1

Generalized Principal Planes

Peter Filzmoser[1]

ABSTRACT We introduce a statistical method for the decomposition of data sets into planes. With the help of principal component analysis or factor analysis the data are presented in a lower-dimensional space. The factors spanning this new co-ordinate system are rotated afterwards in that sense to fulfill the idea of simple structure, extended to two dimensions. The resulting planes which are spanned by two factors each are called generalized principal planes. Dependent on the rotation criterion, these planes are orthogonal or oblique to each other.

1.1 Introduction

Principal component analysis and factor analysis are frequently used statistical tools for the reduction of the dimensionality, following a certain criterion (see e.g. Basilevsky, 1994). Let $\boldsymbol{x} = (x_1, \ldots, x_p)^\top$ be a random vector, and let $\boldsymbol{y} = (y_1, \ldots, y_p)^\top$ be the vector given by the transformation

$$y_i = \frac{x_i - \mathrm{E}(x_i)}{\sqrt{\mathrm{Var}(x_i)}}, \qquad i = 1, \ldots, p$$

to mean zero and variance one (standardization). In factor analysis we take the assumption that aside an error term $\boldsymbol{e}$ the elements of $\boldsymbol{y}$ can be represented by a smaller number $k < p$ of unknown random variables $\boldsymbol{f} = (f_1, \ldots, f_k)^\top$ which are called *factors*. Then the k-factor model may be written in the way $\boldsymbol{y} = \boldsymbol{\Lambda f} + \boldsymbol{e}$, where $\boldsymbol{\Lambda} = [(\lambda_{ij})]$ is a $(p \times k)$-matrix which is called *matrix of loadings* or *factor pattern*, and it describes the connection between factors and variables. The remaining components $\boldsymbol{e} = (e_1, \ldots, e_p)^\top$ are called *unique factors*. In this model we assume that $\mathrm{E}(\boldsymbol{f}) = \mathrm{E}(\boldsymbol{e}) = \boldsymbol{0}$, $\mathrm{Cov}(\boldsymbol{f}, \boldsymbol{e}) = \boldsymbol{O}$, $\mathrm{Cov}(e_i, e_j) = 0$ $(i \neq j)$, and $\mathrm{Var}(f_i) = 1$. With these restrictions, the correlation matrix $\boldsymbol{\rho}$ of the variables $\boldsymbol{y}$ can be expressed by $\boldsymbol{\rho} = \boldsymbol{\Lambda}\mathrm{Cov}(\boldsymbol{f})\boldsymbol{\Lambda}^\top + \mathrm{Cov}(\boldsymbol{e})$. With the additional assumption of orthogonal factors, the covariance matrix of the factors is the identity.

By this factor model a data matrix can be decomposed into factors and loadings (the same is true for principal component analysis). For a better

[1] Vienna University of Technology, Department of Statistics and Probability Theory, Vienna, Austria.

interpretation of the results the new co-ordinate system (factors) is rotated to obtain a so-called simple structure (Thurstone, 1944). We obtain rotated loadings and factor scores which are usually presented in planes spanned by all different pairs of factors, since planes are easy to survey. In this case of two-dimensional representation, the preceding analysis should also be "two-dimensional".

We want to find a two-dimensional decomposition of data sets where the resulting planes should contain a maximum of information. Such planes we call *generalized principal planes* since usual principal planes (planes spanned by pairs of principal components) are special cases of this method. They can be found by rotating the matrix of loadings in this sense to get a so-called two-dimensional simple structure, a generalization to higher dimensions of the usual simple structure. The basic ideas of this method are given in Section 1.2. In Section 1.3 a rotation criterion for obtaining orthogonal planes is introduced, and Section 1.4 is concerned with oblique planes. An example with a practical data set is shown in Section 1.5.

1.2 Method

Let us consider the standardized variables y_i ($i \in \{1, \ldots, p\}$) and two different factors f_a and f_b ($a, b \in \{1, \ldots, k\}$; $a \neq b$) which are obtained by principal component analysis or factor analysis (see previous section). For an extension of the usual rotation criteria to criteria for obtaining two-dimensional simple structure we have to define a distance measure between variables and factors. In the usual case, this distance measure is the correlation between variables and factors, which corresponds to the loadings in the orthogonal case.

For an extension to a two-dimensional rotation criterion we define the distance between variables and factors as the multiple correlation between y_i and the factors $\boldsymbol{f} = (f_a, f_b)^\top$ which is

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

(see e.g. Mardia et al., 1979). $\boldsymbol{\rho}_{\boldsymbol{f}y_i}$ is the correlation matrix (vector) between $\boldsymbol{f}$ and y_i, and $\boldsymbol{\rho}_{\boldsymbol{f}\boldsymbol{f}}$ is the correlation matrix of the factors f_a and f_b. If we suppose the factors to be orthogonal, $\boldsymbol{\rho}_{\boldsymbol{f}\boldsymbol{f}}$ (and therefore $\boldsymbol{\rho}_{\boldsymbol{f}\boldsymbol{f}}^{-1}$) is the identity, and the above equation reduces to

$$\rho_{y_i;\boldsymbol{f}} = \sqrt{\lambda_{ia}^2 + \lambda_{ib}^2} =: \rho_{i;a,b} \ . \tag{1.1}$$

λ_{ia} and λ_{ib} are the loadings of variable y_i on factor f_a and factor f_b, respectively.

