; Velicer's minimum average partial test and the parallel analysis method.

0.2.1 Kaiser-Guttman Rule (K1) The K1 rule [8] is the default option on many statistical packages (i.e., SPSS). It simply states that the number of factors is equal to the number of factors with eigenvalues greater than 1.0. The K1 rule is obviously considered leading to deceptive results [7,9].

0.2.2 Scree Test The scree test was developed by Cattell [10]. The idea in the scree test is that if a factor is important, it will have a large eigenvalues (variance). In the scree plot of eigenvalues extracted from a full correlation matrix against their serial order, start with the smallest eigenvalues to draw a straight line from right to left ensuring that the line capture as many of the eigenvalues as possible within close proximity to the line. If there are still a great number of eigenvalues above the line, start with the smallest eigenvalues not in the first line and draw another straight line according the same rule. This process is stopped when the number of eigenvalues left is small. Then the factors above the lines in the plot are kept. These are the important factors which account for the bulk of the correlations in the matrix.

0.2.3 Velicer's Minimum Average Partial Test (MAP) Velicer's [11] MAP test is a method based on the examination of the matrices of partial correlations following the factor analysis. The idea of the MAP test is that factors should no longer be retained when there are more unsystematic variances than the systematic variances in the partial correlation matrix. Initially, the average squared correlation in the original correlation matrix is calculated. Then, the first factor is partialed out of the correlation matrix of the observed variables and the average of the squared partial correlations in the offdiagonals of the resulting partial correlation matrix is calculated. Similarly, the first two factors are partialed out of the correlation matrix of the observed variables. Likewise,the average of the squared partial correlations is computed. The computation is stopped until all the k-1 factors are partialed out of the correlation matrix,while k is the number of observed variables. Order the k-1 average squared partial correlations. If the average squared correlation in the original correlation matrix is lower than the lowest average squared partial correlation,then there is no factor in the data set. Otherwise,the corresponding number of factors partialed out of the correlation matrix that results in the minimum average squared partial correlation is the number of factors to retain.

0.2.4 Parallel Analysis[12] Parallel analysis (PA) involves randomly simulating the data sets that parallel the real data set in the cases of the number of cases and number of variables. Horn argues that all the eigenvalues are 1.0 when the variables are uncorrelated with each other at a population level. However,in the case of a sample,the eigenvalues will not be the same and the initial eigenvalues will be greater than 1.0. In Horn's original description of this procedure,the eigenvalues obtained from the actual data set are compared with the mean of the eigenvalues simulated from the random parallel data sets. However,the use of the eigenvalues corresponding to an appropriate percentile (typically the 95^{th}) of the distribution of the random data eigenvalues as the comparison baseline is recommended[13-14]. The factor will be retained if the eigenvalue of this factor from the real data set is greater than the corresponding mean or specific percentile eigenvalue from the random data sets.

1 Method

This study uses the data from a national testing program of Canada. It was a quantitative measure consisting of 78 items completed by 9 294 candidates. All the 78 items are dichotomously scored. Thus,the traditional factor analysis is not appropriate for the data. First,the tetrachoric correlation matrix is estimated for the entire 78 items utilizing the TESTFACT program. The matrix for the entire scale is then subjected to factor analysis in the SPSS program using the principle component extraction method. The estimation of the 78 eigenvalues and scree plot are achieved. The number of factors to be retained by the K1 rule and scree test is determined respectively. However,the popular statistical software package,such as SPSS and SAS,doesn't provide the MAP test and parallel analysis for users. The SPSS syntax for these two tests programmed by O'Connor[15] is employed here. The PA decision is based on the 95 percentile eigenvalues of the 100 correlation matrices of the randomly generated data sets with the same number of items and examinees as the real data set. The number of factors in the MAP test is determined by the step number that results in the lowest average squared partial correlation.

The purpose that the simulated data of a known number of factors are used here is to examine the relative accuracy and consistency of the four decision rules with a known criterion in the particular condition of the real data set used in the study. Thus,all the simulated data sets should be compared with the real data set. For each simulated data set,the comparison of the number of factors determined by each of the four rules (K1, scree test,MAP,and PA) with the known number provide us an objective criterion to select the correct rule to determine the number of factors to retain for the real data set.

