Chapter 26

# The FACTOR Procedure

## Chapter Table of Contents

Chapter 26
# The FACTOR Procedure

## Overview

The FACTOR procedure performs a variety of common factor and component analyses and rotations. Input can be multivariate data, a correlation matrix, a covariance matrix, a factor pattern, or a matrix of scoring coefficients. The procedure can factor either the correlation or covariance matrix, and you can save most results in an output data set.

PROC FACTOR can process output from other procedures. For example, it can rotate the canonical coefficients from multivariate analyses in the GLM procedure.

The methods for factor extraction are principal component analysis, principal factor analysis, iterated principal factor analysis, unweighted least-squares factor analysis, maximum-likelihood (canonical) factor analysis, alpha factor analysis, image component analysis, and Harris component analysis. A variety of methods for prior communality estimation is also available.

The methods for rotation are varimax, quartimax, parsimax, equamax, orthomax with user-specified gamma, promax with user-specified exponent, Harris-Kaiser case II with user-specified exponent, and oblique Procrustean with a user-specified target pattern.

Output includes means, standard deviations, correlations, Kaiser's measure of sampling adequacy, eigenvalues, a scree plot, eigenvectors, prior and final communality estimates, the unrotated factor pattern, residual and partial correlations, the rotated primary factor pattern, the primary factor structure, interfactor correlations, the reference structure, reference axis correlations, the variance explained by each factor both ignoring and eliminating other factors, plots of both rotated and unrotated factors, squared multiple correlation of each factor with the variables, and scoring coefficients.

Any topics that are not given explicit references are discussed in Mulaik (1972) or Harman (1976).

## Background

See Chapter 52, "The PRINCOMP Procedure," for a discussion of principal component analysis. See Chapter 19, "The CALIS Procedure," for a discussion of confirmatory factor analysis.

Common factor analysis was invented by Spearman (1904). Kim and Mueller (1978a,b) provide a very elementary discussion of the common factor model. Gorsuch (1974) contains a broad survey of factor analysis, and Gorsuch (1974) and

Cattell (1978) are useful as guides to practical research methodology. Harman (1976) gives a lucid discussion of many of the more technical aspects of factor analysis, especially oblique rotation. Morrison (1976) and Mardia, Kent, and Bibby (1979) provide excellent statistical treatments of common factor analysis. Mulaik (1972) is the most thorough and authoritative general reference on factor analysis and is highly recommended to anyone familiar with matrix algebra. Stewart (1981) gives a nontechnical presentation of some issues to consider when deciding whether or not a factor analysis may be appropriate.

A frequent source of confusion in the field of factor analysis is the term *factor*. It sometimes refers to a hypothetical, unobservable variable, as in the phrase *common factor*. In this sense, *factor analysis* must be distinguished from component analysis since a component is an observable linear combination. *Factor* is also used in the sense of *matrix factor,* in that one matrix is a factor of a second matrix if the first matrix multiplied by its transpose equals the second matrix. In this sense, *factor analysis* refers to all methods of data analysis using matrix factors, including component analysis and common factor analysis.

A *common factor* is an unobservable, hypothetical variable that contributes to the variance of at least two of the observed variables. The unqualified term "factor" often refers to a common factor. A *unique factor* is an unobservable, hypothetical variable that contributes to the variance of only one of the observed variables. The model for common factor analysis posits one unique factor for each observed variable.

The equation for the common factor model is

$$y_{ij} = x_{i1}b_{1j} + x_{i2}b_{2j} + \cdots + x_{iq}b_{qj} + e_{ij}$$

where

| | |
|---|---|
| $y_{ij}$ | is the value of the $i$th observation on the $j$th variable |
| $x_{ik}$ | is the value of the $i$th observation on the $k$th common factor |
| $b_{kj}$ | is the regression coefficient of the $k$th common factor for predicting the $j$th variable |
| $e_{ij}$ | is the value of the $i$th observation on the $j$th unique factor |
| $q$ | is the number of common factors |

It is assumed, for convenience, that all variables have a mean of 0. In matrix terms, these equations reduce to

$$\mathbf{Y} = \mathbf{XB} + \mathbf{E}$$

In the preceding equation, $\mathbf{X}$ is the matrix of factor scores, and $\mathbf{B}'$ is the factor pattern.

There are two critical assumptions:

- The unique factors are uncorrelated with each other.
- The unique factors are uncorrelated with the common factors.

In principal component analysis, the residuals are generally correlated with each other. In common factor analysis, the unique factors play the role of residuals and are defined to be uncorrelated both with each other and with the common factors. Each common factor is assumed to contribute to at least two variables; otherwise, it would be a unique factor.

When the factors are initially extracted, it is also assumed, for convenience, that the common factors are uncorrelated with each other and have unit variance. In this case, the common factor model implies that the covariance $s_{jk}$ between the $j$th and $k$th variables, $j \neq k$, is given by

$$s_{jk} = b_{1j}b_{1k} + b_{2j}b_{2k} + \cdots + b_{qj}b_{qk}$$

or

$$\mathbf{S} = \mathbf{B}'\mathbf{B} + \mathbf{U}^2$$

where $\mathbf{S}$ is the covariance matrix of the observed variables, and $\mathbf{U}^2$ is the diagonal covariance matrix of the unique factors.

If the original variables are standardized to unit variance, the preceding formula yields correlations instead of covariances. It is in this sense that common factors explain the correlations among the observed variables. The difference between the correlation predicted by the common factor model and the actual correlation is the *residual correlation*. A good way to assess the goodness-of-fit of the common factor model is to examine the residual correlations.

The common factor model implies that the partial correlations among the variables, removing the effects of the common factors, must all be 0. When the common factors are removed, only unique factors, which are by definition uncorrelated, remain.

The assumptions of common factor analysis imply that the common factors are, in general, not linear combinations of the observed variables. In fact, even if the data contain measurements on the entire population of observations, you cannot compute the scores of the observations on the common factors. Although the common factor scores cannot be computed directly, they can be estimated in a variety of ways.

The problem of factor score indeterminacy has led several factor analysts to propose methods yielding components that can be considered approximations to common factors. Since these components are defined as linear combinations, they are computable. The methods include Harris component analysis and image component analysis. The advantage of producing determinate component scores is offset by the fact that, even if the data fit the common factor model perfectly, component methods do not generally recover the correct factor solution. You should not use any type of component analysis if you really want a common factor analysis (Dziuban and Harris 1973; Lee and Comrey 1979).

After the factors are estimated, it is necessary to interpret them. Interpretation usually means assigning to each common factor a name that reflects the importance of the factor in predicting each of the observed variables, that is, the coefficients in the pattern matrix corresponding to the factor. Factor interpretation is a subjective process. It can sometimes be made less subjective by *rotating* the common factors, that is, by applying a nonsingular linear transformation. A rotated pattern matrix in which all the coefficients are close to 0 or $\pm1$ is easier to interpret than a pattern with many intermediate elements. Therefore, most rotation methods attempt to optimize a function of the pattern matrix that measures, in some sense, how close the elements are to 0 or $\pm1$.

After the initial factor extraction, the common factors are uncorrelated with each other. If the factors are rotated by an *orthogonal transformation*, the rotated factors are also uncorrelated. If the factors are rotated by an *oblique transformation*, the rotated factors become correlated. Oblique rotations often produce more useful patterns than do orthogonal rotations. However, a consequence of correlated factors is that there is no single unambiguous measure of the importance of a factor in explaining a variable. Thus, for oblique rotations, the pattern matrix does not provide all the necessary information for interpreting the factors; you must also examine the *factor structure* and the *reference structure*.

Rotating a set of factors does not change the statistical explanatory power of the factors. You cannot say that any rotation is better than any other rotation from a statistical point of view; all rotations are equally good statistically. Therefore, the choice among different rotations must be based on nonstatistical grounds. For most applications, the preferred rotation is that which is most easily interpretable.

If two rotations give rise to different interpretations, those two interpretations must not be regarded as conflicting. Rather, they are two different ways of looking at the same thing, two different points of view in the common-factor space. Any conclusion that depends on one and only one rotation being correct is invalid.

## Outline of Use

### *Principal Component Analysis*

One important type of analysis performed by the FACTOR procedure is principal component analysis. The statements

```
proc factor;
run;
```

result in a principal component analysis. The output includes all the eigenvalues and the pattern matrix for eigenvalues greater than one.

Most applications require additional output. For example, you may want to compute principal component scores for use in subsequent analyses or obtain a graphical aid to help decide how many components to keep. You can save the results of the analysis in a permanent SAS data library by using the OUTSTAT= option. (Refer to the *SAS Language Reference: Dictionary* for more information on permanent SAS data libraries and librefs.) Assuming that your SAS data library has the libref save and that the data are in a SAS data set called raw, you could do a principal component analysis as follows:

```
proc factor data=raw method=principal scree mineigen=0 score
     outstat=save.fact_all;
run;
```

The SCREE option produces a plot of the eigenvalues that is helpful in deciding how many components to use. The MINEIGEN=0 option causes all components with variance greater than zero to be retained. The SCORE option requests that scoring coefficients be computed. The OUTSTAT= option saves the results in a specially structured SAS data set. The name of the data set, in this case fact_all, is arbitrary. To compute principal component scores, use the SCORE procedure.

```
proc score data=raw score=save.fact_all out=save.scores;
run;
```

The SCORE procedure uses the data and the scoring coefficients that are saved in save.fact_all to compute principal component scores. The component scores are placed in variables named Factor1, Factor2, . . . , Factor$n$ and are saved in the data set save.scores. If you know ahead of time how many principal components you want to use, you can obtain the scores directly from PROC FACTOR by specifying the NFACTORS= and OUT= options. To get scores from three principal components, specify

```
proc factor data=raw method=principal
     nfactors=3 out=save.scores;
run;
```

To plot the scores for the first three components, use the PLOT procedure.

```
proc plot;
   plot factor2*factor1 factor3*factor1 factor3*factor2;
run;
```

### *Principal Factor Analysis*

The simplest and computationally most efficient method of common factor analysis is principal factor analysis, which is obtained the same way as principal component analysis except for the use of the PRIORS= option. The usual form of the initial analysis is

```
proc factor data=raw method=principal scree
     mineigen=0 priors=smc outstat=save.fact_all;
run;
```

The squared multiple correlations (SMC) of each variable with all the other variables are used as the prior communality estimates. If your correlation matrix is singular, you should specify PRIORS=MAX instead of PRIORS=SMC. The SCREE and MINEIGEN= options serve the same purpose as in the preceding principal component analysis. Saving the results with the OUTSTAT= option enables you to examine the eigenvalues and scree plot before deciding how many factors to rotate and to try several different rotations without re-extracting the factors. The OUTSTAT= data set is automatically marked TYPE=FACTOR, so the FACTOR procedure realizes that it contains statistics from a previous analysis instead of raw data.

After looking at the eigenvalues to estimate the number of factors, you can try some rotations. Two and three factors can be rotated with the following statements:

```
proc factor data=save.fact_all method=principal n=2
     rotate=promax reorder score outstat=save.fact_2;
proc factor data=save.fact_all method=principal n=3
     rotate=promax reorder score outstat=save.fact_3;
run;
```

The output data set from the previous run is used as input for these analyses. The options N=2 and N=3 specify the number of factors to be rotated. The specification ROTATE=PROMAX requests a promax rotation, which has the advantage of providing both orthogonal and oblique rotations with only one invocation of PROC FACTOR. The REORDER option causes the variables to be reordered in the output so that variables associated with the same factor appear next to each other.

You can now compute and plot factor scores for the two-factor promax-rotated solution as follows:

```
proc score data=raw score=save.fact_2 out=save.scores;
proc plot;
   plot factor2*factor1;
run;
```

### Maximum-Likelihood Factor Analysis

Although principal factor analysis is perhaps the most commonly used method of common factor analysis, most statisticians prefer maximum-likelihood (ML) factor analysis (Lawley and Maxwell 1971). The ML method of estimation has desirable asymptotic properties (Bickel and Doksum 1977) and produces better estimates than principal factor analysis in large samples. You can test hypotheses about the number of common factors using the ML method.

The ML solution is equivalent to Rao's (1955) canonical factor solution and Howe's solution maximizing the determinant of the partial correlation matrix (Morrison 1976). Thus, as a descriptive method, ML factor analysis does not require a multivariate normal distribution. The validity of Bartlett's $\chi^2$ test for the number of factors does require approximate normality plus additional regularity conditions that are usually satisfied in practice (Geweke and Singleton 1980).

The ML method is more computationally demanding than principal factor analysis for two reasons. First, the communalities are estimated iteratively, and each iteration takes about as much computer time as principal factor analysis. The number of iterations typically ranges from about five to twenty. Second, if you want to extract different numbers of factors, as is often the case, you must run the FACTOR procedure once for each number of factors. Therefore, an ML analysis can take 100 times as long as a principal factor analysis.

You can use principal factor analysis to get a rough idea of the number of factors before doing an ML analysis. If you think that there are between one and three factors, you can use the following statements for the ML analysis:

```
proc factor data=raw method=ml n=1
     outstat=save.fact1;
run;
proc factor data=raw method=ml n=2 rotate=promax
     outstat=save.fact2;
run;
proc factor data=raw method=ml n=3 rotate=promax
     outstat=save.fact3;
run;
```

The output data sets can be used for trying different rotations, computing scoring coefficients, or restarting the procedure in case it does not converge within the allotted number of iterations.

The ML method cannot be used with a singular correlation matrix, and it is especially prone to Heywood cases. (See the section “Heywood Cases and Other Anomalies” on page 1153 for a discussion of Heywood cases.) If you have problems with ML, the best alternative is to use the METHOD=ULS option for unweighted least-squares factor analysis.

# Getting Started

The following example demonstrates how you can use the FACTOR procedure to perform common factor analysis and use a transformation to rotate the extracted factors.

Suppose that you want to use factor analysis to explore the relationship among assessment scores of a group of students. For each student in the group, you record six homework scores, two midterm examination scores, and the final exam score.

The following DATA step creates the SAS data set Grades:

```
data Grades;
   input HomeWork1 - HomeWork6 MidTerm1 MidTerm2 FinalExam;
   datalines;
15  18  36  29  44  30  78  87  70
15  16  24  30  41  30  71  73  89
15  14  23  34  28  24  84  72  76
15  20  39  35  50  30  74  79  96
15  20  39  35  46  30  76  77  94
15  20  28  30  49  28  40  44  66
15  15  29  25  36  30  88  69  93
15  20  37  35  50  30  97  95  98
14  16  24  30  44  28  57  78  85
15  17  29  26  38  28  56  78  76
15  17  31  34  40  27  72  67  84
11  16  29  34  31  27  83  68  75
15  18  31  18  40  30  75  43  67
14  14  29  25  49  30  71  93  93
15  18  36  29  44  30  85  64  75
;
```

The data set Grades contains the variables representing homework scores (HomeWork1—HomeWork6), the two midterm exam scores (MidTerm1 and MidTerm2), and the final exam score (FinalExam).

The following statements invoke the FACTOR procedure:

```
proc factor data=Grades priors=smc rotate=varimax nfactors=2;
run;
```

The DATA= option in PROC FACTOR specifies the SAS data set Grades as the input data set. The PRIORS= option specifies that the squared multiple correlations (SMC) of each variable with all the other variables are used as the prior communality estimates and also that PROC FACTOR gives a principal factor solution to the common factor model. The ROTATE= option specifies the VARIMAX orthogonal factor rotation method. To see if two latent factors can explain the observed variation in the data, the NFACTOR= option specifies that two factors be retained. All variables in the data set are analyzed.

