ORTHOGONAL PRINCIPAL PLANES

Peter Filzmoser

department of statistics, probability theory and actuarial mathematics
vienna university of technology, austria

Factor analysis and principal component analysis result in computing a new co-ordinate system, which is usually rotated to obtain a better interpretation of the results. In the present paper, the idea of rotation to simple structure is extended to two dimensions. While the classical definition of simple structure is aimed at rotating (one-dimensional) factors, the extension to a simple structure for two dimensions is based on the rotation of planes. The resulting planes (*principal planes*) reveal a better view of the data than planes spanned by factors from classical rotation and hence allow a more reliable interpretation. The usefulness of the method as well as the effectiveness of a proposed algorithm are demonstrated by simulation experiments and an example.

Key words: principal component analysis, factor analysis, orthogonal rotation, simple structure.

1. Introduction

The examination of high-dimensional data is getting a more and more important task in applied multivariate analysis. There are different possibilities for the investigation of large data sets. One way is to reduce the dimension (following any useful criterion). Another way is to search for low-dimensional projections of the full-dimensional data, as done by *projection pursuit* (see e.g. Huber, 1985; Friedman, 1987). In both cases the results should provide the most revealing view of the data.

For data sets including not necessarily a cluster structure (e.g. data sets composed of economical, ecological, social, environmental, health data), a method is required which shows connections between the variables and the relations to the objects in the best way. The results have to permit an extensive interpretation of the data. For this reason the data should be shown graphically in planes since two-dimensional representations are easy to survey. The extracted planes should contain a maximum of information. Baaske (1988) named such planes *principal planes*, and he tried to find triplets of variables spanning a plane which should represent the other variables as good as possible.

The present paper gives another approach for obtaining principal planes. At the basis of a decomposition of the (standardized) data matrix by factor analysis or principal component analysis, the matrix of loadings is rotated to a simple structure for planes. This is a generalization to higher dimensions of the classical simple structure introduced by Thurstone (1944). One could extend these thoughts also to higher dimensions; we would then consider *principal spaces*.

Classical rotation methods extract (one-dimensional) factors which should characterize all variables in a good fashion, and at the same time different factors should include different variables. This provides the factors to be interpreted as non observable quantities, where each factor characterizes other properties. The results are mostly presented by plots where each factor is drawn against another one. Unfortunately, these planes are not constructed to be most meaningful, since by ideal simple structure the variables are close to the axes. The contents of information of a resulting plane could be increased by directly extracting planes instead of vectors (factors).

For rotation to principal planes, two-dimensional factors (pairs of factors spanning a plane) with the same properties as formulated above are to be found. All variables should be well presented by the principal planes, and each principal plane should characterize other variables. This construction enables a configuration, where the variables are close to the principal plane (spanned by two factors), and not just close to the axes. As a consequence, in general more variables are well presented by the resulting plane, and therefore the interpretation of the results can be more extensive and should be more reliable.

This paper is organized as follows: In section 2 basic considerations about simple structure for planes are given. Moreover, the classical varimax-criterion (Kaiser, 1958), in the following abbreviated by VMAX, is extended to a varimax-criterion for planes (VMAX2). Section 3 is concerned with an algorithm for VMAX2. The rotation is based on an iterative procedure. The criterion is improved step-by-step, until no essential change of the solution is visible. Since no explicit solution of the optimization criterion exists, an approximation has to be found in each step of the iteration. The performance of the algorithm is demonstrated by simulation experiments in section 4. Section 5 shows the application of the procedure to a real data set.

2. Method

Basis for rotation to principal planes is the unrotated matrix of loadings obtained by factor analysis or principal component analysis. Let $\mathbf{\Lambda}$ denote the given $(p \times k)$ orthogonal pattern, and let $\boldsymbol{f} = (f_1, \ldots, f_k)^\top$ be the standardized factors arising by linear transformation of the original standardized variables $\boldsymbol{y} = (y_1, \ldots, y_p)^\top$ with the coefficients $\mathbf{\Lambda}$, i.e. $\boldsymbol{y} = \mathbf{\Lambda}\boldsymbol{f} + \boldsymbol{e}$. In more detail, the (i,j)-th element of $\mathbf{\Lambda}$, λ_{ij}, represents the connection between variable y_i and factor f_j, and this relation is used for obtaining classical simple structure: High loadings for a variable on one factor should occur, while at the same time the loadings on the other factors should be low. For the well known and frequently used *varimax*-criterion (Kaiser, 1958), the simple structure is realized by maximizing the sum over the variances of the squared (standardized) factor loadings for each factor, i.e.

$$\frac{1}{p}\sum_{j=1}^{k}\sum_{i=1}^{p}\Big(\frac{\widetilde{\lambda}_{ij}}{\kappa_i}\Big)^4 - \frac{1}{p^2}\sum_{j=1}^{k}\Big[\sum_{i=1}^{p}\Big(\frac{\widetilde{\lambda}_{ij}}{\kappa_i}\Big)^2\Big]^2 = \max \ . \tag{1}$$

$\widetilde{\lambda}_{ij}$ is an element of the rotated matrix of loadings, and $\kappa_i^2 = \sum_{j=1}^{k}\lambda_{ij}^2$ $(i = 1, \ldots, p)$ is called i-th communality. It describes the proportion of variance of the i-th standardized variable explained by all k factors. The communalities are used for standardization to avoid large influence of variables with high communalities.

Simple structure for planes also uses the relationship between variables and planes, expressed by the multiple correlation. Explicitly, for a variable y_i $(i = 1, \ldots, p)$ and a plane spanned by two different factors f_a and f_b $(a, b = 1, \ldots, k;\ a \neq b)$, the multiple correlation between y_i and $\boldsymbol{f} = (f_a, f_b)^\top$ is defined by

$$\rho_{y_i;\boldsymbol{f}} = (\boldsymbol{\rho}_{\boldsymbol{f}y_i}^\top \boldsymbol{\rho}_{\boldsymbol{f}\boldsymbol{f}}^{-1} \boldsymbol{\rho}_{\boldsymbol{f}y_i})^{1/2} \tag{2}$$

(see e.g. Mardia et al., 1979) where $\boldsymbol{\rho}_{\boldsymbol{f}_{y_i}}$ is a (2×1)-matrix with the correlations between factors and the variable at hand, and $\boldsymbol{\rho}_{\boldsymbol{f}\boldsymbol{f}}$ is the correlation matrix of the factors f_a and f_b. If the factors are supposed to be orthogonal, $\boldsymbol{\rho}_{\boldsymbol{f}\boldsymbol{f}}$ (and therefore $\boldsymbol{\rho}_{\boldsymbol{f}\boldsymbol{f}}^{-1}$) is the identity. The correlations between y_i and the factors in the orthogonal case are exactly the loadings, so the above equation reduces to

$$\rho_{y_i;\boldsymbol{f}} = \sqrt{\lambda_{ia}^2 + \lambda_{ib}^2}\ . \tag{3}$$

To simplify the notation, the multiple correlation between y_i and $\boldsymbol{f} = (f_a, f_b)^\top$ is denoted by $\rho_{i;a,b}$.

