EMPIRICAL STATISTICAL SAMPLING DISTRIBUTIONS OF SOME INDICES OF THE NUMBER OF FACTORS TO CONSIDER IN EXPLORATORY FACTOR ANALYSIS

by

PENELOPE JOHNSTON

B. A., University of British Columbia, 1966

A THESIS SUBMITTED IN PARTIAL FULFILLMENT

OF THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF ARTS

in the Department

of

Psychology

SIMON FRASER UNIVERSITY

April, 1970

APPROVAL

Name: Penelope Johnston

Degree: Master of Arts

Title of Thesis: Empirical statistical sampling distributions of some indices of the number of factors to consider in exploratory factor analysis.

Examining Committee:

(C. B. Crawford)
Senior Supervisor

(R. E. Cole)
Examining Committee

(A. H. Elliott)
External Examiner
Assistant Professor
Simon Fraser University
Burnaby, B. C.

Date Approved: April 30, 1970

Abstract

Empirical statistical sampling distributions of the estimates of six number-of-factor indices were generated by Monte Carlo procedures and were inspected for deviation from a known population number of factors. The distributions were based on 100 samples (at each of 3 sample sizes) from Cattell and Sullivan's Cups of Coffee Problem - Sample A (which has psychometric error comparable to that found in the typical psychological study using factor analytic techniques).

Cattell's scree test and Linn's mean square ratio test were implemented too unreliably to be included in the analysis of sampling distributions. Horn's index (following smc and image analyses) and Crawford's index failed to indicate any number of factors for some samples prior to modifications made for the purpose of including them in the analysis of sampling distributions. Sequential analyses of variance indicated that Guttman's Stronger Lower Bound consistently overestimated the population number of factors and that Guttman's Weaker Lower Bound and the modified version of Crawford's index gave the most accurate estimates of the population number of factors.

Table of Contents

List of Tables

List of Figures

Introduction

The Number of Factors Problem

Factor analytic procedures are designed to select a small number of interpretable parameters which best describe a set of variables. Data are collected on a set of observed variables which are thought to relate to the conceptual domain under consideration and then are subjected to procedures which result in the expression of the variables as linear combinations of a smaller set of new hypothetical variables or common factors. These factors represent the aspects of each variable which are common to others in the set and, thus, give a more concise description of the nature of the domain of interest as represented by the variables initially selected.

Generally, the number of common factors influencing a set of variables exceeds the number of variables included in the analysis. In practice, however, it is impossible to extract more common factors than there are variables without placing arbitrary restrictions on the factor matrix (Jöreskog, 1969). Thus, many factor analysts have expanded the concept of common factors to include the following two types: (a) major common factors (which are separable from error-dominated factors in any particular factor analysis), and (b) minor common factors (which are not distinguishable from error-dominated factors in any particular factor analysis). The experimenter must select the factors which best describe his data (viz, the major common factors) while minimizing the inclusion of error variance which is concentrated in the minor common factors.

The selection of the most appropriate factors should be made when the factors are in their most interpretable positions. A factor solution following extraction is only one of an infinite number of mathematically equivalent solutions. Some of these solutions are more psychologically meaningful than others, and very seldom is it the case that the initial solution is the most interpretable. The most meaningful solution can be found by rotating the factors to a new position. The most widely accepted criteria for such solutions (Harman, 1960) are based on Thurstone's (1947) conditions of simple structure. If a simple structure solution exists for a set of data, the expectation of factor stability is also maximized (Horst, 1965). However, most rotation procedures require a prior, independent estimate of the number of factors to be rotated since the rotated factor loadings are dependent on the number of factors rotated (Cliff and Hamburger, 1967). The inclusion of too few factors in the rotation results in a loss of potentially interpretable factors. The inclusion of too many factors in oblique rotations often causes a collapse in the factor space; whereas inclusion of too many factors in orthogonal rotations has the effect of splitting factors, resulting in pseudospecific factors (Cattell, 1966b).

Several general procedures have been developed to estimate the number of factors on the basis of some criterion which is independent of the psychological meaningfulness of the factors. Depending on the type of extraction employed, different indices have been applied to the unrotated factor matrix to exclude from the rotation factors which, on the basis of either their psychometric or their statistical properties, are considered likely to be error factors.

Statistical and Psychometric Approaches to the Problem

The statistical and psychometric approaches to the number-of-factors problem each emphasizes a different source of error variance as the determinant of factor instability. The statistical approach involves the problem of faulty estimation of factors due to the errors arising from the random selection of a certain sample of persons from a total population. Factors derived from a population of variables should be replicable when another sample from the same population of people is used to obtain measures on the variables. Statistical indices of the number of factors (k) actually test the equality of the last (n - k) roots of an (n x n) covariance (or correlation) matrix on the assumptions: (a) that if the roots are very close to equality they may come from a population in which the roots are equal; and (b) that if this is the case, then there is no point in attempting maximization of variation in any particular direction, and k, then, represents a sufficient number of factors. Statistical indices of the number of factors have been developed by Lawley (1940), Rao (1955), Jöreskog (1963, 1969), and others.

The psychometric approach involves the problem of error in the estimation of factors arising from the selection of only some variables from the total domain of variables defined by a set of factors. Thurstone (1947) describes this type of invariance as the invariance of the factorial description of a variable when it is moved from one battery of variables to another which involves the same common factors. The conditions of simple structure are associated with this type of invariance. Many indices have been devised to indicate the number of factors which will most likely satisfy some criterion of this type of invariance, e.g., Guttman (1954), and Kaiser (1960, Kaiser and Caffrey, 1965). These indices assume, however, that the data are based on a population of people.

Any particular factor analysis includes both psychometric and statistical error. Neither the statistical nor the psychometric tests, however, checks for both types of error. Such tests must either (a) assume that the analysis includes a population of variables, and test for errors resulting from the sampling of subjects from a total population of subjects (viz, statistical indices); or (b) assume that the analysis is based on a population of people, and test for error resulting from the sampling of variables from a total domain of variables (viz, psychometric indices). Thus, it is of interest to anyone using one type of test to know how it will behave under conditions in which its assumptions are not met.

It has been found that when the psychometric assumption is met, statistical indices tend to improve their estimation of the number of major common factors as statistical error is reduced; but that when the psychometric assumption is not met, they are apt to indicate increasing numbers of factors (up to $n - 1$, where n is the number of variables included in the analysis) as statistical error is reduced (Browne, 1969). This tendency indicates that statistical tests are apt to take advantage of psychometric sampling error when it is present in the data.

When there is no statistical error in the data, psychometric tests tend to indicate an appropriate number of factors. However, in the presence of statistical sampling error, the effect of sample size on the statistical sampling distribution of psychometric indices is not clear.

Besides the purely statistical and purely psychometric indices of the number of factors, several others have been developed which estimate the number of factors on the basis of some criterion of identifiability of major common as opposed to minor common factors. These tests, which will be termed composite tests, do not include explicit assumptions about the psychometric or statistical properties of the data.

The procedures of several of these composite tests are similar to the statistical tests' procedure of rejecting factors on the basis of the equality of their roots. Cattell's (1966a) scree test and Linn's (1964, 1968) mean square ratio test are of this type. They depart from the statistical tests in their lack of imposition of a statistical test of equality on the latent roots of the minor common factors. Instead, they implement a visual test on the plot of some index of factor size which discriminates between the major and minor common factors. The scree test assumes (a) that the error factors should have an eigenvalue plot which can be described by a straight line, even if this line has a non-zero slope; and (b) that the eigenvalues of the major common factors should not lie on this straight line. The Linn test, described in detail later, involves a plot in which the mean square ratios (indices of factor size) of the minor common factors are assumed to be described by a curve which does not describe similar indices of the major common factors. This differential description is manifest in a break in the mean square ratio plot between the major and minor common factors. Thus, these tests are like statistical tests in their assumption of similarity between error factors, but are unlike the statistical tests in that they are not based on the sampling distribution of the hypothetical error factors.

Another semi-statistical test, Horn's (1965) test, also involves plots of eigenvalues. In this test, however, the eigenvalues of the observed correlation matrix are compared with those of a correlation matrix based on an equal amount of random normal data. The eigenvalues of each matrix are sequenced in descending order and equivalently positioned eigenvalues are compared. Minor common factors are considered to be those factors in the observed data whose eigenvalues are smaller than corresponding eigenvalues of factors in the random matrix.

Crawford's (1966) index, which is described in detail later, is an example of a psychometrically oriented composite test. This test, which is applied after rotation, estimates the number of major common factors to be that number which best satisfies an analytic criterion of simple structure. This test is a psychometric test only insofar as (a) the criterion used is a complete index of Thurstone's (1947) requirements of simple structure, and (b) simple structure provides psychometric invariance.

As with the pure psychometric tests, the behaviour of these composite tests under conditions where different amounts of statistical error are present is not clear. Their behaviour under conditions where varying amounts of psychometric error is present is likewise unclear and is, at least, of equal interest. However, due to problems in realistically sampling degrees of psychometric error, the present study is limited to an examination of the empirical statistical sampling distributions of some purely psychometric indices and some of the composite indices of the number of factors.

The Scope of the Study

The main purpose of the present study is to examine empirical statistical sampling distributions of several commonly used and recently developed non-statistical indices of the number of factors. These tests will be observed as they are applied to 3 sets of 100 random samples of subjects (as generated by standard Monte Carlo procedures) from a given population correlation matrix for which the number of major common factors is known. The study will examine the behaviour of these indices under simulated typical rather than ideal sampling conditions; e.g., the three sets of samples will differ in sample size and the sample sizes will be typical rather than ideal, and the population matrix will include a reasonable amount of psychometric error. (Since the statistical sampling distributions of factor matrices which include psychometric error are not readily calculable, the Monte Carlo procedures were used to generate empirical statistical sampling distributions.) The study will emphasize aspects important to the application of these indices by most psychologists engaging in exploratory factor analysis.

The sampling distributions of the indices will be considered with respect to their deviation from the population number of factors. The extent to which the indices deviate from this population value provides evidence on their usefulness whenever statistical inference is included explicitly or implicitly in the analysis.

