Factor Analysis with SAS

In this document we will outline the SAS procedures for performing the most common types of factor analysis using the SAS procedure PROC FACTOR. By doing so, we will focus on calling the correct SAS options to ensure that we are fitting the correct factor model. While this may seem to be an obvious objective, this simple competency is not transparent in the SAS software, and hence leads to many errors by SAS users.

The General PROC FACTOR Statement

Here is the general syntax that we will consider for the PROC FACTOR procedure.

```
PROC FACTOR
        DATA = value            /* input data set */
        METHOD = value          /* method of estimation - {PRINCIPAL, PRINIT, ML, ULS} */
        PRIORS = value          /* set the prior communality estimates - {MAX, SMC, ONE} */
        NFACTORS = value        /* number of common factors to fit */
        MINEIGEN = 0 (default)      /* minimum eigenvalue criteria for factor selection */
        PROPORTION = value          /* value in (0,1), selects the number of common factors
                                       needed to explain this proportion of the common variance
                                       */
        ROTATE = NONE (default)     /* perform a rotation? – {NONE, VARIMAX, PROMAX} */
        OUT = value             /* output data set with original data and the factor scores */
        OUTSTAT = value         /* output data set containing all associated statistical output */
        PLOTS=ALL               /* produce all plots associated with factor analysis */
        NPLOTS = value          /* number of factors to plot – default is to plot all factors */
        RESIDUALS               /* call by name option for producing the residual correlation
                                   matrix, diagonal elements are the unique variances */
        REORDER                 /* call by name option for reordering variables in the output data
                                   set so that variables are grouped together based on their factor
                                   relationships */
        HEYWOOD                 /* allows the Heywood estimating procedure which sets
                                   communality estimates greater than 1 to 1 */
;
```

Some arguments have default values, as indicated in the example procedure call, and the set of restricted arguments that we will consider is provide in the curly braces in each comment.

Notes:

(1) In almost all uses of PROC FACTOR you will want to explicitly set the value for the PRIORS argument. In some cases SAS will set the default value to ONE, and this default value will not be using the Common Factor Model. Instead it will default to a Principal Components Analysis. In most cases there are only two logical choices for the value of PRIORS: (1) SMC – Squared Multiple Correlation, and (2) MAX – the maximum absolute correlation. If the correlation matrix is deemed to be singular, then SMC will not be a valid option and you will have to try the MAX option.

(2) Use only one of the options NFACTORS, MINEIGEN, or PROPORTION to select the number of common factors to fit. If you specify more than one of these options, then SAS will fit the smallest number of common factors that will satisfy any of the specified options. If you do not specify any option, then SAS will select the number of common factors for you by using the PROPORTION criterion as the default.

(3) The option for the HEYWOOD cases is provided by SAS, but we will not use this option in PREDICT 410.

Principal Factor Analysis

Here is the basic syntax needed to run a Principal Factor Analysis. The arguments corresponding to general options have been suppressed so that we can focus on the arguments that directly determine the Principal Factor Analysis output.

```
PROC FACTOR
        DATA = value
        METHOD = PRINCIPAL
        PRIORS = SMC                        /* {MAX, SMC} */
        ROTATE = NONE (default value)       /* {VARIMAX, PROMAX} */
;
```

Notes:

(1) Note that the SAS default value for PRIORS is ONE, and that this default setting will cause PROC FACTOR to perform a Principal Components Analysis, not a Principal Factor Analysis. In order for SAS to perform a valid Principal Factor Analysis using the Common Factor Model, the user must specify a value of SMC or MAX for the PRIORS argument.

Iterative Principal Factor Analysis

Here is the basic syntax needed to run an Iterative Principal Factor Analysis. The arguments corresponding to general options have been suppressed so that we can focus on the arguments that directly determine the Iterative Principal Factor Analysis output.

```
PROC FACTOR
        DATA = value
        METHOD = PRINIT
        PRIORS = SMC                        /* {MAX, SMC} */
        ROTATE = NONE (default value)       /* {VARIMAX, PROMAX} */
;
```

Notes:

(1) Note that PRINIT, ML, and ULS are all iterative methods, and hence if an estimated communality exceeds 1, then SAS will stop iterating. In this case we do not have a valid factor solution. SAS will allow these options to continue to iterate by specifying the HEYWOOD or ULTRAHEYWOOD options, but we will not consider these to be valid factor solutions either.

Maximum Likelihood Factor Analysis

Here is the basic syntax needed to run a Maximum Likelihood Factor Analysis. The arguments corresponding to general options have been suppressed so that we can focus on the arguments that directly determine the Maximum Likelihood Factor Analysis output.

```
PROC FACTOR
        DATA = value
        METHOD = ML
        PRIORS = SMC (default value)        /* {MAX, SMC} */
        ROTATE = NONE (default value)       /* {VARIMAX, PROMAX} */
;
```

Notes:

(1) Note that PRINIT, ML, and ULS are all iterative methods, and hence if an estimated communality exceeds 1, then SAS will stop iterating. In this case we do not have a valid factor solution. SAS will allow these options to continue to iterate by specifying the HEYWOOD or ULTRAHEYWOOD options, but we will not consider these to be valid factor solutions either.

(2) Maximum Likelihood Factor Analysis is a formal statistical model with formal statistical assumptions that allow statistical inference; hence the output from a ML FA will contain additional information that is not available for the other estimation procedures.

Unweighted Least Squares Factor Analysis

Here is the basic syntax needed to run an Unweighted Least Squares Factor Analysis. The arguments corresponding to general options have been suppressed so that we can focus on the arguments that directly determine the Unweighted Least Squares Factor Analysis output.

```
PROC FACTOR
        DATA = value
        METHOD = ULS
        PRIORS = SMC (default value)        /* {MAX, SMC} */
        ROTATE = NONE (default value)       /* {VARIMAX, PROMAX} */
;
```

Notes:

(1) Note that PRINIT, ML, and ULS are all iterative methods, and hence if an estimated communality exceeds 1, then SAS will stop iterating. In this case we do not have a valid factor solution. SAS will allow these options to continue to iterate by specifying the HEYWOOD or ULTRAHEYWOOD options, but we will not consider these to be valid factor solutions either.