STA 4702 Multivariate Statistical Methods

STA 5701 Applied Multivariate Methods

Factor Analysis

Outline

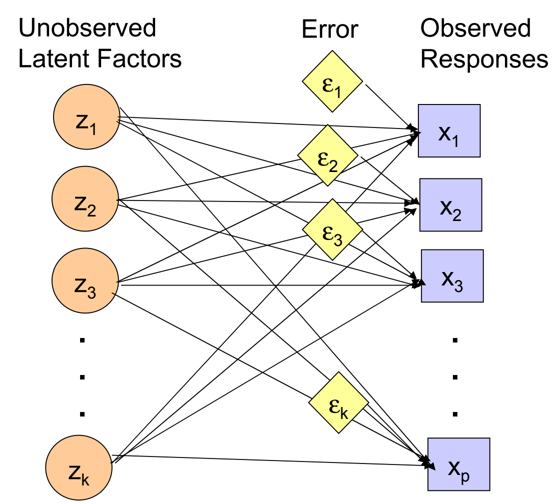
- Analysis Objectives
- Latent Factors.
- Factor Scores.
- Interpretation of Factors.
- Related Practical Issues.
- Rotation.
- An Application.

Analysis Objectives

- Similar to those of PCA.
- Data reduction.
- Identifying the internal relationships among a set of random variables.
- Defining a small set of easily interpretable linear combinations of the original variables, called FACTORS or LATENT FACTORS.
- Defining Latent Factors that can be interpreted as describing some underlying concept or condition whose expression is the observed variables.

Exploratory Factor Analysis = EFA

Conceptual Model



Assumption

- The total variance of a variable reflects the sum of three components.
  - Common Variance - That portion of total variance that correlates or is shared with other variables in the analysis (associated with latent factors).
  - Specific Variance - That portion of the total variance that does not correlate with the other variables (unexplained by latent factors).
  - Error Variance - The inherently unreliable random variation (total unexplained variability).

Comparison to PCA

- PCA assumes total variability can be factored into two components: Explained and Error variances.
- PCA finds eigenvectors that maximize the amount of total variance that can be explained.
- EFA finds factors that maximize the amount of the common variance that is explained.
- Both decompose the correlation matrix into constituent factors (or dimensions or sources of influence).
- PCA is used often in ecological/chemical research.
- EFA used primarily in social science research.

Important Concepts

- **Uniqueness of a variable** - That portion of the total variance that is unrelated to other variables (i.e. uniqueness = specific variance + error variance).
- **Communality** - An index of the portion of the variance of the variables that is accounted for by the set of <u>latent factors</u>.
  - The communality of a variable is equal to the sum over all factors of the squared factor loading for the variable.

PCA models

The PCA approach defines FACTORS as linear combinations of the observed variables.

$$
\begin{aligned}
z_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p = \mathbf{x}'\mathbf{a}_1 \\
z_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p = \mathbf{x}'\mathbf{a}_2 \\
&\vdots \\
z_p &= a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pp}x_p = \mathbf{x}'\mathbf{a}_p
\end{aligned}
$$

We can easily turn this around and define the observed variables as linear combinations of the factors.

$$
\begin{aligned}
x_1 &= b_{11}z_1 + b_{12}z_2 + \cdots + b_{1p}z_p = \mathbf{z}'\mathbf{b}_1 \\
x_2 &= b_{21}z_1 + b_{22}z_2 + \cdots + b_{2p}z_p = \mathbf{z}'\mathbf{b}_2 \\
&\vdots \\
x_p &= b_{p1}z_1 + b_{p2}z_2 + \cdots + b_{pp}z_p = \mathbf{z}'\mathbf{b}_p
\end{aligned}
$$