Inspection of the loading matrix of the real data set,factor saturation,the magnitude of the loading of each item,is low,range from 0.1 to 0.3. Thus,the number of factors to be used for the simulation is set moderately small. 1,3,and 6 factors are built into the simulated data set respectively. Because the test has many open-ended items,all the pseudo-guessing rates are set to zero. The 78 items of the real data set are then fit into 1-,3-,and 6-solution non-chance full information item factor model using TESTFACT to obtain the related parameters subsequently used to simulate the data set of corresponding number of factors. Consequently,three sets of parameters for each solution used as the basis for simulating data are estimated: the standard item difficulties,the factor loadings,and the population means of factor scores.

Next,the three sets of parameters for each solution are used to simulate 5 data sets,$5\times3=15$ data sets totally for all the solutions,with the same number of examinees as the real data set. Each simulated data set was compared to the real data set on item difficulties, factor loadings,and means of factor scores. Then,for each simulated data set,the tetrachoric correlations are estimated using the TESTFACT program. The tetrachoric correlation matrix is subsequently subjected to SPSS utilizing principle component extraction method to obtain the corresponding number of factors determined by each of the four decision rules.

2 Results

For the real data set, each of the four decision rules leads to a different estimate of the number of factors. The principle component extraction results in 20 factors with eigenvalues greater than 1 (K1) . Inspection of the scree plot (Figure 1) indicates the data is essentially unidimensional but shows some indications of minor secondary dimensions since the second line is somewhat steeper than the first line. Figure 2 shows the plot of the observed and the 95 percentile random eigenvalues of 100 generated data sets against factor number, and indicates that 14 factors should be extracted since there are 14 observed eigenvalues greater than the corresponding random eigenvalues. The MAP test shows 4 factors should be retained since the average squared partial correlation obtains the lowest value when the first four factors are partialed out of the tetrachoric correlation matrix.

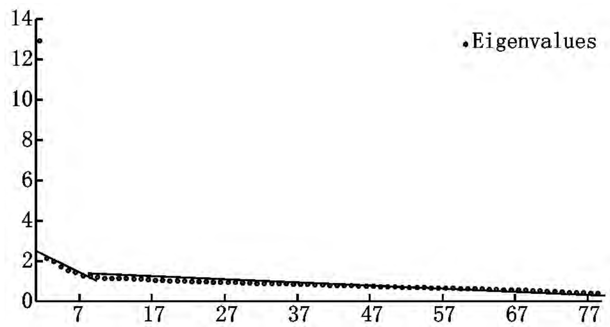

Figure 1　Scree plots for the real data set

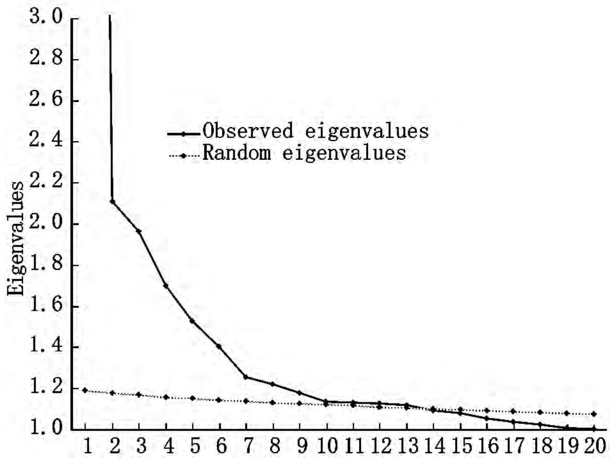

Figure 2　Parallel analysis for the real date set

Similarly, the number of factors determined by each of the four decision rules is estimated foreach set of simulated data. The difference between the number of factors determined by each of the four decision rule and the known population number of factors is calculated respectively. A positive difference shows that the decision rule overestimates the number of factors, while a negative difference indicates the underestimation of the number of factors by the decision rule. A zero difference is a perfect estimate by the decision rule. The average difference and standard deviation among the 5 data sets for each rule for each solution are calculated (Table 1) . On the whole, a low average difference and standard deviation suggests the decision rule is accurate and consistent in the particular condition of the real data set. Consequently, this decision rule can be used to decide the number of factors for the data set. Table 2 presents the performance of four rules for each simulated data set. The first row of table 2 shows the number of factors estimated by the four rules and the difference between the rule-determined number of factors and the known value for the first simulated 1-factor data set. There are 19 eigenvalues greater than 1, so the K1 rule produces a positive difference of 18, an overestimation. Inspection of the scree plot (series 1, figure 3) suggests the data is quite unidimensional. Thus, the scree plot produces a zero difference, a perfect estimation. Likewise, the MAP method (table 1) suggests 1 factor in the data set and present a perfect estimation since the average partial correlation obtains the lowest value, 0.000 581, when only 1 factor is partialed out of the original matrix. Parallel analysis (series 1, figure 4) indicates 6 factors should be extracted from the first simulated 1-factor data set since the first six observed eigenvalues are greater than the corresponding random eigenvalues. Parallel analysis gives a positive difference of 5, an overestimation.

