Exploratory Factor Analysis Using SPSS Syntax

In this document I explain how to use SPSS syntax to run exploratory factor analyses.

The data from this study are based on the Attitudes toward Scientists data from chapter 12 in the text. These data represent scores of 1974 respondents on the nine items shown on page 311 of the text and on the last page of this document. The data are in the file "**Scientist data.sav**."

In SPSS, factor analysis is conducted through the *dimension reduction* command. The basic syntax is shown below.

```
FACTOR
  /VARIABLES ALONE BETTER BORING NOFUN GOOD HELP ODD NORELIGN
  NOINTRST
  /MISSING LISTWISE
  /ANALYSIS ALONE BETTER BORING NOFUN GOOD HELP ODD NORELIGN
NOINTRST
  /PRINT INITIAL CORRELATION KMO REPR EXTRACTION ROTATION
  /FORMAT SORT
  /PLOT EIGEN
  /CRITERIA FACTORS(2) ITERATE(25)
  /EXTRACTION PAF
  /CRITERIA ITERATE(25) DELTA(0)
  /ROTATION OBLIMIN.
```

The subcommand **MISSING LISTWISE** specifies listwise deletion of missing data. This is the default option, so the command can be omitted.

The **ANALYSIS** subcommand specifies the variables to be included in the factor analysis. These need not include all the variables on the **VARIABLES** subcommand, so the same set of commands can be used to obtain factor analyses of different sets of variables, if desired.

The **PRINT** subcommand specifies that the initial solution, observed correlation matrix, KMO and Bartlett's test results, reproduced correlation matrix, and the factors solutions after factor extraction and after rotation should be printed in the output.

The **FORMAT** subcommand specifies that the tables of loadings (both pattern and structure) should be sorted by size, from largest to smallest loadings.

The **PLOT** subcommand specifies that the scree plot should be printed in the output.

The **CRITERIA** subcommand specifies the default number of iterations (25) and the default value of the delta parameter for the oblimin rotation (0).

The **EXTRACTION** command specifies the extraction method as Principal Axis Factoring (PAF), which results in a factor, rather than a component, analysis.

The **ROTATION** subcommand specifies that oblimin rotation should be used.

The first part of the output is the matrix of correlations among the variables:

Correlation Matrix

		ALONE	BETTER	BORING	NOFUN	GOOD	HELP	ODD	NORELIGN	NOINTRST
Correlation	ALONE	1.000	.076	.252	.260	.031	-.010	.216	.096	.351
	BETTER	.076	1.000	-.136	.113	.470	.425	.138	.067	.123
	BORING	.252	-.136	1.000	.261	-.131	-.231	.303	.090	.286
	NOFUN	.260	.113	.261	1.000	.046	.007	.388	.264	.454
	GOOD	.031	.470	-.131	.046	1.000	.402	.029	.051	.049
	HELP	-.010	.425	-.231	.007	.402	1.000	.027	.046	.006
	ODD	.216	.138	.303	.388	.029	.027	1.000	.311	.504
	NORELIGN	.096	.067	.090	.264	.051	.046	.311	1.000	.292
	NOINTRST	.351	.123	.286	.454	.049	.006	.504	.292	1.000

Although there are several pairs of variables with moderate correlations, it is difficult to see an overall pattern. This is where factor analysis can help us.

Next in the output are Bartlett's Test of Sphericity and the Kaiser-Meyer-Olkin (KMO) index. The KMO ranges from 0 to 1, with higher values indicating greater amenability to factoring. According to Kaiser's criteria, the value of .750 shown below is between "middling" and "meritorious." The "middling" KMO value is not surprising, as correlations among variables measured on a four-point Likert scale, as these variables are, will be somewhat attenuated in comparison to correlations among variables measured on more continuous scales.

The chi-square value associated with Bartlett's test is statistically significant. This is a test of the null hypothesis that the correlation matrix is an identity matrix (i.e, that the variables are all correlated at 0). The fact that this hypothesis is rejected tells us that the correlation matrix is not an identity matrix, and therefore may be worth factoring. It should be pointed out, however that a) this test represents a rather low bar, as we would hope that, at the very least, our variables are correlated at greater than 0 levels, and b) this test is very powerful, and nearly always results in rejection with a reasonable sample size.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.750
Bartlett's Test of Sphericity	Approx. Chi-Square	604.585
	df	36
	Sig.	.000

The communalities are shown next. The values in the column labeled "Initial" match those in Table 12.4 of the text. These are the iterated values described on page 313.