Factor Model

$$
\begin{aligned}
x_1 &= \mu_1 + b_{11}z_1 + b_{12}z_2 + \cdots + b_{1k}z_k + \varepsilon_1 \\
x_2 &= \mu_2 + b_{21}z_1 + b_{22}z_2 + \cdots + b_{2k}z_k + \varepsilon_2 \\
&\vdots \\
x_p &= \mu_p + b_{p1}z_1 + b_{p2}z_2 + \cdots + b_{pk}z_k + \varepsilon_p
\end{aligned}
$$

Specific Factors

Common Factors

$z_1 \quad z_2 \quad \cdots \quad z_k$

Factor Loadings

$b_{i1} \quad b_{i2} \quad \cdots \quad b_{ik}$

Find the smallest set of common factors such that the correlations among the components of x are completely accounted for by these factors.

Assumptions/Constraints

Matrix model: $x = \mu + Bz + \varepsilon$

Assumptions:

$$
\begin{aligned}
&E(z) = 0 \\
&E(\varepsilon) = 0 \\
&COV(z,\varepsilon) = 0 \\
&COV(\varepsilon) = \Psi \\
&\Psi = \mathrm{diag}(\psi_1, \psi_2, \cdots, \psi_p),\ \psi_p > 0 \\
&COV(z) = \Delta, \quad \Delta \text{ positive definite}
\end{aligned}
$$

$$COV(x) = COV(Bz) + COV(\varepsilon)$$

$$\Sigma = B\Delta B' + \Psi = BB' + \Psi$$

Specific Variance

Common Variance

Illustration

$$\boldsymbol{\Sigma = BB' + \Psi}$$

k=2, p=4

4 original, 2 latent

$$\Sigma_{4\times4} = \begin{bmatrix} 1.00 & .51 & .35 & .20 \\ .51 & 1.00 & .15 & .06 \\ .35 & .15 & 1.00 & .35 \\ .20 & .06 & .35 & 1.00 \end{bmatrix} \qquad B_{4\times2} = \begin{bmatrix} .8 & -.1 \\ .6 & -.3 \\ .5 & .5 \\ .3 & .4 \end{bmatrix}$$

$$\Psi_{4\times4} = \begin{bmatrix} .35 & 0 & 0 & 0 \\ 0 & .55 & 0 & 0 \\ 0 & 0 & .50 & 0 \\ 0 & 0 & 0 & .75 \end{bmatrix} \qquad BB' = \begin{bmatrix} .65 & .51 & .35 & .20 \\ .51 & .45 & .15 & .06 \\ .35 & .15 & .50 & .35 \\ .20 & .06 & .35 & .25 \end{bmatrix}$$

Communality and Specific Variance

Communality of z_1 $\quad h_1^2 = b_{11}^2 + b_{12}^2 = (.8)^2 + (-.1)^2 = 0.65$

$$B_{4\times2} = \begin{bmatrix} .8 & -.1 \\ .6 & -.3 \\ .5 & .5 \\ .3 & .4 \end{bmatrix} \qquad h = \begin{bmatrix} 0.65 \\ 0.45 \\ 0.50 \\ 0.25 \end{bmatrix}$$

An index of the portion of the variance of the x_i that is accounted for by the set of <u>latent factors</u>.

$$\Psi_{4\times4} = \begin{bmatrix} .35 & 0 & 0 & 0 \\ 0 & .55 & 0 & 0 \\ 0 & 0 & .50 & 0 \\ 0 & 0 & 0 & .75 \end{bmatrix}$$

Specific variance: a measure of the variability of x_i that CANNOT be explained by the latent factors.

Factor (Non) Uniqueness

The factor model is not unique, since any **orthogonal rotation** of the factors will produce the same decomposition of the covariance matrix.

Let Γ be a k by k orthogonal matrix and define a new set of factors by **z*** = Γ **'** **z**. Then the model can be rewritten as:

$$\begin{aligned} x &= \mu + B\Gamma\Gamma' z + \varepsilon \\ &\mu + B^* z^* + \varepsilon \end{aligned}$$

$$\begin{aligned} \Sigma &= BB' + \Psi \\ &= B\Gamma\Gamma' B' + \Psi \\ &= B^* B^{*\prime} + \Psi \end{aligned}$$

The orthogonal rotation does not change the overall covariance matrix, the specific variances nor the communalities.

