A Brief Introduction to PROC FACTOR

The FACTOR procedure in SAS performs common factor and component analyses with rotations. PROC FACTOR also has the flexibility to process output from other SAS procedures such as PROC CORR. To invoke the procedure, start with the PROC FACTOR statement and end with run;

Reading in Data
Data can be read into SAS in the (not exhaustive) following ways:
(1) Raw data
(2) Correlation matrix
(3) Covariance matrix

Raw Data:

Importing data from an excel spreadsheet using PROC IMPORT:

```
proc import out=raw /*name of temporary data file*/
                  datafile = "...\mardia.xls" /*location and name of file*/
                  dbms=excel /*type of file to be imported*/
                  replace; /*overwrite any existing data file called "raw"*/
      sheet="Sheet1"; /*specifies which sheet to import data from*/
      getnames=yes; /*grab variable names in the first row*/
run;
proc print data=raw; run; /*check*/
```

Importing data tab delimited data using the infile statement:

```
data raw; /*name of temporary data file*/
  infile '...\mardia.dat' /*location and name of file*/
  dlm='09'x /*specify that the delimiter is a tab*/
  firstobs=2; /*start reading data from the second line*/
  input mechanics vectors algebra analysis statistics; /*variable names*/
run;
```

Importing data from SPSS portable file using PROC CONVERT:

```
filename raw "...\mardia.por"; /*location and name of SPSS portable file*/
proc convert spss=raw /*name of spss data file defined with filename*/
             out=raw; /*name of temporary data file*/
run;
```

Note, to use PROC CONVERT, the SPSS file has to be saved into a portable format. This can be done in SPSS by going to File → Save As → Save as type → SPSS portable (*.por).

Correlation matrix:
A DATA step may be used to create temporary datasets when reading in full symmetric correlation matrices or triangular correlation matrices. The data file has to be set up such that there are no variable names within.
A print out of the full symmetric correlation matrix and a lower triangle correlation matrix are provided below:

Full symmetric correlation matrix:

```
1.00000 0.55341 0.54675 0.40939 0.38910
0.55341 1.00000 0.60964 0.48508 0.43645
0.54675 0.60964 1.00000 0.71081 0.66474
0.40939 0.48508 0.71081 1.00000 0.60717
0.38910 0.43645 0.66474 0.60717 1.00000
```

Lower triangle correlation matrix:

```
1.00000
0.55341 1.00000
0.54675 0.60964 1.00000
0.40939 0.48508 0.71081 1.00000
0.38910 0.43645 0.66474 0.60717 1.00000
```

```
data corr /*name of temporary data file*/
    (type=corr); /*type of data to be read in*/
  _type_ = 'corr'; /*this tells sas that the elements are from a correlation matrix*/
  infile "...mardia_corr.txt" /*location and name of file*/
  missover; /*tell sas that only a lower triangular matrix is provided*/
  input mechanics vectors algebra analysis statistics; /*variable names*/
run;
proc print data=corr; run;
```

Note that if a full symmetric correlation matrix is being read in, the `missover` option needs to be omitted. Additionally, although there are two statements with `type`, both are essential. The first is specifying that the data is arranged as a correlation matrix. The second `_type_` specifically defines the type of data as a correlation. This is reflected in the printed output below:

Obs	_type_	mechanics	vectors	algebra	analysis	statistics
1	corr	1.00000	.	.	.	.
2	corr	0.55341	1.00000	.	.	.
3	corr	0.54675	0.60964	1.00000	.	.
4	corr	0.40939	0.48508	0.71081	1.00000	.
5	corr	0.38910	0.43645	0.66474	0.60717	1

Covariance matrix:

```
data cov /*name of temporary data file*/
  (type=cov); /*type of data to be read in*/
  _type_ = 'cov'; /*this tells sas that the elements are from a covariance matrix*/
  infile "...\mardia_cov.txt" /*location and name of file*/
  dlm='09'x; /*specify data is tab delimited*/
  input mechanics vectors algebra analysis statistics; /*variable names*/
run;
```

Note that if a triangular covariance matrix is input, the `missover` option needs to be stated on the `infile` statement. The only difference between this code and the one for inputting correlations is the `type=` statement.

