71-12,610

POLLARD, Jr., Nathanael, 1939-
AN EMPIRICAL INVESTIGATION OF SOME EFFECTS OF NON-NORMALITY ON BARTLETT'S TEST OF SIGNIFICANCE IN PRINCIPAL COMPONENT ANALYSIS.

The University of Oklahoma, Ph.D., 1970
Education, theory and practice

University Microfilms, A XEROX Company, Ann Arbor, Michigan

THE UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

AN EMPIRICAL INVESTIGATION OF SOME EFFECTS OF NON-NORMALITY ON BARTLETT'S TEST OF SIGNIFICANCE IN PRINCIPAL COMPONENT ANALYSIS

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

degree of

DOCTOR OF PHILOSOPHY

BY

NATHANAEL POLLARD, JR.

Norman, Oklahoma

1970

AN EMPIRICAL INVESTIGATION OF SOME EFFECTS OF NON-NORMALITY

ON BARTLETT'S TEST OF SIGNIFICANCE IN

PRINCIPAL CONPONENT ANALYSIS

APPROVED BY

____

____

____

____

____

ABSTRACT

The principal component method of factor analysis is a specific probabilistic procedure and permits the construction of tests of significance for testing whether or not factor loadings differ significantly from zero. In addition, the method permits one to test for the number of common factors for a set of random variables.

Thus, if x_i, $i = 1,2,\ldots,p$, are p random variables which can be expressed as linear combinations of q, $q < p$, common factors and p error factors, then a principal component model representing these variables is

$$\text{(I)} \quad x_i = \left(\sum_j f_{ij} u_j \right) + s_i, \ (i = 1,2,\ldots,p;\ j = 1,2,\ldots,q).$$

The symbols f_{ij} represent factor loadings, u_j represent common factor scores, and s_i represent error scores or residuals.

The problem of determining the statistical significance of the factors associated with the correlation matrix of a set of random variables was solved by Bartlett in 1950. The test suggested by Bartlett involves a mathematical consideration of the eigenvalues of the characteristic equation of the correlation matrix.

Existing literature reveals that some methods of factor analysis, for example Lawley's method of maximum likelihood and Whittle's method of least squares, work well when the common factor scores are taken from populations that are not normal. Further, the literature indicates that while the chi square test of completeness associated with Lawley's method of maximum likelihood is insensitive to the population from which the common factor scores were taken, the corresponding chi square test of Bartlett, used in principal component analysis for testing when a component analysis is complete, has not been subjected to a satisfactory test of robustness.

The specific problem investigated in this study may be stated as follows: when the method of principal component analysis is used to estimate factor loadings, what are some effects on Bartlett's test of significance when the common factor scores are taken from a population whose distribution diverges from a normal population. The non-normal populations from which common factor scores were taken for the present study were: 1) the positive half of a normal distribution with zero mean and unit variance, 2) a chi square distribution with three degrees of freedom, and 3) a t-distribution with five degrees of freedom.

Although the results of this study, based on computer generated number populations are necessarily limited, they do strongly indicate that when the method of principal analysis is used, Bartlett's test of significance is relatively insensitive to departure from normality of the distribution of the common factor scores for large numbers of observations.

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank my advisor Dr. Albert D. Smouse for his cooperation and assistance throughout my course of graduate study. Grateful appreciation is also extended to the other members of my committee Dr. Herbert R. Hengst, Dr. Thomas J. Hill, and Professor Eunice M. Lewis.

A special word of thanks is extended to Dr. Larry E. Toothaker for his invaluable criticism of the manuscript of this study. Also, thanks are extended to Dr. John C. Brixey.

Thanks are also extended to the Ford Foundation, the Southern Foundation Fellowships Fund, and Miles College for financial assistance during the period of graduate study in Oklahoma University.

In addition, the writer would like to thank Cass Petre for the typing of this dissertation. To the computer operators at Merrick Computer Center a sizeable debt is owed for running the programs that made this study possible.

Finally, I thank my wife, Sara Ann, for her patience, understanding, encouragement, and assistance. And I thank my parents Rev. and Mrs. H. H. Pollard and my grandmother, Mrs. Cora Smith for their encouragement and patience throughout my life.

DEDICATION

This study is dedicated to my wife, Sara Ann, who has been and will always be my inspiration.

TABLE OF CONTENTS

LIST OF TABLES

Table		Page

AN EMPIRICAL INVESTIGATION OF SOME EFFECTS OF NON-NORMALITY ON BARTLETT'S TEST OF SIGNIFICANCE IN PRINCIPAL COMPONENT ANALYSIS

CHAPTER I

INTRODUCTION AND PROBLEM

Factor analysis has become the generic term for a variety of procedures developed to examine whether the joint variation of p observable random variables can be described approximately in terms of the joint variation of a fewer number, say q, $q < p$, of hypothetical variables called (principal) common factors (Cooley and Lohnes, 1962, p. 151). The procedure to be used in this study is called principal component analysis. Regardless of the method of analysis of the observable random variables, one problem is encountered in the analysis. The problem is when to stop factoring. While there are tests to determine when to stop factoring, the tests usually assume that the observable random variables are measurements taken from populations whose distributions are normally distributed. Bartlett's test is used in principal component analysis to determine when to stop factoring. This study reports results of an investigation of several factor analyses of random variables whose distributions are not normal.

Background for the Study

Principal component analysis. The general factor analysis model, according to Hemmerle (1967, pp. 146-147), may be denoted by

(I) $X = u + Ff + Ss,$

where X, u, and s are column vectors with p components, f is a column vector with k, $k < p$, components, F is a $p \times k$ matrix of constants, and S is a $p \times p$ matrix of constants. The components of f are called factors and the components of F are called factor loadings. The matrix X is a p-dimensional observable random variable so standardized that each of its components has unit variance and zero mean. The restrictions imposed on the model given in (I) are:

a) the vector s is distributed independently of the vector f, and both have multivariate normal distributions;

b) $E(s) = E(f) = 0$; and

c) the individual components of the vector f and of the vector s are distributed independently of each other. That is, $E(ss^t) = I_k$ and $E(ff^t) = I_p$, where I_k and I_p are $k \times k$ and $p \times p$ identity matrices, respectively. The symbols s^t and f^t denote the transpose of s and f, respectively.

A special case of the model given in (I) was developed by Hotelling in 1933. The method associated with the model, called principal component analysis, is described by Hemmerle (1967, pp. 140-141) in the following way: given a set of random variables

x_i, $i = 1,2,\ldots,p$, with the purpose of determining a normalized linear combination, $u_1 = \sum_i f_{i1}x_i$, of these variables, where $\sum_i f_{i1}^2 = 1$ and u_1 has maximum variance; further determining a second normalized linear combination, $u_2 = \sum_i f_{i2}X_i$, of the variables, where $\sum_i f_{i2}^2 = 1$ and u_2 has maximum variance and is uncorrelated with the first, u_1, (i.e. $\mathrm{cov}(u_1,u_2) = 0$); still further determining a third normalized linear combination with maximum variance that is uncorrelated with the first two, and so on, it follows that if the f_{ij}'s can be determined, then the original set of variables can be reduced to a smaller set, u_i, $i = 1,2,\ldots,$ q, $q<p$, for further analysis. In the present study only those q normalized linear combinations with large variances were studied and related to the original data, the variables x_i, $1 = 1,2,\ldots,p$, not being studied independently. The variables u_i, $i = 1,2,\ldots,q$, are called principal components, u_1, is called the first principal component, u_2 is called the second principal component, and so on up to the qth principal component u_q.

Chakravarti, Laha, and Roy (1967, p. 435) point out that if the principal components, u_i, $i = 1,2,\ldots,q$, are interpreted as the common factors of the variables x_i, $i = 1,2,\ldots,p$, then one can write

(II) $$x_i = \sum_j f_{ij}u_j + s_i,$$

where the f_{ij}'s are factor loadings. This, however, is somewhat different from the model given in (I) in that the components of s are not in general mutually uncorrelated.

In the context of model (II), principal component analysis is strictly a mathematical operation designed to accomplish a maximization of

(III) $\quad var_j = \sum_i f_{ij}^2$, ($i = 1,2,\ldots,p$ and $j = 1,2,\ldots,q$), under the conditions

(IV) $\quad r_{jk} = \sum_i f_{ji} f_{ki}$; $\quad$ ($j,k = 1,2,\ldots,p$),

where $r_{jk} = r_{kj}$ and r_{jj} is called the communality of x_j. The conditions in (IV) imply that the correlations between the variables x_j and x_k can be approximated in the manner described, where the correlation coefficient is r_{jk}. This method assumes that the residuals, s_i, $i = 1,2,\ldots,p$, are zero.

The method of Lagrange multipliers is used to maximize the functions given in (III). Cooley and Lohnes (1962, p. 158) show that this method leads to a determination of the factor loadings in principal component analysis. The factor loadings are derived by considering the matrix equation

(V) $\quad (R - t_i I)v_i = \underline{0}$, $i = 1,2,\ldots,p$,

where I is an identity matrix, $\underline{0}$ is a zero vector, and R is the correlation matrix of the variables x_i, $i = 1,2,\ldots,p$. There are p nontrivial solutions of (V) in terms of v_i, and each v_i has components that are the factor loadings for converting the p scores of the variables x_i to one of the new uncorrelated factor scores. The symbol t_i is a constant for each value of i.

If $v_i = \underline{0}$, then v_i is a trivial solution of (V). Cooley and Lohnes (1962, p. 158) state that the nontrivial solutions are found by considering the determinantal equation

(VI) $\quad / R - tI / = 0$,

where / R - tI / denotes the determinant of the matrix R - tI and t is a constant. The determinantal equation given in (VI) is called the characteristic equation of R. For a given root of the characteristic equation of R, the corresponding vector v_i can be obtained by substituting t_i, the given root, in (V).

Cooley and Lohnes (1962, p. 159) state that when the vectors v_j, j = 1,2,...,p, are normalized $\left[\text{i.e. if } v_j^t = (v_{1j}, v_{2j}, \ldots, v_{pj}), \text{ then } \sum_i v_{ij}^2 = 1\right]$, the variance of each set of factors is t_i. The vector v_i produces the factor scores of maximum variance, the variance being the value of the largest root t_i. When principal component analysis is used, the normalized vectors v_i can serve as the factor loadings, at least when numerical values of one have been placed along the diagonal of R.

Bartlett's test of significance. One major characteristic possessed by the principal component method of analysis not present in some of the other procedures in factor analysis is that it is a specific probabilistic procedure and permits the construction of tests of significance for testing whether or not factor loadings differ significantly from zero (Solomon, 1960, p. 312). The method also permits one to test for the number of factors common to a set of random variables.

The problem of determining the statistical significance of the factors associated with the correlation matrix was solved by Bartlett (1950) by considering the significance of the roots t_i associated with the characteristic equation of the correlation matrix. According to Bartlett (1950, pp. 77-78), the test of significance can be described in the following way: let t_i, i = 1, 2,...,p, be the p roots, in descending order, of the determinantal equation given in (VI). If r is the total number of observations of the random variables given in (II), let n = r - 1, n being the total number of degrees of freedom associated with the orginial observations. Bartlett (1950, p. 78) states that the entire correlation structure can be tested for significance by calculating the quantity

$$\text{(VII)} \qquad \chi^2 = -\left[n - (1/6)(2p + 5)\right] \log_e /R/,$$

with $\frac{1}{2}p(p - 1)$ degrees of freedom. If, after the extraction of the largest roots of the characteristic equation of R corresponding to the first factors removed, it is required to test for the significance of the factors remaining, the test statistic given in (VII) takes the form

$$\text{(VIII)} \qquad \chi^2 = -\left[n - (1/6)(2p + 5) - (2/3)k\right] \log_e R_{p-k},$$

with $\frac{1}{2}p'(p' - 1)$ degrees of freedom, after k roots, $t_1, t_2, \ldots, t_k$, have been determined, where $p' = p - k$ and

$$\text{(IX)} \qquad R_{p-k} = /R/ \Big/ \left\{ t_1 t_2 \cdots t_k \left[\frac{p - t_1 - t_2 - \cdots - t_k}{p-k}\right]^{p-k} \right\}$$

The test statistics given in (VII) and (VIII) will be referred to as Bartlett's test of significance.

Bartlett (1951, p. 1) warns that after one or more significant components have been eliminated it is safer to take as the number of degrees of freedom

$\frac{1}{2}(p - k - 1)(p - k + 2)$ instead of $\frac{1}{2}(p - k)(p - k - 1)$ as given in in (VIII) above. The new value for the number of degrees of freedom would increase the estimate given in (VIII) above.

An example of a principal component analysis involving four tests and twenty observations of each test score was presented by Thomson (1951, pp. 124-126). In the final analysis, the first and the third eigenvalues of the correlation matrix were found significant while the second one was not significant. The example is mentioned here to point out that Bartlett's test of significance is valid only if the roots already removed are significant. As soon as one encounters a non-significant factor, the later factors are also non-significant. The last factor encountered in a component analysis is not dealt with. Bartlett (1950, p. 80) states,

> Merely the correlation structure of the variables is being investigated in its relation to variance. For this reason no significance can ever be attached to the last root, for it would be equivalent to asking for the correlation structure of a single variable.

Review of the Literature

Effects of non-normality. Cattell (1952, p. 80) says that the discussion of the significance of real factor variance left in a correlation matrix after extraction of so many factors may well terminate with the general question of significance for any single (rotated or unrotated) factor. However, all of the tests of significance, for this problem and the related problem of when to stop factoring, require the assumption that the scores on the variables are normally distributed, although no such assumption is made for the computation of Pearson's product moment r or for the essential processes of factor analysis itself.

Ferguson (1966) comments on the effects of non-normality on the computation of the Pearson product moment correlation coefficient in a discussion of the assumptions underlying this statistic. He says,

> In calculating the correlation coefficient it need not be assumed that the distribution of the two variables is normal. Correlations can be computed for rectangular and other types of distributions. If the two variables have different shapes, however, this circumstance will impose constraints upon the correlation coefficient.

An investigation of some effects of non-normality on a particular statistic may well commence with a question of the effects of non-normality on the component statistics involved in the computation of the statistic in question. The three basic statistics involved in principal component analysis are 1) Pearson's correlation coefficient, 2) mean, and 3) variance.

Norton (1952), in an empirical investigation of some effects of non-normality on F-distributions, found that unless the departure from normality is very extreme, the departure will probably have no appreciable effect on the validity of the F-test. Scheffe (1959, p. 337) shows that in making inferences about means, the effects of violation of the normality assumption are slight. He also shows that in making inferences about variances, the effects of violation of the normality assumption are dangerous. The findings reported by Scheffe were in agreement with those reported by Norton.

With reference to Lawley's method of maximum likelihood for estimating factor loadings, Thomson (1951, p. 127) points out that it is assumed that both test scores and the factors, of which they are linear combinations, are normally distributed throughout the population of persons to be tested. He concludes, however, that although the assumption of normality has been a subject of some criticism, in practical situations it would seem that departure from strict normality is not serious.

Tests of significance. In factor analysis, there are two related problems regarding tests of significance, both of which assume a knowledge of confidence limits. The first problem is concerned with confidence limits for different loadings in a factor matrix used to determine whether or not a loading is significantly different from zero. The second problem is when to stop factoring (Henrysson, 1960, p. 137). Several methods have been suggested as approximate solutions for each one of these problems. However, the second problem was the one considered in this study.

Cattell (1952, pp. 296-304) and Thomson (1938, pp. 120-126) discuss some of the earlier methods of determining when to stop factoring. Mosier (1939) concluded that of the five plans he investigated, the "best" was to seek an indication that the standard deviations of residuals after the last factor was extracted had become less than the standard error of the mean correlation in the original correlation matrix. Even this method was found to be unsatisfactory in some respects.

Thurston (1938) describes a criterion developed by Ledyard Tucker for deciding when to stop factoring. In this method, the sums of the absolute values of the resulting residuals, including the elements of the diagonal used just before and just after the extraction of a factor, must be less than $(p - 1)/(p + 1)$, where p is the number of tests or variables. Cattell (1952) points out that with this method it is possible to get false results, but that the empiricism of the test leads to some intuitive validity which makes it the most useful of the really quick tests.

Reyburn and Taylor (1939) introduced an alternate criterion to determine when to stop factoring. They advocated dividing the standard errors of each of the original correlation coefficients into the corresponding residual correlation coefficient. This was to be followed by plotting the distribution of these quotients and assuming that if the resulting distribution departs significantly from normality, more factors are still to be factored or extracted. A criticism of this method has been that one does not know how much departure from normality is required for significance.

Coombs (1941, pp. 267-277), assuming that the columns of a residual matrix are distributed according to the binomial distribution $(a + b)^n/2^n$, where a represents a positive entry in the residual matrix and b represents a negative entry in the residual matrix, suggests counting the number of negative signs left in the residual matrix after every attempt has been made to reduce the number of negative signs in the residual matrix. Cattell (1952) points out that this method of testing leads to the extraction of too few factors.

Holzinger and Harman (1941) mention a test due to Swineford. Swineford suggests correlating the original correlation coefficients with the corresponding residual coefficients. If no significant relationship remains, then extraction is considered to be complete. Holzinger and Harman point out that not much research has been done on the effectiveness of this method.

A major criticism of the above methods is that not enough attention is given to the number of observations used in the analysis. McNemar (1942) developed such a criterion suggesting than one extract factors until the standard error of the residual correlation coefficients is less than $1/\sqrt{N}$, where N is the sample size. In McNemar's test the standard error of the residual correlation coefficients is $\sigma_s/(1 - M_h)$, where σ_s is the standard deviation of the residuals after s factors have been extracted and M_h is the mean communality for the s factors extracted.

Saunders' (1952) criterion for deciding when to stop factoring is developed from the same logical foundations as is the criterion proposed by McNemar. Saunders' method takes into account not only the number of observations, the reliabilities of the tests, and the number of variables, but also the number of factors extracted. Saunders' method has been applied to artificial data and has been found to give reasonably good results, but ones which are not always exact. Saunders also proposes a test of significance for any rotated or unrotated factor. This test, according to Saunders (1952), takes the form of a chi square test as follows:

$$(X) \qquad \chi^2 = N(n - 1)/2n \left(\sum_{i=1}^{k} a_i^2 / u_i^2 \right)^2 ,$$

where a_i, $i = 1,2,...,k$, are factor loadings, u_i, $i = 1,2,...,k$, are measures of uniqueness, n is the number of variables, and N is the number of observations. The chi square variable above is assumed to have $(n - k + 1)$ degrees of freedom, where k is the number of factors extracted.

