Factor Analysis Example:
SAS program (in blue) and output (in black)
interleaved with comments (in red)

The following DATA procedure is to read input data. This will create a SAS dataset named CORRMATR whose type is the correlation among variables M, P, C, E, H, and F. Here, you actually type the input data in the program.

```
Data corrmatr (type=corr);
      input M P C E H F;
      Type = 'CORR';
Cards;
0.37 0.62 0.54 0.32 0.284 0.37
0.62 1.00 0.51 0.38 0.351 0.43
0.54 0.51 1.00 0.36 0.336 0.405
0.32 0.38 0.36 1.00 0.686 0.73
0.284 0.351 0.336 0.686 1.00 0.7345
0.37 0.43 0.405 0.73 0.7345 1
run;
```

If you are reading the input data from a file as in your assignment, the program will be:

```
data factor;
      infile 'c:\folders\data\aus85sam.txt';
      input section price lotsize bed bath other stories fireplac cars pool
      fence age;
run;
```

Here, you are reading the input data from a file named AUS85SAM.TXT located in the directory C:\FOLDERS\DATA (or whatever location your data file is). This file contains raw data of the above variables.

The next line of command is to let SAS print the data set CORRMATR just read.

```
proc print data=corrmatr;
run;
```

The SAS output from this is:

Obs	M	P	C	E	H	F	_Type_
1	1.000	0.620	0.540	0.320	0.2840	0.3700	CORR
2	0.620	1.000	0.510	0.380	0.3510	0.4300	CORR
3	0.540	0.510	1.000	0.360	0.3360	0.4050	CORR
4	0.320	0.380	0.360	1.000	0.6860	0.7300	CORR
5	0.284	0.351	0.336	0.686	1.0000	0.7345	CORR
6	0.370	0.430	0.405	0.730	0.7345	1.0000	CORR

The next is for conducting a factor analysis of the read input data with the FACTOR procedure.

```
proc factor method=prinit rotate=v corr msa scree residuals preplot plot;
var M P C E H F;
run;
```

After PROC FACTOR, you are giving options to the FACTOR procedure.

The METHOD= option 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. PRINIT used here yields iterated principal factor analysis.

The ROTATE= option 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. V (or VARIMAX) used here specifies orthogonal varimax rotation. In the VARIMAX rotation the major objective is to have a factor structure in which each variable loads highly on one and only one factor. That is, a given variable should have a high loading on one factor and near zero loading on other factors. Such a factor structure will result in each factor representing a distinct construct.

The CORR option displays the correlation matrix or partial correlation matrix. The MSA option 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. The SCREE option displays a scree plot of the eigenvalues. The RESIDUALS option displays the residual correlation matrix and the associated partial correlation matrix. The diagonal elements of the residual correlation matrix are the unique variances. The PREPLOT option plots the factor pattern before rotation. The PLOT option plots the factor pattern after rotation.

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.

The following is the SAS output.

The FACTOR Procedure

Correlations

	M	P	C
M	**1.00000**	**0.62000**	**0.54000**
P	**0.62000**	**1.00000**	**0.51000**
C	**0.54000**	**0.51000**	**1.00000**
E	**0.32000**	**0.38000**	**0.36000**
H	**0.28400**	**0.35100**	**0.33600**
F	**0.37000**	**0.43000**	**0.40500**

Correlations

	E	H	F
M	**0.32000**	**0.28400**	**0.37000**
P	**0.38000**	**0.35100**	**0.43000**
C	**0.36000**	**0.33600**	**0.40500**
E	**1.00000**	**0.68600**	**0.73000**
H	**0.68600**	**1.00000**	**0.73450**
F	**0.73000**	**0.73450**	**1.00000**

The FACTOR Procedure
Initial Factor Method: Iterated Principal Factor Analysis

Partial Correlations Controlling
all other Variables

	M	P	C
M	**1.00000**	**0.44624**	**0.30877**
P	**0.44624**	**1.00000**	**0.20253**
C	**0.30877**	**0.20253**	**1.00000**
E	**0.01369**	**0.05109**	**0.04784**
H	**-0.03195**	**0.02594**	**0.03159**
F	**0.06094**	**0.09912**	**0.08637**

Partial Correlations Controlling all other Variables

	E	H	F
M	**0.01369**	**-0.03195**	**0.06094**
P	**0.05109**	**0.02594**	**0.09912**
C	**0.04784**	**0.03159**	**0.08637**
E	**1.00000**	**0.31767**	**0.41630**
H	**0.31767**	**1.00000**	**0.45049**
F	**0.41630**	**0.45049**	**1.00000**

The correlation and partial correlation matrices are the result from the CORR and MSA options. An examination of these matrices can be a rule of thumb to decide whether or not the data are appropriate for factor analysis.

