Factor Analysis Notes

Introduction

Factor analysis (FA) supposes that the variables that we observe are actually the result of underlying forces that we do not observe. The objectives of FA are to discover what those underlying forces are, to measure them, and to understand how they affect the variables that we observe.

FA began in psychology with the work of Charles Spearman (1904). Spearman was interested in intelligence. He proposed that there is a single underlying aptitude that largely explains how well people perform on a battery of tests of mental abilities – tests such as vocabulary, memory, analogies, math reasoning, spatial manipulation, etc. He called this underlying aptitude the *g* factor. *G* cannot be directly observed. *G* can be observed only indirectly through its effects on the individual tests. People who have more of *g* tend to do better on all tests than people who have less of *g*. So *g* accounts for why the results on one test correlate highly with the results on another test. The observable variables (the test results) are called **manifest variables**. The unobservable *g* is called a **latent variable**, **hidden variable**, or **factor**. Spearman viewed each manifest variable as a sum of two parts: a part that the manifest variable shares in common with the other manifest variables (this shared part is *g*) and a part that is unique to the manifest variable and not shared with any other manifest variable. Spearman developed FA as a method for estimating *g* and for understanding how *g* affects individual test results. Later, other psychologists challenged Spearman's belief in a single *g* and have proposed multiple aptitude models, with more than one underlying factor, to account for mental test results. FA can also handle these multiple factor models.

Thus, FA intends to discover the underlying, unobservable factors responsible for the interrelationships within a set of manifest variables. Although the factors cannot be observed, they can be estimated. You may rightly wonder how unobservable variables can possibly be estimated if they cannot be measured. There must be strong magic behind this trick. Since only the manifest variables can be measured, if there is to be any chance of estimating the unobservable factors, the estimates must be obtained from the only variables that can be measured – the manifest variables.

But how is that possible? Here is the idea: If two or more manifest variables are observed to vary together – going up and/or down together (or vice-versa) – then presumably something is "causing" them to go up and down together. That cause may be a hidden factor. The more closely the manifest variables move up or down together, the stronger role that common cause plays in the manifest variables. That is, the stronger the correlations among the manifest variables, the larger the factor part (common part) of the manifest variables and the smaller the unique part, to use Spearman's terminology. The weaker the correlations among the manifest variables, the smaller the factor part of the manifest variables and the larger the unique part. So it is plausible that analysis of the correlations among the manifest variables can provide estimates of the size of the factor part and the unique part of each manifest variable.[1]

[1] This discussion does not tell *how* to compute the estimates. But it establishes the plausibility of some kind of estimation.

One manifest variable may feel the effects of a factor more strongly than another manifest variable. Therefore, in estimating a factor, we should give more weight to manifest variables that feel the factor more strongly, and less weight to manifest variables that feel the factor less strongly. This discussion suggests that we might estimate a factor by a linear combination of the manifest variables, applying weights (coefficients) to the manifest variables that are proportional to the effect that the factor has on the manifest variable. The effect that a factor has on a manifest variable can be assessed by the correlation between the factor and the manifest variable. But since the factor cannot be measured, the correlation between the factor and the manifest variable cannot be calculated. But the effect of the factor is present in all manifest variables that feel the factor. And the correlation between two manifest variables that feel the factor *can* be calculated. So we can infer the size of the unobservable correlation between a factor and a manifest variable from the observable correlation between manifest variables. Then we can calculate the weights to apply to the manifest variables to estimate the factors. That is the idea. It still does not provide a formula to estimate factors. The formulas can be computed. They come from a lot of matrix algebra, but have a simple form.

So the estimates of factors are linear combinations of the manifest variables. Often the estimated factors can replace the more numerous manifest variables in further analyses. So users of FA often have in mind a similar objective of dimensionality reduction as users of principal components analysis (PCA).

