Estimation of Factor Loadings and Factor Scores

The Factor Analysis Model

$$\begin{Bmatrix} X_1 &=& \lambda_{11}\xi_1 + \lambda_{12}\xi_2 + \cdots + \lambda_{1m}\xi_m &+& \varepsilon_1 \\ X_2 &=& \lambda_{21}\xi_1 + \lambda_{22}\xi_2 + \cdots + \lambda_{2m}\xi_m &+& \varepsilon_2 \\ X_3 &=& \lambda_{31}\xi_1 + \lambda_{32}\xi_2 + \cdots + \lambda_{3m}\xi_m &+& \varepsilon_3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ X_p &=& \lambda_{p1}\xi_1 + \lambda_{p2}\xi_2 + \cdots + \lambda_{pm}\xi_m &+& \varepsilon_p \end{Bmatrix}$$

The Factor Analysis Model (FAM) specifies that each of *p* manifest variables $X_1, X_2, ..., X_p$ is a linear combination of $m < p$ common factors $\xi_1, \xi_2, ..., \xi_m$ and a single unique factor ε. Under the standard assumptions of FAM, the common factors are all uncorrelated with each other and standardized. The unique factors are all uncorrelated with each other and with all common factors, but are only centered (mean 0), rather than completely standardized.

The manifest variables $X_1, X_2, ..., X_p$ are all measured or observed for each of *n* individuals. So there are *n* observations on each equation. Therefore, the data consist of *pn* values $\{X_{1k}, X_{2k}, ..., X_{pk}, k = 1, ..., n\}$. For the *n* individuals, the FAM can be written:

$$\begin{Bmatrix} X_{1k} &=& \lambda_{11}\xi_{1k} + \lambda_{12}\xi_{2k} + \cdots + \lambda_{1m}\xi_{mk} &+& \varepsilon_{1k} & ,k = 1,...,n \\ X_{2k} &=& \lambda_{21}\xi_{1k} + \lambda_{22}\xi_{2k} + \cdots + \lambda_{2m}\xi_{mk} &+& \varepsilon_{2k} & ,k = 1,...,n \\ X_{3k} &=& \lambda_{31}\xi_{1k} + \lambda_{32}\xi_{2k} + \cdots + \lambda_{3m}\xi_{mk} &+& \varepsilon_{3k} & ,k = 1,...,n \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_{pk} &=& \lambda_{p1}\xi_{1k} + \lambda_{p2}\xi_{2k} + \cdots + \lambda_{pm}\xi_{mk} &+& \varepsilon_{pk} & ,k = 1,...,n \end{Bmatrix}$$

Neither the common factors nor the unique factors can be observed or measured. The common factors and the unique factors are *hidden*, *latent*, or *unobservable*. Nevertheless, values can be estimated for the common and unique factors! The purpose of this document is to discuss how this can be done. There are two steps. The first is the estimation of the factor coefficients (the $\lambda's$), also called the **factor loadings**. For many purposes in FA, that is enough and we can stop there. Estimating the $\lambda's$ can be done using only the correlations among the manifest variables. The second step is to supply estimates of the **factor scores** $(\xi_{1k}, \xi_{2k}, \cdots, \xi_{mk})$ for each individual *k*=1, 2, …, *n*. Once the factor coefficients (the $\lambda's$) have been estimated in step one, the factor scores can be estimated in step two by displaying the formula for the regression of *actual* scores on manifest variables and substituting the estimates from step one. I hasten to add that this regression does not actually need to be run. In any case that would be impossible because *actual* scores are unobservable. However, all that is needed is the *formula* for this regression, which turns out to be a function of the just-calculated factor loadings.

Step 1: Estimation of factor coefficients

Since the manifest variables are observable, their inter-correlations can be calculated. Intuitively, if the correlation between two manifest variables is high, then both variables must be strongly feeling the effects of the common factors. That is, the larger the correlation between two manifest variables, the larger the common factor part and the smaller the unique part. The weaker the correlation, the smaller the common factor part and the larger the unique part. Thus, examination of the pattern of correlations among the manifest variables may provide a basis for estimating the magnitudes of the factor coefficients. This intuition does not tell us how to estimate the factor coefficients, but it provides hope that it may be possible. In fact, estimation of factor coefficients can be done solely with the set of correlations among all manifest variables. I will now explain how this is done.