Using Factor Non Uniqueness

The non-uniqueness of factor loadings can be used to our advantage. Once we obtain one set of factor loadings, we can orthogonally rotate the factors in an attempt to find linear combinations that are "easier to interpret."

Difference between PCA and EFA models

Both PCA (with k factors retained) and EFA can be written as the same model.

$$x = \mu + Bz + \varepsilon$$

In the EFA model we have:

$$\mathrm{COV}(\varepsilon) = \Psi = \mathrm{diag}(\psi_1, \psi_2, \cdots, \psi_p), \psi_p > 0$$

In the PCA model, this condition is not specified in that the matrix is not strictly a diagonal matrix.

Using k principal components to explain variation in x is not the same as using k latent factors in factor analysis.

EFA Issues

Sample Size: No less than 50 observations, better to have 100. Rule of thumb is to have at least 20 observations per variable.

Measurement Scales: Theory assumes continuous, in practice many dummy variables are used. If all variables are binary, special forms of factor analysis must be used.

Original or Standardized Variables: Factor analysis is traditionally done using standardized variables (i.e. assessing the correlation matrix), although it is possible to perform factor analysis on unstandardized values (i.e. using the covariance matrix). Standardized scores aid comparisons among different variables whereas unstandardized scores aid comparisons among different populations.

More Issues

Number or Variables: Minimize the number of variables considered while maintaining a reasonable number from which to derive factors. Identify key variables that, when combined with other variables in the linear function, will produce factors having a chance of being interpretable. Some degree of multi-collinearity among the original variables is desirable.

Impact of Outliers: Can seriously affect definitions of factors. Need to be identified and removed before analysis.

And how do we do this?

Number of Factors to Extract

Somewhat like focusing a microscope - too few or two many factors and the important structures in the data will not be clear.

Trial and error: Start with a predetermined goal then examine fits with +1 or -1 factors. Use multiple criteria.

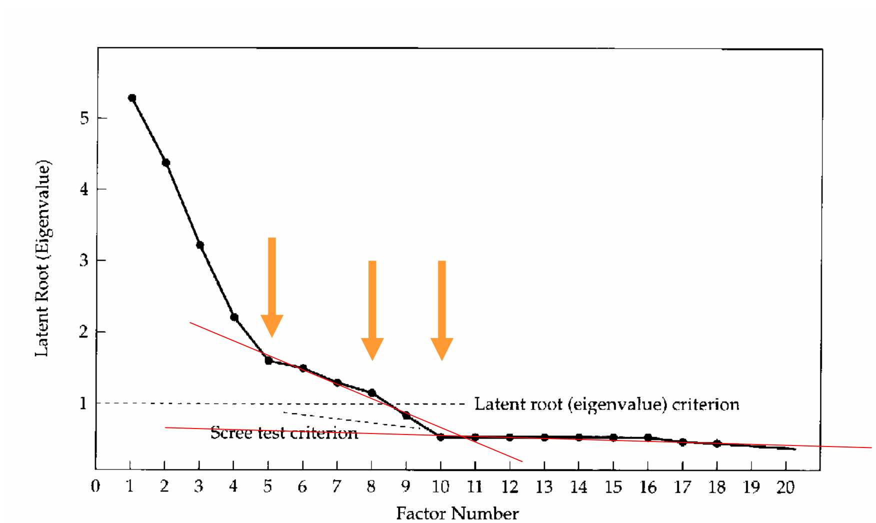
Latent Root Criterion: Only use factors having eigenvalues >1. Use fewer factors if number of variables greater than 50 and more if greater than 50.