The last two rows of Table 2 present the mean and standard deviation of the difference between the rule-determined number and the known number of factors for each rule. The means of the scree plot and MAP methods are close to zero, indicating the two methods perform well, the scree plot slightly better than the MAP method. The two means are both negative, showing the two method incline to underestimate the number of factors. The standard deviations of the two methods are also quite small, indicating that their performances are consistent. The mean of parallel analysis is moderately large and its standard deviation is the largest among the four methods, suggesting this method gives an overestimation and its performance is variable. The mean of the K1 rule is the largest and the standard deviation is quite small, showing this method consistently overestimates the number of factors.

The scree plots and parallel analyses for the three-factor simulated data sets are presented in Figure 5 and 6 respectively. The scree plots indicate that there is essentially one dominant factor in the data sets and the difference between the first and second eigenvalue is quite large compared to the difference between the second and third eigenvalue, suggesting one factor to be retained in the simulated 3-factor data sets. Parallel analyses show that there are 6, 9, 9, 7, and 6 observed eigenvalues greater than the corresponding random eigenvalues respectively for the five data sets.

Table 1 Velicer´s average squared correlations

step	Real data set	1-factor simulated data set					step	Real data set	1-factor simulated data set				
		1	2	3	4	5			1	2	3	4	5
0	0.025 283	0.022 878	0.022 025	0.023 106	0.022 123	0.023 472	11	0.002 996	0.002 783	0.002 779	0.002 802	0.002 786	0.002 771
1	0.002 025	0.000 581	0.000 583	0.000 688	0.000 574	0.000 591	12	0.003 243	0.003 082	0.003 081	0.003 087	0.003 051	0.003 055
2	0.001 908	0.000 749	0.000 774	0.000 863	0.000 751	0.000 764	13	0.003 491	0.003 381	0.003 376	0.003 393	0.003 335	0.003 37
3	0.001 794	0.000 94	0.000 952	0.001 042	0.000 943	0.000 942	14	0.003 742	0.003 694	0.003 65	0.003 707	0.003 636	0.003 686
4	0.001 788	0.001 136	0.001 145	0.001 231	0.001 142	0.001 13	15	0.004 009	0.004 029	0.003 957	0.004 015	0.003 962	0.004 035
5	0.001 865	0.001 344	0.001 348	0.001 422	0.001 349	0.001 331	16	0.004 295	0.004 377	0.004 302	0.004 343	0.004 293	0.004 345
6	0.001 982	0.001 566	0.001 574	0.001 637	0.001 56	0.001 55	17	0.004 623	0.004 685	0.004 641	0.004 686	0.004 627	0.004 712
7	0.002 157	0.001 793	0.001 806	0.001 842	0.001 792	0.001 756	18	0.004 955	0.005 036	0.004 965	0.005 032	0.004 957	0.005 108
8	0.002 338	0.002 025	0.002 037	0.002 065	0.002 018	0.001 989	19	0.005 308	0.005 408	0.005 352	0.005 384	0.005 316	0.005 49
9	0.002 554	0.002 27	0.002 274	0.002 287	0.002 272	0.002 257	20	0.005 649	0.005 828	0.005 717	0.005 735	0.005 711	0.005 922
10	0.002 778	0.002 533	0.002 528	0.002 547	0.002 525	0.002 503							

step	3-factor simulated data set					step	3-factor simulated data set				
	1	2	3	4	5		1	2	3	4	5
0	0.0126 08	0.0121 16	0.0121 96	0.0131 71	0.0128 83	11	0.002 636	0.002 618	0.002 564	0.002 62	0.002 593
1	0.000 743	0.000 717	0.000 732	0.000 726	0.000 766	12	0.002 882	0.002 861	0.002 798	0.002 871	0.002 855
2	0.000 802	0.000 796	0.000 809	0.000 797	0.000 804	13	0.003 135	0.003 125	0.003 051	0.003 141	0.003 11
3	0.000 939	0.000 93	0.000 941	0.000 934	0.000 925	14	0.003 409	0.003 382	0.003 313	0.003 406	0.003 393
4	0.001 117	0.001 116	0.001 123	0.001 117	0.001 105	15	0.003 683	0.003 663	0.003 575	0.003 68	0.003 676
5	0.001 304	0.001 304	0.001 306	0.001 306	0.001 296	16	0.003 975	0.003 95	0.003 873	0.003 965	0.003 957
6	0.001 503	0.001 502	0.001 493	0.001 515	0.001 49	17	0.004 281	0.004 242	0.004 185	0.004 283	0.004 266
7	0.001 712	0.001 708	0.001 683	0.001 717	0.001 69	18	0.004 579	0.004 547	0.004 499	0.004 581	0.004 603
8	0.001 932	0.001 925	0.001 897	0.001 926	0.001 897	19	0.004 889	0.004 875	0.004 82	0.004 89	0.004 941
9	0.002 157	0.002 163	0.002 111	0.002 149	0.002 117	20	0.005 229	0.005 203	0.005 167	0.005 223	0.005 286
10	0.002 386	0.002 378	0.002 332	0.002 384	0.002 351						

