15
Factor analysis

Factor analysis is different from many of the other techniques presented in this book. It is not designed to test hypotheses or to tell you whether one group is significantly different from another. It is included in IBM SPSS as a 'data reduction' technique. It takes a large set of variables and looks for a way the data may be 'reduced' or summarised using a smaller set of factors or components. It does this by looking for 'clumps' or groups among the intercorrelations of a set of variables. This is an almost impossible task to do 'by eye' with anything more than a small number of variables.

This family of factor analytic techniques has a number of different uses. It is used extensively by researchers involved in the development and evaluation of tests and scales. The scale developer starts with a large number of individual scale items and questions and, by using factor analytic techniques, they can refine and reduce these items to form a smaller number of coherent subscales. Factor analysis can also be used to reduce a large number of related variables to a more manageable number, prior to using them in other analyses such as multiple regression or multivariate analysis of variance.

There are two main approaches to factor analysis that you will see described in the literature—exploratory and confirmatory. Exploratory factor analysis is often used in the early stages of research to gather information about (explore) the interrelationships among a set of variables. Confirmatory factor analysis, on the other hand, is a more complex and sophisticated set of techniques used later in the research process to test (confirm) specific hypotheses or theories concerning the structure underlying a set of variables.

The term 'factor analysis' encompasses a variety of different, although related, techniques. One of the main distinctions is between principal components analysis (PCA) and factor analysis (FA). These two sets of techniques are similar in many ways and are often used interchangeably by researchers. Both attempt to produce a smaller number of linear combinations of the original variables in a way that captures (or accounts for) most of the variability in the pattern of correlations. They do differ in a number of ways, however. In principal components analysis the original variables are transformed into a smaller set of linear combinations, with all of the variance in the variables being used. In factor analysis, however, factors are estimated using a mathematical model, whereby only the shared variance is analysed (see Tabachnick & Fidell 2013, Chapter 13, for more information on this).

Although both approaches (PCA and FA) often produce similar results, books on the topic often differ in terms of which approach they recommend. Stevens (1996, pp. 362–3) admits a preference for principal components analysis and gives a number of reasons for this. He suggests that it is psychometrically sound and simpler mathematically, and it avoids some of the potential problems with 'factor indeterminacy' associated with factor analysis (Stevens 1996, p. 363). Tabachnick and Fidell (2013), in their review of PCA and FA, conclude: 'If you are interested in a theoretical solution uncontaminated by unique and error variability … FA is your choice. If, on the other hand, you simply want an empirical summary of the data set, PCA is the better choice' (p. 640).

I have chosen to demonstrate principal components analysis in this chapter. If you would like to explore the other approaches further, see Tabachnick and Fidell (2013).

Note: although PCA technically yields components, many authors use the term 'factor' to refer to the output of both PCA and FA. So don't assume, if you see the term 'factor' when you are reading journal articles, that the author has used FA. Factor analysis is used as a general term to refer to the entire family of techniques.

Another potential area of confusion involves the use of the word 'factor', which has different meanings and uses in different types of statistical analyses. In factor analysis, it refers to the group or clump of related variables; in analysis of variance techniques, it refers to the independent variable. These are very different things, despite having the same name, so keep the distinction clear in your mind when you are performing the different analyses.

STEPS INVOLVED IN FACTOR ANALYSIS

There are three main steps in conducting factor analysis (I am using the term in a general sense to indicate any of this family of techniques, including principal components analysis).

Step 1: Assessment of the suitability of the data for factor analysis

There are two main issues to consider in determining whether a particular data set is suitable for factor analysis: sample size, and the strength of the relationship among the variables (or items). While there is little agreement among authors concerning how large a sample should be, the recommendation generally is: the larger, the better. In small samples, the correlation coefficients among the variables are less reliable, tending to vary from sample to sample. Factors obtained from small data sets do not generalise as well as those derived from larger samples. Tabachnick and Fidell (2013) review this issue and suggest that 'it is comforting to have at least 300 cases for factor analysis' (p. 613). However, they do concede that a smaller sample size (e.g. 150 cases) should be sufficient if solutions have several high loading marker variables (above .80). Stevens (1996, p. 372) suggests that the sample size requirements advocated by researchers have been reducing over the years as more research has been done on the topic. He makes a number of recommendations concerning the reliability of factor structures and the sample size requirements (see Stevens 1996, Chapter 11).

Some authors suggest that it is not the overall sample size that is of concern—rather, the ratio of participants to items. Nunnally (1978) recommends a 10 to 1 ratio; that is, ten cases for each item to be factor analysed. Others suggest that five cases for each item are adequate in most cases (see discussion in Tabachnick & Fidell 2013). I would recommend that you do more reading on the topic, particularly if you have a small sample (smaller than 150) or lots of variables.

The second issue to be addressed concerns the strength of the intercorrelations among the items. Tabachnick and Fidell recommend an inspection of the correlation matrix for evidence of coefficients greater than .3. If few correlations above this level are found, factor analysis may not be appropriate. Two statistical measures are also generated by IBM SPSS to help assess the factorability of the data: Bartlett's test of sphericity (Bartlett 1954), and the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy (Kaiser 1970, 1974). Bartlett's test of sphericity should be significant ($p < .05$) for the factor analysis to be considered appropriate. The KMO index ranges from 0 to 1, with .6 suggested as the minimum value for a good factor analysis (Tabachnick & Fidell 2013).