Next measure, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy is a popular diagnostic measure. It is suggested that KMO measure of below .50 is unacceptable and the overall KMO measure should be greater than .80; however, a measure of above .60 is tolerable. The overall KMO measure can sometimes be increased by deleting the offending variables whose KMO value is low.

Kaiser's Measure of Sampling Adequacy: Overall MSA = 0.81299762

M	P	C	E	H	F
0.76887320	0.81243922	0.86691637	0.83166568	0.81232642	0.79685567

Prior Communality Estimates: ONE

Preliminary Eigenvalues: Total = 6 Average = 1

	Eigenvalue	Difference	Proportion	Cumulative
1	3.36689345	2.17285287	0.5611	0.5611
2	1.19404058	0.68703455	0.1990	0.7602

	Eigenvalue	Difference	Proportion	Cumulative
3	0.50700603	0.13515927	0.0845	0.8447
4	0.37184676	0.05872822	0.0620	0.9066
5	0.31311854	0.06602390	0.0522	0.9588
6	0.24709464		0.0412	1.0000

2 factors will be retained by the MINEIGEN criterion.

The eigenvalue-greater-than-one rule (MINEIGEN criterion), together with the scree plot, is one of the most popular heuristics to determine the number of factors needed to explain correlations among the variables.

```
Scree Plot of Eigenvalues
    ,
    ,
    ,
    ,
    ,
  4 ˆ
    ,
    ,
    ,
    ,               1
    ,
  3 ˆ
E   ,
i   ,
g   ,
e   ,
n   ,
v 2 ˆ
a   ,
l   ,
u   ,
e   ,
s   ,                        2
  1 ˆ
    ,
    ,
    ,                                 3
    ,                                          4        5
    ,                                                            6
  0 ˆ
    ,
    ,
    ,
    ,
    ,
    Šffffffˆffffffffˆffffffffˆffffffffˆffffffffˆffffffffˆffffffffˆffffff
           0        1        2        3        4        5        6

                                   Number
```

The FACTOR Procedure
Initial Factor Method: Iterated Principal Factor Analysis

Iteration	Change	Communalities					
1	0.3594	0.76582	0.71564	0.64061	0.79685	0.81139	0.83061
2	0.1277	0.69839	0.62622	0.51291	0.72453	0.74431	0.78351
3	0.0422	0.67947	0.59762	0.47073	0.69818	0.71876	0.77359
4	0.0135	0.67488	0.58806	0.45722	0.68812	0.70800	0.77395
5	0.0052	0.67444	0.58455	0.45287	0.68398	0.70285	0.77646
6	0.0028	0.67510	0.58304	0.45140	0.68212	0.70004	0.77888
7	0.0019	0.67594	0.58224	0.45084	0.68120	0.69834	0.78075
8	0.0013	0.67671	0.58173	0.45059	0.68071	0.69725	0.78209
9	0.0009	0.67735	0.58136	0.45045	0.68043	0.69652	0.78302

Convergence criterion satisfied.

Eigenvalues of the Reduced Correlation Matrix:
Total = 3.86906994 Average = 0.64484499

	Eigenvalue	Difference	Proportion	Cumulative
1	3.02809318	2.18706637	0.7826	0.7826
2	0.84102682	0.83946479	0.2174	1.0000
3	0.00156202	0.00044406	0.0004	1.0004
4	0.00111796	0.00234028	0.0003	1.0007
5	-.00122232	0.00028539	-0.0003	1.0004
6	-.00150772		-0.0004	1.0000

Factor Pattern

	Factor1	Factor2
M	0.63584	0.52255
P	0.65784	0.38549
C	0.59812	0.30447
E	0.76233	-0.31509
H	0.74908	-0.36797
F	0.83129	-0.30329

The Factor Pattern will help you interpret the factors, i.e., you can identify what the factors represent. From the Factor Pattern, you can see that all the variables have high loadings on Factor 1 and that there is a clear pattern to the sign of loadings on Factor 2. The interpretation will become clearer after the rotation.

