▼ Journal of Research (Science), Bahauddin Zakariya University, Multan, Pakistan.
Vol.13, No.1, June 2002, pp. 87-98 ISSN 1021-1012

THE USE OF CANONICAL PRINCIPLE AND FACTORS ANALYSIS

G. R. Pasha
Chairman, Department of Statistics, Dean, Faculty of Science and Agriculture, Bahauddin Zakariya University, Multan, Pakistan.

Abstract: In this research the question of dimension reducing procedures are of interest, comprises canonical correlation, which investigates the correlation between linear combinations of variables, and principal component analysis that reduces the set of measurement characteristics to fewer components. Factor analysis performs the same reduction but assuming that the observations have an underlying structure. The auxiliary of principal component is taken for a more parsimonious data representation. The deficiencies of principal component analysis are overcome. Rotation of factors application is discussed.

Keywords: Canonical principle, correlation factors, dimensions reducing procedure, deficiencies, linear combinations.

INTRODUCTION

The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set, which consists of a large number of interrelated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and which are ordered so that the first few retain most of the variation present in all of the original variables. Data from multivariate population involve repeated observations on p possibly correlated random variables. Sometimes when p is large, it is however natural that these random variables, although related to each other, may not all contain the same amount of information, and in fact some random variables may be completely redundant. It is obvious to seek ways of rearranging or summarizing the data so that with minimum loss of information as possible, the dimension of the problem is reduced. This task is simplified if the transformations are made in such a way that these p linear functions become uncorrelated (independent, if the original p random variables are normally distributed), for then we could discard from study function reflecting less variability and consider only those functions that have higher variances.

Principal component analysis does not only mean or always work in the sense, the reduction from large number of original variables into small number of transformed variables. If the original variables are uncorrelated then PCA does absolutely nothing. Highly correlated variables surrender best results. And it is quite conceivable that 20 or 30 original variables can be adequately represented by two or three principal components. Nevertheless, it will be also of worth to know that there is a good deal of redundancy in the original variables, with most of them measuring same things.

Principal components analysis was essentially developed by Hoteling [1933] after its origination, Pearson [1901] who studied the problem for the nonstochastic variables. For details and various applications see Rao [1964].

DEFINITION AND DERIVATION OF PRINCIPAL COMPONENTS

Suppose that X is a vector of p random variables and the structure of the covariances or correlations between the p variables are of interest. Unless p is small, or the structure is very simple, it will often not be very helpful to simply look at the p variances and all of the $\frac{1}{2}$ p(p-1) correlations or covariances. An alternative approach is to look for a few (<p) derived variables, which preserve most of the information given by these variances and correlations or covariances.

Although PCA does not ignore covariances and correlations, it concentrates on variances. The first step is to look for a linear function $\alpha_1^t X$ of the elements of *X* which has maximum variances, where α is a vector of p constants $\alpha_{11}, \alpha_{12}, \ldots, \alpha_{1p}$, and t denotes transpose, so that

$$\alpha_1^t X = \alpha_{11} x_1 + \alpha_{12} x_2 + \ldots\ldots + \alpha_{1p} x_p$$

$$= \prod_{j=i}^{P} \alpha_{ij} x_j \qquad (2.1)$$

Next, look for a linear function, uncorrelated with $\alpha_2^t X$, which has maximum variance, and so on, so that at the kth stage a linear function $\alpha_1^t X$ is found which has maximum variance subject to being uncorrelated with $\alpha_1^t X$, $\alpha_2^t X \ldots\ldots\ldots \alpha_{r-1}^t X$. The kth derived variable, $\alpha_k^t X$, is the kth PC. Upto p PCs could be found but it is hoped, in general, that most of the variation in *X* will be accounted by m PCs, where m<p.