Percentage of Variance Criterion: Continue adding factors until the fraction of explained variance exceeds some pre-specified level, say 95% in the natural sciences or 60% in the social sciences, or when the last factor added adds less than 5% (Go back one).

Scree Test Criterion: Examination of the SCREE Plot to identify the number of factors where the curve first begins to straighten out.

Heterogeneity of the Respondents: If there are obvious groups in the data, factors are added until last added factors no longer discriminate groups.

SCREE Plot



Fitting a Factor Model

How do we estimate the latent factor coefficients?

- Non-Iterative Approaches
  - Principal Components Method
  - Principal Factor Method
  - Image Analysis
  - Harris's Non-iterative Canonical Factor Analysis
- Iterative Approaches
  - Maximum Likelihood
  - Unweighted Least Squares
  - Iterative Principal Components
  - Alpha Factor Analysis

Principal Components Method

Select the first k<p principal components and from these determine the factor loadings. If these components explain most of the variation in $\mathbf{x}$, the difference between the observed correlation matrix and estimated correlation matrix should be acceptably small.

Singular Value Decomposition $\quad R = P\Lambda P' \quad \Lambda = \text{diag}\{\lambda_1, \lambda_2, \cdots, \lambda_p\}$

$\hat{B} = \left[\sqrt{\lambda_1}P_1, \cdots \sqrt{\lambda_k}P_k\right]$ Estimate of loadings matrix.

$\hat{\Psi} = \text{diag}\{R - \hat{B}\hat{B}'\}$ Estimate of specific variances.

$\text{Residual} = R - \left(\hat{B}\hat{B}' + \hat{\Psi}\right)$ What does not quite match up with assumptions.

These are small if these $\sum_{i=k+1}^{p}\lambda_i$ are small

$$\text{RMS\_overall} = \sqrt{\frac{1}{p-1}\sum_{i=1}^{p}\sum_{j\neq i}^{p}\text{res}_{ij}^2}$$

Principal factor method

Apply the principal components method to the correlation matrix (or covariance matrix) that has been modified to accommodate prior estimates of the commonalities obtained through a separate external analysis of the covariance matrix.

Prior specification of $\hat{\Psi} = \text{diag}\{\hat{\psi}_1 = 1 - \hat{h}_1^2, \cdots, \hat{\psi}_p = 1 - \hat{h}_p^2\}$

Estimate B by decomposing the matrix $R - \hat{\Psi} = \hat{B}\hat{B}'$

Where to get the prior estimates.

Let r^{ij} be the elements of R^{-1}. Then use

$$\hat{\psi}_i = \frac{1}{r^{ii}}$$

$$\hat{h}_i^2 = 1 - \frac{1}{r^{ii}}$$

Image Analysis

The common part, called the image, of a variable is defined as that part which is predictable by regressing each variable on all other variables. The anti-image is the specific part of the variable that cannot be predicted. First the covariance matrix of the image is computed and then incorporated with the principle components method to extract factors (Not related to GIS).

$M = \left[\text{diag}\left(R^{-1}\right)\right]^{-1}$

$\hat{x} = \left(I - MR^{-1}\right)x = Wx$ Prediction of x by all other x.

$x = Wx + (I - W)x$ = image + anti-image

Use the matrix WRW' as a starting point for the principal components method.

Harris' Non-iterative canonical factor analysis

A canonical correlation analysis of the hypothetical factor variables (common parts) with the observable variables.

