

A. K I E L O C H and W. O K T A B A (Lublin)

FACTOR ANALYSIS BY THE METHOD OF MAXIMUM LIKELIHOOD

1. Introduction. Many naturalists deal in their investigations with multivariate samples that are obtained by observing some characteristics of individuals. For example, the plant breeder records some characteristics of vetch plants, the zootechnician — some characteristics of animals, the apiarist — some characteristics of bees. The agronomist may be interested in the contribution of individual cereal crops to the total agricultural production of the administrative districts of a given province.

In many cases factor analysis solves problems typical to some empirical sciences. It is based on an approximate representation of n correlated characteristics of N individuals as a linear combination of a smaller number m ($m \leq n < N$) of uncorrelated characteristics. The corresponding interpretation is ascribed to new characteristics that are called *common factors*. The investigation consists in choosing the correct mathematical model, in determining its adequacy and, in particular, in finding out the number of common factors and in applying a method as simple as possible to estimate the coefficients (loadings) of common factors. The factors are ordered according to their decreasing contribution to the total variance (cf. (2.13) and further on) or according to other criteria (cf. (4.6)).

Having obtained the relations between the given and the new uncorrelated characteristics, one can investigate the taxonomic problem of characteristic grouping by means of factor analysis. Grouping is reached in two steps: 1. calculation of the matrix of estimators of common factor loadings, and 2. inclusion of a pair of characteristics (this concerns all pairs of characteristics) to the same group in case of approximately equal values of loading estimators of all common factors.

The publications by Anderson and Rubin [2], Gower [7], Holzinger [9], Thurstone [18], Fruchter [6], Mullen [14], and by the others are concerned with factor analysis.

In this paper simple relations between three mathematical models of factor analysis are presented. To each of these models one of the following three methods is applied: 1. \mathcal{H} , i.e. Hotelling's principal components

method [10]; 2. \mathcal{T} , i.e. Thomson's method [17]; 3. \mathcal{L} , i.e. Lawley's method of maximum likelihood. It was noted that the a priori information about the structure of the population with respect to the contributions of unique factors (cf. Section 2) decides on the choice of the \mathcal{T} - or \mathcal{L} -method. Thus, under these different contributions to the populations one should use the restrictions (4.6) which in case of equal contributions are given by (4.2). The method \mathcal{L} of estimating loadings can also be applied to estimating parameters by the \mathcal{T} -method.

The relation between these three methods and the methods of determining eigenvectors of corresponding matrices is shown.

A numerical example of plant breeding is used to illustrate the application of the three methods. Programs for the electronic computer "Odra 1013" based on iterative methods are worked out in order to determine the matrices of estimators of loadings of common factors by the methods \mathcal{H} , \mathcal{T} and \mathcal{L} . Moreover, the matrices of estimators of loadings of unique factors in the three methods are found. In the \mathcal{L} -method, the number of common factors is determined by the test of significance.

Five taxonomic groups are formed out of ten characteristics of vetch by using the matrix of estimators of loadings.

2. Linear model, assumptions and notation. Let

$$(2.1) \quad A = A = [a_{ij}] = [a_1 a_2 \dots a_q]$$

pq

denote the matrix of the elements a_{ij} ($i = 1, 2, \dots, p; j = 1, 2, \dots, q$) arranged in a rectangular array of p rows and q columns, where a_j is the j 'th column vector of the matrix A . The transpose A' of matrix A is indicated by the addition of a prime ($'$) to the letter of the matrix. The transpose of vector a_j is denoted by a_j' . The symbols $O = O$ and $I = I$ represent the zero matrix and the identity matrix, respectively. The symbol $\text{diag}(a_1, a_2, \dots, a_p)$ is used to denote the diagonal matrix with elements a_1, a_2, \dots, a_p . The symbol $\text{Diag}[A]$, where $A = A$, is reserved to denote the matrix $\text{diag}(a_{11}, a_{22}, \dots, a_{pp})$. The sum of the diagonal elements of matrix A is denoted by $\text{tr}A$.

The letters Z, F, S are used to represent random vectors that are called also *random multivariate variables*. Moreover, the symbols $E(Z)$ and $C(ZF)$ are used to denote the vector of the expected values of the random vector $Z = Z$ and the covariance matrix of order $n \times m$ of the

vectors Z and F , respectively. In particular, we use $D(Z) = C(ZZ)$. Let

$z_i = z_i$ ($i = 1, 2, \dots$) represent a random sample of the random vector

$$Z = Z.$$

$n1$

In factor analysis it is assumed that the n -variate $Z = Z$ is connected with the variates $F = F$ and $S = S$ ($m \leq n$) by the relation that can be presented in matrix form as the mathematical model

$$(2.2) \quad Z = AF + BS,$$

where $A = A = [a_1, a_2, \dots, a_m]$ and

$$(2.3) \quad B = \text{diag}(b_1, b_2, \dots, b_n).$$

Vectors F (with components F_1, F_2, \dots, F_m) and S (with components S_1, S_2, \dots, S_n) are called *vector of common factors* and *vector of unique factors*, respectively. The components of vectors F and S are called *common factors* and *unique factors*, respectively. The matrices A and B of loadings of common and unique factors, respectively, are considered as the matrices of population parameters.

The column vector a_j of the matrix A is called the *vector of loadings of factor F_j* . The element b_i of the matrix B is called the *loading of factor S_i* . The estimators of A , B , a_j , b_i , represented by the symbols \hat{A} , \hat{B} , \hat{a}_j , \hat{b}_i respectively, are found by the method of maximum likelihood.

