High Dimensional Data Analysis

Factor Analysis

Factor Analysis

- Introduction
- Factor Model
- Estimation of Factor Loadings
- Factor orthogonal rotation
- Factor Scores

Introduction

- Factor analysis originated in the early 20th century and was a statistical method proposed by scholars such as K. Pearson and C. Spearman to solve the problem of intelligence test scores.
- Factor analysis is an extension and development of principal component analysis, and it is another important dimensionality reduction method in multivariate statistical analysis.
- Factor analysis studies the internal dependencies of the covariance or correlation matrix of multivariate random vectors, expressing multiple variables as being driven by a few common factors, in order to reproduce the structural relationship between the original variables and the common factors.

Content and Requirements

Content:

- Basic ideas of factor analysis model, differences with principal component analysis
- Factor model, basic assumptions, estimation methods of factor loading matrix, factor rotation, factor scores
- Practical statistical significance and mathematical expressions of factor rotation and factor scores
- Algorithm basis of factor analysis in computational programs

Requirements:

- Understand the purpose and practical significance of factor analysis
- Familiar with modeling assumptions of factor analysis mathematical model and practical statistical significance of each component
- Master derivation steps of estimating factor loading matrix using principal component method, and important basic properties
- Able to use computational software programming to solve practical factor analysis problems, provide statistical analysis report

Example 1

To understand students' learning abilities, observations were made on n students' scores in p subjects, denoted as $x_1, x_2, \ldots, x_p$ (e.g., mathematics, Chinese, English, politics, history, etc.). The goal is to analyze which main factors determine students' learning abilities.
Through inductive analysis of each student's scores across subjects, it can be seen that the scores for each subject (i.e., variable) consist of two parts:

$$x_i = a_i f + \epsilon_i, \quad i = 1, \cdots, p,$$

where f is a common factor that affects all $x_i (i = 1, \cdots, p)$, representing the factor of intelligence level. The coefficient a_i is called the factor loading, and ϵ_i is the specific factor related only to the subject. This is a simple factor model.

Example 2

Linden conducted research on the scores of Olympic decathlon events since World War II. He collected 160 sets of data, using $x_1, x_2, \ldots, x_{10}$ to represent the standardized scores of the ten decathlon events. The decathlon events in order are: 100m dash, long jump, shot put, high jump, 400m run, 110m hurdles, discus throw, pole vault, javelin throw, and 1500m run. The goal is to analyze which main factors determine the performance in the decathlon.
Through analysis of the scores from these ten decathlon events, it can be essentially attributed to four factors: sprint speed, explosive arm strength, explosive leg strength, and endurance. Each of these factors is called a factor, therefore this problem can be handled using the following factor model:

$$x_i = a_{i1}f_1 + a_{i2}f_2 + a_{i3}f_3 + a_{i4}f_4 + \varepsilon_i, \quad i = 1, \ldots, 10,$$

where f_1, f_2, f_3, f_4 are the four common factors that affect all variables $x_i (i = 1, \ldots, 10)$, and ε_i is the specific factor related only to variable x_i.

Factor Analysis: Applications and Differences

Factor analysis has two main applications:

1. Seeking basic structure: Reducing objects (variables) with complex interrelationships to a few common factors with clear meanings, to reproduce the intrinsic relationship between common factors and original variables.
2. Classifying variables or samples, and ordering or evaluating samples.

The main differences between factor analysis and principal component analysis are as follows:

1. Principal component analysis only involves general variable transformation and does not require a model for description; while factor analysis requires constructing a factor model that builds the intrinsic relationship between original variables and common factors.
2. In principal component analysis, principal components are expressed as linear combinations of original variables to achieve dimensionality reduction of original variables; while in factor analysis, original variables are expressed as linear combinations of common factors and specific factors to achieve fine characterization and dimensionality reduction of original variables.

