Factor Analysis (FA)

**Introduction:** Factor Analysis (FA) is a method for modeling observed variables, and their covariance structure in terms of a smaller number of underlying unobservable factors. We model the observed variables as a linear combination of a few random variables called *factors*, *constructs* or *latent* variables. Unlike original variables , these factors cannot be measured or observed.

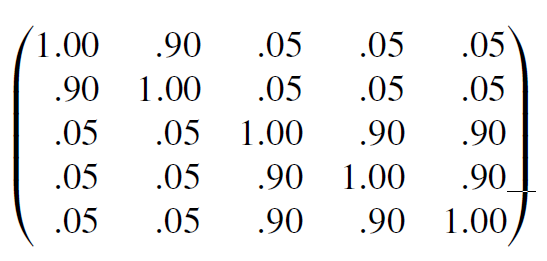
Factor analysis is generally an exploratory/descriptive method that requires many subjective judgments. It is a widely used tool and often controversial because the models, methods, and subjectivity are so flexible that debates about interpretations can occur.

Examples of fields where factor analysis is involved include physiology, health, intelligence, sociology, ecology, etc.

The goal of factor analysis is to reduce the redundancy among the observed variables by using a smaller number of factors. For example, people’s respond to questions about yearly income, education level, occupation, job satisfaction, amount of savings, living status (home owner or renter), marital status, race, gender, age etc. tend to be similar and these variables are associated to latent variables (or factors) ‘socioeconomic status’ and ‘demography’. If we have a subset of y variables that are highly correlated among themselves and have low correlations with all other variables, then the highly correlated variables can be explained or described by some factors.

Factor analysis is a method of dimensionreduction. The basic assumption of factor analysis is that for a collection of observed variables there are a set of **factors** (smaller than the observed variables), that can explain the interrelationships among these variables.

Suppose we have a correlation matrix of five observed variables that has the following form:



Note that variables 1 and 2 seem to have high correlations whereas variables 3, 4, and 5 seem to have high correlations among themselves. Then variables 1 and 2 correspond to a factor, and variables 3, 4, and 5 correspond to another factor.

**Differences in Factor analysis and Principal component analysis**: There are two differences in basic approaches in Factor analysis (FA) and Principal component analysis (PCA):

a. In principal component analysis, we create new variables that are linear combinations of the observed variables. In factor analysis, we model the observed variables as linear functions of the “factors.”  Thus, in one sense, factor analysis is an inversion of principal component analysis. In both PCA and FA, the dimension of the data is reduced.

b. In PCA, we seek for the principal components that explains the most of the variation in the observed variables. In FA, we seek for the factors that account the covariance or correlation structure among the observed variables.

**Orthogonal Factor Model**

We assume a random sample from a population with mean vector and covariance matrix . Our factor model can be thought of as a series of multiple regressions, predicting each of the observable variables from the values of the unobservable common factors , with an accompanying error term that is unique to that variable:

….. ….…(1)

Here, the variable means through can be regarded as the intercept terms for the multiple regression models.

Ideally, we want . The coefficients are called *loadings* and serve as weights. It shows how each depends on the ’s. For example, denotes the importance of factor to the th variable . For a specific , the larger loadings of (in absolute value) relate to the corresponding ’s. From these ’s, we infer a meaning or description of . We describe or interpret for example, by examining it coefficients . The larger loadings of relate to the corresponding ’s. From these ’s, we infer a meaning or description of .

**Model Assumptions**:

Mean

1. The random errors have mean zero: for .
2. The common factors have mean zero: for

The consequence of these assumptions is that the mean response of the ith variable is

Variance

1. The random errors have specific variance: for .
2. The common factors have variance one: for .

Covariance/Correlation

1. The random errors are uncorrelated with one another:
2. The common factors are uncorrelated with one another:
3. The random errors are uncorrelated with the common factors: , ;

These assumptions are necessary to estimate the parameters uniquely. An infinite number of equally well-fitting models with different parameter values may be obtained unless these assumptions are made.

With all the above assumptions, we get

Thus, the variance of variable has two components: the communality is also referred to as the common variance (the variance explained by the common factors ) and the specific variance is also called specificity, noise, unique variance or residual variance.

