A COMPLETE FACTOR ANALYSIS BY AN INDIRECT METHOD

Harry R. Barker Barbara M. Barker

University of Alabama

A novel, indirect method of factoring data matrices was developed by Horst (1965). The method involves first reducing the variable matrix to a few subsets of variables and then deriving a score (such as a total) from each subset for each subject. (Rules for forming subsets are unspecified). The matrix of intercorrelations between subset totals is factored; and, then, taking into account intercorrelation of subset totals with variables, an estimate is made of the factor structure of the variables. The advantage of this method is that it avoids the direct factoring of the larger data matrix. This permits very rapid computer solution and enormously increases the number of variables which can be factored on a computer. Conventional factor methods permit about 80 to 100 variables which can be factored simultaneously on a computer.

During the past several years the investigators have extensively explored the indirect factor method. A summary report of this research by Barker and Barker (1975) indicated two vital requirements for accuracy of the method:

(1) Subsets of variables must be of homogeneous factor composition.

(2) Variables must enter subset totals with appropriate sign (+ or -).

These findings suggested the total impracticality of the indirect solution. In essence, one had to know the factor structure of the set of variables in order to assign the variables properly into subsets and to assign the correct sign for addition. Subsequent research revealed that the indirect method could be useful in testing theories of the factor structure of large data sets.

Three studies demonstrated the theory testing value of the indirect method. Hamlett (1976) evaluated theories of the personal orientation inventory. Barker and Barker (1976A, 1976B) evaluated competing theories of the factor structure of the MMPI on the original normative sample. Interest centered on the degree of association between subset specification (according to theory) and clusters of items identified by factor analysis. The information measure D was used to quantify the degree of agreement between theory and empirical results. Although the theories of the MMPI were successfully rank ordered according to the D measures, none of the D measures was high thereby indicating none of the theories was adequate.

Further attempts were made to extend the usefulness of the indirect factor method, by using results of the indirect method to refine the variable subsets. Starting with an initial clustering of the variables into subsets, an indirect factor solution was obtained and used for the purpose of reassigning the variables to subsets. A second indirect factor solution was obtained, and further refinement of the variable subsets was attempted. This iterative process was continued until computer time was exhausted or convergence on a stable set of variable subsets was reached. Barker and Barker (1977) tested this refined procedure on several computer generated data sets, which varied in strength of factor structure (strong, moderate and mixed) and obtained excellent results. After a few iterations, indirect solutions were virtually identical to those of conventional factor solutions.

The purpose of this study was to replicate the earlier indirect factor analysis of the MMPI normative data (male and female separately) using the iterative factor method in order to arrive at a definitive factor structure.

Method

The data consisted of item answers on the MMPI of 225 males and 325 females. These subjects were originally used as normative groups for the conventional MMPI clinical scales. The data resided on computer tape and were analyzed separately for male and female.

In an attempt to provide an objective and hopefully satisfactory starting place for the indirect factor method, the following steps were taken:

(1) Twenty subjects were selected randomly and then were used in an obverse factor analysis (CORR98, Barker and Barker, 1977). Eigenvalues were examined by a scree test in order to roughly identify the number of factors to retain. Subsequent varimax rotations were performed on successively fewer principal axes factors until the correct number of factors was identified. A varimax factor load equal to or greater than .3 on only one factor was the criterion used for clustering a

---

*Paper presented at American Statistical Association, August 17, 1977 (Chicago, Ill.). The research was made possible by funds from the Research Grants Committee of the University of Alabama.

variable into a subset. The sign of the factor load was used to identify the manner in which the variable entered into the totalling operation.

(2) The initial subsets of variables assigned by CORR98 were used for the first indirect factor solution (CORR99, Barker and Barker, 1977). The computer program (CORR99) utilized the outcome of the factor solution to reassign items to subsets and continued the iterative process for the specified number of interations or until convergence was reached. Convergence was defined as two consecutive identical factor solutions.

The computer program (CORR99) was modified to compute the information measure D between location of items in subsets and in factor clusters for each iteration.

The similarity of factor solutions for male and female was evaluated by (CORR22, Barker and Barker, 1977). This computer program rotates one factor solution to another, and determines degree of contiguity of variables attained in factor space.

Results

The obverse factor solutions, (CORR98) for the male and female samples suggested seven factors in each data set. Clustering of variables on the varimax rotated factors was used to identify the initial subsets of variables and sign for totaling operation for the indirect factor analysis.

Twenty-five iterations of the indirect factor solution were performed on both the male and female data sets. In neither case was convergence between variable subsets and factor clusters attained. Failure of convergence suggests that the factor structure of the data sets is quite weak. Earlier research reported by Barker and Barker (1977) noted that the weaker the factor structure of the data, the greater the number of iterations required to attain convergence.

The twenty-fifth iteration produced a D measure between variable subsets and obtained variable clusters of .60 for males and .70 for females. Although the D measures were lower than expected, the obtained D measures exceed those earlier obtained when evaluating competing theories of the factor structure of the MMPI. Therefore, this solution appears superior to those earlier attained.

The original estimate of seven factors for the males was supported whereas only four factors were retained for the females. Tables 1 and 2 identify items associated with each of the factors for male and female. In order for an item to be identified with a factor, a varimax factor load of .3 or greater on only one factor was required. For males, a preliminary labeling of factors is as follows:

(1) General mental health
(2) Religious
(3) Adventurous, independent
(4) Moralistic
(5) Unclear-items range from fear of catching disease to awareness of ears ringing and dreaming
(6) Neurotic
(7) Phobias, Anxiety

Factor 3 and 5 might well be dropped because they contain so few items.