Lawley (1940) introduced a criterion to be used for testing the significance of common factors when the method of maximum likelihood is used to estimate factor loadings. Harman (1960) commented on the statistic used by Lawley for factor analysis and the statistic used by Bartlett for principal component analysis. He says,

> In making a comparison and distinction between component analysis and factor analysis, Bartlett (22, p. 81) notes that the total ^{2} corresponding to the significance of the unreduced correlation matrix is necessarily the same, and only because of the difference between factors extracted in the two analyses does the analysis of the total ^{2} into its respective components differ p. 382 .

Lawley's test of significance is fundamentally equivalent to the test given by Bartlett (1950), only Bartlett's test was given detailed consideration in this study.

There is no unique test of significance for deciding when to stop factoring. The procedure for determining when a factor analysis is complete depends upon the method of estimating factor loadings and the types of inferences to be drawn. Burt (1952) asserts that when principal component analysis is used to estimate the factor loadings, Bartlett's test is by far the best test of significance for deciding when a factor is significant. When the method of estimating factor loadings is that of maximum likelihood, Lawley's method of maximum likelihood is the method to be employed to test the significance of factors (Henrysson, 1960, p. 137).

Whittle (1952) developed a method of estimating factor loadings using the method of least squares, and states that his method of estimating factor loadings is applicable even when the factor analytic model does not assume residuals and when the factor scores are not normally distributed (1952, pp. 223-225). Henrysson (1960, p. 138) has noted that this method leads to principal component solutions. The only requirement needed to estimate factor loadings using Whittle's method is that the factor scores, including the unique factor scores, are uncorrelated. If the researcher is sure that the factor scores are not normally distributed, Whittle's method may be the most reliable method of estimating factor loadings.

Experimental studies in factor analysis. From 1950 until 1960 there were several studies in factor analysis that began their analyses with a known factor structure. The observation scores used in these studies were usually taken from Wold (1948). These studies can be divided into two general groups: 1) studies in factor analysis concerned with the effectiveness of methods of estimating factor loadings and 2) studies in factor analysis concerned with some effects of violation of the normality assumption. The present study is one of the latter type.

Henrysson (1950, pp. 159-165) conducted an analysis of artificially constructed samples using Lawley's method of maximum likelihood. Henrysson's object was to see if Lawley's test of significance, formulated to be used with large samples, worked with samples of 200 observations. Beginning with a known factor structure of nine variables with one common factor, Henrysson obtained results which were in good agreement with theory. From analyses of twelve samples, Henrysson concluded that all 9 x 9 covariance matrices could be expressed in terms of one significant factor only, with nine specific factors. This conclusion accorded with the predetermined conditions in the artifically constructed samples.

Wold (1953, pp. 43-64) studied the effectiveness of Whittle's method of least squares for estimating factor loadings. The experiments conducted by Wold proceeded in two steps: 1) the construction of artificial samples in accordance with the theory developed by Whittle (1952) and 2) the estimation of the factor structure on the basis of the artificial samples constructed. Wold considered two general types of problems for analysis by the theory developed by Whittle. When the different variables have residuals of equal variances, Wold suggests an analysis of the covariance matrix. He asserts that the correlation matrix should be the object for analysis if the residual variances are proportional to the variances of the variables to be factor analyzed. The results of Wold's experiments, in which he attempted to estimate normalized loadings, supported Whittle's theory. In conclusion, Wold points out that the estimation procedure in factor analysis will not give valid results unless the sample size is so large (or the variances of the residuals so small) that the estimated eigenvalues or roots of the characteristic equation of the correlation matrix, correspond to the order of the true eigenvalues. Moreover, Wold supports his

his assertion by reporting an experiment that failed to give valid results because of this circumstance.

Wold (1953), using artificial samples, found that when the common factor scores were taken from the positive half of a normal distribution and when the residuals were based on independent random samples from a normal distribution, Whittle's method of least squares worked quite well and gave acceptable figures for loadings and for individual factor values. The findings of Wold helped to confirm the conjecture that Whittle's method is more general than either the model proposed by Lawley or the principal component model proposed by Hotelling in that Whittle's theory is distribution free and does not assume the normality assumption stated earlier in this study.

Lawley and Swanson (1954, pp. 75-79) investigated the effectiveness of the method of maximum likelihood for estimating factor loadings and the associated chi square test of significance in an investigation which was similar in design to that of Henrysson. They proceeded from an artificial construction of four hundred sets of observations, divided into sub-samples of fifty observations each. There were seven variables for analysis. This study was more general than the one by Henrysson in that two common factors, instead of one, were introduced into the known factor structure. Each one of the eight subsamples, except one, gave support that the known factor structure indeed had two common factors. In the one instance where support was not given, it was noticed that there was a significantly high correlation between two of the variables for analysis in the residual matrix. On the whole, Lawley and Swanson concluded that the results obtained in their study were in reasonable agreement with the theory of estimating factor loadings by the method of maximum likelihood.

Lawley (1940) presented three methods of estimating factor loadings by the method of maximum likelihood. In a theoretical exposition by Anderson and Rubin (1956, pp. 130-145) it was shown that Lawley's method I works well even when the assumption of normality on the common factor scores is violated.

Fuller and Hemmerle (1966, pp. 225-266) investigated the robustness of the maximum likelihood estimation of the number of common factors necessary in factor analysis. In a comprehensive study based on two hundred observations, they found that the maximum likelihood method was insensitive to the common factors being non-normally distributed. Their investigation was similar to the study by Lawley and Swanson (1954) in that it was based on simulated populations. Starting with two common factors and five specific factors, Fuller and Hemmerle (1966) investigated the effectiveness of Lawley's estimate of the number of common factors by considering factor scores drawn from the following populations: Student's t-distribution, normal distribution, uniform distribution, and bimodal distribution. The method of analysis used in this study was Lawley's maximum likelihood method and the method of simulating populations was the method of Monte Carlo, wherein a game of chance technique is applied to solve certain problems. In the Monte Carlo method random sampling is applied to determine a solution to simulation problems rather than solving the problem analytically or by another method (Martin, 1967, p. 31).

Need for the Study and Statement of the Problem

Advances in computer programming have facilitated estimation of factor loadings regardless of the method used for estimation, including principal component analysis which has become one of the popular methods for estimating factor loadings when using the computer.

Still, Burt (1952, p. 109) points out that a weakness inherent in most methods of factor analysis is the absence of any agreed upon procedure for testing significance of the factors discovered. A perusal of most standard textbooks and articles on factor analysis will reveal that tests of significance are either ignored or not considered in detail.

Since a researcher is not always in a position to test whether or not the population from which factor scores are taken is normally distributed, the implication of violating a normality requirement underlying a test of individual factor significance takes on considerable importance for factor analysts.

In regards to the extent that the normality assumption on the factor scores can be violated, Henrysson (1960, pp. 133-134) says,

> ... it has not yet been determined how strictly the requirement of normality must be satisfied in order for the test of significance to function satisfactorily. It should also be pointed out that the assumption of normality is necessary only for the test of significance but not for the principal component solution itself.

Harman (1960, pp. 382-383) states that while some latitude might be allowed, a variable which is known not to be normally distributed should not be included in a factor analysis. Further, he notes that in the mathematical development leading to the large sample chi square test of significance the variables are assumed to have multivariate normal distributions. Should the chi square test of significance not be sensitive to the type of population from which the factor scores are taken, then non-normal variables can be present without invalidating the test of significance.

While research has been conducted relative to the validity of some methods of testing factor significance, (for example, Whittle's method and Lawley's method of maximum likelihood) the same type of robustness research is needed for the method of principal component analysis and its associated test of significance.

The present study is designed to test the effectiveness of the method of principal component analysis and its associated test of significance when the common factor scores are not taken from normal distributions. Findings in the present study will add support or doubt to the effectiveness of this method of analysis. If no support is given in this area, then further questions can be raised about the effectiveness of Bartlett's test of significance when the common factor scores are not normally distributed.

The specific problem investigated in this study may be stated as follows: when the method of principal component analysis is used to estimate factor loadings, what are some effects on Bartlett's test of significance when the common factor scores are taken from populations whose distributions diverge from normal populations? The null form of the hypotheses to be studied in this investigation are stated in Chapter II of this study, entitled method.

CHAPTER II

METHOD

Overview of the Methodology

In researching the question of the effects of nonnormality on Bartlett's test for factor significance, twelve experiments were conducted in which the shapes of the artificially constructed number populations were systematically varied. Following this, samples from the respective populations, having known shapes and factor structures, were subjected to the principal component method of factor analysis. The results of these analyses were obtained using Bartlett's criterion, the null form of the experimental hypotheses being that Bartlett's test would yield results which did not differ significantly from the predetermined factor structure of the number populations. A detailed statement of the procedure follows.

Description of the Experiment

Each experiment involved one hundred observations of each variable studied in the experiment. The twelve experiments were all carried out in essentially five major steps. The symbol N will be referred to in the rest of this study as the number of observations in each experiment and is equal to one hundred.

After obtaining one hundred observations of each variable to be studied in a given experiment, a correlation matrix was found for the variables. The dimensions of the correlation matrix was 5 x 5, 7 x 7, or 9 x 9 in case the number of variables were five, seven, or nine respectively. In addition to finding the correlation matrix, the means and standard deviations of the variables were also found. The correlation coefficients, means, and standard deviations were found by using subroutine Corre (System/360 Scientific Subroutine Package, 1968, pp. 32-33).

The second step in each experiment was to find the eigenvalues or latent roots and the corresponding eigen vectors of the correlation matrix. The method used to find the eigenvalues was the diagonalization procedure originated by Jacobi and adapted by Von Neumann for large computers (Ralston and Wilf, 1962, pp. 245-279; System/360 Scientific Subroutine Package, 1968, pp. 164-166). This procedure is applicable only to symmetric matrices with real components. Since Bartlett's test of significance depends primarily on the eigenvalues in order to test for the significance of a factor, only eigenvalues were needed for the essential purposes of this study. However, additional information was obtained from the structure of the samples investigated in order to allow for a closer analysis of the theoretical structure of the variables and for future research.

Kaiser (1959) suggested that the number of common factors should be equal to the number of eigenvalues greater than one. He found this number to run from a sixth to a third of the total number of variables. Kaiser stated that this was applicable only when the communalities of the variables were equal to one. Using the suggestion of Kaiser, only those eigenvalues were retained that were greater than one. These were selected using subroutine Trace (System/360 Scientific Subroutine Package, 1968, p. 55). Trace was also used to find the cumulative percentage of eigenvalues greater than one. This terminated the third part of the experimentation.

The fourth step of each experiment involved finding the factor loadings. The loadings were found by multiplying the elements of each normalized eigen vector by the square root of the corresponding eigenvalue. The resulting loadings were elements of the unrotated factor matrix. This part of the computation was achieved by using subroutine Load (System/360 Scientific Subroutine Package, 1968, p. 56).

The final major step of each experiment was to find the rotated factor matrix. The method used for rotation was the varimax method, originated by Kaiser (1959, pp. 413-420), and accomplished in this study by use of subroutine Varmx (System/360 Scientific Subroutine Package, 1968, pp. 56-57). If a_{ij} is used to denote the ith loading on the jth factor, then a_{ij} is normalized by dividing a_{ij} by the square root of the communality of the ith variable. Normalized loadings were obtained by this method. When the resulting structure consisted of p variables and q factors, an orthogonal rotation was performed on the p x q factor matrix such that the variance of the squared normalized loadings was a maximum. That is, such that

$$\text{(I)} \qquad \sum_j \left\{ p \sum_i (a_{ij}^2/h_i^2)^2 - \left[\sum_i (a_{ij}^2/h_i^2) \right]^2 \right\}$$

is a maximum. The symbol h_i in (I) above is the communality of the ith variable.

Following the above five steps the null hypotheses of each experiment were tested at the 0.01 level of significance.

Theoretical Description of the Number Populations

Suppose that z_{1i}, i = 1,2, and 3, are independent normal random variables with zero means and unit variances and that e_{1i}, i = 1, 2, ..., 9, are normal random variables with zero means and variances equal to one such that they are not mutually uncorrelated. Let the variables x_{1i}, i = 1,2, ..., 9, be defined in the following way:

$$x_{11} = 0.98489z_{11} + 0.17321e_{11}$$

$$x_{12} = 0.98995z_{12} + 0.14142e_{12}$$

$$x_{13} = 0.99499z_{13} + 0.10000e_{13}$$

$$x_{14} = 0.57446z_{11} + 0.50000z_{12} + 0.64031z_{13} + 0.10000e_{14}$$

$$x_{15} = 0.80000z_{11} + 0.43589z_{12} + 0.40000z_{13} + 0.10000e_{15}$$

$$x_{16} = 0.50000z_{11} + 0.60827z_{12} + 0.60000z_{13} + 0.14142e_{16}$$

$$x_{17} = 0.70000z_{11} + 0.60000z_{12} + 0.37417z_{13} + 0.10000e_{17}$$

$$x_{18} = 0.42426z_{11}+0.90000z_{12}+ \qquad 0.10000e_{18}$$

$$x_{19} = 0.59161z_{11}+ \qquad 0.80000z_{13}+0.10000e_{19}$$

Then the random variables x_{1i}, $i = 1,2,\ldots,9$, may be thought of as normally distributed variables depending on three common factors z_{1i}, $i = 1,2$, and 3, and upon residuals which are not mutually uncorrelated. Let the variables x_{2i}, $i = 1,2,\ldots,7$, be defined in the following way:

$$x_{21} = x_{11}$$

$$x_{22} = x_{12}$$

$$x_{23} = 0.73485z_{11}+0.67082z_{12}+0.10000e_{14}$$

$$x_{24} = 0.65574z_{11}+0.74162z_{12}+0.14142e_{16}$$

$$x_{25} = 0.59161z_{11}-0.80000z_{12}-0.10000e_{17}$$

$$x_{26} = x_{18}$$

$$x_{27} = 0.74833z_{11}-0.65574z_{12}+0.10000e_{19}$$

Then the random variables x_{2i}, $i = 1,2,\ldots,7$, may be thought of as normally distributed random variables depending on the two common factors z_{11} and z_{12} and upon residuals which are not mutually uncorrelated. Let the random variables x_{3i}, $i = 1,2,\ldots,5$, be defined as follows:

$$x_{31} = 0.99499z_{11}+0.10000e_{11}$$

$$x_{32} = 0.99499z_{11}+0.10000e_{12}$$

$$x_{33} = 0.98489z_{11}+0.17321e_{33}$$

$$x_{34} = 0.98995z_{11}+0.14142e_{34}$$

$$x_{35} =-0.98995z_{11}+0.14142e_{35}$$

Then the random variables x_{3i}, $i = 1,2,\ldots,5$, may be thought of as normally distributed random variables depending on one common factor z_{11} and upon residuals which are not mutually uncorrelated. The factor loadings for all of the above theoretical structures were chosen such that the communality of each varialbe was approximately one since this is a basic assumption for principal component analysis.

Construction of the Numerical Samples

Generation of uniform random variables on the unit interval. The populations described above were generated by the power residue method of generating uniform random numbers on the unit interval. Hamming (1962, pp. 384-388) states that this is the most commonly used method on a binary machine or computer and that the routine for generating uniform random numbers is as follows: Let y_0 and t be given initially. If the machine is a k-digit binary machine, define recursively $y_{n+1} \equiv ty_n$ (modulo the word length of the machine), where n ranges over the set of non-negative integers. The symbolic notation in number theory $x \equiv a$ (modulo m) means that the integer $x - a$ is divisible by m. From this iteration process one multiplication per number was performed and the lower order digits of the product were taken as the next number in the iteration. The consideration encountered in the construction of the samples in this phase of the study was what to choose for t and y_0. Hamming (1962, p. 385) shows that in order to take full advantage of the machine capacity and to assure a long period for the repeating sequence $\{y_n\}$, y_0 and t should be odd numbers and in addition y_0 must be relatively prime to the word length of the machine. Subroutine Randu from (System/360 Scientific Subroutine Package, 1968, p. 77) simulates uniform random numbers according to the above description. Using Subroutine Randu, the uniform random numbers were simulated for this study. It is suggested that the initial value of t should be of the form $8p\pm3$, where p is an integer. The value of t chosen for this study was 65539 (IBM Data Processing Techniques, 1959, p. 5).

The calculation of uniform random numbers is done by fixed point integer arithmetic, and division by the word length of the computer, retaining only the remainder, is implied by the (modulo the word length of the computer) reduction. In order to convert the retained remainder to a point on the unit interval an additional division by the word length of the computer is required. The modulus for the present study was chosen to be the word size of the computer for two reasons: reduction mod m, where m is the word size of the computer, involves only keeping the lower order bits and conversion to the unit interval involves merely assigning the binary point to the left of the number; therefore both divisions are avoided.

<u>Generation of other random variables</u>. A version of the central limit theorem asserts that the sum of identically distributed independent random variables X_1, X_2, ..., X_k, is approximately distributed as a normal distribution with expectation $kE(X_i)$ and variance $kV(X_i)$, where $E(X_i)$ and $V(X_i)$ denote the expectation and variance respectively of either of the random variables X_i, $i = 1,2,\ldots,k$ (Lehman and Bailey, 1968, pp. 226-229). Lehman and Bailey also point out that by the central limit theorem, if Y is a random variable with expectation m and variance s, then the random variable $Z = (Y - m)/\sqrt{s}$ is approximately normally distributed with expectation equal to zero and variance equal to one.

Populations having distributions with zero means and unit variances were constructed in the following way: since a uniformly distributed random variable on the unit interval has expectation equal to 0.5 and variance equal to 1/12, if Y is the sum of k identically distributed independent uniform random variables on the unit interval, then the expectation of Y is 0.5k and the variance of Y is k/12. Therefore, the random variable $Z = (Y - 0.5k)/\sqrt{k/12}$ is approximately normally distributed with $E(Z) = 0$ and $V(Z) = 1.0$. Elements from the population of the random variable Z were generated by first generating k uniformly distributed random variables on the unit interval as described above. According to Lehman and Bailey (1968, p. 227) k = 12 is large enough for a good approximation to the normally distributed random variable Z. However, in the investigation considered here, Subroutine Gauss was refined with k = 48 for a better approximation of elements from the population of Z. Subroutine Gauss is a computer subroutine that generates normally distributed random numbers according to the procedure described above with $k = 12$. In the case where $k = 48$, then $Z = 0.5(Y - 0.5k) = \frac{1}{2}(Y - 24)$.