The output from this analysis is displayed in the following figures.

```
                          The SAS System

                       The FACTOR Procedure
               Initial Factor Method: Principal Factors

                 Prior Communality Estimates: SMC

  HomeWork1       HomeWork2       HomeWork3       HomeWork4       HomeWork5

 0.27602335      0.86733312      0.82222517      0.79295256      0.80742053

         HomeWork6        MidTerm1        MidTerm2       FinalExam

        0.83330706      0.67135234      0.64889405      0.68860512


              Eigenvalues of the Reduced Correlation Matrix:
                 Total = 6.40811331  Average = 0.71201259

                Eigenvalue    Difference    Proportion    Cumulative

           1    3.00212450    1.21898414        0.4685        0.4685
           2    1.78314036    0.71888817        0.2783        0.7468
           3    1.06425218    0.34974843        0.1661        0.9128
           4    0.71450375    0.55643869        0.1115        1.0243
           5    0.15806506    0.10471212        0.0247        1.0490
           6    0.05335294    0.15681933        0.0083        1.0573
           7    -.10346639    0.01266761       -0.0161        1.0412
           8    -.11613399    0.03159110       -0.0181        1.0231
           9    -.14772509                     -0.0231        1.0000

            2 factors will be retained by the NFACTOR criterion.
```

**Figure 26.1.** Table of Eigenvalues from PROC FACTOR

As displayed in Figure 26.1, the prior communality estimates are set to the squared multiple correlations. Figure 26.1 also displays the table of eigenvalues, which are the variances of the principal factors, of the reduced correlation matrix. Each row of the table pertains to a single eigenvalue. Following the column of eigenvalues are three measures of each eigenvalue's relative size and importance. The first of these displays the difference between the eigenvalue and its successor. The last two columns display the individual and cumulative proportions that the corresponding factor contributes to the total variation. The last line displayed in Figure 26.1 states that two factors are retained, as specified by the NFACTORS= option in the PROC FACTOR statement.

```
                    The FACTOR Procedure
          Initial Factor Method: Principal Factors

                       Factor Pattern

                          Factor1         Factor2

         HomeWork1        0.31105        -0.26516
         HomeWork2        0.70521        -0.42151
         HomeWork3        0.83281        -0.01966
         HomeWork4        0.23315         0.54773
         HomeWork5        0.79715        -0.29570
         HomeWork6        0.73831        -0.24142
         MidTerm1         0.21725         0.58751
         MidTerm2         0.39266         0.64770
         FinalExam        0.52745         0.56953
```

**Figure 26.2.** Factor Pattern Matrix from PROC FACTOR

Figure 26.2 displays the factor pattern matrix. The factor pattern matrix is the matrix of correlations between variables and the common factors. When the factors are orthogonal, the pattern matrix is also equal to the matrix of standardized regression coefficients for predicting the variables using the extracted factors.

The pattern matrix suggests that the first factor represents general ability, with positive loadings from all variables. The second factor is more difficult to interpret, but it may represent a contrast between exam and homework scores, with the exception of the score for HomeWork4.

```
                    The FACTOR Procedure
          Initial Factor Method: Principal Factors

              Variance Explained by Each Factor

                    Factor1         Factor2

                  3.0021245       1.7831404


        Final Communality Estimates: Total = 4.785265

  HomeWork1       HomeWork2       HomeWork3       HomeWork4       HomeWork5

 0.16706002      0.67498965      0.69395408      0.35436611      0.72288063

          HomeWork6        MidTerm1        MidTerm2       FinalExam

         0.60338875      0.39236928      0.57369380      0.60256254
```

**Figure 26.3.** Variance Explained and Final Communality Estimates

Figure 26.3 displays the variance explained by each factor and the final communality estimates, including the total communality. The final communality estimates are the proportion of variance of the variables accounted for by the common factors. When the factors are orthogonal, the final communalities are calculated by taking the sum of squares of each row of the factor pattern matrix. For example, the final communality estimate for the variable FinalExam is computed as follows:

$$0.60256254 = (0.52745)^2 + (0.56953)^2$$

Figure 26.4 displays the results of the VARIMAX rotation of the two extracted factors and the final communality estimates of the rotated factors.

The rotated factor pattern matrix is calculated by postmultiplying the original factor pattern matrix (Figure 26.2) by the orthogonal transformation matrix (Figure 26.4).

```
                The FACTOR Procedure
              Rotation Method: Varimax

          Orthogonal Transformation Matrix

                         1               2

         1         0.89675         0.44254
         2        -0.44254         0.89675


              Rotated Factor Pattern

                    Factor1         Factor2

   HomeWork1        0.39628        -0.10013
   HomeWork2        0.81893        -0.06590
   HomeWork3        0.75552         0.35093
   HomeWork4       -0.03332         0.59435
   HomeWork5        0.84570         0.08761
   HomeWork6        0.76892         0.11024
   MidTerm1        -0.06518         0.62299
   MidTerm2         0.06549         0.75459
   FinalExam        0.22095         0.74414
```

**Figure 26.4.** Transformation Matrix and Rotated Factor Pattern

The rotated factor pattern matrix is somewhat simpler to interpret: the rotated Factor1 can now be interpreted as general ability in homework performance. The homework variables load higher on Factor1 (with the single exception of the variable HomeWork4), with small loadings for the exam score variables. The rotated Factor2 seems to measure exam performance or test-taking ability. The exam score variables load heavily on Factor2, as does HomeWork4.

```
                           The FACTOR Procedure
                         Rotation Method: Varimax

                     Variance Explained by Each Factor

                           Factor1         Factor2

                         2.7633918       2.0218731


                 Final Community Estimates: Total = 4.785265

 HomeWork1       HomeWork2       HomeWork3       HomeWork4       HomeWork5

0.16706002      0.67498965      0.69395408      0.35436611      0.72288063

         HomeWork6        MidTerm1        MidTerm2       FinalExam

        0.60338875      0.39236928      0.57369380      0.60256254
```

**Figure 26.5.** Variance Explained and Final Communality Estimates after Rotation

Figure 26.5 displays the variance explained by each factor and the final communality estimates. Even though the variance explained by the rotated Factor1 is less than that explained by the unrotated factor (compare with Figure 26.3), the cumulative variance explained by both common factors remains the same after the orthogonal rotation. Also note that the VARIMAX rotation, as with any orthogonal rotation, has not changed the final communalities.

# Syntax

You can specify the following statements with the FACTOR procedure.

**PROC FACTOR** < *options* > **;**
  **VAR** *variables* **;**
  **PRIORS** *communalities* **;**
  **PARTIAL** *variables* **;**
  **FREQ** *variable* **;**
  **WEIGHT** *variable* **;**
  **BY** *variables* **;**

Usually only the VAR statement is needed in addition to the PROC FACTOR statement. The descriptions of the BY, FREQ, PARTIAL, PRIORS, VAR, and WEIGHT statements follow the description of the PROC FACTOR statement in alphabetical order.

## PROC FACTOR Statement

**PROC FACTOR** < *options* > **;**

The options available with the PROC FACTOR statement are listed in the following table and then are described in alphabetical order.

**Table 26.1.** Options Available in the PROC FACTOR Statement

| **Task** | **Option** |
|---|---|
| Data sets | DATA=<br>OUT=<br>OUTSTAT=<br>TARGET= |
| Extract factors and communalities | HEYWOOD<br>METHOD=<br>PRIORS=<br>RANDOM=<br>ULTRAHEYWOOD |
| Analyze data | COVARIANCE<br>NOINT<br>VARDEF=<br>WEIGHT |
| Specify number of factors | MINEIGEN=<br>NFACTORS=<br>PROPORTION= |
| Specify numerical properties | CONVERGE=<br>MAXITER=<br>SINGULAR= |
| Specify rotation method | GAMMA=<br>HKPOWER=<br>NORM=<br>POWER=<br>PREROTATE=<br>ROTATE= |
| Control displayed output | ALL<br>CORR<br>EIGENVECTORS<br>MSA<br>NOPRINT<br>NPLOT=<br>PLOT<br>PREPLOT<br>PRINT<br>REORDER |

**Table 26.1.** (continued)

| Task | Option |
|---|---|
| | RESIDUALS<br>SCORE<br>SCREE<br>SIMPLE |
| Exclude the correlation matrix from the OUTSTAT= data set | NOCORR |
| Miscellaneous | NOBS= |

**ALL**
displays all optional output except plots. When the input data set is TYPE=CORR, TYPE=UCORR, TYPE=COV, TYPE=UCOV or TYPE=FACTOR, simple statistics, correlations, and MSA are not displayed.

**CONVERGE=*p***
**CONV=*p***
specifies the convergence criterion for the METHOD=PRINIT, METHOD=ULS, METHOD=ALPHA, or METHOD=ML option. Iteration stops when the maximum change in the communalities is less than the value of the CONVERGE= option. The default value is 0.001. Negative values are not allowed.

**CORR**
**C**
displays the correlation matrix or partial correlation matrix.

**COVARIANCE**
**COV**
requests factoring of the covariance matrix instead of the correlation matrix. The COV option can be used only with the METHOD=PRINCIPAL, METHOD=PRINIT, METHOD=ULS, or METHOD=IMAGE option.

**DATA=*SAS-data-set***
specifies the input data set, which can be an ordinary SAS data set or a specially structured SAS data set as described in the section "Input Data Set" beginning on page 1146. If the DATA= option is omitted, the most recently created SAS data set is used.

**EIGENVECTORS**
**EV**
displays the eigenvectors. PROC FACTOR chooses the solution that makes the sum of the elements of each eigenvector nonnegative. If the sum of the elements is equal to zero, then the sign depends on how the number is rounded off.

**GAMMA=*p***
specifies the orthomax weight used with the option ROTATE=ORTHOMAX or PREROTATE=ORTHOMAX. There is no restriction on valid values.

**HEYWOOD**
**HEY**

sets to 1 any communality greater than 1, allowing iterations to proceed.

**HKPOWER=*p***
**HKP=*p***

specifies the power of the square roots of the eigenvalues used to rescale the eigenvectors for Harris-Kaiser (ROTATE=HK) rotation. Values between 0.0 and 1.0 are reasonable. The default value is 0.0, yielding the independent cluster solution (each variable tends to have a large loading on only one factor). A value of 1.0 is equivalent to a varimax rotation. You can also specify the HKPOWER= option with the ROTATE=QUARTIMAX, ROTATE=VARIMAX, ROTATE=EQUAMAX, or ROTATE=ORTHOMAX option, in which case the Harris-Kaiser rotation uses the specified orthogonal rotation method.

**MAXITER=*n***

specifies the maximum number of iterations. You can use the MAXITER= option with the PRINIT, ULS, ALPHA, or ML methods. The default is 30.

**METHOD=*name***
**M=*name***

specifies the method for extracting factors. The default is METHOD=PRINCIPAL unless the DATA= data set is TYPE=FACTOR, in which case the default is METHOD=PATTERN. Valid values for *name* are as follows:

| | |
|---|---|
| ALPHA \| A | produces alpha factor analysis. |
| HARRIS \| H | yields Harris component analysis of $\mathbf{S}^{-1}\mathbf{R}\mathbf{S}^{-1}$ (Harris 1962), a noniterative approximation to canonical component analysis. |
| IMAGE \| I | yields principal component analysis of the image covariance matrix, not Kaiser's (1963, 1970) or Kaiser and Rice's (1974) image analysis. A nonsingular correlation matrix is required. |
| ML \| M | performs maximum-likelihood factor analysis with an algorithm due, except for minor details, to Fuller (1987). The option METHOD=ML requires a nonsingular correlation matrix. |
| PATTERN | reads a factor pattern from a TYPE=FACTOR, TYPE=CORR, TYPE=UCORR, TYPE=COV or TYPE=UCOV data set. If you create a TYPE=FACTOR data set in a DATA step, only observations containing the factor pattern (_TYPE_='PATTERN') and, if the factors are correlated, the interfactor correlations (_TYPE_='FCORR') are required. |
| PRINCIPAL \| PRIN \| P | yields principal component analysis if no PRIORS option or statement is used or if you specify PRIORS=ONE; if you specify a PRIORS statement or a PRIORS= value other than PRIORS=ONE, a principal factor analysis is performed. |
| PRINIT | yields iterated principal factor analysis. |
| SCORE | reads scoring coefficients (_TYPE_='SCORE') from a TYPE=FACTOR, TYPE=CORR, TYPE=UCORR, TYPE=COV, |

or TYPE=UCOV data set. The data set must also contain either a correlation or a covariance matrix. Scoring coefficients are also displayed if you specify the OUT= option.

ULS | U produces unweighted least squares factor analysis.

**MINEIGEN=*p***
**MIN=*p***

specifies the smallest eigenvalue for which a factor is retained. If you specify two or more of the MINEIGEN=, NFACTORS=, and PROPORTION= options, the number of factors retained is the minimum number satisfying any of the criteria. The MINEIGEN= option cannot be used with either the METHOD=PATTERN or the METHOD=SCORE option. Negative values are not allowed. The default is 0 unless you omit both the NFACTORS= and the PROPORTION= options and one of the following conditions holds:

- If you specify the METHOD=ALPHA or METHOD=HARRIS option, then MINEIGEN=1.
- If you specify the METHOD=IMAGE option, then

$$\text{MINEIGEN} = \frac{\text{total image variance}}{\text{number of variables}}$$

- For any other METHOD= specification, if prior communality estimates of 1.0 are used, then

$$\text{MINEIGEN} = \frac{\text{total weighted variance}}{\text{number of variables}}$$

  When an unweighted correlation matrix is factored, this value is 1.

**MSA**

produces the partial correlations between each pair of variables controlling for all other variables (the negative anti-image correlations) and Kaiser's measure of sampling adequacy (Kaiser 1970; Kaiser and Rice 1974; Cerny and Kaiser 1977).

**NFACTORS=*n***
**NFACT=*n***
**N=*n***

specifies the maximum number of factors to be extracted and determines the amount of memory to be allocated for factor matrices. The default is the number of variables. Specifying a number that is small relative to the number of variables can substantially decrease the amount of memory required to run PROC FACTOR, especially with oblique rotations. If you specify two or more of the NFACTORS=, MINEIGEN=, and PROPORTION= options, the number of factors retained is the minimum number satisfying any of the criteria. If you specify the option NFACTORS=0, eigenvalues are computed, but no factors are extracted. If you specify the option NFACTORS=−1, neither eigenvalues nor factors are computed. You can use the NFACTORS= option with the METHOD=PATTERN or METHOD=SCORE option to specify a smaller number of factors than are present in the data set.

**NOBS=*n***

specifies the number of observations. If the DATA= input data set is a raw data set, *nobs* is defined by default to be the number of observations in the raw data set. The NOBS= option overrides this default definition. If the DATA= input data set contains a covariance, correlation, or scalar product matrix, the number of observations can be specified either by using the NOBS= option in the PROC FACTOR statement or by including a _TYPE_='N' observation in the DATA= input data set.

**NOCORR**

prevents the correlation matrix from being transferred to the OUTSTAT= data set when you specify the METHOD=PATTERN option. The NOCORR option greatly reduces memory requirements when there are many variables but few factors. The NOCORR option is not effective if the correlation matrix is required for other requested output; for example, if the scores or the residual correlations are displayed (using SCORE, RESIDUALS, ALL options).

**NOINT**

omits the intercept from the analysis; covariances or correlations are not corrected for the mean.

**NOPRINT**

suppresses the display of all output. Note that this option temporarily disables the Output Delivery System (ODS). For more information, see Chapter 15, "Using the Output Delivery System."