By (2.2) we have

$$(2.4) \quad D(Z) = AD(F)A' + AC(FC)B + BC(SF)A' + BD(S)B,$$

$$(2.5) \quad C(ZF) = AD(F) + BC(SF),$$

$$(2.6) \quad C(ZS) = AC(FS) + BD(S).$$

Let us assume that the n -dimensional variate Z is normally distributed. Without loss of generality we assume

$$(2.7) \quad E(Z) = \mathbf{O} \quad \text{and} \quad D(Z) = \Sigma,$$

where $\text{Diag}[\Sigma] = I$ and Σ is a positive definite matrix. We may write this assumption in the short form

$$(2.8) \quad Z \rightarrow N_n(\mathbf{O}; \Sigma).$$

For (2.7) and (2.2) we may assume that

$$(2.9) \quad E(F) = \mathbf{O}, \quad E(S) = \mathbf{O},$$

i.e. that the expected values of common and unique factors are equal to zero. We also assume that

$$(2.10) \quad D(F) = I, \quad D(S) = I, \quad C(FS) = \mathbf{O}.$$

From (2.4), (2.7) and (2.10) it follows that

$$(2.11) \quad \Sigma = AA' + B^2,$$

and from (2.5) and (2.6) we obtain

$$(2.12) \quad C(ZF) = \mathbf{A}, \quad C(ZS) = \mathbf{B}.$$

The sum of variances of the variates Z_i ($i = 1, 2, \dots, n$) of vector Z is denoted by ψ .

By (2.11) and (2.7), we have

$$(2.13) \quad \psi = \text{tr } \Sigma = \text{tr}(\mathbf{A}\mathbf{A}' + \mathbf{B}^2) = \text{tr}\mathbf{A}'\mathbf{A} + \text{tr}\mathbf{B}^2 = \sum_{j=1}^m \mathbf{a}'_j \mathbf{a}_j + \sum_{i=1}^n b_i^2 = n.$$

The components $\text{tr}\mathbf{A}'\mathbf{A}$, $\text{tr}\mathbf{B}^2$, $\mathbf{a}'_j \mathbf{a}_j$ ($j = 1, 2, \dots, m$) and b_i^2 ($i = 1, 2, \dots, n$) are variances and therefore are denoted by σ_F^2 , σ_S^2 , $\sigma_{F_j}^2$, $\sigma_{S_i}^2$. We call them in succession: σ_F^2 — total contribution of the variates F_j to ψ , σ_S^2 — total contribution of the variates S_i to ψ , $\sigma_{F_j}^2$ — contribution of the variate F to σ_F^2 , $\sigma_{S_i}^2$ — contribution of the variate S_i to the variance of the variate Z_i .

The matrix $\text{Diag}[\mathbf{A}\mathbf{A}']$ is called *communality matrix* and is denoted by $\text{diag}(\sigma_{FZ_1}^2, \sigma_{FZ_2}^2, \dots, \sigma_{FZ_n}^2)$. The element $\sigma_{FZ_i}^2$ ($i = 1, 2, \dots, n$) of this matrix is called *communality of variable Z_i* .

It is easily seen that

$$(2.14) \quad \text{Diag}[\Sigma] = \text{diag}(\sigma_{FZ_1}^2, \sigma_{FZ_2}^2, \dots, \sigma_{FZ_n}^2) + \text{diag}(\sigma_{S_1}^2, \sigma_{S_2}^2, \dots, \sigma_{S_n}^2)$$

and

$$(2.15) \quad \sigma_F^2 = \text{tr}\mathbf{A}\mathbf{A}' = \text{tr}(\text{Diag}[\mathbf{A}\mathbf{A}']) = \sum_{i=1}^n \sigma_{FZ_i}^2.$$

The variance estimators σ_F^2 , σ_S^2 , $\sigma_{F_j}^2$, $\sigma_{S_i}^2$ are represented by $\text{tr}\hat{\mathbf{A}}\hat{\mathbf{A}}'$, $\text{tr}\hat{\mathbf{B}}$, $\hat{\mathbf{a}}'_j \hat{\mathbf{a}}_j$, \hat{b}_i^2 , respectively, and denoted by $\hat{\sigma}_F^2$, $\hat{\sigma}_S^2$, $\hat{\sigma}_{F_j}^2$, $\hat{\sigma}_{S_i}^2$.

3. Estimation of the matrix of factor loadings by the maximum likelihood method. Suppose that (2.8) and (2.11) hold. Then the density function of the random variable Z is equal to (cf. [1])

$$(3.1) \quad f(\mathbf{z}; \mathbf{A}, \mathbf{B}) = c |\Sigma|^{-1/2} \exp - \frac{1}{2} \mathbf{z}' \Sigma^{-1} \mathbf{z},$$

where $c = (2\pi)^{-n/2}$.

Let

$$(3.2) \quad \underset{nN}{\mathbf{Z}} = \underset{nN}{\mathbf{Z}} = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N]$$

denote a matrix, where \mathbf{z}_k ($k = 1, 2, \dots, N$) represent a random sample of the random vector $\underset{n1}{\mathbf{Z}} = \mathbf{Z}$.