For simplicity, we denote the multiple correlation between y_i and $\boldsymbol{f} = (f_a, f_b)^\top$ by $\rho_{i;a,b}$.

Before we come to a formal definition of the rotation criterion, we want to describe the background of the ideas for rotation. If the number of factors is k ($k \geq 4$), the number of different planes (spanned by two different factors) is $\lfloor k/2 \rfloor =: N$ where $\lfloor \cdot \rfloor$ denotes rounding down to the nearest integer. Since both, the sequence of the planes and the sequence of the two factors spanning a plane does not matter, with 4 factors we can find $\binom{4}{2}/2 = 3$ different combinations of factors for 2 planes. For $k = 5$ we can exchange each number in the previous case. So we have $3 \times 5 = 15$ possibilities to build 2 planes, one factor is left. The case $k = 6$ gives the same result as before since the remaining factor (the fifth number) of the previous case just has to be combined with the new one (the sixth number). If the number of factors is 7 we may exchange each number of the combinations from the case $k = 6$. This are $15 \times 5 = 3 \times 5 \times 7 = 105$ different combinations.

Let us denote S the set of all different combinations E of planes, $S = \{s_l, l = 1, \ldots, E\}$, where s_l denotes a set of pairs of indices of factors defining $N = \lfloor k/2 \rfloor$ planes.

Using these preliminary considerations, the number of different combinations E of N planes (k factors, $k \geq 4$) can be proved to be $E = \prod_{K=1}^{\lceil k/2 \rceil} (2K - 1)$, were $\lceil \cdot \rceil$ denotes rounding up to the nearest integer.

For our rotation criterion we have to find the optimum over all E possible combinations of different planes. In order to obtain simple structure on the planes, we would like to have high loadings for a variable on one plane and at the same time low loadings on the other planes. This means, the closer one variable comes to a plane (spanned by two factors), the smaller gets the product of the multiple correlations.

So we can define a criterion of simplicity as

$$\min_{s \in S} \left\{ \sum_{i=1}^{p} \sum_{\substack{\{(a,b),(c,d)\} \in s \\ a \neq c}} \widetilde{\rho}^2_{i;a,b} \widetilde{\rho}^2_{i;c,d} = \min \right\} \tag{1.2}$$

where $\widetilde{\rho}^2_{i;a,b} = \widetilde{\lambda}^2_{ia} + \widetilde{\lambda}^2_{ib}$, and $\widetilde{\lambda}_{ij}$ is an element of the orthogonally rotated matrix of loadings $\widetilde{\mathbf{\Lambda}}$. So the expression within the brackets has to be minimized by a transformation. The outer sum is over all variables $i = 1, \ldots, p$ and the inner sum is over all planes spanned by two factors. Furthermore, we have to find the minimum over all sets $s \in S$.

Once we have found the index set $s \in S$ and the rotation of the loadings which minimize (1.2), the factors $\{f_a, f_b\}$ ($\{(a, b)\} \in s$) span the final generalized principal planes. So we obtain N different planes which should be separated and contain as much information as possible.

We can use criterion (1.2) to develop extensions of usual orthogonal and oblique rotation criteria to criteria for obtaining generalized principal planes.

1.3 Orthogonal Rotation

A frequently used "one-dimensional" rotation criterion is the *varimax-criterion*, defined by

$$p \sum_{j=1}^{k} \sum_{i=1}^{p} \Big(\frac{\widetilde{\lambda}_{ij}}{\kappa_i} \Big)^4 - \sum_{j=1}^{k} \Big[\sum_{i=1}^{p} \Big(\frac{\widetilde{\lambda}_{ij}}{\kappa_i} \Big)^2 \Big]^2 = \max \tag{1.3}$$

(Kaiser, 1958). $\kappa_i^2 = \sum_{j=1}^{k} \lambda_{ij}^2$ $(i = 1, \ldots, p)$ is called i-th *communality*, and it describes the proportion of variance of the i-th standardized variable explained by all k factors. The idea of this criterion is to maximize the sum over the variances of the squared factor loadings for each factor. Since variables with high communalities would have more influence, we divide by this expression.