A secondary purpose, which developed from the main purpose, is an examination of the inter-rater reliability of the two visual tests, the scree test and the mean square ratio test, mentioned above. In order to find typical results of these tests for each sample, they will be implemented by five raters. Inter-rater reliabilities will be found and if the indices prove to be significantly unreliable, their sampling distributions will not be analysed since reliable measures of central tendency would be required for the examination of the statistical sampling distributions of these two indices. Results of unreliability provide evidence on the utility of these tests to the relatively unsophisticated psychologist who uses factor analytic techniques.

Procedure

Indices Considered in the Study

One hundred samples at each of three sample-size levels were generated by Monte Carlo procedures (which are described in detail later) and the number of factors estimated by each index for each sample was recorded as the basis for the indices' sampling distributions. The specific indices considered and the specific estimate of the number of factors collected on each sample for each index are as follows.

Guttman's Stronger Lower Bound

Guttman (1954) shows that the minimum number of common factors which can exactly account for the off-diagonal elements of a population correlation matrix, R, equals the minimum rank of a reduced Gramian correlation matrix,

$$G = R - U^2, \tag{1}$$

where U^2 is a diagonal matrix with $0 \leq u_j^2 \leq 1$; ($j = 1,2,\ldots,n$; where n is the number of variables). He goes on to show that if squared multiple correlations (smc's) are used as estimates of the communalities in the analysis, since each smc, r^2, is the lower bound for the communality of its corresponding variable, i.e.,

$$r_j^2 \leq h_j^2 = 1 - u_j^2 \; ; \; (j = 1,2,\ldots,n); \tag{2}$$

then a criterion for the minimum rank of the correlation matrix is the number of factors whose eigenvalues are greater than zero.

Guttman's Weaker Lower Bound

If variables are standardized and a factor analysis is implemented with unities in the diagonal of the correlation matrix (i.e., a principal component analysis), since unity is an upper bound estimate of the communalities, Guttman (1954) shows that the lower bound for the rank of a reduced Gramian correlation matrix,

$$G = R - U^2 \tag{3}$$

where

$$1 \leq h_j^2 = 1 - u_j^2 \tag{4}$$

and

$$0 \leq u_j^2 \leq 1 \; ; \; (j = 1,2,\ldots,n), \tag{5}$$

is the number of eigenvalues of R which are greater than unity. Kaiser (1960; Kaiser and Caffrey, 1965) proved that any factor whose eigenvalue is less than one (following a principal components extraction) has negative Kuder-Richardson-20 internal consistency (or alpha reliability) and argued that it should, therefore, be eliminated from the analysis. He also noted that in a great many cases, this criterion specifies the same number of factors as experimenters have found to be interpretable.

Cattell's Scree Test

This test (Cattell, 1966a) follows the factoring of a correlation matrix with unities in the diagonal and involves the extraction of as many factors as there are variables. Rejection of factors is based on the assumption that any factors with proportionally small eigenvalues are error factors and account only for trivial variance. Error factors with large eigenvalues before rotation, which this test would not exclude from the rotation procedures, are assumed to become apparent after the rotation since the error variance may be separated from the non-error variance by the rotation procedure. Plots of the eigenvalues of principal components typically show an initial curvature and a later linearity of slope. Cattell (1966a, 1966b) notes that when the ratio of the number of people to the number of variables is small, the latter part of the slope may be resolved into two or three straight lines, none of which describes the eigenvalues of the initial factors. The scree test rejects those factors which lie on the scree, or linear part (or parts, in the cases of double or triple screes) of the slope.

As the determination of the scree is a subjective operation, five raters (two undergraduate and three graduate students) who had no knowledge of the number of factors in the population, inspected the sample plots of the principal component eigenvalues for the scree. Following a test of inter-rater reliability, if the results proved to be significantly reliable, the modal number of factors for each sample was taken to be the typical estimate of the number of factors using the scree test. If the curve was bimodal, the smaller number of factors was to be taken under the assumption that it represented the top of a double scree missed by the raters who gave the larger estimate. In cases where all five raters responded differently, the index was considered to have failed to indicate the number of factors.

Horn's Test

Horn (1965) proposes comparing the eigenvalues of an observed correlation matrix to the averaged eigenvalues of a large number of random correlation matrices. The random correlation matrices are derived from random normal data matrices of the same sample size as the observed data matrix and are of the same order as the observed correlation matrix. The point at which the plots of the two sets of eigenvalues cross is taken to be the cutting point for the number of factors. Horn (1965) used this procedure following a complete component analysis but used only a single random correlation matrix. Humphreys and Ilgen (1969) empirically showed that this test tends to indicate too few factors following principal components extraction; but that results closer to the maximum likelihood estimates may be obtained following principal axis extraction with smc's rather than unities, as communality estimates.

Comparison with one random sample. For each sample generated, a sample corresponding in sample size but from an identity population matrix was also generated. Each pair of correlation matrices was subjected to principal components, smc, image covariance (Harris, 1962; Kaiser, 1963), and Harris (Harris, 1962) extractions; and each of the plots of the four sets of eigenvalue-pairs was inspected for crosses. The number of eigenvalues preceding the first cross of each plot was considered to equal the number of factors indicated for that particular extraction method for that sample. In cases where the plots did not cross, the index was considered to indicate the number of factors to equal the number of variables.

Comparison with 100 random samples. Corresponding eigenvalues for the 100 samples from the identity matrix for each extraction procedure and sample size were averaged. The eigenvalues of the 100 samples from the structured population matrix (whose selection is described in detail later) of the same sample size and type of extraction were compared with the averaged eigenvalues for crosses. As above, the first cross was taken to indicate the number of factors for that particular extraction method for that sample; and where the plots did not cross, the index was considered to indicate the number of factors to be equal to the number of variables.

Linn's Test

To obtain an estimate of the amount of error in an observed correlation matrix, this test (Linn, 1964, 1968) uses a correlation matrix based on a random normal data matrix of the same sample size as the observed data matrix. In this test, however, the random matrix is used to augment the observed matrix before factoring. The augmented matrix includes the observed intercorrelations, the appended intercorrelations, and the cross-correlations between the observed and appended variables. The extent to which the appended matrix deviates from an identity matrix provides an estimate of the error variance in the factor loading matrix. Linn (1968) did principal axis factoring of (a) the augmented correlation matrix, R_{aug}, with unities in the diagonal, (b) the augmented correlation matrix, R_{aug}, with smc's in the diagonal, and (c) B - I, where $B = S^{-1}R_{aug}S^{-1}$, where $S^2 = (\mathrm{diag}(R_{aug}^{-1}))^{-1}$; i.e., Harris (1962) rescaled factoring. He then computed a mean square ratio for each factor in the factor loading matrix, A_{aug}, such that for each factor, the

$$\text{Mean Square Ratio}_k = \frac{\dfrac{\sum_{i=1}^{n} a_{ik}^2}{n}}{\dfrac{\sum_{i=n+1}^{n+m} a_{ik}^2}{m}} \qquad (6)$$

where n is the number of observed variables, m is the number of appended variables, and a_{ik} is the principal axis factor loading of the i^{th} variable of the augmented correlation matrix of order (n + m) on the k^{th} factor. The mean square ratios are then plotted and rejection of factors is made on the basis of a break in the curve of the plot. Linn (1964) found that the best choice for m was n/2 since (a) the inclusion of too many appended variables obscures the break in the mean square ratio plot; and (b) the inclusion of too few appended variables tends to give unstable results.

The mean square ratios were found for each of four extractions (principal components, smc, image covariance, and Harris) of each of the one hundred samples from the three sample sizes. Since this index, like the scree test, requires subjective operations, five new raters (two undergraduate and three graduate students), without knowledge of the population value of the number of factors, inspected plots of the mean square ratios for breaks. They had the aid of plots of the first order differences between the mean square ratios of successive factors to assist in making their decisions. Before plotting the ratios, the factors were reordered according to the numerators of the ratios. The number of factors was indicated by the number of mean square ratios before the break. As with the scree test, following a test of inter-rater reliability, if the results proved to be significantly reliable, the modal number of factors was considered to be the typical number indicated by this index. If the curve was bimodal, the larger number was considered to be the more conservative estimate (since the effects of underfactoring are generally considered worse than those of overfactoring) and thus, was taken to be the number of factors indicated. In cases where all five raters responded differently, the index was deemed to have failed to indicate the number of factors.

Crawford's Test

This test has been devised (Crawford, 1966) to provide an index of the interpretability of factors rather than of their significance, It is applied to the row-normalized factor loading matrix following any adequate method of orthogonal rotation and provides an index of combined test and factor parsimony (or simplicity of structure). If

$$T(m) = (n - 1) \sum_{i}^{n} \sum_{p}^{m} \sum_{\substack{q \\ p \neq q}}^{m} a_{ip}^{2} a_{iq}^{2} \tag{7}$$

and

$$F(m) = (m - 1) \sum_{p}^{m} \sum_{i}^{n} \sum_{\substack{j \\ i \neq j}}^{n} a_{ip}^{2} a_{jp}^{2} \tag{8}$$

where p and q refer to factors, i and j refer to variables, and a is a factor loading, then when the sequence of sums

$$I(m) = T(m) + F(m); \ (m = 2,3,\ldots,k; \ k \leq n) \tag{9}$$

reaches a definite minimum, maximum parsimony and interpretability are reached.

Crawford's index of interpretability function was found following quartimax (Carroll, 1953), varimax (Kaiser, 1958), and equamax (Saunders, 1962) rotations of each of 2 to n/2 smc factors. The rotation procedure used required that at least 2 factors be rotated and since there are seldom more than n /2 interpretable common factors for any set of n variables, that particular range was selected for the rotations to the three criteria. For each sample, the three sets of (n/2 - 1) indices were inspected for minima. The number of factors rotated to generate the minimum value of the index for each type of rotation was considered to be the number of factors indicated by this index for each rotation of each sample.

Table 1 indicates the types of extraction and rotation which preceded the application of the number-of-factor indices under consideration.

Monte Carlo Methods

Procedures are available by which correlation matrices can be generated from a population matrix. These procedures eliminate the necessity of generating random normal sample data matrices which are then correlated, thereby reducing computation time to a small fraction of that required by the longer procedure.

