Chapter 17: Exploratory factor analysis

Smart Alex's Solutions

Task 1

Rerun the analysis in this chapter using principal component analysis and compare the results to those in the chapter. (Set the iterations to convergence to 30.)

Running the analysis

Access the main dialog box (Figure 1) by selecting Analyze Dimension Reduction ▸ Factor... . Simply select the variables you want to include in the analysis (remember to exclude any variables that were identified as problematic during the data screening) and transfer them to the box labelled *Variables* by clicking on ➔.

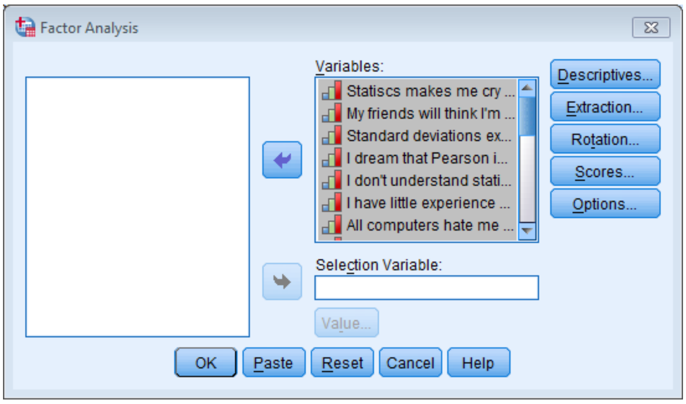


Figure 1: Main dialog box for factor analysis

There are several options available, the first of which can be accessed by clicking on Descriptives... to access the dialog box in Figure 2. The *Univariate descriptives* option provides means and standard deviations for each variable. Most of the other options relate to the correlation matrix of variables (the *R*-matrix). The *Coefficients* option produces the *R*-matrix, and selecting the *Significance levels* option will include the significance value of each correlation in the *R*-matrix. You can also ask for the *Determinant* of this matrix, and this option is useful for testing for multicollinearity or singularity.

KMO and Bartlett's test of sphericity produces the Kaiser–Meyer–Olkin measure of sampling adequacy and Bartlett's test. We have already stumbled across KMO and Bartlett's test and have seen the various criteria for adequacy, but with a sample of 2571 we shouldn't have cause to worry.

The *Reproduced* option produces a correlation matrix based on the model (rather than the real data). Differences between the matrix based on the model and the matrix based on the observed data indicate the residuals of the model. SPSS produces these residuals in the lower table of the reproduced matrix, and we want relatively few of these values to be greater than .05. Luckily, to save us scanning this matrix, SPSS produces a summary of how many residuals lie above .05. The *Reproduced* option should be selected to obtain this summary. The *Anti-image* option produces an anti-image matrix of covariances and correlations. These matrices contain measures of sampling adequacy for each variable along the diagonal and the negatives of the partial correlation/covariances on the off-diagonals. The diagonal elements, like the KMO measure, should all be greater than 0.5 at a bare minimum if the sample is adequate for a given pair of variables. If any pair of variables has a value less than this, consider dropping one of them from the analysis. The off-diagonal elements should all be very small (close to zero) in a good model. When you have finished with this dialog box click on Continue to return to the main dialog box.

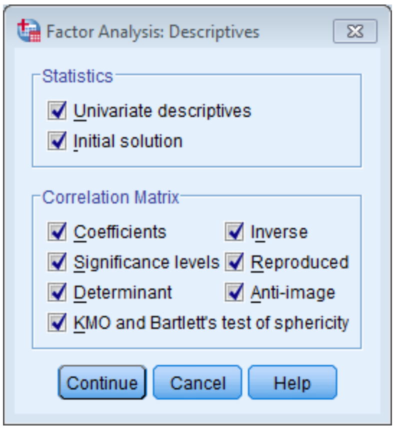


Figure 2: Descriptives in factor analysis

To access the *extraction* dialog box (Figure 3), click on Extraction... in the main dialog box. There are several ways of conducting a factor analysis, and when and where you use the various methods will depend on numerous things. For our purposes we will use *principal component* analysis (Principal components) which, strictly speaking, isn't factor analysis; however, the two procedures may often yield similar results.