By (3.1) and (3.2), the likelihood function of the matrix \mathbf{Z} is of the form

$$(3.3) \quad L(\mathbf{Z}, \mathbf{A}, \mathbf{B}) = c^N |\Sigma|^{N/2} \exp - \frac{N}{2} \text{tr} \Sigma^{-1} \mathbf{R},$$

where

$$(3.4) \quad \mathbf{R} = \underset{nn}{\mathbf{R}} = \frac{1}{N} \mathbf{Z}\mathbf{Z}'$$

is a sample covariance matrix of the variables Z_1, Z_2, \dots, Z_n . According to the maximum likelihood method we obtain the estimators of the matrices \mathbf{A} and \mathbf{B} by differentiating $\ln L(\mathbf{Z}; \mathbf{A}, \mathbf{B})$ with respect to all elements of matrix \mathbf{A} , and with respect to all the diagonal elements of matrix \mathbf{B} . After putting the derivatives equal to zero, we obtain equations in the form (cf. [13])

$$(3.5) \quad -N\hat{\mathbf{A}}' \hat{\Sigma}^{-1} + N\hat{\mathbf{A}}' \hat{\Sigma}^{-1} \underset{mn}{\mathbf{R}} \hat{\Sigma}^{-1} = \mathbf{O},$$

$$(3.6) \quad -N \text{Diag}[\hat{\mathbf{B}} \hat{\Sigma}^{-1} - \hat{\mathbf{B}} \hat{\Sigma}^{-1} \underset{nn}{\mathbf{R}} \hat{\Sigma}^{-1}] = \mathbf{O},$$

where

$$(3.7) \quad \hat{\Sigma} = \hat{\mathbf{A}}\hat{\mathbf{A}}' + \hat{\mathbf{B}}^2.$$

Instead of (3.5) and (3.6) we write

$$(3.8) \quad \hat{\mathbf{A}} \hat{\Sigma}^{-1} (\mathbf{R} - \hat{\Sigma}) = \mathbf{O},$$

$$(3.9) \quad \text{Diag}[\mathbf{R} - \hat{\Sigma}] = \mathbf{O},$$

where $\hat{\Sigma}$ is given in (3.7).

Equations (3.8) and (3.9) have no unique solution.

Note that every pair of matrices $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$, being the solution of the equation.

$$(3.10) \quad \mathbf{R} - \hat{\Sigma} = \mathbf{O},$$

is the solution of equations (3.8) and (3.9).

4. Hotelling's, Thomson's and Lawley's methods of factor analysis.

In this section relations are given that bind together the three following fundamental factor analysis methods: 1. in the \mathcal{H} method of principal components [10], 2. in the \mathcal{T} method of principal components [17] as a modification of the \mathcal{H} method, and 3. in the \mathcal{L} method of maximum likelihood [12]. According to (3.10), the maximum likelihood estimator of covariance matrix Σ is given by the matrix \mathbf{R} . By the known matrix \mathbf{R} and, by (3.7), each of the three methods requires that the equations

$$(4.1) \quad \mathbf{R} = \hat{\mathbf{A}}\hat{\mathbf{A}}' + \hat{\mathbf{B}}^2$$

should be solved with respect to the estimators of the elements of the matrices \mathbf{A} and \mathbf{B} .

Note that the estimators mentioned above can be obtained by the maximum likelihood method for the following three mathematical models.

Model 1 is given by (2.2) under the condition that $\mathbf{B} = \mathbf{0}$ and under restrictions of the form ($m = n$)

$$(4.2) \quad \mathbf{A}\mathbf{A}' = \mathbf{A},$$

where $\mathbf{A} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m)$ and $\lambda_1 > \lambda_2 > \dots > \lambda_m$.

Models 2 and 3 are given by (2.2) under restrictions (4.2) ($m < n$) and

$$(4.3) \quad \mathbf{A}'\mathbf{B}^{-2}\mathbf{A} = \mathbf{E},$$

respectively, where $\mathbf{E} = \text{diag}(\xi_1, \xi_2, \dots, \xi_m)$ and $\xi_1 > \xi_2 > \dots > \xi_m$. The meaning of the symbols is given in Section 2.

As in the preceding section, n denotes the number of variables under observation. By the restrictions given above, the loading estimators which are the elements of the matrices \mathbf{A} and \mathbf{B} are unique.

The *a priori* information on the contributions of unique factor in a population structure implies the choice of the type of the restriction, i.e. it implies the choice of the \mathcal{L} or \mathcal{T} method. In case of equal contributions, the restrictions are given by (4.2).

Let us characterize in short the three methods under discussion.

1° The symbols \mathbf{A} and \mathbf{F} of Hotelling's model

$$(4.4) \quad \begin{matrix} \mathbf{Z} & = & \mathbf{A}\mathbf{F} \\ n1 & & mn1 \end{matrix}$$

are designated in Section 2. Moreover, let us take assumptions (2.7) and (2.10) for the random variable \mathbf{F} . Then the estimators $\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, \dots, \hat{\mathbf{a}}_n$ are found from the equations (cf. 3.10)

$$(4.5) \quad \mathbf{R} = \hat{\mathbf{A}}\hat{\mathbf{A}}'$$

under the first condition that the contribution of $\hat{\sigma}_{F_1}^2$ to the variance sum ψ (cf. § 2) of all n variables Z_1, Z_2, \dots, Z_n has as great a total as possible. Then the second condition is that the contribution of $\hat{\sigma}_{F_2}^2$ to the residual part of ψ after elimination of $\hat{\sigma}_{F_1}^2$ has as great a total as possible. This process is continued until the total communality is analysed. This gives the following equations (cf. (4.2)):

$$(4.6) \quad \hat{\mathbf{A}}'\mathbf{R} = \mathbf{A}\hat{\mathbf{A}}', \quad \hat{\mathbf{A}}'\hat{\mathbf{A}} = \mathbf{A}.$$

From equations (4.6) it follows that the estimator of the communality matrix is given by the matrix \mathbf{A} in the form

$$(4.7) \quad \text{diag}(\hat{\sigma}_{F_1}^2, \hat{\sigma}_{F_2}^2, \dots, \hat{\sigma}_{F_n}^2) = \mathbf{A},$$

where $\hat{\sigma}_{F_1}^2 > \hat{\sigma}_{F_2}^2 > \dots > \hat{\sigma}_{F_n}^2$.