These observations suggest that there may be a relationship between FA and PCA. Both methods calculate linear combinations of observable variables that explain much of the variance of those observables. In FA those linear combinations are called factors; in PCA they are called principal components. Both methods are used for dimensionality reduction. Both methods can be interpreted as rotations of coordinate systems. One FA method is, in fact, mathematically identical to PCA!

However, the philosophy of FA is very different from the philosophy of PCA. In PCA, the components (factors) are functions of the observable variables. But in FA, it is the reverse: The observables are considered to be functions of the factors. In FA, these factors are unobservable constructs[2] that are responsible for the observables; in PCA there are no unobservables. In PCA, the number of principal components = number of observable variables.[3] In FA, the number of factors need not equal the number of manifest variables – the number of factors may be unknown. In PCA the statistical model is quite definite. In FA, the form of the FA model may not be known.

[2] To be sure, the factors must be estimated. And since the only way to estimate them is to use the observable variables, then the *estimated* factors must be functions of the observables. But there are indeterminacies in FA that are absent in PCA. For example, in FA the number of factors is unknown; in PCA the number of components = number of variables -- the question is how many components to keep. In FA, the form of the FA model is also unknown; the PCA model is quite definite.

[3] Although not all of the principal components may be retained.

In summary, FA is based on the assumption that some underlying unobservable factors are responsible for the observed correlation among the manifest variables. The objectives of FA are

- to estimate those underlying factors
- to study their structure and effects

The major issues in the use of FA:

1) Do the assumptions of the FA model apply?
2) If so, how many factors are there?
3) How to estimate ("extract") the factors?
4) How to interpret the meaning of the factors?

Complicating the matter is the fact that the factors cannot be observed or measured directly. This is a major conceptual difference between FA and principal components analysis (PCA). In PCA, the components are linear combinations of the observed variables; so only the weights (coefficients) of the observed variables need to be estimated. In FA, the observed variables are linear combinations of the factors; not only do the coefficients of the factors need to be estimated, so also do the factor scores -- even the number of factors in the linear combination must be estimated. FA is like trying to do a regression of observed variables on predictors (factors) without knowing the values of the predictors, or even how many predictors there are!

The FACTOR ANALYSIS MODEL and ASSUMPTIONS.

FA makes some assumptions about the data. I have collected these assumptions together as a set of specifications, called the **factor analysis model**, presented below:

1. There are m **common factors**, denoted $\xi_1, ..., \xi_m$, that underlie p ($p > m$) manifest variables $X_1, X_2, ..., X_p$.
2. Each manifest variable X_i is a linear combination of one or more common factors and a **unique factor** ε_i (unique to X_i):
$$X_i = \lambda_{i1}\xi_1 + \lambda_{i2}\xi_2 + \cdots + \lambda_{im}\xi_m + \varepsilon_i$$
3. Each factor ξ_j contributes to at least two manifest variables (i.e., at least two of $\lambda_{1j}, \lambda_{2j}, \lambda_{3j}, ..., \lambda_{pj}$ are nonzero). This is a natural assumption because if only one $\lambda_{ij} \neq 0$, for some $i = 1, 2, ..., p$ (i.e., if ξ_j contributes to only one manifest variable), then ξ_j would really be part of the unique factor for that X_i.
4. All common factors ξ's and manifest variables X's are standardized to mean 0, stdev 1.[4]
5. corr($\varepsilon_i, \varepsilon_j$) = 0 for all i,j, $i \neq j$. This is a natural assumption because if two different ε's could be correlated, then part of their effects would be in common to their two X-variables and thus not unique to their separate X-variables.
6. corr(ξ_i, ε_j) = 0 for all i,j. This is a natural assumption because each common factor affects at least two manifest variables – so if a unique factor correlates with a common factor, then that unique factor affects at least two manifest variables, contradicting the uniqueness of the unique factor.
7. corr(ξ_i, ξ_j) = 0 for all i,j, $i \neq j$. I.e., the common factors are uncorrelated. Since the common factors are also standardized, then they are **orthonormal**. However, since there