Doing a Factor Analysis

```
proc factor data=cov /*name of data file*/
 (type=cov)/*type of data*/
  nobs=88 /*number of observations*/
  corr /*print correlation matrix*/
  priors=smc /*types of priors to be used*/
  method=u /*method of extraction*/
  nfactors=5 /*number of factors to retain*/
  maxiter = 50 /*maximum number of iterations*/
  rotate=varimax /*type of rotation*/
  scree /*print of scree plot*/
  res /*display residual correlation matrix*/
 heywood;/*sets to 1 any communality greater than 1, allowing iterations to proceed*/
 var mechanics--statistics; /*variables to be included*/
run;
```

Note that these options may be omitted. However, SAS will run on defaults and it may need a little sleuthing to find out what they are.

The code above has SAS conduct an exploratory factor analysis. Here is a quick and dirty run down and explanation of the available options with this procedure:

(a) `data`
Specifies the dataset to be used; it may be a raw dataset, a correlation or covariance matrix.

(b) `(type = name )`
If the dataset is not in raw format, this option is required. `cov` is for covariance matrix and `corr` is for correlation matrix.

(c) `nobs = n`
Specifies the number of observations used to generate the data. This is required when using correlations or covariances since this information is not obtainable from the data. By default, if raw data is read in, SAS will compute this value.

(d) `corr`
prints out the correlation matrix in the output

(e) `priors = name`
Specifies what type of priors used. `max` = sets prior communality estimate for each variable to its maximum absolute correlation with any other variable. `one` = sets all prior communalities to 1.0. `smc` = sets the prior communality estimate for each variable to its squared multiple correlation with all other variables.

(f) `method = name`
Specifies the method of extraction used. `p` = principle components. `u` = unweighted least squares. `ml` = maximum likelihood. Note that if both the priors and method options are used but are inconsistent – for example, `priors = smc` and `method = p`, SAS will default and use principle factors.

(g) `nfactors = n`
Specifies the number of factors to be extracted. Note that omitting this option, SAS will extract the number of factors with eigen values greater than 1.0. In addition, SAS will only extract a maximum of factors based on eigen values greater than 1.0. Hence, if `nfactors = 5` but 3 eigen values are negative, only 2 factors will be extracted.

(h) `maxiter = n`
Specifies the maximum number of iterations during factor extraction. The default number is 30.

(i) `rotate = name`
Specifies the type of rotation used. `none` = no rotation is performed. Other specifiable rotations are `varimax`, `promax`, and `procrustes`.

(j) `scree`
Specifies that the scree plot be output.

(k) `res`
Displays the residual correlation matrix.

(l) `var`
Specifies the variables to be used in the factor analysis. This statement may be used to choose specific variables from the dataset. If this statement is omitted, SAS will use all the variables present in the dataset by default.

Annotated Output

```
The FACTOR Procedure
Initial Factor Method: Unweighted Least Squares

                   Prior Communality Estimates: SMC
 mechanics         vectors         algebra        analysis      statistics
0.37641463      0.44512334      0.67135511      0.54086250      0.47932294


Preliminary Eigenvalues: Total = 2.51307851  Average = 0.5026157
        Eigenvalue    Difference    Proportion    Cumulative
   1    2.70286694    2.51846085        1.0755        1.0755
   2    0.18440609    0.27912086        0.0734        1.1489
   3    -.09471476    0.03565853       -0.0377        1.1112
   4    -.13037330    0.01873316       -0.0519        1.0593
   5    -.14910646                     -0.0593        1.0000

   2 factors will be retained by the MINEIGEN criterion.

              Factor Pattern
                   Factor1         Factor2
mechanics          0.64273         0.34232
vectors            0.70821         0.28694
algebra            0.89660        -0.08719
analysis           0.77102        -0.23504
statistics         0.71798        -0.22820
```

Note that although 5 factors were specified, SAS only extracts 2.

These are the unrotated factor loadings

```
The FACTOR Procedure
Rotation Method: Varimax

    Orthogonal Transformation Matrix
                      1              2
       1        0.76553        0.64340
       2       -0.64340        0.76553

          Rotated Factor Pattern

                   Factor1         Factor2
mechanics          0.27179         0.67559
vectors            0.35754         0.67532
algebra            0.74247         0.51013
analysis           0.74147         0.31615
statistics         0.69646         0.28726
```

This is the transformation matrix applied to the unrotated factor pattern matrix above to obtain the rotated factor pattern matrix below.