Differences Between Factor Analysis and PCA (Continued)

3. Principal component analysis strives to use a few principal components to explain the total variance, while factor analysis strives to use a few common factors to describe covariance or correlation relationships; although they share some common goals, their mathematical mechanisms and levels of refinement differ.
4. Except for possible positive/negative sign differences in coefficient vectors, the solution of principal components is unique; while due to the unobservability of latent factors, the solution of factors is not unique and can undergo rotational transformation. This flexibility allows us to obtain common factors with clear meanings, which is an important reason why factor analysis has broader applications than principal component analysis.
5. Principal components do not change with the number of extracted principal components, but factors change with the number of factors in the factor model.

Factor Model

Let $x = (x_1, \cdots, x_p)$ be an observable p-dimensional random vector with mean vector and covariance matrix $\mu = E(x), \Sigma = Var(x)$. In factor analysis, data are generally centered and standardized, where $\mu = 0, \Sigma = R$. The general form of the factor model is:

$$\begin{cases} x_1 = a_{11}f_1 + a_{12}f_2 + \cdots + a_{1m}f_m + \varepsilon_1 = a'_{(1)}f + \varepsilon_1, \\ x_2 = a_{21}f_1 + a_{22}f_2 + \cdots + a_{2m}f_m + \varepsilon_2 = a'_{(2)}f + \varepsilon_2, \\ \vdots \\ x_p = a_{p1}f_1 + a_{p2}f_2 + \cdots + a_{pm}f_m + \varepsilon_p = a'_{(p)}f + \varepsilon_p, \end{cases}$$

This model can be expressed in matrix form:

$$x = Af + \varepsilon,$$

where $\varepsilon = (\varepsilon_1, \cdots, \varepsilon_p)$ is the specific factor vector,
$A = (a_{ij})_{p\times m} = (a_1, \cdots, a_m) = (a_{(1)}, \cdots, a_{(p)})$ is the factor loading matrix.

Factor Model

The basic assumptions of the factor model are:

$$E(f) = 0, \quad Var(f) = I_m,$$

$$Cov(f, \epsilon) = 0, \quad E(\epsilon) = 0,$$

$$Var(\epsilon) = D = diag(\sigma_1^2, \cdots, \sigma_p^2).$$

From the above assumptions, it can be seen that the common factors are uncorrelated with each other and have unit variance, the specific factors are uncorrelated with each other, and the specific factors are uncorrelated with the common factors. A factor model satisfying the above assumptions is called an orthogonal factor model.
The orthogonal factor model has the following properties:
(1) The covariance matrix has the decomposition form

$$\Sigma = AA^T + D.$$

If the original variables are centered and standardized, then the matrix on the left side is the correlation matrix R.

Factor Model

(2) The model is not affected by the units of the variables.
(3) The factor loading matrix A is not unique. Because, if T is any orthogonal matrix, let $A^* = AT, f^* = T^T f$, then the orthogonal factor model can be expressed as

$$x = A^* f^* + \epsilon.$$

The above non-uniqueness of the factor loading matrix is a good property. Using this property, through rotational transformation of the factor loading matrix, we can obtain a factor model with common factors having clear meanings.
(4) Statistical meaning of factor loadings. Since

$$Cov(x_i, f_j) = a_{ij},$$

if x_i is a standardized variable, then the factor loading a_{ij} is the correlation coefficient between the i-th original variable and the j-th common factor. That is, the loading matrix A of the factor model is the correlation matrix between the original variables and the common factors.

Factor Model

The squared sum of elements in the i-th row of matrix A, $a_{(i)}^T = (a_{i1}, a_{i2}, \cdots, a_{im})$,

$$h_i^2 = a_{(i)}^T a_{(i)} = \sum_{j=1}^{m} a_{ij}^2$$

is the communality of variable x_i, also called common variance. It reflects the influence of common factors on the original variable and can be viewed as the variance contribution of common factors to x_i.
The squared sum of elements in the j-th column of matrix A,

$$g_j^2 = \sum_{i=1}^{p} a_{ij}^2$$

reflects the influence of common factor f_j on the original variables and is a measure of the importance of this common factor, which can be regarded as the variance contribution of this common factor to the original variables.