In order to generate random numbers from the population N(p,0,1), a random variable distributed as the population of N(0,1) was generated first. From the population N(0,1) samples were selected according to the rule that for each element from the population of N(0,1) the absolute value of this element was taken to be a member of the population of N(p,0,1).

Wold (1953, pp. 43-44) points out that the population N(p,0,1) obtained from N(0,1) has mean equal to $\sqrt{2/\pi} = 0.798$ and standard deviation equal to $\sqrt{(\pi - 2)/\pi} = 0.603$. In order to convert the population N(p,0,1) to a population with zero mean and unit variance, each member of N(p,0,1) was multiplied by 1/0.603. This gave rise to a population with unit variance and mean equal to 1.3234. Then each member of the derived population was decreased by 1.3234. This did not change the variance but it gave rise to a population with zero mean. The final population would be considered to diverge from normality because the population is not symmetrical. Although the standardized population obtained here is no longer distributed as N(p,0,1), the same terminology will be used to designate this distribtuion.

Two other types of populations of random numbers were simulated for this study. One type of random numbers were taken from a population approximately distributed as chi square random variable with three degrees of freedom and the other population was approximately distributed as t-distribution with five degrees of freedom. These populations were generated to investigate whether Bartlett's test was sensitive to skewness and the kurtosis of a distribution that is not normal. A chi square distribution with three degrees of freedom is considerably skewed and a t-distribution with five degrees of freedom has kurtosis greater than zero. These distributions were then considered to diverge from a normal distribution since the kurtosis for a normal distribution is zero and a normal distribution is not skewed.

In general if Z_i, $i = 1,2,\ldots,k$, are k samples of random numbers with unit standard normal distributions, then $X = \sum_{i=1}^{k} Z_i^2$ is distributed as a chi square random variable with k degrees of freedom. A random variable distributed as a chi square distribution with three degrees of freedom was generated by first simulating three unit standard normal random variables by the method described for generating a population distributed normally with zero mean and unit variance. Taking the sum of the squares of these unit standard normal random variables gave a chi square distribution with three degrees of freedom.

The first three moments of a chi square distribution with k degrees of freedom are k, 2k, and 8k (Kendall and Stuart, 1958, p. 370). If the rth moment of a distribution is denoted by u_r, then the first three moments of a chi square distribution with three degrees of freedom are three, six, and twenty four, and the measure of skewness $sk = \sqrt{u_3^2 / u_2^3}$ is then approximately 1.633.

After the chi square distribution was generated, the numbers were then scaled so that the variance of the skewed population was unity and its mean was zero. The variance of a chi square distribution is not changed if each chi square value is increased by the same constant value or decreased by the same constant value. Each chi square value was multiplied by $C = 1/\sqrt{6}$, yielding a population with mean equal to $3/\sqrt{6}$ and variance equal to one. The constant $A = -3/\sqrt{6}$ was then added to each chi square value, yielding a skewed population with zero mean and unit variance. The skewness sk remains invariant under a multiplicative or additive transformation, so sk is still approximately 1.633. Although the standardized distribution obtained here is no longer distributed as a chi square distribution with three degrees of freedom, the same terminology will be used to designate this distribution.

Lehman and Bailey (1968, p. 228) state that if Z is a unit standard normal random variable and X_n is a chi square variable with n degrees of freedom, then the random variable $T = Z/\sqrt{(X_n / n)}$ is distributed approximately as a t-distribution with n degrees of freedom. Kendall and Stuart (1958, p. 375) point out that the moments u_r of a t-distribution are known only for $r < n$. In order to be able to determine the kurtosis of the distribution, a t-distribution with five degrees of freedom was considered for investigation. The distribution was generated by generating a chi square distribution with five degrees of freedom by the procedure described above and also by generating a unit standard normal distribution. Taking the quotient of the unit standard normal random variable with the square root of the chi square variable with five degrees of freedom divided by its number of degrees of freedom gave a distribution distributed approximately as a t-distribution with five degrees of freedom.

The moments u_r of a t-distribution with n degrees of freedom exist only when $r < n$, and are zero when r is odd and

(II) $$u_{2r} = n^r \frac{\Gamma(r + \frac{1}{2})\ \Gamma(\frac{1}{2}n - r)}{\Gamma(\frac{1}{2})\ \Gamma(\frac{1}{2}n)},$$ for $2r < n$ and where

r is a positive integer. The kurtosis measure $ku = (u_4/u_2^2) - 3$ is then 6.0. The distribution is then leptokurtic since $ku > 3$. Scheffe (1959, pp. 331-334) points out that the skewness and the kurtosis of a population remain invariant under a multiplicative or an additive transformation.

After the t-distribution was generated, the numbers were then scaled so that the variance of the leptokurtic population was unity and its mean was zero. Each value of the distribution was multiplied by $C = \sqrt{3/5}$, yielding a population with zero mean and unit variance. The kurtosis was unchanged and therefore remained 6.0. Although the standardized distribution obtained here is no longer distributed as a chi distribution with five degrees of freedom, the same terminology will be used to designate this distribution.

In order to generate the random variables e_{1i}, i = 1,2, ..., 9, which are not mutually uncorrelated, a method used by

Fieller, Lewis, and Pearson (1955, pp. vii - viii) was used. The method used may be described briefly in the following way: if in a finite sample of the random variable e_{1i} which is a unit standard normal random variable, one denotes the jth member of the sample by e^j_{1i}, let x_{0j} denote the jth member of a sample taken from the population of a unit standard normal random variable, and let z_{ij} denote the jth member of the ith random rearrangement of the set of x_{0j}'s, then the jth member e^j_{1i} may be defined as

(II) $$e^j_{1i} = r_i x_{0j} + \sqrt{1 - r_i^2}\; z_{ij}.$$

Fieller, Lewis, and Pearson (1955, p. xii) state that the pair of generated values given by the sample values e^j_{1t} and e^j_{1s} can be regarded as being taken from a bivarate normal distribution with correlation $r_t r_s$.

A refinement of subroutine Gauss was used to generate a sample of the random variables e_{1i}, $i = 1, 2, \ldots, 9$. This was achieved by generating a unit standard normal random variable and from this population a sample was taken representing the x_{0j} scores. A random rearrangement of this sample was obtained using random permutations taken from Moses and Oakford (1963, pp. 93-120). There was nine rearrangements of the set of x_{0j}'s generated. Using the set of x_{0j}'s and the sets of rearrangements of these values, (III) of this chapter was used to generate the random variables e_{1i}, $i = 1, 2, \ldots, 9$.

The two essential requirements for generating random numbers for the above populations were met by using the power residue method. These requirements are that first, the values of a uniformly distributed random variable distributed on the unit interval be uniformly distributed over the interval regardless of the number of values computed; and secondly, that successive values of the random variable be independent. It is pointed out in (IBM Data Processing Techniques, Random Number Generation and Testing, 1959, p. 7) that the above requirements are met by the power residue method of generating random numbers.

A factor analysis usually commences with the covariance matrix of the underlying variables. With the scaling described above all factor scores were taken from populations having zero means and unit variances. In this case the covariance matrix and the correlation matrix of the underlying variables are equal and the correlation matrix is the appropriate matrix for analysis.

Statistical Tests and Null Hypotheses

Chakravarti, Laha, and Roy (1967, p. 436) give a formulation of the hypothesis usually tested in principal component analysis by Bartlett's test of significance. The hypothesis may be stated as follows: if $t_1, t_2, \ldots, t_p$, are all of the eigenvalues of the correlation matrix of a p-variate normal population such that $t_1 \geq t_2 \geq \ldots \geq t_p$, then it is not necessary to extract more than k factors if $t_k > t_{k+1}$ but $t_{k+1} = t_{k+2} = \ldots = t_p$. Several formulations of this hypothesis were tested with the populations described earlier. One should note that equality of the latter p - k eigenvalues means that the eigenvalues are equal in magnitude.

Experiment I. In experiment I the factor scores were taken from a population that was normally distributed with zero mean and unit variance. The variables that were factor analyzed were x_{1i}, $i = 1, 2, \ldots, 9$. After finding the eigenvalues for the correlation matrix of these variables, Bartlett's test of significance was applied to test the following null hypotheses:

H_0 - 1N. The variables x_{1i}, $i = 1, 2, \ldots, 9$, do not have one common factor.

H_0 - 2N. The variables x_{1i}, $i = 1, 2, \ldots, 9$, do not have two common factors.

H_0 - 3N. The variables x_{2i}, $i = 1, 2, \ldots, 9$, do not have three common factors.

Experiment II. In experiment II the factor scores were taken from a population that was distributed as the upper half of a unit standard normal population. The variables analyzed were the same as in experiment I. Bartlett's test was used to test the following null hypotheses:

H_0 - $1N_p$. The variables x_{1i}, $i = 1, 2, \ldots, 9$, do not have one common factor.

H_0 - $2N_p$. The variables x_{1i}, $i = 1, 2, \ldots, 9$, do not have two common factors.

H_0 - $3N_p$. The variables x_{1i}, $i = 1, 2, \ldots, 9$, do not have three common factors.

Experiment III. In experiment III the population from which the common factor scores were taken was distributed approximately as a chi square distribution with three degrees of freedom.

A factor analysis was performed on the variables x_{1i}, $i = 1, 2, \ldots, 9$, and Bartlett's test was used to test the following null hypotheses:

H_{01} - 1Cs. The variables x_{1i}, $i = 1, 2, \ldots, 9$, do not have one common factor.

H_{01} - 2Cs. The variables x_{2i}, $i = 1, 2, \ldots, 9$, do not have two common factors.

H_{01} - 3Cs. The variables x_{1i}, $i = 1, 2, \ldots, 9$, do not have three common factors.

<u>Experiment IV</u>. Experiment IV was a factor analysis of the variables x_{1i}, $i = 1, 2, \ldots, 9$, where the common factor scores were taken from a population that was distributed approximately as a t-distribution with five degrees of freedom. The following null hypotheses were tested with Bartlett's test:

H_{01} - 1T. The variables x_{1i}, $i = 1, 2, \ldots, 9$, do not have one common factor.

H_{01} - 2T. The variables x_{1i}, $i = 1, 2, \ldots, 9$, do not have two common factors.

H_{01} - 3T. The variables x_{1i}, $i = 1, 2, \ldots, 9$, do not have three common factors.

<u>Experiment V</u>. In experiment V an investigation was begun of the random variables x_{2i}, $i = 1, 2, \ldots, 7$. A factor analysis was made of these variables, where the common factor scores were taken from a population that was approximately distributed as a normal distribution with zero mean and unit variance. The following null hypotheses were tested using Bartlett's test:

H_{02} - 1N. The variables x_{2i}, i = 1, 2, ..., 7, do not have one common factor.

H_{02} - 2N. The variables x_{2i}, i = 1, 2, ..., 7, do not have two common factors.

<u>Experiment VI</u>. Experiment VI was a factor analysis of the variables x_{2i}, i = 1, 2, ..., 7, where the common factor scores were taken from a population that was distributed approximately as the positive half of a unit standard normal population. Bartlett's test of significance was applied to test the following null hypotheses:

H_{02} - 1Np. The variables x_{2i}, i = 1, 2, ..., 7, do not have one common factor.

H_{02} - 2Np. The variables z_{2i}, i = 1, 2, ..., 7, do not have two common factors.

<u>Experiment VII</u>. An investigation was made in experiment VII of the variables x_{2i}, i = 1, 2, ..., 7, where the common factor scores were taken from a population that was distributed as a chi square distribution with three degrees of freedom. The null hypotheses tested in experiment VII were as follows:

H_{02} - 1Cs. The variables x_{2i}, i = 1, 2, ..., 7, do not have one common factor.

H_{02} - 2Cs. The variables x_{2i}, i = 1, 2, ..., 7, do not have two common factors.

<u>Experiment VIII</u>. The final investigation of the variables x_{2i}, i = 1, 2, ..., 7, was carried out in experiment VIII. In this experiment the variables were considered to have factor scores

taken from a population that was distributed approximately as a t-distribution with five degrees of freedom. The following null hypotheses were tested:

H_{02} - 1T. The variables x_{2i}, i = 1, 2, ..., 7, do not have one common factor.

H_{02} - 2T. The variables x_{2i}, i = 1, 2, ..., 7, do not have two common factors.

Experiment IX. Experiment IX was a factor analysis of the variables x_{3i}, i = 1, 2, ..., 5, where the common factor scores were taken from a unit standard normal population. The following null hypothesis was tested:

H_{03} - 1N. The variables x_{3i}, i = 1, 2, ..., 5, do not have one common factor.

Experiment X. Experiment X was a factor analysis of the random variables considered in experiment IX, where the common factor scores were taken from a population that was approximately distributed as the positive half of a normal distribution with zero mean and unit variance. The following null hypothesis was tested:

H_{03} - 1Np. The variables x_{3i}, i = 1, 2, ..., 5, do not have one common factor.

Experiment XI. An investigation was made in experiment XI of the variables x_{3i}, i = 1, 2, ..., 5, where the common factor scores were taken from a population that was distributed approximately as a chi square distribution with three degrees of freedom. The null hypothesis tested in this experiment was:

H_{03} - 1Cs. The variables x_{3i}, i = 1, 2, ..., 5, do not have one common factor.

Experiment XII. The final experiment of the present study was a factor analysis of the variables x_{3i}, $i = 1, 2, \ldots, 5$, where the common factor scores were taken from a population that was distributed approximately as a t-distribution with five degrees of freedom. The following null hypothesis was tested using Bartlett's test of significance:

H_{03} - 1T. The variables x_{3i}, $i = 1, 2, \ldots, 5$, do not have one common factor.

The twenty four null hypotheses listed above were tested at the 0.01 level of significance. In the cases where a factor was found not to be significant in a given experiment, no comment was made for any factor extracted after the insignificant factor.

CHAPTER III

RESULTS

For each of the null hypotheses tested in which the Bartlett's chi square test was used, a two-tailed test was used. The number of degrees of freedom for the chi square value was taken to be as Bartlett suggested, that is, $\frac{1}{2}(p - k - 1)(p - k + 1)$, where p is the number of variables and k is the number of factors that have been extracted (1951, p. 1). The criterion value for all of the tables involving the chi square test of Bartlett was set at the 0.01 level of significance. The chi square values used most often in this study are given in Table 1 for reference. Therefore, a null hypothesis was rejected when the chi square test resulted in a value greater than the value given in Table 1, which shows critical chi square values for the 0.01 level with the appropriate degrees of freedom, i.e., p (probability of chance occurrence) less than 0.01.

The critical values in Table 1 were taken from Owen (1962, p. 51). In the majority of the cases reported in this study, significance was obtained when the number of degrees of freedom was taken from Bartlett (1950, p. 78) or Bartlett (1951, p. 1).

Table 26 and Table 27, found at the end of this chapter, contain the means and standard deviations respectively for experiments

Table 1

Chi-Square Values for the 0.01 Level of Significance*

D.F.	Chi-square value
5	16.750
9	23.589
10	25.188
14	31.319
20	39.997
21	41.401
27	49.645
35	60.275
36	61.581

*Chi-Square values given above are for a two-tailed test

I - XII. In Table 28, also found at the end of this chapter, are found the eigenvalues for the correlation matrices for the variables investigated in experiments I - XII.

Experiment I

In experiment I a principal component analysis was made of the variables x_{1i}, i = 1, 2, ..., 9, where the common factor scores were taken from N(0,1). The means and standard deviations of these variables were found to be approximately zero and one respectively. The upper triangular portion of the correlation matrix for these variables is given in Table 2.

In null hypothesis H_0-1N it is stated that the variables x_{1i}, i = 1, 2, ..., 9, based on a normal population, do not have one common factor. With k = 0, Table 3 shows a chi square value that is statistically significant. Therefore, the null hypothesis is rejected at the 0.01 level of significance. This conclusion is in agreement with what is known to be theoretically true.

In null hypothesis H_0-2N it is stated that the variables x_{1i}, i = 1, 2, ..., 9, based on a normal population, do not have two common factors, Table 3, with k = 1, shows a chi square value that is statistically significant. The conclusion was to reject this null hypothesis. This is in agreement with the theoretical structure of the variables investigated in experiment I.

With k = 2 the null hypothesis H_0-3N was tested. The null hypothesis states that the variables x_{1i}, i = 1, 2, ..., 9, based on a normal population, do not have three common factors.

Table 2

Correlation Matrix for the Variables in Experiment I

1.0000	-0.0474	0.0118	0.6108	0.8310	0.5555	0.7149	0.3943	0.6178
	1.0000	-0.1620	0.3725	0.3135	0.5159	0.5183	0.8852	-0.1826
		1.0000	0.5892	0.3461	0.5232	0.3004	-0.1364	0.7801
			1.0000	0.9326	0.9753	0.9386	0.6249	0.8295
				1.0000	0.9104	0.9653	0.6646	0.7764
					1.0000	0.9501	0.7238	0.7347
						1.0000	0.8039	0.6661
							1.0000	0.1199
								1.0000

Table 3

Chi Square Analysis of Common Factors in Experiment I

k	D.F.	Coef(k)	$\log_e(R_{p-k})$	Chi Square	p
0	36	-(95.1659)	-(25.4611)	2423.0317	< 0.01
1	35	-(94.4992)	-(19.8333)	1874.2319	< 0.01
2	27	-(93.8325)	-(14.9970)	1407.2026	< 0.01
3	20	-(93.1658)	-(0.6180)	57.5798	< 0.01
4	14	-(92.4991)	-(0.1388)	12.8420	> 0.01

Table 3 shows a chi square value that is significant at the 0.01 level of significance. Therefore, the null hypothesis was rejected. This was in agreement with what was known to be theoretically true.

Experiment II

A principal component analysis was made of the variables x_{1i}, i = 1, 2, ..., 9, where the common factor scores were taken from N(p,0,1). The means and standard deviations of these variables were found to be zero and one respectively. The upper triangular portion of the correlation matrix of these variables is given in Table 4.