**NORM=COV | KAISER | NONE | RAW | WEIGHT**

specifies the method for normalizing the rows of the factor pattern for rotation. If you specify the option NORM=KAISER, Kaiser's normalization is used $(\sum_j p_{ij}^2 = 1)$. If you specify the option NORM=WEIGHT, the rows are weighted by the Cureton-Mulaik technique (Cureton and Mulaik 1975). If you specify the option NORM=COV, the rows of the pattern matrix are rescaled to represent covariances instead of correlations. If you specify the option NORM=NONE or NORM=RAW, normalization is not performed. The default is NORM=KAISER.

**NPLOT=*n***

specifies the number of factors to be plotted. The default is to plot all factors. The smallest allowable value is 2. If you specify the option NPLOT=$n$, all pairs of the first $n$ factors are plotted, producing a total of $n(n-1)/2$ plots.

**OUT=*SAS-data-set***

creates a data set containing all the data from the DATA= data set plus variables called Factor1, Factor2, and so on, containing estimated factor scores. The DATA= data set must contain multivariate data, not correlations or covariances. You must also specify the NFACTORS= option to determine the number of factor score variables. If you want to create a permanent SAS data set, you must specify a two-level name. Refer to "SAS Files" in *SAS Language Reference: Concepts* for more information on permanent data sets.

**OUTSTAT=***SAS-data-set*
specifies an output data set containing most of the results of the analysis. The output data set is described in detail in the section “Output Data Sets” on page 1149. If you want to create a permanent SAS data set, you must specify a two-level name. Refer to “SAS Files” in *SAS Language Reference: Concepts* for more information on permanent data sets.

**PLOT**
plots the factor pattern after rotation.

**POWER=***n*
specifies the power to be used in computing the target pattern for the option ROTATE=PROMAX. Valid values must be integers $\geq 1$. The default value is 3.

**PREPLOT**
plots the factor pattern before rotation.

**PREROTATE=***name*
**PRE=***name*
specifies the prerotation method for the option ROTATE=PROMAX. Any rotation method other than PROMAX or PROCRUSTES can be used. The default is PREROTATE=VARIMAX. If a previously rotated pattern is read using the option METHOD=PATTERN, you should specify the PREROTATE=NONE option.

**PRINT**
displays the input factor pattern or scoring coefficients and related statistics. In oblique cases, the reference and factor structures are computed and displayed. The PRINT option is effective only with the option METHOD=PATTERN or METHOD=SCORE.

**PRIORS=***name*
specifies a method for computing prior communality estimates. You can specify numeric values for the prior communality estimates by using the PRIORS statement. Valid values for *name* are as follows:

| | |
|---|---|
| ASMC \| A | sets the prior communality estimates proportional to the squared multiple correlations but adjusted so that their sum is equal to that of the maximum absolute correlations (Cureton 1968). |
| INPUT \| I | reads the prior communality estimates from the first observation with either _TYPE_='PRIORS' or _TYPE_='COMMUNAL' in the DATA= data set (which must be TYPE=FACTOR). |
| MAX \| M | sets the prior communality estimate for each variable to its maximum absolute correlation with any other variable. |
| ONE \| O | sets all prior communalities to 1.0. |
| RANDOM \| R | sets the prior communality estimates to pseudo-random numbers uniformly distributed between 0 and 1. |
| SMC \| S | sets the prior communality estimate for each variable to its squared multiple correlation with all other variables. |

The default prior communality estimates are as follows.

| METHOD= | PRIORS= |
|---|---|
| PRINCIPAL | ONE |
| PRINIT | ONE |
| ALPHA | SMC |
| ULS | SMC |
| ML | SMC |
| HARRIS | (not applicable) |
| IMAGE | (not applicable) |
| PATTERN | (not applicable) |
| SCORE | (not applicable) |

By default, the options METHOD=PRINIT, METHOD=ULS, METHOD=ALPHA, and METHOD=ML stop iterating and set the number of factors to 0 if an estimated communality exceeds 1. The options HEYWOOD and ULTRAHEYWOOD allow processing to continue.

**PROPORTION=***p*
**PERCENT=***p*
**P=***p*

specifies the proportion of common variance to be accounted for by the retained factors using the prior communality estimates. If the value is greater than one, it is interpreted as a percentage and divided by 100. The options PROPORTION=0.75 and PERCENT=75 are equivalent. The default value is 1.0 or 100%. You cannot specify the PROPORTION= option with the METHOD=PATTERN or METHOD=SCORE option. If you specify two or more of the PROPORTION=, NFACTORS=, and MINEIGEN= options, the number of factors retained is the minimum number satisfying any of the criteria.

**RANDOM=***n*

specifies a positive integer as a starting value for the pseudo-random number generator for use with the option PRIORS=RANDOM. If you do not specify the RANDOM= option, the time of day is used to initialize the pseudo-random number sequence. Valid values must be integers $\geq 1$.

**REORDER**
**RE**

causes the rows (variables) of various factor matrices to be reordered on the output. Variables with their highest absolute loading (reference structure loading for oblique rotations) on the first factor are displayed first, from largest to smallest loading, followed by variables with their highest absolute loading on the second factor, and so on. The order of the variables in the output data set is not affected. The factors are not reordered.

**RESIDUALS**
**RES**

displays the residual correlation matrix and the associated partial correlation matrix. The diagonal elements of the residual correlation matrix are the unique variances.

**ROTATE=***name*
**R=***name*

specifies the rotation method. The default is ROTATE=NONE. The following orthogonal rotation methods are available in the FACTOR procedure: EQUAMAX, ORTHOMAX, QUARTIMAX, PARSIMAX, and VARIMAX.

After the initial factor extraction, the common factors are uncorrelated with each other. If the factors are rotated by an *orthogonal transformation,* the rotated factors are also uncorrelated. If the factors are rotated by an *oblique transformation,* the rotated factors become correlated. Oblique rotations often produce more useful patterns than do orthogonal rotations. However, a consequence of correlated factors is that there is no single unambiguous measure of the importance of a factor in explaining a variable. Thus, for oblique rotations, the pattern matrix does not provide all the necessary information for interpreting the factors; you must also examine the *factor structure* and the *reference structure*. Refer to Harman (1976) and Mulaik (1972) for further information.

Valid values for *name* are as follows:

EQUAMAX | E — specifies orthogonal equamax rotation. This corresponds to the specification ROTATE=ORTHOMAX with GAMMA=*number of factors*/2.

HK — specifies Harris-Kaiser case II orthoblique rotation. You can use the HKPOWER= option to set the power of the square roots of the eigenvalues by which the eigenvectors are scaled.

NONE | N — specifies that no rotation be performed.

ORTHOMAX — specifies general orthomax rotation with the weight specified by the GAMMA= option.

PARSIMAX — specifies orthogonal Parsimax rotation. This corresponds to the specification ROTATE=ORTHOMAX with

$$\text{GAMMA} = \frac{nvar \times (nfact - 1)}{nvar + nfact - 2}$$

where *nvar* is the number of variables, and *nfact* is the number of factors.

PROCRUSTES — specifies oblique Procrustes rotation with target pattern provided by the TARGET= data set. The unrestricted least squares method is used with factors scaled to unit variance after rotation.

PROMAX | P — specifies oblique promax rotation. The PREROTATE= and POWER= options can be used with the option ROTATE=PROMAX.

QUARTIMAX | Q specifies orthogonal quartimax rotation. This corresponds to the specification ROTATE=ORTHOMAX with GAMMA=0.

VARIMAX | V specifies orthogonal varimax rotation. This corresponds to the specification ROTATE=ORTHOMAX with GAMMA=1.

**SCORE**

displays the factor scoring coefficients. The squared multiple correlation of each factor with the variables is also displayed except in the case of unrotated principal components.

**SCREE**

displays a scree plot of the eigenvalues (Cattell 1966, 1978; Cattell and Vogelman 1977; Horn and Engstrom 1979).

**SIMPLE**
**S**

displays means, standard deviations, and the number of observations.

**SINGULAR=*p***
**SING=*p***

specifies the singularity criterion, where $0 < p < 1$. The default value is 1E$-$8.

**TARGET=*SAS-data-set***

specifies an input data set containing the target pattern for Procrustes rotation (see the description of the ROTATE= option). The TARGET= data set must contain variables with the same names as those being factored. Each observation in the TARGET= data set becomes one column of the target factor pattern. Missing values are treated as zeros. The _NAME_ and _TYPE_ variables are not required and are ignored if present.

**ULTRAHEYWOOD**
**ULTRA**

allows communalities to exceed 1. The ULTRAHEYWOOD option can cause convergence problems because communalities can become extremely large, and ill-conditioned Hessians may occur.

**VARDEF=DF | N | WDF | WEIGHT | WGT**

specifies the divisor used in the calculation of variances and covariances. The default value is VARDEF=DF. The values and associated divisors are displayed in the following table where $i$= 0 if the NOINT option is used and $i$= 1 otherwise, and where $k$ is the number of partial variables specified in the PARTIAL statement.

| **Value** | **Description** | **Divisor** |
|---|---|---|
| DF | degrees of freedom | $n - k - i$ |
| N | number of observations | $n - k$ |
| WDF | sum of weights DF | $\sum_i w_i - k - i$ |
| WEIGHT \| WGT | sum of weights | $\sum_i w_i - k$ |

**WEIGHT**

factors a weighted correlation or covariance matrix. The WEIGHT option can be used only with the METHOD=PRINCIPAL, METHOD=PRINIT, METHOD=ULS, or METHOD=IMAGE option. The input data set must be of type CORR, UCORR, COV, UCOV or FACTOR, and the variable weights are obtained from an observation with _TYPE_='WEIGHT'.

## BY Statement

**BY** *variables* **;**

You can specify a BY statement with PROC FACTOR to obtain separate analyses on observations in groups defined by the BY variables. When a BY statement appears, the procedure expects the input data set to be sorted in order of the BY variables.

If your input data set is not sorted in ascending order, use one of the following alternatives:

- Sort the data using the SORT procedure with a similar BY statement.
- Specify the BY statement option NOTSORTED or DESCENDING in the BY statement for the FACTOR procedure. The NOTSORTED option does not mean that the data are unsorted but rather that the data are arranged in groups (according to values of the BY variables) and that these groups are not necessarily in alphabetical or increasing numeric order.
- Create an index on the BY variables using the DATASETS procedure (in Base SAS software). For more information on creating indexes and using the BY statement with indexed datasets, refer to "SAS Files" in *SAS Language Reference: Concepts*.

If you specify the TARGET= option and the TARGET= data set does not contain any of the BY variables, then the entire TARGET= data set is used as a Procrustean target for each BY group in the DATA= data set.

If the TARGET= data set contains some but not all of the BY variables, or if some BY variables do not have the same type or length in the TARGET= data set as in the DATA= data set, then PROC FACTOR displays an error message and stops.

If all the BY variables appear in the TARGET= data set with the same type and length as in the DATA= data set, then each BY group in the TARGET= data set is used as a Procrustean target for the corresponding BY group in the DATA= data set. The BY groups in the TARGET= data set must be in the same order as in the DATA= data set. If you specify the NOTSORTED option in the BY statement, there must be identical BY groups in the same order in both data sets. If you do not specify the NOTSORTED option, some BY groups can appear in one data set but not in the other.

For more information on the BY statement, refer to the discussion in *SAS Language Reference: Concepts*. For more information on the DATASETS procedure, refer to the discussion in the *SAS Procedures Guide*.

## FREQ Statement

**FREQ** *variable* **;**

If a variable in the data set represents the frequency of occurrence for the other values in the observation, include the variable's name in a FREQ statement. The procedure then treats the data set as if each observation appears $n$ times, where $n$ is the value of the FREQ variable for the observation. The total number of observations is considered to be equal to the sum of the FREQ variable when the procedure determines degrees of freedom for significance probabilities.

If the value of the FREQ variable is missing or is less than one, the observation is not used in the analysis. If the value is not an integer, the value is truncated to an integer.

The WEIGHT and FREQ statements have a similar effect, except in determining the number of observations for significance tests.

## PARTIAL Statement

**PARTIAL** *variables* **;**

If you want the analysis to be based on a partial correlation or covariance matrix, use the PARTIAL statement to list the variables that are used to partial out the variables in the analysis.

## PRIORS Statement

**PRIORS** *communalities* **;**

The PRIORS statement specifies numeric values between 0.0 and 1.0 for the prior communality estimates for each variable. The first numeric value corresponds to the first variable in the VAR statement, the second value to the second variable, and so on. The number of numeric values must equal the number of variables. For example,

```
proc factor;
   var     x  y  z;
   priors .7 .8 .9;
run;
```

You can specify various methods for computing prior communality estimates with the PRIORS= option of the PROC FACTOR statement. Refer to the description of that option for more information on the default prior communality estimates.

## VAR Statement

**VAR** *variables* **;**

The VAR statement specifies the numeric variables to be analyzed. If the VAR statement is omitted, all numeric variables not specified in other statements are analyzed.

## WEIGHT Statement

**WEIGHT** *variable* **;**

If you want to use relative weights for each observation in the input data set, specify a variable containing weights in a WEIGHT statement. This is often done when the variance associated with each observation is different and the values of the weight variable are proportional to the reciprocals of the variances. If a variable value is negative or is missing, it is excluded from the analysis.

# Details

## Incompatibilities with Earlier Versions of PROC FACTOR

PROC FACTOR no longer supports the FUZZ, FLAG, and ROUND options. However, a more flexible form of formatting is available. For an example of creating customized output, see Example 26.2.

## Input Data Set

The FACTOR procedure can read an ordinary SAS data set containing raw data or a special data set specified as a TYPE=CORR, TYPE=UCORR, TYPE=SSCP, TYPE=COV, TYPE=UCOV, or TYPE=FACTOR data set containing previously computed statistics. A TYPE=CORR data set can be created by the CORR procedure or various other procedures such as the PRINCOMP procedure. It contains means, standard deviations, the sample size, the correlation matrix, and possibly other statistics if it is created by some procedure other than PROC CORR. A TYPE=COV data set is similar to a TYPE=CORR data set but contains a covariance matrix. A TYPE=UCORR or TYPE=UCOV data set contains a correlation or covariance matrix that is not corrected for the mean. The default VAR variable list does not include Intercept if the DATA= data set is TYPE=SSCP. If the Intercept variable is explicitly specified in the VAR statement with a TYPE=SSCP data set, the NOINT option is activated. A TYPE=FACTOR data set can be created by the FACTOR procedure and is described in the section “Output Data Sets” on page 1149.

If your data set has many observations and you plan to run FACTOR several times, you can save computer time by first creating a TYPE=CORR data set and using it as input to PROC FACTOR.

```
proc corr data=raw out=correl;     /* create TYPE=CORR data set */
proc factor data=correl method=ml; /* maximum likelihood        */
proc factor data=correl;           /* principal components      */
```

The data set created by the CORR procedure is automatically given the TYPE=CORR data set option, so you do not have to specify TYPE=CORR. However, if you use a DATA step with a SET statement to modify the correlation data set, you must use the TYPE=CORR attribute in the new data set. You can use a VAR statement with PROC FACTOR when reading a TYPE=CORR data set to select a subset of the variables or change the order of the variables.

Problems can arise from using the CORR procedure when there are missing data. By default, PROC CORR computes each correlation from all observations that have values present for the pair of variables involved (pairwise deletion). The resulting correlation matrix may have negative eigenvalues. If you specify the NOMISS option with the CORR procedure, observations with any missing values are completely omitted from the calculations (listwise deletion), and there is no danger of negative eigenvalues.