Step 2: Factor extraction

Factor extraction involves determining the smallest number of factors that can be used to best represent the interrelationships among the set of variables. There are a variety of approaches that can be used to identify (extract) the number of underlying factors or dimensions. Some of the most commonly available extraction techniques (this always conjures up the image for me of a dentist pulling teeth!) are: principal components; principal factors; image factoring; maximum likelihood factoring; alpha factoring; unweighted least squares; and generalised least squares.

The most commonly used approach is principal components analysis. This will be demonstrated in the example given later in this chapter. It is up to the researcher to determine the number of factors that he/she considers best describes the underlying relationship among the variables. This involves balancing two conflicting needs: the need to find a simple solution with as few factors as possible; and the need to explain as much of the variance in the original data set as possible. Tabachnick and Fidell (2013) recommend that researchers adopt an exploratory approach, experimenting with different numbers of factors until a satisfactory solution is found.

There are a number of techniques that can be used to assist in the decision concerning the number of factors to retain: Kaiser's criterion; scree test; and parallel analysis.

Kaiser's criterion

One of the most commonly used techniques is known as Kaiser's criterion, or the eigenvalue rule. Using this rule, only factors with an eigenvalue of 1.0 or more are retained for further investigation (this will become clearer when you see the example presented in this chapter). The eigenvalue of a factor represents the amount of the total variance explained by that factor. Kaiser's criterion has been criticised, however, as resulting in the retention of too many factors in some situations.

Scree test

Another approach that can be used is Catell's scree test (Catell 1966). This involves plotting each of the eigenvalues of the factors (IBM SPSS does this for you) and inspecting the plot to find a point at which the shape of the curve changes direction and becomes horizontal. Catell recommends retaining all factors above the elbow, or break in the plot, as these factors contribute the most to the explanation of the variance in the data set.

Parallel analysis

An additional technique gaining popularity, particularly in the social science literature (e.g. Choi, Fuqua & Griffin 2001; Stober 1998), is Horn's parallel analysis (Horn 1965). Parallel analysis involves comparing the size of the eigenvalues with those obtained from a randomly generated data set of the same size. Only those eigenvalues that exceed the corresponding values from the random data set are retained. This approach to identifying the correct number of components to retain has been shown to be the most accurate, with both Kaiser's criterion and Catell's scree test tending to overestimate the number of components (Hubbard & Allen 1987; Zwick & Velicer 1986). If you intend to publish your results in a journal article in the psychology or education fields you will need to use, and report, the results of parallel analysis. Many journals (e.g. *Educational and Psychological Measurement*, *Journal of Personality Assessment*) are now making it a requirement before they will consider a manuscript for publication. These three techniques are demonstrated in the worked example presented later in this chapter.

Step 3: Factor rotation and interpretation

Once the number of factors has been determined, the next step is to try to interpret them. To assist in this process, the factors are 'rotated'. This does not change the underlying solution—rather, it presents the pattern of loadings in a manner that is easier to interpret. IBM SPSS does not label or interpret each of the factors for you. It just shows you which variables 'clump together'. From your understanding of the content of the variables (and underlying theory and past research), it is up to you to propose possible interpretations.

There are two main approaches to rotation, resulting in either orthogonal (uncorrelated) or oblique (correlated) factor solutions. According to Tabachnick and Fidell (2013), orthogonal rotation results in solutions that are easier to interpret and to report; however, they do require the researcher to assume (usually incorrectly) that the underlying constructs are independent (not correlated). Oblique approaches allow for the factors to be correlated, but they are more difficult to interpret, describe and report (Tabachnick & Fidell 2013, p. 642). In practice, the two approaches (orthogonal and oblique) often result in very similar solutions, particularly when the pattern of correlations among the items is clear (Tabachnick & Fidell 2013). Many researchers conduct both orthogonal and oblique rotations and then report the clearest and easiest to interpret. I always recommend starting with an oblique rotation to check the degree of correlation between your factors.

Within the two broad categories of rotational approaches there are a number of different techniques provided by IBM SPSS (orthogonal: Varimax, Quartimax, Equamax; oblique: Direct Oblimin, Promax). The most commonly used orthogonal approach is the Varimax method, which attempts to minimise the number of variables that have high loadings on each factor. The most commonly used oblique technique is Direct Oblimin. For a comparison of the characteristics of each of these approaches, see Tabachnick and Fidell (2013, p. 643). In the example presented in this chapter, Oblimin rotation will be demonstrated.

Following rotation you are hoping for what Thurstone (1947) refers to as 'simple structure'. This involves each of the variables loading strongly on only one component, and each component being represented by a number of strongly loading variables. This will help you interpret the nature of your factors by checking the variables that load strongly on each of them.