Suggested labels for the female factors are:

(1) General mental health
(2) Moralistic
(3) Neurotic
(4) Phobias, anxiety

Factor 4 contains only 6 items and could be dropped.

Rotation of the varimax factor structure for females to maximum contiguity with the varimax factor structure of the males resulted in relatively good factor alignment; however item locations in factor space were quite separated. This further supports the apparant lack of similarity in male and female factor structure.

Data were processed on a UNIVAC 1110 system with 128K core allocation. The required computer time and costs for the indirect solutions were as follows:

(1) The male data required a computer run time of 25 min. 50 sec. and cost $176.42.

(2) The female data required about 34 min. at a cost in excess of $210.00.

Discussion

It appears that a weak factor structure characterizes the original MMPI normative data for both male and female. In view of the decided slant of MMPI items towards measuring pathology and the alleged normalcy of the subject samples, this is to be expected. A finding of considerable interest is the difference in factor structure for male and female.

In interpreting the findings of the study several cautions should be observed. The

data were obtained in 1957, on principally Mid-Western rural subjects. Cultural biases are very likely reflected in the data. The ratio of subjects to variables is grossly inadequate according to several criteria for multivariate work. For example, Cattel's rule which suggests that the number of subjects equal or exceed the number of variables by 100 is far from met. The data set would be of less interest except that the standard norms for the MMPI were obtained on these two samples. These norms have remained unchanged throughout the test's history.

Considering the large number of items on the MMPI which are not scored on the regular clinical scales, it was anticipated that additional useful factors might emerge which could be used to measure dimensions among normals. The obtained results support this view.

Summary

Application of a refined version of Horst's indirect factor method to the original normative data of the MMPI disclosed weak factor structure for both male and female. Seven factors for male and 4 factors for females were extracted and rotated to a varimax criterion. Factors extracted for male did not closely resemble factors extracted for female.

References

Barker, H. R. and Barker, B. M. Behavioral Sciences Statistics Program Library. University of Alabama, 2nd. rev. edition, 1977.

Barker, B. M. and Barker, H. R. Evaluation of theorized factor structure of the MMPI for male and female populations. Proceedings of the Social Statistics Section, American Statistical Association, 1976A, 174-178.

Barker, H. R. and Barker, B. M. An indirect method for testing the dimensionality of large data sets. Proceedings of the Social Statistics Section, American Statistical Association, 1976B, 179-183.

Barker, H. R., and Barker, B. M. Summary of research on a novel, indirect factor analytic solution. Proceedings of the Social Statistics Section, American Statistical Association, 1975, 298-301.

Barker, H. R. and Barker, B. M. Further refinement of an indirect method for factoring large data sets. Paper presented at Southeastern Psychological Association, Hollywood, Fla., May 6, 1977.

Hamlett, C. C. Validation of the theorized factor structure of the personal orientation inventory. Proceedings of the Social Statistics Section, American Statistical Association, 1976, 366-370.

Hathaway, S. R. and Briggs, P. F. Some normative data on new MMPI scales. Journal of Clinical Psychology 1957, 13, 364-368.

Horst, P. Factor analysis of data matrices. New York: Holt, 1965.

Table 1
Identification of Item Numbers with
Seven MMPI Factor Scales (Male)
(MMPI Normative Data)

I						II	III	IV	V	VI	VII
13	123	244	328	375	506	46	257	6	37	2	3
15	129	248	331	381	507	50	289	30	131	20	7
24	136	252	332	382	509	58	497	45	281	60	8
27	138	259	333	383	511	98	501	90	302	114	9
28	139	266	337	384	517	115	502	95	329	119	36
32	146	269	338	386	518	249		111	432	130	79
33	147	275	339	388	526	287		118	524	152	91
34	158	278	341	389	530	483		135		161	153
35	162	280	343	390	531	558		198		174	160
40	171	282	345	392	543			215		190	163
41	172	284	346	395	551			225		214	170
42	182	290	348	397	553			231		242	176
44	184	291	349	398				255		310	178
48	191	292	350	404				427		330	188
61	194	293	351	406				430		533	230
66	197	296	352	418				446		540	243
82	200	299	354	426				488			318
84	202	301	355	438				490			353
85	205	303	357	442				548			367
97	210	305	358	448							379
100	224	307	359	469							399
104	226	312	365	472							401
109	234	314	366	473							412
117	239	323	368	499							479
121	241	325	374	505							521
											522

Table 2
Identification of Item Numbers with
Four MMPI Factor Scales (Female)
(MMPI Normative Data)

I						II	III	IV
3	129	247	335	385	506	2	37	170
5	136	251	337	388	509	21	55	348
13	138	252	338	389	511	30	60	429
15	142	259	340	390	525	45	68	521
16	147	265	343	392	526	80	103	534
22	148	266	344	396	530	99	130	546
24	157	267	345	397	531	111	133	
29	158	273	350	398	535	135	153	
32	163	278	351	407	543	181	154	
40	166	284	352	411	551	208	175	
41	171	290	354	414	553	231	187	
43	172	292	356	416	560	285	193	
62	179	301	357	418	564	308	214	
67	186	303	359	421		378	281	
72	189	305	360	439		391	294	
76	190	307	361	442		427	302	
82	191	314	362	443		446	330	
84	201	315	366	448		452	460	
86	214	316	368	465		457	464	
94	217	317	374	468		481	478	
106	224	321	375	475		490	486	
108	234	322	377	487		527	496	
109	241	326	382	489		548	540	
117	244	328	383	492				
120	245	333	384	499				