PROC FACTOR can also create a TYPE=FACTOR data set, which includes all the information in a TYPE=CORR data set, and use it for repeated analyses. For a TYPE=FACTOR data set, the default value of the METHOD= option is PATTERN. The following statements produce the same PROC FACTOR results as the previous example:

```
proc factor data=raw method=ml outstat=fact; /* max. likelihood */
proc factor data=fact method=prin;      /* principal components */
```

You can use a TYPE=FACTOR data set to try several different rotation methods on the same data without repeatedly extracting the factors. In the following example, the second and third PROC FACTOR statements use the data set fact created by the first PROC FACTOR statement:

```
proc factor data=raw outstat=fact; /* principal components */
proc factor rotate=varimax;        /* varimax rotation     */
proc factor rotate=quartimax;      /* quartimax rotation   */
```

You can create a TYPE=CORR, TYPE=UCORR, or TYPE=FACTOR data set in a DATA step. Be sure to specify the TYPE= option in parentheses after the data set name in the DATA statement and include the _TYPE_ and _NAME_ variables. In a TYPE=CORR data set, only the correlation matrix (_TYPE_='CORR') is necessary. It can contain missing values as long as every pair of variables has at least one nonmissing value.

```
data correl(type=corr);
   _TYPE_='CORR';
   input _NAME_ $ x y z;
   datalines;
x  1.0  .   .
y   .7 1.0  .
z   .5  .4 1.0
;
proc factor;
run;
```

You can create a TYPE=FACTOR data set containing only a factor pattern (_TYPE_='PATTERN') and use the FACTOR procedure to rotate it.

```
data pat(type=factor);
   _TYPE_='PATTERN';
   input _NAME_ $ x y z;
   datalines;
factor1  .5  .7  .3
factor2  .8  .2  .8
;
proc factor rotate=promax prerotate=none;
run;
```

If the input factors are oblique, you must also include the interfactor correlation matrix with _TYPE_='FCORR'.

```
data pat(type=factor);
   input _TYPE_ $ _NAME_ $ x y z;
   datalines;
pattern factor1  .5  .7  .3
pattern factor2  .8  .2  .8
fcorr   factor1 1.0  .2  .
fcorr   factor2  .2 1.0  .
;
proc factor rotate=promax prerotate=none;
run;
```

Some procedures, such as the PRINCOMP and CANDISC procedures, produce TYPE=CORR or TYPE=UCORR data sets containing scoring coefficients (_TYPE_='SCORE' or _TYPE_= 'USCORE'). These coefficients can be input to PROC FACTOR and rotated by using the METHOD=SCORE option. The input data set must contain the correlation matrix as well as the scoring coefficients.

```
proc princomp data=raw n=2 outstat=prin;
run;
proc factor data=prin method=score rotate=varimax;
run;
```

## Output Data Sets

### *The OUT= Data Set*

The OUT= data set contains all the data in the DATA= data set plus new variables called Factor1, Factor2, and so on, containing estimated factor scores. If more than 99 factors are requested, the new variable names are Fact1, Fact2, and so on. Each estimated factor score is computed as a linear combination of the standardized values of the variables that are factored. The coefficients are always displayed if the OUT= option is specified and are labeled "Standardized Scoring Coefficients."

### *The OUTSTAT= Data Set*

The OUTSTAT= data set is similar to the TYPE=CORR or TYPE=UCORR data set produced by the CORR procedure, but it is a TYPE=FACTOR data set and it contains many results in addition to those produced by PROC CORR. The OUTSTAT= data set contains observations with _TYPE_='UCORR' and _TYPE_='USTD' if you specify the NOINT option.

The output data set contains the following variables:

- the BY variables, if any
- two new character variables, _TYPE_ and _NAME_
- the variables analyzed, that is, those in the VAR statement, or, if there is no VAR statement, all numeric variables not listed in any other statement.

Each observation in the output data set contains some type of statistic as indicated by the _TYPE_ variable. The _NAME_ variable is blank except where otherwise indicated. The values of the _TYPE_ variable are as follows:

| _TYPE_ | Contents |
|---|---|
| MEAN | means |
| STD | standard deviations |
| USTD | uncorrected standard deviations |
| N | sample size |
| CORR | correlations. The _NAME_ variable contains the name of the variable corresponding to each row of the correlation matrix. |
| UCORR | uncorrected correlations. The _NAME_ variable contains the name of the variable corresponding to each row of the uncorrected correlation matrix. |
| IMAGE | image coefficients. The _NAME_ variable contains the name of the variable corresponding to each row of the image coefficient matrix. |
| IMAGECOV | image covariance matrix. The _NAME_ variable contains the name of the variable corresponding to each row of the image covariance matrix. |

| | |
|---|---|
| COMMUNAL | final communality estimates |
| PRIORS | prior communality estimates, or estimates from the last iteration for iterative methods |
| WEIGHT | variable weights |
| SUMWGT | sum of the variable weights |
| EIGENVAL | eigenvalues |
| UNROTATE | unrotated factor pattern. The _NAME_ variable contains the name of the factor. |
| RESIDUAL | residual correlations. The _NAME_ variable contains the name of the variable corresponding to each row of the residual correlation matrix. |
| PRETRANS | transformation matrix from prerotation. The _NAME_ variable contains the name of the factor. |
| PREROTAT | factor pattern from prerotation. The _NAME_ variable contains the name of the factor. |
| TRANSFOR | transformation matrix from rotation. The _NAME_ variable contains the name of the factor. |
| FCORR | interfactor correlations. The _NAME_ variable contains the name of the factor. |
| PATTERN | factor pattern. The _NAME_ variable contains the name of the factor. |
| RCORR | reference axis correlations. The _NAME_ variable contains the name of the factor. |
| REFERENC | reference structure. The _NAME_ variable contains the name of the factor. |
| STRUCTUR | factor structure. The _NAME_ variable contains the name of the factor. |
| SCORE | scoring coefficients. The _NAME_ variable contains the name of the factor. |
| USCORE | scoring coefficients to be applied without subtracting the mean from the raw variables. The _NAME_ variable contains the name of the factor. |

## Missing Values

If the DATA= data set contains data (rather than a matrix or factor pattern), then observations with missing values for any variables in the analysis are omitted from the computations. If a correlation or covariance matrix is read, it can contain missing values as long as every pair of variables has at least one nonmissing entry. Missing values in a pattern or scoring coefficient matrix are treated as zeros.

## Cautions

- The amount of time that FACTOR takes is roughly proportional to the cube of the number of variables. Factoring 100 variables, therefore, takes about 1000 times as long as factoring 10 variables. Iterative methods (PRINIT, ALPHA, ULS, ML) can also take 100 times as long as noniterative methods (PRINCIPAL, IMAGE, HARRIS).
- No computer program is capable of reliably determining the optimal number of factors since the decision is ultimately subjective. You should not blindly accept the number of factors obtained by default; instead, use your own judgment to make a decision.
- Singular correlation matrices cause problems with the options PRIORS=SMC and METHOD=ML. Singularities can result from using a variable that is the sum of other variables, coding too many dummy variables from a classification variable, or having more variables than observations.
- If you use the CORR procedure to compute the correlation matrix and there are missing data and the NOMISS option is not specified, then the correlation matrix may have negative eigenvalues.
- If a TYPE=CORR, TYPE=UCORR or TYPE=FACTOR data set is copied or modified using a DATA step, the new data set does not automatically have the same TYPE as the old data set. You must specify the TYPE= data set option in the DATA statement. If you try to analyze a data set that has lost its TYPE=CORR attribute, PROC FACTOR displays a warning message saying that the data set contains _NAME_ and _TYPE_ variables but analyzes the data set as an ordinary SAS data set.
- For a TYPE=FACTOR data set, the default is METHOD=PATTERN, not METHOD=PRIN.

## Factor Scores

The FACTOR procedure can compute estimated factor scores directly if you specify the NFACTORS= and OUT= options, or indirectly using the SCORE procedure. The latter method is preferable if you use the FACTOR procedure interactively to determine the number of factors, the rotation method, or various other aspects of the analysis. To compute factor scores for each observation using the SCORE procedure,

- use the SCORE option in the PROC FACTOR statement
- create a TYPE=FACTOR output data set with the OUTSTAT= option
- use the SCORE procedure with both the raw data and the TYPE=FACTOR data set
- do not use the TYPE= option in the PROC SCORE statement

For example, the following statements could be used:

```
proc factor data=raw score outstat=fact;
run;
proc score  data=raw score=fact out=scores;
run;
```

or

```
proc corr   data=raw out=correl;
run;
proc factor data=correl score outstat=fact;
run;
proc score  data=raw score=fact out=scores;
run;
```

A component analysis (principal, image, or Harris) produces scores with mean zero and variance one. If you have done a common factor analysis, the true factor scores have mean zero and variance one, but the computed factor scores are only estimates of the true factor scores. These estimates have mean zero but variance equal to the squared multiple correlation of the factor with the variables. The estimated factor scores may have small nonzero correlations even if the true factors are uncorrelated.

## Variable Weights and Variance Explained

A principal component analysis of a correlation matrix treats all variables as equally important. A principal component analysis of a covariance matrix gives more weight to variables with larger variances. A principal component analysis of a covariance matrix is equivalent to an analysis of a weighted correlation matrix, where the weight of each variable is equal to its variance. Variables with large weights tend to have larger loadings on the first component and smaller residual correlations than variables with small weights.

You may want to give weights to variables using values other than their variances. Mulaik (1972) explains how to obtain a maximally reliable component by means of a weighted principal component analysis. With the FACTOR procedure, you can indirectly give arbitrary weights to the variables by using the COV option and rescaling the variables to have variance equal to the desired weight, or you can give arbitrary weights directly by using the WEIGHT option and including the weights in a TYPE=CORR data set.

Arbitrary variable weights can be used with the METHOD=PRINCIPAL, METHOD=PRINIT, METHOD=ULS, or METHOD=IMAGE option. Alpha and ML factor analyses compute variable weights based on the communalities (Harman 1976, pp. 217-218). For alpha factor analysis, the weight of a variable is the reciprocal of its communality. In ML factor analysis, the weight is the reciprocal of the uniqueness. Harris component analysis uses weights equal to the reciprocal of one minus the squared multiple correlation of each variable with the other variables.

For uncorrelated factors, the variance explained by a factor can be computed with or without taking the weights into account. The usual method for computing variance accounted for by a factor is to take the sum of squares of the corresponding column of the factor pattern, yielding an unweighted result. If the square of each loading is multiplied by the weight of the variable before the sum is taken, the result is the weighted variance explained, which is equal to the corresponding eigenvalue except in image analysis. Whether the weighted or unweighted result is more important depends on the purpose of the analysis.

In the case of correlated factors, the variance explained by a factor can be computed with or without taking the other factors into account. If you want to ignore the other factors, the variance explained is given by the weighted or unweighted sum of squares of the appropriate column of the factor structure since the factor structure contains simple correlations. If you want to subtract the variance explained by the other factors from the amount explained by the factor in question (the Type II variance explained), you can take the weighted or unweighted sum of squares of the appropriate column of the reference structure because the reference structure contains semipartial correlations. There are other ways of measuring the variance explained. For example, given a prior ordering of the factors, you can eliminate from each factor the variance explained by previous factors and compute a Type I variance explained. Harman (1976, pp. 268-270) provides another method, which is based on direct and joint contributions.

## Heywood Cases and Other Anomalies

Since communalities are squared correlations, you would expect them always to lie between 0 and 1. It is a mathematical peculiarity of the common factor model, however, that final communality estimates may exceed 1. If a communality equals 1, the situation is referred to as a Heywood case, and if a communality exceeds 1, it is an ultra-Heywood case. An ultra-Heywood case implies that some unique factor has negative variance, a clear indication that something is wrong. Possible causes include

- bad prior communality estimates
- too many common factors
- too few common factors
- not enough data to provide stable estimates
- the common factor model is not an appropriate model for the data

An ultra-Heywood case renders a factor solution invalid. Factor analysts disagree about whether or not a factor solution with a Heywood case can be considered legitimate.

Theoretically, the communality of a variable should not exceed its reliability. Violation of this condition is called a quasi-Heywood case and should be regarded with the same suspicion as an ultra-Heywood case.

Elements of the factor structure and reference structure matrices can exceed 1 only in the presence of an ultra-Heywood case. On the other hand, an element of the factor pattern may exceed 1 in an oblique rotation.

The maximum-likelihood method is especially susceptible to quasi- or ultra-Heywood cases. During the iteration process, a variable with high communality is given a high weight; this tends to increase its communality, which increases its weight, and so on.

It is often stated that the squared multiple correlation of a variable with the other variables is a lower bound to its communality. This is true if the common factor model fits the data perfectly, but it is not generally the case with real data. A final communality estimate that is less than the squared multiple correlation can, therefore, indicate poor fit, possibly due to not enough factors. It is by no means as serious a problem as an ultra-Heywood case. Factor methods using the Newton-Raphson method can actually produce communalities less than 0, a result even more disastrous than an ultra-Heywood case.

The squared multiple correlation of a factor with the variables may exceed 1, even in the absence of ultra-Heywood cases. This situation is also cause for alarm. Alpha factor analysis seems to be especially prone to this problem, but it does not occur with maximum likelihood. If a squared multiple correlation is negative, there are too many factors retained.

With data that do not fit the common factor model perfectly, you can expect some of the eigenvalues to be negative. If an iterative factor method converges properly, the sum of the eigenvalues corresponding to rejected factors should be 0; hence, some eigenvalues are positive and some negative. If a principal factor analysis fails to yield any negative eigenvalues, the prior communality estimates are probably too large. Negative eigenvalues cause the cumulative proportion of variance explained to exceed 1 for a sufficiently large number of factors. The cumulative proportion of variance explained by the retained factors should be approximately 1 for principal factor analysis and should converge to 1 for iterative methods. Occasionally, a single factor can explain more than 100 percent of the common variance in a principal factor analysis, indicating that the prior communality estimates are too low.

If a squared canonical correlation or a coefficient alpha is negative, there are too many factors retained.

Principal component analysis, unlike common factor analysis, has none of these problems if the covariance or correlation matrix is computed correctly from a data set with no missing values. Various methods for missing value correlation or severe rounding of the correlations can produce negative eigenvalues in principal components.

## Time Requirements

$n$ = number of observations
$v$ = number of variables
$f$ = number of factors
$i$ = number of iterations during factor extraction
$r$ = length of iterations during factor rotation

| **The time required to compute...** | **is roughly proportional to** |
|---|---|
| an overall factor analysis | $iv^3$ |
| the correlation matrix | $nv^2$ |
| PRIORS=SMC or ASMC | $v^3$ |
| PRIORS=MAX | $v^2$ |
| eigenvalues | $v^3$ |
| final eigenvectors | $fv^2$ |
| ROTATE=VARIMAX, QUARTIMAX, EQUAMAX, ORTHOMAX, PARSIMAX, PROMAX, or HK | $rvf^2$ |
| ROTATE=PROCRUSTES | $vf^2$ |

Each iteration in the PRINIT or ALPHA method requires computation of eigenvalues and $f$ eigenvectors.

Each iteration in the ML or ULS method requires computation of eigenvalues and $v - f$ eigenvectors.

The amount of time that PROC FACTOR takes is roughly proportional to the cube of the number of variables. Factoring 100 variables, therefore, takes about 1000 times as long as factoring 10 variables. Iterative methods (PRINIT, ALPHA, ULS, ML) can also take 100 times as long as noniterative methods (PRINCIPAL, IMAGE, HARRIS).