Additional resources

In this chapter only a very brief overview of factor analysis is provided. Although I have attempted to simplify it here, factor analysis is a sophisticated and complex family of techniques. If you are intending to use factor analysis with your own data, I suggest that you read up on the technique in more depth. For a thorough, but easy-to-follow, book on the topic I recommend Pett, Lackey and Sullivan (2003). For a more complex coverage, see Tabachnick and Fidell (2013).

DETAILS OF EXAMPLE

To demonstrate the use of factor analysis, I will explore the underlying structure of one of the scales included in the **survey.sav** data file provided on the website accompanying this book. One of the scales used was the Positive and Negative Affect Scale (PANAS: Watson, Clark & Tellegen 1988) (see Figure 15.1). This scale consists of twenty adjectives describing different mood states, ten positive (e.g. proud, active, determined) and ten negative (e.g. nervous, irritable, upset). The authors of the scale suggest that the PANAS consists of two underlying dimensions (or factors): positive affect and negative affect. To explore this structure with the current community sample the items of the scale will be subjected to principal components analysis

(PCA), a form of factor analysis that is commonly used by researchers interested in scale development and evaluation.

If you wish to follow along with the steps described in this chapter, you should start IBM SPSS and open the file labelled **survey.sav** on the website that accompanies this book. The variables that are used in this analysis are labelled pn1 to pn20. The scale used in the survey is presented in Figure 15.1. You will need to refer to these individual items when attempting to interpret the factors obtained. For full details and references for the scale, see the Appendix.

Example of research question: What is the underlying factor structure of the Positive and Negative Affect Scale? Past research suggests a two-factor structure (positive affect/negative affect). Is the structure of the scale in this study, using a community sample, consistent with this previous research?

What you need: A set of correlated continuous variables.

What it does: Factor analysis attempts to identify a small set of factors that represents the underlying relationships among a group of related variables.

Assumptions:

1. *Sample size*. Ideally, the overall sample size should be 150+ and there should be a ratio of at least five cases for each of the variables (see discussion in Step 1 earlier in this chapter).
2. *Factorability of the correlation matrix*. To be considered suitable for factor analysis, the correlation matrix should show at least some correlations of $r = .3$ or greater. Bartlett's test of sphericity should be statistically significant at $p < .05$ and the Kaiser-Meyer-Olkin value should be .6 or above. These values are presented as part of the output from factor analysis.
3. *Linearity*. Because factor analysis is based on correlation, it is assumed that the relationship between the variables is linear. It is certainly not practical to check scatterplots of all variables with all other variables. Tabachnick and Fidell (2013) suggest a 'spot check' of some combination of variables. Unless there is clear evidence of a curvilinear relationship, you are probably safe to proceed provided you have an adequate sample size and ratio of cases to variables (see Assumption 1).
4. *Outliers among cases*. Factor analysis can be sensitive to outliers, so as

part of your initial data screening process (see Chapter 6) you should check for these and either remove or recode to a less extreme value.

This scale consists of a number of words that describe different feelings and emotions. For each item indicate to what extent you have felt this way during the past few weeks. Write a number from 1 to 5 on the line next to each item.

very slightly or not at all	a little	moderately	quite a bit	extremely
1	2	3	4	5
1. interested________		8. distressed________		15. excited________
2. upset________		9. strong________		16. guilty________
3. scared________		10. hostile________		17. enthusiastic_______
4. proud________		11. irritable________		18. alert________
5. ashamed________		12. inspired________		19. nervous________
6. determined_______		13. attentive________		20. jittery________
7. active________		14. afraid________		

Figure 15.1
Positive and Negative Affect Scale (PANAS)

PROCEDURE FOR FACTOR ANALYSIS

Before you start the following procedure, choose **Edit** from the menu, select **Options**, and make sure there is a tick in the box **No scientific notation for small numbers in tables**.

Procedure (Part 1)

1. From the menu at the top of the screen, click on **Analyze**, then select **Dimension Reduction**, and then **Factor.**
2. Select all the required variables (or items on the scale). In this case, I would select the items that make up the PANAS Scale (pn1 to pn20). Move them into the **Variables** box.
3. Click on the **Descriptives** button.
   In the **Statistics** section, make sure that **Initial Solution** is ticked.
   In the section marked **Correlation Matrix**, select the options **Coefficients** and **KMO and Bartlett's test of sphericity**. Click on **Continue.**
4. Click on the **Extraction** button.
   In the **Method** section, make sure **Principal components** is shown, or choose one of the other factor extraction techniques (e.g. Maximum likelihood).
   In the **Analyze** section, make sure the **Correlation matrix** option is selected. In the **Display** section, select **Screeplot** and make sure the **Unrotated factor solution** option is also selected.
   In the **Extract** section, select **Based on Eigenvalue** or, if you want to force a specific number of factors, click on **Fixed number of factors** and type in the number. Click on **Continue.**
5. Click on the **Rotation** button. Choose **Direct Oblimin** and press **Continue.**
6. Click on the **Options** button.
   In the **Missing Values** section, click on **Exclude cases pairwise**.
   In the **Coefficient Display Format** section, click on **Sorted by size** and **Suppress small coefficients.** Type the value of .3 in the box next to **Absolute value below.** This means that only loadings above .3 will be displayed, making the output easier to interpret.
7. Click on **Continue** and then **OK** (or on **Paste** to save to **Syntax Editor**).