The null hypothesis H_0-1Np states that the variables x_{1i}, i = 1, 2, ..., 9, based on a population distributed as the positive half of a normal population with zero mean and unit variance, do not have one common factor. With k = 0, Table 5 shows a chi square value that is in the rejection region. Therefore,

Table 4

Correlation Matrix for the Variables in Experiment II

1.0000	-0.0380	0.2223	0.6631	0.8342	0.6014	0.7385	0.3966	0.7001
	1.0000	-0.0956	0.4144	0.3526	0.5401	0.5299	0.8892	-0.1088
		1.0000	0.6739	0.5080	0.6197	0.4422	0.0097	0.8403
			1.0000	0.9535	0.9795	0.9531	0.6765	0.8474
				1.0000	0.9301	0.9733	0.6932	0.8206
					1.0000	0.9565	0.7571	0.7669
						1.0000	0.8166	9.7197
							1.0000	0.2139
								1.0000

Table 5

Chi Square Analysis of Common Factors in Experiment II

k	D.F.	Coef(k)	$\log_e(R_{p-k})$	Chi Square	p
0	36	-(95.1659)	-(26.5169)	2523.5049	$<$ 0.01
1	35	-(94.4992)	-(20.0851)	1898.0254	$<$ 0.01
2	37	-(93.8325)	-(14.7142)	1380.6702	$<$ 0.01
3	20	-(93.1658)	-(0.7574)	70.5596	$<$ 0.01
4	14	-(92.4991)	-(0.1470)	13.5981	$>$ 0.01

the null hypothesis was rejected.

In null hypothesis H_0-2Np it is stated that the variables x_{1i}, i = 1, 2, ..., 9, based on a population distributed as the positive half of a unit standard normal population, do not have two common factors. Table 5 shows a chi square value that is statistically significant. The null hypothesis was rejected at the 0.01 level of significance.

Experiment III

In experiment III a principal component analysis was made of the variables x_{1i}, i = 1, 2, ..., 9, where the common factor scores were taken from a chi square distribution with three degrees of freedom (a non-symmetric and skewed population). The means and standard deviations of these variables were found to be approximately zero and one respectively. The upper triangular portion of the correlation matrix of the variables is given in Table 6.

Table 6

Correlation Matrix for the Variables in Experiment III

1.0000	0.0304	-0.1678	0.4769	0.7661	0.4260	0.6609	0.4382	0.5071
	1.0000	0.1000	0.6004	0.4982	0.6962	0.6580	0.9000	0.0851
		1.0000	0.6005	0.3174	0.5672	0.3050	-0.0007	0.7506
			1.0000	0.9188	0.9816	0.9366	0.7300	0.8337
				1.0000	0.8950	0.9733	0.7624	0.7793
					1.0000	0.9354	0.7883	0.7619
						1.0000	0.8648	0.6972
							1.0000	0.2775
								1.0000

The null hypothesis H_{01}-1Cs states that the variables x_{1i}, i = 1, 2, ..., 9, based on a chi square distribution with three degrees of freedom, do not have one common factor. With k = 0, Table 7 shows a chi square value that is statistically

Table 7

Chi Square Analysis of Common Factors in Experiment III

k	D.F.	Coef(k)	$\log_e(R_{p-k})$	Chi Square	p
0	36	-(95.1659)	-(26.3122)	2504.0278	$<$ 0.01
1	35	-(94.4992)	-(19.5044)	1843.1553	$<$ 0.01
2	27	-(93.8325)	-(15.7694)	1479.6812	$<$ 0.01
3	20	-(93.1658)	-(0.8843)	82.3863	$<$ 0.01
4	14	-(92.4991)	-(0.1200)	11.0988	$>$ 0.01

significant at the 0.01 level of significance. Therefore, the null hypothesis was rejected.

In null hypothesis H_{01}-3Cs it is stated that the variables x_{1i}, i = 1, 2, ..., 9, based on a chi square distribution with three degrees of freedom, do not have three common factors. With k = 2, Table 7 shows a chi square value that is significant at the 0.01 level of significance. The conclusion was to reject the null hypothesis.

Experiment IV

A principal component analysis was made of the variables x_{1i}, i = 1, 2, ..., 9, where the common factor scores were taken from $t_{(5)}$, a leptokurtic population. The means and standard deviations of these variables were found to be approximately equal to zero and one respectively. The upper triangular portion of the correlation matrix of these variables is given in Table 8.

The null hypothesis H_{01}-1T states that the variables x_{1i}, i = 1, 2, ..., 9, based on a t-distribution with five degrees of freedom, do not have one common factor. With k = 0, Table 9 shows a chi square value that is statistically significant. The conclusion was to reject the null hypothesis at the 0.01 level of significance.

The null hypothesis H_{01}-2T states that the variables x_{1i}, i = 1, 2, ..., 9, based on a t-distribution with five degrees of freedom, do not have two common factors. With k = 1, Table 9 shows a chi square value that is large enough to reject this null hypothesis at the 0.01 level of significance.

In null hypothesis H_{01}-3T it is stated that the variables x_{1i}, i = 1, 2, ..., 9, based on a t-distribution with five degrees of freedom, do not have three common factors. With k = 2, Table 9 shows a chi square value that is statistically significant. Therefore, the null hypothesis was rejected at the 0.01 level of significance.

Experiment V

In experiment V a principal component analysis was made of the variables x_{2i}, i = 1, 2, ..., 7, where the common factor scores were taken from N(0,1). The means and standard deviations

Table 8

Correlation Matrix for the Variables in Experiment IV

1.0000	-0.0973	-0.1130	0.6358	0.8531	0.5642	0.7775	0.5109	0.6866
	1.0000	0.0741	0.4207	0.2928	0.5571	0.4360	0.7915	-0.0449
		1.0000	0.5287	0.2556	0.5045	0.2646	-0.0022	0.6302
			1.0000	0.9338	0.9776	0.9511	0.7517	0.8731
				1.0000	0.9009	0.9821	0.7699	0.8438
					1.0000	0.9401	0.8181	0.7910
						1.0000	0.8501	0.7902
							1.0000	0.3828
								1.0000

Table 9

Chi Square Analysis of Common Factors in Experiment IV

k	D.F.	Coef(k)	$\log_e(R_{p-k})$	Chi Square	p
0	36	-(95.1659)	-(26.0958)	2483.4280	$<$ 0.01
1	35	-(94.4992)	-(20.6911)	1955.2966	$<$ 0.01
2	27	-(93.8325)	-(16.3806)	1537.0281	$<$ 0.01
3	20	-(93.1658)	-(0.6713)	62.5467	$<$ 0.01
4	14	-(92.4991)	-(0.1294)	11.9702	$>$ 0.01

of these variables were found to be approximately equal to zero and one respectively. The upper triangular portion of the correlation matrix of these variables is given in Table 10.

Table 10

Correlation Matrix for the Variables in Experiment V

1.0000	-0.0474	0.7345	0.6689	0.6298	0.3943	0.7677
	1.0000	0.6180	0.6977	-0.7950	0.8852	-0.6669
		1.0000	0.9839	-0.0331	0.9000	0.1518
			1.0000	-0.1374	0.9365	0.0519
				1.0000	-0.4435	0.9734
					1.0000	-0.2687
						1.0000

In null hypothesis H_{02}-1N it is stated that the variables

x_{21}, i = 1, 2, ..., 7, based on a unit standard normal population, do not have one common factor. With k = 0, Table 11 shows a chi

Table 11

Chi Square Analysis of Common Factors in Experiment V

k	D.F.	Coef(k)	$\log_e(R_{p-k})$	Chi Square	p
0	21	-(95.8327)	-(20.8619)	1999.2527	< 0.01
1	20	-(95.1660)	-(18.2134)	1733.2939	< 0.01
2	14	-(94.4993)	-(0.4577)	43.2537	< 0.01
3	9	-(93.8326)	-(0.0905)	8.4925	> 0.01

square value that is statistically significant. The conclusion was to reject this null hypothesis at the 0.01 level of significance.

In null hypothesis H_{02}2N it is stated that the variables x_{2i}, i = 1, 2, ..., 7, based on a unit standard normal population, do not have two common factors. With k = 1, Table 11 shows a chi square value that is in the rejection region. Therefore, the null hypothesis was rejected.

Experiment VI

In experiment VI a principal component analysis was made of the variables x_{2i}, i = 1, 2, ..., 7, where the common factor scores were taken from N(p,0,1). The means and standard deviations of these variables were found to be approximately equal to zero and one respectively. The upper triangular portion of the correlation matrix of these variables is given in Table 12.

Table 12

Correlation Matrix for the Variables in Experiment VI

1.0000	-0.0380	0.7138	0.6454	0.6069	0.3966	0.7570
	1.0000	-0.6496	0.7266	-0.8066	0.8892	-0.6719
		1.0000	0.9840	-0.0928	0.9142	0.1058
			1.0000	-0.1970	0.9472	0.0042
				1.0000	-0.4673	0.9707
					1.0000	-0.2816
						1.0000

In null hypothesis H_{02}-1NP it is stated that the variables x_{2i}, i = 1, 2, ..., 7, based on a population distributed as the positive half of a unit standard normal population, do not have one common factor. With k = 0, Table 13 shows a chi square

Table 13

Chi Square Analysis of Common Factors in Experiment VI

k	D.F.	Coef(k)	$\log_e(R_{p-k})$	Chi Square	p
0	21	-(95.8327)	-(20.9959)	1999.2527	<0.01
1	20	-(95.1660)	-(18.1941)	1731.4600	<0.01
2	14	-(94.4993)	-(0.4519)	42.7062	<0.01
3	9	-(93.8326)	-(0.0731)	6.8580	>0.01

value that is statistically significant. The null hypothesis was rejected at the 0.01 level of significance.

In null hypothesis H_{02}-2Np it is stated that the variables x_{2i}, i = 1, 2, ..., 7, based on a population distributed as the positive half of a unit standard normal population, do not have two common factors. With k = 1, the chi square value found in Table 13 is statistically significant. Therefore, the null hypothesis was rejected at the 0.01 level of significance.

Experiment VII

In experiment VII a principal component analysis was made of the variables x_{2i}, i = 1, 2, ..., 7, where the common factor scores were taken from a chi square distribution with three degrees of freedom. The means and standard deviations of these variables were found to be equal approximately to zero and one respectively. The upper triangular portion of the correlation matrix of these variables is given in Table 14.

Table 14

Correlation Matrix for the Variables in Experiment VII

1.0000	0.0304	0.7291	0.6718	0.5523	0.4382	0.7306
	1.0000	0.6848	0.7500	-0.8060	0.9000	-0.6507
		1.0000	0.9861	-0.1380	0.9238	0.0882
			1.0000	-0.2288	0.9511	-0.0009
				1.0000	-0.4864	0.9648
					1.0000	-0.2762
						1.0000

In null hypothesis H_{02}-1Cs it is stated that the variables x_{2i}, i = 1, 2, ..., 7, based on a chi square distribution with three degrees of freedom, do not have one common factor. With k = 0, Table 15 shows a chi square value that is statistically

Table 15

Chi Square Analysis of Common Factors in Experiment VII

k	D.F.	Coef(k)	$\log_e(R_{p-k})$	Chi Square	p
0	21	-(95.8327)	-(20.0838)	1924.6816	< 0.01
1	20	-(95.1660)	-(16.7294)	1592.0728	< 0.01
2	14	-(94.4993)	-(0.4996)	47.2119	< 0.01
3	9	-(93.8326)	-(0.0565)	5.3007	> 0.01

significant at the 0.01 level of significance. The conclusion was to reject the null hypothesis.

In null hypothesis H_{02}-2Cs it is stated that the variables x_{2i}, i = 1, 2, ..., 7, do not have two common factors. Table 15, with k = 1, shows a chi square value that is statistically significant at the 0.01 level of significance. Therefore, the null hypothesis was rejected at the 0.01 level of significance.

Experiment VIII

A principal component analysis was made of the variables x_{2i}, i = 1, 2, ..., 7, where the common factor scores were taken from $t_{(5)}$, a letpokurtic distribution. The means and standard deviations of these variables were found to be approximately equal to zero and one respectively. The upper triangular portion of the correlation matrix of the variables is given in Table 16.

Table 16

Correlation Matrix for the Variables in Experiment VIII

1.0000	-0.0973	0.8204	0.7572	0.7275	0.5109	0.8462
	1.0000	0.4660	0.5642	-0.7433	0.7915	-0.6039
		1.0000	0.9847	0.2309	0.8994	0.4080
			1.0000	0.1189	0.9388	0.3047
				1.0000	-0.1972	0.9747
					1.0000	-0.0098
						1.0000

In null hypothesis H_{02}-1T it is stated that the variables x_{2i}, i = 1, 2, ..., 7, do not have one common factor. With k = 0, the chi square value in Table 17 is large enough to support the

Table 17

Chi Square Analysis of Common Factors in Experiment VIII

k	D.F.	Coef(k)	$\log_e(R_{p-k})$	Chi Square	p
0	21	-(95.8327)	-(21.3959)	2050.4241	$<$ 0.01
1	20	-(95.1660)	-(18.3177)	1743.2178	$<$ 0.01
2	14	-(94.4993)	-(0.4887)	46.1862	$<$ 0.01
3	9	-(93.8326)	-(0.0684)	6.4187	$>$ 0.01

conclusion to reject this null hypothesis at the 0.01 level of significance.

In null hypothesis H_{02}-2T it is stated that the variables x_{2i}, $i = 1, 2, \ldots, 7$, do not have two common factors. With $k = 1$, Table 17 shows a chi square value that is statistically significant. Therefore, the null hypothesis was rejected at the 0.01 level of significance.

Experiment IX

In experiment IX a principal component analysis was made of the variables x_{3i}, $i = 1, 2, \ldots, 5$, where the common factor scores were taken from N(0,1). The means and standard deviations of these variables were found to be approximately equal to zero and one respectively. The upper triangular portion of the correlation matrix of these variables is given in Table 18.

Table 18

Correlation Matrix for the Variables in Experiment IX

1.0000	-0.9872	0.9901	0.9905	-0.9813
	1.0000	-0.9785	-0.9841	0.9937
		1.0000	0.9819	-0.9697
			1.0000	-0.9788
				1.0000

In null hypothesis H_{03}-1N it is stated that the variables x_{3i}, $i = 1, 2, \ldots, 5$, do not have one common factor. With $k = 0$,

Table 19 shows a chi square value that is statistically significant

Table 19

Chi Square Analysis of Common Factors in Experiment IX

k	D.F.	Coef(k)	$\log_e(R_{p-k})$	Chi Square	p
0	10	-(96.4995)	-(16.0649)	1550.2512	< 0.01

at the 0.01 level of significance. Therefore, the null hypothesis was rejected.

Experiment X

In experiment X a principal component analysis was made of the variables x_{3i}, $i = 1, 2, \ldots, 5$, where the common factor scores were taken from $N(p,0,1)$. The means and standard deviations were found to be approximately equal to zero and one respectively. The upper triangular portion of the correlation matrix of these variables is given in Table 20.

Table 20

Correlation Matrix for the Variables in Experiment X

1.0000	-0.9871	0.9912	0.9905	-0.9812
	1.0000	-0.9802	-0.9841	0.9937
		1.0000	0.9832	-0.9716
			1.0000	-0.9788
				1.0000

In null hypothesis H_{03}-1NP it is stated that the variables x_{3i}, i = 1, 2, ..., 5, do not have one common factor. With k = 0, Table 21 shows a chi square value that is statistically significant. Therefore, the null hypothesis was rejected at the 0.01 level of significance.

Table 21

Chi Square Analysis of Common Factors in Experiment X

k	D.F.	Coef(k)	$\log_e(R_{p-k})$	Chi Square	p
0	10	-(96.4995)	-(16.1886)	1562.1899	<0.01

Experiment XI

A principal component analysis was made of the variables x_{3i}, i = 1, 2, ..., 5, where the common factor scores were taken from a chi square distribution with three degrees of freedom. The means and standard deviations of these variables were found to be equal approximately to zero and one respectively. The upper triangular portion of the correlation matrix of these variables is given in Table 22.

The null hypothesis H_{03}-1Cs states that the variables x_{3i}, i = 1, 2, ..., 5, do not have one common factor. With k = 0, Table 23 shows a chi square value that is statistically significant. Therefore, the conclusion was to reject the null hypothesis at the 0.01 level of significance.

Table 22

Correlation Matrix for the Variables in Experiment XI

1.0000	-0.9870	0.9902	0.9902	-0.9808
	1.0000	-0.9785	-0.9837	0.9937
		1.0000	0.9819	-0.9697
			1.0000	-0.9780
				1.0000

Table 23

Chi Square Analysis of Common Factors in Experiment XI

k	D.F.	Coef(k)	$\log_e(R_{p-k})$	Chi Square	p
0	10	-(96.4995)	-(14.7069)	1419.2073	< 0.01

Experiment XII

In experiment XII a principal component analysis was made of the variables x_{3i}, $i = 1, 2, \ldots, 5$, where the common factor scores were taken from $t_{(5)}$. The means and standard deviations of the variables x_{3i}, $i = 1, 2, \ldots, 5$, were found to be approximately equal to zero and one respectively. The upper triangular portion of the correlation matrix of these variables is given in Table 24.

In null hypothesis H_{03}-1T it is stated that the variables x_3i, $i = 1, 2, \ldots, 5$, do not have one common factor. With

Table 24

Correlation Matrix for the Variables in Experiment XII

1.0000	-0.9922	0.9938	0.9940	-0.9885
	1.0000	-0.9867	-0.9903	0.9962
		1.0000	0.9887	-0.9812
			1.0000	-0.9869
				1.0000

k = 0, Table 25 shows a chi square value that is statistically significant. Therefore, the conclusion was to reject this null hypothesis at the 0.01 level of significance.