A next point for developing a rotation criterion for planes is to define more precisely how the planes are constructed. The aim is to obtain planes spanned by different and disjoint pairs of factors. If the number of factors, k, is even, $q = k/2$ pairs of factors are forming the planes. For an odd number k, one factor is omitted and the results are $q = (k-1)/2$ planes. Each set of pairs can be defined by simply ordering the k factors in a particular way (calling the first two factors the first pair, the second two the second pair, etc.). Hence, there are $k!$ such sets of pairs. However, among these $k!$ sets of pairs many are essentially the same: First of all, all pairs appear twice (as (a, b) and (b, a)), hence, to correct for this, the number $k!$ has to be divided by 2 for each pair, hence by 2^q for the q pairs. Furthermore, sets of q pairs that only differ with respect to the order in which the pairs are given are equal, hence the result has to be divided by $q!$, the number of permutations in which the pairs can occur. This gives a total number of combinations of different planes with disjoint factors of

$$E = \frac{k!}{2^q q!} \ . \tag{4}$$

Let s_l $(l = 1, \ldots, E)$ denote the indices of the factors for one particular combination of planes, and let $S = \{s_l \mid l = 1, \ldots, E\}$ be the set of these combinations. E.g. for $k = 6$ factors, one element of S might be the set $\{(1, 2), (3, 4), (5, 6)\}$.

A rotation criterion for simple structure for planes has to find the optimum over all E combinations of planes, i.e. the best result over all factor combinations $s \in S$. The simplicity of the structure is defined at the basis of the relation between variables and planes, given by (2) or, for orthogonal factors, by (3). Similar to the classical case, this value should be high for a variable at one plane and at the same time low at the other planes.

With this knowledge the ideas of VMAX (1) can be easily extended to a varimax-criterion for planes. The simple structure for planes can be realized by considering the variance of the squared "loadings on the planes" which are defined by (2). If the planes are spanned by the factors $\{f_a, f_b\}$ $(\{(a, b)\} \in s)$, this variance is

$$s^2_{a,b} = \frac{1}{p} \sum_{i=1}^{p} \left(\widetilde{\rho}^2_{i;a,b} \right)^2 - \frac{1}{p^2} \left[\sum_{i=1}^{p} \widetilde{\rho}^2_{i;a,b} \right]^2 \tag{5}$$

or, since the factors are supposed to be orthogonal,

$$s^2_{a,b} = \frac{1}{p} \sum_{i=1}^{p} \left(\widetilde{\lambda}^2_{ia} + \widetilde{\lambda}^2_{ib} \right)^2 - \frac{1}{p^2} \left[\sum_{i=1}^{p} \left(\widetilde{\lambda}^2_{ia} + \widetilde{\lambda}^2_{ib} \right) \right]^2 \tag{6}$$

where $\widetilde{\rho}^2_{i;a,b} = \widetilde{\lambda}^2_{ia} + \widetilde{\lambda}^2_{ib}$. Equation (6) describes the variance of the squared multiple correlations (SMC) of the variables y_i with the plane $\{f_a, f_b\}$. In analogy to VMAX, the variances given by (6) have to be summarized over all planes of a combination s, and the loadings are to be divided by the corresponding communalities. The resulting expression has to be maximized by an orthogonal transformation. Finally, the maximum over all different combinations $s \in S$ has to be found. Expressed by a formula, the *varimax-criterion for planes* is defined by

$$VMAX2 = \max_{s \in S} \left\{ \sum_{\{(a,b)\} \in s} \left[p \sum_{i=1}^{p} \left(\frac{\widetilde{\rho}_{i;a,b}}{\kappa_i} \right)^4 - \left(\sum_{i=1}^{p} \left(\frac{\widetilde{\rho}_{i;a,b}}{\kappa_i} \right)^2 \right)^2 \right] = \max \right\} . \tag{7}$$

The results are principal planes spanned by the factors $\{f_a, f_b\}$ ($\{(a,b)\} \in s$). Note that criterion (7) can easily be modified to obtain a two-dimensional extension of the *quartimax-criterion* (Carroll, 1953), the *quartimax-criterion for planes:*

$$QMAX2 = \max_{s \in S} \left\{ \sum_{\{(a,b)\} \in s} \sum_{i=1}^{p} \left(\frac{\widetilde{\rho}_{i;a,b}}{\kappa_i} \right)^4 = \max \right\} . \tag{8}$$

3. Algorithm

A numerical solution of the varimax-criterion for planes can be found by an iterative process. In each iteration two different planes spanned by the factors $\{f_a, f_b\}$ and $\{f_c, f_d\}$ ($\{(a,b),(c,d)\} \in s, a \neq c$), respectively, are considered. In this 4-dimensional space the varimax-criterion for the two planes is defined by

$$\begin{aligned} VMAX2_{a,b;c,d} &= p \sum_{i=1}^{p} \left[\left(\frac{\widetilde{\rho}_{i;a,b}}{\kappa_i} \right)^4 + \left(\frac{\widetilde{\rho}_{i;c,d}}{\kappa_i} \right)^4 \right] \\ & \quad - \left[\sum_{i=1}^{p} \left(\frac{\widetilde{\rho}_{i;a,b}}{\kappa_i} \right)^2 \right]^2 - \left[\sum_{i=1}^{p} \left(\frac{\widetilde{\rho}_{i;c,d}}{\kappa_i} \right)^2 \right]^2 . \end{aligned} \tag{9}$$

Since an orthogonal rotation in 4 dimensions would be rather complicated, the rotation is performed in 4 steps in the planes $\{f_a, f_c\}$, $\{f_b, f_d\}$, $\{f_a, f_d\}$ and $\{f_b, f_c\}$. Steps 1-4 are repeated until (9) cannot be further increased. If this is done, the four steps are applied to another two planes of the combination s, and so on. The entire process is started again until convergence, this means until the varimax-criterion for planes (7) cannot be further improved. In order to avoid that the algorithm converges to a local maximum, several orthogonal random starts of the whole procedure have to be done (see also Gebhardt, 1968; ten Berge, 1984). More information about the practical performance of this procedure is given in the next section.