The syntax from this procedure is:

```
FACTOR
  /VARIABLES pn1 pn2 pn3 pn4 pn5 pn6 pn7 pn8 pn9 pn10 pn11 pn12 pn13
  pn14 pn15 pn16 pn17 pn18 pn19 pn20
  /MISSING PAIRWISE
  /ANALYSIS pn1 pn2 pn3 pn4 pn5 pn6 pn7 pn8 pn9 pn10 pn11 pn12 pn13
  pn14 pn15 pn16 pn17 pn18 pn19 pn20
  /PRINT INITIAL CORRELATION KMO EXTRACTION ROTATION
  /FORMAT SORT BLANK(.3)
  /PLOT EIGEN
  /CRITERIA MINEIGEN(1) ITERATE(25)
  /EXTRACTION PC
  /CRITERIA ITERATE(25) DELTA(0)
  /ROTATION OBLIMIN
  /METHOD=CORRELATION.
```

Selected output generated from this procedure is shown below:

Correlation Matrix

		PN1	PN2	PN3	PN4	PN5	PN6	PN7	PN8	PN9	PN10
	PN1	1.000	-.139	-.152	.346	-.071	.352	.407	-.250	.416	-.122
	PN2	-.139	1.000	.462	-.141	.271	-.127	-.197	.645	-.188	.411
	PN3	-.152	.462	1.000	-.102	.247	-.097	-.255	.494	-.200	.234
	PN4	.346	-.141	-.102	1,000	-.156	.295	.331	-.152	.396	-.056
	PN5	-.071	.271	.247	-.156	1.000	-.067	-.248	.278	-.201	.258
	PN6	.352	-.127	-.097	.295	-.067	1.000	.329	-.048	.426	.077
	PN7	.407	-.197	-.255	.331	-.248	.329	1.000	-.232	.481	-.093
	PN8	-.250	.645	.494	-.152	.278	-.048	-.232	1.000	-.181	.380
	PN9	.416	-.188	-.200	.396	-.201	.426	.481	-.181	1.00	-.070
Correl	PN10	-.122	.411	.234	-.056	.258	.077	-.093	.380	-.070	1.000
	PN11	-.210	.500	.333	-.179	.266	-.043	-.214	.464	-.210	.583
	PN12	.482	-.204	-.135	.315	-.063	.401	.400	-.175	.407	-.074
	PN13	.491	-.171	-.165	.329	-.137	.336	.391	-.199	.427	-.114
	PN14	-.151	.406	.810	-.107	.302	-.090	-.271	.459	-.198	.263
	PN15	.413	-.136	-.085	.317	-.062	.276	.329	-.098	.362	-.067
	PN16	-.177	.314	.330	-.121	.539	-.099	-.221	.378	-.164	.314
	PN17	.562	-.208	-.190	.368	-.156	.396	.484	-.218	.465	-.134
	PN18	.466	-.196	-.181	.338	-.189	.451	.458	-.234	.462	-.066
	PN19	-.148	.459	.560	-.124	.285	-.050	-.234	.480	-.198	.339
	PN20	-.176	.425	.424	-.171	.245	-.025	-.204	.431	-.219	.367

KMO and Bartlett's Test

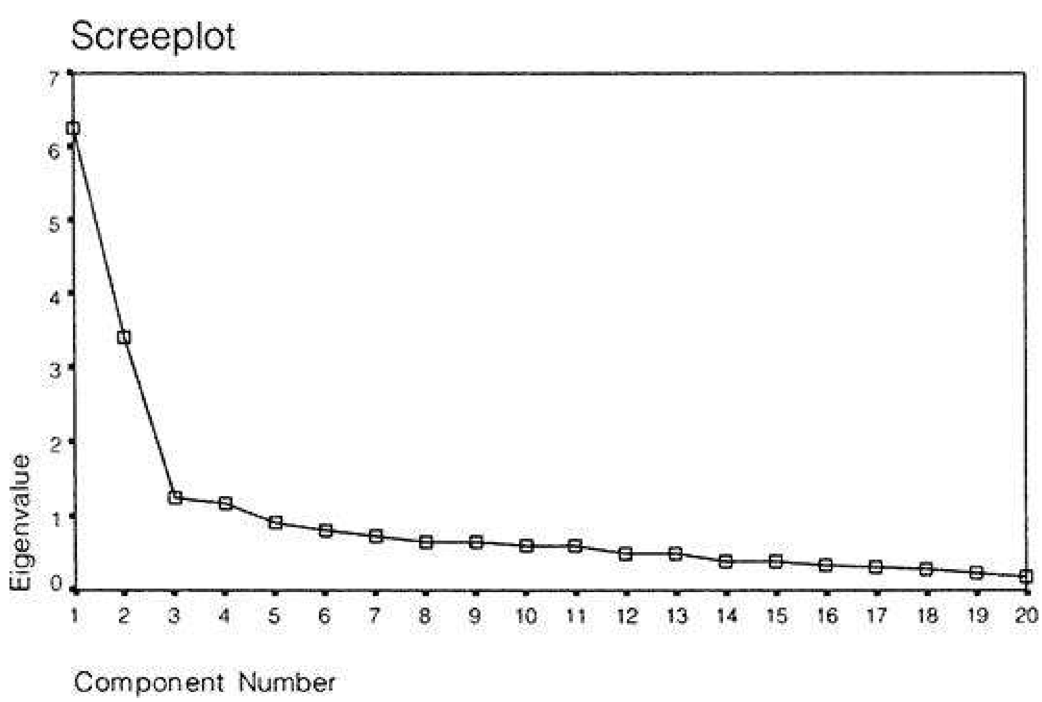
Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.874
Bartlett's Test of Sphericity	Approx Chi-Square	3966.539
	df	190
	Sig.	.000