PRINCIPAL COMPONENTS BASED ON POPULATION

We shall assume that the vector $X = (X_1, \ldots\ldots, X_p)^t$ are measured in the same or comparable units. And if the sample size is large so that the final results of a PCA will be meaningful, new standardized variable should by formed. The relevant covariance matrix will then be a sample correlation matrix, and now all the components will be measured in the same units. Let *X* be a p-component vector with covariance matrix $\sum$. Since $\sum$ is a symmetric and positive definite matrix. Let α denote a p-variate vector of unknown weights for each of the components of *X*, and let Z_1 denote the scalar

$$Z_1 = \alpha^t \; X$$

$$= \prod_{k=1}^{p} \alpha_k \; x_k \tag{3.1}$$

Then there exists an orthogonal matrix $\alpha = (\alpha_1 \; .....\alpha_p)$, $\alpha \; \alpha^t = I_p$ such that

$$\alpha^t \; \sum \alpha = D_\lambda \tag{3.2}$$

where D_λ = diag $(\lambda_1.....,\lambda_p)$.

Let $Z = \alpha^t \; X$

$$= \begin{bmatrix} \alpha_1^t \\ \vdots \\ \alpha_p^t \end{bmatrix} X$$

$$= \begin{bmatrix} \alpha_1^t X \\ \vdots \\ \alpha_p^t X \end{bmatrix} \tag{3.3}$$

Then cov(Z)=$\alpha^t \sum \alpha = D_\lambda$, and components $Z_1 = \alpha_1^t X,.....,Z_p = \alpha_p^t X$ are uncorrelated. These p-variate uncorrelated linear function components are the principal components.

Now maximize var (Z_1). Since $Z_1 = \alpha_1^t X$, var $(Z_1) = \alpha^t \sum \alpha$.

$$\max_{\alpha} \left(\alpha^t \; \sum \alpha\right) \tag{3.4}$$

It is required to fulfill the condition

$$\alpha^t \alpha = \prod_{1}^{p} \alpha^2$$

$$\alpha^t \alpha = 1 \tag{3.5}$$

If $A = A^t$, and λ_j denote the latent roots of A

$$\min_{i \le j \le p} (\lambda_j) \le \frac{X^t AX}{X^t X} \le \min_{i \le j \le p} (\lambda_j) \tag{3.6}$$

there exists an orthogonal matrix α such that

$$A = \alpha \; D_\lambda \; \alpha^t \tag{3.7}$$

where D_λ = diag $(\lambda_1.....,\lambda_p)$ are latent roots of A and if the latent roots are distinct.

The representation with the elements of D_λ are arranged to that $\lambda_1 > , \ldots\ldots\ldots\ldots , > \lambda_p$ is unique. For proof of the main results, see Mirsky [1955].

The algebraic result from equation (3.6) shows that maximizing the *var*(Z_i) is the largest latent root of $\sum$. Using Lagrangian

$$W = \alpha^t \sum \alpha - \theta \left(\alpha^t \alpha - 1\right) \tag{3.8}$$

where θ is a Lagrangian multiplier, and maximize *W* using the constraint. Differentiating *W* with respect to α, and equating to zero yields

$$\frac{\partial W}{\partial \alpha} = 2\sum\alpha - 2\theta\alpha = 0 \tag{3.9}$$

$$\left(\sum - \theta\ I\right)\alpha = 0 \tag{3.10}$$

since $\alpha \neq 0$, there can be a solution only if

$$\left|\sum - \theta\ I\right| = 0 \tag{3.11}$$

From equation (3.10) , $\sum \alpha = \theta\ \alpha$. Pre-multiplication by α^t gives

$$\alpha^t \sum \alpha = \theta\ \alpha^t \alpha$$

$$= \theta\ = \mathrm{var}\left(Z_1\right) \tag{3.12}$$

To maximize $\mathrm{var}\left(Z_1\right)$ take θ as large as possible. The solutions to equation (3.10) and equation (3.11) are

$\left(\alpha_1\ , \theta_1\right)$;

$$Z_1 = \alpha_1^t X \tag{3.13}$$

The result is summarized in the following theorem.