step	6-factor simulated data set					step	6-factor simulated data set				
	1	2	3	4	5		1	2	3	4	5
0	0.002 853	0.022 756	0.002 745	0.002 987	0.002 756	11	0.002 528	0.003 426	0.002 52	0.002 707	0.002 591
1	0.001 952	0.003 903	0.001 895	0.002 039	0.001 936	12	0.002 759	0.003 661	0.002 748	0.002 93	0.002 815
2	0.001 643	0.002 68	0.001 593	0.001 844	0.001 7	13	0.002 993	0.003 898	0.002 985	0.003 169	0.003 047
3	0.001 385	0.002 3	0.001 439	0.001 72	0.001 496	14	0.003 237	0.004 192	0.003 234	0.003 417	0.003 283
4	0.001 344	0.002 34	0.001 357	0.001 713	0.001 431	15	0.003 494	0.004 483	0.003 486	0.003 674	0.003 551
5	0.001 417	0.002 436	0.001 412	0.001 82	0.001 513	16	0.003 768	0.004 768	0.003 758	0.003 94	0.003 827
6	0.001 503	0.002 473	0.001 496	0.002 024	0.001 615	17	0.004 055	0.005 086	0.004 034	0.004 213	0.004 106
7	0.001 694	0.002 637	0.001 687	0.001 871	0.001 757	18	0.004 354	0.005 388	0.004 316	0.004 5	0.004 381
8	0.001 892	0.002 785	0.001 884	0.002 069	0.001 959	19	0.004 67	0.005 716	0.004 611	0.004 801	0.004 676
9	0.002 095	0.002 998	0.002 087	0.002 275	0.002 157	20	0.004 978	0.006 062	0.004 919	0.005 125	0.004 993
10	0.002 305	0.003 198	0.002 303	0.002 485	0.002 37						

In step 0: the average squared partial correlation is calculated in the original correlation matrix. 0In step $i(i>0)$: the first i factors are partialed out of the original correlation matrix.

Table 2 Comparison of the number of factors determined by the four decision rules

Data set		Estimation of number of factors				Difference betweenthe estimation of number of factors with the known number of factors			
		K1	Scree	MAP	PA	K1	Scree	MAP	PA
1-factor	1	19	1	1	6	18	0	0	5
	2	20	1	1	4	19	0	0	3
	3	19	1	1	5	18	0	0	4
	4	19	1	1	3	18	0	0	2
	5	19	1	1	3	18	0	0	2
3-factor	1	22	1	1	6	19	-2	-2	3
	2	23	1	1	9	20	-2	-2	6
	3	21	1	1	9	18	-2	-2	6
	4	21	1	1	7	18	-2	-2	4
	5	21	1	1	6	18	-2	-2	3
6-factor	1	26	6	4	11	20	0	-2	5
	2	25	7	4	15	19	1	-2	9
	3	25	6	4	13	19	0	-2	7
	4	25	7	4	14	19	1	-2	8
	5	25	7	4	11	19	1	-2	5
Mean						18.666 67	-0.466 67	-1.333 33	4.8
S. D.						0.699 206	1.146 977	0.942 809	2.039 608