The values in the column labeled "extraction" are those obtained after extracting the 2 specified factors. These values are obtained by squaring the unrotated loading coefficients and summing these across the 2 factors.

Communalities

	Initial	Extraction
ALONE	.160	.175
BETTER	.314	.504
BORING	.217	.280
NOFUN	.272	.372
GOOD	.273	.408
HELP	.259	.405
ODD	.337	.441
NORELIGN	.138	.150
NOINTRST	.387	.565

Extraction Method: Principal Axis Factoring.

The table below shows the eigenvalues and percentages of explained variance. The latter values are calculated as the eigenvalue divided by the total number of variables (here, 9) and multiplied by 100. The values in the column labeled "cumulative %" are obtained by summing down the "% of variance" column.

As explained in the text (pp. 326-327), values in the set of columns under "Initial Eigenvalues" are based on the full rather than the reduced correlation matrix. This is somewhat confusing because factor analyses such as principal axis factoring are based on the reduced matrix.

Given the confusing nature of the initial eigenvalues, we will simply ignore these values and concentrate on those under the heading "Extraction Sums of Squared Loadings." The two factors account for about 21% and 15% of the variance, respectively, for a total of nearly 37%. These values are disappointingly small and indicate that much of the variance in the set of variables is not accounted for by the factors. The low values are likely due, at least in part, to the coarse nature of the variables' 1 – 4 Likert scale. Such scales, as noted previously, attenuate the correlations among the variables relative to what would be obtained from a more continuous scale.

The values under the heading "Rotation Sums of Squared Loadings" represent the values obtained after the 2 factors have been rotated. Because the rotation process spreads the variances across the factors in a different way than in the unrotated solution, these values are somewhat different than their unrotated counterparts.

As a final note, the values under the "total" columns in these last two sets are not eigenvalues, which is why they are not labeled as such. This is because they are calculated from the reduced correlation matrix, as explained on page 327 of the text. And, as indicated by note *a* below the table, the rotation sums of squared loadings cannot be added together to obtain a total. This is because the values obtained from a rotated solution overlap, to some extent, due to the correlation among the factors. The overlapping part of the variance is included in the value of each factor and is thus counted twice. To illustrate, adding together the values of 1.924 and 1.376 for the unrotated solution yields 3.3, whereas adding together the two values for the rotated solution yields a slightly higher value of 3.303. The latter value would be higher if the variables were more highly correlated. However, their correlation is a mere .023.