Model (1) can be written in matrix notation as

where , , , , and

|  |  |
| --- | --- |
|  | ………..(2) |

For instance, with variables and factors, the model in (1) becomes

which is in matrix notation becomes

…..(3)

Or,

The assumptions listed above can be expressed in vector and matrix notation as:

For , becomes ;

and for gives , an identity matrix;

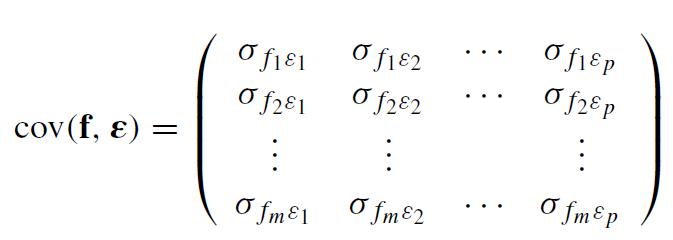
For , becomes ;

and for become

|  |  |
| --- | --- |
|  | ……(4) |

And for all and becomes . [note that denotes the scalar covariance value between two variables and , whereas, denotes the covariance matrix of two random vectors and of random variables.]

The notation indicates an rectangular matrix containing the covariances of the ’s and **’s.**

****

As noted earlier, the emphasis of FA is to explain the variances and covariances among variables in terms of a simplified structure involving the loadings and specific variances ; that is, we wish to express the population covariance matrix of in terms ofand.

From , we can write

Or,

Since and are uncorrelated and is a matrix of constant terms, we can write

Or,

Note that in equation (5), the covariances of ’s are modeled only by since is a diagonal matrix.For example, with variables and factors, , where and are first and second row of , respectively, and .

Now, let us find the covariance of ’s with ’s in terms of ’s.

Consider, for example, . From (1) we have,

Since is uncorrelated with all other ’s and for all and ,

Similarly,

Hence, the loadings themselves represent the covariance of the variables with the factors. In general,

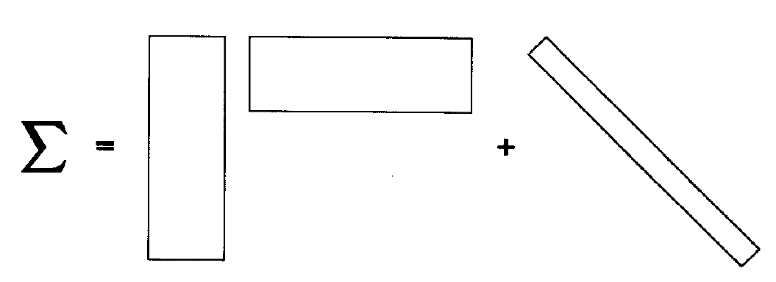
, , . …….(6)

We can write this in the form .

If standardized variables are used, then the correlation matrix can be expressed as , and the loadings become correlations between the variables and the factors:

, , . …….(7)

In schematic form, has the following appearance:



**Checking Validity of Assumptions**: One advantage of FA model is that when it does not fit the data, the estimate of clearly reflects this failure. In such cases, there are two problems in the estimate: (1) it is unclear how many factors there should be, and (2) it is unclear what the factors are. In other statistical procedures, failure of assumptions may not lead to such obvious consequences in the estimation or tests. In FA, the assumptions are essentially self-checking, whereas, in other procedures, we typically have to check the assumptions with residual plot, formal tests, and so on.

**Estimation of Loadings and Communalities**: There are many different approaches to estimation of the loadings and communalities, for example, principal component method, principal axis or factor method, iterated principal factor method, maximum likelihood method, generalized least squares, etc. Here we will present the principal component method.

*Principal Component method*: From a random sample , first we obtain the sample covariance matrix and then we try to find an estimator that will approximate the fundamental expression with in place of :

…….(8)

In the principal component approach, first we neglect and factor .

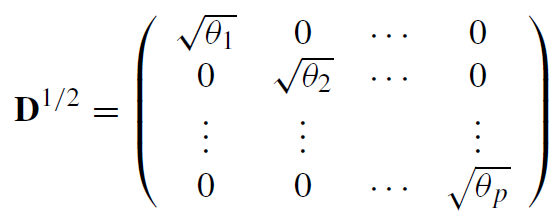
In order to factor , we use the spectral decomposition: …… (9)

where is the orthogonal matrix constructed with normalized eigenvectors of as columns and is a diagonal matrix with the eigenvalues of on the diagonal:

|  |  |
| --- | --- |
|  | …….(10) |

Since the eigenvalues of the positive semidefinite matrix are all positive or zero, we can factor **D** into

, where is the square root matrix:



With this factoring of , (9) becomes

This is of the form , but we do not define to be because is matrix and we are seeking that is with . We, therefore, define with largest eigenvalues of and containing the corresponding eigenvectors. We then estimate by the first columns of , , where is , is and is .

We illustrate the structure on the for and :

|  |  |
| --- | --- |
|  | ………..(12) |

Note that the diagonal element of is the sum of squares of the row of , or . Hence to complete the approximation of in (8), we write

Or, ………(13)

In this method of estimation, the sums of squares of the rows of is equal to communalities, and sums of squares of the columns of is equal to the eigenvalue. This can be easily shown.