Let us consider the optimization in one particular plane, say, in the plane spanned by the factors $\{f_a, f_c\}$, in more detail. The rotated loadings are computed by the orthogonal transformation

$$\widetilde{\lambda}_{ia} = \lambda_{ia}\cos\vartheta + \lambda_{ic}\sin\vartheta \tag{10}$$
$$\widetilde{\lambda}_{ic} = -\lambda_{ia}\sin\vartheta + \lambda_{ic}\cos\vartheta \tag{11}$$
$$\widetilde{\lambda}_{ij} = \lambda_{ij} \qquad \text{for} \quad j \neq a, c \tag{12}$$

and $i = 1, \ldots, p$. The rotation angle in this plane is ϑ. Insertion of the rotated loadings into (9) gives

$$\begin{aligned} VMAX2_{\underline{a},b;\underline{c},d} = {} & p\sum_{i=1}^{p}\left[\frac{[(\lambda_{ia}\cos\vartheta + \lambda_{ic}\sin\vartheta)^2 + \lambda_{ib}^2]^2}{\kappa_i^4}\right. \\ & \left. + \frac{[(-\lambda_{ia}\sin\vartheta + \lambda_{ic}\cos\vartheta)^2 + \lambda_{id}^2]^2}{\kappa_i^4}\right] \\ & - \left[\sum_{i=1}^{p}\frac{(\lambda_{ia}\cos\vartheta + \lambda_{ic}\sin\vartheta)^2 + \lambda_{ib}^2}{\kappa_i^2}\right]^2 \\ & - \left[\sum_{i=1}^{p}\frac{(-\lambda_{ia}\sin\vartheta + \lambda_{ic}\cos\vartheta)^2 + \lambda_{id}^2}{\kappa_i^2}\right]^2 . \end{aligned} \tag{13}$$

Two indices are underlined in the above formula to indicate, in which plane the rotation is performed. For maximizing (13), the first derivative with respect to ϑ is calculated and the result is set equal to zero. Since this derivative has no explicit solution for the parameter ϑ, an approximation can be found for example by the *regula-falsi method* (see e.g. Golub and Ortega, 1993) described below.

Equation (13) may be rewritten in the form

$$f(\vartheta) = c_1\sin(4\vartheta) + c_2\cos(4\vartheta) + c_3\sin(2\vartheta) + c_4\cos(2\vartheta) + c_5 \tag{14}$$

where c_1 to c_5 are constants depending on the loadings. In general, the periodicity of the function f is π. This means that for finding the maximum one is interested in a rotation of the factors within the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$. The maximum can be found approximately by performing the following steps:

Step 1: Select P fixed regularly distributed values $\vartheta_1, \ldots, \vartheta_P$ within the interval $[-\frac{\pi}{2}, \frac{\pi}{2}]$ and compute the function values $f(\vartheta_1), \ldots, f(\vartheta_P)$.

Step 2: Search for the maximum $f(\vartheta_m)$ of the P function values and take the neighbors of ϑ_m, which are ϑ_{m-1} and ϑ_{m+1}. For $m = 1$ the neighbors are $-\frac{\pi}{2}$ and ϑ_2, for $m = P$ take ϑ_{P-1} and $\frac{\pi}{2}$.

Step 3: Compute the values $f'(\vartheta_{m-1})$ and $f'(\vartheta_{m+1})$, which are the starting values for the *regula falsi* method. Since f is continuous, $\text{sgn} f'(\vartheta_{m-1}) \neq \text{sgn} f'(\vartheta_{m+1})$ if P is large enough.

Step 4: Compute the zero point ϑ_z of the connection line between $f'(\vartheta_{m-1})$ and $f'(\vartheta_{m+1})$ by

$$\vartheta_z = \vartheta_{m-1} - f'(\vartheta_{m-1}) \frac{\vartheta_{m+1} - \vartheta_{m-1}}{f'(\vartheta_{m+1}) - f'(\vartheta_{m-1})} \; . \tag{15}$$

Step 5: If $f'(\vartheta_z) = 0$ (or ≈ 0), a real zero of f' within the interval $(-\frac{\pi}{2}, \frac{\pi}{2})$ has been found. In the other case $\text{sgn} f'(\vartheta_z) \neq \text{sgn} f'(\vartheta_{m-1})$ or $\text{sgn} f'(\vartheta_z) \neq \text{sgn} f'(\vartheta_{m+1})$, and the procedure can be started with the corresponding new interval from *step 4*.

Figure 1 illustrates the proposed procedure. For $P = 10$ fixed points (which would be too small in practice) the function values are computed. After performing the regula falsi method, the first approximation of the maximum of f at ϑ_z is obtained. The second approximation is already very close to the real maximum.

Increasing P implies a longer computation time for one iteration of the regula-falsi method. On the other hand, the number of iterations in general decreases since the maximum $f(\vartheta_m)$ for larger P (*step 2*) will in general be closer to the final maximum. It turned out in practice that a choice of $P = 50$ is a good compromise.

A computer program with the proposed algorithm in the language GAUSS can be obtained from the author.

Figure 1 is inserted about here.

---

4. Simulation

The proposed rotation method is tested in the following by simulation experiments. A matrix of loadings with $p = 100$ variables and $k = 4$ factors is generated randomly. The first 50 variables have two nonzero loadings at the first two factors, the last 50 variables have two nonzero loadings at the last two factors, i.e. the target pattern has complexity 2. The nonzero loadings are uniformly distributed in the interval $[-1, 1]$. For each variable, the loadings are standardized by their communality.

The simulated pattern is scrambled in each of 50 replications by random orthogonal rotations, and the rotation methods quartimax and varimax (classically and for planes) are applied. For each analysis the iterative process was considered to have converged if the relative difference of the objective function of two consecutive cycles was below 10^{-6}. The convergence was reached in each of the following simulations already after a few iterations.

The resulting rotated loadings are compared with the original ones by permuting and reflecting the columns to obtain optimal congruence with the truly simple pattern. For each row of the loading matrices, the Euclidean norm of the difference to the corresponding row of the truly simple pattern was computed. The mean of these values has been taken as a measure of goodness of recovering the original pattern. Expressed by a formula, this measure which is called *congruence index* is defined by

$$C = \frac{1}{p} \sum_{i=1}^{p} \left(\sum_{j=1}^{k} \left(\lambda_{ij} - \sum_{l=1}^{k} a_{il} t_{lj} \right)^2 \right)^{1/2} \tag{16}$$

where λ_{ij} is the (i, j)-th element of the target pattern, a_{il} is the (i, l)-th element of the randomly rotated pattern, and t_{lj} is the (l, j)-th element of the best transformation matrix obtained by each rotation method at hand.

The result of this study is shown in Figure 2. The horizontal axis shows the number of replications of the simulation, the vertical axis represents the congruence index (16) as the measure of deviation from the original pattern. The results of QMAX are labeled by ○, those of QMAX2 by $+$, VMAX by $\triangle$, and VMAX2 by $\times$.