THEOREM

Suppose for $X\left(p \times 1\right),\ E\left(X\right) = \phi$, and

$\mathrm{var}\ \left(X\right) = \sum$; *let* $\theta_1 \geq \theta_2 \geq \ldots\ldots\ldots\ldots \geq \theta_p \geq 0$ denotes the latent roots of $\sum$, and let $\alpha_1 , \ldots\ldots\ldots\ldots , \alpha_p$ denote the corresponding normalized latent vectors of $\sum$; let $Z_1 = \alpha_1^t X$, and constraint α so that $\alpha^t \alpha = 1$. Then, if $\alpha = \alpha_1$, Z_1 is called the first principal component of X. and $\mathrm{var}\left(Z_1\right) = \theta_1$.

Define $Z_2 = \alpha_2^t X$, where now not only $\alpha^t \alpha = 1$, but in addition, it is required Z_2 be orthogonal to Z_1. It is straightforward given by using two Lagrangian multipliers for the two constraints that $\mathrm{var}\left(Z_2\right)$ is maximized for $\alpha = \alpha_2$.

Nevertheless, define $Z_3 = \alpha_3^t X$, and require for $\alpha^t \alpha = 1$, Z_3 be orthogonal to both Z_1 and Z_2 moreover generate a set of p orthogonal axes that are obtained by applying the normalized latent vectors of $\sum$.

EXAMPLE ON PRINCIPAL COMPONENTS

A work for determining whether the process was under 'control' was done by two statisticians [Jackson and Morris 1957]. There are many variables involved and especially each of them is checked simultaneously to sure whether that variable is still under control. To measure the quality 9 variables were included and basically these variables were correlated. Computations on the covariance matrix were taken up in an attempt to reduce the dimension of the problem so that only a few variables need to be examined for control.

Table 4.1: Variance – Covariance Matrix

X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
177								
179	419							
95	245	302						
96	131	60	158					
53	181	109	102	137				
32	127	142	42	96	128			
-7	-2	4	4	4	2	34		
-4	1	4	3	5	2	31	39	
-3	4	11	2	6	8	33	39	48

Table 4.2: Eigen values Based on Table 4.1

Component	Eigen value	Cumulative proportion of total variance
1	878.519228	0.609237
2	196.095890	0.745225
3	128.643106	0.834437
4	103.430205	0.906164
5	81.261078	0.962517
6	37.848834	0.988764
7	6.976383	0.993602
8	5.706471	0.997560
9	3.518806	1.000000

Table 4.3: Characteristic Vectors of the Var-Covariance Matrix shown in Table 4.1

Components	Variables								
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
I	0.305	0.654	0.482	0.261	0.323	0.271	0.002	0.006	0.014
II	-0.485	-0.150	0.587	-0.491	-0.038	0.376	0.057	0.053	0.088
III	-0.412	-0.182	-0.235	0.457	0.495	0.268	0.256	0.266	0.282

Table 4.4: Correlation between Components and Variables

Components	Variables								
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9
I	0.679	0.946	0.824	0.616	0.819	0.710	0.012	0.028	0.062
II	-0.511	-0.103	0.473	-0.517	-0.045	0.462	0.137	0.120	0.179
III	-0.351	-0.101	-0.154	0.412	0.480	0.269	0.497	0.483	0.462
r^2	0.845	0.916	0.928	0.849	0.903	0.791	0.266	0.249	0.249

GENERAL FACTOR MODEL

Suppose we wish to examine the aptitude of graduates going for admission in Department of Statistics, B.Z. University Multan. We may give them a test of 30 questions, yet may fall into few categories, such as reading comprehension, mathematical approach, and general knowledge etc. These categories are called factors. If there are only three factors, then the score on the *i*th question may be modeled in the form

$$Y_i = \theta_i + \lambda_{i1} F_1 + \lambda_{i2} F_2 + \lambda_{i3} F_3 + e_i \qquad i = 1, \ldots\ldots 30 \; ; \; j = 1, 2,3$$

More generally, if there are *k* factors and *p* score $(p \geq k)$ in the test, then we can write

$$Y_i = \theta_i + \prod_{j=1}^{k} \lambda_{ij} F_i + e_i \tag{5.1}$$

where the parameter λ_{ij}, relating the *i*th variable with the *j*th factor, is called a factor loading.