The sum of squares of the row of is

, which is the estimated communality and is denoted by . Thus,

…….(14).

The sum of squares of the column of is the eigenvalue of :

since the normalized eigenvectors (columns of ) have length 1.

By (13) and (14), the variance of the variable is partitioned into a part due to the factors and a part due uniquely to the variable:

Thus the factor contributes to .

Note that the total sample variance is .

The total variance due to the factor is the sum of squares of the column of

……(16)

which is the eigenvalue by (15).

The proportion of total sample variance due to the factor is, therefore,

If the variables are not commensurate (meaning they do not have same measurement units), we can use standardized variables and work with the correlation matrix . The eigenvalues and eigenvectors of are then used in place of those of in (11) to obtain estimates of loadings. In practice, is used more often than and is the default in most software packages.

If we choose matrix, then (17) becomes

, where is the number of variables.

Clearly, and have different values for **S** than for **R**.

**Choosing the number of factors,**

Several criteria have been proposed for choosing , the number of factors. Below we consider three criteria.

1. Choose equal to the number of factors necessary for the variance accounted for to achieve a predetermined percentage, say, 80%, of the total variance or .

2. Choose equal to the number of eigenvalues greater than the average eigenvalue. If we factor the sample correlation matrix , the average is 1; for it is .

3. Use the scree plot which is a plot of eigenvalues of or . If the graph falls sharply, followed by a straight line with much smaller slop, choose equal to the number of eigenvalues before the straight line begins.

4. Test the hypothesis that m is the correct number of factors.

To test , we use , which is approximately where degrees of freedom and and are maximum likelihood estimators. Rejection of implies that m is too small and more factors are needed. For the hypothesis test, we assume that the observations constitute a random sample from a multivariate normal distribution with the population mean vector and the population covariance matrix

Estimation of factor loadings are described above using principal component method. Other available methods are principal factor method, iterated principal factor method, and maximum likelihood method.

Below we present the maximum likelihood method for estimating factor loadings.

We assume that the observations constitute a random sample from where is the population mean vector and is the population covariance matrix. Then, andcan beestimated by the method of maximum likelihood. It can be shown that the estimates and satisfy the following:

is diagonal.

These equations must be solved iteratively until a convergence occurs.

**Interpretation of factors: Varimax Rotation**

In real life, there may be situations where most of the factor loadings are of the moderate sizes. As such, analysts encounter difficulties in forming significant factors. One could overcome this problem by using matrix rotation. The idea is to simplify interpretational aspects of factors.

The purpose of a rotation is to produce factors with a mix of high and low loadings and few moderate-sized loadings. From a mathematical viewpoint, there is no difference between a rotated and unrotated matrix.

Varimax rotation (also called Kaiser-Varimax rotation) is a statistical technique used at one level of factor analysis as an attempt to clarify the interpretation of factors. Varimax rotation maximizes the sum of the variance of the squared loadings, where ‘loadings’ are the correlations between variables and factors (for factoring correlation matrix ). If the loadings in a column were nearly equal, the variance would be close to 0. As the squared loadings approach 0 and 1, the variance will approach a maximum. Thus, the varimax method attempts to make the loadings either large or small to facilitate interpretation.

The varimax rotation usually results in high factor loadings for a smaller number of variables and low factor loadings for the rest. In simple terms, the result is a small number of important variables, which makes it easier to interpret results.

Under varimax rotation, our original model becomes

where, andfor some orthogonal matrix where.We plan to find an appropriate rotation defined through matrix **T**, that yields the most easily interpretable factors. Varimax rotation involves scaling the loadings by dividing them by the corresponding communality as shown below:

The varimax rotation finds the rotation that maximizes the quantity given below:

This is the sample variances of the standardized loadings for each factor summed over the  factors.

Note that the varimax procedure cannot guarantee that all variables will load highly on only one factor. In fact, no procedure could do this for all possible data sets.

**Assessing size of loadings to be significant:**

To assess significance of factor loadings obtained from the correlation matrix R, a threshold value of 0.3 has been advocated by many writers. For most successful applications, however, a critical value of 0.3 is too low and will result in variables of complexity greater than 1. A target value of 0.5 or 0.6 is typically more useful.

Notes:

1. In many situations, the grouping of variables are not so logical, and a revision can be tried, such as adjusting the size of loadings deemed to be important, changing , using a different method of estimating the laodings, or employing another type of rotation.
2. When sample covariance matrix is used to estimate loadings, the sample variance for the variable is
3. When sample correlation matrix is used, then the following relationship holds:

,