To translate criterion (1.3) into a two-dimensional varimax-criterion we have to consider the variance of the squared factor loadings for the planes spanned by the factors $\{f_a, f_b\}$ $(\{(a, b)\} \in s)$:

$$s_{a,b}^2 = \frac{1}{p} \sum_{i=1}^{p} \Big(\widetilde{\lambda}_{ia}^2 + \widetilde{\lambda}_{ib}^2 \Big)^2 - \frac{1}{p^2} \left[\sum_{i=1}^{p} \Big(\widetilde{\lambda}_{ia}^2 + \widetilde{\lambda}_{ib}^2 \Big) \right]^2 . \tag{1.4}$$

With the notation introduced in (1.2) of the previous section, (1.4) is equivalent to

$$s_{a,b}^2 = \frac{1}{p} \sum_{i=1}^{p} \Big(\widetilde{\rho}_{i;a,b}^2 \Big)^2 - \frac{1}{p^2} \left[\sum_{i=1}^{p} \widetilde{\rho}_{i;a,b}^2 \right]^2 \tag{1.5}$$

which describes the variance of the squared multiple correlations of the variables y_i with the plane $\{f_a, f_b\}$. Similar to the usual varimax-criterion we build the sum of (1.5) for all planes of a combination s and divide the loadings by the corresponding communalities. The resulting expression has to be maximized by an orthogonal transformation. Afterwards we have to find the maximum over all different combinations $s \in S$. So the *varimax-criterion for planes* is defined by

$$\max_{s \in S} \left\{ \sum_{\{(a,b)\} \in s} \left[p \sum_{i=1}^{p} \left(\frac{\widetilde{\rho}_{i;a,b}}{\kappa_i} \right)^4 - \left(\sum_{i=1}^{p} \left(\frac{\widetilde{\rho}_{i;a,b}}{\kappa_i} \right)^2 \right)^2 \right] = \max \right\} . \tag{1.6}$$

The results are generalized principal planes spanned by the factors $\{f_a, f_b\}$ $(\{(a, b)\} \in s)$. Note that criterion (1.6) can easily be modified to obtain a two-dimensional extension of the *quartimax-criterion*.

The practical solution of the maximization problem (1.6) can be found by an iterative process. We consider two different planes $\{f_a, f_b\}$ and $\{f_c, f_d\}$

$(\{(a,b),(c,d)\} \in s,\ a \neq c)$. For these two planes criterion (1.6) is

$$\begin{aligned} VMAX2_{a,b;c,d} &= p\sum_{i=1}^{p}\left[\left(\frac{\widetilde{\rho}_{i;a,b}}{\kappa_i}\right)^4 + \left(\frac{\widetilde{\rho}_{i;c,d}}{\kappa_i}\right)^4\right] \\ &\quad - \left[\sum_{i=1}^{p}\left(\frac{\widetilde{\rho}_{i;a,b}}{\kappa_i}\right)^2\right]^2 - \left[\sum_{i=1}^{p}\left(\frac{\widetilde{\rho}_{i;c,d}}{\kappa_i}\right)^2\right]^2 . \end{aligned} \tag{1.7}$$

The maximization has to be done by an orthogonal transformation. Since an orthogonal rotation in 4 dimensions would be rather complicated, it is approximated iteratively by performing the rotation 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 to 4 are repeated until (1.7) cannot be further increased. Afterwards, the 4 steps are performed onto the next two planes defined by the index set s, and so on. This procedure is repeated until convergence. Finally, we have to repeat the whole process for all different combinations $s \in S$ and take the maximum.

The rotation in the plane $\{f_a, f_c\}$ can be done by the orthogonal transformation

$$\widetilde{\lambda}_{ia} = \lambda_{ia}\cos\theta + \lambda_{ic}\sin\theta \tag{1.8}$$

$$\widetilde{\lambda}_{ic} = -\lambda_{ia}\sin\theta + \lambda_{ic}\cos\theta \tag{1.9}$$

$$\widetilde{\lambda}_{ij} = \lambda_{ij} \quad \text{for} \quad j \neq a, c \tag{1.10}$$

with the rotation angle θ. $\widetilde{\lambda}_{ij}$ and λ_{ij} denote the (i,j)-th element of the rotated and unrotated matrices of loadings $\widetilde{\boldsymbol{\Lambda}}$ and $\boldsymbol{\Lambda}$, respectively. Insertion into (1.7) gives

$$\begin{aligned} VMAX2_{\underline{a},b;\underline{c},d} &= p\sum_{i=1}^{p}\left[\frac{[(\lambda_{ia}\cos\theta + \lambda_{ic}\sin\theta)^2 + \lambda_{ib}^2]^2}{\kappa_i^4}\right. \\ &\quad \left. + \frac{[(-\lambda_{ia}\sin\theta + \lambda_{ic}\cos\theta)^2 + \lambda_{id}^2]^2}{\kappa_i^4}\right] \\ &\quad - \left[\sum_{i=1}^{p}\frac{(\lambda_{ia}\cos\theta + \lambda_{ic}\sin\theta)^2 + \lambda_{ib}^2}{\kappa_i^2}\right]^2 \\ &\quad - \left[\sum_{i=1}^{p}\frac{(-\lambda_{ia}\sin\theta + \lambda_{ic}\cos\theta)^2 + \lambda_{id}^2}{\kappa_i^2}\right]^2 . \end{aligned} \tag{1.11}$$

The underlined indices emphasize the factors which are rotated. The maximum of (1.11) can be found by setting the first derivative with respect to θ equal to zero. Since this derivative has no explicit solution for the rotation angle θ, an approximation has to be found e.g. with the *regula-falsi method* (see e.g. Golub and Ortega, 1993).