The basic equation in factor analysis,

$$Z = AP, \tag{10}$$

can be converted to

$$R = ZZ' = \frac{1}{N}APP'A' \tag{11}$$

where Z is the standardized score matrix, A is the common factor loading matrix, P is the common factor score matrix, R is the correlation matrix, N is the number of people, and ' denotes a matrix transpose. Since, for orthogonal factors, the expected value of $\frac{1}{N}PP'$ is an identity matrix, the latter equation can be reduced to its more familiar form

$$R = AA'. \tag{12}$$

When each sample correlation matrix, R*, is generated, deviations from this identity provide sampling deviations in the sample correlation and factor matrices such that

$$R^* = DAFA'D \tag{13}$$

where F is the sample factor covariance matrix and D is the diagonal matrix used to standardize the variances of the sample variables.

TABLE 1

Application of the Indices to Their Respective Types of Extraction and Rotation*

TYPE OF EXTRACTION	TYPE OF ROTATION	INDEX
Principal Components	none	Cattell
	none	Guttman's W. L. B.
	none	Horn
	none	Linn
SMC Analysis	none	Guttman's S. L. B.
	none	Horn
	none	Linn
	Quartimax	Crawford
	Varimax	Crawford
	Equamax	Crawford
Image Analysis	none	Horn
	none	Linn
Harris Analysis	none	Horn
	none	Linn

* This procedure was followed for each of the 100 samples at each of the 3 sample size levels.

Such sampling procedures have been used by Hamburger (1965), Browne (1968), Cliff and Pennell (1967), Linn (1968), and others in studying such sampling problems as rotational stability and appropriateness, criteria for the number of factors, methods of factor extraction, stability of factor loadings, etc.

Browne's (1968) method of generating sample correlation matrices was used in the present study. The procedure involves the square root factoring of the population correlation matrix, R, into AA' where A is lower triangular. A random matrix, T, is generated where

t_{ij} = random elements distributed as N(0,1), $i > j$;

t_{ii} = random elements distributed as Chi with $(N - i)$ degrees of freedom; and

$t_{ij} = 0$, $i < j$.

The sample covariance matrix, C, is then produced such that

$$C = ATT'A' \tag{14}$$

which is then rescaled to a sample correlation matrix, R*, such that

$$R^* = (\text{diag } C)^{-\frac{1}{2}} C (\text{diag } C)^{-\frac{1}{2}}. \tag{15}$$

In the present study this method was also employed to generate the random matrices needed for Horn's and Linn's indices.

The required normal and Chi elements of the T matrix were obtained as follows. Uniform random numbers on the interval $0 < u < 1$ were generated by means of the Tausworthe pseudorandom number generator which Whittlesey (1968) has shown to be appropriate for the IBM 360 computer used in this study. Each uniform random number was then normalized by the function,

$$z = \frac{\log_e \frac{1-u}{u}}{\pi / \sqrt{3}}, \tag{16}$$

which is approximately distributed N(0,1). The square root of squares of k such random normal elements was then used to generate the Chi elements with k degrees of freedom.

Selection of a Population Correlation Matrix

The population correlation matrix selected for the present study is that from Cattell and Sullivan's (1962) Cups of Coffee Problem - Sample A (N = 80, n = 15). The correlation matrix is presented in Table 2. In selecting a population correlation matrix for the present study, it was necessary to choose one for which the number of major common factors was known and in which there was psychometric error comparable with that found in typical exploratory psychological factor analyses. The investigators in the above study hypothesized a factor matrix and selected variables with hypothesized factor loadings on the factor matrix. The resulting factor loading matrix (based on 80 measures of each of 15 variables) closely approximated the hypothesized matrix but displayed error comparable to that found in factor analyses of psychological data. This matrix is well suited to orthogonal transformations (which is of importance in the study of Crawford's index) and has a rather large ratio of the number of major common factors (5) to the number of variables (15) (which provides a wider opportunity for a large number of factor comparisons with a minimum of computation time).

Selection of Sample Sizes

To test the effects of sample size on the statistical sampling dis-

TABLE 2

Population Correlation Matrix *'

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	1.																						
2	91	1.																					
3	88	94	1.																				
4	-07	03	03	1.																			
5	-14	-04	-13	54	1.																		
6	-12	-03	-05	30	60	1.																	
7	07	10	10	-21	-04	07	1.																
8	02	-03	-07	-33	-21	-07	44	1.															
9	-05	-05	-08	15	14	-10	-72	-46	1.														
10	-26	-25	-18	24	-07	02	-08	01	02	1.													
11	-15	-15	-23	03	-09	04	-18	-04	05	54	1.												
12	04	08	09	-03	12	-10	08	-19	-03	-71	-56	1.											
13	-01	03	04	16	43	95	02	-05	-09	-04	03	-12	1.										
14	36	38	37	-04	-12	-22	-04	10	17	07	18	-17	-18	1.									
15	03	-07	-15	28	36	27	08	06	01	06	03	-05	21	05	1.								
16	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	1.							
17	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	1.						
18	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	1.					
19	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	1.				
20	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	1.			
21	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	1.		
22	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	1.	
23	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	00	1.

*The first 15 rows are taken from Cattell and Sullivan's (1962) Cups of Coffee Problem; the last 8 rows are the identity matrix needed for Linn's test.

'Correlations of less than 1.0 are entered without the decimal point.

tributions of the six indices under consideration, sample sizes of 2n, 4n, and 8n (n = 15) were chosen to generate sample matrices by means of Browne's (1968) procedure. These sample sizes were chosen because they approximate those used in most factor analyses of psychological data. They represent typical, rather than ideal, sampling conditions. Most experimenters have used very large sample sizes to investigate the psychometric stability of number of factors indices (e.g., Humphreys, 1969; Linn 1964, 1968). Browne's (1968) study of the statistical sampling distributions of several indices of the number of factors also used large sample sizes (viz, 6n, 12.5n, and 125n). The sample sizes in these studies are closer to the ideal than the typical factor analysis conditions.

As statistical error variance is proportional to the reciprocal of sample size, the selection of 2n, 4n, and 8n allows a linear comparison of expected error variances; i.e., the expected error variance of the 2n group is twice that of the 4n and four times that of the 8n groups, etc. Use of a sample size of 2n presents severe problems for replication, while use of 8n gives reasonably stable results in the replication situation.

Data Analysis and Results

Reliability of the Scree Test and Mean Square Ratio Test

As both the scree test and mean square ratio test involve some degree of subjective interpretation of plots, these indices were checked for their inter-rater reliability prior to inspection of their sampling distributions on the assumption that if these results were significantly unreliable, there was no justification for choosing the modal value for each sample above any other value. Inter-rater reliability can be described by either the specific or generic reliability statistic (Lord and Novick, Chapter 9). When the raters are only nominally, rather than strictly, parallel, however, the generic reliability statistic is more appropriate. In the present case, strict parallelism implies that in the statistical population being rated (of which the sets of 100 samples are, themselves, samples), each rater would give the same average estimate of the number of factors; whereas nominal parallelism implies that in the population being tested each rater may give a different average estimate. Both types of parallelism also imply that the within-rater variances and inter-rater covariances are equal. As in most inter-rater reliability situations which appear in the psychological literature, the raters in the present study could be considered only nominally parallel.

The generic reliabilities for the three cases of Cattell's scree test and twelve cases of Linn's mean square ratio test are presented in Table 3. In all cases the generic reliabilities are extremely low.

TABLE 3

Generic Reliabilities of the Scree Test and Mean Square Ratio Test

INDEX	SAMPLE SIZE 30	60	120
LINN: Principal Components	.087	.113	.167
SMC	.064	.115	.308
Image	.070	.395	.222
Harris	.139	.111	.069
CATTELL	.071	.054	.054

Since significance tests of generic reliability are not readily available, the significance test of specific reliability (Gulliksen, 1950, Ch.14) was made. The significance test of generic reliability would be more stringent than that of specific reliability. Thus, a finding of significant specific unreliability guarantees significant generic unreliability, but a finding of significant specific reliability does not guarantee significant generic reliability. In all 15 cases of these subjective methods of selecting a number of factors, the specific reliability hypothesis was rejected at the .05 level. Thus, on the basis of their unreliability, the results of the scree test and mean square ratio test were not included in the investigation of sampling distributions.

The Sampling Distributions

Frequency Distributions of the Raw Data

Table 4 (a-i) gives the frequency distributions of the number of factors indicated by each case of the indices at the three sample-size levels. It should be noted at this point that modifications were made to two of the indices for purposes of analysis. The analytic procedures used in the study would not admit missing data.

Thus, Horn's test, which relies on crosses in the plots of eigenvalue-pairs to give results, was modified to indicate 15 factors when this cross did not occur. As is indicated in parts (b) and (c) of Table 4, this happened in several samples of the smc and image analyses. Although some of the frequency distributions took on a peculiar shape due to this modification (particularly in the smc distributions), and although, in the later analyses, the deviation from the population of the number of factors (or the bias) of the sampling distributions was increased (as is illustrated in Figure 1), it was thought that this bias might be an appropriate quantification of the tendency of the test to fail to indicate any number of factors at all.

As is illustrated in Figure 2, the tendency to fail increased as sample size increased. For the smc extraction, this tendency also increased as the number of random eigenvalues used in the comparison with the observed eigenvalues increased, but for the image extraction, this tendency was reversed.