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	6.250	31.249	31.249	6.250	31.249	31.249	5.254
2	3.396	16.979	48.228	3.396	16.979	48.228	4.410
3	1.223	6.113	54.341	1.223	6.113	54.341	2.639
4	1.158	5.788	60.130	1.158	5.788	60.130	2.428
5	.898	4.490	64.619				
6	.785	3.926	68.546				
7	.731	3.655	72.201				
8	.655	3.275	75.476				
9	.650	3.248	78.724				
10	.601	3.004	81.728				
11	.586	2.928	84.656				
12	.499	2.495	87.151				
13	.491	2.456	89.607				
14	.393	1.964	91.571				
15	.375	1.875	93.446				
16	.331	1.653	95.100				
17	.299	1.496	96.595				
18	.283	1.414	98.010				
19	.223	1.117	99.126				
20	.175	.874	100.000				

Extraction Method: Principal Component Analysis.

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



Component Matrix[a]

	Component			
	1	2	3	4
PN17	.679	.474		
PN18	.639	.404		
PN7	.621			
PN8	-.614	.420		
PN9	.609	.323		
PN13	.607	.413		
PN1	.600	.381		
PN2	-.591	.408		
PN 3	-.584	.449	-.457	
PN14	-.583	.456	-.451	
PN12	.582	.497		
PN19	-.569	.545		
PN11	-.554	.366	.462	
PN20	-.545	.459		
PN4	.474			
PN15	.477	.483		
PN6	.432	.437		
PN10	-.416	.426	.563	
PN5	-.429			.649
PN16	-.474	.357		.566

a. Extraction Method: Principal Component Analysis.

Pattern Matrixa

	Component			
	1	2	3	4
pn17	.836			
pn12	.795			
pn13	.735			
pn18	.732			
pn15	.721			
pn1	.708			
pn9	.638			
pn6	.604			
pn7	.578			
pn4	.531			
pn3		.909		
pn14		.888		
pn19		.799		
pn20		.677		
pn8		.477	.413	
pn10			.808	
pn11			.707	
pn2		.434	.473	
pn5				.830
pn16				.773

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization

a. Rotation converged in 8 iterations.

INTERPRETATION OF OUTPUT

As with most IBM SPSS procedures, there is a lot of output generated. In this section, I will take you through the key pieces of information that you need.

Interpretation of output—Part 1

Step 1

To verify that your data set is suitable for factor analysis, check that the **Kaiser-Meyer-Olkin Measure of Sampling Adequacy** (KMO) value is .6 or above and that the **Bartlett's Test of Sphericity** value is significant (i.e. the Sig. value should be .05 or smaller). In this example the KMO value is .874 and Bartlett's test is significant (p = .000), therefore factor analysis is appropriate. In the **Correlation Matrix** table (not shown here for space reasons), look for correlation coefficients of .3 and above (see Assumption 2). If you don't find many in your matrix, you should reconsider the use of factor analysis.

Step 2

To determine how many components (factors) to 'extract', we need to consider a few pieces of information provided in the output. Using Kaiser's criterion, we are interested only in components that have an eigenvalue of 1 or more. To determine how many components meet this criterion, we need to look in the **Total Variance Explained** table. Scan down the values provided in the first set of columns, labelled **Initial Eigenvalues**. The eigenvalues for each component are listed. In this example, only the first four components recorded eigenvalues above 1 (6.25, 3.396, 1.223, 1.158). These four components explain a total of 60.13 per cent of the variance (see **Cumulative** % column).

Step 3

Often, using the Kaiser criterion, you will find that too many components are extracted, so it is important to also look at the **Screeplot**. What you look for is a change (or elbow) in the shape of the plot. Only components above this point are retained. In this example, there is quite a clear break between the second and third components. Components 1 and 2 explain or capture much more of the variance than the remaining components. From this plot, I would recommend retaining (extracting) only two components. There is also another little break after the fourth component. Depending on the research context, this might also be worth exploring. Remember, factor analysis is used as a data exploration technique, so the interpretation and the use you put it to is up to your judgment rather than any hard and fast statistical rules.