[4] We will not suppose that the unique term ε is standardized, although we will assume that ε has mean zero. Thus Var(ε) need not be 1.

are fewer common factors (*m*) than manifest variables (*p*), then the transformation from manifest variable space to common factor space does not preserve the geometry of manifest variable space. The common factor transformation omits the unique factors.[5] But the factor transformation does provide "interesting" perspectives on the data. The assumption of uncorrelated factors is not nearly as natural as other assumptions. Why should unobserved underlying factors necessarily operate orthogonally to each other? To be sure, you could always resolve non-orthogonal forces into their orthogonal components. However, the strongest action of the underlying factors may be angled toward each other. In some FA methods, the assumption of uncorrelated factors is dropped. Non-orthogonal factor rotations are frequently used to explore further interesting features of the data interrelationships.

When the coefficients of the factor transformation are arranged in a table, with manifest variables as rows and common factors as columns, as below, they form the **pattern matrix**, also called the **factor pattern**.

Pattern Matrix				
	ξ_1	ξ_2	...	ξ_m
X_1	λ_{11}	λ_{12}	...	λ_{1m}
X_2	λ_{21}	λ_{22}	...	λ_{2m}
	...	...	...	...
X_p	λ_{p1}	λ_{p2}	...	λ_{pm}

To interpret the meaning of the factors, it is useful to arrange a corresponding table, called the **structure matrix**, with manifest variables as rows and common factors as columns, but with **factor loadings** (correlations between manifest variables and common factors) as cell entries. Under the FA model (above), if assumption 7 is adopted (uncorrelated common factors), then the pattern matrix = structure matrix, i.e., the coefficients are the correlations. Thus, the meaning of the factors may be divined from either the pattern matrix or the structure matrix.

Structure Matrix				
	ξ_1	ξ_2	...	ξ_m
X_1	$corr_{11}$	$corr_{12}$	...	$corr_{1m}$
X_2	$corr_{21}$	$corr_{22}$	...	$corr_{2m}$
	...	...	...	...
X_p	$corr_{p1}$	$corr_{p2}$	...	$corr_{pm}$

[5] This also tells us that the reverse transformation, from common factor space, back to manifest variable space cannot successfully recreate the manifest variables exactly.

From the FA model and assumptions, the following results can be mathematically deduced:

[1] The partial correlation[6] of X_i and X_j , controlling for the common factors is zero. I.e., remove the effects of the common factors from X_i and X_j, and the correlation between what is left is zero. After you take out the common factors from X_i and X_j, what is left is only the unique factors ε_i and ε_j, which have zero correlation by assumption 5.

[2] corr(X_i, ξ_j) = λ_{ij} = slope of regression of X_i on ξ_j. In order to see this, consider the factor representation of $X_i = \lambda_{i1}\xi_1 + \lambda_{i2}\xi_2 + \cdots + \lambda_{im}\xi_m + \varepsilon_i$. This will be the multiple regression of X_i on $\xi_1, \xi_2, \cdots, \xi_m, \varepsilon_i$ since every x_i exactly satisfies this relationship. By the FA model, the predictors $\xi_1, \xi_2, \cdots, \xi_m, \varepsilon_i$ are all uncorrelated, so the multiple regression coefficients are the same as they would be in simple regression, i.e., $\lambda_{ij} = corr(X_i, \xi_j)\dfrac{std(X_i)}{std(\xi_j)} = corr(X_i, \xi_j)$ since both X_i and ξ_j are standardized. This says that the factor coefficients are the factor loadings, so the pattern matrix = structure matrix. Also, λ_{ij}^2 = R-square = proportion of variance of X_i explained by ξ_j.

[3] $Var(X_i) = \underbrace{\lambda_{i1}^2 + \lambda_{i2}^2 + \lambda_{i3}^2 + \cdots + \lambda_{ip}^2}_{\text{communality}} + \underbrace{Var(\varepsilon_i)}_{\text{uniqueness}} = 1$

[You should be able to derive this property.]