Since no noise has been added to the target pattern, and since the pattern is ideally plane-wise clustered, a congruence index of 0 would indicate the best solution (global minimum). Figure 2 shows that the results for the classical rotation methods QMAX and VMAX do not change by orthogonally rotating the target. Moreover, these methods lead to the same results (up to the given precision of the optimization procedure of the algorithm). QMAX2 and VMAX2 are strongly depending on orthogonal random starts. About half of the solutions of the rotation methods for planes are better than the classical methods, some are much better. It is interesting that VMAX2 and QMAX2 give about the same result if the solution comes closer to the global optimum; otherwise VMAX2 is slightly better than QMAX2.

Figure 2 is inserted about here.

The second simulation experiment uses the same target pattern as before. This pattern is contaminated in each of 50 simulations with random noise, the elements of which are distributed according to $N(0, 1/10)$. The simulated patterns are scrambled by random orthogonal rotations and the same rotation methods as in the previous experiment are used for pattern recovery. Since QMAX2 and VMAX2 are sensitive to local optima, 10 randomized starting positions are chosen by orthogonal random rotations of each simulated pattern, and the best solution is taken as the best approximation of the global optimum. The result of this study is presented in Figure 3 where the number of replications of the simulation is drawn against the congruence index (16). A big difference between the classical methods and rotation methods for planes is visible. Just for 2 out of 50 simulations the classical methods lead to better results. QMAX gives in general a lower value of the congruence index than VMAX. Like in the previous experiment, the results of QMAX2 and VMAX2 are very similar. It is interesting to compute the average of the congruence indices over all 50 simulations. The value for QMAX is 0.64, VMAX gives 0.66, QMAX2 and VMAX2 are much lower with a value of 0.24.

Figure 3 is inserted about here.

A third simulation experiment is also based on a matrix of loadings of complexity two (i.e. two nonzero loadings on at most two factors), but the rows of this target pattern are not ideally plane-wise clustered. In more detail, the target pattern is of the same size as before ($p = 100$ variables and $k = 4$ factors), the first 40 variables have two nonzero loadings at the first two factors, the next 8 variables have two nonzero loadings at factors 1 and 3, the next 7 variables at factors 1 and 4, the next 6 variables at factors 2 and 3, and the remaining 39 variables have two nonzero loadings at the last two factors. The nonzero loadings are constructed in the same way as in the first simulation experiment. The pattern is contaminated in each of 50 simulations with randomly distributed noise (according to $N(0, 1/10)$), and the resulting loadings are scrambled by random orthogonal rotations. For QMAX2 and VMAX2 the best solutions out of 20 randomized starting positions are taken. (The results of QMAX and VMAX are varying only marginally for different orthogonal random starts.) Figure 4 presents the resulting congruence index (vertical axis) for each simulated pattern (horizontal axis). Although the solutions of the classical methods and the methods for rotation to principal planes are much closer now, the results of QMAX2 and VMAX2 are clearly better. The congruence index is lower than 0.2 for about 75% of the results of QMAX2 and VMAX2, but just for 15% of VMAX and not even for 10% of QMAX. Like in the previous experiments, the results for QMAX2 and VMAX2 are very similar. The average congruence indices are: 0.30 for QMAX, 0.28 for VMAX, and 0.19 for both QMAX2 and VMAX2. When taking the best solutions out of 10 instead of 20 orthogonal random starts, the average congruence increases to 0.23 for QMAX2 and VMAX2.

Figure 4 is inserted about here.

The simulation was done on a Pentium PC with 120 MHz. Considering the first simulation experiment, the mean computation time over all 50 replications for QMAX, QMAX2, VMAX and VMAX2 is reported in Table 1. The values for QMAX2 and VMAX2 correspond to the time needed for one out of E different plane combinations (see (4)). Table 1 also summarizes the mean computation time for the rotation of 6, 8, 10 and 12 factors. VMAX2 needs about twice as much computation time as QMAX2, and there is a big difference between the time for classical rotation and rotation to planes (almost factor 100 for VMAX2). Since the values for QMAX2 and VMAX2 have to be multiplied by the number E of different plane combinations, the proposed algorithm becomes impractical for a larger number of factors.

Table 1 is inserted about here.

The previous simulation experiments have shown how well QMAX2 and VMAX2 perform with matrices of loadings with complexity two. However, what would happen when rotating matrices of loadings with complexity one? As an example we consider the pattern shown in the left part of Table 2 ($p = 8$ and $k = 4$). The middle part of Table 2 shows the result for both QMAX and VMAX rotation. The true pattern is perfectly recovered by these methods. Finally, the right part of Table 2 shows the general result for QMAX2 and VMAX2, where the first two factors span the first principal plane and the last two factors the second principal plane. a, b, c

and d are real numbers within the interval -1 and 1 with the restrictions $a^2 + b^2 = c^2 + d^2 = 1$. This means that an infinite number of solutions for QMAX2 and VMAX2 is possible, each obtained by orthogonal rotation of the principal planes. The reason for this phenomenon can be found by considering the rotation algorithm (section 3). The pairwise rotations are only performed for pairs of factors of different planes; rotations within a principal plane are not necessary because the resulting plane will not give any new information.

Table 2 is inserted about here.

---

5. Example

In this section a data set is considered which origins from a cooperation between the research institutes *Studia* (Schlierbach, Austria) and *Albtum* (Weihenstephan, Germany). More than 800 variables from different fields were measured in the 96 Bavarian districts and cities, mainly in the year 1987. Detailed information can be found in a technical report (Studia and Albtum, 1993).

For reason of clarity nine variables are selected: rate of unemployment (1), median salary (2), percentage of the population in the age 6-15 years (3), percentage of the population older that 60 years (4), percentage of divorced persons (5), percentage of commuters (6), average number of persons per household (7), percentage of larceny delicts (8), and percentage of stomach cancer as cause of death (9).

The choice of four factors gives a proportion of explained variation of about 93%. Principal common factor analysis without rotation of the factors results in the loadings presented in the left part of Table 3. The loadings after classical varimax rotation are shown in the right part of Table 3.

Table 3 is inserted about here.

The unrotated matrix of loadings is basis for varimax rotation to principal planes. The best result (maximum of the objective function) of the loadings for 10 orthogonal random starts is shown in the left part of Table 4. The first principal plane (PPL1) is spanned by the factors F1 and F2, the second principal plane (PPL2) by F3 and F4. The right part of Table 4 summarizes the contribution of each variable to the VMAX2 criterion and the sum over all single contributions for both, the resulting principal planes from VMAX2 rotation and the planes spanned by F1-F3 and F2-F4 resulting from VMAX rotation. The latter factor combinations give the maximum sum of the single contributions, and hence these planes allow the best two-dimensional representation of the VMAX loadings.

The contribution of a variable to the VMAX2 criterion is high if this variable is well presented by only one plane. Variable 1 hence has high contribution to PPL2 but low (negative) contribution to PPL1 where the loadings are close to zero. Variables 2 and 6 are not well presented by the principal planes. All other variables (except variable 4) have higher contributions with VMAX2 rotation than with VMAX rotation. Also the sum over the single contributions is higher for VMAX2 which means that VMAX2 rotation separates the variables into planes in a better way.