In the *Analyze* box there are two options: to analyse the *Correlation matrix* or to analyse the *Covariance matrix*. The *Display* box has two options within it: to display the *Unrotated factor solution* and a *Scree plot*. The scree plot is a useful way of establishing how many factors should be retained in an analysis. The unrotated factor solution is useful in assessing the improvement of interpretation due to rotation. If the rotated solution is little better than the unrotated solution then it is possible that an inappropriate (or less optimal) rotation method has been used.

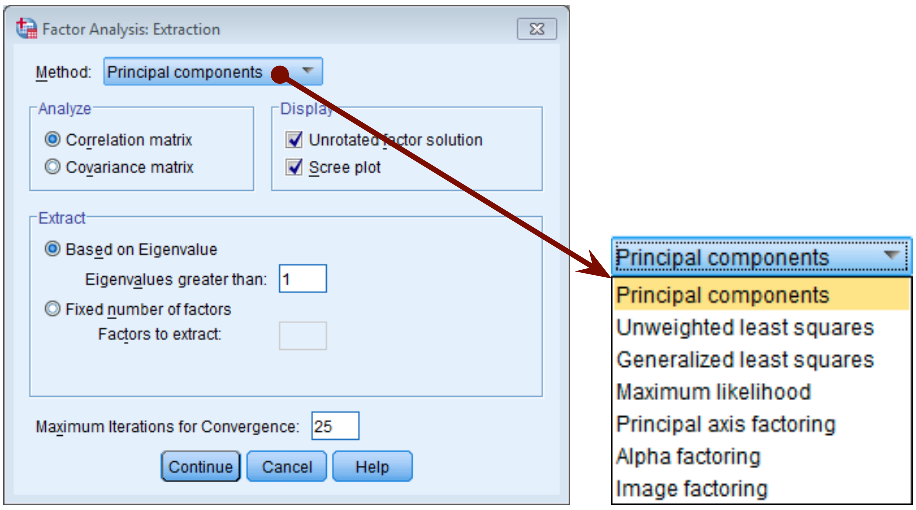
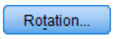


Figure 3: Dialog box for factor extraction

The *Extract* part of the dialog box provides options pertaining to the retention of factors. You have the choice of either selecting factors with eigenvalues greater than a user-specified value or retaining a fixed number of factors. For the *Eigenvalues greater than* option the default is Kaiser's recommendation of eigenvalues over 1. It is probably best to run a primary analysis with the *Eigenvalues greater than* 1 option selected, select a scree plot and compare the results.

Click on 

to access the dialog box in Figure 4. The exact choice of rotation will depend on whether or not you think that the underlying factors should be related. If there are theoretical grounds to think that the factors are independent (unrelated) then you should choose one of the orthogonal rotations (I recommend varimax). However, if theory suggests that your factors might correlate then you should go for one of the oblique rotations (direct oblimin or promax). In this example I've selected varimax.

The dialog box also has options for displaying the *Rotated solution* and a *Loading plot*. The rotated solution is displayed by default and is essential for interpreting the final rotated analysis. The loading plot will provide a graphical display of each variable plotted against the extracted factors up to a maximum of three factors. With two factors these plots are fairly interpretable, and you should hope to see one group of variables clustered close to the *x*-axis and a different group of variables clustered around the *y*-axis. If all variables are clustered between the axes, then the rotation has been relatively unsuccessful in maximizing the loading of a variable onto a single factor. With three factors these plots can become quite messy and certainly put considerable strain on the visual system! However, they can still be a useful way to determine the underlying structures within the data.