1.4 Oblique Rotation

The purpose of factor rotation is to obtain a simpler pattern of the matrix of loadings which should enable a better interpretation of the results. With the restriction of orthogonal factors it is in general not possible to achieve this simplicity in an optimal sense. Therefore it is suggested to rotate the factors according to oblique criteria.

A well-known oblique rotation criterion in classical factor analysis is the *oblimin-criterion*, which is defined by

$$\sum_{s<t=1}^{k} \left(\sum_{i=1}^{p} \widetilde{\lambda}_{is}^2 \widetilde{\lambda}_{it}^2 - \frac{\gamma}{p} \left(\sum_{i=1}^{p} \widetilde{\lambda}_{is}^2 \right) \left(\sum_{i=1}^{p} \widetilde{\lambda}_{it}^2 \right) \right) = \min \tag{1.12}$$

(see e.g. Harman, 1967).

An extension of criterion (1.12) to a criterion for two-dimensional simple structure can be found with similar thoughts as shown in Section 1.3. So we may define the *oblimin-criterion for planes* by

$$\min_{s\in S} \left\{ \sum_{\substack{\{(a,b),(c,d)\}\in s \\ a\neq c}} \left[\sum_{i=1}^{p} (\widetilde{\lambda}_{ia}^2 + \widetilde{\lambda}_{ib}^2)(\widetilde{\lambda}_{ic}^2 + \widetilde{\lambda}_{id}^2) - \frac{\gamma}{p} \sum_{i=1}^{p} (\widetilde{\lambda}_{ia}^2 + \widetilde{\lambda}_{ib}^2) \sum_{i=1}^{p} (\widetilde{\lambda}_{ic}^2 + \widetilde{\lambda}_{id}^2) \right] = \min \right\} . \tag{1.13}$$

Choosing the parameter $\gamma = 0$ gives the *quartimin-criterion for planes*, $\gamma = 1$ gives the so-called *covarimin-criterion for planes*.

In criterion (1.13) the indices (a, b) and (c, d) refer to different planes spanned by the factors $\{f_a, f_b\}$ and $\{f_c, f_d\}$, and they are included in a combination s of different planes. We have to minimize with respect to a transformation which is in general not orthogonal. Finally, the minimum over all different combinations E of planes has to be found.

The solution of the above minimization problem can be found iteratively. We consider two different planes $\{f_a, f_b\}$ and $\{f_c, f_d\}$ ($\{(a, b), (c, d)\} \in s$, $a \neq c$). For these two planes, criterion (1.13) is

$$OMIN2_{a,b;c,d} = \sum_{i=1}^{p} (\widetilde{\lambda}_{ia}^2 + \widetilde{\lambda}_{ib}^2)(\widetilde{\lambda}_{ic}^2 + \widetilde{\lambda}_{id}^2) - \frac{\gamma}{p} \sum_{i=1}^{p} (\widetilde{\lambda}_{ia}^2 + \widetilde{\lambda}_{ib}^2) \sum_{i=1}^{p} (\widetilde{\lambda}_{ic}^2 + \widetilde{\lambda}_{id}^2) . \tag{1.14}$$

This expression has to be minimized by a transformation. Similar to Section 1.3, 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 to 4 are repeated until (1.14) cannot be further decreased. Afterwards, the 4 steps are performed onto the next two planes of a combination in s, and so on, and this procedure is repeated until convergence. Finally, we have to repeat the whole process for all different combinations $s \in S$ and take the minimum.

The rotation in one of these 4 steps, for example in the plane $\{f_a, f_c\}$, can be done with the help of the transformation matrix

$$\boldsymbol{T}_{ac} = \begin{pmatrix} t_{aa} & 0 & t_{ac} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1.15}$$

(see Jennrich and Sampson, 1966). The rotated loadings are then determined by

$$\widetilde{\boldsymbol{\Lambda}} = \boldsymbol{\Lambda}\boldsymbol{T}_{ac} \ , \tag{1.16}$$

or expressed as elements of $\widetilde{\boldsymbol{\Lambda}}$ and $\boldsymbol{\Lambda}$,

$$\begin{aligned} \widetilde{\lambda}_{ia} &= t_{aa}\lambda_{ia} \\ \widetilde{\lambda}_{ic} &= t_{ac}\lambda_{ia} + \lambda_{ic} \\ \widetilde{\lambda}_{ij} &= \lambda_{ij} \qquad \text{for} \quad j = b, d \end{aligned}$$

$(i = 1, \ldots, p)$. Since the minimization of criterion (1.14) is independent from b and d, we may write the reduced problem as

$$\begin{aligned} OMIN2^*_{\underline{a},b;\underline{c},d} \;=\; & \sum_{i=1}^{p}(t_{aa}\lambda_{ia})^2(t_{ac}\lambda_{ia}+\lambda_{ic})^2 + \sum_{i=1}^{p}(t_{aa}\lambda_{ia})^2\lambda_{id}^2 \\ & + \sum_{i=1}^{p}(t_{ac}\lambda_{ia}+\lambda_{ic})^2\lambda_{ib}^2 - \frac{\gamma}{p}\sum_{i=1}^{p}\sum_{j=1}^{p}\Big[(t_{aa}\lambda_{ia})^2(t_{ac}\lambda_{ja}+\lambda_{jc})^2 \\ & + (t_{aa}\lambda_{ia})^2\lambda_{jd}^2 + (t_{ac}\lambda_{ia}+\lambda_{ic})^2\lambda_{jb}^2\Big] \ . \end{aligned} \tag{1.17}$$