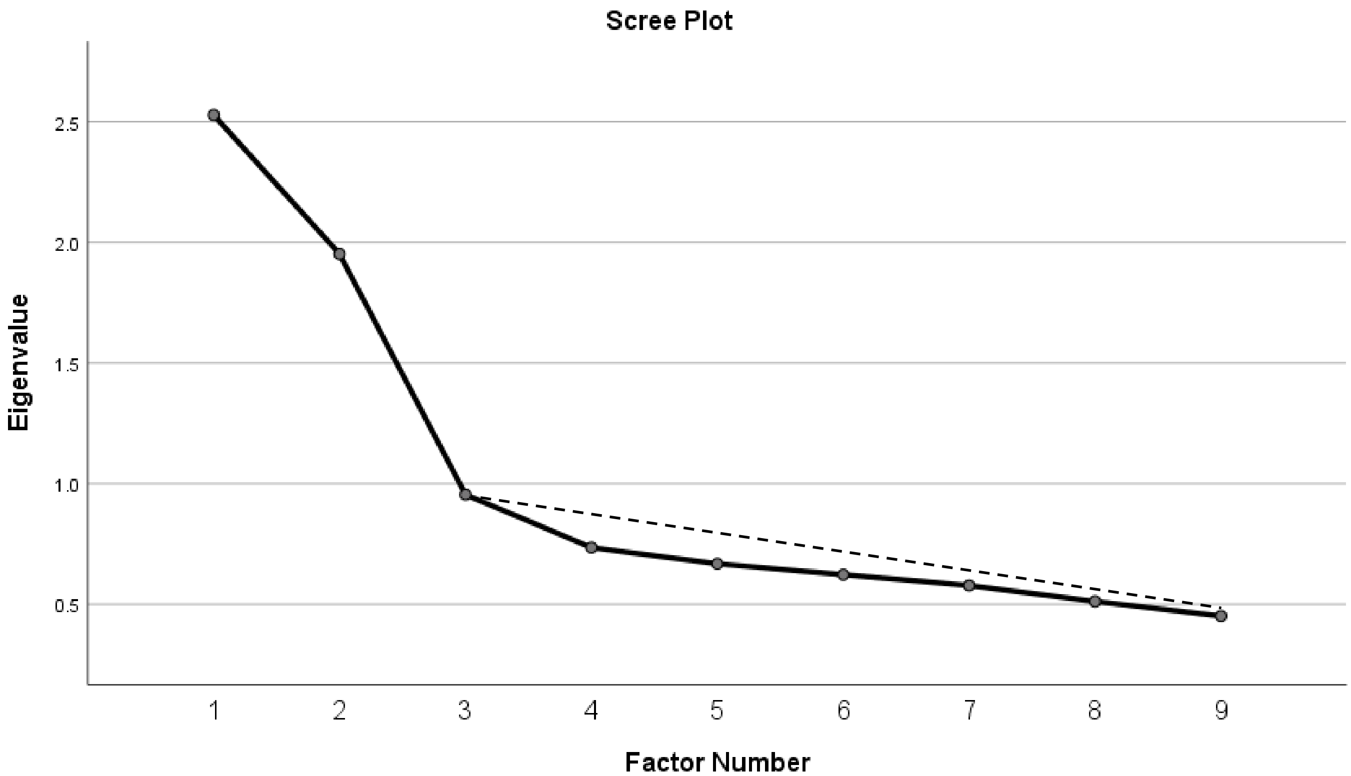
Total Variance Explained

	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
Factor	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	2.528	28.089	28.089	1.924	21.374	21.374	1.910
2	1.952	21.689	49.778	1.376	15.285	36.658	1.393
3	.954	10.604	60.382				
4	.735	8.168	68.550				
5	.668	7.420	75.969				
6	.622	6.915	82.885				
7	.577	6.416	89.301				
8	.512	5.685	94.985				
9	.451	5.015	100.000				

Extraction Method: Principal Axis Factoring.

a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.

The scree plot is shown below:



The number of factors is determined as the number before the line based on the eigenvalues levels off to become relatively straight. I have superimposed a dashed straight line along the eigenvalue line beginning at factor 3. Although the dots representing factors 3 – 9 do not fall directly on the straight line, they are fairly close. However, we may wish to examine a 3 factor solution.

One way of determining whether the correct number of factors has been extracted is to examine a matrix of residual correlations. These residuals are the differences between the observed, or actual, correlations and the correlations reproduced from the factor model (see pp.305-306 in the text for an explanation of the calculations for reproduced correlations).

If the number of factors extracted is incorrect, the factor model (in our example, a 2-factor model) will not be able to reproduce all the correlations sufficiently, and there will be some large residuals. In SPSS, the number of residual correlations greater than .05 is printed below the table.

Some researchers use rough rules of thumb based on residual correlations to assess whether the number of factors is adequate. For example, if no more than 10% of the residuals correlations are greater than .05, the number of factors may be considered adequate.

A more useful way to use the residual correlations to assess model fit is to examine the variable pairs with large residuals in an attempt to determine why the model was unable to account for the observed correlation. Examples of this process are provided in Chapter 13 for confirmatory factor analysis, and the same can be done for exploratory factor analyses.

The reproduced correlations are shown in the top half of the table below and residual correlations are in the bottom half. To take an example, consider the correlation between the variables ALONE and BORING.

From the correlation matrix shown previously, we know that the observed correlation between the two is .252.

The reproduced correlation is .185, so the actual correlation has been underestimated by .252-.185 = .067 (note that the value for the reproduced correlation is slightly different: this is because the actual calculations are carried out to more decimal places than those shown in the table).

Overall, the residual correlations for the 2-factor model are quite small, indicating 2 factors are probably sufficient.