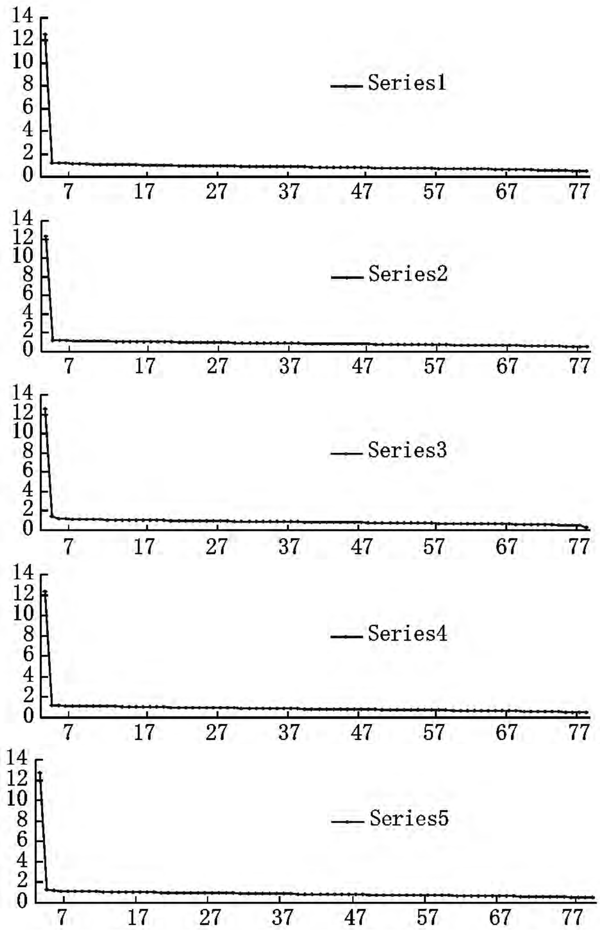

Figure 3　Scree plots for the 5 1-factor simulated data sets

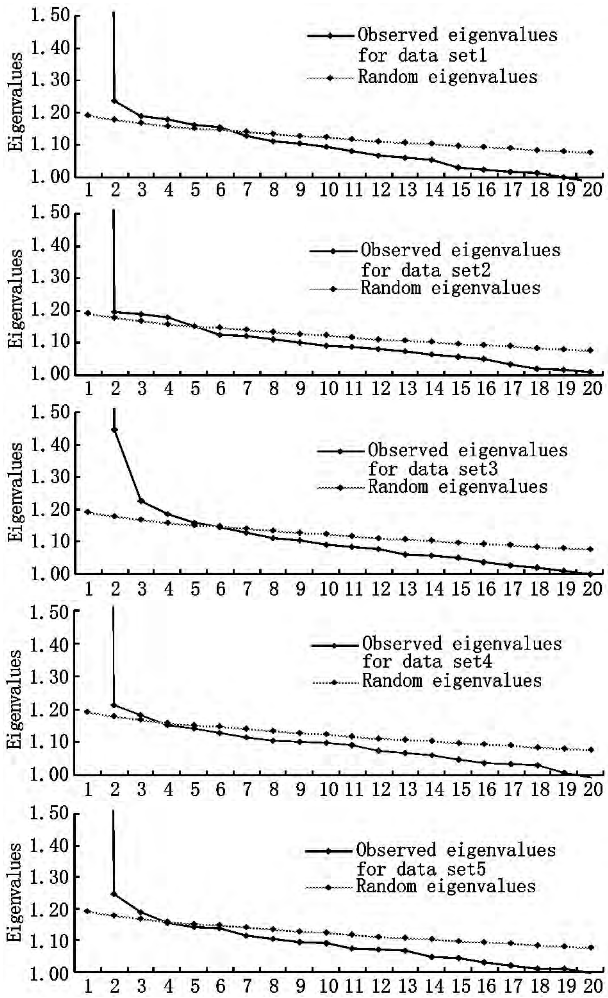

Figure 4　Scree plots for the 5 3-factor simulated data sets

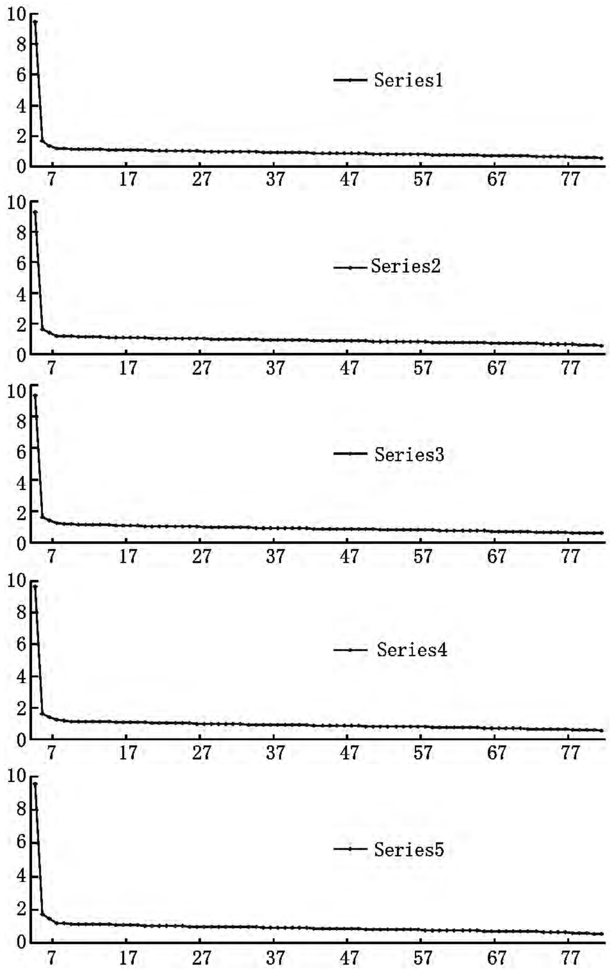

Figure 5　Scree plots for the 5 3-factor simulated data sets

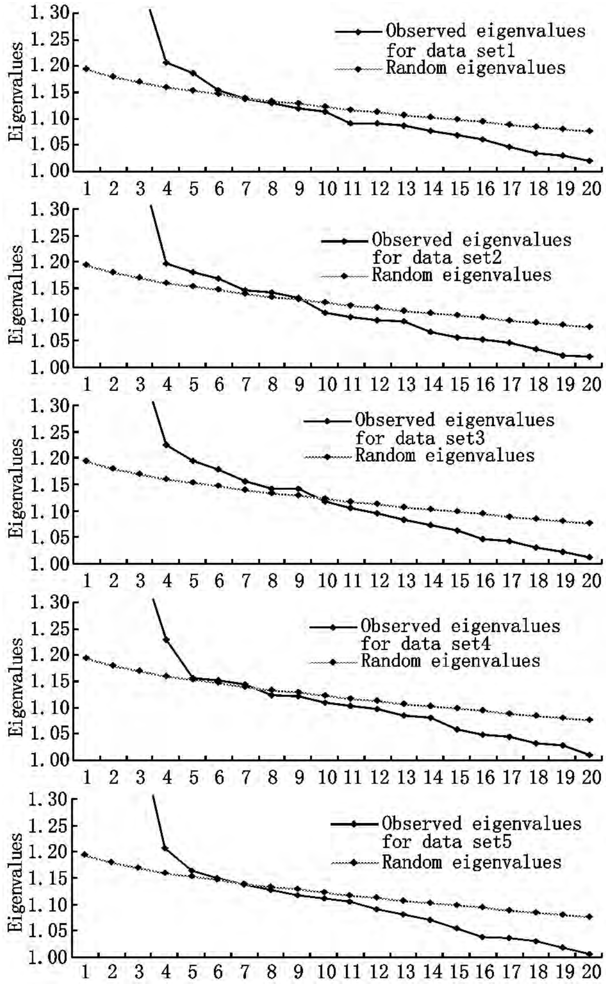

Figure 6　Parallel analysis for the 5 1-factor simulated data sets

Similarly, the scree plots and parallel analyses for the six-factor simulated data sets are presented in Figure 7 and 8 respectively. The scree plots indicate there are 6, 7, 6, 7 and 7 factors respectively in each of the five data sets. Parallel analyses shows there are 11, 15, 13, 14, and 11 observed eigenvalues greater than the corresponding random eigenvalues respectively for the five data sets.

3　Discussion & Conclusions

The purpose of this study is to examine the accuracy and consistency of the four rules (K1, scree test, MAP, and PA) of determining the number of factors to retain in the factor analysis. It should be noted that there are a number of variables that contribute to the inaccuracy and inconsistency of the decision rules. This study focuses on a particular condition of a real data set, which has 78 items, 9 294 examinees, and low factor saturations, and suggests a simulation method of selecting the appropriate decision rule in each particular condition. Simulation study is conducted using the TESTFACT program to simulate data of a known number of factors that compared to the real data on the number of items and examinees, item difficulties, factor loadings, and means of factor scores. Subsequently, the four decision rules are assessed by comparing the rule-determined number of factors with the known number of factors for the simulated data set in the particular condition. The K1 rule consistently overestimates the number of factors to retain. This finding is consistent with that of Zwick and Velicer[9]. The number of factors falls in the one-forth to one-third of the number of items, which is also fundamentally consistent with the finding of Gorsuch[10]. Thus, the K1 rule should be used with caution.