The other index modified for the same reason was Crawford's index. Theoretically, this index may fail if simple structure is not evidenced in the factor matrix. In such cases, the Crawford index may not reach a definite minimum but show several local minima. As is shown in Figure 3,

TABLE 4(a)

Frequency Distributions of the Number of Factors Indicated by Horn's Test on Principal Components

Number of Factors	SINGLE RANDOM EIGENVALUES Sample Size 30	60	120	AVERAGE RANDOM EIGENVALUES Sample Size 30	60	120
1	0	0	0	0	0	0
2	3	0	0	1	0	0
3	25	2	0	24	2	0
4	63	73	56	69	76	60
5	8	25	42	6	22	40
6	1	0	2	0	0	0
7	0	0	0	0	0	0
8	0	0	0	0	0	0
9	0	0	0	0	0	0
10	0	0	0	0	0	0
11	0	0	0	0	0	0
12	0	0	0	0	0	0
13	0	0	0	0	0	0
14	0	0	0	0	0	0
15	0	0	0	0	0	0
Total	100	100	100	100	100	100

TABLE 4(b)

Frequency Distributions of the Number of Factors Indicated by Horn's Test on SMC Factors

Number of Factors	SINGLE RANDOM EIGENVALUES Sample Size 30	60	120	AVERAGE RANDOM EIGENVALUES Sample Size 30	60	120
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	4	0	0	3	0	0
4	36	1	0	32	0	0
5	27	8	0	31	3	0
6	19	24	2	24	24	1
7	10	12	5	8	14	1
8	1	5	3	0	6	0
9	1	1	2	0	0	0
10	0	0	0	0	0	0
11	0	0	0	0	0	0
12	0	0	0	0	0	0
13	0	0	0	0	0	0
14	0	0	0	0	0	0
15	2	49	88	2	53	98
Total	100	100	100	100	100	100

TABLE 4(c)

Frequency Distributions of the Number of Factors Indicated by Horn's Test on Image Factors

Number of Factors	SINGLE RANDOM EIGENVALUES Sample Size 30	60	120	AVERAGE RANDOM EIGENVALUES Sample Size 30	60	120
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	2	0	0	0	0	0
4	18	0	0	21	0	0
5	34	2	0	32	0	0
6	27	1	0	33	0	0
7	8	8	0	8	6	0
8	4	19	0	5	15	0
9	5	17	1	1	14	0
10	1	11	3	0	20	0
11	0	9	5	0	12	1
12	1	10	7	0	10	3
13	0	2	11	0	7	15
14	0	7	28	0	9	45
15	0	14	45	0	7	36
Total	100	100	100	100	100	100

TABLE 4(d)

Frequency Distributions of the Number of Factors Indicated by Horn's Test on Harris Factors

Number of Factors	SINGLE RANDOM EIGENVALUES Sample Size 30	60	120	AVERAGE RANDOM EIGENVALUES Sample Size 30	60	120
1	0	0	0	0	0	0
2	0	0	0	0	0	0
3	0	0	0	0	0	0
4	0	0	0	0	0	0
5	0	0	0	0	0	0
6	0	0	0	2	0	0
7	1	0	0	1	0	0
8	6	1	0	7	0	0
9	12	18	37	18	19	36
10	26	53	54	25	53	56
11	33	26	9	22	23	8
12	14	1	0	16	4	0
13	7	1	0	6	1	0
14	1	0	0	3	0	0
15	0	0	0	0	0	0
Total	100	100	100	100	100	100

TABLE 4(e)

Frequency Distributions of the Number of Factors Indicated by Crawford's Test on the Quartimax Rotation

Number of Factors	SAMPLE SIZE 30	60	120
2	0	0	0
3	3	0	0
4	21	22	17
5	24	43	70
6	29	26	9
7	14	8	4
8	9	1	0
Total	100	100	100

TABLE 4(f)

Frequency Distributions of the Number of Factors Indicated by Crawford's Test on the Varimax Rotation

Number of Factors	SAMPLE SIZE 30	60	120
2	0	0	0
3	3	0	0
4	21	21	14
5	23	43	72
6	28	26	10
7	15	9	3
8	10	1	1
Total	100	100	100

TABLE 4(g)

Frequency Distributions of the Number of Factors Indicated by Crawford's Test on the Equamax Rotation

Number of Factors	SAMPLE SIZE 30	60	120
2	0	0	0
3	3	0	0
4	25	23	18
5	28	43	72
6	26	26	9
7	17	8	1
8	1	0	0
Total	100	100	100

TABLE 4(h)

Frequency Distributions of the Number of Factors Indicated by Guttman's Weaker Lower Bound

Number of Factors	SAMPLE SIZE 30	60	120
1	0	0	0
2	0	0	0
3	0	0	0
4	23	9	14
5	69	75	77
6	8	16	9
7	0	0	0
8	0	0	0
9	0	0	0
10	0	0	0
11	0	0	0
12	0	0	0
13	0	0	0
14	0	0	0
15	0	0	0
Total	100	100	100

TABLE 4(i)

Frequency Distributions of the Number of Factors Indicated by Guttman's Stronger Lower Bound

Number of Factors	SAMPLE SIZE 30	60	120
1	0	0	0
2	0	0	0
3	0	0	0
4	0	0	0
5	0	0	0
6	0	0	0
7	0	0	0
8	0	0	0
9	14	35	66
10	62	62	34
11	20	3	0
12	4	0	0
13	0	0	0
14	0	0	0
15	0	0	0
Total	100	100	100

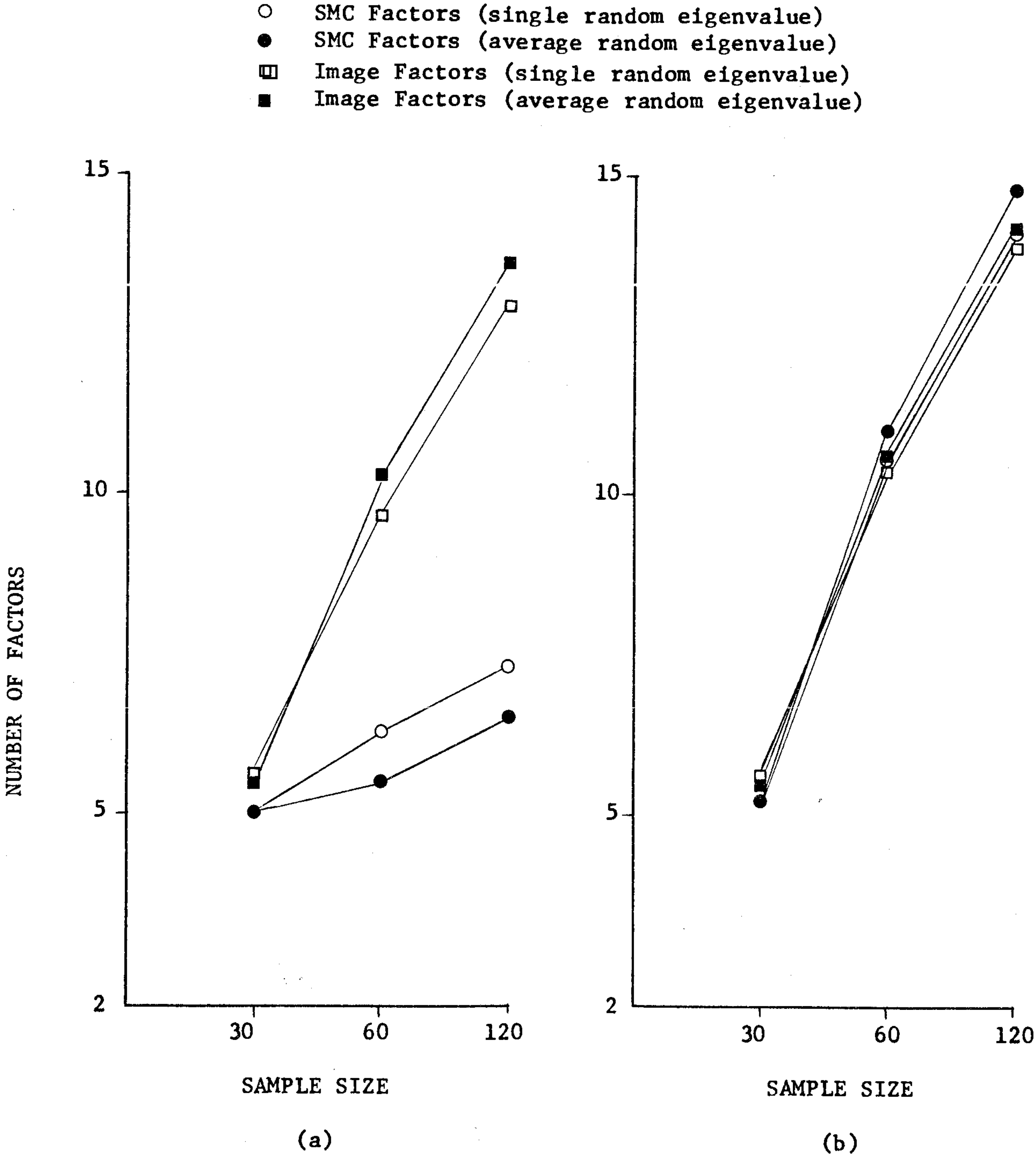

Fig. 1. Mean number of factors indicated by Horn's test following two methods of extraction and two number-of-random-eigenvalues levels (a) before and (b) after modification.

