

Factor Analysis in SAS Using a Correlation Matrix as Input

In SAS, factor (or component) analysis can be estimated by inputting a correlation (or covariance) matrix rather than the raw data as input. This can be useful when rerunning an analysis from data supplied in an article or a report.

To create a correlation matrix as a data file in SAS, use syntax such as that shown below.

```
data corr (type = corr);  
infile cards missover;  
_type_ = 'CORR';  
input _name_ $ x1 x2 x3 x4 x5 x6;  
datalines;  
x1 1.000  
x2 0.70 1.00  
x3 0.65 0.70 1.00  
x4 0.20 0.05 0.25 1.00  
x5 0.15 0.10 0.15 0.65 1.00  
x6 0.10 0.20 0.20 0.75 0.60 1.00  
;
```

In the syntax above, the **data** command names the correlation matrix as “corr.” The specification **(type = corr)** indicates that the data file is a correlation matrix.

In the next line, **infile cards** indicates that “cards” or rows of data will be entered inline. The **missover** option must be used when entering a correlation matrix. This is because SAS will expect as many variables in each row as there are variables, but this will not be the case with correlation input.

The line **_type_ = 'CORR'** specifies that the data entered will be correlations.

The **input** line specifies that the names of the variables (**_name_**) will be entered first, followed by the correlations. This line also provides the variable names and the order of the variables.

The **datalines** statement indicates that the data will begin on the next line. The correlation matrix should be entered as shown, with the variable name at the beginning of each row.

Running these commands will result in the working dataset “corr” shown below:

	type	_name_	x1	x2	x3	x4	x5	x6
1	CORR	x1	1
2	CORR	x2	0.7	1
3	CORR	x3	0.65	0.7	1	.	.	.
4	CORR	x4	0.2	0.05	0.25	1	.	.
5	CORR	x5	0.15	0.1	0.15	0.65	1	.
6	CORR	x6	0.1	0.2	0.2	0.75	0.6	1

The data can now be used in SAS analyses such as **proc factor** that accept matrix data using the usual commands:

```
proc factor data=corr corr residuals method=prinit priors=smc nfactors=2 msa  
rotate=oblimin plot=scree;  
var x1-x6;  
run;
```


These commands and the associated output are discussed in the document “SAS Syntax for EFA” so I will not repeat them here.

However, I comment on one aspect of this analysis. When entering a correlation matrix as shown above, there is not a way to include sample sizes. This is not typically a problem because the sample size is relevant only for calculating standard errors and tests of significance, and these are not usually of interest in EFA.

By default, SAS assumes a sample size of 10,000 when a correlation matrix is entered using the commands shown previously. This can be seen in the output below, which was obtained from a factor analysis of the correlation matrix in “corr.”


Input Data Type	Correlations
N Set/Assumed in Data Set	10000
N for Significance Tests	10000

It is possible to add sample sizes to the working data set “corr” after it has been created, if desired. To do this, open “corr” through the File Explorer in SAS and double-click on the *edit* icon:




	type	_name_	x1	x2	x3	x4	x5	x6
1	CORR	x1	1
2	CORR	x2	0.7	1
3	CORR	x3	0.65	0.7	1	.	.	.
4	CORR	x4	0.2	0.05	0.25	1	.	.
5	CORR	x5	0.15	0.1	0.15	0.65	1	.
6	CORR	x6	0.1	0.2	0.2	0.75	0.6	1

Double-clicking the *edit* icon should open the *add row* icon.



	type	_name_	x1	x2	x3	x4	x5	x6
1	CORR	x1	1
2	CORR	x2	0.7	1
3	CORR	x3	0.65	0.7	1	.	.	.
4	CORR	x4	0.2	0.05	0.25	1	.	.
5	CORR	x5	0.15	0.1	0.15	0.65	1	.
6	CORR	x6	0.1	0.2	0.2	0.75	0.6	1


Double-clicking on the *add row* icon will create a new blank row. Enter the “_type_” as “N” and enter the sample sizes under each variable name. Then click on the *save* icon.



VIEWTABLE: Work.Corr

	type	_name_	x1	x2	x3	x4	x5	x6
NEW	N		100	100	100	100	100	100

The dataset “corr” should now contain the sample sizes as shown below:



VIEWTABLE: Work.Corr

	type	_name_	x1	x2	x3	x4	x5	x6
1	CORR	x1	1
2	CORR	x2	0.7	1
3	CORR	x3	0.65	0.7	1	.	.	.
4	CORR	x4	0.2	0.05	0.25	1	.	.
5	CORR	x5	0.15	0.1	0.15	0.65	1	.
6	CORR	x6	0.1	0.2	0.2	0.75	0.6	1
7	N		100	100	100	100	100	100

Running **proc factor** will now result in the following data information:

Input Data Type	Correlations
N Set/Assumed in Data Set	100
N for Significance Tests	100