2° In the \mathcal{T} method the matrix estimators $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ can be obtained from the equations

$$(4.8) \quad \begin{cases} \hat{\mathbf{A}}'(\mathbf{R} - \hat{\mathbf{B}}^2) = \Lambda \hat{\mathbf{A}}', \\ \hat{\mathbf{A}}' \hat{\mathbf{A}} = \Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_m), \\ \text{Diag}[\mathbf{R} - \hat{\mathbf{A}} \hat{\mathbf{A}}' - \hat{\mathbf{B}}^2] = \mathbf{O}, \end{cases}$$

where $\lambda_1 > \lambda_2 > \dots > \lambda_m > 0$. It can be verified that, similarly as in 1°, the solutions of equations (3.8) and (3.9) are given by $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}^2$.

From equations (4.8) and then from the restrictions for λ_i it follows that

$$(4.9) \quad \text{diag}(\hat{\sigma}_{F_1}^2, \hat{\sigma}_{F_2}^2, \dots, \hat{\sigma}_{F_m}^2) = \Lambda$$

and $\hat{\sigma}_{F_1}^2 > \hat{\sigma}_{F_2}^2 > \dots > \hat{\sigma}_{F_m}^2$.

It means that the factors F_i are ordered according to their decreasing contribution to the variance $\hat{\sigma}_{F_i}^2$. From the formal point of view the \mathcal{T} -method can be considered as the \mathcal{H} -method applied to the matrix $\mathbf{R} - \hat{\mathbf{B}}^2$.

3° If we use restriction (4.3), we note that the \mathcal{L} -method requires the solution of equations (3.8) and (3.9).

By using these restrictions in equation (3.8) we obtain

$$(4.10) \quad \hat{\mathbf{A}}' \hat{\mathbf{B}}^{-2} (\mathbf{R} - \hat{\mathbf{B}}^2) = \Xi \hat{\mathbf{A}}', \quad \hat{\mathbf{A}}' \hat{\mathbf{B}}^{-2} \hat{\mathbf{A}} = \Xi, \quad \text{Diag}[\mathbf{R} - \hat{\mathbf{A}} \hat{\mathbf{A}}' - \hat{\mathbf{B}}^2] = \mathbf{O}.$$

The estimators obtained by using each of the methods under assumptions (2.8) can be considered as the maximum likelihood estimators.

5. Solution of the maximum likelihood equations under two types of restrictions. From the point of view of calculations, the \mathcal{H} , \mathcal{T} and \mathcal{L} -methods can be reduced to finding the eigenvectors of matrices \mathbf{R} , $\mathbf{R} - \hat{\mathbf{B}}^2$, and $\hat{\mathbf{B}}^2 (\mathbf{R} - \hat{\mathbf{B}}^2)$, respectively. There are several methods, both exact and approximate, to find the eigenvectors of a given matrix. In the \mathcal{H} -method the exhaustion method is applied under the requirement that restriction (4.2) is fulfilled by the eigenvectors.

Let us note that the matrices $\mathbf{R} - \hat{\mathbf{B}}^2$ and $\hat{\mathbf{B}}^2 (\mathbf{R} - \hat{\mathbf{B}}^2)$ are unknown matrices since $\hat{\mathbf{B}}$ is unknown. $\hat{\mathbf{B}}$ is a diagonal matrix with non-negative elements, therefore, in order to solve equations (4.8) and (4.10), the approximate methods for estimating $\hat{\mathbf{B}}^2$ and $\hat{\mathbf{A}}$ with the accuracy given *a priori* are used.

Lawley's iterative method to solve equations (4.10) under a given number m of common factors is used.

This method is as follows. Let $\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, \dots, \hat{\mathbf{a}}_m$ denote the columns of the matrix $\hat{\mathbf{A}}$ of equations (4.10). The symbols $B_{(k)}^2$ and $\mathbf{a}_{1(k)}, \mathbf{a}_{2(k)}, \dots$,

..., $\mathbf{a}_{m(k)}$ are reserved to denote the k -th approximation of the matrix \mathbf{B}^2 and of the vectors $\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, \dots, \hat{\mathbf{a}}_m$, respectively. The iterative process is based on the formulas

$$(5.1) \quad \mathbf{B}_{(k)}^2 = \text{Diag} \left[\mathbf{R} - \sum_{j=1}^m \mathbf{a}_{j(k)} \mathbf{a}'_{j(k)} \right],$$

$$\mathbf{a}_{j(k+1)} = \mathbf{d}_{j(k)} / \sqrt{\mathbf{d}'_{j(k)} \mathbf{c}_{j(k)}},$$

where

$$\mathbf{c}_{j(k)} = \mathbf{a}_{j(k)} \mathbf{B}_{(k)}^{-2}, \quad \mathbf{d}_{j(k)} = \left(\mathbf{R} - \mathbf{B}_{(k)}^2 - \sum_{i=1}^{j-1} \mathbf{a}_{i(k)} \mathbf{a}'_{i(k)} \right) \mathbf{c}_{j(k)}.$$