In the matrix form equation (5.1) reduces to

$$Y = \theta + \wedge F + e$$

(5.2)

where *Y* is a vector of $(p \times 1), \theta = (\theta_1, \ldots\ldots, \theta_p)^t, \Lambda$ is a matrix of coefficient parameter with order $(p \times k), F = (F_1, \ldots\ldots, F_k)^t$ and $e = (e_1, \ldots\ldots, e_p)^t$.

Each element of *Y* may be thought of as having been generated by a linear combination of orthogonal, unobservable factors upon which some distribution term has been superimposed. The general factor model restrains the following assumptions:

$$W(F) = N(0,1) \; , \; k \leq p; \tag{5.3}$$

$$W(e) = N(0, \psi) \tag{5.4}$$

where $\psi = diag \; (\psi_1, \ldots\ldots \psi_p)$ *e* and *F* are independent;

$$\Lambda^t \psi^{-1} \Lambda = A : k \times k \tag{5.5}$$

$$= diag \, (\Lambda_1, \ldots\ldots \Lambda_k).$$

where $\Lambda_1 > \ldots\ldots > A_k$.

MAXIMUM LIKELIHOOD ESTIMATES

The original derivations of maximum likelihood estimates (MLE) of Λ and ψ are given by Lawley [1940] and Lawley and Maxwell [1970]. We derive the estimators as solution of these equations.

$$diag \; S = diag \left(\Lambda \; \Lambda^t + \psi \right) \tag{6.1}$$

and

$$\Lambda = S \left(\Lambda \; \Lambda^t + \psi \right)^{-1} \Lambda \tag{6.2}$$

The method of MLE requires $N > P$. If $N \leq P$, then one can estimate the factor loadings by such method as principal factor analysis. For detail of solution of above equation (6.1) and (6.2) [Joreskog and Von Thillo 1971]. It should be noted that adding the same constant to each variable has no effect on the covariance matrix. Similarly, if any variable is multiplied by a constant, then the factor loadings of the variable are multiplied by that constant. However in short, we shall assume that all variables are standardized to have 0 mean and variance 1. Now equation (5.2) becomes

$$Y^* = \Lambda F + e \tag{6.3}$$

Hence equation (6.1) can be written as

$$\prod_{j=1}^{k} \lambda_{ij}^2 + \psi_i = 1 \tag{6.4}$$

and equation (6.2) remains

$$\Lambda = S\left(\Lambda\ \Lambda^t + \psi\right)^{-1}\ \Lambda$$

where S is the simple covariance matrix of the standardized variables or the correlation matrix of the original variables.

PRINCIPAL FACTOR ANALYSIS

The standardization of equation (5.2) will result in equation (6.3). Furthermore, equation (6.1) deduces to

$$\prod_{j=1}^{k} \lambda_{ij}^2 + \psi_i = 1$$

or equivalently,

$$\psi_i = 1 - \prod_{j=1}^{k} \lambda_{ij}^2 \tag{7.1}$$

where $i = 1, \ldots\ldots, p$. If these residual variances are known, then by performing a principal component analysis on $\left(S - \psi\right)$, we may obtain an estimate of Λ. We can then estimate ψ_i from equation (7.1). The process is however iterated until the final solution is achieved. Details are given in Kaiser and Rice [1974].