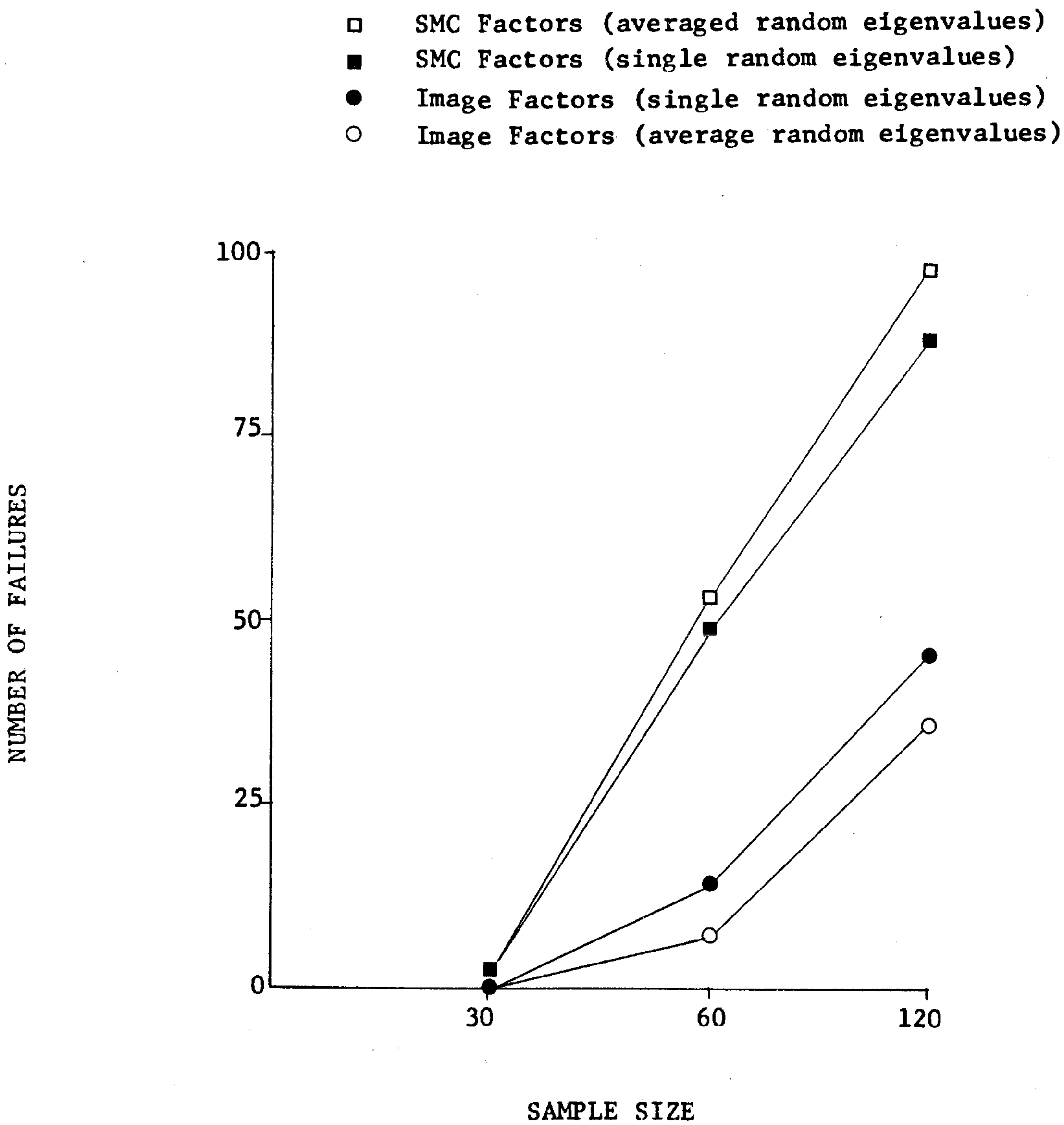

Fig. 2. Number of failures of Horn's test before modification.

the tendency to fail decreased as sample size increased. The test was modified such that the number of factors considered to be indicated was the number of factors which generated the smallest of the local minima when they occurred. A comparison of the results of the original and modified procedures, as illustrated in Figure 4, indicated that although there was not a large effect on the frequency distributions, if any existed the modified version would show more bias in the later analyses.

Analysis of Variance Designs

To assess the significance of the effects of (a) sample size, (b) type of index, and (c) the several cases within Horn's and Crawford's indices, two sets of analyses of variance were carried out on the data. One set of analyses was made on the raw data results of the number-of-factors indices. This set of analyses allowed for examination of the direction of the bias of the indices, i.e., for the examination of whether they tended to indicate fewer or more factors than the population value. The second set of analyses involved a transformation of the raw data to absolute deviations from the population value of the number of factors in order to analyse differences in the indices' magnitude of bias.

Each set of analyses was composed of four separate analyses, the last of which depended on the results of the first three. The designs were as follows.

Horn analysis. This was a three-way analysis with two within- and one between-samples variables. The samples were nested within three sample-size levels and crossed with the four types of extraction (or types of communality estimate), and with the two levels of number of random eigenvalues used in the average-comparison eigenvalues.

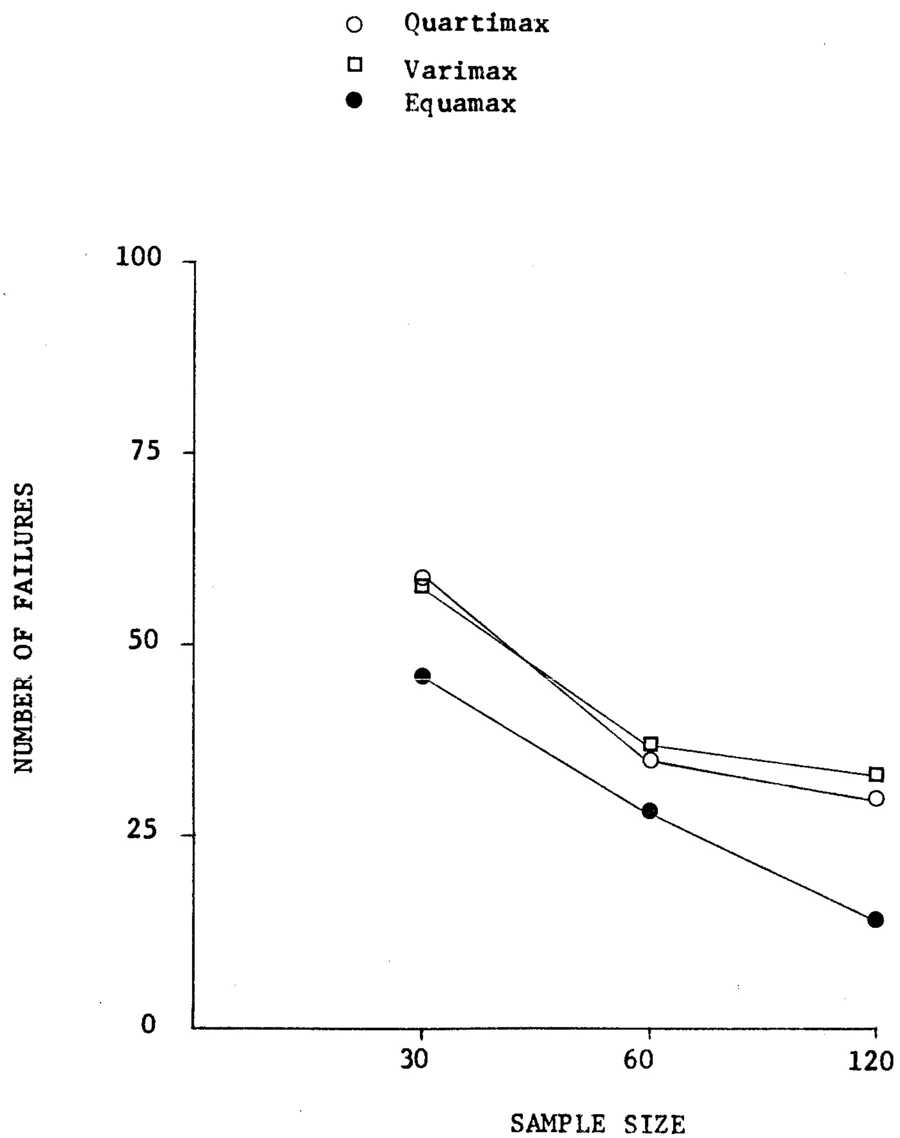

Fig. 3. Number of failures of Crawford's index before modification.

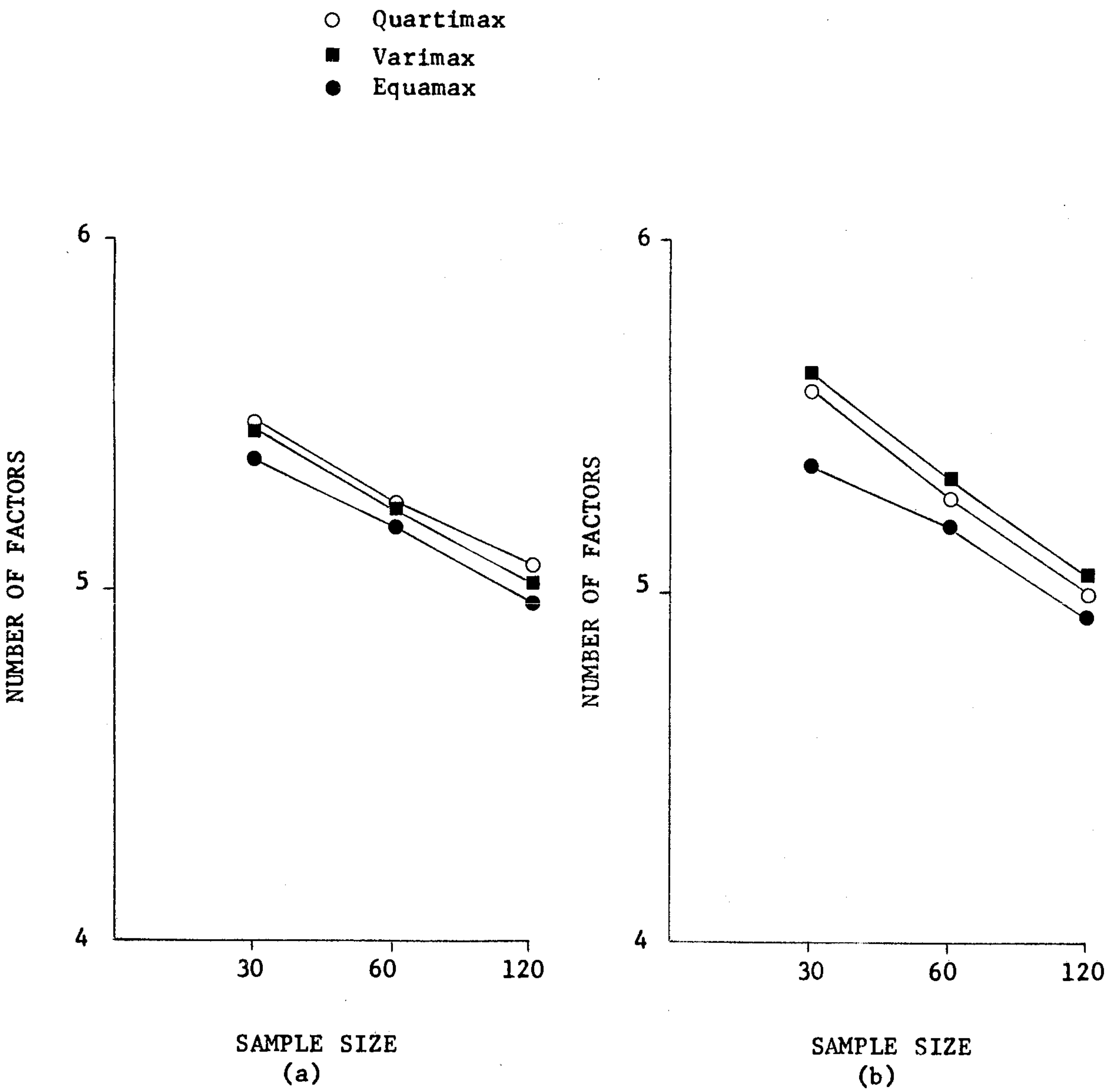

Fig. 4. Mean number of factors indicated by Crawford's test following three rotation methods (a) before and (b) after modification.