For each variable X_i , **communality** + **uniqueness** = 1. The communality of X_i is the proportion of the variance of X_i explained by all of the common factors = multiple R-square from regressing X_i on all of the common factors.

The **factorial complexity** of a variable is the number of factors having high (positive or negative) loadings on that variable.

The **factorial determination** of the variables is the degree to which the observed variables are determined by the common factors, and can be measured by the average communality: $\Sigma\Sigma\ \lambda_{ij}^2 / p$.

[6] The partial correlation between two variables U1 and U2, controlling for U3 conceptually means that in a 3D plot of U1, U2, U3, you set U3 = constant and then correlate U1 and U2 on the plane U3 = constant. The partial correlation may or may not depend upon which constant you choose for U3. With multivariate normally distributed data, the constant does not matter. There are formulas for calculating partial correlation.

[4] $corr(X_i, X_j) = \lambda_{i1}\lambda_{j1} + \lambda_{i2}\lambda_{j2} + \cdots + \lambda_{im}\lambda_{jm}$. *[You should be able to derive this property.]*

This says that the factors determine the correlation structure among the manifest variables. However, the correlation structure among the manifest variables does not determine the factors: Given a correlation matrix of the variables, there can be multiple factor patterns that yield the correlation matrix, where

- Each factor pattern has the same number of orthogonal factors satisfying the FA model and assumptions, or
- Each factor pattern has the same number of oblique [7] factors satisfying the FA model and assumptions (except assumption 7, of course), or
- Each factor pattern has a different number of orthogonal or oblique factors satisfying the FA model, or
- The factor analysis model assumptions are violated.

Since the observed variables (and more particularly their correlation structure) are all we can directly observe, there is a fundamental indeterminacy in FA that is unavoidable.

Fundamentally, it is not possible to test whether the FA model applies. The indeterminacies mentioned above imply that

- the appropriateness of the FA model can never be proven,
- the number of factors can never be known,
- the factor loadings can never be known even if the correlation matrix of the variables were known exactly and without sampling or measurement error.

The most that can be obtained from FA is that the analysis is *consistent* with the FA model.

The applicability of FA is an assumption. Common assumptions include

- the **postulate of factorial causation** -- that the FA model and assumptions hold,
- the **postulate of parsimony** -- that fewer factors are to be preferred over more factors.

HOW MANY FACTORS ARE THERE?

As indicated above, this question cannot be answered definitively, because of the indeterminacies in going from observed correlation matrix to factors. Often, one has a theory that suggests that (say) 3 factors are present. Then one can try 2, 3, and 4 factor models to see which is most satisfactory. With certain factor extraction methods (maximum likelihood) and distributional assumptions (multivariate normal for the X's), you can test whether more factors are needed than were extracted (see METHOD=ML below). All this assumes, however, the truth of the FA model.

HOW to ESTIMATE the FACTORS?

The estimation of factors is called **extraction of factors**. Because the only data available for estimation of factors are the observables X_1, ..., X_p, the factors are estimated to be linear combinations of X_1, ..., X_p. However, remember that FA conceptualizes the observables to be

[7] Oblique factors may be nonorthogonal.

linear combinations of the factors, rather the reverse. SAS offers several types of factor extraction methods, selected by the **METHOD=** option within **PROC FACTOR**. (See examples at end of these notes.) A few comments on some of them:

(1) **METHOD=PRINCIPAL** The most common factor extraction method.[8] This method uses the techniques of principal component analysis (PCA). The (possibly adjusted) correlation matrix of the observed variables is analyzed into orthogonal linear combinations (factors) in such a way as to maximize the variance of each factor subject to the factor being uncorrelated with all preceding factors. This loads as much of the total variance of the variables as possible onto the first few factors.