Table 4 is inserted about here.

Figure 5 shows the resulting planes from VMAX and VMAX2 rotation. Figure 5a and 5b represent the planes spanned by F1-F3 and F2-F4 from VMAX, respectively (see Table 3), and Figure 5c and 5d show PPL1 and PPL2 from VMAX2, respectively (see Table 4). The first planes from VMAX and VMAX2 (Figure 5a and 5c) show some similarity, whereas the second planes (Figure 5b and 5d) reveal big differences. Variables 1, 2, 4, 6 and 9 are well presented by PPL2 (large distance to the center), but not well presented by PPL1 (close to the center). Similarly, variables 3, 5, 7 and 8 are very well presented by PPL1, but at the same time they are close to the center in PPL2. The variables in the second plane from VMAX rotation are mainly arranged along the direction of factor 2, whereas in PPL2 the variables are spread over the whole plane. PPL2 hence makes the connections between variables 1, 2, 4, 6 and 9 visible and allows interpretations for these relationships. In this example, however, one still has to be careful with interpretations, since connections can also be caused by other variables which have not been analyzed. Therefore it is advisable to take the whole data set for a detailed investigation.

Figure 5 is inserted about here.

---

6. Summary

The idea of simple structure for planes is an extension of the classical simple structure. In the classical case the rotated factors should be as separated as possible, i.e. each row of the rotated loadings matrix should contain only a few high loadings, the remaining loadings should be low. Simple structure for planes is constructed in that way to obtain planes which are as separated as possible: For each variable there should be only a few high “loadings on planes”, which are expressed by the multiple correlation (2), otherwise low values are desired. The advantage of this extension is that the resulting principal planes are aimed to be disjoint. Hence, by just presenting the principal planes, most information of the matrix of loadings is given. Since graphical presentations of the loadings by planes are easy to survey, especially in connection with factor scores like in a biplot (Gabriel, 1971), this is a desired property. For classical rotation methods, usually all pairs of factors have to be shown for a detailed (2-dimensional) presentation of the loadings.

Another property which results from the construction of two-dimensional simple structure is that for each principal plane a maximum number of variables is not near the origin. Hence, these variables are well presented by the plane and the interpretation of relationships between these variables are more reliable.

Analytic solutions of the rotation methods QMAX2 and VMAX2 can be found by the proposed algorithm. In each step the objective function increases monotonically. However, the algorithm often leads to local optimality, and the global optimum has to be approximated by applying the algorithm with several orthogonal random starts. The simulation examples have shown that 10 to 20 orthogonal random starts are sufficient for recovering complicated target patterns. However, for applications it is more advisable not to fix the number of randomizations but to continue on searching so long as additional tries have a good chance of improving the solutions already in store.

The proposed algorithm becomes impractical for a larger number of factors. Since the mean computation times given in Table 1 have to be multiplied by the number of different plane combinations, VMAX2 rotation for 10 factors would need about 100.000 seconds for one orthogonal random start at the computer used for this study. However, it turned out that in most cases different plane combinations did not result in new solutions. Further investigation has to be done in examining how many combinations are sufficient for finding the best result.

In section 2 it was suggested that for an odd number of factors one factor should be omitted for principal plane rotation. Since the number of factors to be retained in factor analysis is often a hard decision, it could be more advisable to retain always an even number of factors.

The simulation has shown that QMAX2 and VMAX2 often give the same result. In most applications one would prefer VMAX2 because it turned out that QMAX2 leads to a main first plane, as the classical quartimax-criterion has the tendency to converge over a main first vector (Kaiser, 1956, 1958).

If the complexity of a pattern is one (or close to one), the rotated pattern obtained by QMAX2 and VMAX2 can be transformed to the pattern obtained by QMAX and VMAX by an orthogonal rotation (see Table 2). The solution from QMAX2 and VMAX2 can be seen as an alternative representation of the original pattern. Hence, in that case QMAX2 and VMAX2 are not competitors of the classical rotation criteria.

It is possible to extend the ideas and formulas of simple structure for planes to simple structure for higher dimensions. The results which we call *principal spaces* have the same properties as principal planes.

References

Baaske, W. (1988). Principal planes: visual structure analysis by cluster biplot projection pursuit in a multinational survey. *Applied Mathematics and Computation*, *26*, 303-314.

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

Friedman, J.H. (1987). Exploratory projection pursuit. *Journal of the American Statistical Association*, *82(397)*, 249-266.

Gabriel, K.R. (1971). The biplot graphic display of matrices with application to principal component analysis. *Biometrika, 58(3)*, 453-467.

Gebhardt, F. (1968). A counterexample to two-dimensional varimax-rotation. *Psychometrika*, *33*, 35-36.

Golub, G.H., and Ortega, J.M. (1993). *Scientific computing, An introduction with parallel computing*. New York: Academic Press.

Huber, P.J. (1985). Projection Pursuit. *The Annals of Statistics*, *13/2*, 435-475.

Kaiser, H.F. (1956). Note on Carroll's analytic simple structure. *Psychometrika*, *21*, 89-92.

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

Mardia, K.V., Kent J.T., and Bibby, J.M. (1979). *Multivariate Analysis*. London: Academic Press.

Studia and Albtum (1993). Externe Leistungen der bäuerlichen Landwirtschaft in Bayern. Technical report. STUDIA - Research Group for International Analyses, Schlierbach; ALBTUM - Professorship for Applied Agricultural Business Economics, Freising-Weihenstephan, Munich University of Technology. In German.

ten Berge, J.M.F. (1984). A joint treatment of varimax rotation and the problem of diagonalizing symmetric matrices simultaneously in the least-squares sense. *Psychometrika*, *48*, 347-358.

Thurstone, L.L. (1944). Second-order factors. *Psychometrika*, *9*, 71-100.

Figure 1

Regula-falsi algorithm for maximizing the VMAX2 criterion in one plane. f is the function defined by the criterion, and f' is the derivative of f.

Figure 2

Simulation experiment with an ideally plane-wise clustered target pattern of complexity two. The horizontal axis shows the number of replications (orthogonal random starts), and the vertical axis represents the congruence index (16) as a measure of deviation from the original pattern. Results are shown for QMAX (○), QMAX2(+), VMAX (△) and VMAX2 (×).

Figure 3

Simulation with the same target pattern as in Figure 2 with additional random noise. The horizontal axis shows the number of simulated patterns, and the vertical axis represents the congruence index. Results are shown for QMAX (○), QMAX2(+), VMAX (△) and VMAX2 (×). For QMAX2 and VMAX2 the best results out of 10 orthogonal random starts are shown.