SELECTING THE NUMBER OF FACTORS

The model (5.2) may be tested by means of involving a sequence of likelihood ratio test for testing the number of factors in a factor model. We test the hypothesis

$$H_\circ : \sum\ = \Lambda\ \Lambda^t + \psi \quad \text{against} \quad H_1 : \sum\ \neq \Lambda\ \Lambda^t + \psi$$

The likelihood ratio tests reject H_o if $\lambda = \dfrac{|S|^{N/2}}{\left|\hat{\Sigma}\right|^{N/2}}$

Asymptotic theory gives the distribution of (-2 In λ) as Chi-Square with *g* degree of freedom, where

$$g=\frac{1}{2}\left[\left(p-k\right)^2-\left(p+k\right)\right]$$

Bartlett [1954] proposed the approximation

$$-\left[n-\left(2p+5\right)/6-\left(2k/3\right)\right]\ \text{In}\left[|S|\ /\ |\hat{\Sigma}|\right]=\chi_g^2$$

Hence the hypothesis is rejected if $\chi_g^2 \geq \chi_{g.\alpha}^2$.

ROTATION PROBLEM

After the determination of provisional factor loadings, it can be observed that they are not unique. If $F_1, \ldots F_k$ are provisional factors and the initial selection of the factor loading Λ is completed, the next step is to rotate the factors. Thus the provisional factors are transformed into new factors that are easier to interpret. That is

$$F \sim N\left(0, I\right) \tag{9.1}$$

Premultiplication of orthogonal matrix $\Gamma\left(p\times p\right)$ equation (9.1) deduces to

$$F^* = \Gamma F \sim N\left(0, I\right) \tag{9.2}$$

Hence equation (6.3) can be written as

$$\begin{aligned} Y^* &= \Lambda\ F + e \\ &= \Lambda\ \Gamma^t\ \Gamma\ F + e \\ &= \Lambda\ \Gamma^t\ \ F^* + e \end{aligned} \tag{9.3}$$

The rotation of factors will result in the postmultiplication of the factor loading by an orthogonal matrix, If the condition of independent factors is relaxed, we may premultiply our factors by an arbitrary matrix *B*.

$$F^{**} = B\ \ F \tag{9.4}$$

Hence equation (6.3) becomes

$$\begin{aligned} Y^* &= \Lambda\ F + e \\ &= \Lambda\ B^{-1}\ BF + e \\ &= \Lambda\ B^{-1}\ F^{**} + e \end{aligned} \tag{9.5}$$

where $F^{**} \sim N\left(0,\ BB^t\right)$.

VARIMAX ROTATION

The most common of the orthogonal rotation methods is varimax rotation. It is based on the assumption that the interpretability of *j*th factor can be measured by the variance of the square of factor loadings, which is maximized. For detail see Maxwell and Lawley [1971].

Let

$$d_j = \prod_{i=1}^{p} 1_{ij}^2 \qquad (j = 1,...., k) \tag{10.1}$$

Then the following expression is maximized

$$\prod_{j=1}^{k} \prod_{i=1}^{p} \left(1_{ij}^2 - p^{-1} d_j\right)^2 \tag{10.2}$$

Such procedure attempts to give the values either close to zero or close to unity. This approach was first proposed by H.F. Kaiser. Later he modified it by normalization to improve his results [Kaiser 1958]. Despite a lot of other methods, varimax is suggested as standard approach.

QUARTIMAX ROTATION

There are several methods for choosing the final factor loading. Nauhous and Wrigley [1954] had shown that if the variables are to be treated as few as possible, then this method is used. As its name suggests, the minimizing of equation (10.1) is to maximize the sum of fourth powers of 1_{ij} . The cross products of the factor loadings L= 1_{ij} are also minimized. If 1_{ij} are the estimated factor loadings, then maximization will take place on

$$\prod_{1 \le e < f \le k} \prod_{i=1}^{p} \left(1_{ie}\ 1_{if}\right) \tag{11.1}$$