Let $\mathbf{R}_{pxp} = \begin{bmatrix} corr(X_1,X_1) & corr(X_1,X_2) & \cdots & corr(X_1,X_p) \\ corr(X_2,X_1) & corr(X_2,X_2) & \cdots & corr(X_2,X_p) \\ \vdots & \vdots & \ddots & \vdots \\ corr(X_p,X_1) & corr(X_p,X_2) & \cdots & corr(X_p,X_p) \end{bmatrix}$ denote the theoretical correlation matrix of the manifest variables, according to the FAM. It is possible to write each of these correlations solely in terms of the factor coefficients:

Now, the correlation of two standardized variables = their covariance. So $corr(X_i,X_j) = \text{cov}(X_i,X_j) = \text{cov}(\lambda_{i1}\xi_1 + \lambda_{i2}\xi_2 + \cdots + \lambda_{im}\xi_m + \varepsilon_i, \lambda_{j1}\xi_1 + \lambda_{j2}\xi_2 + \cdots + \lambda_{jm}\xi_m + \varepsilon_j)$, in which the FAM expressions have been substituted for X_i and X_j. Let us apply to this covariance a general formula from mathematical statistics for the covariance of two linear combinations: $\text{cov}(a_1A_1 + \ldots + a_hA_h, b_1B_1 + \ldots + b_hB_h) = \sum_{i=1}^{h}\sum_{j=1}^{h} a_ib_j \,\text{cov}(A_i,B_j)$. Since common and unique factors are uncorrelated except for identical factors, then $\text{cov}(\lambda_{i1}\xi_1 + \lambda_{i2}\xi_2 + \cdots + \lambda_{im}\xi_m + \varepsilon_i, \lambda_{j1}\xi_1 + \lambda_{j2}\xi_2 + \cdots + \lambda_{jm}\xi_m + \varepsilon_j) = \lambda_{i1}\lambda_{j1} + \lambda_{i2}\lambda_{j2} + \cdots + \lambda_{im}\lambda_{jm}$ if $i \neq j$ and $corr(X_i,X_i) = \lambda_{i1}^2 + \lambda_{i2}^2 + \cdots + \lambda_{im}^2 + \text{var}(\varepsilon_i) = 1$. That is, for off-diagonal terms in $\mathbf{R}_{pxp}$, the entry in row i, column j is the inner product of $(\lambda_{i1},\lambda_{i2},\cdots,\lambda_{im})$ and $(\lambda_{j1},\lambda_{j2},\cdots,\lambda_{jm})$, and the on-diagonal terms add in the variance of the unique factor.

That is, $\mathbf{R}_{pxp} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{pm} \end{bmatrix} \begin{bmatrix} \lambda_{11} & \lambda_{21} & \cdots & \lambda_{p1} \\ \lambda_{12} & \lambda_{22} & \cdots & \lambda_{p2} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{1m} & \lambda_{2m} & \cdots & \lambda_{pm} \end{bmatrix} + \begin{bmatrix} \text{var}(\varepsilon_1) & 0 & \cdots & 0 \\ 0 & \text{var}(\varepsilon_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \text{var}(\varepsilon_p) \end{bmatrix}$

or

$$\mathbf{R}_{pxp} = \mathbf{\Lambda}_{pxm}\mathbf{\Lambda}^{\mathrm{T}}{}_{pxm} + V(\boldsymbol{\varepsilon}),$$

where $\mathbf{\Lambda}_{pxm}$ is the matrix of factor coefficients in the FAM:

$\boldsymbol{\Lambda}_{pxm} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{pm} \end{bmatrix}$ and $V(\boldsymbol{\varepsilon}) = \begin{bmatrix} \text{var}(\varepsilon_1) & 0 & \cdots & 0 \\ 0 & \text{var}(\varepsilon_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \text{var}(\varepsilon_p) \end{bmatrix}$ and superscript T denotes transpose. This representation is a decomposition of the manifest variable correlation matrix into a **reduced correlation matrix** $\boldsymbol{\Lambda}_{pxm}\boldsymbol{\Lambda}^{\mathrm{T}}{}_{pxm}$, which is equal to the actual correlation matrix, except for the diagonal elements, and the **uniqueness matrix** $V(\boldsymbol{\varepsilon})$.