Prior specification of $\hat{\Psi} = \text{diag}\{\hat{\psi}_1 = 1 - \hat{h}_1^2, \cdots, \hat{\psi}_p = 1 - \hat{h}_p^2\}$

Estimate B by decomposing the matrix $R - \hat{\Psi} = \hat{B}\hat{B}'$

R_a is the R matrix with diagonal elements replaced by $\hat{h}_i^2$

The first k positive eigenvalues λ_1 to λ_k and eigenvectors $P=\{P_1,\ldots, P_k\}$ of the matrix $\hat{\Psi}^{-1/2}R_a\hat{\Psi}^{-1/2}$ are used to estimate B as

$$\hat{B} = \hat{\Psi}^{1/2}P\hat{\Lambda}^{1/2}$$

$$\hat{\Psi}_{new} = R - \hat{B}\hat{B}'$$

$$\hat{\Lambda}^{1/2} = \text{diag}\left(\sqrt{\hat{\lambda}_1}, \cdots, \sqrt{\hat{\lambda}_k}\right)$$

Maximum Likelihood

Assumes a p-variate multivariate normal distribution for the $\mathbf{x}$ vector with covariance structure given as for the factor model. Traditional MLE (maximum likelihood estimation) is used to find the elements of **B** and $\boldsymbol{\Psi}$.

$$\max_{B,\Psi}\left(\frac{n-1}{2}\right)\left\{\operatorname{trace}\left[(BB'+\Psi)^{-1}S\right]-\ln\left|(BB'+\Psi)^{-1}S\right|\right\}$$

Subject to the constraint that the ψ_{ii} are positive.

This method has the advantage of a built-in likelihood-ratio test of whether the k-factor model is adequate.

A two-stage estimation process is available for more efficient determination of the MLE estimates.

Unweighted Least Squares

B and $\boldsymbol{\Psi}$ are estimated using iteration to minimize the least squares function below subject to the constraint that B'B is a diagonal matrix of order k.

$$\min_{B,\Psi}\{F_u(B,\Psi)\}=\min_{B,\Psi}\left[\frac{1}{2}\operatorname{trace}[(S-BB'-\Psi)(S-BB'-\Psi)]\right]$$

Essentially the same as the iterative two-stage MLE estimation method mentioned previously.

Iterated Principal Factor Method

Iterate the principal factor method until the change in the final communality estimates is smaller than a pre-specified value. Similar to unweighted least squares only $\mathbf{B'B}$ not constrained to being a diagonal matrix.

$$\hat{\psi}_i=\frac{1}{r^{ii}}$$
$$\hat{h}_i^2=1-\frac{1}{r^{ii}}$$

First specification of

$$\hat{\Psi}_1=\operatorname{diag}\left\{\hat{\psi}_1=1-\hat{h}_1^2,\cdots,\hat{\psi}_p=1-\hat{h}_p^2\right\}$$

Estimate $\mathbf{B}_1$ by decomposing the matrix: $R-\hat{\Psi}_1$

New estimate $\hat{\Psi}_2=\operatorname{diag}(R-\hat{B}_1\hat{B}_1{}')$

Alpha Factor Analysis

A psychometric method of doing factor analysis that seeks to determine common factors such that <u>observed variables have maximum correlation with the common factors</u>. The squared correlation used is call the **generalizability** or **Cronboch's Alpha** and measures the extent to which the variables (usually testing results) are generalizable to the population.

$$\hat{B}=H^{1/2}P\hat{\Lambda}^{1/2}\qquad \hat{\Lambda}^{1/2}=\operatorname{diag}\left(\sqrt{\hat{\lambda}_1},\cdots,\sqrt{\hat{\lambda}_k}\right)$$

Eigenvectors of   Eigenvalues of

$$\hat{H}^{-1/2}\left(R_a-\hat{\Psi}\right)\hat{H}^{-1/2}$$

$$H=\operatorname{diag}(R-\hat{\Psi})\qquad \hat{\psi}_i=\frac{1}{r^{ii}}\ \text{Initial estimates}\qquad \alpha_i=\frac{p}{p-1}\left\{1-\frac{1}{\lambda_i}\right\}\quad i=1,\cdots,k$$

Significance of Factor Loadings

What is an important or significantly non-zero loading?

Rule of Thumb: Factor loadings that meet the criteria below:

- Greater than ±.30 (9% variance explained) is considered minimally significant,
- Greater than ±.40 (16%) are more important, and,
- Greater than ±.50 (25%) are quite significant.