L L^t remains unchanged if any orthogonal rotation on *L* is performed. Under orthogonal transformation the diagonal term *L* L^t are invariant. It is therefore

$$\left(\prod_{j=1}^{k} \prod_{i=1}^{p} 1_{ij}^2 \right)^2 = \text{Constant} \tag{11.2}$$

Since, equation (11.1) is equivalent to

$$\prod_{j=1}^{k} \prod_{i=1}^{p} 1_{ij}^4 + 2 \prod_{1 \le e\, f \le k} \prod_{i=1}^{p} \left(1_{ie}\ 1_{if}\right)^2 \tag{11.3}$$

Hence, following expression is maximized.

$$\prod_{j=1}^{k} \prod_{i=1}^{p} 1_{ij}^4 \tag{11.4}$$

EXAMPLE OF FACTOR ANALYSIS

The data for this example is taken from Euromonitor [1979], which shows the percentage of the labour force in nine different types of industry groups in 26 European countries. Where, *AGR* = agriculture, MIN = mining, MAN = manufacturing, PS = power supplies, CON = construction, SER = service industries, FIN = finance, SPS = social and personal services, TC = transport and communications. The data can be visualized in Table 12.1.

Table 12.1: Percentage of People Employed in Various Industries.

Country	AGR	MIN	MAN	PS	CON	SER	FIN	SPS	TC
Belgium	3.3	0.9	27.6	0.9	8.2	19.1	6.2	26.6	7.2
Denmark	9.2	0.1	21.8	0.6	8.3	14.6	6.5	32.2	7.1
France	10.8	0.8	27.5	0.9	8.9	16.8	6.0	22.6	5.7
W. Germany	6.7	1.3	35.8	0.9	7.3	14.4	5.0	22.3	6.1
Ireland	23.2	1.0	20.7	1.3	7.5	16.8	2.8	20.8	6.1
Italy	15.9	0.6	27.6	0.5	10.0	18.1	1.6	20.1	5.7
Luxemburg	7.7	3.1	30.8	0.8	9.2	18.5	4.6	19.2	6.2
Netherland	6.3	0.1	22.5	1.0	9.9	18.0	6.8	28.5	6.8
U.K.	2.7	1.4	30.2	1.4	6.9	16.9	5.7	28.3	6.4
Austria	12.7	1.1	30.2	1.4	9.0	16.8	4.9	16.8	7.0
Finland	13.0	0.4	25.9	1.3	7.4	14.7	5.5	24.3	7.6
Greece	41.4	0.6	17.6	0.6	8.1	11.5	2.4	11.0	6.7
Norway	9.0	0.5	22.4	0 .8	8.6	16.9	4.7	27.6	9.4
Portugal	27.8	0.3	24.5	0.6	8.4	13.3	2.7	16.7	5.7
Spain	22.9	0.8	28.5	0.7	11.5	9.7	8.5	11.8	5.5
Sweden	6.1	0.4	25.9	0.8	7.2	14.4	6.0	32.4	6.8
Switzerland	7.7	0.2	37.8	0.8	9.5	17.5	5.3	15.4	5.7
Turkey	66.8	0.7	7.9	0.1	2.8	5.2	1.1	11.9	3.2
Bulgaria	23.6	1.9	32.3	0.6	7.9	8.0	0.7	18.2	6.7
Czechoslovakia	16.5	2.9	35.5	1.2	8.7	9.2	0.9	17.9	7.0
E. Germany	4.2	2.9	41.2	1.3	7.6	11.2	1.2	22.1	8.4
Hungary	21.7	3.1	29.6	1.9	8.2	9.4	0.9	17.2	8.0
Poland	31.1	2.5	25.7	0.9	8.4	7.5	0.9	16.1	6.9
Romania	34.7	2.1	30.1	0.6	8.7	5.9	1.3	11.7	5.0
U.S.S R	23.7	1.4	25.8	0.6	9.2	6.1	0.5	23.6	9.3
Yugoslavia	48.7	1.5	16.8	1.1	4.9	6.4	11.3	5.3	4.0

Table 12.2: Correlation Matrix in Lower Diagonal Form Calculated from Table 12.1.