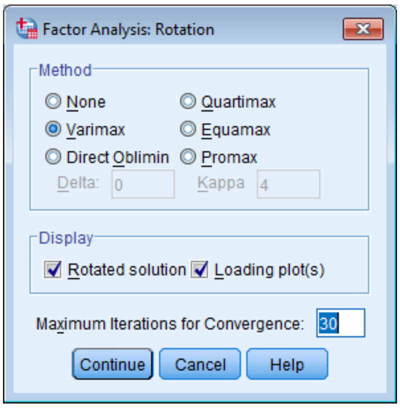
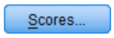


Figure 4: *Factor Analysis: Rotation* dialog box

A final option is to set the *Maximum Iterations for Convergence*, which specifies the number of times that the computer will search for an optimal solution. In most circumstances the default of 25 is more than adequate for SPSS to find a solution for a given data set. However, if you have a large data set (like we have here) then the computer might have difficulty finding a solution (especially for oblique rotation). To allow for the large data set we are using, change the value to 30.

The *factor scores* dialog box (Figure 5) can be accessed by clicking on 

in the main dialog box. This option allows you to save factor scores for each case in the data editor. SPSS creates a new column for each factor extracted and then places the factor score for each case within that column. These scores can then be used for further analysis, or simply to identify groups of participants who score highly on particular factors. There are three methods of obtaining these scores. If you want to ensure that factor scores are uncorrelated then select the *Anderson-Rubin* method; if correlations between factor scores are acceptable then choose the *Regression* method. As a final option, you can ask SPSS to produce the factor score coefficient matrix.

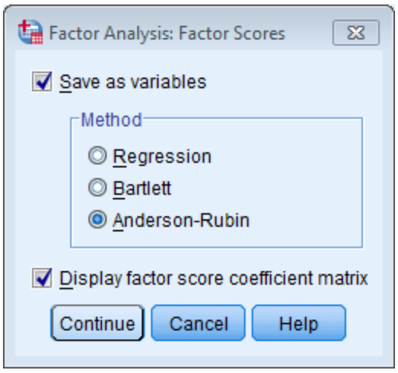


Figure 5: *Factor scores* dialog box

The final two options relate to how coefficients are displayed. By default SPSS will list variables in the order in which they are entered into the data editor. Usually, this format is most convenient. However, when interpreting factors it is sometimes useful to list variables by size. If you select *Sorted by size*, SPSS will order the variables by their factor loadings. In fact, it does this sorting fairly intelligently so that all of the variables that load highly onto the same factor are displayed together. The second option is to *Suppress small coefficients: Absolute value below* a specified value (by default .1). This option ensures that factor loadings within ± .1 are not displayed in the output. Again, this option is useful for assisting in interpretation. The default value is probably sensible, but on your first analysis I recommend changing it either to .4 (for interpretation purposes) or to a value reflecting the expected value of a significant factor loading given the sample size. This will make interpretation simpler. You can, if you like, rerun the analysis and set this value lower just to check you haven't missed anything (like a loading of .39). For this example set the value at .4.

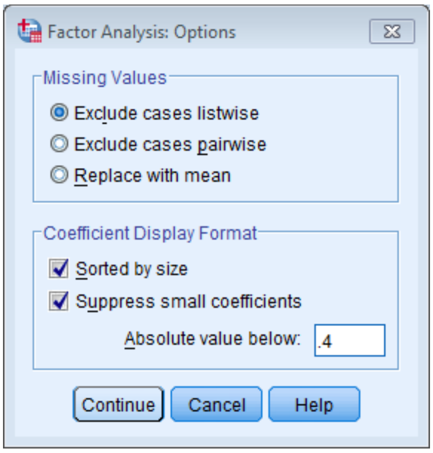


Figure 6: Factor analysis *options* dialog box

Interpreting output from SPSS

Select the same options as I have in the screenshots and run a PCA with orthogonal rotation. Repeat, but using direct oblimin rotation. For the purposes of saving space in this section I set the default SPSS options such that each variable is referred to only by its label on the data editor (e.g., Q12). On the output *you* obtain, you should find that the SPSS uses the value label (the question itself) in all of the output. When using the output in this chapter just remember that Q1 represents question 1, Q2 represents question 2 and Q17 represents question 17.