Crawford analysis. This was a two-way analysis with samples nested within the three sample-size levels and crossed with the three types of rotation.

Guttman analysis. This also was a two-way analysis. Samples were nested within sample-size levels and crossed with the two types of index (viz, Guttman's Weaker Lower Bound and Guttman's Stronger Lower Bound).

Overall analysis. This was a two-way analysis in which the indices which showed the smallest magnitude of error in the above three analyses were selected for comparison with each other. Samples were nested within the three sample-size levels and crossed with the three types of number of factors index.

Information was gathered regarding the relative magnitudes of the different types of effects in the eight analyses of variance. For any significant effect, E,

$$\hat{\theta}^2_E = \frac{MS_E - MS_e}{N} \tag{17}$$

where MS_E represents the mean square for the E-effect, MS_e is the mean square for the error term, and N is the number of observations at each level of E. $\hat{\theta}^2_E$ represents an estimate of the magnitude of the effect of E on the data and is comparable to the other $\hat{\theta}^2$'s (of significant effects) in the analyses.

Following the analyses of variance, several a posteriori comparisons of means were made using the Scheffé procedure presented by Ferguson (1959, p. 296). These comparisons were made to determine whether (in the magnitude-of-bias analyses) there was a significant difference between the most accurate and the second most accurate indices when averaged over sample-size levels.

Results of the Analyses of Variance

Horn analyses. As is indicated in Table 5, the type of extraction for the Horn test had a significant and large effect. Table 6 indicates that the Horn test most accurately approximates the population number of factors after a principal components extraction and most inaccurately approximates the population number of factors after extraction with smc's. Table 5 also shows a significant and large sample-size effect, the direction of which is indicated in Table 7. Increasing sample size tends to increase the bias of the Horn indices. However, there is a very large interaction effect between sample size and type of extraction indicated in Table 5 which is illustrated in Figure 5. Whereas increasing the sample size tends to decrease the raw-score and absolute deviations of the Horn test following Harris and principal components analyses, it tends to increase both types of bias following smc and image analysis. Table 6 and Figure 5 also indicate that this index tends to underestimate the population number of factors following principal components extraction and tends to overestimate the population number of factors following the other types of extraction.

The effect of using averaged eigenvalues rather than single-sample eigenvalues from the random data matrices in the comparison with the eigenvalues from the observed sample matrices was not significant.

To select the best case of the Horn test for the overall anlysis, the means of the principal components and Harris extractions (in the absolute-deviation analysis) were chosen as showing the least magnitude of error and were compared with each other using the Scheffé method for a posteriori comparisons. It was found that the principal components results were significantly better than those following the Harris extraction at $p \leq .01$.

TABLE 5

Relative Magnitudes of Significant Effects ($p \leq .05$) in the Analyses of Variance as Represented by $\hat{\theta}^2$

HORN ANALYSES	SAMPLE SIZE (SS)	TYPE OF EXTRACTION (E)	SINGLE VS AVERAGE EIGENVALUES (SR)	INTERACTIONS SSXE	EXSR	SSXSR
Raw Scores	4.863	8.850	ns	6.780	.014	ns
Absolute Deviation	3.539	4.794	ns	6.654	ns	.016

CRAWFORD ANALYSES	SAMPLE SIZE (SS)	TYPE OF ROTATION (R)	INTERACTION SSXR
Raw Scores	.060	.005	.002
Absolute Deviation	.139	.002	.002

GUTTMAN ANALYSES	SAMPLE SIZE (SS)	TYPE OF TEST (T)	INTERACTION SSXT
Raw Scores	.030	11.344	.107
Absolute Deviation	.048	9.930	.062

OVERALL ANALYSES	SAMPLE SIZE (SS)	TYPE OF TEST (T)	INTERACTION SSXT
Raw Scores	.004	.294	.064
Absolute Deviation	.050	.092	.025

TABLE 6(a)

Means of Extraction Groups for the Horn Analyses

	TYPES OF EXTRACTION			
ANALYSES	PRINCIPAL COMPONENT	SMC	IMAGE	HARRIS
Raw Scores	4.15	10.15	10.02	10.12
Absolute Deviation	.86	5.43	5.17	5.12

TABLE 6(b)

Means of Rotation Groups for the Crawford Analyses

	TYPES OF ROTATION		
ANALYSES	QUARTIMAX	VARIMAX	EQUAMAX
Raw Score	5.27	5.31	5.16
Absolute Deviation	.71	.72	.63

TABLE 6(c)

Means of the Test Type Groups for the Guttman Analyses

	TYPES OF TEST	
ANALYSES	GUTTMAN'S WEAKER LOWER BOUND	GUTTMAN'S STRONGER LOWER BOUND
Raw Scores	4.96	9.72
Absolute Deviation	.26	4.72

TABLE 6(d)

Means of the Test Type Groups for the Overall Analyses

	TYPES OF TEST		
ANALYSES	HORN-PC-100	CRAWFORD-E	G.W.L.B.
Raw Scores	4.13	5.16	4.96
Absolute Deviation	.87	.63	.26

TABLE 7

Means of Sample Size Groups in the Analyses of Variance

ANALYSES	SAMPLE SIZES 30	60	120
HORN:			
Raw Scores	6.27	8.91	10.65
Absolute Deviation	2.19	4.30	5.94
CRAWFORD:			
Raw Scores	5.52	5.23	4.99
Absolute Deviation	1.07	.67	.32
GUTTMAN:			
Raw Scores	7.50	7.38	7.15
Absolute Deviation	2.73	2.47	2.19
OVERALL:			
Raw Scores	4.67	4.82	4.76
Absolute Deviation	.82	.57	.38

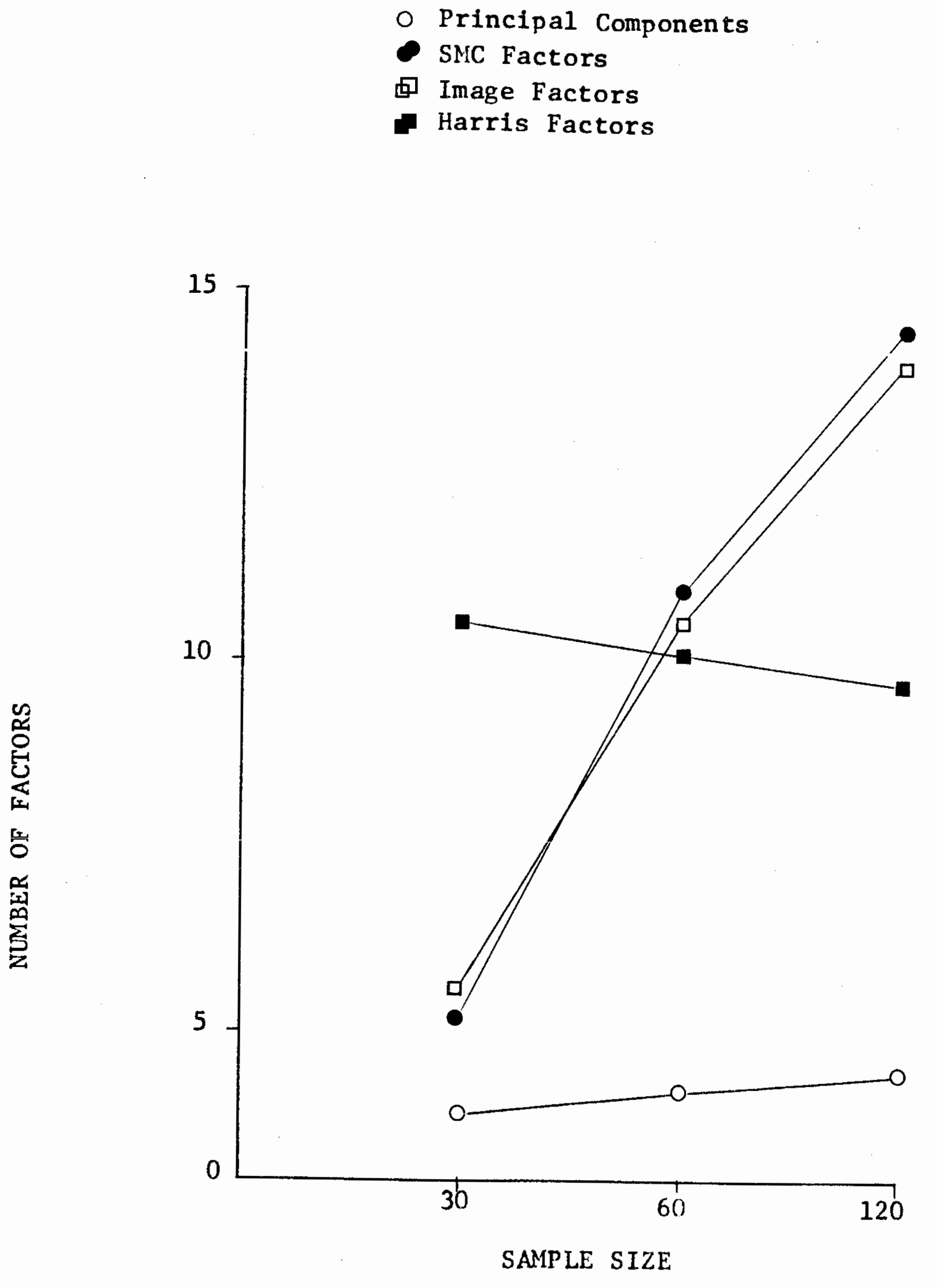

Fig. 5. Mean number of factors indicated by Horn's test following four methods of extraction.

Since there was no significant difference between the single- and averaged-eigenvalue cases, and since the averaged-eigenvalue case is expected to be more accurate on theoretical grounds, the Horn test following principal components analysis with averaged-random eigenvalues was selected for the overall analysis.