The preliminary approximation of the vectors $\hat{\mathbf{a}}_1, \hat{\mathbf{a}}_2, \dots, \hat{\mathbf{a}}_m$ can be reached by m successive eigenvectors $\mathbf{a}_{1(0)}, \mathbf{a}_{2(0)}, \dots, \mathbf{a}_{m(0)}$ of the matrix \mathbf{R} under the conditions

$$(5.2) \quad \mathbf{a}'_{i(0)} \mathbf{a}_{j(0)} = \begin{cases} 0 & \text{for } i \neq j, \\ \lambda_i & \text{for } i = j \quad (i, j = 1, 2, \dots, m, \dots, n), \end{cases}$$

where $\lambda_1 > \lambda_2 > \dots > \lambda_m > \dots > \lambda_n$ are the eigenvalues of the matrix \mathbf{R} . The iterative process can be ended when there are no differences between the successive approximations, i.e. they agree within errors of rounding. Let us note that the vectors $\mathbf{a}_{i(0)}$ ($i = 1, 2, \dots$) underlying conditions (5.2) are the solutions of equations (4.6), i.e. they are the vectors of loadings of the corresponding factors F_i in the \mathcal{H} -method.

The iterative process based on formulas (5.1) can also be applied to solve equations (4.8) if we put

$$(5.3) \quad \mathbf{c}_{j(k)} = \mathbf{a}_{j(k)},$$

where $\mathbf{a}_{j(k)}$ denotes the k -th approximation of the vector $\hat{\mathbf{a}}_j$ (the j -th column of the matrix $\hat{\mathbf{A}}$ of equations (4.8)).

6. Determination of the number of common factors. In order to determine the number of common factors in a model by applying the \mathcal{T} and \mathcal{L} -methods it is suggested to use the method of vanishing of the residual matrix (cf. (6.1)) and Bartlett's approximate chi-square [3] test, respectively.

Let $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ denote such estimators of the matrices \mathbf{A} and \mathbf{B} , respectively, that are the solutions of equations (4.8) for a number $p \leq n$ ($p = 1, 2, \dots$) given *a priori*.

The matrix

$$(6.1) \quad \mathbf{R}_p = \mathbf{R} - \hat{\mathbf{B}}^2 - \hat{\mathbf{A}} \hat{\mathbf{A}}'$$

is called the p -th residual matrix.

The number of common factors $m = p$, when $R_p \approx O$, is determined by the method of vanishing the residual matrix. In practice the system (4.8) is solved for successive values $p = 1, 2, \dots$ until the corresponding matrix R_p is the null matrix within errors of rounding.

Bartlett's approximate chi-square test is based on the test function [13] given by

$$(6.2) \quad \chi^2 = N' \sum_{j < i}^n (r_{ij}^p)^2 / \hat{b}_j^2 \hat{b}_i^2$$

with $\nu = \frac{1}{2}[(n-p)^2 - (n+p)]$ degrees of freedom, where

$$(6.3) \quad N' = N - \frac{1}{6}(2n+5) - \frac{2}{3}p - 1$$

and where r_{ij}^p denotes the (i, j) -element of the matrix R_p .

The number m is determined by such a value $p = 1, 2, \dots$ that gives the value of chi-square in (6.2) and that is for the first time smaller than the critical value χ^2 at a given level of confidence, say

$$(6.4) \quad \chi_p^2 < \chi_{0.05}^2,$$

where $\chi_{0.05}^2$ denotes the 5% critical value χ^2 .

7. Numerical example. For illustration of the methods of factor analysis given in Section 4 let us use the following example. The data representing the observations of $n = 10$ characteristics of in $N = 331$ vetch plants have been obtained from the Chair of Plant Breeding and Seed Production of the Agricultural College, Lublin. The following successive characteristics are numbered from 1 to 10: height of a plant (1), height of the position of the first pod (2), weight of seeds from one plant (3), weight of 1000 seeds (4), length of leaf (5), length of leaflet (6), breadth of leaflet (7), number of days from germination to flowering (8), length of pod (9), and breadth of pod (10).

Numerical calculations have been done on the electronic computer "Odra 1013".

Let us assume that the random variables X_1, X_2, \dots, X_{10} corresponding to the characteristics under consideration are normally distributed with means μ_i and variances σ_i^2 ($i = 1, 2, \dots, 10$).

The sample mean

$$\bar{x}_i = \hat{\mu} = \frac{1}{N} \sum_{j=1}^N x_{ij}$$

is the estimator of the mean μ_i of the i -th characteristic. The estimator of the variance σ_i^2 is given by the sample variance

$$S_i^2 = \hat{\sigma}_i^2 = \frac{1}{N} \sum_{j=1}^N (x_{ij} - \bar{x}_i)^2.$$

TABLE 1. Observational data

Charac- teristics	Plants					
	1	2	...	j	...	$N = 331$
1	$x_{11} = 104$	$x_{12} = 70$...	x_{1j}	...	$x_{1,331} = 102$
2	$x_{21} = 34$	$x_{22} = 30$...	x_{2j}	...	$x_{2,331} = 41$
...
i	x_{i1}	x_{i2}	...	x_{ij}	...	$x_{i,331}$
.
.
.
$n = 10$	$x_{10,1} = 9$	$x_{10,2} = 10$...	$x_{10,j}$...	$x_{10,331} = 10$

In view of the great number of plants ($N = 331$) we assume that the values of random variables Z_i are of the form

$$(7.1) \quad z_{ij} = \frac{x_{ij} - \mu}{\sigma_i} \approx \frac{x_{ij} - \bar{x}_i}{S_i} \quad (i = 1, 2, \dots, 10; j = 1, 2, \dots, 331),$$

i.e. they are approximately the values of the random variables with zero means and unit variances.