Origin of the name *factor analysis*

The expression $\mathbf{R}_{pxp} = \boldsymbol{\Lambda}_{pxm}\boldsymbol{\Lambda}^{\mathrm{T}}{}_{pxm} + V(\boldsymbol{\varepsilon})$ explains the origin of the term "factor analysis" for this methodology, because the correlation matrix is "factored" into a matrix $\boldsymbol{\Lambda}_{pxm}$ and its transpose (plus the variance matrix of the unique factors). This factorization permits the estimation of factor coefficients (and subsequently, in Step 2, factor scores) from the manifest data. Factor analysis estimates $\boldsymbol{\Lambda}_{pxm} = \begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{p1} & \lambda_{p2} & \cdots & \lambda_{pm} \end{bmatrix}$ to make the theoretical correlations in $\mathbf{R}_{pxp}$ as close as possible to the empirically calculated correlations. Denote the empirical correlation matrix, as calculated from the actual data, by $\hat{R}_{pxp} = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1p} \\ r_{21} & r_{22} & \cdots & r_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ r_{p1} & r_{p2} & \cdots & r_{pp} \end{bmatrix}$.

Factor analysis uses a least-squares estimation method to make the *off-diagonal* elements of the reduced correlation matrix $\boldsymbol{\Lambda}_{pxm}\boldsymbol{\Lambda}^{\mathrm{T}}{}_{pxm}$ as close as possible to the corresponding empirical correlations in $\hat{R}_{pxp}$. In some cases, it is possible to reproduce the manifest correlations exactly. As you can imagine, getting the manifest correlations exactly correct is unusual and applies only rarely. For example, with one common factor, the off-diagonal correlation matrix of three manifest variables can be matched exactly (barely), but that of four manifests cannot. In this regard, note that the number of off-diagonal terms in the correlation matrix grows in proportion to the square of the number of manifest variables. However, the number of λ coefficients grows only linearly with the number of manifest variables. Therefore, with a given number of common factors, the correlation matrix cannot be matched exactly once the number of manifest variables exceeds a small number. In fact, the number of off-diagonal terms in the empirical correlation matrix is $p(p\text{-}1)/2$. The number of distinct λ coefficients is mp. If $p(p\text{-}1)/2 \leq mp$, that is, if $p \leq 2m + 1$, then there are at least enough λ coefficients to match all empirical correlations exactly.

Least-squares estimation of factor coefficients

What optimization procedure does FA employ to make the estimated manifest correlation matrix as close as possible to the observed correlation matrix? FA seeks to minimize the sum of squared differences between the empirically observed off-diagonal elements and the theoretical off-diagonal elements.[1] The minimization criterion is

$$\min_{\lambda's} \sum_{i \neq j} \sum [r_{ij} - (\lambda_{i1}\lambda_{j1} + \lambda_{i2}\lambda_{j2} + \cdots + \lambda_{im}\lambda_{jm})]^2$$

This least-squares optimization is best left to statistical software.

The main take-away of all of the above discussion is that the implausible program of estimating factor coefficients when the factors cannot be observed can actually be carried out.

Rotation of solution

Note that the solution for $\mathbf{\Lambda}_{pxm}$ is not unique: Suppose $\mathbf{\Lambda}_{pxm}$ achieves the minimum of the minimization criterion. Let $\mathbf{P}_{mxm}$ be any orthonormal matrix. Then $\mathbf{\Lambda}_{pxm}\,\mathbf{P}_{mxm}$ also achieves the minimum, since $\mathbf{\Lambda}_{pxm}\mathbf{\Lambda}^{\mathrm{T}}{}_{pxm} = \mathbf{\Lambda}_{pxm}\mathbf{P}_{mxm} \cdot \mathbf{P}^{T}_{mxm}\mathbf{\Lambda}^{T}_{pxm}$ yields the same reduced correlation matrix. Thus there is a fundamental indeterminacy in the FA solution, and an FA method must specify a particular one. This is actually quite a useful property, for it says that we may "rotate" any initial factor solution by any orthonormal transformation that we choose in order to gain more insight (better interpretation), and still reproduce the observed correlations as closely as the original solution. That is, an orthonormally rotated solution loses no explanatory power to account for the observed manifest variables, but may gain interpretability or insight.