## Displayed Output

PROC FACTOR output includes

- Mean and Std Dev (standard deviation) of each variable and the number of observations, if you specify the SIMPLE option
- Correlations, if you specify the CORR option
- Inverse Correlation Matrix, if you specify the ALL option
- Partial Correlations Controlling all other Variables (negative anti-image correlations), if you specify the MSA option. If the data are appropriate for the common factor model, the partial correlations should be small.

- Kaiser's Measure of Sampling Adequacy (Kaiser 1970; Kaiser and Rice 1974; Cerny and Kaiser 1977) both overall and for each variable, if you specify the MSA option. The MSA is a summary of how small the partial correlations are relative to the ordinary correlations. Values greater than 0.8 can be considered good. Values less than 0.5 require remedial action, either by deleting the offending variables or by including other variables related to the offenders.
- Prior Communality Estimates, unless 1.0s are used or unless you specify the METHOD=IMAGE, METHOD=HARRIS, METHOD=PATTERN, or METHOD=SCORE option
- Squared Multiple Correlations of each variable with all the other variables, if you specify the METHOD=IMAGE or METHOD=HARRIS option
- Image Coefficients, if you specify the METHOD=IMAGE option
- Image Covariance Matrix, if you specify the METHOD=IMAGE option
- Preliminary Eigenvalues based on the prior communalities, if you specify the METHOD=PRINIT, METHOD=ALPHA, METHOD=ML, or METHOD=ULS option. The table produced includes the Total and the Average of the eigenvalues, the Difference between successive eigenvalues, the Proportion of variation represented, and the Cumulative proportion of variation.
- the number of factors that are retained, unless you specify the METHOD=PATTERN or METHOD=SCORE option
- the Scree Plot of Eigenvalues, if you specify the SCREE option. The preliminary eigenvalues are used if you specify the METHOD=PRINIT, METHOD=ALPHA, METHOD=ML, or METHOD=ULS option.
- the iteration history, if you specify the METHOD=PRINIT, METHOD=ALPHA, METHOD=ML, or METHOD=ULS option. The table produced contains the iteration number (Iter); the Criterion being optimized (Joreskog 1977); the Ridge value for the iteration if you specify the METHOD=ML or METHOD=ULS option; the maximum Change in any communality estimate; and the Communalities
- Significance tests, if you specify the option METHOD=ML, including Bartlett's Chi-square, df, and Prob $> \chi^2$ for $H_0$: No common factors and $H_0$: factors retained are sufficient to explain the correlations. The variables should have an approximate multivariate normal distribution for the probability levels to be valid. Lawley and Maxwell (1971) suggest that the number of observations should exceed the number of variables by fifty or more, although Geweke and Singleton (1980) claim that as few as ten observations are adequate with five variables and one common factor. Certain regularity conditions must also be satisfied for Bartlett's $\chi^2$ test to be valid (Geweke and Singleton 1980), but in practice these conditions usually are satisfied. The notation Prob>chi**2 means "the probability under the null hypothesis of obtaining a greater $\chi^2$ statistic than that observed." The Chi-square value is displayed with and without Bartlett's correction.
- Akaike's Information Criterion, if you specify the METHOD=ML option. Akaike's information criterion (AIC) (Akaike 1973, 1974, 1987) is a general

criterion for estimating the best number of parameters to include in a model when maximum-likelihood estimation is used. The number of factors that yields the smallest value of AIC is considered best. Like the chi-square test, AIC tends to include factors that are statistically significant but inconsequential for practical purposes.

- Schwarz's Bayesian Criterion, if you specify the METHOD=ML option. Schwarz's Bayesian Criterion (SBC) (Schwarz 1978) is another criterion, similar to AIC, for determining the best number of parameters. The number of factors that yields the smallest value of SBC is considered best; SBC seems to be less inclined to include trivial factors than either AIC or the chi-square test.
- Tucker and Lewis's Reliability Coefficient, if you specify the METHOD=ML option (Tucker and Lewis 1973)
- Squared Canonical Correlations, if you specify the METHOD=ML option. These are the same as the squared multiple correlations for predicting each factor from the variables.
- Coefficient Alpha for Each Factor, if you specify the METHOD=ALPHA option
- Eigenvectors, if you specify the EIGENVECTORS or ALL option, unless you also specify the METHOD=PATTERN or METHOD=SCORE option
- Eigenvalues of the (Weighted) (Reduced) (Image) Correlation or Covariance Matrix, unless you specify the METHOD=PATTERN or METHOD=SCORE option. Included are the Total and the Average of the eigenvalues, the Difference between successive eigenvalues, the Proportion of variation represented, and the Cumulative proportion of variation.
- the Factor Pattern, which is equal to both the matrix of standardized regression coefficients for predicting variables from common factors and the matrix of correlations between variables and common factors since the extracted factors are uncorrelated
- Variance explained by each factor, both Weighted and Unweighted, if variable weights are used
- Final Communality Estimates, including the Total communality; or Final Communality Estimates and Variable Weights, including the Total communality, both Weighted and Unweighted, if variable weights are used. Final communality estimates are the squared multiple correlations for predicting the variables from the estimated factors, and they can be obtained by taking the sum of squares of each row of the factor pattern, or a weighted sum of squares if variable weights are used.
- Residual Correlations with Uniqueness on the Diagonal, if you specify the RESIDUAL or ALL option
- Root Mean Square Off-diagonal Residuals, both Over-all and for each variable, if you specify the RESIDUAL or ALL option
- Partial Correlations Controlling Factors, if you specify the RESIDUAL or ALL option

- Root Mean Square Off-diagonal Partials, both Over-all and for each variable, if you specify the RESIDUAL or ALL option
- Plots of Factor Pattern for unrotated factors, if you specify the PREPLOT option. The number of plots is determined by the NPLOT= option.
- Variable Weights for Rotation, if you specify the NORM=WEIGHT option
- Factor Weights for Rotation, if you specify the HKPOWER= option
- Orthogonal Transformation Matrix, if you request an orthogonal rotation
- Rotated Factor Pattern, if you request an orthogonal rotation
- Variance explained by each factor after rotation. If you request an orthogonal rotation and if variable weights are used, both weighted and unweighted values are produced.
- Target Matrix for Procrustean Transformation, if you specify the ROTATE=PROCRUSTES or ROTATE=PROMAX option
- the Procrustean Transformation Matrix, if you specify the ROTATE=PROCRUSTES or ROTATE=PROMAX option
- the Normalized Oblique Transformation Matrix, if you request an oblique rotation, which, for the option ROTATE=PROMAX, is the product of the prerotation and the Procrustean rotation
- Inter-factor Correlations, if you specify an oblique rotation
- Rotated Factor Pattern (Std Reg Coefs), if you specify an oblique rotation, giving standardized regression coefficients for predicting the variables from the factors
- Reference Axis Correlations if you specify an oblique rotation. These are the partial correlations between the primary factors when all factors other than the two being correlated are partialled out.
- Reference Structure (Semipartial Correlations), if you request an oblique rotation. The reference structure is the matrix of semipartial correlations (Kerlinger and Pedhazur 1973) between variables and common factors, removing from each common factor the effects of other common factors. If the common factors are uncorrelated, the reference structure is equal to the factor pattern.
- Variance explained by each factor eliminating the effects of all other factors, if you specify an oblique rotation. Both Weighted and Unweighted values are produced if variable weights are used. These variances are equal to the (weighted) sum of the squared elements of the reference structure corresponding to each factor.
- Factor Structure (Correlations), if you request an oblique rotation. The (primary) factor structure is the matrix of correlations between variables and common factors. If the common factors are uncorrelated, the factor structure is equal to the factor pattern.
- Variance explained by each factor ignoring the effects of all other factors, if you request an oblique rotation. Both Weighted and Unweighted values are produced if variable weights are used. These variances are equal to the (weighted)

sum of the squared elements of the factor structure corresponding to each factor.

- Final Communality Estimates for the rotated factors if you specify the ROTATE= option. The estimates should equal the unrotated communalities.
- Squared Multiple Correlations of the Variables with Each Factor, if you specify the SCORE or ALL option, except for unrotated principal components
- Standardized Scoring Coefficients, if you specify the SCORE or ALL option
- Plots of the Factor Pattern for rotated factors, if you specify the PLOT option and you request an orthogonal rotation. The number of plots is determined by the NPLOT= option.
- Plots of the Reference Structure for rotated factors, if you specify the PLOT option and you request an oblique rotation. The number of plots is determined by the NPLOT= option. Included are the Reference Axis Correlation and the Angle between the Reference Axes for each pair of factors plotted.

If you specify the ROTATE=PROMAX option, the output includes results for both the prerotation and the Procrustean rotation.

## ODS Table Names

PROC FACTOR assigns a name to each table it creates. You can use these names to reference the table when using the Output Delivery System (ODS) to select tables and create output data sets. These names are listed in the following table. For more information on ODS, see Chapter 15, "Using the Output Delivery System."

**Table 26.2.** ODS Tables Produced in PROC FACTOR

| ODS Table Name | Description | Option |
|---|---|---|
| AlphaCoef | Coefficient alpha for each factor | METHOD=ALPHA |
| CanCorr | Squared canonical correlations | METHOD=ML |
| CondStdDev | Conditional standard deviations | SIMPLE w/PARTIAL |
| ConvergenceStatus | Convergence status | METHOD=PRINIT, =ALPHA, =ML, or =ULS |
| Corr | Correlations | CORR |
| Eigenvalues | Eigenvalues | default, SCREE |
| Eigenvectors | Eigenvectors | EIGENVECTORS |
| FactorWeightRotate | Factor weights for rotation | HKPOWER= |
| FactorPattern | Factor pattern | default |
| FactorStructure | Factor structure | ROTATE= any oblique rotation |
| FinalCommun | Final communalities | default |
| FinalCommunWgt | Final communalities with weights | METHOD=ML, METHOD=ALPHA |
| FitMeasures | Measures of fit | METHOD=ML |
| ImageCoef | Image coefficients | METHOD=IMAGE |

**Table 26.2.** (continued)

| ODS Table Name | Description | Option |
|---|---|---|
| ImageCov | Image covariance matrix | METHOD=IMAGE |
| ImageFactors | Image factor matrix | METHOD=IMAGE |
| InputFactorPattern | Input factor pattern | PRINT |
| InputScoreCoef | Standardized input scoring coefficients | METHOD=SCORE |
| InterFactorCorr | Inter-factor correlations | ROTATE= any oblique rotation |
| InvCorr | Inverse correlation matrix | ALL |
| IterHistory | Iteration history | METHOD=PRINIT, =ALPHA, =ML, or =ULS |
| MultipleCorr | Squared multiple correlations | METHOD=IMAGE or METHOD=HARRIS |
| NormObliqueTrans | Normalized oblique transformation matrix | ROTATE= any oblique rotation |
| ObliqueRotFactPat | Rotated factor pattern | ROTATE= any oblique rotation |
| ObliqueTrans | Oblique transformation matrix | HKPOWER= |
| OrthRotFactPat | Rotated factor pattern | ROTATE= any orthogonal rotation |
| OrthTrans | Orthogonal transformation matrix | ROTATE= any orthogonal rotation |
| ParCorrControlFactor | Partial correlations controlling factors | RESIDUAL |
| ParCorrControlVar | Partial correlations controlling other variables | MSA |
| PartialCorr | Partial correlations | MSA, CORR w/PARTIAL |
| PriorCommunalEst | Prior communality estimates | PRIORS=, METHOD=ML, METHOD=ALPHA |
| ProcrustesTarget | Target matrix for Procrustean transformation | ROTATE=PROCRUSTES, ROTATE=PROMAX |
| ProcrustesTrans | Procrustean transformation matrix | ROTATE=PROCRUSTES, ROTATE=PROMAX |
| RMSOffDiagPartials | Root mean square off-diagonal partials | RESIDUAL |
| RMSOffDiagResids | Root mean square off-diagonal residuals | RESIDUAL |
| ReferenceAxisCorr | Reference axis correlations | ROTATE= any oblique rotation |
| ReferenceStructure | Reference structure | ROTATE= any oblique rotation |
| ResCorrUniqueDiag | Residual correlations with uniqueness on the diagonal | RESIDUAL |
| SamplingAdequacy | Kaiser's measure of sampling adequacy | MSA |
| SignifTests | Significance tests | METHOD=ML |
| SimpleStatistics | Simple statistics | SIMPLE |
| StdScoreCoef | Standardized scoring coefficients | SCORE |
| VarExplain | Variance explained | default |

**Table 26.2.** (continued)

| ODS Table Name | Description | Option |
|---|---|---|
| VarExplainWgt | Variance explained with weights | METHOD=ML, METHOD=ALPHA |
| VarFactorCorr | Squared multiple correlations of the variables with each factor | SCORE |
| VarWeightRotate | Variable weights for rotation | NORM=WEIGHT, ROTATE= |

# Examples

## Example 26.1. Principal Component Analysis

The following example analyzes socioeconomic data provided by Harman (1976). The five variables represent total population, median school years, total employment, miscellaneous professional services, and median house value. Each observation represents one of twelve census tracts in the Los Angeles Standard Metropolitan Statistical Area.

The first analysis is a principal component analysis. Simple descriptive statistics and correlations are also displayed. This example produces Output 26.1.1:

```
data SocioEconomics;
   title 'Five Socioeconomic Variables';
   title2 'See Page 14 of Harman: Modern Factor Analysis, 3rd Ed';
   input Population School Employment Services HouseValue;
   datalines;
5700     12.8      2500      270       25000
1000     10.9      600       10        10000
3400     8.8       1000      10        9000
3800     13.6      1700      140       25000
4000     12.8      1600      140       25000
8200     8.3       2600      60        12000
1200     11.4      400       10        16000
9100     11.5      3300      60        14000
9900     12.5      3400      180       18000
9600     13.7      3600      390       25000
9600     9.6       3300      80        12000
9400     11.4      4000      100       13000
;
proc factor data=SocioEconomics simple corr;
   title3 'Principal Component Analysis';
run;
```

There are two large eigenvalues, 2.8733 and 1.7967, which together account for 93.4% of the standardized variance. Thus, the first two principal components provide an adequate summary of the data for most purposes. Three components, explaining 97.7% of the variation, should be sufficient for almost any application.

PROC FACTOR retains two components on the basis of the eigenvalues-greater-than-one rule since the third eigenvalue is only 0.2148.

The first component has large positive loadings for all five variables. The correlation with Services (0.93239) is especially high. The second component is a contrast of Population (0.80642) and Employment (0.72605) against School ($-0.54476$) and HouseValue ($-0.55818$), with a very small loading on Services ($-0.10431$).

The final communality estimates show that all the variables are well accounted for by two components, with final communality estimates ranging from 0.880236 for Services to 0.987826 for Population.