The oblique factor analysis model for the two planes spanned by the factors $\boldsymbol{f} = (f_a, f_b, f_c, f_d)^\top$ is

$$\boldsymbol{\rho} - \boldsymbol{\Psi} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^\top \tag{1.18}$$

(see Section 1.1), where $\boldsymbol{\Psi}$ is the covariance matrix of the error term (diagonal matrix) and $\boldsymbol{\Phi}$ is the covariance matrix of the factors. Insertion of (1.16) into the model (1.18) gives

$$\boldsymbol{\rho} - \boldsymbol{\Psi} = \widetilde{\boldsymbol{\Lambda}}\boldsymbol{T}_{ac}^{-1}\boldsymbol{\Phi}(\boldsymbol{T}_{ac}^{-1})^\top\widetilde{\boldsymbol{\Lambda}}^{\,\top} = \widetilde{\boldsymbol{\Lambda}}\mathrm{Cov}(\boldsymbol{T}_{ac}^{-1}\boldsymbol{f})\widetilde{\boldsymbol{\Lambda}}^{\,\top} \ .$$

The rotated factors are

$$\widetilde{\boldsymbol{f}} = \boldsymbol{T}_{ac}^{-1}\boldsymbol{f} \ , \tag{1.19}$$

and the correlation matrix $\widetilde{\boldsymbol{\Phi}}$ of the rotated factors is

$$\widetilde{\boldsymbol{\Phi}} = \mathrm{Cov}(\boldsymbol{T}_{ac}^{-1}\boldsymbol{f}) \ . \tag{1.20}$$

If $t_{aa} \neq 0$ in (1.15), the inverse of $\boldsymbol{T}_{ac}$ is

$$\boldsymbol{T}_{ac}^{-1} = \begin{pmatrix} \frac{1}{t_{aa}} & 0 & -\frac{t_{ac}}{t_{aa}} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} ,$$

and (1.19) can be rewritten as

$$\begin{aligned} \widetilde{f}_a &= \frac{1}{t_{aa}} f_a - \frac{t_{ac}}{t_{aa}} f_c \\ \widetilde{f}_j &= f_j \quad \text{for} \quad j = b, c, d \ . \end{aligned}$$

Since we want the factors to have unit length,

$$\widetilde{f}_a^2 = \frac{1}{t_{aa}^2} f_a^2 + \frac{t_{ac}^2}{t_{aa}^2} f_c^2 - 2\frac{t_{ac}}{t_{aa}^2} f_a f_c = 1 \ ,$$

we get the condition

$$t_{aa}^2 = 1 + t_{ac}^2 - 2t_{ac}\varphi_{ac} \tag{1.21}$$

with $\varphi_{ac} = f_a f_c$ being the correlation between the factors f_a and f_c. It can be shown by an elementary calculation that insertion of condition (1.21) into (1.17) gives

$$\begin{aligned} OMIN2^*_{\underline{a},b;\underline{c},d} &= (c_0 + d_0) + (c_1 + d_1)t_{ac} + (c_2 + d_2)t_{ac}^2 \\ &\quad +(c_3 + d_3)t_{ac}^3 + (c_4 + d_4)t_{ac}^4 \end{aligned} \tag{1.22}$$

with

$$\begin{aligned} c_0 &= \sum_{i=1}^{p} (\lambda_{ia}^2\lambda_{ic}^2 + \lambda_{ia}^2\lambda_{id}^2 + \lambda_{ib}^2\lambda_{ic}^2) \\ c_1 &= \sum_{i=1}^{p} (-2\varphi_{ac}\lambda_{ia}^2\lambda_{ic}^2 + 2\lambda_{ia}^3\lambda_{ic} - 2\varphi_{ac}\lambda_{ia}^2\lambda_{id}^2 + 2\lambda_{ia}\lambda_{ib}^2\lambda_{ic}) \\ c_2 &= \sum_{i=1}^{p} (\lambda_{ia}^4 + \lambda_{ia}^2\lambda_{ic}^2 - 4\varphi_{ac}\lambda_{ia}^3\lambda_{ic} + \lambda_{ia}^2\lambda_{id}^2 + \lambda_{ia}^2\lambda_{ib}^2) \\ c_3 &= \sum_{i=1}^{p} (-2\varphi_{ac}\lambda_{ia}^4 + 2\lambda_{ia}^3\lambda_{ic}) \\ c_4 &= \sum_{i=1}^{p} \lambda_{ia}^4 \end{aligned}$$