Crawford analyses. Table 5 indicates that the sample-size effect, the type-of-rotation effect, and the interaction between these effects were significant in both of the Crawford analyses. The magnitude of the sample-size effect is greater than the other two effects. As is indicated in Table 7, increasing sample size decreases the bias in the Crawford index in both analyses. The type-of-rotation effect, as is shown in Table 6, indicates that the Crawford index best approximated the population number of factors following the equamax rotation and was most biased following the varimax rotation. The significant but small interaction between sample size and type of rotation in the raw-score analysis is illustrated in Figure 4b. This illustration indicates a tendency for Crawford's index to underestimate the population number of factors following the equamax rotation in the large sample-size group.

For the purpose of selecting the rotation which gave the most accurate results, the means of the quartimax and equamax rotations from the absolute-deviation analysis were compared. It was found that the results following the equamax rotation were significantly more accurate than the quartimax results at $p \leq .01$.

Guttman analyses. In both analyses of the Guttman tests, as is shown in Table 5, the sample-size, type-of-index, and interaction effects were all significant. The largest effect was due to the type of test. The sample-size effect was smaller than the interaction effect. As is illustrated in Figure 6, although increasing sample size tends to produce less bias in Guttman's Stronger Lower Bound, it has no consistent effect on Guttman's Weaker Lower Bound. Table 7 and Figure 6 both indicate that Guttman's Stronger Lower Bound tends to overestimate the population number of factors and Guttman's Weaker Lower Bound tends to underestimate the number of factors.

The means of the absolute deviations for the two tests were compared for selection to the final analyses, and it was found that Guttman's Weaker Lower Bound showed significantly less bias than Guttman's Stronger Lower Bound at $p \leq .01$, and thus, the former was included in the overall analyses.

The overall analyses. The indices included in the final analyses - (a) Horn's test with averaged eigenvalues following principal components analysis (Horn - PC - 100), (b) Crawford's index following the equamax rotation (Crawford - E), and (c) Guttman's Weaker Lower Bound (G.W.L.B.) - had a significant effect in both of the overall analyses as is indicated in Table 5. From the same table it can be seen that there were also small but significant sample-size- and interaction effects. As is indicated in Table 7, in the absolute-deviation analysis, increasing sample size produces decreasing magnitude of bias; but in the raw-score analysis, the interaction effect, illustrated in Figure 7, indicates that increasing sample size improves the results of the Crawford and Horn indices (and has no apparent effect on Guttman's Weaker Lower Bound), but that at small sample sizes, whereas the Crawford test tends to overestimate the number of factors, the Horn index tends to underestimate the number of factors.

From Table 6 it can be seen that in both analyses, Guttman's Weaker

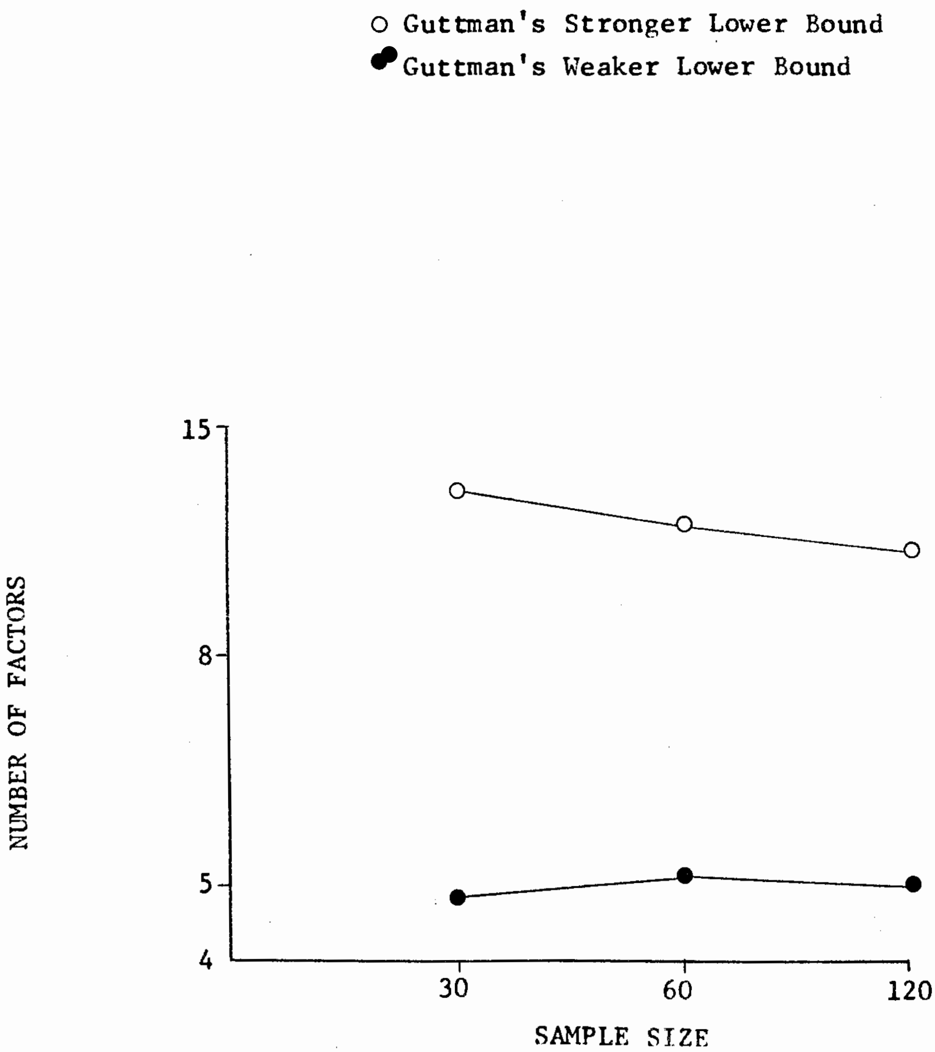

Fig. 6. Mean number of factors indicated by the Guttman tests.

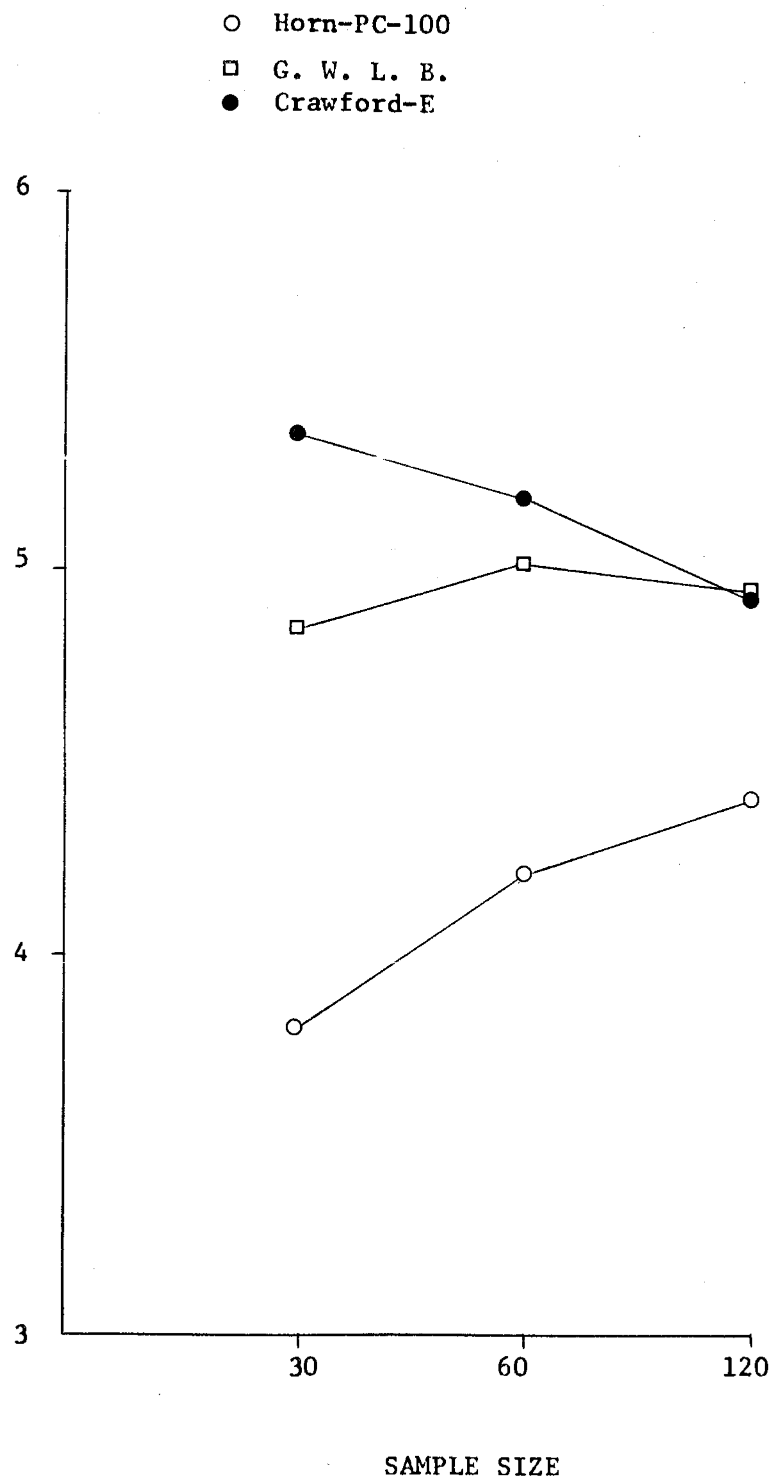

Fig. 7. Mean number of factors indicated by the tests included in the overall analysis.

Lower Bound is the least biased and the Horn index, the most biased. A comparison was made on the means of the Crawford index and Guttman's Weaker Lower Bound from the absolute-deviation analysis, and it was found that Guttman's Weaker Lower Bound is significantly better at approximating the population number of factors than the Crawford index (following the equamax rotation) at the $p \leq .01$ level.

Discussion

Reliability

The scree test was devised as an aid in determining the number of factors to consider for people who have had a relatively large amount of practical experience working with factor analytic techniques (Cattell, 1966a). Apparently, the mean square ratio test was developed for a similar population of factor analysts. Thus, it might be argued that the relatively naive raters in the present study should show less reliability than those for whom the indices were designed. However, both tests are highly susceptible to being biased toward any hypothesis the user has regarding the number of major common factors in a set of data. A user with a hypothesis regarding the number of factors would probably interpret the scree or root break to be closer to the hypothesized number of factors whenever subjective estimation of where the scree begins or root break occurs enters into the implementation of these indices. Thus, these indices should be used with extreme caution (a) by naive factor analysts, and (b) when the analyst has some implicit or explicit hypothesis of the number of factors to consider.