**Output 26.1.1.** Principal Component Analysis

```
                    Five Socioeconomic Variables
          See Page 14 of Harman: Modern Factor Analysis, 3rd Ed
                    Principal Component Analysis

                       The FACTOR Procedure

           Means and Standard Deviations from 12 Observations

                Variable            Mean       Std Dev

                Population      6241.667     3439.9943
                School            11.442        1.7865
                Employment      2333.333     1241.2115
                Services         120.833      114.9275
                HouseValue     17000.000     6367.5313


                           Correlations

              Population         School     Employment       Services     HouseValue

Population       1.00000        0.00975        0.97245        0.43887        0.02241
School           0.00975        1.00000        0.15428        0.69141        0.86307
Employment       0.97245        0.15428        1.00000        0.51472        0.12193
Services         0.43887        0.69141        0.51472        1.00000        0.77765
HouseValue       0.02241        0.86307        0.12193        0.77765        1.00000
```

**Principal Component Analysis**

**The FACTOR Procedure**
**Initial Factor Method: Principal Components**

**Eigenvalues of the Correlation Matrix: Total = 5 Average = 1**

| | Eigenvalue | Difference | Proportion | Cumulative |
|---|---|---|---|---|
| 1 | 2.87331359 | 1.07665350 | 0.5747 | 0.5747 |
| 2 | 1.79666009 | 1.58182321 | 0.3593 | 0.9340 |
| 3 | 0.21483689 | 0.11490283 | 0.0430 | 0.9770 |
| 4 | 0.09993405 | 0.08467868 | 0.0200 | 0.9969 |
| 5 | 0.01525537 | | 0.0031 | 1.0000 |

**Factor Pattern**

| | Factor1 | Factor2 |
|---|---|---|
| Population | 0.58096 | 0.80642 |
| School | 0.76704 | -0.54476 |
| Employment | 0.67243 | 0.72605 |
| Services | 0.93239 | -0.10431 |
| HouseValue | 0.79116 | -0.55818 |

**Variance Explained by Each Factor**

| Factor1 | Factor2 |
|---|---|
| 2.8733136 | 1.7966601 |

**Final Communality Estimates: Total = 4.669974**

| Population | School | Employment | Services | HouseValue |
|---|---|---|---|---|
| 0.98782629 | 0.88510555 | 0.97930583 | 0.88023562 | 0.93750041 |

## Example 26.2. Principal Factor Analysis

The following example uses the data presented in Example 26.1, and performs a principal factor analysis with squared multiple correlations for the prior communality estimates (PRIORS=SMC).

To help determine if the common factor model is appropriate, Kaiser's measure of sampling adequacy (MSA) is requested, and the residual correlations and partial correlations are computed (RESIDUAL). To help determine the number of factors, a scree plot (SCREE) of the eigenvalues is displayed, and the PREPLOT option plots the unrotated factor pattern.

The ROTATE= and REORDER options are specified to enhance factor interpretability. The ROTATE=PROMAX option produces an orthogonal varimax prerotation followed by an oblique rotation, and the REORDER option reorders the variables according to their largest factor loadings. The PLOT procedure is used to produce a plot of the reference structure. An OUTSTAT= data set is created by PROC FACTOR and displayed in Output 26.2.15.

This example also demonstrates how to define a picture format with the FORMAT procedure and use the PRINT procedure to produce customized factor pattern output. Small elements of the Rotated Factor Pattern matrix are displayed as '.'. Large values are multiplied by 100, truncated at the decimal, and flagged with an asterisk '*'. Intermediate values are scaled by 100 and truncated. For more information on picture formats, refer to "Formats" in *SAS Language Reference: Dictionary*.

```
ods output ObliqueRotFactPat = rotfacpat;
proc factor data=SocioEconomics
     priors=smc msa scree residual preplot
     rotate=promax reorder plot
     outstat=fact_all;
   title3 'Principal Factor Analysis with Promax Rotation';

proc print;
   title3 'Factor Output Data Set';
run;

proc format;
   picture FuzzFlag
   low  - 0.1  = '   .   '
   0.10 - 0.90 = '009   ' (mult = 100)
   0.90 - high = '009 *' (mult = 100);
run;

proc print data = rotfacpat;
   format factor1-factor2 FuzzFlag.;
run;
```

**Output 26.2.1.** Principal Factor Analysis

```
                 Principal Factor Analysis with Promax Rotation

                              The FACTOR Procedure
                        Initial Factor Method: Principal Factors

              Partial Correlations Controlling all other Variables

              Population        School    Employment      Services    HouseValue

Population       1.00000      -0.54465       0.97083       0.09612       0.15871
School          -0.54465       1.00000       0.54373       0.04996       0.64717
Employment       0.97083       0.54373       1.00000       0.06689      -0.25572
Services         0.09612       0.04996       0.06689       1.00000       0.59415
HouseValue       0.15871       0.64717      -0.25572       0.59415       1.00000


         Kaiser's Measure of Sampling Adequacy: Overall MSA = 0.57536759

   Population          School      Employment        Services      HouseValue

   0.47207897      0.55158839      0.48851137      0.80664365      0.61281377

            2 factors will be retained by the PROPORTION criterion.
```

```
                 Principal Factor Analysis with Promax Rotation

                              The FACTOR Procedure
                        Initial Factor Method: Principal Factors

                          Prior Communality Estimates: SMC

   Population          School      Employment        Services      HouseValue

   0.96859160      0.82228514      0.96918082      0.78572440      0.84701921


                 Eigenvalues of the Reduced Correlation Matrix:
                    Total = 4.39280116  Average = 0.87856023

                  Eigenvalue    Difference    Proportion    Cumulative

             1    2.73430084    1.01823217        0.6225        0.6225
             2    1.71606867    1.67650586        0.3907        1.0131
             3    0.03956281    0.06408626        0.0090        1.0221
             4    -.02452345    0.04808427       -0.0056        1.0165
             5    -.07260772                     -0.0165        1.0000

            2 factors will be retained by the PROPORTION criterion.
```

Output 26.2.1 displays the results of the principal factor extraction.

If the data are appropriate for the common factor model, the partial correlations controlling the other variables should be small compared to the original correlations. The partial correlation between the variables School and HouseValue, for example, is 0.65, slightly less than the original correlation of 0.86. The partial correlation between Population and School is -0.54, which is much larger in absolute value than the original correlation; this is an indication of trouble. Kaiser's MSA is a summary, for each variable and for all variables together, of how much smaller the partial correlations are than the original correlations. Values of 0.8 or 0.9 are considered good, while MSAs below 0.5 are unacceptable. The variables Population, School, and Employment have very poor MSAs. Only the Services variable has a good MSA. The overall MSA of 0.58 is sufficiently poor that additional variables should be included in the analysis to better define the common factors. A commonly used rule is that there should be at least three variables per factor. In the following analysis, there seems to be two common factors in these data, so more variables are needed for a reliable analysis.

The SMCs are all fairly large; hence, the factor loadings do not differ greatly from the principal component analysis.

The eigenvalues show clearly that two common factors are present. There are two large positive eigenvalues that together account for 101.31% of the common variance, which is as close to 100% as you are ever likely to get without iterating. The scree plot displays a sharp bend at the third eigenvalue, reinforcing the preceding conclusion.

```
                 Principal Factor Analysis with Promax Rotation

                              The FACTOR Procedure
                    Initial Factor Method: Principal Factors

Scree Plot of Eigenvalues
     |
   3 +
     |
     |                  1
     |
     |
     |
   2 +
E    |
i    |                              2
g    |
e    |
n    |
v  1 +
a    |
l    |
u    |
e    |
s    |
   0 +                                          3           4           5
     |
     |
     |
     |
     |
  -1 +
     |
     -------+-----------+-----------+-----------+-----------+-----------+-------
            0           1           2           3           4           5

                                        Number
```

**Output 26.2.2.** Factor Pattern Matrix and Communalities

```
              Principal Factor Analysis with Promax Rotation

                           The FACTOR Procedure
                  Initial Factor Method: Principal Factors

                              Factor Pattern

                                   Factor1         Factor2

                Services           0.87899        -0.15847
                HouseValue         0.74215        -0.57806
                Employment         0.71447         0.67936
                School             0.71370        -0.55515
                Population         0.62533         0.76621


                     Variance Explained by Each Factor

                           Factor1         Factor2

                         2.7343008       1.7160687


               Final Communality Estimates: Total = 4.450370

Population          School      Employment        Services      HouseValue

0.97811334      0.81756387      0.97199928      0.79774304      0.88494998
```

As displayed in Output 26.2.2, the principal factor pattern is similar to the principal component pattern seen in Example 26.1. For example, the variable Services has the largest loading on the first factor, and the Population variable has the smallest. The variables Population and Employment have large positive loadings on the second factor, and the HouseValue and School variables have large negative loadings.

The final communality estimates are all fairly close to the priors. Only the communality for the variable HouseValue increased appreciably, from 0.847019 to 0.884950. Nearly 100% of the common variance is accounted for. The residual correlations (off-diagonal elements) are low, the largest being 0.03 (Output 26.2.3). The partial correlations are not quite as impressive, since the uniqueness values are also rather small. These results indicate that the SMCs are good but not quite optimal communality estimates.

**Output 26.2.3.** Residual and Partial Correlations

Principal Factor Analysis with Promax Rotation

The FACTOR Procedure
Initial Factor Method: Principal Factors

Residual Correlations With Uniqueness on the Diagonal

| | Population | School | Employment | Services | HouseValue |
|---|---|---|---|---|---|
| Population | 0.02189 | -0.01118 | 0.00514 | 0.01063 | 0.00124 |
| School | -0.01118 | 0.18244 | 0.02151 | -0.02390 | 0.01248 |
| Employment | 0.00514 | 0.02151 | 0.02800 | -0.00565 | -0.01561 |
| Services | 0.01063 | -0.02390 | -0.00565 | 0.20226 | 0.03370 |
| HouseValue | 0.00124 | 0.01248 | -0.01561 | 0.03370 | 0.11505 |

Root Mean Square Off-Diagonal Residuals: Overall = 0.01693282

| Population | School | Employment | Services | HouseValue |
|---|---|---|---|---|
| 0.00815307 | 0.01813027 | 0.01382764 | 0.02151737 | 0.01960158 |

Partial Correlations Controlling Factors

| | Population | School | Employment | Services | HouseValue |
|---|---|---|---|---|---|
| Population | 1.00000 | -0.17693 | 0.20752 | 0.15975 | 0.02471 |
| School | -0.17693 | 1.00000 | 0.30097 | -0.12443 | 0.08614 |
| Employment | 0.20752 | 0.30097 | 1.00000 | -0.07504 | -0.27509 |
| Services | 0.15975 | -0.12443 | -0.07504 | 1.00000 | 0.22093 |
| HouseValue | 0.02471 | 0.08614 | -0.27509 | 0.22093 | 1.00000 |

**Output 26.2.4.** Root Mean Square Off-Diagonal Partials

Principal Factor Analysis with Promax Rotation

The FACTOR Procedure
Initial Factor Method: Principal Factors

Root Mean Square Off-Diagonal Partials: Overall = 0.18550132

| Population | School | Employment | Services | HouseValue |
|---|---|---|---|---|
| 0.15850824 | 0.19025867 | 0.23181838 | 0.15447043 | 0.18201538 |

**Output 26.2.5.** Unrotated Factor Pattern Plot

```
                 Principal Factor Analysis with Promax Rotation

                              The FACTOR Procedure
                    Initial Factor Method: Principal Factors

Plot of Factor Pattern for Factor1 and Factor2

                                     Factor1
                                        1

                                   D   .9

                                       .8
                      E
                       B               .7                   C
                                                              A
                                       .6

                                       .5

                                       .4

                                       .3

                                       .2
                                                                        F
                                       .1                               a
                                                                        c
        -1 -.9-.8-.7-.6-.5-.4-.3-.2-.1  0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1.0t
                                                                        o
                                      -.1                               r
                                                                        2
                                      -.2

                                      -.3

                                      -.4

                                      -.5

                                      -.6

                                      -.7

                                      -.8

                                      -.9

                                       -1

       Population=A  School=B      Employment=C  Services=D    HouseValue=E
```

As displayed in Output 26.2.5, the unrotated factor pattern reveals two tight clusters of variables, with the variables HouseValue and School at the negative end of Factor2 axis and the variables Employment and Population at the positive end. The Services variable is in between but closer to the HouseValue and School variables. A good rotation would put the reference axes through the two clusters.

**Output 26.2.6.** Varimax Rotation: Transform Matrix and Rotated Pattern

```
               Principal Factor Analysis with Promax Rotation

                         The FACTOR Procedure
                     Prerotation Method: Varimax

                   Orthogonal Transformation Matrix

                                  1              2

                     1         0.78895        0.61446
                     2        -0.61446        0.78895


                        Rotated Factor Pattern

                             Factor1        Factor2

                 HouseValue    0.94072       -0.00004
                 School        0.90419        0.00055
                 Services      0.79085        0.41509
                 Population    0.02255        0.98874
                 Employment    0.14625        0.97499
```

**Output 26.2.7.** Varimax Rotation: Variance Explained and Communalities

```
               Principal Factor Analysis with Promax Rotation

                         The FACTOR Procedure
                     Prerotation Method: Varimax

                   Variance Explained by Each Factor

                         Factor1        Factor2

                       2.3498567      2.1005128


               Final Community Estimates: Total = 4.450370

  Population        School    Employment      Services    HouseValue

  0.97811334    0.81756387    0.97199928    0.79774304    0.88494998
```

**Output 26.2.8.** Varimax Rotated Factor Pattern Plot

```
                 Principal Factor Analysis with Promax Rotation

                              The FACTOR Procedure
                          Prerotation Method: Varimax

Plot of Factor Pattern for Factor1 and Factor2

                                    Factor1
                                       1
                                      E
                                      .B

                                      .8           D

                                      .7

                                      .6

                                      .5

                                      .4

                                      .3

                                      .2
                                                                C  F
                                      .1                           a
                                                                   c
        -1 -.9-.8-.7-.6-.5-.4-.3-.2-.1  0 .1 .2 .3 .4 .5 .6 .7 .8 .9 A.0t
                                                                   o
                                     -.1                           r
                                                                   2
                                     -.2

                                     -.3

                                     -.4

                                     -.5

                                     -.6

                                     -.7

                                     -.8

                                     -.9

                                      -1

       Population=A  School=B      Employment=C  Services=D    HouseValue=E
```

Output 26.2.6, Output 26.2.7 and Output 26.2.8 display the results of the varimax rotation. This rotation puts one axis through the variables HouseValue and School but misses the Population and Employment variables slightly.