$$\begin{aligned}
d_0 &= -\frac{\gamma}{p}\sum_{i=1}^{p}\sum_{j=1}^{p}(\lambda_{ia}^2\lambda_{jc}^2+\lambda_{ia}^2\lambda_{jd}^2+\lambda_{ic}^2\lambda_{jb}^2)\\
d_1 &= -\frac{\gamma}{p}\sum_{i=1}^{p}\sum_{j=1}^{p}\Big[2\lambda_{ia}^2(-\varphi_{ac}\lambda_{jc}^2+\lambda_{ja}\lambda_{jc}-\varphi_{ac}\lambda_{jd}^2)+2\lambda_{ia}\lambda_{ic}\lambda_{jb}^2\Big]\\
d_2 &= -\frac{\gamma}{p}\sum_{i=1}^{p}\sum_{j=1}^{p}\lambda_{ia}^2(\lambda_{ja}^2+\lambda_{jc}^2-4\varphi_{ac}\lambda_{ja}\lambda_{jc}+\lambda_{jd}^2+\lambda_{jb}^2)\\
d_3 &= -\frac{\gamma}{p}\sum_{i=1}^{p}\sum_{j=1}^{p}2\lambda_{ia}^2(-\varphi_{ac}\lambda_{ja}^2+\lambda_{ja}\lambda_{jc})\\
d_4 &= -\frac{\gamma}{p}\sum_{i=1}^{p}\sum_{j=1}^{p}\lambda_{ia}^2\lambda_{ja}^2\ .
\end{aligned}$$

The first derivative of (1.22) with respect to t_{ac} is

$$(c_1+d_1)+2(c_2+d_2)t_{ac}+3(c_3+d_3)t_{ac}^2+4(c_4+d_4)t_{ac}^3 \tag{1.23}$$

which is set to zero in order to find the minimum. From (1.23) t_{ac} can be determined. t_{aa} is, aside the sign, definitely determined by condition (1.21). The rotated loadings $\widetilde{\mathbf{\Lambda}}$ can be seen from (1.16). The rotated correlation matrix $\widetilde{\mathbf{\Phi}}$ between the factors $\widetilde{f}_a$ and $\widetilde{f}_c$ is calculated in analogy to (1.20), namely

$$\widetilde{\mathbf{\Phi}} = \boldsymbol{T}_{ac}^{-1}\mathbf{\Phi}\left(\boldsymbol{T}_{ac}^{-1}\right)^{\top}$$

or expressed in another notation

$$\begin{aligned}
\widetilde{\varphi}_{aj} &= \frac{1}{t_{aa}}\varphi_{aj}-\frac{t_{ac}}{t_{aa}}\varphi_{cj} \qquad \text{for} \qquad j=b,c,d\\
\widetilde{\varphi}_{ij} &= \varphi_{ij} \qquad \text{for} \qquad i,j=b,c,d\ .
\end{aligned}$$

$\widetilde{\varphi}_{aa}=1$ follows from condition (1.21).

A comparison between classical rotation and rotation to generalized principal planes was done by simulation (see Filzmoser, 1996). For simulated data sets which include variables approximately arranged in a plane it was shown, that the variables of the plane are represented in much a better way when using “two-dimensional” criteria.

1.5 Example

The data set we want to investigate in this section was prepared by H. Rauth and G. Sedlacek. We fell obliged to them for the permission to work with these data. For a detailed description see Rauth et al. (1996).

The concern is measurements of the year 1991 in all 99 Austrian districts (Vienna is treated as one district). The following variables have been considered (the abbreviations in the brackets refer to Figures 1.1 to 1.3): percentage of children ($<$ 15 years) (`Children`), youth and adult people (15 to 60 years) (`Adult+Youth`), and old people ($>$ 60 years) (`Old`) in the resident population; portion of employees in the manufacturing industry, in the industry, and in the building trade (`Industry`), in the trade (`Trade`), in the lodging and catering trade (`Tourism`), in service (`Service`), and in the agriculture and forest (`Agriculture`), in percent of the total employees; portion of unemployed people (`Unemployed`) in percent of the resident population; portion of people with university education (`University`), secondary school (`Sec.School`), and primary school, technical college, or apprenticeship (`Prim.School`) in percent of the resident population; number of total employees in workshop places divided by the number of the total employees of the resident population (`Im/migration`) (this number should express the economic situation of the district, and it shows if there is immigration or migration); portion of commuters commuting not daily (`Comm.n.daily`), portion of commuters commuting to another district (`Comm.distr.`) in percent of the total employees of the resident population; portion of mountain farms in percent of all farms (`Mountain-farm`); number of overnight stays per year in the tourism (`Tourist-stays`).