Example 3

Let the covariance matrix of random vector $x = (x_1, \cdots, x_4)$ be

$$\Sigma = \begin{pmatrix} 6 & 6 & 2 & 12 \\ 6 & 11 & 1 & 17 \\ 2 & 1 & 6 & 8 \\ 12 & 17 & 8 & 16 \end{pmatrix}.$$

Σ can be decomposed as

$$\Sigma = AA^T + D = \begin{pmatrix} 2 & 0 \\ 3 & -1 \\ 1 & 2 \\ 6 & 1 \end{pmatrix} \begin{pmatrix} 2 & 3 & 1 & 6 \\ 0 & -1 & 2 & 1 \end{pmatrix} + \begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix},$$

From the elements of matrix A, we can obtain the communality of each original variable and the variance contribution of each common factor. For example: The communality of variable x_3 is

$$a_{31}^2 = a_{31}^2 + a_{32}^2 = 1^2 + 2^2 = 5.$$

The variance contribution of common factor f_2 to $x = (x_1, \cdots, x_4)'$ is

$$g_2^2 = \sum_{i=1}^{4} a_{i2}^2 = 0^2 + (-1)^2 + 2^2 + 1^2 = 6.$$

Estimation of Factor Loadings

Similar to principal component analysis, we can perform factor analysis starting from the covariance matrix or correlation matrix. However, in practice, the covariance matrix or correlation matrix is often unknown and needs to be estimated from samples. Therefore, we use the sample covariance matrix as an estimate of the population covariance matrix, and the sample correlation matrix as an estimate of the population correlation matrix. Next, we consider estimating the factor loading matrix A and the diagonal matrix D through the decomposition of the sample correlation matrix R.

Principal Component Method Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p \geq 0$ be the eigenvalues of the sample correlation matrix R, and $u_1, u_2, \cdots, u_p$ be the corresponding unit orthogonal eigenvectors. Then R has the spectral decomposition:

$$R = \sum_{i=1}^{p} \lambda_i u_i u_i' = \left(\sqrt{\lambda_1} u_1, \cdots, \sqrt{\lambda_p} u_p\right) \begin{pmatrix} \sqrt{\lambda_1} u_1' \\ \vdots \\ \sqrt{\lambda_p} u_p' \end{pmatrix}$$

Estimation of Factor Loadings

When the last p-m eigenvalues of R are small, it can be approximately decomposed as:

$$R \approx \left(\sqrt{\lambda_1}u_1, \cdots, \sqrt{\lambda_m}u_m\right) \begin{pmatrix} \sqrt{\lambda_1}u_1' \\ \vdots \\ \sqrt{\lambda_m}u_m' \end{pmatrix} + \begin{pmatrix} \sigma_1^2 & & \\ & \ddots & \\ & & \sigma_p^2 \end{pmatrix}$$

$$= AA' + D,$$

where

$$A = \left(\sqrt{\lambda_1}u_1, \cdots, \sqrt{\lambda_m}u_m\right) = (a_{ij})_{p\times m},$$

$$\sigma_i^2 = 1 - \sum_{j=1}^{m} a_{ij}^2 = 1 - h_i^2, \quad i = 1, 2, \cdots, p,$$

$$D = diag(\sigma_1^2, \sigma_2^2, \cdots, \sigma_p^2).$$

The A and D given above are a solution of the factor model, called the principal component solution of the factor model.

Estimation of Factor Loadings

Starting from the sample correlation matrix, we further improve or modify the principal component solution. If we have an estimate of the initial communality h_i^{*2}, we can obtain the following correlation matrix:

$$R^* = \begin{pmatrix} h_1^{*2} & r_{12} & \cdots & r_{1p} \\ r_{21} & h_2^{*2} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & h_p^{*2} \end{pmatrix}$$

Let $\lambda_1^*, \lambda_2^*, \ldots, \lambda_m^* > 0$ be the first m eigenvalues of R^*, and their corresponding unit orthogonal eigenvectors be $u_1^*, u_2^*, \ldots, u_m^*$. Then we have the approximate decomposition:

$$R^* = \left(\sqrt{\lambda_1^*}u_1^*, \ldots, \sqrt{\lambda_m^*}u_m^*\right) \begin{pmatrix} \sqrt{\lambda_1^*}u_1^{*'} \\ \vdots \\ \sqrt{\lambda_m^*}u_m^{*'} \end{pmatrix} = AA',$$

Then $A = \left(\sqrt{\lambda_1^*}u_1^*, \ldots, \sqrt{\lambda_m^*}u_m^*\right)$ and $D = diag(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_p^2)$ are an improved solution of the factor model, and this solution is called the principal factor solution.