**Output 26.2.9.** Promax Rotation: Procrustean Target and Transform Matrix

```
              Principal Factor Analysis with Promax Rotation

                          The FACTOR Procedure
                        Rotation Method: Promax

               Target Matrix for Procrustean Transformation

                                 Factor1         Factor2

                  HouseValue         1.00000        -0.00000
                  School             1.00000         0.00000
                  Services           0.69421         0.10045
                  Population         0.00001         1.00000
                  Employment         0.00326         0.96793


                    Procrustean Transformation Matrix

                                      1               2

                          1      1.04116598      -0.0986534
                          2      -0.1057226      0.96303019
```

**Output 26.2.10.** Promax Rotation: Oblique Transform Matrix and Correlation

```
              Principal Factor Analysis with Promax Rotation

                          The FACTOR Procedure
                        Rotation Method: Promax

                 Normalized Oblique Transformation Matrix

                                      1               2

                          1         0.73803         0.54202
                          2        -0.70555         0.86528


                       Inter-Factor Correlations

                              Factor1         Factor2

                   Factor1         1.00000         0.20188
                   Factor2         0.20188         1.00000
```

**Output 26.2.11.** Promax Rotation: Rotated Factor Pattern and Correlations

```
                Principal Factor Analysis with Promax Rotation

                            The FACTOR Procedure
                          Rotation Method: Promax

       Rotated Factor Pattern (Standardized Regression Coefficients)

                                  Factor1         Factor2

                 HouseValue      0.95558485      -0.0979201
                 School          0.91842142      -0.0935214
                 Services        0.76053238      0.33931804
                 Population      -0.0790832      1.00192402
                 Employment         0.04799      0.97509085


                        Reference Axis Correlations

                                Factor1         Factor2

                  Factor1        1.00000        -0.20188
                  Factor2       -0.20188         1.00000
```

**Output 26.2.12.** Promax Rotation: Variance Explained and Factor Structure

```
                Principal Factor Analysis with Promax Rotation

                            The FACTOR Procedure
                          Rotation Method: Promax

                Reference Structure (Semipartial Correlations)

                                  Factor1         Factor2

                 HouseValue         0.93591        -0.09590
                 School             0.89951        -0.09160
                 Services           0.74487         0.33233
                 Population        -0.07745         0.98129
                 Employment         0.04700         0.95501


        Variance Explained by Each Factor Eliminating Other Factors

                          Factor1         Factor2

                        2.2480892       2.0030200


                       Factor Structure (Correlations)

                                  Factor1         Factor2

                 HouseValue         0.93582         0.09500
                 School             0.89954         0.09189
                 Services           0.82903         0.49286
                 Population         0.12319         0.98596
                 Employment         0.24484         0.98478
```

**Output 26.2.13.** Promax Rotation: Variance Explained and Final Communalities

```
                 Principal Factor Analysis with Promax Rotation

                              The FACTOR Procedure
                            Rotation Method: Promax

            Variance Explained by Each Factor Ignoring Other Factors

                              Factor1         Factor2

                            2.4473495       2.2022803


                  Final Community Estimates: Total = 4.450370

   Population          School      Employment        Services      HouseValue

   0.97811334      0.81756387      0.97199928      0.79774304      0.88494998
```

**Output 26.2.14.** Promax Rotated Factor Pattern Plot

```
                 Principal Factor Analysis with Promax Rotation

                              The FACTOR Procedure
                            Rotation Method: Promax

Plot of Reference Structure for Factor1 and Factor2
Reference Axis Correlation = -0.2019  Angle = 101.6471

                                     Factor1
                                        1
                                     E
                                     B .9

                                       .8
                                                 D
                                       .7

                                       .6

                                       .5

                                       .4

                                       .3

                                       .2
                                                                        F
                                       .1                               a
                                                                    C   c
        -1 -.9-.8-.7-.6-.5-.4-.3-.2-.1  0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1.0t
                                                                        o
                                      -.1                            A  r
                                                                        2
                                      -.2

                                      -.3

                                      -.4

                                      -.5

                                      -.6

                                      -.7

                                      -.8

                                      -.9

                                       -1

       Population=A  School=B      Employment=C  Services=D    HouseValue=E
```

The oblique promax rotation (Output 26.2.9 through Output 26.2.14) places an axis through the variables Population and Employment but misses the HouseValue and School variables. Since an independent-cluster solution would be possible if it were not for the variable Services, a Harris-Kaiser rotation weighted by the Cureton-Mulaik technique should be used.

**Output 26.2.15.** Output Data Set

Factor Output Data Set

| Obs | _TYPE_ | _NAME_ | Population | School | Employment | Services | House Value |
|---|---|---|---|---|---|---|---|
| 1 | MEAN | | 6241.67 | 11.4417 | 2333.33 | 120.833 | 17000.00 |
| 2 | STD | | 3439.99 | 1.7865 | 1241.21 | 114.928 | 6367.53 |
| 3 | N | | 12.00 | 12.0000 | 12.00 | 12.000 | 12.00 |
| 4 | CORR | Population | 1.00 | 0.0098 | 0.97 | 0.439 | 0.02 |
| 5 | CORR | School | 0.01 | 1.0000 | 0.15 | 0.691 | 0.86 |
| 6 | CORR | Employment | 0.97 | 0.1543 | 1.00 | 0.515 | 0.12 |
| 7 | CORR | Services | 0.44 | 0.6914 | 0.51 | 1.000 | 0.78 |
| 8 | CORR | HouseValue | 0.02 | 0.8631 | 0.12 | 0.778 | 1.00 |
| 9 | COMMUNAL | | 0.98 | 0.8176 | 0.97 | 0.798 | 0.88 |
| 10 | PRIORS | | 0.97 | 0.8223 | 0.97 | 0.786 | 0.85 |
| 11 | EIGENVAL | | 2.73 | 1.7161 | 0.04 | -0.025 | -0.07 |
| 12 | UNROTATE | Factor1 | 0.63 | 0.7137 | 0.71 | 0.879 | 0.74 |
| 13 | UNROTATE | Factor2 | 0.77 | -0.5552 | 0.68 | -0.158 | -0.58 |
| 14 | RESIDUAL | Population | 0.02 | -0.0112 | 0.01 | 0.011 | 0.00 |
| 15 | RESIDUAL | School | -0.01 | 0.1824 | 0.02 | -0.024 | 0.01 |
| 16 | RESIDUAL | Employment | 0.01 | 0.0215 | 0.03 | -0.006 | -0.02 |
| 17 | RESIDUAL | Services | 0.01 | -0.0239 | -0.01 | 0.202 | 0.03 |
| 18 | RESIDUAL | HouseValue | 0.00 | 0.0125 | -0.02 | 0.034 | 0.12 |
| 19 | PRETRANS | Factor1 | 0.79 | -0.6145 | . | . | . |
| 20 | PRETRANS | Factor2 | 0.61 | 0.7889 | . | . | . |
| 21 | PREROTAT | Factor1 | 0.02 | 0.9042 | 0.15 | 0.791 | 0.94 |
| 22 | PREROTAT | Factor2 | 0.99 | 0.0006 | 0.97 | 0.415 | -0.00 |
| 23 | TRANSFOR | Factor1 | 0.74 | -0.7055 | . | . | . |
| 24 | TRANSFOR | Factor2 | 0.54 | 0.8653 | . | . | . |
| 25 | FCORR | Factor1 | 1.00 | 0.2019 | . | . | . |
| 26 | FCORR | Factor2 | 0.20 | 1.0000 | . | . | . |
| 27 | PATTERN | Factor1 | -0.08 | 0.9184 | 0.05 | 0.761 | 0.96 |
| 28 | PATTERN | Factor2 | 1.00 | -0.0935 | 0.98 | 0.339 | -0.10 |
| 29 | RCORR | Factor1 | 1.00 | -0.2019 | . | . | . |
| 30 | RCORR | Factor2 | -0.20 | 1.0000 | . | . | . |
| 31 | REFERENC | Factor1 | -0.08 | 0.8995 | 0.05 | 0.745 | 0.94 |
| 32 | REFERENC | Factor2 | 0.98 | -0.0916 | 0.96 | 0.332 | -0.10 |
| 33 | STRUCTUR | Factor1 | 0.12 | 0.8995 | 0.24 | 0.829 | 0.94 |
| 34 | STRUCTUR | Factor2 | 0.99 | 0.0919 | 0.98 | 0.493 | 0.09 |

The output data set displayed in Output 26.2.15 can be used for Harris-Kaiser rotation by deleting observations with _TYPE_='PATTERN' and _TYPE_='FCORR', which are for the promax-rotated factors, and changing _TYPE_='UNROTATE' to _TYPE_='PATTERN'.

Output 26.2.16 displays the rotated factor pattern output formatted with the picture format 'FuzzFlag'.

**Output 26.2.16.** Picture Format Output

| Obs | RowName | Factor1 | Factor2 |
|---|---|---|---|
| 1 | HouseValue | 95 * | . |
| 2 | School | 91 * | . |
| 3 | Services | 76 | 33 |
| 4 | Population | . | 100 * |
| 5 | Employment | . | 97 * |

The following statements produce Output 26.2.17:

```
data fact2(type=factor);
   set fact_all;
   if _TYPE_ in('PATTERN' 'FCORR') then delete;
   if _TYPE_='UNROTATE' then _TYPE_='PATTERN';

proc factor rotate=hk norm=weight reorder plot;
   title3 'Harris-Kaiser Rotation with Cureton-Mulaik Weights';
run;
```

The results of the Harris-Kaiser rotation are displayed in Output 26.2.17:

**Output 26.2.17.** Harris-Kaiser Rotation

```
              Harris-Kaiser Rotation with Cureton-Mulaik Weights

                            The FACTOR Procedure
                       Rotation Method: Harris-Kaiser

                        Variable Weights for Rotation

Population          School      Employment        Services      HouseValue

0.95982747      0.93945424      0.99746396      0.12194766      0.94007263


                       Oblique Transformation Matrix

                                   1               2

                        1        0.73537         0.61899
                        2       -0.68283         0.78987


                         Inter-Factor Correlations

                                Factor1         Factor2

                    Factor1     1.00000         0.08358
                    Factor2     0.08358         1.00000
```

**Harris-Kaiser Rotation with Cureton-Mulaik Weights**

**The FACTOR Procedure**
**Rotation Method: Harris-Kaiser**

**Rotated Factor Pattern (Standardized Regression Coefficients)**

| | Factor1 | Factor2 |
|---|---|---|
| HouseValue | 0.94048 | 0.00279 |
| School | 0.90391 | 0.00327 |
| Services | 0.75459 | 0.41892 |
| Population | -0.06335 | 0.99227 |
| Employment | 0.06152 | 0.97885 |

**Reference Axis Correlations**

| | Factor1 | Factor2 |
|---|---|---|
| Factor1 | 1.00000 | -0.08358 |
| Factor2 | -0.08358 | 1.00000 |

**Reference Structure (Semipartial Correlations)**

| | Factor1 | Factor2 |
|---|---|---|
| HouseValue | 0.93719 | 0.00278 |
| School | 0.90075 | 0.00326 |
| Services | 0.75195 | 0.41745 |
| Population | -0.06312 | 0.98880 |
| Employment | 0.06130 | 0.97543 |

**Variance Explained by Each Factor Eliminating Other Factors**

| Factor1 | Factor2 |
|---|---|
| 2.2628537 | 2.1034731 |

**Harris-Kaiser Rotation with Cureton-Mulaik Weights**

**The FACTOR Procedure**
**Rotation Method: Harris-Kaiser**

**Factor Structure (Correlations)**

| | Factor1 | Factor2 |
|---|---|---|
| HouseValue | 0.94071 | 0.08139 |
| School | 0.90419 | 0.07882 |
| Services | 0.78960 | 0.48198 |
| Population | 0.01958 | 0.98698 |
| Employment | 0.14332 | 0.98399 |

**Variance Explained by Each Factor Ignoring Other Factors**

| Factor1 | Factor2 |
|---|---|
| 2.3468965 | 2.1875158 |

**Final Communality Estimates: Total = 4.450370**

| Population | School | Employment | Services | HouseValue |
|---|---|---|---|---|
| 0.97811334 | 0.81756387 | 0.97199928 | 0.79774304 | 0.88494998 |

```
               Harris-Kaiser Rotation with Cureton-Mulaik Weights

                              The FACTOR Procedure
                         Rotation Method: Harris-Kaiser

Plot of Reference Structure for Factor1 and Factor2
Reference Axis Correlation = -0.0836  Angle = 94.7941

                                    Factor1
                                       1
                                       E
                                      .B

                                      .8
                                                  D
                                      .7

                                      .6

                                      .5

                                      .4

                                      .3

                                      .2
                                                                    F
                                      .1                            a
                                                                 C  c
        -1 -.9-.8-.7-.6-.5-.4-.3-.2-.1  0 .1 .2 .3 .4 .5 .6 .7 .8 .9 1.0t
                                                                 A  o
                                     -.1                            r
                                                                    2
                                     -.2

                                     -.3

                                     -.4

                                     -.5

                                     -.6

                                     -.7

                                     -.8

                                     -.9

                                      -1

       Population=A  School=B      Employment=C  Services=D    HouseValue=E
```

In the results of the Harris-Kaiser rotation, the variable Services receives a small weight, and the axes are placed as desired.

## Example 26.3. Maximum-Likelihood Factor Analysis

This example uses maximum-likelihood factor analyses for one, two, and three factors. It is already apparent from the principal factor analysis that the best number of common factors is almost certainly two. The one- and three-factor ML solutions reinforce this conclusion and illustrate some of the numerical problems that can occur. The following statements produce Output 26.3.1:

```
proc factor data=SocioEconomics method=ml heywood n=1;
   title3 'Maximum-Likelihood Factor Analysis with One Factor';
run;
proc factor data=SocioEconomics method=ml heywood n=2;
   title3 'Maximum-Likelihood Factor Analysis with Two Factors';
run;
proc factor data=SocioEconomics method=ml heywood n=3;
   title3 'Maximum-Likelihood Factor Analysis with Three Factors';
run;
```

**Output 26.3.1.** Maximum-Likelihood Factor Analysis

```
               Maximum-Likelihood Factor Analysis with One Factor

                              The FACTOR Procedure
                      Initial Factor Method: Maximum Likelihood

                         Prior Communality Estimates: SMC

   Population          School      Employment        Services      HouseValue

   0.96859160      0.82228514      0.96918082      0.78572440      0.84701921


        Preliminary Eigenvalues: Total = 76.1165859  Average = 15.2233172

                  Eigenvalue    Difference    Proportion    Cumulative

             1    63.7010086    50.6462895        0.8369        0.8369
             2    13.0547191    12.7270798        0.1715        1.0084
             3     0.3276393     0.6749199        0.0043        1.0127
             4    -0.3472805     0.2722202       -0.0046        1.0081
             5    -0.6195007                     -0.0081        1.0000

              1 factor will be retained by the NFACTOR criterion.


Iteration    Criterion    Ridge   Change              Communalities

    1        6.5429218   0.0000   0.1033   0.93828   0.72227   1.00000   0.71940
                                           0.74371
    2        3.1232699   0.0000   0.7288   0.94566   0.02380   1.00000   0.26493
                                           0.01487


        Convergence criterion satisfied.
```

```
              Maximum-Likelihood Factor Analysis with One Factor

                             The FACTOR Procedure
                    Initial Factor Method: Maximum Likelihood

                 Significance Tests Based on 12 Observations

                                                                  Pr >
                  Test                        DF    Chi-Square    ChiSq

       H0: No common factors                  10       54.2517    <.0001
       HA: At least one common factor
       H0: 1 Factor is sufficient              5       24.4656    0.0002
       HA: More factors are needed


           Chi-Square without Bartlett's Correction       34.355969
           Akaike's Information Criterion                 24.355969
           Schwarz's Bayesian Criterion                   21.931436
           Tucker and Lewis's Reliability Coefficient      0.120231


                        Squared Canonical Correlations

                                    Factor1

                                  1.0000000


Eigenvalues of the Weighted Reduced Correlation Matrix: Total = 0  Average = 0

                              Eigenvalue    Difference

                         1         Infty         Infty
                         2    1.92716032    2.15547340
                         3    -.22831308    0.56464322
                         4    -.79295630    0.11293464
                         5    -.90589094
```

```
              Maximum-Likelihood Factor Analysis with One Factor

                           The FACTOR Procedure
                  Initial Factor Method: Maximum Likelihood

                              Factor Pattern

                                         Factor1

                     Population         0.97245
                     School             0.15428
                     Employment         1.00000
                     Services           0.51472
                     HouseValue         0.12193


                   Variance Explained by Each Factor

                  Factor        Weighted    Unweighted

                  Factor1     17.8010629    2.24926004


                Final Communality Estimates and Variable Weights
        Total Communality: Weighted = 17.801063   Unweighted = 2.249260

                Variable      Communality        Weight

                Population     0.94565561    18.4011648
                School         0.02380349     1.0243839
                Employment     1.00000000         Infty
                Services       0.26493499     1.3604239
                HouseValue     0.01486595     1.0150903
```

Output 26.3.1 displays the results of the analysis with one factor. The solution on the second iteration is so close to the optimum that PROC FACTOR cannot find a better solution, hence you receive this message:

```
Convergence criterion satisfied.
```

When this message appears, you should try rerunning PROC FACTOR with different prior communality estimates to make sure that the solution is correct. In this case, other prior estimates lead to the same solution or possibly to worse local optima, as indicated by the information criteria or the Chi-square values.