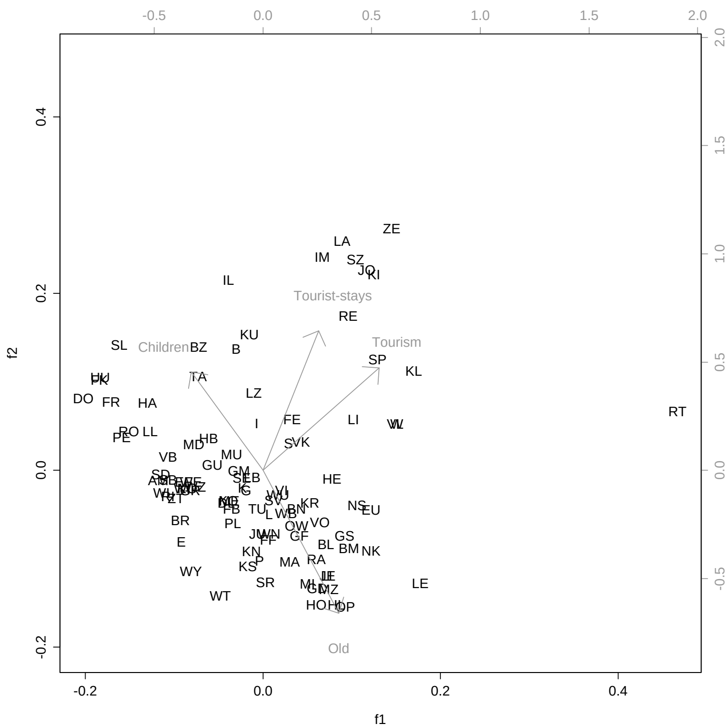
A description of some objects of interest is given in the interpretations of the final figures. The objects with one-letter abbreviations are the capitals (cities) of the 9 Austrian political districts.

We standardize the data matrix (variables) and calculate the principal components. The first 6 components explain about 86% of the total variation. At next a rotation of the components according to the *varimax-criterion for planes* (1.6) is performed. We obtain 3 planes which are orthogonal to each other. Since we are also interested in the values of the objects in the new co-ordinate system, the factor scores are estimated (see e.g. Basilevsky, 1994).

At this stage we have loadings and scores of the three principal planes. A useful tool for the presentation of these results are biplots (Gabriel, 1971) where both variables and objects can be shown in one plot. For a better interpretation of the biplots we present just variables with a higher squared multiple correlation coefficient than 0.4 with the plane (see (1.1)). This means that only those variables are shown in the plots which are "close" to the plane.

The biplot representations of the three principal planes are shown in Figures 1.1 to 1.3. We also give interpretations to the planes, which are short by intention. The cosine between the variable vectors approximates the correlation, the projection of the objects onto the variable vectors approximate the data values.

FIGURE 1.1. First Principal Plane



Interpretation of Figure 1.1:

This plane shows the connections between tourism and population. The districts in the north of the plane are rural touristic regions. There is a large portion in children and in the tourism. Reutte (`RT`) would be the only district with many tourists and a large portion in old people, but, possibly, the estimation for this district was not very reliable.

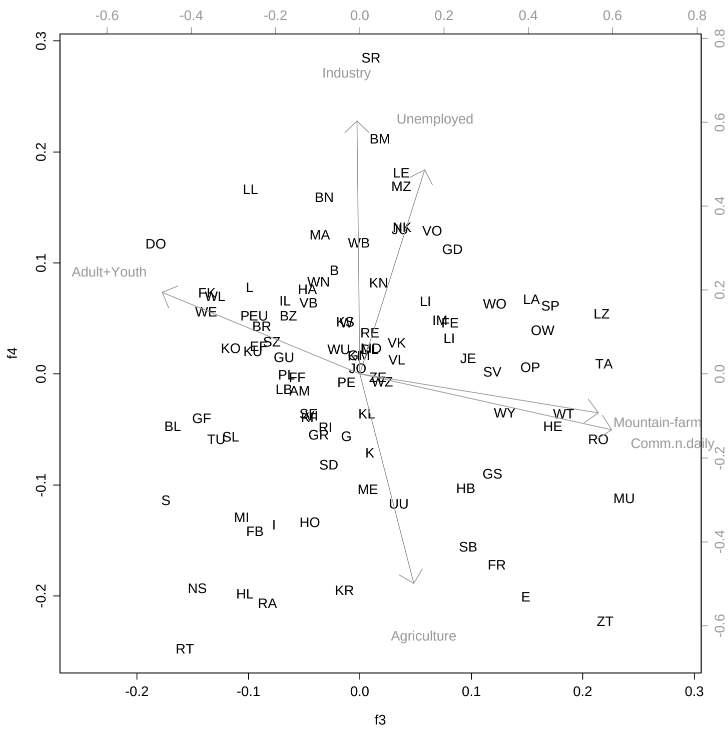
It is clear that the variables `Tourist-stays` and `Tourism`, are positively correlated. The variables describing the portion in children and old people are highly negative correlated. The districts with a large portion in these variables may be seen easily.

Interpretation of Figure 1.2:

In the north of this plane we find a high correlation between unemployment and portion of employed in the industry. A typical region with these properties is the so-called *Mur-Mürz-Furche* which consists of the districts Bruck/Mur (`BM`), Leoben (`LE`), and Mürzzuschlag (`MZ`).

In the east we find a very high correlation between the variable describing the number of mountain farms, and the portion of employed people

FIGURE 1.2. Second Principal Plane



commuting not daily.

In the south of this plane we have the regions with a large portion of employment in the agriculture. This variable is highly negative correlated with unemployment and employment in the industry.