Estimation of Factor Loadings

The solutions obtained above are all approximate solutions. To obtain solutions with higher approximation accuracy, the iterative principal factor method is often used, which utilizes the $\hat{D} = diag(\hat{\sigma}_1^2, \cdots, \hat{\sigma}_p^2)$ obtained above as the initial estimate of specific variances, repeating the above steps until a stable solution is obtained.

Maximum Likelihood Method

Assume the distributions of original variables, common factors, and specific factors are all multivariate normal distributions. In this case, the likelihood function can be expressed as $L(A, D)$. It can be shown that the solution $\hat{A}, \hat{D}$ that maximizes $L(A, D)$ satisfies the following system of equations:

$$S\hat{D}^{-1}\hat{A} = \hat{A}(I + \hat{A}^T\hat{D}^{-1}\hat{A}),$$

$$\hat{D} = diag(S - \hat{A}\hat{A}^T),$$

where S is the sample covariance matrix. This system of equations does not have an exact solution in analytical form, and multiple iterations are required to obtain satisfactory results.

Example 4

Comprehensive Development Data of a Certain Region Over 20 Years

To study the comprehensive development status of a certain region, statistical data on economic development and social conditions over 20 years were collected, including total population (x1), GDP (x2), social fixed asset investment (x3), urbanization level (x4), per capita living area (x5), and passenger traffic volume (x6). Use the principal component method, principal factor method, and maximum likelihood method respectively to estimate the factor loading matrix for this data.

No.	x_1	x_2	x_3	x_4	x_5	x_6
1	7296.0	292.13	69.97	0.095	5.7	12208.0
2	7395.0	346.57	79.6	0.099	6.1	12682.0
3	7494.0	395.38	85.0	0.103	6.11	13109.0
4	7564.0	459.83	96.46	0.107	6.65	14839.0
5	7637.0	581.56	140.15	0.123	6.9	17309.0
6	7711.0	680.46	194.33	0.132	7.77	19772.0
7	7818.0	742.05	223.08	0.125	9.15	26459.0
8	7958.0	892.29	297.77	0.132	9.61	25209.0
9	8061.0	1117.66	369.82	0.162	9.96	29035.0
10	8160.0	1293.94	305.54	0.182	10.25	30718.0
11	8493.0	1511.19	335.66	0.186	10.05	29798.0
12	8570.0	1810.54	439.82	0.193	10.49	31931.0
13	8610.0	2196.53	601.5	0.205	10.8	33920.0
14	8642.0	2779.49	892.48	0.219	11.2	33634.0
15	8671.0	3872.18	1108.0	0.24	11.88	34467.0
16	8765.0	5002.34	1320.97	0.248	12.35	36425.0
17	8738.0	5960.42	1558.01	0.259	12.13	39199.0
18	8785.0	6650.02	1792.22	0.263	12.7	43218.0
19	8838.0	7162.2	2056.97	0.26	12.82	50558.0
20	8883.0	7662.1	2243.9	0.261	13.1	59350.0

Estimation of Factor Loadings

Solution: First determine the number of factors m using principal component analysis. When $m = 1$, the cumulative contribution rate is 0.93; when $m = 2$, the cumulative contribution rate reaches 0.98. In this factor analysis, we take $m = 2$. The initial estimates of the factor loading matrix are obtained using three methods below.