Correlation Tests: While standard tests for significance of the correlation coefficient can be used, because of the larger standard errors associated with factor loadings, significance is assumed only for loadings with associated very small p-values.

Sample Size and Factor Loadings

Based on Sample Size: What one considers significant depends on the size of the sample.

Factor Loading	Sample Size	Factor Loading	Sample Size
.30	350	.55	100
.35	250	.60	85
.40	200	.65	70
.45	150	.70	60
.50	120	.75	50

Small sample size requires that the loading has to be large.

Factor Rotation

Factor Rotation: If the factor loadings initially obtained are not readily interpretable, it is customary to transform the loadings by post-multiplying using an orthogonal matrix so that interpretable loadings may result.

What is our objective in rotation?

Have each factor, $\mathbf{z}$, highly correlated with just a few $\mathbf{x}$ and uncorrelated with the others.

$$\mathrm{COV}(x_i, z_j) = b_{ij}$$

Rotate so that a variable loads high on one factor and as low as possible on all other factors (**QUARTIMAX**).
Try to get factor scores close to -1, or 0 or +1 (**VARIMAX**).
Try to accomplish both at the same time (**EQUIMAX**).

Objective of Rotation

$$\begin{aligned}
\hat{x}_1 &= \mu_1 + b_{11}z_1 + b_{12}z_2 + \cdots + b_{1k}z_k \\
\hat{x}_2 &= \mu_2 + b_{21}z_1 + b_{22}z_2 + \cdots + b_{2k}z_k \\
&\vdots \\
\hat{x}_p &= \mu_p + b_{p1}z_1 + b_{p2}z_2 + \cdots + b_{pk}z_k
\end{aligned}$$

On each line, only a couple of the factor loadings (b_{ij}) are non-zero. Also, for example, if $\mathbf{b}_{.1}$ is non-zero on one row, $\mathbf{b}_{.2}$ to $\mathbf{b}_{.p}$ are zero.

$$\begin{aligned}
z_1 &= a_{11}x_1 + a_{12}x_2 + \cdots + a_{1p}x_p \\
z_2 &= a_{21}x_1 + a_{22}x_2 + \cdots + a_{2p}x_p \\
&\vdots \\
z_k &= a_{k1}x_1 + a_{k2}x_2 + \cdots + a_{kp}x_p
\end{aligned}$$

Thus each factor, $\mathbf{z}$, can be defined as a linear combination of just a few of the $\mathbf{x}$ variables.

Accomplished by defining a function measuring the variability of the loadings and searching for rotations that maximize this function.

a_{ij} are called the scores.

Quartimax, Raw Varimax, Equimax and Parsimax Rotation

Find that orthogonal rotation of the factor loadings that maximizes the variance of the squared loadings.

$$\max\left[\frac{1}{p}\sum_{j=1}^{k}\left\{\sum_{i=1}^{p} b_{ij}^{*4} - \frac{\gamma}{p}\left(\sum_{i=1}^{p} b_{ij}^{*2}\right)^2\right\}\right]$$

Simply a function of the loadings.

Quartimax rotation: $\gamma = 0$
Raw Varimax rotation: $\gamma = 1$
Equimax rotation: $\gamma = k/2$
Parsimax rotation: $\gamma = p(k-1)/(p+k-2)$

Varimax Rotation

Find that orthogonal rotation of the factor loadings that maximizes the normalized sum of the variance of the squared loadings.

$$\max\left[\sum_{j=1}^{k}\left\{\frac{1}{p}\sum_{i=1}^{p}\left(\frac{b_{ij}^{*2}}{h_i}\right)^2 - \left[\frac{1}{p}\sum_{i=1}^{p}\frac{b_{ij}^{*2}}{h_i}\right]^2\right\}\right]$$

Typically, this is the default rotation method in most computer packages because in practice it seems to provide a clearer separation among the factors and tends to be more invariant than other rotations (in simulation studies).