Figure 4

Simulation with a target pattern of complexity two which is not ideally plane-wise clustered. The horizontal axis shows the number of simulated patterns (target pattern plus normally distributed noise), and the vertical axis represents the deviation. Results are shown for QMAX (○), QMAX2(+), VMAX (△) and VMAX2 (×). For QMAX2 and VMAX2 the best results out of 20 orthogonal random starts are shown.

Figure 5

Factor analysis results of the Bavarian data set: (a) plane F1-F3 from VMAX, (b) plane F2-F4 from VMAX, (c) plane F1-F2 from VMAX2 (first principal plane), (d) plane F3-F4 from VMAX2 (second principal plane).

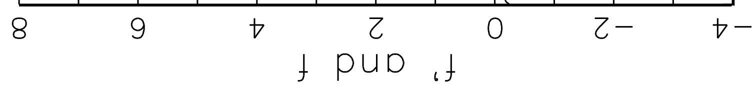

FIGURE 1

Regula-falsi algorithm for maximizing the VMAX2 criterion in one plane.
f is the function defined by the criterion, and f' is the derivative of f.

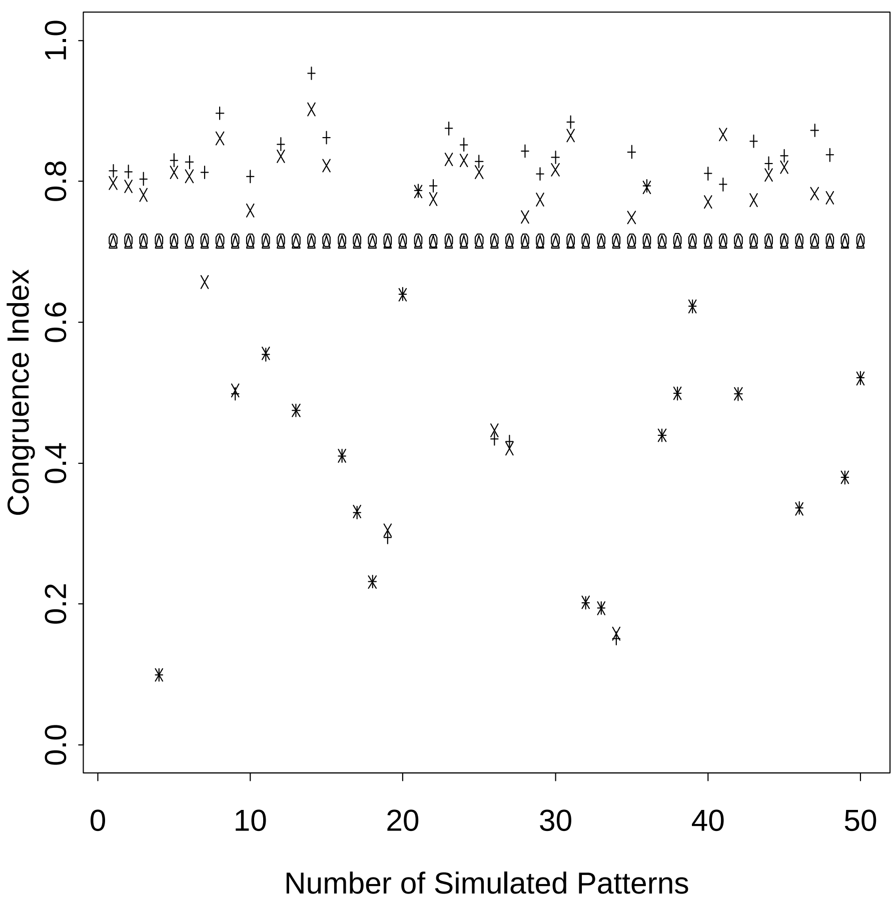

FIGURE 2

Simulation experiment with an ideally plane-wise clustered target pattern of complexity two. The horizontal axis shows the number of replications (orthogonal random starts), and the vertical axis represents the congruence index (16) as a measure of deviation from the original pattern. Results are shown for QMAX (○), QMAX2(+), VMAX (△) and VMAX2 (×).

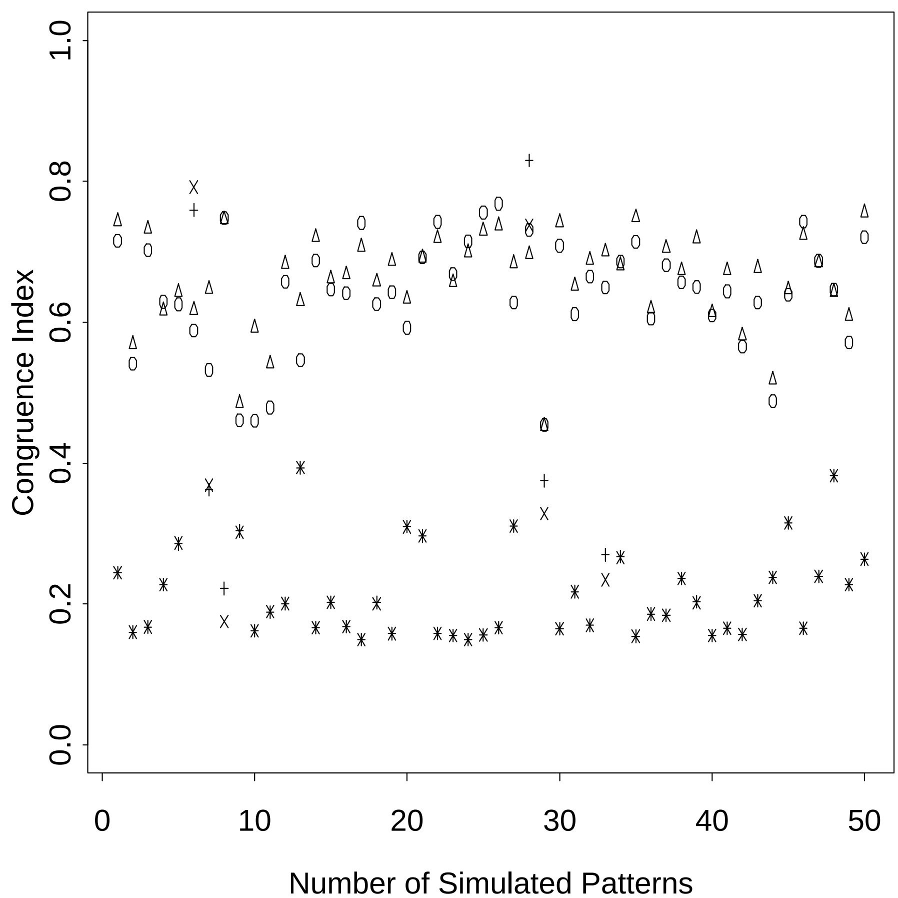

Figure 3

Simulation with the same target pattern as in Figure 2 with additional random noise. The horizontal axis shows the number of simulated patterns, and the vertical axis represents the congruence index. Results are shown for QMAX (○), QMAX2(+), VMAX (△) and VMAX2 (×). For QMAX2 and VMAX2 the best results out of 10 orthogonal random starts are shown.