The variable Employment has a communality of 1.0 and, therefore, an infinite weight that is displayed next to the final communality estimate as a missing/infinite value. The first eigenvalue is also infinite. Infinite values are ignored in computing the total of the eigenvalues and the total final communality.

**Output 26.3.2.** Maximum-Likelihood Factor Analysis: Two Factors

```
              Maximum-Likelihood Factor Analysis with Two Factors

                              The FACTOR Procedure
                     Initial Factor Method: Maximum Likelihood

                        Prior Communality Estimates: SMC

   Population          School      Employment        Services      HouseValue

   0.96859160      0.82228514      0.96918082      0.78572440      0.84701921


        Preliminary Eigenvalues: Total = 76.1165859  Average = 15.2233172

                  Eigenvalue    Difference    Proportion    Cumulative

             1    63.7010086    50.6462895        0.8369        0.8369
             2    13.0547191    12.7270798        0.1715        1.0084
             3     0.3276393     0.6749199        0.0043        1.0127
             4    -0.3472805     0.2722202       -0.0046        1.0081
             5    -0.6195007                     -0.0081        1.0000

              2 factors will be retained by the NFACTOR criterion.



Iteration    Criterion    Ridge   Change              Communalities

    1        0.3431221   0.0000   0.0471   1.00000   0.80672   0.95058   0.79348
                                           0.89412
    2        0.3072178   0.0000   0.0307   1.00000   0.80821   0.96023   0.81048
                                           0.92480
    3        0.3067860   0.0000   0.0063   1.00000   0.81149   0.95948   0.81677
                                           0.92023
    4        0.3067373   0.0000   0.0022   1.00000   0.80985   0.95963   0.81498
                                           0.92241
    5        0.3067321   0.0000   0.0007   1.00000   0.81019   0.95955   0.81569
                                           0.92187


        Convergence criterion satisfied.
```

```
              Maximum-Likelihood Factor Analysis with Two Factors

                             The FACTOR Procedure
                    Initial Factor Method: Maximum Likelihood

                   Significance Tests Based on 12 Observations

                                                                     Pr >
           Test                        DF    Chi-Square     ChiSq

    H0: No common factors              10       54.2517    <.0001
    HA: At least one common factor
    H0: 2 Factors are sufficient        1        2.1982    0.1382
    HA: More factors are needed


        Chi-Square without Bartlett's Correction       3.3740530
        Akaike's Information Criterion                 1.3740530
        Schwarz's Bayesian Criterion                   0.8891463
        Tucker and Lewis's Reliability Coefficient     0.7292200


                        Squared Canonical Correlations

                             Factor1         Factor2

                           1.0000000       0.9518891


                  Eigenvalues of the Weighted Reduced Correlation
                 Matrix: Total = 19.7853157  Average = 4.94632893

                   Eigenvalue    Difference    Proportion    Cumulative

              1         Infty         Infty
              2    19.7853143    19.2421292        1.0000        1.0000
              3     0.5431851     0.5829564        0.0275        1.0275
              4    -0.0397713     0.4636411       -0.0020        1.0254
              5    -0.5034124                     -0.0254        1.0000
```

**Maximum-Likelihood Factor Analysis with Two Factors**

**The FACTOR Procedure**
**Initial Factor Method: Maximum Likelihood**

**Factor Pattern**

| | Factor1 | Factor2 |
|---|---|---|
| Population | 1.00000 | 0.00000 |
| School | 0.00975 | 0.90003 |
| Employment | 0.97245 | 0.11797 |
| Services | 0.43887 | 0.78930 |
| HouseValue | 0.02241 | 0.95989 |

**Variance Explained by Each Factor**

| Factor | Weighted | Unweighted |
|---|---|---|
| Factor1 | 24.4329707 | 2.13886057 |
| Factor2 | 19.7853143 | 2.36835294 |

**Final Communality Estimates and Variable Weights**
**Total Communality: Weighted = 44.218285 Unweighted = 4.507214**

| Variable | Communality | Weight |
|---|---|---|
| Population | 1.00000000 | Infty |
| School | 0.81014489 | 5.2682940 |
| Employment | 0.95957142 | 24.7246669 |
| Services | 0.81560348 | 5.4256462 |
| HouseValue | 0.92189372 | 12.7996793 |

Output 26.3.2 displays the results of the analysis using two factors. The analysis converges without incident. This time, however, the Population variable is a Heywood case.

**Output 26.3.3.** Maximum-Likelihood Factor Analysis: Three Factors

```
            Maximum-Likelihood Factor Analysis with Three Factors

                             The FACTOR Procedure
                    Initial Factor Method: Maximum Likelihood

                       Prior Communality Estimates: SMC

   Population          School      Employment        Services      HouseValue

   0.96859160      0.82228514      0.96918082      0.78572440      0.84701921


        Preliminary Eigenvalues: Total = 76.1165859  Average = 15.2233172

                  Eigenvalue    Difference    Proportion    Cumulative

             1    63.7010086    50.6462895        0.8369        0.8369
             2    13.0547191    12.7270798        0.1715        1.0084
             3     0.3276393     0.6749199        0.0043        1.0127
             4    -0.3472805     0.2722202       -0.0046        1.0081
             5    -0.6195007                    -0.0081        1.0000

              3 factors will be retained by the NFACTOR criterion.


                WARNING: Too many factors for a unique solution.


Iteration    Criterion    Ridge    Change              Communalities

    1        0.1798029   0.0313   0.0501   0.96081   0.84184   1.00000   0.80175
                                           0.89716
    2        0.0016405   0.0313   0.0678   0.98081   0.88713   1.00000   0.79559
                                           0.96500
    3        0.0000041   0.0313   0.0094   0.98195   0.88603   1.00000   0.80498
                                           0.96751
    4        0.0000000   0.0313   0.0006   0.98202   0.88585   1.00000   0.80561
                                           0.96735


        ERROR: Converged, but not to a proper optimum.
                      Try a different 'PRIORS' statement.
```

```
             Maximum-Likelihood Factor Analysis with Three Factors

                             The FACTOR Procedure
                   Initial Factor Method: Maximum Likelihood

                  Significance Tests Based on 12 Observations

                                                                 Pr >
             Test                         DF    Chi-Square     ChiSq

    H0: No common factors                 10       54.2517    <.0001
    HA: At least one common factor
    H0: 3 Factors are sufficient          -2        0.0000     .
    HA: More factors are needed


        Chi-Square without Bartlett's Correction       0.0000003
        Akaike's Information Criterion                 4.0000003
        Schwarz's Bayesian Criterion                   4.9698136
        Tucker and Lewis's Reliability Coefficient     0.0000000


                       Squared Canonical Correlations

                  Factor1         Factor2         Factor3

                1.0000000       0.9751895       0.6894465


                 Eigenvalues of the Weighted Reduced Correlation
                Matrix: Total = 41.5254193  Average = 10.3813548

                  Eigenvalue    Difference    Proportion    Cumulative

             1         Infty         Infty
             2    39.3054826    37.0854258        0.9465        0.9465
             3     2.2200568     2.2199693        0.0535        1.0000
             4     0.0000875     0.0002949        0.0000        1.0000
             5    -0.0002075                     -0.0000        1.0000
```

```
         Maximum-Likelihood Factor Analysis with Three Factors

                        The FACTOR Procedure
               Initial Factor Method: Maximum Likelihood

                           Factor Pattern

                        Factor1         Factor2         Factor3

   Population           0.97245        -0.11233        -0.15409
   School               0.15428         0.89108         0.26083
   Employment           1.00000         0.00000         0.00000
   Services             0.51472         0.72416        -0.12766
   HouseValue           0.12193         0.97227        -0.08473


                  Variance Explained by Each Factor

                 Factor        Weighted    Unweighted

                 Factor1     54.6115241    2.24926004
                 Factor2     39.3054826    2.27634375
                 Factor3      2.2200568    0.11525433


           Final Communality Estimates and Variable Weights
 Total Communality: Weighted = 96.137063   Unweighted = 4.640858

               Variable      Communality        Weight

               Population     0.98201660    55.6066901
               School         0.88585165     8.7607194
               Employment     1.00000000         Infty
               Services       0.80564301     5.1444261
               HouseValue     0.96734687    30.6251078
```

The three-factor analysis displayed in Output 26.3.3 generates this message:

**`WARNING: Too many factors for a unique solution.`**

The number of parameters in the model exceeds the number of elements in the correlation matrix from which they can be estimated, so an infinite number of different perfect solutions can be obtained. The Criterion approaches zero at an improper optimum, as indicated by this message:

**`Converged, but not to a proper optimum.`**

The degrees of freedom for the chi-square test are $-2$, so a probability level cannot be computed for three factors. Note also that the variable Employment is a Heywood case again.

The probability levels for the chi-square test are 0.0001 for the hypothesis of no common factors, 0.0002 for one common factor, and 0.1382 for two common factors. Therefore, the two-factor model seems to be an adequate representation. Akaike's information criterion and Schwarz's Bayesian criterion attain their minimum values at two common factors, so there is little doubt that two factors are appropriate for these data.

# References

Akaike, H. (1973), "Information Theory and the Extension of the Maximum Likelihood Principle," in *Second International Symposium on Information Theory,* eds. V.N. Petrov and F. Csaki, Budapest: Akailseoniai-Kiudo, 267–281.

Akaike, H. (1974), "A New Look at the Statistical Identification Model," *IEEE Transactions on Automatic Control*, 19, 716–723.

Akaike, H. (1987), "Factor Analysis and AIC," *Psychometrika* 52, 317–332.

Bickel, P.J. and Doksum, K.A. (1977), *Mathematical Statistics,* San Francisco: Holden-Day.

Cattell, R.B. (1966), "The Scree Test for the Number of Factors," *Multivariate Behavioral Research,* 1, 245–276.

Cattell, R.B. (1978), *The Scientific Use of Factor Analysis,* New York: Plenum.

Cattell, R.B. and Vogelman, S. (1977), "A Comprehensive Trial of the Scree and KG Criteria for Determining the Number of Factors," *Multivariate Behavioral Research*, 12, 289–325.

Cerny, B.A. and Kaiser, H.F. (1977), "A Study of a Measure of Sampling Adequacy for Factor-Analytic Correlation Matrices," *Multivariate Behavioral Research,* 12, 43–47.

Cureton, E.E. (1968), *A Factor Analysis of Project TALENT Tests and Four Other Test Batteries,* (Interim Report 4 to the U.S. Office of Education, Cooperative Research Project No. 3051.) Palo Alto: Project TALENT Office, American Institutes for Research and University of Pittsburgh.

Cureton, E.E. and Mulaik, S.A. (1975), "The Weighted Varimax Rotation and the Promax Rotation," *Psychometrika,* 40, 183–195.

Dziuban, C.D. and Harris, C.W. (1973), "On the Extraction of Components and the Applicability of the Factor Model," *American Educational Research Journal,* 10, 93–99.

Fuller (1987), *Measurement Error Models*, New York: John Wiley & Sons, Inc.

Geweke, J.F. and Singleton, K.J. (1980), "Interpreting the Likelihood Ratio Statistic in Factor Models When Sample Size Is Small," *Journal of the American Statistical Association,* 75, 133–137.

Gorsuch, R.L. (1974), *Factor Analysis,* Philadelphia: W.B. Saunders Co.

Harman, H.H. (1976), *Modern Factor Analysis,* Third Edition, Chicago: University of Chicago Press.

Harris, C.W. (1962), "Some Rao-Guttman Relationships," *Psychometrika,* 27, 247–263.

Horn, J.L. and Engstrom, R. (1979), "Cattell's Scree Test in Relation to Bartlett's Chi-Square Test and Other Observations on the Number of Factors Problem," *Multivariate Behavioral Research,* 14, 283–300.

Joreskog, K.G. (1962), "On the Statistical Treatment of Residuals in Factor Analysis," *Psychometrika*, 27, 335–354.

Joreskog, K.G. (1977), "Factor Analysis by Least-Squares and Maximum Likelihood Methods," in *Statistical Methods for Digital Computers*, eds. K. Enslein, A. Ralston, and H.S. Wilf, New York: John Wiley & Sons, Inc.

Kaiser, H.F. (1963), "Image Analysis," in *Problems in Measuring Change*, ed. C.W. Harris, Madison, WI: University of Wisconsin Press.

Kaiser, H.F. (1970), "A Second Generation Little Jiffy," *Psychometrika*, 35, 401–415.

Kaiser, H.F. and Cerny, B.A. (1979), "Factor Analysis of the Image Correlation Matrix," *Educational and Psychological Measurement*, 39, 711–714.

Kaiser, H.F. and Rice, J. (1974), "Little Jiffy, Mark IV," *Educational and Psychological Measurement*, 34, 111–117.

Kerlinger, F.N. and Pedhazur, E.J. (1973), *Multiple Regression in Behavioral Research*, New York: Holt, Rinehart & Winston, Inc.

Kim, J.O. and Mueller, C.W. (1978a), *Introduction to Factor Analysis: What It Is and How To Do It*, Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-013, Beverly Hills: Sage Publications.

Kim, J.O. and Mueller, C.W. (1978b), *Factor Analysis: Statistical Methods and Practical Issues*, Sage University Paper Series on Quantitative Applications in the Social Sciences, series no. 07-014, Beverly Hills: Sage Publications.

Lawley, D.N. and Maxwell, A.E. (1971), *Factor Analysis as a Statistical Method*, New York: Macmillan Publishing Co., Inc.

Lee, H.B. and Comrey, A.L. (1979), "Distortions in a Commonly Used Factor Analytic Procedure," *Multivariate Behavioral Research*, 14, 301–321.

Mardia, K.V., Kent, J.T., and Bibby, J.M. (1979), *Multivariate Analysis*, London: Academic Press.

McDonald, R.P. (1975), "A Note on Rippe's Test of Significance in Common Factor Analysis," *Psychometrika*, 40, 117–119.

McDonald, R.P. (1985), *Factor Analysis and Related Methods*, New Jersey: Lawrence Erlbaum Associates, Publishers.

Morrison, D.F. (1976), *Multivariate Statistical Methods*, Second Edition, New York: McGraw-Hill Book Co.

Mulaik, S.A. (1972), *The Foundations of Factor Analysis*, New York: McGraw-Hill Book Co.

Rao, C.R. (1955), "Estimation and Tests of Significance in Factor Analysis," *Psychometrika*, 20, 93–111.

Schwarz, G. (1978), "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–464.

Spearman, C. (1904), "General Intelligence Objectively Determined and Measured," *American Journal of Psychology*, 15, 201–293.

Stewart, D.W. (1981), “The Application and Misapplication of Factor Analysis in Marketing Research,” *Journal of Marketing Research*, 18, 51–62.

Tucker, L.R. and Lewis, C. (1973), “A Reliability Coefficient for Maximum Likelihood Factor Analysis,” *Psychometrika*, 38, 1–10.

The correct bibliographic citation for this manual is as follows: SAS Institute Inc., *SAS/STAT® User's Guide, Version 8*, Cary, NC: SAS Institute Inc., 1999.

**SAS/STAT® User's Guide, Version 8**



ISBN 1–58025–494–2



SAS Institute Inc., SAS Campus Drive, Cary, North Carolina 27513.

1st printing, October 1999



The Institute is a private company devoted to the support and further development of its software and related services.