Variance Explained by Each Factor

Factor1	Factor2
3.0280932	0.8410268

Final Communality Estimates: Total = 3.869120

M	P	C	E	H	F
0.67735436	0.58135639	0.45044651	0.68042611	0.69651690	0.78301972

Residual Correlations With
Uniqueness on the Diagonal

	M	P	C
M	0.32265	0.00028	0.00059
P	0.00028	0.41864	-0.00084
C	0.00059	-0.00084	0.54955
E	-0.00007	-0.00003	-0.00003
H	-0.00001	0.00007	-0.00000
F	-0.00008	0.00006	0.00013

Residual Correlations With Uniqueness on the Diagonal

	E	H	F
M	-0.00007	-0.00001	-0.00008
P	-0.00003	0.00007	0.00006
C	-0.00003	-0.00000	0.00013
E	0.31957	-0.00099	0.00072
H	-0.00099	0.30348	0.00020
F	0.00072	0.00020	0.21698

Root Mean Square Off-Diagonal Residuals: Overall = 0.00042458

M	P	C	E	H	F
0.00029656	0.00039665	0.00046210	0.00054819	0.00045119	0.00034450

The residuals are all small and the overall RMSR is .0004, indicating that the factor structure explains most of the correlations.

Partial Correlations Controlling Factors

	M	P	C
M	1.00000	0.00076	0.00141
P	0.00076	1.00000	-0.00174
C	0.00141	-0.00174	1.00000
E	-0.00021	-0.00008	-0.00007
H	-0.00004	0.00020	-0.00001
F	-0.00030	0.00019	0.00039

Partial Correlations Controlling Factors

	E	H	F
M	-0.00021	-0.00004	-0.00030
P	-0.00008	0.00020	0.00019
C	-0.00007	-0.00001	0.00039
E	1.00000	-0.00317	0.00275
H	-0.00317	1.00000	0.00079
F	0.00275	0.00079	1.00000

Root Mean Square Off-Diagonal Partials: Overall = 0.00126957

M	P	C	E	H	F
0.00073393	0.00086048	0.00101659	0.00187834	0.00146169	0.00130138

The partial correlations are close to zero and the overall RMSR for partial correlations, .001, is also small, so the estimated factor model seems appropriate according to the RMSRs of the residual and the partial correlation matrices.

This is before the rotation.

```
Plot of Factor Pattern for Factor1 and Factor2

                             Factor1
                                1

                               .9
                      F
                               .8
                    E D
                               .7
                                           B   A
                               .6        C

                               .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

M=A           P=B           C=C           E=D           H=E
F=F
```

The FACTOR Procedure
Rotation Method: Varimax

Orthogonal Transformation Matrix

	1	2
1	0.76668	0.64202
2	-0.64202	0.76668

Rotated Factor Pattern

	Factor1	Factor2
M	0.15200	0.80886
P	0.25687	0.71790
C	0.26309	0.61744
E	0.78676	0.24786
H	0.81055	0.19881
F	0.83205	0.30118

After the rotation, you can see that each variable has a high loading on one factor and a low loading on the other. Therefore, it is easier to interpret the factors.

Variance Explained by Each Factor

Factor1	Factor2
2.1265947	1.7425253

Final Communality Estimates: Total = 3.869120

M	P	C	E	H	F
0.67735436	0.58135639	0.45044651	0.68042611	0.69651690	0.78301972

You can compare this result with the result before the rotation and find out that the final communality estimates are the same although the eigenvalues for each factor (same as variance explained by each factor) has changed.

The FACTOR Procedure
Rotation Method: Varimax

This is after the rotation. You can see that most of the variables are close to either Factor 1 axis or Factor 2 axis. Variables A, B, and C have are highly loaded on Factor 2, whereas variables D, E, and F are highly loaded on Factor 1.

Plot of Factor Pattern for Factor1 and Factor2

```
                             Factor1
                                1

                               .9
                                         F
                               .8    E D

                               .7

                               .6

                               .5

                               .4

                               .3
                                                  C  B
                               .2
                                                        A       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

M=A           P=B           C=C           E=D           H=E
F=F
```