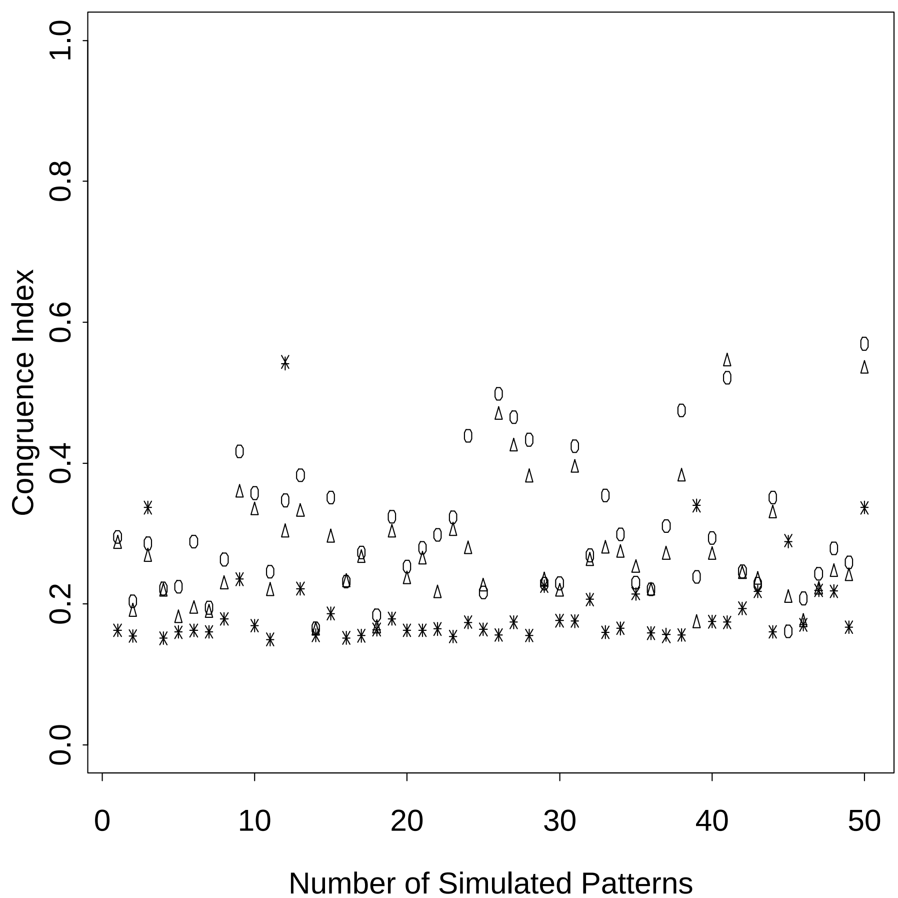

Figure 4
Simulation with a target pattern of complexity two which is not ideally plane-wise clustered. The horizontal axis shows the number of simulated patterns (target pattern plus normally distributed noise), and the vertical axis represents the deviation. Results are shown for QMAX (○), QMAX2(+), VMAX (△) and VMAX2 (×). For QMAX2 and VMAX2 the best results out of 20 orthogonal random starts are shown.

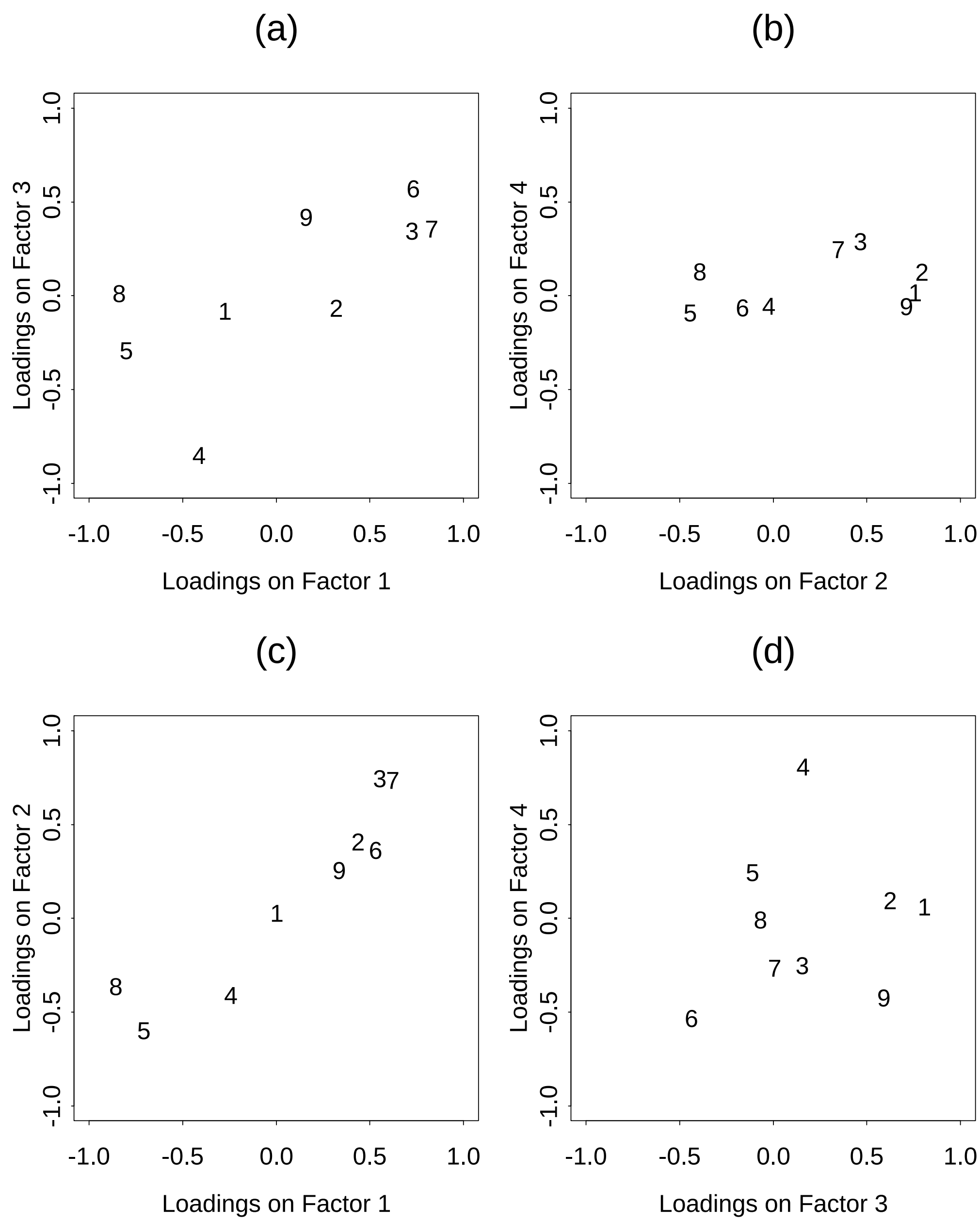

FIGURE 5

Factor analysis results of the Bavarian data set: (a) plane F1-F3 from VMAX, (b) plane F2-F4 from VMAX, (c) plane F1-F2 from VMAX2 (first principal plane), (d) plane F3-F4 from VMAX2 (second principal plane).