Step 2: Estimation of factor scores

Factor scores are calculated for individual observations. Factor scores are necessary only if it is desired to score individuals (people, companies, products, etc.) to see where they fall on the dimensions defined by the *m* common factors. This could be useful to help interpret the meaning of a factor or to evaluate individual performance on it.

The general approach

The idea for estimating factor scores is to do a regression of the factors on the manifest variables and predict the factor scores by plugging the observed values of the manifest variables into those regressions (one regression for each factor). However, you immediately see that there is a problem with this idea: We cannot run the contemplated regressions! There are no known factor scores for the proposed dependent variable! They are what we are trying to estimate – and if we had them to use in the regression, we would not need to estimate them!

The way around this problem is to realize that the formulas for the regression equations that we need can be written down in terms of the $\lambda's$ that we estimated in Step 1. Factor score values are not necessary to get the formulas. The estimates from Step 1 can then be substituted into Step 2.

[1] You may be interested to know that, by comparison, PCA minimizes the sum of squared differences for *all* elements of the correlation matrix, including diagonal terms.

To see that this is really possible, think about simple linear regression. The slope is $\rho_{y,x}\frac{\sigma_y}{\sigma_x}$. We did not need any data to write down this formula. If we have estimates of the correlation and the two standard deviations, we do not need to run the regression. The actual data are needed only to provide those estimates. If we can get the estimates otherwise, then we do not need the data at all.

In the general case, the regression equations that result from regressing each of a number of response variables (call them U's) on a number of predictor variables (call them V's) are the conditional means of the U's given the V's. For example, in simple linear regression, the conditional mean of Y given X is $\mu_y - \rho_{y,x}\frac{\sigma_y}{\sigma_x}(x-\mu_x)$, which is the regression equation. In the general case, there is a theorem in mathematical statistics that provides the coefficients of the regression equations: The coefficients of the predictor V's are the rows of the matrix $\mathbf{CS}^{-1}$, in which C is the matrix of covariances between the U's and V's, and **S** is the covariance matrix between the V's and themselves. For example, in the case of simple linear regression, $\mathbf{C} = \mathbf{C}_{1x1} =$ Cov(Y,X) = $\rho_{y,x}\sigma_y\sigma_x$ and $\mathbf{S} = \mathbf{S}_{1x1} = \sigma_x^2$. So $\mathbf{CS}^{-1} = \rho_{y,x}\sigma_y\sigma_x\frac{1}{\sigma_x^2} = \rho_{y,x}\frac{\sigma_y}{\sigma_x}$.

Details

In our case, we want the coefficients from regressing $(\xi_1,\xi_2,\cdots,\xi_m)$ as responses on $(X_1,X_2,...,X_p)$ as predictors. Thus, $\mathbf{C} = \mathbf{C}_{mxp} = Cov((\xi_1,\xi_2,\cdots,\xi_m),(X_1,X_2,...,X_p))$. Now the general term in row *i*, column *j* of $\mathbf{C}_{mxp}$ is $Cov(\xi_i,X_j)$ = (substituting the FAM expression for X_j) $Cov(\xi_i,\ \lambda_{j1}\xi_1+\lambda_{j2}\xi_2+\cdots+\lambda_{jm}\xi_m+\varepsilon_j)$ = (by uncorrelated assumptions of the FAM) $Cov(\xi_i,\lambda_{ji}\xi_i)=\lambda_{ji}Cov(\xi_i,\xi_i)=\lambda_{ji}$. Thus $\mathbf{C}_{mxp} = \mathbf{\Lambda}_{pxm}{}^{T}$. Furthermore, $\mathbf{S} = \mathbf{S}_{pxp} =$ $Cov((X_1,X_2,...,X_p),(X_1,X_2,...,X_p)) = \mathbf{R}_{pxp} = \mathbf{\Lambda}_{pxm}\mathbf{\Lambda}^{\mathrm{T}}{}_{pxm} + V(\boldsymbol{\varepsilon})$. Thus the coefficients from regressing $(\xi_1,\xi_2,\cdots,\xi_m)$ on $(X_1,X_2,...,X_p)$ are the rows of $\mathbf{CS}^{-1} = \mathbf{C}_{mxp}\ \mathbf{S}_{pxp}{}^{-1} =$ $\mathbf{\Lambda}^{\mathrm{T}}{}_{pxm}\left(\mathbf{\Lambda}_{pxm}\mathbf{\Lambda}^{\mathrm{T}}{}_{pxm}+V(\boldsymbol{\varepsilon})\right)^{-1}$. Since $(\xi_1,\xi_2,\cdots,\xi_m)$ and $(X_1,X_2,...,X_p)$ are assumed to be standardized, then the regression intercepts are all zeroes. Thus, the *m* factor scores of individual *k* can be obtained as the inner product of the rows of $\mathbf{\Lambda}^{\mathrm{T}}{}_{pxm}\left(\mathbf{\Lambda}_{pxm}\mathbf{\Lambda}^{\mathrm{T}}{}_{pxm}+V(\boldsymbol{\varepsilon})\right)^{-1}$ with a column vector $\mathbf{x}_{px1}^{(k)}$ containing the *p* values of individual *k* on the manifest variables. That is,