The first body of output concerns data screening, assumption testing and sampling adequacy. You'll find several large tables (or matrices) that tell us interesting things about our data. If you selected the *Univariate descriptives* option in Figure 2 then the first table will contain descriptive statistics for each variable (the mean, standard deviation and number of cases). This table is not included here, but you should have enough experience to be able to interpret it. The table also includes the number of missing cases; this summary is a useful way to determine the extent of missing data.

The top half of the *R*-matrix (or correlation matrix) shows the Pearson correlation coefficient between all pairs of questions, whereas the bottom half contains the one-tailed significance of these coefficients (Output 1). We can use this correlation matrix to check the pattern of relationships. First, scan the matrix for correlations greater than .3, then look for variables that only have a small number of correlations greater than this value. Then scan the correlation coefficients themselves and look for any greater than .9. If any are found then you should be aware that a problem could arise because of multicollinearity in the data.

In summary, all questions in the SAQ correlate reasonably well with all others and none of the correlation coefficients are excessively large; therefore, we won't eliminate any questions at this stage.

[illegible]

Output 1

Output shows the inverse of the correlation matrix (R^{-1}), which is used in various calculations (including factor scores). This matrix is produced using the *Inverse* option in Figure 2 but in all honesty is useful only if you want some insight into the calculations that go on in a factor analysis. Most of us have more interesting things to do than gain insight into the workings of factor analysis and the practical use of this matrix is minimal—so ignore it!

Output shows several very important parts of the output: the KMO measure of sampling adequacy, Bartlett's test of sphericity and the anti-image correlation and covariance matrices (note that these matrices have been edited down to contain only the first and last five variables). The anti-image correlation and covariance matrices provide similar information (remember the relationship between covariance and correlation) and so only the anti-image correlation matrix need be studied in detail as it is the most informative.

Inverse of Correlation Matrix

	Q01	Q02	Q03	Q04	Q05	Q19	Q20	Q21	Q22	Q23
Q01	1.595	-.028	.087	-.268	-.233	.017	-.024	.011	.002	-.078
Q02	-.028	1.232	-.224	-.057	.013	-.037	.076	.062	-.148	-.003
Q03	.087	-.224	1.661	.138	.057	-.175	.118	.122	-.009	-.103
Q04	-.268	-.057	.138	1.626	-.203	-.049	-.006	-.149	-.045	-.023
Q05	-.233	.013	.057	-.203	1.410	-.024	-.016	-.074	.045	-.006
Q06	.034	-.078	-.072	-.011	-.055	-.023	.080	.069	.058	.025
Q07	.039	.025	.127	-.152	-.072	.105	.077	-.386	.019	-.012
Q08	-.087	-.051	-.013	-.134	-.045	.074	.034	-.039	-.035	.003
Q09	-.023	-.242	-.208	.043	-.027	-.141	.050	-.047	-.156	-.110
Q10	-.017	-.015	-.023	.009	-.124	-.012	.056	.026	.023	.017
Q11	-.075	.061	.121	-.041	.000	-.010	-.140	-.009	.055	.015
Q12	-.011	.046	.147	-.259	-.091	.060	-.100	-.141	.026	-.038
Q13	-.145	-.011	-.055	.040	.007	.014	.028	-.061	.077	-.042
Q14	-.064	.033	.115	-.007	-.040	.063	.002	-.110	.041	-.034
Q15	.138	.050	.013	-.098	.021	.013	-.054	.058	.034	-.030
Q16	-.454	-.017	.142	-.063	-.155	.071	-.008	-.158	-.005	.033
Q17	-.084	-.045	.063	-.064	-.030	-.074	.025	-.077	.015	.080
Q18	-.041	.028	.070	-.044	.004	.047	-.004	-.136	-.037	.033
Q19	.017	-.037	-.175	-.049	-.024	1.264	.120	.048	-.141	-.045
Q20	-.024	.076	.118	-.006	-.016	.120	1.370	-.511	-.014	-.034
Q21	.011	.062	.122	-.149	-.074	.048	-.511	1.830	-.036	.018
Q22	.002	-.148	-.009	-.045	.045	-.141	-.014	-.036	1.200	-.202
Q23	-.078	-.003	-.103	-.023	-.006	-.045	-.034	.018	-.202	1.094