Table 25

Chi Square Analysis of Common Factors in Experiment XII

k	D.F.	Coef(k)	$\log_e(R_{p-k})$	Chi Square	p
0	10	-(96.4995)	-(15.8255)	1527.1533	< 0.01

Table 26

Means for the Variables in Experiments I-XII

Experiment	Means								
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
1	-0.0904	-0.0601	-0.1177	-0.1514	-0.1399	-0.1486	-0.1371	-0.0855	-0.1450
2	0.0186	-0.0495	-0.1108	-0.0779	-0.0438	-0.0825	-0.0505	-0.0288	-0.0738
3	-0.1487	-0.0316	0.0858	-0.0400	-0.0929	-0.0379	-0.0847	-0.0846	-0.0163
4	0.0708	-0.1062	-0.0877	-0.0613	-0.0172	-0.0770	-0.0391	-0.0579	-0.0240
5	-0.0904	-0.0601	-0.1010	-0.1004	-0.0029	-0.0855	-0.0327		
6	0.0186	-0.0495	-0.0124	-0.0198	0.0540	-0.0288	0.0431		
7	-0.1487	-0.0316	-0.1252	-0.1178	-0.0610	-0.0846	-0.0959		
8	0.0708	-0.1062	-0.0119	-0.0276	0.1312	-0.0579	0.1203		
9	-0.0878	0.0774	-0.0903	-0.0889	0.0760				
10	0.0223	-0.0328	0.0187	0.0207	-0.0336				
11	-0.1468	0.1363	-0.1486	-0.1476	0.1346				
12	0.0750	-0.0855	0.0709	0.0731	-0.0860				

Table 27

Standard Deviations for the Variables in Experiments I-XII

Experiment	Standard Deviations								
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
1	1.0414	0.9660	0.9474	0.9116	0.9523	0.9132	0.9246	0.9524	0.9729
2	1.0317	1.0232	0.9243	1.0207	1.0448	1.0025	1.0178	1.0098	1.0614
3	0.8418	0.9947	1.0845	1.0417	0.9581	1.0634	0.9768	0.9706	1.0243
4	1.0025	1.1054	1.0324	0.9675	0.9635	0.9564	0.9675	1.0612	1.0124
5	1.0414	0.9660	0.9649	0.9822	1.0498	0.9524	1.0485		
6	1.0317	1.0232	1.0054	1.0042	1.0510	1.0098	1.0521		
7	0.8418	0.9947	0.9236	0.9401	0.9280	0.9706	0.8909		
8	1.0025	1.1054	0.9950	0.9849	1.1264	1.0612	1.1104		
9	1.0357	1.0189	1.0330	1.0313	1.0208				
10	1.0313	1.0266	1.0656	1.0316	1.0222				
11	0.8442	0.8648	0.8639	0.8497	0.8824				
12	0.9999	0.9933	1.0047	1.0028	0.9999				

Table 28

Eigenvalues of Correlation Matrices in Experiments I-XII

Experiment	Eigenvalues								
	t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
1	5.8237	2.0767	1.0369	0.0216	0.0114	0.0091	0.0073	0.0072	0.0059
2	6.1467	1.9435	0.8541	0.0207	0.0099	0.0077	0.0062	0.0060	0.0050
3	6.2853	1.5982	1.0581	0.0233	0.0097	0.0074	0.0065	0.0061	0.0052
4	5.7270	2.0435	1.1735	0.0201	0.0098	0.0078	0.0069	0.0064	0.0050
5	3.9280	3.0204	0.0194	0.0105	0.0080	0.0077	0.0058		
6	4.0166	2.9332	0.0191	0.0099	0.0078	0.0074	0.0059		
7	4.3110	2.6278	0.0243	0.0114	0.0095	0.0088	0.0072		
8	4.1685	2.7847	0.0183	0.0088	0.0076	0.0066	0.0055		
9	4.9343	0.0374	0.0164	0.0065	0.0053				
10	4.9366	0.0360	0.0158	0.0063	0.0053				
11	4.9076	0.0522	0.0236	0.0095	0.0072				
12	4.0303	0.0397	0.0173	0.0070	0.0057				

CHAPTER IV

DISCUSSION

According to Kaiser (1959), in principal component analysis, the number of common factors needed for a complete factor analysis is equal to the number of eigenvalues of the correlation matrix that are greater than or equal to one. The results of this study support that assumption. In each experiment, except experiment II, Table 28 shows that the number of eigenvalues greater than or equal to one is equal to the number of common factors assumed initially in the given experiment.

In each experiment, except those in which one common factor was assumed, Bartlett's test of significance showed that there was one significant common factor more than was assumed initially. That is, in those cases where two and three common factors were assumed to exist among the variables, Bartlett's test showed the presence of exactly three and four common factors respectively. Since it was assumed initially in each experiment that each variable had an error factor or a residual present and since the percentage of variance explained by the additional factor was small, the additional factor may be assumed to be due to error. In those experiments for which one common factor was assumed, correlation coefficients among the variables were extremely high. No correlation coefficient was less than 0.9.

Delimitations of the Study

The empirical investigation reported in the present study were limited in the following ways:

1) Only artificial samples were used. Measurements were random variates simulated on an IBM 360 computer.

2) The number of common factors for a set of observable random variables was limited to one, two, or three common factors.

3) The study was concerned with statistical significance and not with practical significance.

4) The populations from which the common factor scores were taken were limited to one of the following types:

a) A t-distribution with five degrees of freedom

b) The population that is distributed as the positive half of a unit standard normal distribution

c) A chi square distribution with three degrees of freedom

d) A normal distribution with zero mean and unit variance

5) Only Bartlett's test of significance was used to test for the significance of the number of factors.

6) The procedure for computation of factor loadings was limited to the method of principal component analysis.

Suggestions for Further Research

With the above list of delimitations, it is clear that this investigation holds no final answers to the question of how robust Bartlett's test of significance is in principal component analysis. More research is needed for cases where the number of observations is less than one hundred and for the cases where the number of observations is greater than one hundred. The study by Fuller and Hemmerle (1966) may hold some of the answers for the latter case. In that experiment, they showed that Lawley's chi square test, when used to test whether or not a factor analysis is complete in maximum likelihood factor analysis, was insensitive to the type of population from which common factor scores were taken. The number of observations in their study was two hundred.

One of the initial assumptions of the method of principal component analysis is that the communality of the variables is a numerical value of one, or at least that numerical values of one are used along the diagonal of the correlation matrix of the variables that are to be factor analyzed. Although this assumption was made in the present study, further research is needed in cases where the communality of each variable is not approximately one, since in most practical situations this is not the case.

Bartlett (1950) points out that his justification for testing the significance of factors in principal component analysis is not as complete as one would desire. The justification would be more realistic, according to Bartlett, if the researcher were considering the analysis of test scores known to have true equal variances, but standardized to unity only for the mean variance. In this case Bartlett's test could be given more justification, but would involve a variation in the number of degrees of freedom. Bartlett further notes that pending a more detailed investigation, some doubt remains as to whether or not the reduction in degrees of freedom that ensues from the individual standardization of the tests is automatically felt in the residual factor components (an assumption implicit in the proposed test), or is mainly absorbed by the larger roots of the characteristic equation of the correlation matrix.

The chi square analyses reported here showed significance of the factors extracted to be unquestionable for the first one or two factors extracted. It was for the factors later extracted that significance was not as positive. It appears that research on later-extracted factors, based on chi square distributions would provide some further data on Bartlett's test in addition to that presented in the present investigation. If sampling distributions for the Bartlett chi square statistic were obtained for the cases where the factor scores were normally distributed, distributed as the positive half of a unit standard normal distribution, distributed as a chi distribution with three degrees of freedom, and distributed as a t-distribution with five degrees of freedom, then one could use some statistical test to determine whether or not the sampling distributions were taken from the same population. The Kruskal and Wallis multi-sample test for identical populations would be an appropriate test.

The investigation reported here was concerned, in each of the cases where one, two, and three common factors were assumed initially, with only one theoretical model. Another apparent limitation of the present study is that stronger inferences could be made if models were varied. This was one of the important features of the study by Fuller and Hemmerle (1966).

With regard to the number of observations needed to make inferences about the significance of the factors extracted in a factor analysis, Burt (1952) points out that previous investigations in factor analysis suggest that generally speaking, when the number of variables is from ten to twenty, at least twenty observations are needed to establish one factor, fifty observations are needed to establish two factors, one hundred observations are needed to establish three factors, and between two hundred and four hundred observations are needed to establish four factors. Burt further suggests that it is not usually wise to try to extract more than four factors with any single battery of tests or in any single research. Ultimately, research on the robustness of Bartlett's test must include an estimate of the number of observations needed to establish the number of observations needed to establish any number of factors. Research in this area is needed.

Finally, no mention is made in any of the articles in the literature reported in this study of any tests made on artificial data to see if the variables are linear combinations of the factors. This is an essential assumption on the model for factor analysis in general and principal component analysis in particular. Research is needed to determine whether or not the failure of Bartlett's test of significance being sensitive to the population from which the common factor scores are taken is due to the variables not being linear combinations of the common factors.

The above mentioned areas of needed research in the area of tests of significance are merely a few of the ones mentioned in the literature. Tests of significance in factor analysis are quite nebulous and no general agreement seems to exist as to how one should establish the significance of a factor extracted in a given analysis.

Implications for Theory and Practice

Theoretically, in Bartlett's test of significance, the statistics given in (VII) and (VIII) of chapter one are approximately distributed as a chi square variable, if the number of observations is large. Bartlett (1950, pp. 82-83) proves this assertion on the basis of the factor scores and hence the variables being distributed normally. The results of this study point out that the test derived by Bartlett may not be distribution free and may depend, without exception, on the normality assumption.

The researcher is not always able to determine whether the variables, common factors, or specific factors come from populations that are normally distributed. The research reported here and in similar related research give evidence that the researcher should not make inferences using Bartlett's test of significance unless he is sure that the normality assumption is sufficiently satisfied. Therefore, practically speaking the test derived by Bartlett should be used with caution.

Solomon (1960, p. 310) states that the structure of the model for principal component analysis is the same as the general factor analysis model except that, at least at first, one does not distinguish between common and specific factors. The assumption in a model proposed by Henrysson (1950, p. 125) for example, is the existence of error factors caused by errors of measurements. Therefore, no inferences should be made about the presence of specific factors on the basis of Bartlett's test and the method of principal component analysis. Such inferences should be made on the basis of other information inherent in the data for analysis.

CHAPTER V

SUMMARY AND CONCLUSIONS

The basis of factor analysis is that if two random variables or activities involve a common factor or element, then there will be a correlation between them. If one gives two psychological tests, for example, to a group of people, then the correlation between these tests will be a function of the extent to which the tests are calling on common abilities (Adcock, 1954, p. 19).

The principal component model assumes, among other assumptions given earlier, that the populations from which the common factor scores are taken are normally distributed. This implies that the random variables or activities possessing these common factors are also normally distributed.

Bartlett's test of significance is formulated to indicate the significance of the factors not yet extracted. Since only the correlation structure of the variables is being investigated in its relation to variance, no significance can ever be attached to the factor corresponding to the smallest eigenvalue of the correlation matrix of the variables, for this would be equivalent to asking for the correlation structure of a single variable (Bartlett, 1950, p 80).

The experiments performed in the present study were designed to study some effects on Bartlett's test of significance when the common factor scores were not taken from populations that were normally distributed. The populations considered were 1) a t-distribtuion with five degrees of freedom ($t_{(5)}$), 2) a chi square distribution with three degrees of freedom ($\chi_{(3)}$), and 3) a population distributed as the positive half of a unit standard normal variable (N(p,0,1)). The populations were artificially simulated on an IBM 360 computer using concepts of power residues. The number of observations for each experiment was one hundred, the number of variables was five, seven, or nine, and the dimensions of the correlation matrices investigated were 5 x 5, 7 x 7, and 9 x 9.

There were three models considered for analysis. In one model there were five random variables, where one common factor was assumed to exist among the variables. A second model involved seven random variables, where two common factors were assumed to exist among the variables. The third model considered consisted of nine random variables, where three common factors were assumed to exist among the variables. For each variable, in a given model, the existence of an error factor was assumed initially. Bartlett's test of significance was used to test the significance of factors extracted in the following cases:

1) the common factor scores were taken from N(0,1)

2) the common factor scores were taken from $t_{(5)}$

3) the common factor scores were taken from N(p,0,1)

4) the common factor scores were taken from a chi square distribution with three degrees of freedom.

The error factors in each experiment were assumed to be taken from N(0,1).

Conclusions

Although the results of this study are necessarily limited, they do strongly indicate that when the method of principal component analysis is used, Bartlett's test of significance is relatively insensitive to departure from normality of the distribution of the common factor scores for large numbers of observations.

The results of the present study point out that the number of common factors in principal component analysis is approximately equal to the number of eigenvalues of the correlation matrix of the variables that are factor analyzed. This approximation of the number of common factors is a good initial guess of the number of common factors. Further inferences about the number of common factors can be made with Bartlett's test of significance if the researcher is sure that the variables are normally distributed.

REFERENCES

Adcock, C. J. Factorial Analysis for Non-Mathematicians. New York: Cambridge University Press, 1954.

Ahmavaara, Yrjo and Markkanen, Touko. The Unified Factor Model. Helsinki Finland: Uudenmaan Kirjapaino, 1958.

American Psychological Association. Publication Manual of the American Psychological Association. 1967 Revision. Washington, D. C.: American Psychological Association, Inc., 1967.

Anderson, Barry F. The Psychology Experiment. Belmont, California: Wadsworth Publishing Co., Inc., 1966.

Anderson, T. W. and Rubin, Herman. Statistical Inference in Factor Analysis. Third Berkeley Symposium on Mathematical Statistics and Probability. Edited by Jerzy Neyman, New York: University of California Press, 1956.

Bartlett, M. S. The Vector Representation of a Sample. Proceedings of the Cambridge Philosophical Society, 1934, 30, 327-340.

Bartlett, M. S. Further Aspects of the Theory of Multiple Regression. Proceedings of the Cambridge Philosophical Society, 1937, 34, 33-40.

Bartlett, M. S. Internal and External Factor Analysis. British Journal of Statistical Psychology, 1948, 1, 73-81.

Bartlett, M. S. Tests of Significance in Factor Analysis. British Journal of Statistical Psychology, 1950, 3, 77-85.

Bartlett, M. S. A Further Note on Tests of Significance in Factor Analysis. British Journal of Statistical Psychology, 1951, 4, 1-2.

Bevington, Philip R. Data Reduction and Error Analysis for the Physical Sciences. New York: McGraw-Hill Book Co., 1969.

Box, G. E. A General Distribution Theory for a Class of Likelihood Criteria. Biometrika, 1949, 36, 317.

Bradley, James V. Distribution-Free Statistical Tests. Englewood Cliffs, New Jersey: Prentice Hall Inc., 1968.

Burford, Roger L. Statistics: A Computer Approach. Columbus, Ohio: Charles E. Merrill Publishing Company, 1968.

Burt, Cyril. Tests of Significance in Factor Analysis. British Journal of Statistical Psychology, 1952, 5, 109-133.

Cattell, Raymond B. Factor Analysis. New York: Harper & Brothers, Publishers, 1952.

Chakravarti, I. M., Laha, R. G., and Roy, J. Handbook of Methods of Applied Statistics. New York: John Wiley & Sons, Inc., 1967.

Cooley, William W. and Lohnes, Paul R. Multivariate Procedures for the Behavioral Sciences. New York: John Wiley & Sons, Inc., 1962.

Coombs, Clyde. A Criterion for Significant Common Factor Variance. Psychometrika. 1941, 6, 267-272.

Coombs, Clyde. A Theory of Data. New York: John Wiley & Sons, Inc., 1964.

Cramer, Harald. Mathematical Methods of Statistics. Princeton, New Jersey: Princeton University Press, 1963.

Dempster, A. P. Continuous Multivariate Analysis. Reading, Massachusetts: Addison-Wesley Publishing Co., 1969.

Downie, N. M. and Heath, R. W Basic Statistical Methods. New York: Harper & Row, Publishers, 1965.

Dubois, Philip H. Multivariate Correlational Analysis. New York: Harper & Brothers, Publishers, 1957.

Ferguson, George. Statistical Analysis in Psychology and Education. New York: McGraw-Hill Book Company, 1966.

Fieller, E. C., Lewis, T., and Pearson, E. S. Correlated Random Normal Deviates. In E. S. Pearson (Ed.), Tracts for Computers. New York: Cambridge University Press, 1955.

Fruchter, Benjamin. Introduction to Factor Analysis. Princeton, New Jersey: D. Van Nostrand Company, Inc., 1954.

Guilford, J. P. Psychometric Methods. New York: McGraw-Hill Book Company, 1954.

Gulliksen, Harold. Theory of Mental Tests. New York: John Wiley & Sons, Inc., 1958.

Hamming, R. W. Numerical Methods for Scientists and Engineers. New York: McGraw-Hill Book Company, 1962.

Harman, Harry H. Modern Factor Analysis. Chicago: The University of Chicago Press, 1960.

Hays, William L. Statistics for Psychologists. New York: Holt, Rinehart and Winston, 1964.

Hemmerle, William J. Statistical Computations on a Digital Computer. Waltham, Massachusetts: Blaisdell Publishing Company, 1967.

Henrysson, Sten. The Significance of Factor Loadings. British Journal of Statistical Psychology, 1950, 3, 159-165.

Henrysson, Sten. Factor Analysis in the Behavioral Sciences. Stockholm Sweden: Almqvist & Wiksell, 1960.

Higman, Bryan. Applied Group-Theoretic and Matrix Methods. New York: Dover Publications, Inc., 1964.

Hodge, Bartow. Computers for Engineers. New York: McGraw-Hill Book Company, 1969.

Holzinger, Karl J. and Harman, Harry H. Factor Analysis. Chicago, Illinois: University of Chicago Press, 1941.

Horn. Use of Random Variables in Factor Analysis. British Journal of Statistical Psychology, 1966, 19, 127-129.

Hotelling, Harold. Analysis of a Complex of Statistical Variables into Principal Components. Journal of Educational Psychology, 1933, 24, 417-441, 498-520.

IBM Data Processing Techniques. Random Number Generation and Testing. White Plains, New York: Technical Publications Department, 1959.

Ito, Koichi. On the Effects of Heteroscedasticity and Nonnormality upon Some Multivariate Test Procedures. In Paruchuri R. Krishnaiah (Ed.), Multivariate Analysis II. New York: Academic Press, 1969.

Jansson, Birger. Random Number Generators. Stockholm, Sweden: Victor Pettersons Bokindustri Aktiebolag, 1966.

Joreskog, K. G. Some Contributions to Maximum Likelihood Factor Analysis. Psychometrika, 1967, 32, 443-482.

Joreskog, K. G New Methods in Maximum Likelihood Factor Analysis. *British Journal of Statistical Psychology*, 1968, 21, 85-96.

Kaiser, Henry F. Computer Program for Varimax Rotation in Factor Analysis. *Educational and Psychological Measurement*, 1951, 4, 1-2.

Kendall, M. G. and Stuart, A. *The Advanced Theory of Statistics*. London: Charles Griffin & Company Limited, 1958. 3 vols.

Kendall, M. G. and Smith, Babington, B. Random Sampling Numbers. In E. S. Pearson (Ed.), *Tracts for Computers*. New York: Cambridge University Press, 1951.

Kerlinger, Fred N. *Foundations of Behavioral Research*. New York: Holt, Rinehart and Winston, Inc., 1964.

Lancaster, H. O. *The Chi-Squared Distribution*. New York: John Wiley & Sons, Inc., 1969.

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.

Lawley, D. N. Further Investigation in Factor Estimation. *Proceedings of the Royal Society of Edinburgh*, 1941, 16, 176-185.

Lawley, D. N. The Application of the Maximum Likelihood Method to Factor Analysis. *British Journal of Statistical Psychological*, 1943, 33, 172-175.