Step 4

The third way of determining the number of factors to retain is parallel analysis (see discussion earlier in this chapter). For this procedure, you need to use the list of eigenvalues provided in the **Total Variance Explained** table and some additional information that you must get from another little statistical program (developed by Marley Watkins, 2000) that is available from the website for this book. Follow the links to the **Additional Material** site and download the zip file (**parallel analysis.zip**) onto your computer. Unzip this onto your hard drive and click on the file **MonteCarloPA.exe**.

A program will start that is called Monte Carlo PCA for Parallel Analysis. You will be asked for three pieces of information: the number of variables you are analysing (in this case, 20); the number of participants in your sample (in this case, 435); and the number of replications (specify 100). Click on **Calculate**. Behind the scenes, this program will generate 100 sets of random data of the same size as your real data file (20 variables × 435 cases). It will calculate the average eigenvalues for these 100 randomly generated samples and print these out for you. See Table 15.1. Your numbers will not necessarily match my values shown in Table 15.1 as each of the samples are randomly generated, so will vary each time you use the program.

Your job is to systematically compare the first eigenvalue you obtained in IBM SPSS with the corresponding first value from the random results generated by parallel analysis. If your value is larger than the criterion value from parallel analysis, you retain this factor; if it is less, you reject it. The results for this example are summarised in Table 15.2. The results of parallel analysis support our decision from the screeplot to retain only two factors for further investigation.

Step 5

Moving back to our IBM SPSS output, the final table we need to look at is the **Component Matrix**. This shows the unrotated loadings of each of the items on the four components. IBM SPSS uses the Kaiser criterion (retain all components with eigenvalues above 1) as the default. You will see from this table that most of the items load quite strongly (above .4) on the first two components. Very few items load on Components 3 and 4. This suggests that a two-factor solution is likely to be more appropriate.

Step 6

Before we make a final decision concerning the number of factors, we should have a look at the rotated four-factor solution that is shown in the **Pattern Matrix** table. This shows the items loadings on the four factors with ten items loading above .3 on Component 1, five items loading on Component 2, four items on Component 3 and only two items loading on Component 4. Ideally, we would like three or more items loading on each component so this solution is not optimal, further supporting our decision to retain only two factors.

Using the default options in IBM SPSS, we obtained a four-factor solution. It is now necessary to go back and 'force' a two-factor solution.

7/03/2004	11:58:37 AM		
Number of variables:	20		
Number of subjects:	435		
Number of replications:	100		
Eigenvalue		Random Eigenvalue	Standard Dev
1		1.3984	.0422
2		1.3277	.0282
3		1.2733	.0262
4		1.2233	.0236
5		1.1832	.0191
6		1.1433	.0206
7		1.1057	.0192
8		1.0679	.0193
9		1.0389	.0186
10		1.0033	.0153
11		0.9712	.0180
12		0.9380	.0175
13		0.9051	.0187
14		0.8733	.0179
15		0.8435	.0187
16		0.8107	.0185
17		0.7804	.0190
18		0.7449	.0194
19		0.7090	.0224
20		0.6587	.0242
7/03/2004	11:58:50 AM		

Montecarlo PCA for Parallel Analysis

Watkins, M. W. (2000). MonteCarlo PCA for parallel analysis [computer software]. State College, PA: Ed & Psych Associates.

Table 15.1
Output from parallel analysis

Component number	Actual eigenvalue from PCA	Criterion value from parallel analysis	Decision
1	6.250	1.3984	accept
2	3.396	1.3277	accept
3	1.223	1.2733	reject
4	1.158	1.2233	reject
5	.898	1.1832	reject

Table 15.2
Comparison of eigenvalues from PCA and criterion values from parallel analysis

Procedure (Part 2)

1. Repeat all steps in Procedure (Part 1), but when you click on the **Extraction** button click on **Fixed number of factors**. In the box next to **Factors to extract** type in the number of factors you would like to extract (e.g. 2).
2. Click on **Continue** and then **OK.**

Some of the output generated is shown below.

Communalities

	Initial	Extraction
pn1	1.000	.505
pn2	1.000	.516
pn3	1.000	.543
pn4	1.000	.308
pn5	1.000	.258
pn6	1.000	.377
pn7	1.000	.445
pn8	1.000	.553
pn9	1.000	.475
pn10	1.000	.355
pn11	1.000	.440
pn12	1.000	.586
pn13	1.000	.538
pn14	1.000	.548
pn15	1.000	.462
pn16	1.000	.352
pn17	1.000	.686
pn18	1.000	.572
pn19	1.000	.620
pn20	1.000	.507

Extraction Method: Principal Component Analysis.

Total Variance Explained

	Initial Eigenvalues	Extraction Sums of Squared	Rotation Sums of

Component				Loadings			Squared Loadingsa
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	6.250	31.249	31.249	6.25	31.249	31.249	5.277
2	3.396	16.979	48.228	3.39	16.979	48.228	5.157
3	1.223	6.113	54.341				
4	1.158	5.788	60.130				
5	.898	4.490	64.619				
6	.785	3.926	68.546				
7	.731	3.655	72.201				
8	.655	3.275	75.476				
9	.650	3.248	78.724				
10	.601	3.004	81.728				
11	.586	2.928	84.656				
12	.499	2.495	87.151				
13	.491	2.456	89.607				
14	.393	1.964	91.571				
15	.375	1.875	93.446				
16	.331	1.653	95.100				
17	.299	1.496	96.595				
18	.283	1.414	98.010				
19	.223	1.117	99.126				
20	.175	.874	100.000				

Extraction Method: Principal Component Analysis.