(1) Principal component method estimate. The factor loading matrix and specific variance matrix obtained by the principal component method are:

$$A_1 = \begin{pmatrix} 0.96 & -0.27 \\ 0.95 & 0.30 \\ 0.95 & 0.30 \\ 0.98 & -0.10 \\ 0.97 & -0.22 \\ 0.97 & 0.00 \end{pmatrix}, \quad D_1 = \begin{pmatrix} 0.01 & 0 & 0 & 0 & 0 \\ 0 & 0.00 & 0 & 0 & 0 \\ 0 & 0 & 0.00 & 0 & 0 \\ 0 & 0 & 0 & 0.03 & 0 \\ 0 & 0 & 0 & 0 & 0.02 \\ 0 & 0 & 0 & 0 & 0.06 \end{pmatrix}$$

Estimation of Factor Loadings

(2) Principal factor method estimate. The factor loading matrix and specific variance matrix obtained by the principal factor method are:

$$A_2 = \begin{pmatrix} 0.95 & -0.27 \\ 0.95 & 0.29 \\ 0.95 & 0.30 \\ 0.98 & -0.10 \\ 0.96 & -0.21 \\ 0.96 & -0.01 \end{pmatrix}, \quad D_2 = \begin{pmatrix} 0.02 & 0 & 0 & 0 & 0 \\ 0 & 0.01 & 0 & 0 & 0 \\ 0 & 0 & 0.00 & 0 & 0 \\ 0 & 0 & 0 & 0.03 & 0 \\ 0 & 0 & 0 & 0 & 0.03 \\ 0 & 0 & 0 & 0 & 0.08 \end{pmatrix}$$

Estimation of Factor Loadings

(3) Maximum likelihood estimation. The factor loading matrix and specific variance matrix obtained by the maximum likelihood method are:

$$A_3 = \begin{pmatrix} 0.93 & 0.36 \\ 0.98 & -0.21 \\ 0.97 & -0.22 \\ 0.97 & 0.19 \\ 0.94 & 0.28 \\ 0.95 & 0.00 \end{pmatrix}, \quad D_3 = \begin{pmatrix} 0.005 & 0 & 0 & 0 & 0 \\ 0 & 0.005 & 0 & 0 & 0 \\ 0 & 0 & 0.005 & 0 & 0 \\ 0 & 0 & 0 & 0.019 & 0 \\ 0 & 0 & 0 & 0 & 0.045 \\ 0 & 0 & 0 & 0 & 0.084 \end{pmatrix}$$

From the calculation results, it can be seen that the principal component method estimate and the principal factor method estimate are similar, while the maximum likelihood estimate differs somewhat from the first two estimates in the second factor loading vector.

Estimation of Factor Loadings

The factor variances, variance contribution rates, and cumulative variance contribution rates for the three methods are shown in table below. It can be seen that under the factor analysis principal component method, the cumulative variance contribution rate of the first two factors is the same as that of the first two principal components in principal component analysis, which is 98%.

	Principal component			Principal factor			Maximum likelihood		
Factor	Var.	Contr.	Cum.	Var.	Contr.	Cum.	Var.	Contr.	Cum.
1	5.57	0.93	0.93	5.54	0.92	0.92	5.500	0.917	0.917
2	0.31	0.05	0.98	0.30	0.05	0.97	0.338	0.056	0.973

Factor Orthogonal Rotation

The purpose of factor analysis is not only to find common factors, but more importantly to understand the practical meaning of each factor. The meaning of factors is related to the factor loading matrix. The loading matrices obtained by the previously introduced estimation methods may not satisfy the "simple structure criterion", meaning the typical representativeness of each common factor is not prominent, and their practical meaning is not clear. Therefore, it is necessary to perform rotation transformation on the factor loading matrix. There are many methods for factor rotation, including orthogonal rotation and oblique rotation. The most commonly used method is the Varimax orthogonal rotation method, which aims to polarize the absolute values of factor loadings in the factor loading matrix towards 0 and 1, achieving the purpose of simplifying its structure.

In practical calculations of factor analysis, after obtaining the initial factor loading matrix using the principal component method, principal factor method, or maximum likelihood method, it can be repeatedly right-multiplied by an orthogonal matrix, causing the elements of the column vectors to polarize in absolute value towards 0 and 1, thereby making the corresponding common factors have clearer practical meanings. This method of transforming the loading matrix is called orthogonal rotation of factor axes.

Factor Scores

The main purpose of factor analysis is to compress high-dimensional data by establishing the intrinsic relationship between high-dimensional data variables and low-dimensional factor variables through the factor model. In this process, obtaining clearly interpretable common factors through factor rotation has very important practical significance. Next, the original data can be analyzed using the obtained common factors and factor loading matrix, and the estimated values of common factors can be calculated for each sample, which are called factor scores. Factor scores can be used for model diagnosis and can also serve as a basis for analyzing or evaluating the original data.
Below are two commonly used methods for calculating factor scores.