Using the values x_{ij} of Table 1, we calculate the corresponding standardized values z_{ij} from formula (7.1).

The covariance matrix \mathbf{R} (cf. (4.6), (4.8) and (4.10)) is calculated by using (3.4) and given in Table 2, where $\mathbf{Z} = \mathbf{Z} = \mathbf{Z}$ (cf. (3.2)) denotes the nN $10,331$ matrix with the elements z_{ij} ($i = 1, 2, \dots, n; j = 1, 2, \dots, N$). Let us note that, by (7.1), \mathbf{R} is the matrix of the correlation coefficients with unities in diagonal.

TABLE 2. Correlation matrix for $n = 10$ characteristics in $N = 331$ vetch plants

Charac- teristics	1	2	3	4	5	6	7	8	9	10
1	1,00									
2	0,35	1,00								
3	0,36	0,15	1,00							
4	0,06	0,02	0,39	1,00						
5	0,28	0,25	0,11	0,10	1,00					
6	0,25	0,25	0,08	0,07	0,58	1,00				
7	0,22	0,18	0,04	0,07	0,52	0,75	1,00			
8	0,05	-0,05	-0,09	0,09	-0,04	0,07	0,09	1,00		
9	0,13	0,12	0,14	0,16	0,13	0,14	0,12	0,07	1,00	
10	0,13	0,21	0,03	0,29	0,23	0,24	0,25	0,16	0,45	1,00

The system of equations (4.6) is solved by the method of exhaustion (cf. [4] and [5]) that has been introduced to factor analysis by Hotelling.

The vectors of estimators of loadings for ten common factors are given in Table 3.

The estimates $\hat{\sigma}_{F_j}^2$ ($j = 1, 2, \dots, 10$) of contribution of the successive common factors to the variance sum of all variables are given at the bottom of the table.

TABLE 3. Loading estimates and their contributions to the variance of the common factors in the \mathcal{H} -method

Charac- teristic	Factor									
	1	2	3	4	5	6	7	8	9	10
1	0,53	0,17	-0,45	-0,25	0,40	-0,27	-0,35	-0,18	0,19	-0,02
2	0,49	0,04	-0,33	-0,55	0,09	0,51	0,27	0,10	0,05	0,03
3	0,33	0,59	-0,51	0,28	0,06	-0,18	0,23	0,01	-0,34	0,02
4	0,31	0,63	0,09	0,52	-0,07	0,34	-0,07	0,02	0,30	-0,01
5	0,72	-0,30	-0,09	0,12	-0,15	0,01	-0,30	0,49	-0,10	0,05
6	0,78	-0,41	0,04	0,19	-0,02	-0,05	0,18	-0,11	0,02	-0,37
7	0,74	-0,43	0,10	0,23	-0,01	-0,06	0,17	-0,25	0,06	0,33
8	0,11	0,07	0,55	0,06	0,80	0,00	0,09	0,17	-0,06	0,01
9	0,41	0,42	0,38	-0,35	-0,29	-0,44	0,21	0,17	0,18	0,02
10	0,54	0,32	0,51	-0,24	-0,17	0,19	-0,26	-0,26	-0,30	-0,02
Estimate $\hat{\sigma}_{F_j}^2$	2,87	1,50	1,31	1,01	0,95	0,71	0,53	0,48	0,38	0,25

Using the iterative process based on formulas (5.1), equations (4.8) and (4.10) are solved successively for $p = 1, 2, \dots, m = 5$. Equations (4.8) are solved under conditions (5.3). The p initial estimates of corresponding vectors, which have been obtained by the \mathcal{H} -method and are given in Table 3, are considered as a preliminary approximation of the vector estimators $\hat{\mathbf{a}}_i$ ($i = 1, 2, \dots, p$).

In both methods the iterative process is ended for a given p if

$$|b_{i(k)}^2 - b_{i(k+1)}^2| < 0,0001,$$

where $b_{i(k)}$ and $b_{i(k+1)}$ ($i = 1, 2, \dots, n$) denote the elements of matrices $\mathbf{B}_{(k)}$ and $\mathbf{B}_{(k+1)}$, respectively.

The estimates are given by the k -th approximation in the form

$$\hat{\mathbf{A}} = [\mathbf{a}_{1(k)}, \mathbf{a}_{2(k)}, \dots, \mathbf{a}_{p(k)}] \quad \text{and} \quad \hat{\mathbf{B}}^2 = \mathbf{B}_{(k)}^2 \quad \text{each.}$$

The residual matrix \mathbf{R}_p is calculated by (6.1) for $p = 1, 2, 3, 4, 5$. Let us note that the matrix \mathbf{R} given in Table 2 with unities in diagonal

is, according to Section 2, of the form $\text{Diag}[\mathbf{R}] = \mathbf{I}$. Hence the third equality of (4.8) and (4.10) takes the form

$$(7.2) \quad \hat{\mathbf{B}}^2 = \mathbf{I} - \text{Diag}[\hat{\mathbf{A}}\hat{\mathbf{A}}'].$$

By (7.2) and (6.1), the elements of the residual matrix are zeros.