TABLE 1
Mean computation time for one out of E combinations of planes (in seconds), for 2 planes (4 factors) to 6 planes (12 factors) at the basis of 50 replications of a simulation.

Rotation Method	Number of Factors 4	6	8	10	12
QMAX	0.16	0.44	0.94	2.72	6.81
QMAX2	2.39	19.23	34.88	46.67	96.70
VMAX	0.11	0.33	0.82	1.65	2.47
VMAX2	4.70	34.44	76.84	105.66	199.49

Table 2

Example of a matrix of loadings with complexity one: True pattern, QMAX or VMAX rotation, and QMAX2 or VMAX2 rotation ($a^2 + b^2 = c^2 + d^2 = 1$).

	True Pattern				QMAX or VMAX				QMAX2 or VMAX2			
	F1	F2	F3	F4	F1	F2	F3	F4	F1	F2	F3	F4
V1	1	0	0	0	1	0	0	0	a	b	0	0
V2	1	0	0	0	1	0	0	0	a	b	0	0
V3	0	1	0	0	0	1	0	0	b	a	0	0
V4	0	1	0	0	0	1	0	0	b	a	0	0
V5	0	0	1	0	0	0	1	0	0	0	c	d
V6	0	0	1	0	0	0	1	0	0	0	c	d
V7	0	0	0	1	0	0	0	1	0	0	d	c
V8	0	0	0	1	0	0	0	1	0	0	d	c

Table 3
Unrotated loadings and loadings after varimax rotation of the Bavarian data set.

Variable	No Rotation				VMAX			
	F1	F2	F3	F4	F1	F2	F3	F4
Unemployment rate	0.11	0.77	0.21	-0.02	-0.27	0.76	-0.08	0.01
Salary	0.60	0.62	-0.09	0.04	0.32	0.79	-0.07	0.12
Population 6-15 (%)	0.95	0.05	-0.03	0.19	0.73	0.47	0.34	0.29
Population >60 (%)	-0.69	0.44	-0.48	0.00	-0.41	-0.02	-0.85	-0.06
Divorced (%)	-0.96	-0.02	0.12	0.00	-0.80	-0.44	-0.29	-0.09
Commuter (%)	0.72	-0.60	0.04	-0.12	0.73	-0.16	0.57	-0.07
Persons/Household	0.98	-0.09	-0.10	0.15	0.83	0.35	0.35	0.25
Larceny (%)	-0.82	-0.07	0.39	0.21	-0.84	-0.39	0.01	0.13
Stomach cancer (%)	0.62	0.39	0.38	-0.14	0.16	0.71	0.42	-0.06

Table 4
Loadings after varimax rotation for planes (VMAX2) of the Bavarian data set (F1 to F4). Contribution of each variable to the VMAX2 criterion for the principal planes F1-F2 and F3-F4 resulting from VMAX2 rotation and the planes F1-F3, F2-F4 from VMAX rotation, and sum over all single contributions.

	VMAX2				Contr. VMAX2		Contr. VMAX	
Variable	F1	F2	F3	F4	PPL1	PPL2	F1-F3	F2-F4
Unemployment rate	0.00	0.03	0.81	0.06	-3.00	7.37	-3.34	5.61
Salary	0.44	0.41	0.62	0.09	-0.99	0.89	-3.29	5.32
Population 6-15 (%)	0.55	0.74	0.15	-0.25	4.40	-1.53	0.71	-0.38
Population >60 (%)	-0.24	-0.41	0.16	0.81	-2.42	3.41	5.44	-1.29
Divorced (%)	-0.71	-0.60	-0.11	0.24	4.68	-1.56	1.99	-0.85
Commuter (%)	0.53	0.36	-0.44	-0.54	-1.08	1.00	4.89	-1.28
Persons/Household	0.62	0.73	0.01	-0.27	4.76	-1.56	2.54	-0.99
Larceny (%)	-0.86	-0.37	-0.07	-0.01	5.91	-1.61	2.35	-0.95
Stomach cancer (%)	0.34	0.25	0.59	-0.43	-2.43	3.44	-2.76	3.35
					9.84	9.84	8.54	8.54

Table 1

Mean computation time for one out of E combinations of planes (in seconds), for 2 planes (4 factors) to 6 planes (12 factors) at the basis of 50 replications of a simulation.

Table 2

Example of a matrix of loadings with complexity one: True pattern, QMAX or VMAX rotation, and QMAX2 or VMAX2 rotation ($a^2 + b^2 = c^2 + d^2 = 1$).

Table 3

Unrotated loadings and loadings after varimax rotation of the Bavarian data.

Table 4

Loadings after varimax rotation for planes (VMAX2) of the Bavarian data set (F1 to F4). Contribution of each variable to the VMAX2 criterion for the principal planes F1-F2 and F3-F4 resulting from VMAX2 rotation and the planes F1-F3, F2-F4 from VMAX rotation, and sum over all single contributions.

ORTHOGONAL PRINCIPAL PLANES

Peter Filzmoser[1]

DEPARTMENT OF STATISTICS, PROBABILITY THEORY AND ACTUARIAL MATHEMATICS
VIENNA UNIVERSITY OF TECHNOLOGY, AUSTRIA

[1]Department of Statistics, Probability Theory and Actuarial Mathematics, Vienna University of Technology, Wiedner Hauptstr. 8-10, A–1040 Vienna, Austria.
The author is most grateful to the referees, the Associate Editor, and the Editor for helpful suggestions, which resulted in a completely improved version of an earlier manuscript.

Abstract

Factor analysis and principal component analysis result in computing a new co-ordinate system, which is usually rotated to obtain a better interpretation of the results. In the present paper, the idea of rotation to simple structure is extended to two dimensions. While the classical definition of simple structure is aimed at rotating (one-dimensional) factors, the extension to a simple structure for two dimensions is based on the rotation of planes. The resulting planes (*principal planes*) reveal a better view of the data than planes spanned by factors from classical rotation and hence allow a more reliable interpretation. The usefulness of the method as well as the effectiveness of a proposed algorithm are demonstrated by simulation experiments and an example.

Key words: principal component analysis, factor analysis, orthogonal rotation, simple structure.