Weighted Least Squares Method

Suppose x satisfies the orthogonal factor model $x = Af + \varepsilon$. Assuming the factor loading matrix A and specific variance matrix D are known, and treating the specific factor vector ε as the error vector of the model. Since the error variances are generally not equal, we use the weighted least squares method to estimate the value of the common factor f.

Factor Scores

The weighted error sum of squares is

$$Q(f) = \varepsilon^\top D^{-1}\varepsilon = (x - Af)^\top D^{-1}(x - Af),$$

where A, D are known, and x is the observable variable whose value is also known.
From $\frac{\partial Q(f)}{\partial f} = 0$, we obtain the estimate of f:

$$\hat{f} = (A^\top D^{-1} A)^{-1} A^\top D^{-1} x,$$

This is the weighted least squares estimate of the factor score, also known as the Bartlett factor score.
For sample $x_{(t)}$, its factor score is

$$\hat{f}_{(t)} = (A^\top D^{-1} A)^{-1} A^\top D^{-1} x_{(t)}, \quad t = 1, 2, \cdots, n.$$

The factor score matrix of the data matrix $X_{n\times p} = (x_{(1)}, \cdots, x_{(n)})$ is

$$F_{n\times m} = (\hat{f}_{(1)}, \cdots, \hat{f}_{(n)})^\top = XD^{-1}A(A^\top D^{-1}A)^{-1}.$$

Factor Scores

The factor score estimated by the regression method is:

$$\hat{f} = Bx = A'R^{-1}x.$$

Thus, the factor score for sample $x_{(t)}$ is:

$$\hat{f}_{(t)} = A'R^{-1}x_{(t)}, \quad t = 1, 2, \cdots, n.$$

The score matrix for the data matrix:

$$X_{n,p} = (x_{(1)}, \cdots, x_{(n)})^T$$

is:

$$F_{n,m} = (\hat{f}_1, \cdots, \hat{f}_m) = (\hat{f}_{(1)}, \cdots, \hat{f}_{(n)})^T = XR^{-1}A.$$

The regression method for calculating factor scores was first proposed by Thompson, hence this method is also called the Thompson method. To date, there is no consensus on which of the two methods for calculating factor scores is better.

Factor Scores

Here we introduce a comprehensive scoring calculation method, which computes the weighted average of each factor score. Among the m score vectors, if there is an inverse relationship between the component size of the i-th score vector $\hat{f}_i$ and the samples, i.e., the smallest score component corresponds to the "best" sample and the largest score component corresponds to the "worst" sample, then set $\hat{f}_i = -\hat{f}_1$, until the relationship between the component size of each vector in $\hat{f}_1, \ldots, \hat{f}_m$ and the samples is all in positive order. At this point, the comprehensive factor score vector is:

$$\hat{f}_c = \omega_1 f_1 + \cdots + \omega_m f_m,$$

where ω_i is the variance contribution rate of factor $\hat{f}_i$.

Example 5

Comprehensive Development Data of a Certain Region Over 20 Years

(1) Perform varimax rotation on the initial estimates of the factor loading matrix obtained using three methods in Example 4, explain the meaning of each factor based on the rotated factor loadings, and classify the variables through factor loading scatter plots.

(2) Provide the expressions of factor score functions based on the first two methods, and draw scatter plots of the first two factor score data based on the three methods respectively.