Lawley, D. N. and Swanson, Z Tests of Significance in a Factor Analysis of Artificial Data. *British Journal of Statistical Psychology*, 1954, 7, 75-79.

Lehman, Richard S. and Bailey, Daniel E. *Digital Computing*. New York: John Wiley & Sons, Inc., 1968.

Linquist, E. F. *Design and Analysis of Experiments in Psychology* and Education. Boston: Houghton Mifflin Company, 1953.

Magnusson, David. *Test Theory*. Reading, Massachusetts: Addison-Wesley Publishing Company, 1967.

Martin, Francis F. *Computer Modeling and Simulation*. New York: John Wiley & Sons, Inc., 1968.

McNemar, Q. On the Number of Factors. *Psychometrika*, 1942, 7, 9-18.

Mood, Alexander M. and Graybill, Franklin A. Introduction to the Theory of Statistics. New York: McGraw-Hill Book Company, 1963.

Moses, Lincoln E. and Oakford, Robert V. Tables of Random Permutations. Stanford, California: Stanford University Press, 1963.

Mosier, C. I. Influence of Chance Error on Simple Structure: An Empirical Investigation of the Effect of Chance Error and Estimated Communalities on Simple Structure in Factorial Analysis. Psychometrika, 1939, 4, 33-44.

Norton, D. W. An Empirical Investigation of Some Effects of Non-normality and Heterogeneity on the F-distribution. Unpublished doctoral dissertation, State University of Iowa, 1952.

Noether, Gottfried E. Elements of Nonparametric Statistics. New York: John Wiley & Sons, Inc., 1967.

Owen, D. B. Handbook of Statistical Tables. Reading, Massachusetts: Addison-Wesley Publishing Company, Inc., 1962.

Organick, Elliott I. A Fortran IV Primer. Reading, Massachusetts: Addison-Wesley Publishing Company, 1966.

Parzen, Emanuel. Modern Probability Theory and Its Applications. New York: John Wiley & Sons, Inc., 1960.

Ralston, A. and Wilf, H. S. Mathematical Methods for Digital Computers. New York: John Wiley and Sons, Inc., 1962.

Rao, Radhakrishna C. Estimation and Tests of Significance in Factor Analysis. Psychometrika, 1955, 20, 93-111.

Reyburn, H. A. and Taylor, J. G. Some Factors of Personality. British Journal of Psychology, 1939, 30, 151-165.

Scheffé, Henry. The Analysis of Variance. New York: John Wiley & Sons, Inc., 1959.

Saunders, D. R. Factor Analysis III. Desiderata in Experimental Design. Cited by Raymond B. Cattell, Factor Analysis. New York: Harper & Brothers, Publishers, 1952, p. 304.

Siegel, Sidney. Nonparametric Statistics. New York: McGraw-Hill Book Company, 1956.

Singh, Jagjit. Great Ideas of Operations Research. New York: Dover Publications, Inc., 1968.

Smouse, A. D. A Writing Style Supplement. Norman, Oklahoma: Oklahoma University Printing Press, 1969.

Solomon, Herbert. A Survey of Mathematical Models in Factor Analysis. In Herbert Solomon (Ed.), Mathematical Thinking in the Measurement of Behavior. Glencoe, Illinois: The Free Press of Glencoe, 1960.

System/360 Scientific Subroutine Package. Programmer's Manual. White Plains, New York: Technical Publications Department, 1968.

Thompson, Godfrey. The Factorial Analysis of Human Ability. Warwick Square, London: University of London Press, 1951.

Thurston, L. I. The Vectors of the Mind. Chicago, Illinois: University of Chicago Press, 1947.

Thurston, L. L. Primary Mental Abilities. Chicago, Illinois: University of Chicago Press, 1957.

Teichroew, Daniel and Sitgreaves, Rosedith. Computation of an Empirical Sampling Distribution for the W-Classification Statistic. In Herbert Solomon (Ed.), Studies in Item Analysis and Prediction. Stanford, California: Stanford University Press, 1961.

Weinberg, George H. and Schumaker, John A. Statistics an Intuitive Approach. Belmont, California: Wadsworth Publishing Company, Inc., 1966.

Whittle, P. On Principal Components and Least Square Methods of Factor Analysis. Skandinavisk Aktuarietidskrift, 1952, 35, 223-239.

Wilks, S. S. Certain Generalizations in the Analysis of Variance. Biometrika, 1932, 24, 471-494.

Wilks, S. S. Mathematical Statistics. New York: John Wiley & Sons, Inc., 1962.

Wishart, John. The Generalized Product Moment Distribution in Samples from a Normal Multivariate Population. Biometrika, 1928, 20A, 32-52.

Wold, Herman. Random Normal Deviates. In E. S. Pearson (Ed.), Tracts for Computers. London: Cambridge University Press, 1948.

Wold, Herman. Some Artificial Experiments in Factor Analysis. _Nordisk Psykologi's_, 1953, No. 3.

Wolfle, Dael. _Factor Analysis to 1940_. Chicago: The University of Chicago Press, 1940.

APPENDIX I

DEFINITION OF TERMS AND MEANING OF SYMBOLS

Definition of Terms

Bartlett's Test of Significance. The chi square statistics given in (VII) and (VIII) of Chapter I.

Common Factor. A factor involved in more than one variable of a given set of variables (Harman, 1960, p. 12).

Common Factor Score. A numerical value taken on by a common factor.

Communality. The total variance of a variable due to factors which the variable has in common with other variables in a set of variables (Cattell, 1952, p. 423).

Corre. A computer scientific subroutine that computes means, standard deviations, sums of cross-products of deviations from means, and Pearson product moment correlation coefficients. (System/360 Scientific Subroutine Package, 1968, p. 164).

Eigen. A computer scientific subroutine that computes the eigenvalues and eigen vectors of a real symmetric matrix (System/360 Scientific Subroutine Package, 1968, p. 164).

Factor. A hypothetical construct (Harman, 1960, p. 12).

Gauss. A computer scientific subroutine that computes a normally distributed random number with a given variance and a given mean (System/360 Scientific Subroutine Package, 1968, p. 77).

Load. A computer scientific subroutine that computes a factor matrix (loadings) from eigenvalues and associated eigen vectors (System/360 Scientific Subroutine Package, 1968, p. 56).

Principal Component Analysis. A method of factor analysis used to reduce the dimensionality of a problem in terms of the number of variables to be analyzed. Also a method of factor analysis used to classify the initial variables of a problem into sets such that variables within a given set have certain characteristics in common (Hemmerle, 1967, p. 140).

Random Variable. A symbol X standing for any of a range of number events, each having a probability p(X) (Hays, 1964, p. 109).

Randu. A computer scientific subroutine that computes a uniformly distributed random number between zero and one (System/360 Scientific Subroutine Package, 1968, p. 77).

Residual. A factor common to a variable due to error (Henrysson, 1950, p. 125).

Specific Factor. A factor involved in a single variable of a given set of variables (Harman, 1960, p. 12).

Trace. A computer scientific subroutine that computes the cumulative percentage of eigenvalues greater than or equal to a constant (System/360 Scientific Subroutine Package, 1968, p. 55).

Unit Standard Normal Random Variable. A random variable whose range of values are taken from a normal population with zero mean and unit variance.

Varmx. A computer scientific subroutine that performs orthogonal rotations on a factor matrix (System/360 Scientific Subroutine Package, 1968, p. 56).

Meaning of Symbols

N(u,v). A population that is normally distributed with mean u and variance v.

$t_{(n)}$. A population that is distributed as a t-distribution with n degrees of freedom.

N(p,u,v). A population that is distributed as the positive half of N(u,v).

K. The number of factors extracted.

Coef(k). The coefficient of $\log_e(R_{p-k})$ given in (VIII) of Chapter I.

E(X). The expected value of a random variable X.

V(X). The variance of a random variable V.

X_i. The ith variable considered in a given experiment.

t_i. The ith eigenvalue of a set of eigenvalues of a correlation matrix, where the eigenvalues have been arranged in descending order.

sk. The measure of the skewness of a distribution.

ku. The measure of the kurtosis of a distribution.

$$\Gamma(n + \tfrac{1}{2}) = (1.3.5\ldots(n - 1))\sqrt{\pi}/2^n$$

$$\Gamma\left[(n + 1)/2\right] = (1.3.5\ldots.(n - 1))\sqrt{\pi}/2^{\frac{1}{2}n}$$

$$\Gamma(\tfrac{1}{2}) = \sqrt{\pi}$$

$\chi^2_{(n)}$. A population that is distributed as a chi square distribution with n degrees of freedom.

$\sum_i$. Summation as i ranges from 1 to p.

$\sum_j$. Summation as j ranges from 1 to q.