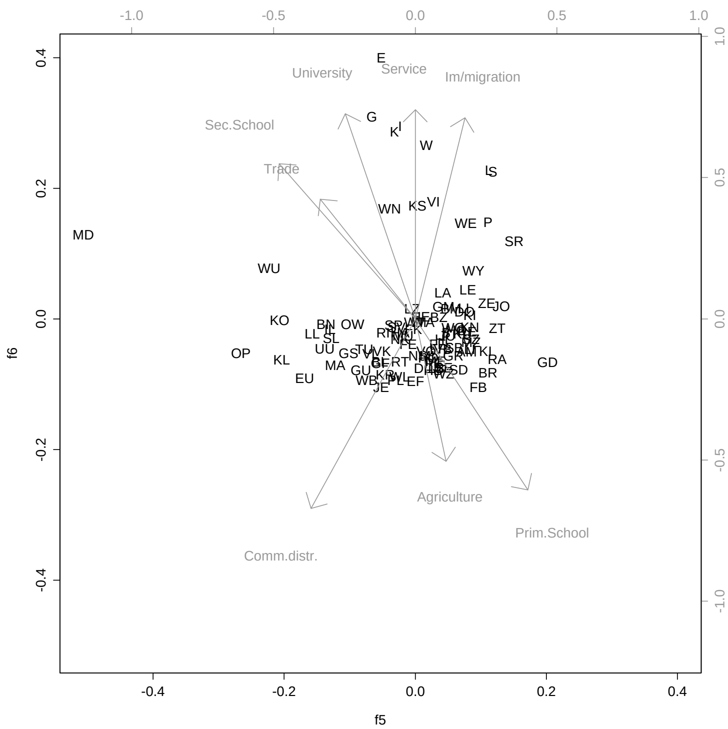
In the west we have non-rural regions (they are negatively correlated with `Mountain-farm`) with a high portion in people in the age able to work.

Interpretation of Figure 1.3:

This plane shows the contrast between city (north) and country (south). In the north we have high education, a large portion of employment in trade and service, and large immigration. Typical regions with these properties are the main cities (except Bregenz (`B`)), the small city Eisenstadt (`E`) has the highest values. Districts with the same properties are also Wiener Neustadt (`WN`), Krems (`KS`), Villach (`VI`), Wels (`WE`), and perhaps Steyr (`SR`) and Wien-Umgebung (`WU`). These districts are close to the main cities with good connections to them. The district Mödling (`MD`) also belongs to these regions, the difference is, that there is a low immigration.

In the south we have districts with large portion of employment in the

FIGURE 1.3. Third Principal Plane



agriculture. One mark of those regions is low education. There are a lot of districts with a large portion of employed people commuting to another district.

In the west there are regions with a large portion of employed people commuting to another district. These regions are mainly districts close to main cities.

1.6 Summary

This multivariate method for obtaining principal planes is very helpful specially for data sets where relations between the variables, and also relations between variables and objects, are to be investigated. At the basis of factors arising from principal component analysis or factor analysis, a rotation of these factors is performed in that sense to obtain simple structure on planes. This means that the resulting planes should be close to a maximum number of variables, and the planes should be as separated as possible (which is helpful for the interpretation). As a measure of "closeness" we use the multiple correlation between each variable and two factors spanning a plane. In the example (Section 1.5) we were able to present all 17 variables in the 3 planes quite well, and just the variable `Agriculture` arose in two planes.

The investigation of the results can be facilitated if the generalized principal planes are represented with biplots. So especially relations between variables and objects can be analyzed in a better way. It is useful for the interpretation to present just those variables which are "close" to the planes. As a measure of closeness we can again use the multiple correlation between each variable and the two factors spanning the plane.

This method can easily be extended to three-dimensional simple structure (*generalized principal spaces*), but possibly two-dimensional representations are easier to survey.

1.7 References

[Bas94] A. Basilevsky. *Statistical Factor Analysis and Related Methods: Theory and Applications.* Wiley & Sons, New York, 1994.

[Fil96] P. Filzmoser. *Principal Planes.* PhD dissertation, Dept. of Statistics and Probability Theory, Univ. of Technology, Vienna, 1996. Unpublished.

[Gab71] K.R. Gabriel. The biplot graphic display of matrices with application to principal component analysis. *Biometrika*, 58(3):453–467, 1971.

[GO93] G.H. Golub and J.M. Ortega. *Scientific Computing, An Introduction with Parallel Computing.* Academic Press, New York, 1993.

[Har67] H.H. Harman. *Modern Factor Analysis.* The University of Chicago Press, Chicago and London, 2nd edition, 1967.

[JS66] R.I. Jennrich and P.F. Sampson. Rotation for simple loadings. *Psychometrika*, 31(3):313–323, 1966.

[Kai58] H.F. Kaiser. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3):187–200, 1958.

[MKB79] K.V. Mardia, J.T. Kent, and J.M. Bibby. *Multivariate Analysis.* Acad. Press, London, 1979.

[RS96] H. Rauth and G. Sedlacek. Classification of Austria into homogeneous mathematical regions. Technical report, Dept. of Statistics and Probability Theory, Univ. of Technology, Vienna, 1996. (In German), Unpublished.

[Thu44] L.L. Thurstone. Second-order factors. *Psychometrika*, 9:71–100, 1944.