In the case of the \mathcal{L} -method, the values of the test function have been calculated from formula (6.2) for $p = 1, 2, 3, 4, 5$ and compared with the corresponding chi-square critical values. In virtue of $p = 5$, $N' = 322,24$, the test function value is equal to 11,60. By the critical value $\chi_{0,05}^2 = 11,70$ for five degrees of freedom we take $m = 5$.

The estimates of the loadings of common factors and the estimates of their contribution to the variance of common factors are given in Tables 4 and 5. Moreover, the values of ξ_j ($j = 1, 2, \dots, 5$) (cf. (4.3)), indicating how to order the common factors of the \mathcal{L} -method, are shown in Table 5. The corresponding symmetrical residual matrices \mathbf{R}_5 in the case of the \mathcal{H} and \mathcal{L} -methods are given in Tables 6 and 7. Zeros, as diagonal elements (cf. relations (6.1) and (7.2)), have been replaced by the estimates of the contributions to the unique factors.

TABLE 4. Loading estimates and their contributions to the variance of the common factors in the \mathcal{F} -method

Characteristic	Factor				
	1	2	3	4	5
1	0,47	0,09	-0,32	0,31	0,41
2	0,38	0,01	-0,17	0,01	0,29
3	0,29	0,35	-0,56	0,29	-0,20
4	0,26	0,43	-0,14	0,03	-0,38
5	0,62	-0,23	-0,05	-0,02	-0,02
6	0,79	-0,39	0,10	0,03	-0,14
7	0,72	-0,35	0,14	0,02	-0,14
8	0,13	0,32	0,71	0,56	0,02
9	0,32	0,30	0,03	-0,16	0,05
10	0,58	0,54	0,27	-0,50	0,12
Estimate $\sigma_{F_j}^2$	2,50	1,13	1,07	0,77	0,49

Let us compare the estimates of loadings and estimates of variances obtained in the three methods and let us consider the ordered arrangement of common factors.

In the \mathcal{H} and \mathcal{F} methods, the ordered arrangement of common factors F_j ($j = 1, 2, 3, 4, 5$) is determined by their decreasing contribution to the variance $\sigma_{F_j}^2$ (cf. Section 4), while in the \mathcal{L} -method it is defined by

TABLE 5. Loading estimates and their contribution to the variance of common factors in the \mathcal{L} -method

Characteristic	Factor				
	1	2	3	4	5
1	0,07	-0,37	0,01	0,49	0,45
2	-0,02	-0,36	-0,06	0,16	0,32
3	-0,08	-0,21	-0,11	0,76	-0,18
4	0,11	-0,21	-0,32	0,33	-0,40
5	0,00	-0,64	0,16	0,02	0,05
6	0,12	-0,82	0,35	-0,07	-0,06
7	0,14	-0,75	0,30	-0,11	-0,06
8	0,98	0,07	0,03	0,02	0,00
9	0,10	-0,31	-0,38	0,04	0,01
10	0,22	-0,54	-0,68	-0,17	0,02
Estimate $\hat{\sigma}_{F_j}^2$	1,08	2,39	0,96	1,00	0,51
ξ_j	24,23	8,90	4,16	2,79	1,03

TABLE 6. Residual matrix R_5 and the estimates of contributions to unique factors in the \mathcal{T} -method

Characteristics	1	2	3	4	5	6	7	8	9	10
1	0,41									
2	0,00	0,75								
3	0,00	0,00	0,35							
4	0,00	0,00	0,00	0,59						
5	0,00	0,02	-0,02	0,02	0,56					
6	-0,01	0,01	0,01	-0,01	0,00	0,18				
7	0,01	-0,02	0,00	0,00	0,00	0,00	0,32			
8	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,06		
9	0,00	-0,01	0,02	-0,02	0,00	0,01	-0,01	0,00	0,78	
10	0,00	0,01	-0,01	0,01	-0,01	0,00	0,01	0,00	0,01	0,04

the relation $\xi_1 > \xi_2 > \xi_3 > \xi_4 > \xi_5$, where ξ_i are defined in (4.3) and are shown in Table 5.

The estimates $\hat{\sigma}_{F_j}^2$ ($j = 1, 2, 3, 4, 5$), obtained in the \mathcal{H} -method, are greater than the corresponding estimates of the \mathcal{T} -method (cf. Tables 4 and 5).

The differences of the corresponding estimates $\hat{\sigma}_{S_i}^2$ ($i = 1, 2, \dots, 10$) in the \mathcal{T} and \mathcal{L} -methods are small, except $\hat{\sigma}_{S_{10}}^2$ (cf. Tables 6 and 7). The estimates $\hat{\sigma}_F^2$ in the \mathcal{T} and \mathcal{L} -methods are equal to 5,96 and 5,94, respectively, and can be considered as equal, while the estimate of the contribution of five initial common factors in the \mathcal{H} -method is equal to 7,64.