$\{y_n\}$. The sequence $y_1, y_2, y_3, \ldots$

APPENDIX II

COMPUTER PROGRAM FOR FACTOR ANALYSIS

USED IN THIS STUDY

```
//FACTO     JOB ED010301,POLLARD,MSGLEVEL=1
// EXEC FORTGCLG,PARM.FORT='BCD'
//FORT.SYSIN DD *
      SUBROUTINE JOYCE(IX,S,AM,V)                                       JOYCE
C     THIS SUBROUTINE IS A REFINEMENT OF GAUSS WHERE K = 48             JOYCE
      A = 0.0                                                           JOYCE
      DO 50 I =1,48                                                     JOYCE
      CALL RANDU(IX,IY,Y)                                               JOYCE
      IX = IY                                                           JOYCE
   50 A = A + Y                                                         JOYCE
      BAY = A - 24.0                                                    JOYCE
      V = (BAY/2.0)*S + AM                                              JOYCE
      RETURN                                                            JOYCE
      END                                                               JOYCE
      SUBROUTINE DATA(M,D)                                              DATA
      DIMENSION D(1)                                                    DATA
   54 FORMAT(12F6.0)                                                    DATA
C      READ AN OBSERVATION FROM INPUT DEVICE                            DATA
      READ(5,54) (D(I),I=1,M)                                           DATA
      RETURN                                                            DATA
      END                                                               DATA
      DIMENSION BB(9),D(9),S(9),T(9),XBAR(9),R(45),V(81),TV(51),        FACTO
     1RHO(9),FACT(9),CHI(4),RPK(4),NDF(4),NNDF(4),BETA(4),TAU(9),       FACTO
C      ADD THESE DIMENSION CARDS FOR N = 100                            FACTO
     2TMAIN(100),INDEX(100),E(100,9),PM(100,9),X(100,9),XX(100,7),      FACTO
     3XXX(100,5),COEF(4),ANN(100),USER(100),A(100)                      FACTO
      DIMENSION ALPHA(4)                                                FACTO
      DIMENSION SFACT(5),TEMP(5)                                        FACTO
C     PROJECT NORMAL                                                    FACTO
C     RESEARCH IN FACTOR ANALYSIS                                       FACTO
C     STATEMENT OF THE PROBLEM IS AS FOLLOWS.                           FACTO
C     WHAT ARE SOME EFFECTS OF NON-NORMALITY ON BARTLETT'S TEST OF      FACTO
C     SIGNIFICANCE WHEN COMMON FACTOR SCORES ARE NOT NORMAL IN          FACTO
C     PRINCIPAL COMPONENT ANALYSIS                                      FACTO
C      PERMUTATION 1 FROM PAGE 93 BLOCK 1 MOSES AND OAKFORD TABLES      FACTO
```

```
C      PERMUTATION 2 FROM PAGE 93 BLOCK 7 MOSES AND OAKFORD TABLES       FACTO
C      PERMUTATION 3 FROM PAGE 95 BLOCK 1 MOSES AND OAKFORD TABLES       FACTO
C      PERMUTATION 4 FROM PAGE 97 BLOCK 1 MOSES AND OAKFORD TABLES       FACTO
C       PERMUTATION 5 FROM PAGE 97 BLOCK7 MOSES AND OAKFORD TABLES       FACTO
C      PERMUTATION 6 FROM PAGE 99 BLOCK 8 MOSES AND OAKFORD TABLES       FACTO
C      PERMUTATION 7 FROM PAGE 98 BLOCK 8 MOSES AND OAKFORD TABLES       FACTO
C      PERMUTATION 8 FROM PAGE 98 BLOCK 1 MOSES AND OAKFORD TABLES       FACTO
C       PERMUTATION 9 FROM PAGE 99 BLOCK 8 MOSES AND OAKFORD TABLES      FACTO
C      BEGIN GENERATION OF FACTOR SCORES                                 FACTO
C      GENERATE SPECIFIC FACTOR SCORES                                   FACTO
C      NS3 IS THE INPUT IX FOR GAUSS                                     FACTO
C      NS2 IS THE NO. OF COMMON FACTORS                                  FACTO
C      NS1 IS THE NO. OF VARIABLES-SAME AS  M                            FACTO
 1075 FORMAT(1H , 'ACCEPT AT UREF 0.05 LEVEL',I3,E20.8)                  FACTO
 1076 FORMAT(1H , 'REJECT AT UREF 0.05 LEVEL' ,I3,E20.8)                 FACTO
 1077 FORMAT(1H , 'ACCEPT AT UREF 0.01 LEVEL',I3,E20.8)                  FACTO
 1078 FORMAT(1H , 'REJECT AT UREF 0.01 LEVEL',I3,E20.8)                  FACTO
 1175 FORMAT(1H ,'ACCEPT AT REF 0.05 LEVEL',I3,E20.8)                    FACTO
 1176 FORMAT(1H ,'REJECT AT REF 0.05 LEVEL',I3,E20.8)                    FACTO
 1178 FORMAT (1H ,'REJECT AT REF 0.01 LEVEL',I3,E20.8)                   FACTO
  971 FORMAT(10H AAAAAAAAA,E20.10,4I3)                                   FACTO
  972 FORMAT(1H , ' ITER = ', I3)                                        FACTO
 1177 FORMAT(1H ,'ACCEPT AT REF 0.01 LEVEL',I3,E20.8)                    FACTO
      NS3 = 65537                                                        FACTO
      IX =NS3                                                            FACTO
      SX=1.0                                                             FACTO
      AM=0.0                                                             FACTO
      N=100                                                              FACTO
      DO 200 ID=1,N                                                      FACTO
      CALL JOYCF(IX,SX,AM,VX)                                            FACTO
      TMAIN(ID) = VX                                                     FACTO
  200 CONTINUE                                                           FACTO
      RHO(1) = 0.8                                                       FACTO
      RHO(2) =0.75                                                       FACTO
      RHO(3) =0.54                                                       FACTO
```

```
      RHO(4) =0.39                                                      FACTO
      RHO(5) =0.65                                                      FACTO
      RHO(6) = 0.7                                                      FACTO
      RHO(7) = 0.5                                                      FACTO
      RHO(8) = 0.4                                                      FACTO
      RHO(9) =0.49                                                      FACTO
C     MM IS THE INITIAL VALUE OF M                                      FACTO
      MM=9                                                              FACTO
      DO 206 J=1,MM                                                     FACTO
  202 READ(5,201)   (INDEX(L), L=1,N)                                   FACTO
      DO 205 I=1,N                                                      FACTO
  201 FORMAT(10I5)                                                      FACTO
      JF=INDEX(I)                                                       FACTO
      PM(I,J) = TMAIN(JF)                                               FACTO
      E(I,J) = RHO(J)*TMAIN(I)+SQRT(1.0 - RHO(J)**2)*PM(I,J)            FACTO
  205 CONTINUE                                                          FACTO
  206 CONTINUE                                                          FACTO
C     GENERATE COMMON FACTOR SCORES                                     FACTO
      IX = 65537
      INDIA = 4                                                         FACTO
      DO 9031 NIX = 1,INDIA                                             FACTO
      IF(NIX - 1) 109,8212,8220                                         FACTO
 8220 IF(NIX - 2) 109,8214,8221                                         FACTO
 8221 IF(NIX - 3) 109,8216,8222                                         FACTO
 8222 IF(NIX - 4) 109,8218,109                                          FACTO
 8212 SX = 1.0                                                          FACTO
      DO 8210 IP = 1,N                                                  FACTO
      NS2 = 3                                                           FACTO
      DO 211 I=1,NS2                                                    FACTO
      CALL JOYCE(IX,SX,AM,VX)                                           FACTO
      FACT(I)=VX                                                        FACTO
  211 CONTINUE                                                          FACTO
      X(IP,1)=0.98489*FACT(1) +0.17321*E(IP,1)                          FACTO
      X(IP,2)=0.98995*FACT(2) + 0.14142*E(IP,2)                         FACTO
      X(IP,3)=0.99499*FACT(3) + 0.10000*E(IP,3)                         FACTO
```

```
      X(IP,4)=0.57446*FACT(1)+0.50000*FACT(2)+0.64031*FACT(3)+0.10000   FACTO
     1*E(IP,4)                                                           FACTO
      X(IP,5)=0.80000*FACT(1)+0.43589*FACT(2)+0.40000*FACT(3)+0.10000   FACTO
     1*E(IP,5)                                                           FACTO
      X(IP,6)=0.50000*FACT(1)+0.60828*FACT(2)+0.60000*FACT(3)+0.14142   FACTO
     1*E(IP,6)                                                           FACTO
      X(IP,7)=0.70000*FACT(1)+0.60000*FACT(2)+0.37417*FACT(3)+0.10000   FACTO
     1*E(IP,7)                                                           FACTO
      X(IP,8)=0.42426*FACT(1)+0.90000*FACT(2)+0.10000*E(IP,8)           FACTO
      X(IP,9)=0.59161*FACT(1)+0.80000*FACT(3)+0.10000*E(IP,9)           FACTO
      XX(IP,1)=0.98489*FACT(1)+0.17321*E(IP,1)                          FACTO
      XX(IP,2)=0.98995*FACT(2)+0.14142*E(IP,2)                          FACTO
      XX(IP,3)=0.73485*FACT(1)+0.67082*FACT(2)+0.10000*E(IP,4)          FACTO
      XX(IP,4)=0.65574*FACT(1)+0.74162*FACT(2)+0.14142*E(IP,6)          FACTO
      XX(IP,5)=0.59161*FACT(1)-0.80000*FACT(2)-0.10000*E(IP,7)          FACTO
      XX(IP,6)=0.42426*FACT(1)+0.90000*FACT(2)+0.10000*E(IP,8)          FACTO
      XX(IP,7)=0.74833*FACT(1)-0.65574*FACT(2)+0.10000*E(IP,9)          FACTO
      XXX(IP,1)=0.99499*FACT(1)+0.10000*E(IP,1)                         FACTO
      XXX(IP,2)=-0.99499*FACT(1)+0.10000*E(IP,2)                        FACTO
      XXX(IP,3)=0.98489*FACT(1)+0.17321*E(IP,3)                         FACTO
      XXX(IP,4)=0.98995*FACT(1)+0.14142*E(IP,4)                         FACTO
      XXX(IP,5)=-0.98995*FACT(1)+0.14142*E(IP,5)                        FACTO
 8210 CONTINUE                                                          FACTO
C     WRITE STATEMENTS FOR THE RANDOM VARIABLES                         FACTO
      GO TO 8213                                                        FACTO
C     GENERATE SCORES FROM POSITIVE HALF OF NORMAL CURVE                FACTO
 8214 SX=1.0                                                            FACTO
      DO 8211 IP=1,N                                                    FACTO
      DO 251 ICOUNT=1,NS2                                               FACTO
  249 CALL JOYCE(IX,SX,AM,VX)                                           FACTO
      FACT(ICOUNT)=ABS(VX)                                              FACTO
      FACT(ICOUNT) = (1.6583)*(FACT(ICOUNT)) - 1.3234                   FACTO
  251 CONTINUE                                                          FACTO
      X(IP,1)=0.98489*FACT(1) +0.17321*E(IP,1)                          FACTO
      X(IP,2)=0.98995*FACT(2) + 0.14142*E(IP,2)                         FACTO
```

```
      X(IP,3)=0.99499*FACT(3) + 0.10000*E(IP,3)                       FACTO
      X(IP,4)=0.57446*FACT(1)+0.50000*FACT(2)+0.64031*FACT(3)+0.10000 FACTO
     1*E(IP,4)                                                        FACTO
      X(IP,5)=0.80000*FACT(1)+0.43589*FACT(2)+0.40000*FACT(3)+0.10000 FACTO
     1*E(IP,5)                                                        FACTO
      X(IP,6)=0.50000*FACT(1)+0.60828*FACT(2)+0.60000*FACT(3)+0.14142 FACTO
     1*E(IP,6)                                                        FACTO
      X(IP,7)=0.70000*FACT(1)+0.60000*FACT(2)+0.37417*FACT(3)+0.10000 FACTO
     1*E(IP,7)                                                        FACTO
      X(IP,8)=0.42426*FACT(1)+0.90000*FACT(2)+0.10000*E(IP,8)         FACTO
      X(IP,9)=0.59161*FACT(1)+0.80000*FACT(3)+0.10000*E(IP,9)         FACTO
      XX(IP,1)=0.98489*FACT(1)+0.17321*E(IP,1)                        FACTO
      XX(IP,2)=0.98995*FACT(2)+0.14142*E(IP,2)                        FACTO
      XX(IP,3)=0.73485*FACT(1)+0.67082*FACT(2)+0.10000*E(IP,4)        FACTO
      XX(IP,4)=0.65574*FACT(1)+0.74162*FACT(2)+0.14142*E(IP,6)        FACTO
      XX(IP,5)=0.59161*FACT(1)-0.80000*FACT(2)-0.10000*E(IP,7)        FACTO
      XX(IP,6)=0.42426*FACT(1)+0.90000*FACT(2)+0.10000*E(IP,8)        FACTO
      XX(IP,7)=0.74833*FACT(1)-0.65574*FACT(2)+0.10000*E(IP,9)        FACTO
      XXX(IP,1)=0.99499*FACT(1)+0.10000*E(IP,1)                       FACTO
      XXX(IP,2)=-0.99499*FACT(1)+0.10000*E(IP,2)                      FACTO
      XXX(IP,3)=0.98489*FACT(1)+0.17321*E(IP,3)                       FACTO
      XXX(IP,4)=0.98995*FACT(1)+0.14142*E(IP,4)                       FACTO
      XXX(IP,5)=-0.98995*FACT(1)+0.14142*E(IP,5)                      FACTO
 8211 CONTINUE                                                        FACTO
C     WRITE STATEMENTS FOR THE RANDOM VARIABLES                       FACTO
      GO TO 8213                                                      FACTO
C     COMPUTE CHI SQUARE DISTRIBUTION                                 FACTO
 8216 SX = 1.0                                                        FACTO
      DO 8230 IP = 1,N                                                FACTO
      KINDI=3                                                         FACTO
      LKIND=3                                                         FACTO
      MLK = 5                                                         FACTO
      DO 261 I = 1,MLK                                                FACTO
      FACT(I) = 0.0                                                   FACTO
  261 CONTINUF                                                        FACTO
```

```
      DO 259 LSP = 1, KINDI                                                FACTO
      DO 260 NSP = 1, LKIND                                                FACTO
      CALL JOYCE(IX,SX,AM,VX)                                              FACTO
      TEMP(NSP) = VX                                                       FACTO
      FACT(LSP)=FACT(LSP) + TEMP(NSP)**2                                   FACTO
  260 CONTINUE                                                             FACTO
      FACT(LSP) = 0.4082 * FACT(LSP) - 1.2246                              FACTO
  259 CONTINUE                                                             FACTO
C     COMPUTE VARIABLES                                                    FACTO
      X(IP,1)=0.98489*FACT(1) +0.17321*E(IP,1)                             FACTO
      X(IP,2)=0.98995*FACT(2) + 0.14142*E(IP,2)                            FACTO
      X(IP,3)=0.99499*FACT(3) + 0.10000*E(IP,3)                            FACTO
      X(IP,4)=0.57446*FACT(1)+0.50000*FACT(2)+0.64031*FACT(3)+0.10000      FACTO
     1*E(IP,4)                                                             FACTO
      X(IP,5)=0.80000*FACT(1)+0.43589*FACT(2)+0.40000*FACT(3)+0.10000      FACTO
     1*E(IP,5)                                                             FACTO
      X(IP,6)=0.50000*FACT(1)+0.60828*FACT(2)+0.60000*FACT(3)+0.14142      FACTO
     1*E(IP,6)                                                             FACTO
      X(IP,7)=0.70000*FACT(1)+0.60000*FACT(2)+0.37417*FACT(3)+0.10000      FACTO
     1*E(IP,7)                                                             FACTO
      X(IP,8)=0.42426*FACT(1)+0.90000*FACT(2)+0.10000*E(IP,8)              FACTO
      X(IP,9)=0.59161*FACT(1)+0.80000*FACT(3)+0.10000*E(IP,9)              FACTO
      XX(IP,1)=0.98489*FACT(1)+0.17321*E(IP,1)                             FACTO
      XX(IP,2)=0.98995*FACT(2)+0.14142*E(IP,2)                             FACTO
      XX(IP,3)=0.73485*FACT(1)+0.67082*FACT(2)+0.10000*E(IP,4)             FACTO
      XX(IP,4)=0.65574*FACT(1)+0.74162*FACT(2)+0.14142*E(IP,6)             FACTO
      XX(IP,5)=0.59161*FACT(1)-0.80000*FACT(2)-0.10000*E(IP,7)             FACTO
      XX(IP,6)=0.42426*FACT(1)+0.90000*FACT(2)+0.10000*E(IP,8)             FACTO
      XX(IP,7)=0.74833*FACT(1)-0.65574*FACT(2)+0.10000*E(IP,9)             FACTO
      XXX(IP,1)=0.99499*FACT(1)+0.10000*E(IP,1)                            FACTO
      XXX(IP,2)=-0.99499*FACT(1)+0.10000*E(IP,2)                           FACTO
      XXX(IP,3)=0.98489*FACT(1)+0.17321*E(IP,3)                            FACTO
      XXX(IP,4)=0.98995*FACT(1)+0.14142*E(IP,4)                            FACTO
      XXX(IP,5)=-0.98995*FACT(1)+0.14142*E(IP,5)                           FACTO
 8230 CONTINUE                                                             FACTO
```

```
C     WRITE STATEMENTS FOR THE RANDOM VARIABLES                         FACTO
      GO TO 8213                                                        FACTO
C     COMPUTE T- DISTRIBUTION                                           FACTO
 8218 SX = 1.0                                                          FACTO
      DO 8240 IP = 1,N                                                  FACTO
      LBJ=3                                                             FACTO
      JFK=5                                                             FACTO
      NPJ = 5                                                           FACTO
      DO 262 I = 1,NPJ                                                  FACTO
      SFACT(I) = 0.0                                                    FACTO
      FACT(I) = 0.0                                                     FACTO
  262 CONTINUE                                                          FACTO
      DO 269 LSP = 1,LBJ                                                FACTO
      DO 270 KAP = 1,JFK                                                FACTO
      CALL JOYCE(IX,SX,AM,VX)                                           FACTO
      TEMP(KAP)=VX                                                      FACTO
      SFACT(LSP)=SFACT(LSP)+TEMP(KAP)**2                                FACTO
  270 CONTINUE                                                          FACTO
      CALL JOYCE(IX,SX,AM,VX)                                           FACTO
      FACT(LSP)=(VX) / (SQRT(SFACT(LSP)))                               FACTO
      FACT(LSP)= 1.7321*FACT(LSP)                                       FACTO
  269 CONTINUE                                                          FACTO
C     COMPUTE VARIABLES                                                 FACTO
      X(IP,1)=0.98489*FACT(1) +0.17321*E(IP,1)                          FACTO
      X(IP,2)=0.98995*FACT(2) + 0.14142*E(IP,2)                         FACTO
      X(IP,3)=0.99499*FACT(3) + 0.10000*E(IP,3)                         FACTO
      X(IP,4)=0.57446*FACT(1)+0.50000*FACT(2)+0.64031*FACT(3)+0.10000   FACTO
     1*E(IP,4)                                                          FACTO
      X(IP,5)=0.80000*FACT(1)+0.43589*FACT(2)+0.40000*FACT(3)+0.10000   FACTO
     1*E(IP,5)                                                          FACTO
      X(IP,6)=0.50000*FACT(1)+0.60828*FACT(2)+0.60000*FACT(3)+0.14142   FACTO
     1*E(IP,6)                                                          FACTO
      X(IP,7)=0.70000*FACT(1)+0.60000*FACT(2)+0.37417*FACT(3)+0.10000   FACTO
     1*E(IP,7)                                                          FACTO
      X(IP,8)=0.42426*FACT(1)+0.90000*FACT(2)+0.10000*E(IP,8)           FACTO
```

```
      X(IP,9)=0.59161*FACT(1)+0.80000*FACT(3)+0.10000*E(IP,9)           FACTO
      X(IP,9)=0.59161*FACT(1)+0.80000*FACT(3)+0.10000*E(IP,9)           FACTO
      XX(IP,1)=0.98489*FACT(1)+0.17321*E(IP,1)                          FACTO
      XX(IP,2)=0.98995*FACT(2)+0.14142*E(IP,2)                          FACTO
      XX(IP,3)=0.73485*FACT(1)+0.67082*FACT(2)+0.10000*E(IP,4)          FACTO
      XX(IP,4)=0.65574*FACT(1)+0.74162*FACT(2)+0.14142*E(IP,6)          FACTO
      XX(IP,5)=0.59161*FACT(1)-0.80000*FACT(2)-0.10000*E(IP,7)          FACTO
      XX(IP,6)=0.42426*FACT(1)+0.90000*FACT(2)+0.10000*E(IP,8)          FACTO
      XX(IP,7)=0.74833*FACT(1)-0.65574*FACT(2)+0.10000*E(IP,9)          FACTO
      XXX(IP,1)=0.99499*FACT(1)+0.10000*E(IP,1)                         FACTO
      XXX(IP,2)=-0.99499*FACT(1)+0.10000*E(IP,2)                        FACTO
      XXX(IP,3)=0.98489*FACT(1)+0.17321*E(IP,3)                         FACTO
      XXX(IP,4)=0.98995*FACT(1)+0.14142*E(IP,4)                         FACTO
      XXX(IP,5)=-0.98995*FACT(1)+0.14142*E(IP,5)                        FACTO
 8240 CONTINUE                                                          FACTO
C     WRITE STATEMENTS FOR THE RANDOM VARIABLES                         FACTO
 8213 NEXP = 3                                                          FACTO
      DO 9000 NUM=1,NEXP                                                FACTO
      IF(NUM - 1) 9001,9002,9001                                        FACTO
 9001 IF(NUM - 2) 9003,9004,9003                                        FACTO
 9003 IF(NUM - 3) 109,9006,109                                          FACTO
 9002 READ(5,71) PR,PR1,N,M,CON                                         FACTO
      IO=1                                                              FACTO
      NS2=1                                                             FACTO
      NSM2 = 2                                                          FACTO
      NDF(1) = 6                                                        FACTO
      NDF(2) = 3                                                        FACTO
      CALL CORRE(N,M,IO,XXX,XBAR,S,V,R,D,BB,T)                          FACTO
      GO TO 34                                                          FACTO
 3001 NDFT=10                                                           FACTO
      IF(CHIO - 18.307) 1054,1054,1055                                  FACTO
 1054 WRITE(6,1075) NDFT, CHIO                                          FACTO
      GO TO 1056                                                        FACTO
 1055 WRITE(6,1076) NDFT,CHIO                                           FACTO
 1056 IF(CHIO - 23.209)  1057,1057,1058                                 FACTO
```

```
1057 WRITE(6,1077) NDFT,CHIO                                          FACTO
     GO TO 1059                                                       FACTO
1058 WRITE(6,1078) NDFT,CHIO                                          FACTO
1059 GO TO 8888                                                       FACTO
9004 READ(5,71) PR,PR1,N,M,CON                                        FACTO
     IO=1                                                             FACTO
     NS2=2                                                            FACTO
     NSM2=3                                                           FACTO
     NDF(1) = 15                                                      FACTO
     NDF(2) =10                                                       FACTO
     NDF(3) =6                                                        FACTO
     CALL CORRE(N,M,IO,XX,XBAR,S,V,R,D,BB,T)                          FACTO
     GO TO 34                                                         FACTO
4001 NDFT=21                                                          FACTO
     IF(CHIO - 32.671) 1061,1061,1062                                 FACTO
1061 WRITE(6,1075) NDFT,CHIO                                          FACTO
     GO TO 1063                                                       FACTO
1062 WRITE(6,1076) NDFT,CHIO                                          FACTO
1063 IF(CHIO - 38.932) 1071,1071,1072                                 FACTO
1071 WRITE(6,1077) NDFT,CHIO                                          FACTO
     GO TO 1973                                                       FACTO
1072 WRITE(6,1078) NDFT,CHIO                                          FACTO
1973 GO TO 8888                                                       FACTO
9006 READ(5,71) PR,PR1,N,M,CON                                        FACTO
     IO=1                                                             FACTO
     NS2=3                                                            FACTO
     NSM2=4                                                           FACTO
     NDF(1) =28                                                       FACTO
     NDF(2) = 21                                                      FACTO
     NDF(3) = 15                                                      FACTO
     NDF(4) = 10                                                      FACTO
     CALL CORRE(N,M,IO,X,XBAR,S,V,R,D,BB,T)                           FACTO
     GO TO 34                                                         FACTO
 701 NDFT=36                                                          FACTO
     IF(CHIO - 50.998) 8501,8501,8502                                 FACTO
```

```
 8501 WRITE(6,1075) NDFT,CHIO                                           FACTO
      GO TO 8505                                                        FACTO
 8502 WRITE(6,1076) NDFT,CHIO                                           FACTO
 8505 IF(CHIO - 58.619) 8503,8503,8504                                  FACTO
 8503 WRITE(6,1077) NDFT,CHIO                                           FACTO
      GO TO 8506                                                        FACTO
 8504 WRITE(6,1078) NDFT,CHIO                                           FACTO
 8506 GO TO 8888                                                        FACTO
C     BEGIN FACTO                                                       FACTO
C     BEGIN FACTO                                                       FACTO
   71 FORMAT(A4,A2,I5,I2,F6.0)                                          FACTO
C     PRINT MEANS                                                       FACTO
   34 WRITE(6,35) (XBAR(J),J=1,M)                                       FACTO
   35 FORMAT(//6H MEANS/(3E20.8))                                       FACTO
C     PRINT STANDARD DEVIATIONS                                         FACTO
      WRITE(6,43)  (S(J),J=1,M)                                         FACTO
   43 FORMAT(//20H STANDARD DEVIATIONS/(3E20.8  ))                      FACTO
      WRITE(6,45)                                                       FACTO
   45 FORMAT   (//25H CORRELATION COEFFICIENTS)                         FACTO
      DO 120 I=1,M                                                      FACTO
      DO 110 J=1,M                                                      FACTO
  203 IF (I-J) 102,104,104                                              FACTO
  102 L=I+(J*J-J)/2                                                     FACTO
      GO TO 110                                                         FACTO
  104 L=J+(I*I-I)/2                                                     FACTO
  110 D(J)=R(L)                                                         FACTO
  992 FORMAT(I2,9F8.5)                                                  FACTO
  120 WRITE(6,50) I,(D(J),J=1,M)                                        FACTO
   50 FORMAT(//4H ROW,I3/(11F12.5))                                     FACTO
      MV=0                                                              FACTO
      CALL EIGEN(R,V,M,MV)                                              FACTO
C     BEGIN TESTS OF NULL HYPOTHESES                                    FACTO
C     BEGIN TESTS OF NULL HYPOTHESES                                    FACTO
      DO 1029 I=1,M                                                     FACTO
      L = I + (I*I - I)/2                                               FACTO
```

```
      TAU(I) = R(L)                                                     FACTO
 1029 CONTINUE                                                          FACTO
      DETR=1.0                                                          FACTO
      DO 1091 I=1,M                                                     FACTO
      DETR=TAU(I)*DETR                                                  FACTO
 1091 CONTINUE                                                          FACTO
      WRITE(6,990)DETR,(TAU(I),I=1,M)                                   FACTO
  990 FORMAT(4E20.10/4E20.10/4E20.10)                                   FACTO
      TSUM=0.0                                                          FACTO
      DO 1092 I=1,M                                                     FACTO
      TSUM=TSUM + TAU(I)                                                FACTO
 1092 CONTINUE                                                          FACTO
      DO 1094 I=1,NSM2                                                  FACTO
      BETA(I) = 0.0                                                     FACTO
      ALPHA(I) = 1.0                                                    FACTO
 1094 CONTINUE                                                          FACTO
      DO 1093 K=1,NSM2                                                  FACTO
C     NSM2 = NS2 + 1                                                    FACTO
      DO 1095 J = 1,K                                                   FACTO
      BETA(K) = BETA(K) + TAU(J)                                        FACTO
      ALPHA(K) = ALPHA(K) * TAU(J)                                      FACTO
 1095 CONTINUE                                                          FACTO
      GAMMA = DETR / ALPHA(K)                                           FACTO
      DELTA = M - BETA(K)                                               FACTO
      NEXPN = M - K                                                     FACTO
      RPK(K) = GAMMA*(NEXPN/DELTA)**NEXPN                               FACTO
 1093 CONTINUE                                                          FACTO
C  RPK(I) = R( P - I)                                                   FACTO
      CHIO = -((N-1.0)-(0.1667)*(2.0*M + 5.0))*ALOG(DETR)               FACTO
      NDFT = 0.5*M*(M-1)                                                FACTO
      WRITE(7,971) CHIO,NDFT,NIX,NUM,ITER                               FACTO
      DO 1098 I=1,NSM2                                                  FACTO
      COEF(I) =(  (N-1.0) - (0.1667)*(2.0*M + 5.0)-(0.6667)*(I))        FACTO
      CHI(I) = -COEF(I)*ALOG(RPK(I))                                    FACTO
      NDF(I) = (0.5)*(M-I)*(M-(I+1.0))                                  FACTO
```

```
      NNDF(I) = (0.5)*(M - (I+1.0))*(M-I+2.0)                           FACTO
      WRITE(7,971) CHI(I),NNDF(I),NIX,NUM,ITER                          FACTO
C     IF THE FOLLOWING CARD IS REMOVED RPK(I)'S ARE PRINTED AS OUTPUT   FACTO
      RPK(I) = ALOG(RPK(I))                                             FACTO
 1098 CONTINUE                                                          FACTO
  983 FORMAT(24H CHI(I),NNDF(I),I=1,NSM2)                               FACTO
  991 FORMAT(E20.10,I3)                                                 FACTO
  982 FORMAT(16H RPK(I),I=1,NSM2)                                       FACTO
  981 FORMAT(17H COEF(I),I=1,NSM2)                                      FACTO
      CON=1.0                                                           FACTO
      CALL TRACE(M,R,CON,K,D)                                           FACTO
C     PRINT EIGEN VALUES                                                FACTO
      DO 130 I=1,K                                                      FACTO
      L=I+(I*I-I)/2                                                     FACTO
  130 S(I)=R(L)                                                         FACTO
      WRITE (6,91) ( S(J),J=1,K)                                        FACTO
   91 FORMAT(///12H EIGENVALUES/(10E14.9))                              FACTO
C     PRINT CUMULATIVE PERCENTAGE OF EIGEN VALUES                       FACTO
      WRITE(6,92)  (  D(J),J=1,K)                                       FACTO
   92 FORMAT(//37H CUMULATIVE PERCENTSGE OF EIGENVALUES/(10F12.5))      FACTO
C     PRINT EIGENVECTORS                                                FACTO
      WRITE(6,93)                                                       FACTO
   93 FORMAT(13H EIGENVECTORS)                                          FACTO
      L=0                                                               FACTO
      DO 150 J=1,K                                                      FACTO
      DO 140 I=1,M                                                      FACTO
      L=L+1                                                             FACTO
  140 D(I)=V(L)                                                         FACTO
  150 WRITE(6,94) J, (D(I),I=1,M)                                       FACTO
   94 FORMAT(//7H VECTOR,I3/(10F12.5))                                  FACTO
      CALL LOAD(M,K,R,V)                                                FACTO
C     PRINT FACTOR MATRIX                                               FACTO
      WRITE(6,95)K                                                      FACTO
   95 FORMAT(///16H FACTOR MATRIX (,I3,9H FACTORS))                     FACTO
      DO 180 I=1,M                                                      FACTO
```

```
      DO 170 J=1,K                                                      FACTO
      L=M*(J-1)+I                                                       FACTO
  170 D(J)=V(L)                                                         FACTO
  180 WRITE(6,96)I,(D(J),J=1,K)                                         FACTO
   96 FORMAT(//9H VARIABLE,I3/(10F12.5))                                FACTO
      IF(K-1) 185,185,188                                               FACTO
  185 WRITE(6,97)K                                                      FACTO
   97 FORMAT(//5H ONLY,I2,30H FACTORS RETAINED. NO ROTATION )           FACTO
  188 CALL VARMX(M,K,V,NC,TV,BB,T,D)                                    FACTO
C     PRINT VARIANCES                                                   FACTO
      NV=NC+1                                                           FACTO
      WRITE(6,98)                                                       FACTO
   98 FORMAT(///10H ITERATION,7X, 9HVARIANCES/8H   CYCLE   )            FACTO
      DO 190 I=1,NV                                                     FACTO
      NC=I-1                                                            FACTO
  190 WRITE(6,99)NC,TV(I)                                               FACTO
   99 FORMAT(I6,F20.6)                                                  FACTO
C     PRINT ROTATED FACTOR MATRIX                                       FACTO
      WRITE(6,81)K                                                      FACTO
   81 FORMAT(///24H ROTATED FACTOR MATRIX (,I3,9H FACTORS))             FACTO
      DO 220 I=1,M                                                      FACTO
      DO 210 J=1,K                                                      FACTO
      L=M*(J-1)+I                                                       FACTO
  210 S(J)=V(L)                                                         FACTO
  220 WRITE(6,82)I,(S(J),J=1,K)                                         FACTO
   82 FORMAT(//9H VARIABLE,I3,(10F12.5))                                FACTO
C     PRINT COMMUNALITIES                                               FACTO
      WRITE(6,83)                                                       FACTO
   83 FORMAT(///23H CHECK ON COMMUNALITIES//9H VARIABLE,7X,8HORIGINAL,  FACTO
     112X,5HFINAL,10X,10HDIFFERENCE)                                    FACTO
      DO 230 I=1,M                                                      FACTO
  230 WRITE(6,88)I,BB(I),T(I),D(I)                                      FACTO
   88 FORMAT(I6,3F18.5)                                                 FACTO
 1501 IF(NSM2 - 2) 1502,2000,1502                                       FACTO
 1502 IF(NSM2 - 3) 1503,3000,1503                                       FACTO
```

```
1503 GO TO 4000                                         FACTO
2000 DO 1050 I=1,NSM2                                   FACTO
     INT = NDF(I)                                       FACTO
1049 GO TO 1019                                         FACTO
1040 CONTINUE                                           FACTO
1050 CONTINUE                                           FACTO
1060 GO TO 3001                                         FACTO
1019 IF(I - 1 ) 1020,2081,1020                          FACTO
1020 IF(I-2) 109,2091,109                               FACTO
2081 IF(INT - 6) 109,1082,109                           FACTO
1082 IF(CHI(I) - 12.592) 1083,1083,1084                 FACTO
1083 WRITE(6,1075) NDF(I),CHI(I)                        FACTO
     GO TO 1081                                         FACTO
1084 WRITE(6,1076) NDF(I),CHI(I)                        FACTO
1081 IF(CHI(I) - 16.812) 1073,1073,1074                 FACTO
1073 WRITE(5,1077) NDF(I),CHI(I)                        FACTO
     GO TO 2271                                         FACTO
1074 WRITF(6,1078) NDF(I), CHI(I)                       FACTO
2271 GO TO 1040                                         FACTO
2091 IF(INT-3) 109,8892,109                             FACTO
8892 IF(CHI(I)-7.815) 8893,8893,8894                    FACTO
8893 WRITE(6,1075)  NDF(I),CHI(I)                       FACTO
     GO TO 8891                                         FACTO
8894 WRITE(6,1076)    NDF(I),CHI(I)                     FACTO
8891 IF(CHI(I) - 11.345) 1193,1193,1194                 FACTO
1193 WRITE(6,1077)  NDF(I),CHI(I)                       FACTO
     GO TO 1195                                         FACTO
1194 WRITE(6,1078) NDF(I),CHI(I)                        FACTO
1195 GO TO 1040                                         FACTO
3000 DO 2050 I=1,NSM2                                   FACTO
     INT=NDF(I)                                         FACTO
2049 GO TO 2019                                         FACTO
2040 CONTINUE                                           FACTO
2050 CONTINUE                                           FACTO
2051 GO TO 4001                                         FACTO
```

```
2019 IF(I-1) 2020,3081,2020                                             FACTO
2020 IF (I-2) 2021,3091,2021                                            FACTO
2021 IF( I - 3 ) 109,3555,109                                           FACTO
3081 IF(INT-15) 109,2082,109                                            FACTO
2082 IF(CHI(I) - 24.996) 2083,2083,2084                                 FACTO
2083 WRITE(6,1075) NDF(I),CHI(I)                                        FACTO
     GO TO 5991                                                         FACTO
2084 WRITE(6,1076) NDF(I),CHI(I)                                        FACTO
5991 IF(CHI(I) - 30.578) 2773   ,2773,2774                              FACTO
2773 WRITE(6,1077) NDF(I),CHI(I)                                        FACTO
     GO TO 2271                                                         FACTO
2774 WRITE(6,1078) NDF(I),CHI(I)                                        FACTO
2771 GO TO 2040                                                         FACTO
3091 IF(INT-10   ) 109,2092,109                                         FACTO
2092 IF(CHI(I) - 18.307) 2093,2093,2094                                 FACTO
2093 WRITE(6,1075) NDF(I),CHI(I)                                        FACTO
     GO TO 2781                                                         FACTO
2094 WRITE(6,1076) NDF(I),CHI(I)                                        FACTO
2781 IF(CHI(I)-23.209) 2073,2073,2074                                   FACTO
2073 WRITE(6,1077) NDF(I),CHI(I)                                        FACTO
     GO TO 2071                                                         FACTO
2074 WRITE(6,1078) NDF(I),CHI(I)                                        FACTO
2071 GO TO 2040                                                         FACTO
3555 IF(INT - 6) 109,3002,109                                           FACTO
3002 IF(CHI(I)-12.592) 3003,3003,3004                                   FACTO
3003 WRITE(6,1075) NDF(I),CHI(I)                                        FACTO
     GO TO 3454                                                         FACTO
3004 WRITE(6,1076) NDF(I),CHI(I)                                        FACTO
3454 IF(CHI(I)-16.812) 2473,2473,2475                                   FACTO
2473 WRITE(6,1077) NDF(I),CHI(I)                                        FACTO
     GO TO 2471                                                         FACTO
2475 WRITE(6,1078) NDF(I),CHI(I)                                        FACTO
2471 GO TO 2040                                                         FACTO
4000 DO 3050 I=1,NSM2                                                   FACTO
     INT=NDF(I)                                                         FACTO
```

```
3049 GO TO 3019                                                         FACTO
3040 CONTINUE                                                           FACTO
3050 CONTINUE                                                           FACTO
3051 GO TO 701                                                          FACTO
3019 IF(I-1) 3020,4081,3020                                             FACTO
3020 IF(I-2) 3021,4091,3021                                             FACTO
3021 IF(I-3) 3022,4111,3022                                             FACTO
3022 IF(I-4) 109,4011,109                                               FACTO
4081 IF(INT-28) 109,3082,109                                            FACTO
3082 IF(CHI(I) - 41.337) 3083,3083,3084                                 FACTO
3083 WRITE(6,1075) NDF(I),CHI(I)                                        FACTO
     GO TO 3333                                                         FACTO
3084 WRITE(6,1076) NDF(I),CHI(I)                                        FACTO
3333 IF(CHI(I)-48.278) 3073,3073,3074                                   FACTO
3073 WRITE(6,1077) NDF(I),CHI(I)                                        FACTO
     GO TO 3071                                                         FACTO
3074 WRITE(6,1078) NDF(I),CHI(I)                                        FACTO
3071 GO TO 3040                                                         FACTO
4091 IF(INT-21) 109,3092,109                                            FACTO
3092 IF(CHI(I) - 32.671) 3093,3093,3094                                 FACTO
3093 WRITE(6,1075) NDF(I),CHI(I)                                        FACTO
     GO TO 5555                                                         FACTO
3094 WRITE(6,1076) NDF(I),CHI(I)                                        FACTO
5555 IF(CHI(I)-38.932) 6073,6073,6074                                   FACTO
6073 WRITE(6,1077) NDF(I),CHI(I)                                        FACTO
     GO TO 5071                                                         FACTO
6074 WRITE(6,1078) NDF(I),CHI(I)                                        FACTO
5071 GO TO 3040                                                         FACTO
4111 IF(INT-15) 109,4002,109                                            FACTO
4002 IF(CHI(I) - 24.996) 4003,4003,4004                                 FACTO
4003 WRITE(6,1075) NDF(I),CHI(I)                                        FACTO
     GO TO 5081                                                         FACTO
4004 WRITE(6,1076) NDF(I),CHI(I)                                        FACTO
5081 IF(CHI(I) - 30.578) 5083,5083,5084                                 FACTO
5083 WRITE(6,1077) NDF(I),CHI(I)                                        FACTO
```

```
      GO TO 5091                                                        FACTO
 5084 WRITE(6,1078) NDF(I),CHI(I)                                       FACTO
 5091 GO TO 3040                                                        FACTO
 4011 IF(INT-10) 109,6002,109                                           FACTO
 6002 IF(CHI(I)-18.307) 6083,6083,6084                                  FACTO
 6083 WRITE(6,1075) NDF(I),CHI(I)                                       FACTO
      GO TO 7072                                                        FACTO
 6084 WRITE(6,1076) NDF(I),CHI(I)                                       FACTO
 7072 IF(CHI(I) - 23.209) 7073,7073,7074                                FACTO
 7073 WRITE(6,1077) NDF(I),CHI(I)                                       FACTO
      GO TO 8081                                                        FACTO
 7074 WRITE(6,1078) NDF(I),CHI(I)                                       FACTO
 8081 GO TO 3040                                                        FACTO
 8888 CONTINUE                                                          FACTO
 9000 CONTINUE                                                          FACTO
 9031 CONTINUE                                                          FACTO
      WRITE(6,10001) IX                                                 FACTO
10001 FORMAT(I10)                                                       FACTO
  109 CONTINUE                                                          FACTO
      END                                                               FACTO
/*
//GO.SYSIN DD *
    3   92   64   82   40   95   20   28   62   43
   25    8   23   41   85    7   81   54    6   39
   59   79   70   18   71   55   66   15   72   75
   96   34   24   93   51   63   77   94   45   98
   27   31   32   91   21   73   68   50   26   90
   42   61   60   14   17   12   99   69   11   97
   19   84   44   56   49   80   58   83   88   76
   53   52    5   47   16   29   57    2   10   67
   78   33   48   86   37   35   89   38   87   65
   13   46   36    1  100   74   30    4   22    9
   42   30   28   59    4   76   44   87   58   13
   97   27   23   47   98   72   33   94   21    1
   45   55    3   35   39   38   66   84   70   69
```