SAS Implemented Rotations

SAS does orthogonal rotation of the factor loadings that maximizes the normalized variance of the squared loadings.

$$\max\left[\frac{1}{p}\sum_{j=1}^{k}\left\{\sum_{i=1}^{p}\left(\frac{b_{ij}^{*2}}{h_i}\right)^2 - \frac{\gamma}{p}\left(\sum_{i=1}^{p}\frac{b_{ij}^{*2}}{h_i}\right)^2\right\}\right]$$

Quartimax rotation: $\gamma = 0$
Raw Varimax rotation: $\gamma = 1$
Equimax rotation: $\gamma = k/2$
Parsimax rotation: $\gamma = p(k-1)/(p+k-2)$

Oblique Rotations

If after rotation, the factors are still not easily interpretable, oblique or <u>non-orthogonal</u> rotations may be attempted. In this case, the final factors are no longer independent of each other.

$$B^* = BT$$

Harris and Kaiser (HK) rotation: The oblique transformation matrix is taken as the product of an orthogonal matrix and diagonal matrices. The diagonal matrices, D_1 and D_2, are chosen to correspond to acceptable levels of correlation among the final latent factors.

$$T = D_2 B D_1$$

You specify!
But How?

Promax Rotation

PROMAX rotation: A matrix Q(m), called the target matrix, is defined as a power (of order m) function of the (usually VARIMAX) rotated loadings.

SAS Default
m=3

$$q_{ij}(m) = \left|b_{ij}^{*}\right|^{m-1} b_{ij}^{*}$$

$$U = \left(B^{*\prime} B^{*}\right)^{-1} B^{*\prime} Q(m)$$
$$T = U\left\{\operatorname{diag}(U'U)^{-1/2}\right\}$$
$$B^{*} = BT$$

Factor Scores

Factor scores are the predicted values of the factors for each observation.

Weighted Least Squares Method: Assume the estimated values of B and Ψ are the true values, the factor scores for individual i are given by:

$$\hat{z}_i = (\hat{B}'\hat{\Psi}^{-1}\hat{B})^{-1}\hat{B}'\hat{\Psi}^{-1}(x_i - \bar{x})$$

Regression Method (SAS): Use the formula corresponding to the linear regression coefficient formula:

$$\hat{z}_i = \hat{B}'S^{-1}(x_i - \bar{x})$$

Bi-Plot of Factor Scores?

Other Use of Factor Analysis
Analysis of Regression Residuals

In some cases, a factor analysis may be performed on the residuals from a multivariate multiple regression.

Typically used when data need to first be adjusted for a set of covariates or a time trend must be removed ("partialed out").

Q and R Factor Analysis

R Factor Analysis: Analyzes relationships among variables to identify groups of variables forming latent dimensions. Uses correlations among the variables.

Q Factor Analysis: Analyzes relationships among respondents to form groups or clusters based on their similarity on a set of characteristics. Uses correlations among the cases or observations.

Q factor analysis is a type of **cluster analysis** in which clusters are formed based upon the intercorrelations among observations.

Traditional cluster analysis forms groups based on a distance-based similarity measure between the respondents' responses.

Summary

Factor Analysis

- Is a powerful and useful multivariate technique for extracting information from high dimensional data in large data sets.
- Extracts relationships not usually visible directly from viewing the data.
- Can be used to develop measures (latent factors) capable of representing a number of observed variables.

Summary(2)

Factor Analysis

- Is complex and has many techniques with associated controversy as to which technique is best.
- Incorporates a high degree of subjectivity in deciding the number of factors, factor rotations to use and factor interpretation.
- Has reliability issues that are unresolved - Small changes in a data set can lead to big changes in the form of the latent factors computed.
- Provides no guarantee of the validity, stability or plausibility of results.