AGR	MIN	MAN	PS	CON	SER	FIN	SPS	TC
1.000								
0.036	1.000							
-0.671	0.445	1.000						
-0.400	0.406	0.385	1.000					
-0.538	-0.026	0.495	0.060	1.000				
-0.737	-0.397	0.204	0.202	0.356	1.000			
-0.220	-0.443	-0.156	0.110	0.016	0.366	1.000		
-0.747	-0.281	0.154	0.132	0.158	0.572	0.108	1.000	
-0.565	-0.157	0.351	0.375	0.388	0.188	-0.246	0.568	1.000

Table 12.3: Eigen values and Vectors for the Correlation Matrix in Table 12.2.

Eigen Values	Eigenvector coefficients								
	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7	Y_8	Y_9
3.487	0.524	0.001	-0.348	-0.256	-0.325	-0.379	-0.074	-0.387	-0.367
2.130	0.054	0.618	0.355	0.261	0.051	-0.350	-0.454	-0.222	0.203
1.099	-0.049	0.201	0.151	0.561	-0.153	0.115	0.587	-0.312	-0.378
0.995	0.029	0.064	-0.346	0.393	-0.668	-0.050	-0.052	0.412	0.314
0.543	0.213	-0.164	-0.385	0.295	0.472	-0.283	0.280	-0.220	0.513
0.383	-0.153	0.101	0.289	-0.357	-0.130	-0.615	0.526	0.263	0.124
0.226	0.021	-0.726	0.479	0.256	-0.211	0.229	-0.188	-0.191	0.068
0.137	0.008	0.088	0.126	-0.341	0.356	0.388	0.174	-0.506	0.545
0	-0.806	-0.049	-0.366	0.019	-0.083	-0.238	-0.145	-0.351	-0.072

The data of Table 12.1 is isolated with similar employment distribution in different European countries and is generally aiding the comprehension of the relationship between the countries. The correlation matrix for the 9 variable composed on percentages of employment in European countries

is given in Table 12.2. For the purpose of factor analysis based on principal factor analysis, the eigen values and eigenvectors are computed and are shown in Table 12.3.

There are three eigenvalues greater than unity so the "rule of thumb" proposes that three factors should be carried out. Nevertheless, the fourth eigenvalue is almost equal to unit, therefore we suggest the FA with four factors.

The eigenvectors in Table 12.3 gives the coefficients λ_{ij}, and these are changed into factor loadings. The factor model will be as follows:

$$\begin{aligned} Y_1 &= \quad 0.98\ F_1 + 0.08\ F_2 - 0.05\ F_3 + 0.03\ F_4 \\ Y_2 &= \quad 0.00\ F_1 + 0.90\ F_2 - 0.21\ F_3 + 0.06\ F_4 \\ &\vdots \\ Y_9 &= -0.69\ F_1 + 0.30\ F_2 - 0.39\ F_3 + 0.31\ F_4 \end{aligned}$$

After varimax rotation Kaiser normalization was applied. Now the model deduced to

$$\begin{aligned} Y_1 &= \quad 0.68\ F_1 + 0.27\ F_2 - 0.31\ F_3 + 0.57\ F_4 \\ Y_2 &= \quad 0.22\ F_1 + 0.70\ F_2 - 0.55\ F_3 + 0.31\ F_4 \\ &\vdots \\ Y_9 &= -0.77\ F_1 + 0.23\ F_2 - 0.33\ F_3 - 0.23\ F_4 \end{aligned}$$

We are now in a position to declare the final results for computing factor score.