No.	x_1	x_2	x_3	x_4	x_5	x_6
1	7296.0	292.13	69.97	0.095	5.7	12208.0
2	7395.0	346.57	79.6	0.099	6.1	12682.0
3	7494.0	395.38	85.0	0.103	6.11	13109.0
4	7564.0	459.83	96.46	0.107	6.65	14839.0
5	7637.0	581.56	140.15	0.123	6.9	17309.0
6	7711.0	680.46	194.33	0.132	7.77	19772.0
7	7818.0	742.05	223.08	0.125	9.15	26459.0
8	7958.0	892.29	297.77	0.132	9.61	25209.0
9	8061.0	1117.66	369.82	0.162	9.96	29035.0
10	8160.0	1293.94	305.54	0.182	10.25	30718.0
11	8493.0	1511.19	335.66	0.186	10.05	29798.0
12	8570.0	1810.54	439.82	0.193	10.49	31931.0
13	8610.0	2196.53	601.5	0.205	10.8	33920.0
14	8642.0	2779.49	892.48	0.219	11.2	33634.0
15	8671.0	3872.18	1108.0	0.24	11.88	34467.0
16	8765.0	5002.34	1320.97	0.248	12.35	36425.0
17	8738.0	5960.42	1558.01	0.259	12.13	39199.0
18	8785.0	6650.02	1792.22	0.263	12.7	43218.0
19	8838.0	7162.2	2056.97	0.26	12.82	50558.0
20	8883.0	7662.1	2243.9	0.261	13.1	59350.0

(1) Varimax Rotation

The rotation matrix T_1 for principal component estimate A_1 and the rotated factor loading matrix $\hat{A}_1$ are:

$$T_1 = \begin{pmatrix} 0.7278591 & 0.6857267 \\ -0.6857267 & 0.7278591 \end{pmatrix}, \quad \hat{A}_1 = \begin{pmatrix} 0.88 & 0.46 \\ 0.49 & 0.87 \\ 0.48 & 0.87 \\ 0.78 & 0.60 \\ 0.85 & 0.50 \\ 0.71 & 0.67 \end{pmatrix}$$

The rotation matrix T_2 for principal factor estimate A_2 and the rotated factor loading matrix $\hat{A}_2$ are:

$$T_2 = \begin{pmatrix} 0.7281238 & 0.6854456 \\ -0.6854456 & 0.7281238 \end{pmatrix}, \quad \hat{A}_2 = \begin{pmatrix} 0.88 & 0.46 \\ 0.50 & 0.86 \\ 0.49 & 0.87 \\ 0.78 & 0.60 \\ 0.85 & 0.50 \\ 0.70 & 0.65 \end{pmatrix}$$

The rotation matrix T_3 for maximum likelihood estimate A_3 and the rotated factor loading matrix $\hat{A}_3$ are:

$$T_3 = \begin{pmatrix} 0.7467324 & 0.6651245 \\ -0.6651245 & 0.7467324 \end{pmatrix}, \quad \hat{A}_3 = \begin{pmatrix} 0.89 & 0.46 \\ 0.49 & 0.88 \\ 0.49 & 0.87 \\ 0.79 & 0.60 \\ 0.83 & 0.50 \\ 0.69 & 0.66 \end{pmatrix}$$

For comparison, the table below shows the variance, contribution rate, and cumulative contribution rate of factors before and after orthogonal rotation using the maximum likelihood method for initial estimation.

	Before rotation			After rotation		
Factor	Variance	Contr. rate	Cum. contr.	Variance	Contr. rate	Cum. contr.
1	5.500	0.917	0.917	3.050	0.508	0.508
2	0.338	0.056	0.973	2.787	0.465	0.973

Factor Scores

The factor analysis results show two clear dimensions:
The first factor, representing urbanization, loads highly on total population, per capita living area, urbanization rate, and passenger traffic.
The second factor, representing economic development, loads highly on GDP and fixed asset investment.
The scatter plot of factor loadings confirms this classification, showing distinct clusters for the variables associated with each factor.