The **PRIORS** option permits you to select preliminary estimates of the communalities. Some common choices for PRIORS:

- **PRIORS=ONE** With all communalities equal to one, this implies that the factors determine all of the variables completely (no uniqueness factors).[9] With the selection of PRIORS=ONE, the METHOD=PRINCIPAL factor extraction technique becomes identical to PCA. In PCA, there is no uniqueness – all of the variables are completely determined by the set of principal components. Doing PCA with PROC FACTOR rather than PROC PRINCOMP gains the ability to rotate the principal components.
- **PRIORS=SMC** (Squared Multiple Correlation) means that the elements in the diagonal of the ordinary correlation matrix are adjusted before extracting factors from the adjusted correlation matrix. The ones in the diagonal of the correlation matrix (corr(X_i,X_i)=1) are replaced by estimates of the communalities. The estimated communality of each X_i is equal to the R-square from regressing X_i on all of the other X's. This reasonably views the communality of X_i as approximately the part of X_i in common with the other X's.[10] Choosing PRIORS other than ONE is **Principal Factor Analysis**.

(2) **METHOD=PRINIT** (Iterated Principal method)[11] This is an iterated method involving a new factor extraction at each step. The first step is a factor extraction with METHOD=PRINCIPAL using PRIORS=ONE, i.e., PCA. For subsequent steps, the estimated communalities from the preceding step replace the diagonal elements of the correlation matrix, and a new factor extraction is performed with METHOD=PRINCIPAL. So each step after the first is a Principal Factor Analysis. After each step, new estimated communalities are obtained for each X_i using the above mathematical relation [3] communality of $X_i = \lambda_{i1}^2 + \lambda_{i2}^2 + \lambda_{i3}^2 + \cdots + \lambda_{ip}^2$, where the new estimates of the loadings λ are substituted into the right-hand-side. Iterations continue until the communality estimates stabilize, defined as maximum change in communalities of X_1, ..., X_p from previous step less than the convergence criterion (default = .001), or until the limit for iterations is reached (default limit = 30).

[8] Called "Principal Component Factoring".
[9] Recall that communality + uniqueness = 1 for each variable.
[10] If the FA model is correct, SMC is a theoretical lower bound for the communality. However, with real data, sampling variation and the unlikelihood that FA modeling assumptions hold precisely render this lower bound much fuzzier and make SMC a reasonable first guess at communality.
[11] Called "Principal Axis Factoring."

(3) **METHOD=ML** (Maximum Likelihood). To use this method, the number of common factors is assumed known, and the *X*'s (and factors) are assumed to have a multivariate normal distribution. By default, METHOD=ML uses PRIORS=SMC. The method estimates the factor loadings most likely to have produced the observed correlation matrix. The method is iterative and time-consuming. It is susceptible to anomalies such as estimated communalities exceeding one (**ultra-Heywood cases**), which may destabilize the computational routines and yield unreliable results[12] unless the **HEYWOOD** option is used (resets the over-one communalities to 1 and continues). ML produces approximate significance tests of

1. H_0: *there are no common factors* versus H_a: *there is at least one common factor*, and
2. H_0: *the number of factors extracted is correct* versus H_a: *more factors are needed*.

Assuming the FA model is appropriate, these tests are OK. However, the tests might reject H_0 not only b/c more factors are necessary, but also because the FA model does not really apply. So if you reject H_0, you can not be sure whether the reason is that you need more factors or that FA is inappropriate.

(4) **METHOD=ULS** (Unweighted Least Squares). To use this method, the number of factors is assumed known, but no distribution is assumed. The method chooses the factor loadings to minimize the squared differences between the observed *X*-correlations and the correlations resulting from the estimated factor loadings. The solution is iterative. By default, METHOD=ULS uses PRIORS=SMC. After the first iteration, estimates of the correlations between observables are computed using the above mathematical relation [4] $corr(X_i, X_j) = \lambda_{i1}\lambda_{j1} + \lambda_{i2}\lambda_{j2} + \cdots + \lambda_{im}\lambda_{jm}$, where the new estimates of the loadings λ are substituted into the right hand side. The matrix of estimated correlations is called the **reproduced correlation matrix**. The difference between the observed correlations and the reproduced correlation matrix is the **residual correlation matrix**. ULS chooses loadings λ to minimize the sum of squared off-diagonal terms in the residual correlation matrix. Iterations continue until the convergence criterion is met or until the limit of iterations is reached.