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

Component Matrixa

	Component	
	1	2
pn 17	.679	.474
pn18	.639	.404
pn7	.621	
pn8	-.614	.420
pn9	.609	.323
pn13	.607	.413
pn1	.600	.381
pn2	-.591	.408
pn3	-.584	.449
pn14	-.583	.456
pn12	.582	.497
pn19	-.569	.545
pn11	-.554	.366
pn20	-.545	.459
pn4	.474	
pn16	-.474	.357
pn5	-.429	
pn15	.477	.483
pn6	.432	.437

pn10	-.416	.426

Extraction Method: Principal Component Analysis,
a. 2 components extracted.

Pattern Matrixa

	Component	
	1	2
pn17	.825	
pn12	.781	
pn18	.742	
pn13	.728	
pn15	.703	
pn1	.698	
pn9	.656	
pn6	.635	
pn7	.599	
pn4	.540	
pn19		.806
pn14		.739
pn3		.734
pn8		.728
pn20		.718
pn2		.704
pn11		.645
pn10		.613
pn16		.589
pn5		.490

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.
a. Rotation converged in 6 iterations.

Structure Matrix

	Component	
	1	2
pn17	.828	
pn12	.763	
pn18	.755	
pn13	.733	
pn1	.710	
pn9	.683	
pn15	.670	
pn7	.646	-.338
pn6	.605	
pn4	.553	
pn19		.784
pn8		.742
pn14		.740
pn3		.737
pn2		.717

pn20			.712
pn11			.661
pn16			.593
pn10			.590
pn5			.505

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

Component Correlation Matrix

Component	1	2
1	1.000	-.277
2	-.277	1.000

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.

Interpretation of output—Part 2: Oblimin rotation of two-factor solution

The first thing we need to check is the percentage of variance explained by this two-factor solution shown in the **Total Variance Explained** table. For the two-factor solution only 48.2 per cent of the variance is explained, compared with over 60 per cent explained by the four-factor solution.

After rotating the two-factor solution, there are three new tables at the end of the output you need to consider. First, have a look at the **Component Correlation Matrix** (at the end of the output). This shows you the strength of the relationship between the two factors (in this case the value is quite low, at –.277). This gives us information to decide whether it was reasonable to assume that the two components were not related (the assumption underlying the use of Varimax rotation) or whether it is necessary to use, and report, the Oblimin rotation solution shown here.

In this case the correlation between the two components is quite low, so we would expect very similar solutions from the Varimax and Oblimin rotation. If, however, your components are more strongly correlated (e.g. above .3), you may find discrepancies between the results of the two approaches to rotation. If that is the case, you need to report the Oblimin rotation.

Oblimin rotation provides two tables of loadings. The **Pattern Matrix** shows the factor loadings of each of the variables. Look for the highest loading items on each component to identify and label the component. In this example, the main loadings on Component 1 are items 17, 12, 18 and 13. If you refer back to the actual items themselves (presented earlier in this chapter), you will see that these are all positive affect items (enthusiastic, inspired, alert, attentive). The main items on Component 2 (19, 14, 3, 8) are negative affect items (nervous, afraid, scared, distressed). In this case, identification and labelling of the two components is easy. This is not always the case, however.

The **Structure Matrix** table, which is unique to the Oblimin output, provides information about the correlation between variables and factors. If you need to present the Oblimin rotated solution in your output, you must present both of these tables.

Earlier in the output a table labelled **Communalities** is presented. This gives information about how much of the variance in each item is explained. Low values (e.g. less than .3) could indicate that the item does not fit well with the other items in its component. For example, item pn5 has the lowest communality value (.258) for this two-factor solution, and it also shows the lowest loading (.49) on Component 2 (see **Pattern Matrix**). If you are interested in improving or refining a scale, you could use this information to remove items from the scale. Removing items with low communality values tends to increase the total variance explained. Communality values can change dramatically depending on how many factors are retained, so it is often better to interpret the communality values after you have chosen how many factors you should retain using the screeplot and parallel analysis.

Warning: the output in this example is a very 'clean' result. Each of the variables loaded strongly on only one component, and each component was represented by a number of strongly loading variables (an example of 'simple structure'). For a discussion of this topic, see Tabachnick and Fidell (2013, p. 652). Unfortunately, with your own data you will not always have such a straightforward result. Often you will find that variables load moderately on a number of different components, and some components will have only one or two variables loading on them. In cases such as this, you may need to consider rotating a different number of components (e.g. one more and one less) to see whether a more optimal solution can be found. If you find that some variables just do not load on the components obtained, you may also need to consider removing them and repeating the analysis. You should read as much as you can on the topic to help you make these decisions. An easy-to-follow book to get you started is Pett, Lackey & Sullivan (2003).