Modifications

Horn's Index

The modifications to the Horn test resulted from failures of the test to indicate the number of factors using Humphreys' (Humphreys and Ilgen, 1969) expansion of Horn's (1965) original procedure. The surprising failures of the eigenvalues to cross in the smc and image analyses indicates that although the eigenvalues from the factoring of two different correlation matrices of the same order and sample size may be comparable following principal components analysis, they are not necessarily comparable following smc or image extractions. In the principal components case, the sums of the eigenvalues of the comparable correlation matrices described above are equal to n, the number of variables. Thus, the eigenvalues must either be identical or show a cross in their plot. Similarly, in the Harris case, the sums of the reciprocals of the two sets of eigenvalues are equal to n and the plots of both the reciprocals and the eigenvalues show crosses. In other extractions, however, no similar restriction is placed on the sum of the eigenvalues, and their plots need not necessarily cross. Thus, the Horn test need not necessarily give an estimate of the number of factors following factor extractions which do not place appropriate restrictions on the sum of the eigenvalues.

Crawford's Index

The findings (a) that the modifications to the Crawford index had little effect on the sampling distributions, (b) that the failure rate decreased as sample size increased in the unmodified version, and (c) that there was little effect of the type of rotation on the sampling distributions, suggest that this index probably estimates the number of factors well even when the correlation matrix does not contain, and the factor matrix does not exhibit, the conditions of simple structure.

The Sampling Distributions

Increasing sample size (a) had no effect on Guttman's Weaker Lower Bound; (b) improved the estimates of the number of factors in all cases of Crawford's index, Horn's index following principal components analysis and the Harris extraction, and Guttman's Stronger Lower Bound; and (c) impaired the estimates of Horn's index following smc and image analyses. This last tendency indicates that the theoretical basis for the Horn test is weak, and that outside the principal components model, there may be a tendency for the test to fail entirely. If one considers the factor analytic model to include three sources of error variance - (a) psychometric, (b) statistical, and (c) an interaction between the psychometric and statistical sources - the improvement of the non-purely statistical test with increasing sample size (or decreasing amounts of statistical error) is to be expected. It is interesting, however, that Guttman's Weaker Lower Bound appears to be robust with respect to variations in statistical sampling error. This apparent robustness, however, may be peculiar to the population matrix used in the study since for this matrix, Guttman's Weaker Lower Bound gave good results even at the smallest sample size.

The tendency of Guttman's Weaker Lower Bound, Horn's index (following principal component, smc, and image extractions), and Crawford's index to specify too few factors in some samples (as is shown in Table 4) must be noted. In general, the underestimation of the number of factors is considered worse than its overestimation. Usually the investigator would rather risk overinterpretation of his data than not identify a possibly interpretable factor which was rejected by one of the number-of-factor indices. However, the indices which consistently overestimated the number of factors in the present study overestimated it to such an extent that any rotation of the number of factors indicated by these tests would have been badly distorted. The rotation would have taken advantage of too much information specific to the particular sample for the analysis to be useful.

Concluding Comments

It must be remembered that only one population matrix was included in the study. This implies that only one level of psychometric error was present in all of the samples. Such a restriction precludes the examination of the effects on the number-of-factors indices (a) of interactions between different levels of statistical and psychometric error; (b) of the number of variables (e.g., Cattell (1966b) notes that Guttman's Weaker Lower Bound tends to underestimate the number of factors when the number of variables, n, is less than 20 and to overestimate the number of factors when n is greater than 50); and (c) of the ratio of the number of major common factors to the number of variables (e.g., Kaiser (1960) discusses the tendency of Guttman's Stronger Lower Bound to indicate more than n/2 factors. If the true number of major common factors is close to n/2, Guttman's Stronger Lower Bound would probably give better estimates of the number of factors than many of the other indices.) The effects of all these parameters on the sampling distributions of the number-of-factors indices are probably of even greater interest to the psychologist using factor analytic techniques than is knowledge of their statistical sampling distributions alone. Given this warning, however, in the present study, Guttman's Weaker Lower Bound and Crawford's index (on equamax factors) appeared to give better estimates of the population number of factors than did the other indices considered.

Given that, in practice, factor analytic procedures include both psychometric and statistical error, it seems that the traditional indices of the number of factors, which assume the observed error to be totally psychometric or totally statistical, are not really appropriate. An objective test which operates on some criterion of the identifiability of the factors, but which operates independently of the statistical or psychometric models of factor analysis, such as the Horn index (after principal components and Harris analyses) and the Crawford index, may be more appropriate in most practical situations.

I feel I must conclude this discussion with a comment which has appeared in every study of this type: "None of the indices considered in the study appears to be a completely adequate estimator of the number of factors to consider in exploratory factor analysis".

Summary

For the purpose of examining the behaviour of non-statistical indices of the number of factors under conditions in which various degrees of statistical sampling error were present, empirical statistical sampling distributions of the estimates of six number-of-factors indices were generated by Monte Carlo procedures and inspected for deviation from a known population number of factors. The distributions were based on 100 samples (at each of 3 sample sizes) from Cattell and Sullivan's Cups of Coffee Problem - Sample A (which has a fairly well defined number of factors but includes psychometric error comparable to that found in typical applications of factor analytic techniques to psychological data).

Cattell's scree test and Linn's mean square ratio test, which require subjective operations, were implemented too unreliably by five raters to be included in the analysis of sampling distributions. Horn's index (following smc and image analyses) and Crawford's index failed to indicate any number of factors for some samples. In order to include them in the analysis of sampling distributions, modifications were made which (a) had little effect on the mean estimate of the number of factors of the Crawford test, and (b) biased the estimates of the Horn test to an extent indicating the number of failures. For the particular population matrix used in the study, sequential analyses of variance indicated that Guttman's Stronger Lower Bound consistently overestimated the population number of factors whereas Guttman's Weaker Lower Bound and the modified version of Crawford's index gave the most accurate estimates of the population number of factors.

Bibliography

Browne, M. W. A comparison of factor analytic techniques. Psychometrika, 1968, 33, 267-334.

Browne, M. W. Fitting the factor analysis model. Psychometrika, 1969, 34, 375-394.

Carroll, J. B. An analytic solution for approximating simple structure in factor analysis. Psychometrika, 1953, 18, 23-38.

Cattell, R. B. The scree test for the number of factors. Multivariate Behavioral Research, 1966, 1, 245-276. (a)

Cattell, R. B. The strategic use and meaning of factor analysis. In R. B. Cattell, (Ed.), Handbook of multivariate experimental psychology. Chicago: Rand McNally, 1966. Pp. 174-243. (b)

Cattell, R. B. & Sullivan, W. The scientific nature of factors: a demonstration by cups of coffee. Behavioral Science, 1962, 7, 184-193.

Cliff, N. & Hamburger, C. D. The study of sampling errors in factor analysis by means of artificial experiments. Psychological Bulletin, 1967, 68, 430-445.

Cliff, N. & Pennell, R. The influence of communality, factor strength, and loading size on the sampling characteristics of factor loadings. Psychometrika, 1967, 32, 309-326.

Crawford, C. B. Analytic methods of rotation in the determination of the number of factors. Unpublished doctoral dissertation, McGill University, 1966.

Ferguson, G. A. Statistical analysis in psychology and education. New York: McGraw-Hill, 1959.

Gulliksen, H. Theory of mental tests. New York: John Wiley & Sons, 1950

Guttman, L. Some necessary conditions for common factor analysis. Psychometrika, 1954, 19, 149-161.

Hamburger, C. D. Factorial stability as a function of analytic rotation method, type of factor pattern, and size of sample. (Doctoral dissertation, University of Southern California) Ann Arbor, Mich.: University Microfilms, 1965. No. 65-3106.

Harris, C. W. Some Rao-Guttman relationships. Psychometrika, 1962, 27, 247-263.

Horn, J. L. A rationale and test for the number of factors in factor analysis. Psychometrika, 1965, 30, 179-185.

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

Humphreys, L. G. & Ilgen, D. R. Note on a criterion for the number of common factors. Educational and Psychological Measurement, 1969, 29, 571-578.

Jöreskog, K. G. Statistical estimation in factor analysis. Uppsala: Almqvist & Wiksell, 1963.

Jöreskog, K. G. A general approach to confirmatory maximum likelihood factor analysis. Psychometrika, 1969, 34, 183-202.

Kaiser, H. F. The varimax criterion for analytic rotation in factor analysis. Psychometrika, 1958, 23, 187-200.

Kaiser, H. F. The application of electronic computers to factor analysis. Educational and Psychological Measurement, 1960, 20, 141-151.

Kaiser, H. F. Image analysis. In C. W. Harris (Ed.) Problems in measuring change. University of Wisconsin Press, 1963. Pp. 156-166.

Kaiser, H. F. & Caffrey, J. Alpha factor analysis. Psychometrika, 1965, 30, 1-14.

Lawley, D. N. The estimation of factor loadings by the method of maximum likelihood. Proceedings of the Royal Society of Edinburgh, 1940, 60, 64-82.

Linn, R. L. Use of random normal deviates to determine the number of factors in factor analysis. Unpublished Master's thesis, University of Illinois, Urbana, Ill., 1964.

Linn, R. L. A Monte Carlo approach to the number of factors problem. Psychometrika, 1968, 33, 37-71.

Lord, F. M. & Novick, M. R. Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968.

Rao, C. R. Estimation and tests of significance in factor analysis. Psychometrika, 1955, 20, 93-111.

Saunders, D. R. Transvarimax: some properties of the ratiomax and equamax criteria for blind orthogonal rotation. American Psychologist, 1962, 17, 395. (Abstract)

Thurstone, L. L. Multiple factor analysis. Chicago: University of Chicago Press, 1947.

Whittlesey, J. R. B. A comparison of the correlational behavior of random number generators for the IBM 360. Communications of the ACM, 1968, 11, 641-644.