Reproduced Correlations

		ALONE	BETTER	BORING	NOFUN	GOOD	HELP	ODD	NORELIGN	NOINTRST
Reproduced Correlation	ALONE	.175[a]	.057	.185	.254	.015	-.012	.277	.159	.313
	BETTER	.057	.504[a]	-.140	.112	.449	.439	.124	.101	.133
	BORING	$\boxed{.185}$	-.140	.280[a]	.258	-.168	-.197	.280	.148	.320
	NOFUN	.254	.112	.258	.372[a]	.048	.007	.405	.235	.459
	GOOD	.015	.449	-.168	.048	.408[a]	.404	.055	.059	.055
	HELP	-.012	.439	-.197	.007	.404	.405[a]	.011	.033	.005
	ODD	.277	.124	.280	.405	.055	.011	.441[a]	.255	.499
	NORELIGN	.159	.101	.148	.235	.059	.033	.255	.150[a]	.289
	NOINTRST	.313	.133	.320	.459	.055	.005	.499	.289	.565[a]
Residual[b]	ALONE		.019	.066	.006	.016	.003	-.061	-.063	.038
	BETTER	.019		.004	.002	.020	-.014	.013	-.034	-.010
	BORING	$\boxed{.066}$	.004		.002	.037	-.034	.023	-.058	-.034
	NOFUN	.006	.002	.002		-.002	-.001	-.017	.030	-.005
	GOOD	.016	.020	.037	-.002		-.003	-.026	-.008	-.006
	HELP	.003	-.014	-.034	-.001	-.003		.017	.012	.001
	ODD	-.061	.013	.023	-.017	-.026	.017		.056	.005
	NORELIGN	-.063	-.034	-.058	.030	-.008	.012	.056		.003
	NOINTRST	.038	-.010	-.034	-.005	-.006	.001	.005	.003	

Extraction Method: Principal Axis Factoring.

a. Reproduced communalities

b. Residuals are computed between observed and reproduced correlations. There are 5 (13.0%) nonredundant residuals with absolute values greater than 0.05.

The pattern and structure matrices are shown next. Values in the pattern matrix are the correlations of the variables with each factor, holding constant or partialing out all other factors. For example, the pattern loading of NOINTRST with factor 1 its correlation with factor 1, holding constant its correlation with factor 2.

Pattern Matrix[a]

	Factor 1	Factor 2
NOINTRST	.750	.039
ODD	.662	.044
NOFUN	.608	.037
BORING	.448	-.291
ALONE	.418	-.002
NORELIGN	.379	.068
BETTER	.125	.696
GOOD	.026	.637
HELP	-.041	.636

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.

a. Rotation converged in 4 iterations.

Values in the structure matrix are the correlations of the variables with the factors. For these coefficients, other factors are not partialed out.

In this example, values of the pattern and structure coefficients are very similar. This is because of the low (.023) correlation between the 2 factors. This indicates that the factors are essentially uncorrelated, so partialing out the other factor has very little effect.

Structure Matrix

	Factor 1	Factor 2
NOINTRST	.751	.056
ODD	.663	.059
NOFUN	.609	.051
BORING	.442	-.281
ALONE	.418	.007
NORELIGN	.381	.077
BETTER	.141	.699
GOOD	.040	.638
HELP	-.026	.635

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.

The factor correlation matrix is shown last in the SPSS output and shows the very low correlation between the 2 factors.

Factor Correlation Matrix

Factor	1	2
1	1.000	.023
2	.023	1.000

Extraction Method: Principal Axis Factoring.
Rotation Method: Oblimin with Kaiser Normalization.

Attitudes Toward Scientists items

1. A scientist usually works alone. (ALONE)
2. Scientific researchers are dedicated people who work for the good of humanity (GOOD)
3. Scientists don't get as much fun out of life as other people do. (NOFUN)
4. Scientists are helping to solve challenging problems. (HELP)
5. Scientists are apt to be odd and peculiar people. (ODD)
6. Most scientists want to work on things that will make life better for the average person. (BETTER)
7. Scientists are not likely to be very religious people. (NORELIGN)
8. Scientists have few interests other than their work. (NOINTRST)
9. A job as a scientist would be boring. (BORING)