(5) **METHOD=ALPHA**. The cases are considered to be a population rather than a sample; the variables are considered to be a sample of variables from a larger population with common factors underlying all of them. The method is supposed to maximize the correlation of the extracted factors with the corresponding population factors. By default, METHOD=ALPHA uses PRIORS=SMC.

(6) **METHOD=IMAGE**. The **image** of a variable is the part of it that is explainable by a linear combination of the other variables. The **anti-image** is the part that is not thus explainable. The philosophy of image FA differs from regular FA in viewing the common part as a function of observed variables rather than unobservable factors. The variables are considered to be a sample and the common part based on the available sample of variables is called the **partial image**. The PRIORS option is not applicable for METHOD=IMAGE.

[12] The existence of an ultra-Heywood case implies a unique factor with negative variance. Since this is theoretically impossible, something is wrong: possibly the wrong number of common factors, or not enough data, or bad prior communality estimates, or the FA model is not appropriate for these data.

How to INTERPRET the MEANING of the FACTORS?

The interpretation of factor loadings is very similar to the interpretation of principal components. The basic idea is to examine the factor pattern (loadings). The meaning of a factor is obtained from the meanings of the variables that "load high" on that factor. A variable loads high if the magnitude of the correlation (loading) between the variable and the factor is high. The following guides may be helpful in interpreting factor patterns.

- The meaning of a factor is derived from the common meaning of the variables that load high on the factor.
- If no variables load very high, but several share comparable lower values, then the factor is probably an average or general factor in common to the affected variables.
- If the signs of the high loading variables are mixed (+ and -), then the factor probably represents a contrast between the variables of opposite sign.
- Often a contrast will have high loadings with mixed signs on variables which all had the same sign in an earlier factor; in this case, the contrast is probably picking up the secondary explanatory power of those variables that was not picked up by the earlier factor.

Often, examination of the initial factor pattern reveals no clear interpretation. In such cases, a **rotation** may improve interpretability. Geometrically, a rotation is a shift in the angular orientation of the factor axes. Analytically, a rotation is a set of linear combinations of the factors, yielding new factors with new interpretations. The rotation is applied to the initial extraction and changes its factor pattern. The purpose of most rotations is to simplify the pattern matrix by getting the loadings as close to 0 and 1 as possible. The effect is to spread out the loadings, increasing their variability.[13]

A rotation can be **orthogonal**, meaning that as the factor axes are rotated, they retain their perpendicular orientations with each other. A rotation can be **oblique**, meaning that some axes end up at angles of more or less than 90 degrees with each other. SAS offers several types of orthogonal and oblique rotations selected by the **ROTATE=** option. A few notes on some orthogonal rotations:

[1] **ROTATE=VARIMAX**. The most popular option. This rotation maximizes the variance of squared loadings for each *factor*, subject to the rotation being orthogonal. Thus, varimax simplifies the *columns* of the pattern matrix.

[2] **ROTATE=QUARTIMAX**. This rotation maximizes the variance of the squared loadings for each *variable* subject to the rotation being orthogonal. Thus quartimax, simplifies the *rows* of the pattern matrix.

[3] **ROTATE=EQUAMAX**. This orthogonal rotation maximizes a weighted average of the variance of squared loadings for each factor and each variable. So it tries to combine varimax and quartimax.

Oblique rotations are more complicated. When the original factor axes are rotated, the axes do not remain perpendicular. This means that the factors are now correlated. In this case, the factor structure (correlations) is no longer the same as the factor pattern (coefficients).