81	24	34	46	77	99	86	96	7	91
2	50	5	17	10	92	57	95	52	32
49	85	40	14	79	29	19	65	60	83
89	88	37	48	90	12	8	6	62	20
45	25	11	22	56	80	9	26	93	36
78	51	82	15	61	63	54	100	68	31
64	53	67	41	74	16	18	73	43	71
39	10	34	46	84	5	47	85	70	4
76	11	74	66	19	100	13	48	62	72
57	44	94	90	77	38	80	60	36	54
31	2	21	53	92	17	15	67	25	83
71	41	56	42	26	37	27	86	18	14
65	75	28	61	59	8	43	89	73	55
96	68	82	79	50	51	87	29	64	95
58	69	91	35	7	45	9	22	20	30
32	81	33	52	99	88	24	63	40	6
1	23	98	97	3	93	78	12	49	16
94	12	99	73	79	30	25	4	6	55
56	64	74	78	95	13	28	62	1	24
70	57	47	7	9	26	31	82	10	45
41	15	71	27	40	39	85	34	49	61
65	87	67	72	88	29	18	54	8	63
60	48	84	58	66	96	97	22	77	52
36	2	14	23	68	91	69	32	50	33
3	92	81	86	5	83	75	93	46	90
20	21	17	11	19	80	44	51	100	89
76	59	35	42	98	37	38	53	16	43
63	78	20	16	97	7	71	46	13	57
54	5	11	85	45	33	44	37	61	10
53	42	32	100	88	69	55	25	64	93
18	52	9	95	24	90	98	68	31	8
66	60	28	15	29	75	40	38	47	79
43	14	92	81	2	41	19	89	59	83
30	12	67	51	87	27	49	48	74	91
72	56	21	6	99	62	65	77	70	58

35	76	1	50	36	34	73	17	86	22
82	94	26	80	4	84	23	96	3	39
21	41	74	99	67	89	47	62	98	52
19	55	95	2	7	29	90	17	59	56
53	24	87	12	97	68	42	94	48	28
60	3	92	66	88	23	38	78	84	35
26	81	37	82	77	25	91	54	44	61
13	27	40	49	20	36	75	80	33	58
72	5	83	73	79	50	39	30	85	65
69	14	1	8	10	15	9	76	6	57
51	45	70	93	32	100	18	63	86	31
71	64	11	96	16	4	43	34	22	46
14	95	60	73	2	22	77	45	42	21
25	87	10	68	47	40	13	56	51	36
26	18	86	3	17	65	67	20	91	44
33	50	9	1	7	57	52	96	30	35
4	76	90	62	37	28	99	70	55	34
11	46	84	83	89	6	79	8	48	71
38	66	32	97	53	88	100	98	43	69
92	74	19	82	78	31	23	93	41	72
27	54	94	5	63	49	61	59	85	15
24	80	81	12	29	64	75	39	16	58
36	63	78	18	26	68	24	1	96	4
98	25	91	82	27	43	37	42	28	19
95	55	87	66	61	65	67	88	16	47
14	41	23	89	62	92	94	21	79	44
39	3	93	72	31	56	69	76	48	90
7	13	50	60	74	12	70	35	80	49
46	54	22	5	77	32	81	73	34	33
52	51	75	57	29	99	45	100	20	71
38	10	17	58	9	8	59	6	92	53
15	40	86	85	83	30	64	11	2	84
5	9	34	74	53	59	29	38	65	61
93	20	88	19	78	33	87	14	60	41
40	24	52	43	81	92	58	15	99	22

```
   35   76    1   50   36   34   73   17   86   22
   54   44   84   90   47   73   50   28   97   49
   10   25   98   95   23   77   89   67   66   35
   68    4   39   72   57   31   76   70   64   51
   86   42   94   56   27    8   69   26   82   62
   85  100    1   30    6   36   21   11    2   16
   12   91   96   71   37   45   18   79    3   48
    7   46   80   17   75   55   13   83   63   32
FACTO 00100050001.0
FACTO 00100070001.0
FACTO 00100090001.0
FACTO 00100050001.0
FACTO 00100070001.0
FACTO 00100090001.0
FACTO 00100050001.0
FACTO 00100070001.0
FACTO 00100090001.0
FACTO 00100050001.0
FACTO 00100070001.0
FACTO 00100090001.0
/*
```