TABLE 7. Residual matrix R_5 and the estimates of contributions to unique factors in the \mathcal{L} -method

Characteristics	1	2	3	4	5	6	7	8	9	10
1	0,42									
2	0,00	0,74								
3	0,00	0,00	0,33							
4	0,00	0,00	0,00	0,58						
5	0,01	0,02	-0,02	0,03	0,56					
6	-0,01	0,01	0,00	-0,01	0,00	0,18				
7	0,01	-0,02	0,00	0,00	0,00	0,00	0,32			
8	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,04		
9	-0,01	-0,03	0,02	-0,04	0,00	0,01	-0,01	0,00	0,74	
10	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,00	0,01	0,17

In the \mathcal{H} and \mathcal{T} -methods the first factor, and in the \mathcal{L} -method the second factor, have the greatest contributions to $\hat{\sigma}_F^2$. The corresponding contributions in the \mathcal{H} , \mathcal{T} and \mathcal{L} -methods are equal to 2,87, 2,50 and 2,39, respectively. These factors have the highest correlation with characteristics 5, 6 and 7, and the smallest correlation with characteristics 8.

The interpretation of common factors may be based on comparisons of loadings shown in Tables 3, 4 and 5. Considering Table 3 we see that the highest correlation coefficients for first factor F_1 (first column) equal to 0,72, 0,78 and 0,74, correspond to characteristics 5 (length of leaf), 6 (length of leaflet) and 7 (breadth of leaflet). They define the shape of the leaf of vetch. We also state that the highest correlation coefficients for factor F_2 (second column) equal to 0,59 and 0,63 correspond to characteristics 3 (weight of seeds) and 4 (weight of 1000 seeds). We follow this way for the remaining columns.

The characteristics with the highest correlation coefficients can serve a plant breeder for determining the names of the successive common factors.

Analysing the results shown in Tables 3, 4 and 5, we divide nine out of the ten characteristics under consideration into four groups. The characteristics contained in the same group have similar correlations with the common factors. Thus the height of plant (1) and the height of position of the first pod (2) form the first group; then the weight of seeds of one plant (3) and the weight of 1000 seeds (4) form the second group, the length of leaf (5), the length of leaflet (6) and the breadth of leaflet (7) form the third group, and the length (9) and breadth of pod (10) form the fourth group. Characteristic (8), i.e. the number of days from germination to the end of flowering, should be treated separately because of a small degree of agreement of this characteristic in correlation with the

individual common factors, in contradistinction to the remaining nine characteristics. The agreement of the correlation between this characteristic and the others is rather small and not so high as the accordance with respect to the nine remaining characteristics.

References

- [1] T. W. Anderson, *Introduction to multivariate statistical analysis*, New York 1952.
- [2] — and H. Rubin, *Statistical inference in factor analysis*, Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability 5 (1956), p. 111-150.
- [3] M. S. Bartlett, *A note on the multiplying factor for various χ^2 approximations*, J. Roy. Statist. Soc. B 16 (1954), p. 296-298.
- [4] B. P. Demidowicz and J. A. Maron, *Metody numeryczne, część I*, Warszawa 1965.
- [5] W. N. Faddiejewa, *Metody numeryczne algebry liniowej*, Warszawa 1955.
- [6] B. Fruchter, *Introduction to factor analysis*, New York 1954.
- [7] J. C. Gower, *Multivariate analysis and multidimensional geometry*, The Statistician 17 (1967), p. 13-28.
- [8] H. H. Harman, *Modern factor analysis*, Chicago 1960.
- [9] K. J. Holzinger and H. H. Harman, *Factor analysis*, Chicago 1941.
- [10] H. Hotelling, *Analysis of a complex of statistical variables into principal components*, J. Educat. Psych. 24 (1933), pp. 417-441, 498-520.
- [11] — *Simplified calculation of principal components*, Psychometrika 1 (1936), p. 27-35.
- [12] D. N. Lawley, *The estimation of factor loadings by the method of maximum likelihood*, Proc. Roy. Soc. Edinburgh 60 (1940), p. 64-82.
- [13] D. N. Lawley and A. E. Maxwell, *Factor analysis as a statistical method*, London 1963.
- [14] F. Mullen, *Factors in the growth of girls seven to seventeen years of age*, Ph. D. dissertation, Dept. of Education, Univ. of Chicago, 1939.
- [15] J. Okóń, *Analiza czynnikowa w psychologii*, Warszawa 1964.
- [16] K. Pearson, *On lines and planes of closest fit to systems of points in spaces*, Phil. Mag. 6 (1901), p. 559-572.
- [17] G. H. Thomson, *Hotelling's method modified to give Spearman's g*, J. Educat. Psych. 25 (1934), p. 366-374.
- [18] L. L. Thurstone, *Multiple-factor analysis*, Chicago 1947.

DEPT. OF MATHEMATICAL STATISTICS
 AGRICULTURAL COLLEGE
 LUBLIN

Received on 15. 7. 1969;
 revised version on 3. 12. 1969



A. KIELOCH i W. OKTABA (Lublin)

ANALIZA CZYNNIKOWA METODĄ NAJWIĘKSZEJ WIAROGODNOŚCI

STRESZCZENIE

W pracy wskazano na proste relacje wiążące trzy modele matematyczne analizy czynnikowej, którym odpowiadają trzy metody: 1. \mathcal{H} składowych głównych Hotellinga, 2. \mathcal{T} Thomsona, 3. \mathcal{L} największej wiarogodności Lawleya.

Do ilustracji stosowania trzech metod posłużył przykład liczbowy z hodowli roślin. W oparciu o metody iteracyjne, dla których ułożono programy na cyfrową maszynę „Odra 1013”, wyznaczono macierze ocen ładunków czynników wspólnych i macierze czynników swoistych w metodach \mathcal{H} , \mathcal{T} i \mathcal{L} .

Porównano oceny parametrów uzyskane trzema metodami.