Parallel analysis performs slightly better than K1, but still consistently overestimates the number of factors to retain. This finding is clearly contrary with that of Zwick and Velicer[9] that parallel analysis is consistently accurate. A possible reason for this finding is that the factor saturations of this study are quite low, range from 0.1 to 0.3, whereas the factor saturations in the Zwick and Velicer's study are set at the 0.5 and 0.8 levels. This finding suggests that parallel analysis performs much better when the factor saturations are high than the low level of factor saturations. Thus, the use of parallel analysis should be limited to the case of high and moderate level of factor saturations.

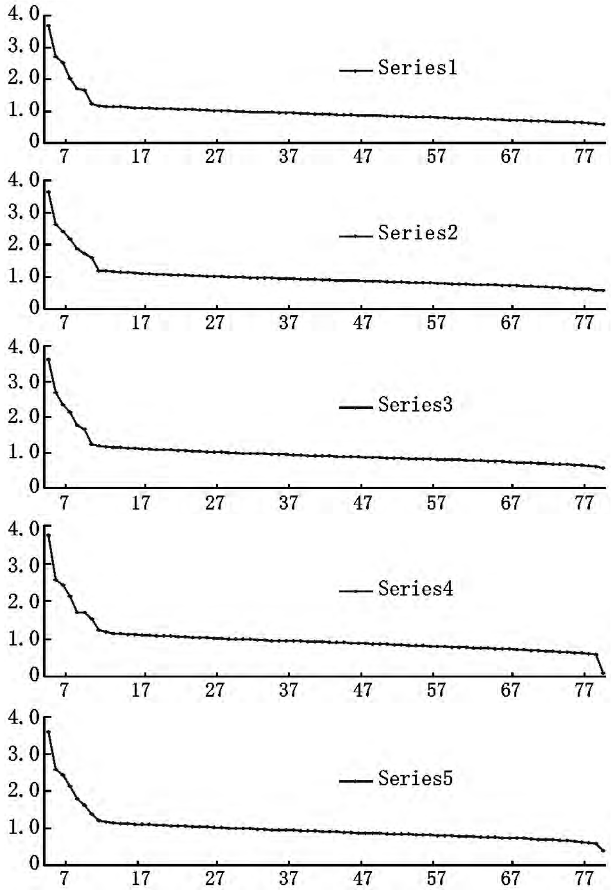

Figure 7　Parallel analysis for the 5 3-factor simulated data sets

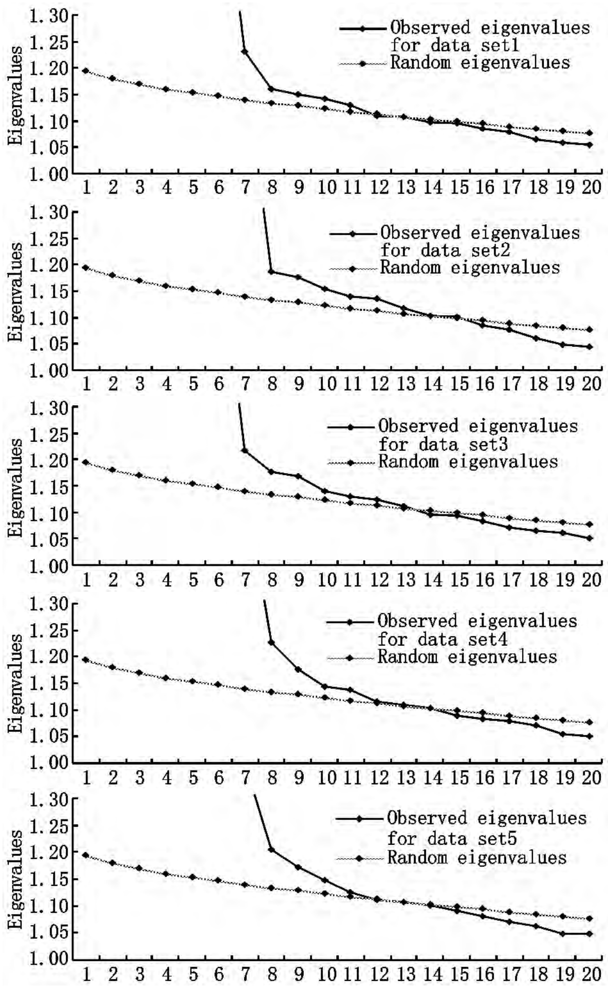

Figure 8　Parallel analysis for the 5 6-factor simulated data sets

On the whole, the MAP method is more accurate than the K1 rule and parallel analysis. It tends to underestimate the number of factors to retain, which is consistent with the findings of Zwick and Velicer.