PRESENTING THE RESULTS FROM FACTOR ANALYSIS

The information you provide in your results section is dependent on your discipline area, the type of report you are preparing and where it will be presented. If you are publishing in the areas of psychology and education particularly there are quite strict requirements for what needs to be included in a journal article that involves the use of factor analysis. You should include details of the method of factor extraction used, the criteria used to determine the number of factors (this should include parallel analysis), the type of rotation technique used (e.g. Varimax, Oblimin), the total variance explained, the initial eigenvalues, and the eigenvalues after rotation.

A table of loadings should be included showing all values (not just those above .3). For the Varimax rotated solution, the table should be labelled 'pattern/structure coefficients'. If Oblimin rotation was used, then both the Pattern Matrix and the Structure Matrix coefficients should be presented in full (these can be combined into one table as shown below), along with information on the correlations among the factors.

The results of the output obtained in the example above could be presented as follows:

> The 20 items of the Positive and Negative Affect Scale (PANAS) were subjected to principal components analysis (PCA) using SPSS version 18. Prior to performing PCA, the suitability of data for factor analysis was assessed. Inspection of the correlation matrix revealed the presence of many coefficients of .3 and above. The Kaiser-Meyer-Olkin value was .87, exceeding the recommended value of .6 (Kaiser 1970, 1974) and Bartlett's Test of Sphericity (Bartlett 1954) reached statistical significance, supporting the factorability of the correlation matrix.
>
> Principal components analysis revealed the presence of four components with eigenvalues exceeding 1, explaining 31.2%, 17%, 6.1% and 5.8% of the variance respectively. An inspection of the screeplot revealed a clear break after the second component. Using Catell's (1966) scree test, it was decided to retain two components for further investigation. This was further supported by the results of Parallel Analysis, which showed only two components with eigenvalues exceeding the corresponding criterion values for a randomly generated data matrix of the same size (20 variables × 435 respondents).
>
> The two-component solution explained a total of 48.2% of the variance, with Component 1 contributing 31.25% and Component 2 contributing 17.0%. To aid in the interpretation of these two components, oblimin rotation was performed. The rotated solution revealed the presence of simple structure (Thurstone 1947), with both components showing a number of strong loadings and all variables loading substantially on only one component. The interpretation of the two components was consistent with previous research on the PANAS Scale, with positive affect items loading strongly on Component 1 and negative affect items loading strongly on Component 2. There was a weak negative correlation between the two factors ($r = -.28$). The results of this analysis support the use of the positive affect items and the negative affect items as separate scales, as suggested by the scale authors (Watson, Clark & Tellegen 1988).

You will need to include both the Pattern Matrix and Structure Matrix in your report, with all loadings showing. To get the full display of loadings, you will need to rerun the analysis that you chose as your final solution (in this case, a two-factor Oblimin rotation), but this time you will need to turn off the option to display only coefficients above .3 (see procedures section). Click on **Options**, and in the **Coefficient Display Format** section remove the tick from the box: **Suppress small coefficients**.

If you are presenting the results of this analysis in your thesis (rather than a journal article), you may also need to provide the screeplot and the table of unrotated loadings (from the **Component Matrix**) in the appendix. This would allow the reader of your thesis to see if they agree with your decision to retain only two components.

Presentation of results in a journal article tends to be much briefer, given space limitations. If you would like to see a published article using factor analysis, go to www.hqlo.com/content/3/1/82 and select the pdf option on the right-hand side of the screen that appears.

Table 1

Pattern and Structure Matrix for PCA with Oblimin Rotation of Two Factor Solution of PANAS Items

Item	Pattern coefficients		Structure coefficients		Communalities
	Component 1	Component 2	Component 1	Component 2	
17. enthusiastic	**.825**	-.012	**.828**	-.241	.686
12. inspired	**.781**	.067	**.763**	-.149	.586
18. alert	**.742**	-.047	**.755**	-.253	.572
13. attentive	**.728**	-.020	**.733**	-.221	.538
15. excited	**.703**	.119	**.710**	-.236	.462
1. interested	**.698**	-.043	**.683**	-.278	.505
9. strong	**.656**	-.097	**.670**	-.076	.475
6. determined	**.635**	.107	**.646**	-.338	.377
7. active	**.599**	-.172	**.605**	-.069	.445
4. proud	**.540**	-.045	**.553**	-.195	.308
19. nervous	.079	**.806**	-.144	**.784**	.620
14. afraid	-.003	**.739**	-.253	**.742**	.548
3. scared	-.010	**.734**	-.207	**.740**	.543
8. distressed	-.052	**.728**	-.213	**.737**	.553
20. jittery	.024	**.718**	-.242	**.717**	.507
2. upset	-.047	**.704**	-.175	**.712**	.516
11. irritable	-.057	**.645**	-.236	**.661**	.440
10. hostile	.080	**.613**	-.176	**.593**	.355
16. guilty	-.013	**.589**	-.090	**.590**	.352
5. ashamed	-.055	**.490**	-.191	**.505**	.258

Note: major loadings for each item are bolded.

For other examples of how to present the results of factor analysis see Chapter 16 in Nicol and Pexman (2010b).

ADDITIONAL EXERCISES