$$\text{Factor scores for individual } k = \mathbf{\Lambda}^{\mathrm{T}}{}_{pxm}\left(\mathbf{\Lambda}_{pxm}\mathbf{\Lambda}^{\mathrm{T}}{}_{pxm}+V(\boldsymbol{\varepsilon})\right)^{-1}\mathbf{x}_{px1}^{(k)}.$$

Since estimates of $\mathbf{\Lambda}_{pxm}$ were obtained in Step 1, those estimates can be substituted in the above to provide estimates of factor scores. $V(\boldsymbol{\varepsilon})$ is a diagonal matrix with the uniquenesses on the diagonal. The uniqueness of factor *i* = 1 – communality of $\xi_i = 1-(\lambda_{i1}{}^2+\lambda_{i2}{}^2+\cdots+\lambda_{im}{}^2)$. Again, this quantity can be estimated using the results of Step 1.

Summary

In the Factor Analysis Model, the value of manifest variable *i* for individual *k* is a combination of factor loadings $(\lambda_{i1}, \lambda_{i2}, \cdots, \lambda_{im})$ for the variable and factor scores $(\xi_{1k}, \xi_{2k}, \cdots, \xi_{mk})$ for the individual:

$$X_{ik} = \lambda_{i1}\xi_{1k} + \lambda_{i2}\xi_{2k} + \cdots + \lambda_{im}\xi_{mk} + \varepsilon_{ik}$$

Only the manifest variables are observed. Neither loadings nor scores are observed. Nevertheless, both loadings and scores can be estimated. The key idea is to separate the estimation of loadings from scores and to estimate the loadings first.

Estimation of loadings can be done separately from scores for two reasons:

1. The sample correlations between manifest variables can be empirically computed;
2. The theoretical correlation between manifest variables is a function only of the loadings:

$$corr(X_i, X_j) = \lambda_{i1}\lambda_{j1} + \lambda_{i2}\lambda_{j2} + \cdots + \lambda_{im}\lambda_{jm} \text{ if } i \neq j$$

Therefore, the least-squares squared difference $\min_{\lambda's} \sum_{i \neq j} \sum [r_{ij} - (\lambda_{i1}\lambda_{j1} + \lambda_{i2}\lambda_{j2} + \cdots + \lambda_{im}\lambda_{jm})]^2$ between this form and the empirical correlations can be minimized to yield estimates of factor loadings.

Having estimates of the loadings now helps in estimating the scores: Conceptually, the scores may be estimated by regressing the true factor scores on the manifest variables. (Yes, you read that right!) This gives a regression equation that predicts factor scores from manifest values. This procedure would be hopeless if you actually had to do that regression because you lack the factor scores to use as response variables. But all we really need is the formula for that regression – which can be calculated. The formula depends upon the factor loadings that we just estimated. Plug the values $\mathbf{x}^{(k)}_{px1}$ of the manifest variables for individual *k* into the formula to yield $\mathbf{\Lambda}^{\mathrm{T}}{}_{pxm}\left(\mathbf{\Lambda}_{pxm}\mathbf{\Lambda}^{\mathrm{T}}{}_{pxm} + V(\boldsymbol{\varepsilon})\right)^{-1}\mathbf{x}^{(k)}_{px1}$ as the factor scores for individual *k*.