In general, the Scree test is the most accurate method in this study. Scree test is a relatively easy method compared to the MAP and parallel analysis. However, a drawback in the application of scree test is the subjectivity of the judgement of the number of factors from a scree plot. Zwick and Velicer [7] find the good interrater reliability both among naïve and among judges. It will be very helpful to use several raters to examine the scree plot in order to obtain a more reliable judgement of number of factors in the future study.

In summary, none of the four decision rules is quite accurate under the particular condition of the real data set considered in this study, somewhat due to condition of the low factor saturations. Relatively speaking, the scree test and MAP method are more accurate compared to K1 and parallel analysis.

This study suggests a simulation method for selecting the appropriate rule to decide the number of factors for a specific data set. A critical problem is the necessity of generating a large number of data sets comparable to the real data set on a number of characteristics. The number of items and examinees, item difficulties, factor loadings, and mean factor scores are considered in the process of simulating the data in this study. Further consideration of the comparability of the reliability of the test, the communality and uniqueness of each item, and the number of items having non-zero loadings on each factor should be very helpful in addressing this issue. Future research that more carefully examine the comparability of the simulated data set with the real data set may provide more accurate conclusions.

4 References

[1] Reckase M D. The past and future of multidimensional item response theory [J]. Applied Psychological Measurement,1997,21(1):25-36.
[2] Ackerman T A,Gierl M J,Walker C M. An NCME instructional module on using multidimensional item response theory to evaluate educational and psychological tests [J]. Educational Measurement: Issues and Practice,2003,22:37-57.
[3] Kane M. Validation. Educational measurement [M]. 4th ed. Westport,CT: American Council on Education and Praeger,2006:17-64.
[4] Bernstein I H,Teng Gary. Factoring items and factoring scales are different: Spurious evidence for multidimensionality due to item categorization [J]. Psychological Bulletin,1989,105(3):467-477.
[5] Panter A T,Swygert K A,Dahlstrom W G,et al. Factor analytic approaches to personality item-level data [J]. Journal of Personality Assessment,1997,68(3):561-589.
[6] Mcleod L D,Swygert K A,Tissen D. Factor analysis for items scored in two categories [A] //Thissen D,Wainer H. Test Scoring [M]. Mahwah,NJ: Lawrence Erlbaum,2001:189-216.
[7] Comrey A L. Common methodological problems in factor analysis studies [J]. Journal of Consulting & Clinical Psychology,1978,46(46):648-659.
[8] Kaiser H F. The application of electronic computers to factor analysis [J]. Educational & Psychological Measurement,1960,20(1):141-151.
[9] Zwick W R,Velicer W F. Comparison of five rules for determining the number of components to retain [J]. Psychological Bulletin,1986,99:432-442.
[10] Cattell R B. The scree test for the number of factors [J]. Multivariate Behavioral Research,1966,1(2):245-276.
[11] Velicer W F. Determining the number of components from the matrix of partial correlations [J]. Psychometrika,1976,41(3):321-327.
[12] Horn J L. A rationale and test for the number of factors in factor analysis [J]. Psychometrika,1965,30(2):179-185.
[13] Cota A A,Longman R S,Holden R R,et al. Interpolating 95th percentile eigenvalues from random data: An empirical example [J]. Educational & Psychological Measurement,1993,53(3):583-596.
[14] Glorfeld L W. An improvement on Horn's parallel analysis methodology for selecting the correct number of factors to retain [J]. Educational & Psychological Measurement,1995,55(55):377-393.
[15] O'Connor B P. SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test [J]. Behavior Research Methods,Instruments,& Computers,2000,32(3):396-402.
[16] Gorsuch R L. Factor analysis [M]. 2nd ed. Hillsdale,NJ: Erlbaum,1983.
[17] Zwick W R,Velicer W F. Factors influencing four rules for determining the number of components to retain [J]. Multivariate Behavior Research,1982,17(2):253-269.

用模拟方法确定因子分析中因子数的实证研究

崔 迎
(艾尔伯塔大学教育学院教育心理学系,阿尔伯塔 埃德蒙顿 T6G2G5)

摘要: 在1维或多维项目反应理论中,正确地决定教育评估考试的维度非常重要. 通过分析考试的潜在结构,因子分析可以帮助分析考试的维度. 在因子分析中有多种方法决定因子数目. 但值得注意的是较多变量可以影响这些方法的准确性. 该文提出了一种在因子分析中对特定的数据最准确决定因子数的模拟方法.
关键词: 结构效度; 项目反应理论; 因子分析; 维度

(责任编辑: 冉小晓)