$$\begin{aligned} \hat{F}_1 &= \quad 0.176\ Y_1 + 0.127\ Y_2 + \ldots\ldots\ldots - 0.430\ Y_9 \\ \hat{F}_2 &= -0.082\ Y_1 + 0.402\ Y_2 + \ldots\ldots\ldots + 0.014\ Y_9 \\ \hat{F}_3 &= -0.122\ Y_1 + 0.203\ Y_2 + \ldots\ldots\ldots - 0.304\ Y_9 \\ \hat{F}_4 &= \quad 0.175\ Y_1 - 0.031\ Y_2 + \ldots\ldots\ldots + 0.088\ Y_9 \end{aligned}$$

CONCLUSIONS

We observe that FA model is identical with a special case of latent structure. There are infinite number of latent classes that are supposed to measure the relationship among the variables. In latent structure analysis, a problem may allow probabilities to be presented in terms of polynomial in the continuous variable. The polynomial may be taken in linear from in each latent variable. Alternatively, if the latent structure is linear in many continuous allowable variables, the model is the FA model. The FA model depends heavily on multivariate normality using maximum likelihood estimation, where the structure became complicated one, while estimating the parameters. We stress mainly on PCA to find initial factors and then FA may be taken up. In this way, the major portion of the calculation (the finding of eigen values and eigenvectors from the correlation matrix) is done by computer using one of the standard statistical packages. One demerit that stands against FA is that it is not as

objective as most statistical methods. Many statisticians are somewhat doubted about its value. Chatfields and Collins [1980] concluded that FA should not be used in most situations.

References

Bartlett, M.S. (**1954**) "A note on the mutiplying factors for various Chi-squared approximation", *JRSS Ser. B*, 16, 296-298.

Chatfield, A.J. and Collins, A.J. (**1980**) "Introduction to Multivariate Analysis", Chapman and Hall, London.

Euromonitor, (**1979**) "European Marketing Data and Statistics", Euromonitor Publication London, pp.76-77.

Hoteling, H. (**1933**) "Analysis of a statistical variable into principal components", *J. Edu. Psychology*, 24, 417-441, 498-520.
"The most predictable criterion", *J. Edu. Psychology*, 26, 139-142.
"Relations between two sets of variates", *Biometrika*, 28, 321-377.

Jackson, J.E. and Moris, R.H. (**1957**) "An application of multivariate quality control to photographic processing", *JASA*, 52, 186-199.

Joreskog, K.G. and Von Thillo, M. (**1971**) "New rapid algorithms for factor analysis by unweighted least squares, generalized least squares and maximum likelihood", Res. Memo 71-75, Educational testing service Princeton, New Jersey.

Kaiser, H.F. (**1958**) "The varimax criterion for analysis rotation in factor analysis", *Psychometrika*, 23, 187-200.

Kaiser, H.F. and Rice, T. (**1974**) "Little Jiffy IV Education", *Psycho. Measurement*, 34, 111-117.

Lawley, D.N. (**1940**) "The estimation of factor loadings by the method of maximum likelihood", *Proc. Roy. Soci.* Edinburgh, 60, 62-82.

Lawley, D.N. and Maxwell, A.E. (**1963**, **1970**) "Factor Analysis as a Statistical Method", Butterworth and Company, London.

Maxwell, A.E. and Lawley, D.E. (**1971**) "Factor Analysis as a Statistical Method", Elsevier, New York.

Mirsky, L. (**1955**) "Introduction to Linear Algebra", Oxford University Press, New York.

Nauhaus, J.O. and Wrigley, C. (**1954**) "The quartimax method an analytical approach to orthogonal simple structure", *Brit. J. Statist. Psy.*, 7, 81-91.

Pearson, K. (**1901**) "On lines an planes of closest fit to a system of points in space", *Phil. Mag.* Ser. 6, 2, 559-572.

Rao, C.R. (**1964**) "The use and interpretation of principal components analysis in applied the research", *Sankhya*, Ser. A, 26, 329-358.