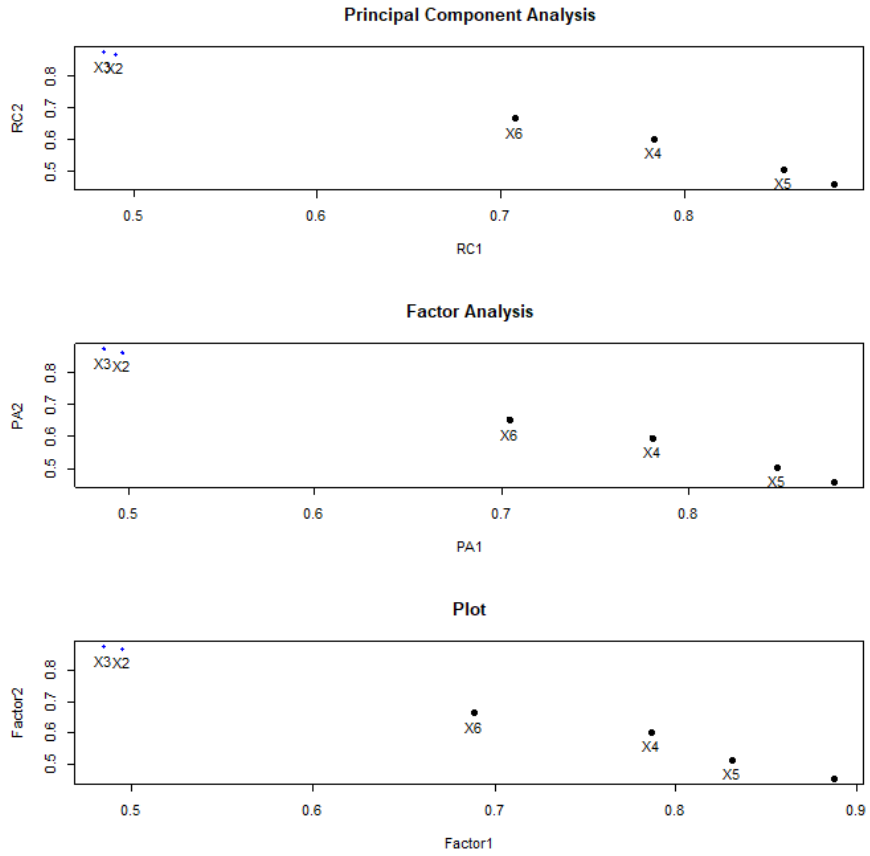


Figure: Factor loading scatter plots after orthogonal rotation

Factor Scores

(2) Factor Scores

The table on the right shows the sample factor scores calculated by the maximum likelihood method and their corresponding rankings (arranged from high to low), as well as the comprehensive scores calculated according to the formula:

$$f = 0.508f_1 + 0.465f_2$$

where the coefficients 0.508 and 0.465 are the variance contribution rates of the first and second factors, respectively. It can be seen that the order of comprehensive scores of each sample is consistent with the order of sample sizes of each variable. This example illustrates the rationality of the comprehensive ranking method because it reasonably integrates the directions of the two factors.

Sample	F1 Score	Order	F2 Score	Order	Comp Score	Order
1	-1.83835746	20	-0.002616359	7	-0.93510220	20
2	-1.58767444	19	-0.126007919	8	-0.86513230	19
3	-1.35172899	18	-0.246623604	9	-0.80135830	18
4	-1.16325508	17	-0.327749626	10	-0.74333716	17
5	-0.97146558	16	-0.364789535	11	-0.66313165	16
6	-0.77397976	15	-0.404999209	12	-0.58150635	15
7	-0.49572413	14	-0.525030538	14	-0.49596706	14
8	-0.19796727	13	-0.605005063	15	-0.38189473	13
9	0.08152153	12	-0.637014429	16	-0.25479877	12
10	0.40916225	8	-0.823676551	17	-0.17515517	11
11	1.13933311	3	-1.182207161	20	0.02905489	10
12	1.24955688	1	-1.092735518	19	0.12665288	9
13	1.23015135	2	-0.860838530	18	0.22462697	8
14	1.08012769	4	-0.410897908	13	0.35763734	7
15	0.92341773	5	0.101745226	6	0.51640774	6
16	0.86500417	6	0.553757874	5	0.69691953	5
17	0.51395250	7	1.160441791	4	0.80069330	4
18	0.40187861	9	1.565807204	3	0.93225469	3
19	0.27680594	10	1.961843771	2	1.05287477	2
20	0.20924094	11	2.266596085	1	1.16026158	1

Factor Scores

The right figure shows the factor scatter plots obtained by calculating initial factor loadings using three methods, performing orthogonal factor rotation using the variance maximization method, and calculating factor scores using regression methods. As can be seen from the figure, the factor score results obtained by the three methods are similar.