SAS prints additional tables to help interpret the oblique factors, such as **reference axis correlations** and **reference structure**. The reference axis correlations are the partial

[13] In mentioning variability here, I do not mean the variance of the observables, nor of the factors. I mean the variability of the factor *coefficients* as a set of numbers. Maximizing some function of the variability of the factor coefficients is the optimization criterion for some of the rotation methods.

correlations among pairs of factors with the effects of all other factors removed from the pair being correlated. The reference structure is the set of semipartial correlations between the variables and the common factors with the effects of all other common factors removed from the factor being correlated but not from the variable being correlated. In the case of uncorrelated factors, the reference axis correlation matrix has 1's in the main diagonal and 0's off the diagonal, and the reference structure is the same as the factor pattern (coefficients).

Interpretation of factors is very much an art.

ILLUSTRATIVE SAS PROGRAMS

(1)
```
PROC FACTOR;
```
Does principal components analysis on all the numeric variables in the most recently created SAS ds, retaining all factors with eigenvalues greater than 1 (the default retention rule).

(2)
```
PROC FACTOR  DATA=MY.INFO  METHOD=PRINCIPAL  MINEIGEN=0;
    VAR X1 X2 X3 X4;
```
Does PCA on X1 X2 X3 X4 in SAS dataset MY.INFO, retaining all factors with eigenvalue greater than 0 (i.e., all 4 factors).

(3)
```
PROC FACTOR  DATA=MY.DATA  METHOD=PRINCIPAL  PRIORS=SMC
NFACTORS=3;
    VAR X1 X2 X3 X4;
```
Does Principal Factor Analysis with squared multiple correlations as initial estimates of communalities, retaining 3 factors, using variables X1 X2 X3 X4 in MY.DATA.

(4)
```
PROC FACTOR  DATA=MY.DATA  METHOD=ML  HEYWOOD  PRIORS=ONE
MINEIGEN=1  NFACTORS=3  SCREE  ROTATE=VARIMAX;
      VAR X1 X2 X3 X4;
```
Does maximum likelihood factor analysis on X1 X2 X3 X4 in MY.DATA with initial estimates of communalities = 1, retaining three factors provided they all have eigenvalues greater than 1, resetting to 1 (HEYWOOD) any communalities set > 1 during the iterations, producing a scree plot of the eigenvalues, and rotating the ML orthogonal factor extraction with the varimax orthogonal rotation. The output will include tests of H_0: *there are no common factors* versus H_a: *there is at least one common factor*, and H_0: *three*[14] *factors are sufficient* versus H_a: *more factors are needed*.

(5)
```
PROC FACTOR  DATA=MY.DATA  METHOD=PRINCIPAL  PRIORS=SMC
PERCENT=80  SCORE  OUTSTAT=MY.STAT;
    VAR X1 X2 X3 X4;
  PROC FACTOR  DATA=MY.STAT  ROTATE=PROMAX;
  PROC SCORE  DATA=YOUR.DATA  SCORE=MY.STAT  OUT=OUR.CASES;
    VAR X1 X2 X3 X4;
```
Does Principal Factor Analysis on X1 X2 X3 X4 in MY.DATA with initial estimates of communalities = squared multiple correlations, retaining enough factors to account for at least 80% of the common variance, printing scoring information for estimating factor scores for cases and storing that information in the MY.STAT output data set. Then (second PROC FACTOR) rotate initial factor extraction stored in MY.STAT through oblique promax rotation. Then (PROC SCORE) apply factor scoring information stored in MY.STAT to estimate factor scores for the observations in YOUR.DATA using X1 X2 X3 X4, and storing the observation scores in OUR.CASES for further analysis or for printing or plotting.

[14] Assuming MINEIGEN=1 does not eliminate some of the three factors, in which case the number of factors actually selected will be tested for sufficiency.