Output 2

For the KMO statistic Kaiser (1974) recommends a bare minimum of .5 and that values between .5 and .7 are mediocre, values between .7 and .8 are good, values between .8 and .9 are great and values above .9 are superb (Hutcheson & Sofroniou, 1999). For these data the value is .93, which falls into the range of being superb, so we should be confident that the sample size is adequate for factor analysis.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.930
Bartlett's Test of Sphericity	Approx. Chi-Square	19334.492
	df	253
	Sig.	.000

Anti-image Matrices

Statistics=Anti-image Correlation

Variables	Question_01	Question_02	Question_03	Question_04	Question_05	Question_19	Question_20	Question_21	Question_22	Question_23
Question_01	.930[a]	-.020	.053	-.167	-.156	.012	-.016	.006	.001	-.059
Question_02	-.020	.875[a]	-.157	-.041	.010	-.029	.059	.041	-.121	-.002
Question_03	.053	-.157	.951[a]	.084	.037	-.121	.078	.070	-.007	-.076
Question_04	-.167	-.041	.084	.955[a]	-.134	-.034	-.004	-.086	-.033	-.017
Question_05	-.156	.010	.037	-.134	.960[a]	-.018	-.011	-.046	.035	-.005
Question_06	.020	-.053	-.042	-.007	-.035	-.015	.051	.039	.040	.018
Question_07	.023	.016	.072	-.087	-.044	.068	.048	-.208	.013	-.008
Question_08	-.049	-.033	-.007	-.075	-.027	.047	.021	-.020	-.023	.002
Question_09	-.016	-.193	-.142	.030	-.020	-.111	.038	-.031	-.126	-.092
Question_10	-.012	-.012	-.016	.006	-.093	-.009	.043	.017	.019	.015
Question_11	-.041	.038	.064	-.022	-3.269E-5	-.006	-.082	-.005	.034	.010
Question_12	-.007	.031	.087	-.154	-.058	.040	-.065	-.079	.018	-.028
Question_13	-.085	-.008	-.032	.023	.004	.009	.018	-.033	.052	-.030
Question_14	-.040	.023	.069	-.004	-.026	.044	.001	-.063	.029	-.026
Question_15	.089	.037	.008	-.062	.014	.009	-.037	.035	.025	-.024
Question_16	-.264	-.011	.081	-.036	-.096	.047	-.005	-.085	-.003	.023
Question_17	-.047	-.029	.035	-.035	-.018	-.047	.015	-.041	.010	.055
Question_18	-.023	.018	.039	-.025	.002	.030	-.003	-.072	-.024	.023
Question_19	.012	-.029	-.121	-.034	-.018	.941[a]	.091	.031	-.115	-.038
Question_20	-.016	.059	.078	-.004	-.011	.091	.889[a]	-.323	-.011	-.028
Question_21	.006	.041	.070	-.086	-.046	.031	-.323	.929[a]	-.024	.013
Question_22	.001	-.121	-.007	-.033	.035	-.115	-.011	-.024	.878[a]	-.176
Question_23	-.059	-.002	-.076	-.017	-.005	-.038	-.028	.013	-.176	.766[a]

a. Measures of Sampling Adequacy(MSA)

Output 3

I mentioned that KMO can be calculated for multiple and individual variables. The KMO values for individual variables are produced on the diagonal of the anti-image correlation matrix (I have highlighted these cells). These values make the anti-image correlation matrix an extremely important part of the output (although the anti-image covariance matrix can be ignored). As well as checking the overall KMO statistic, it is important to examine the diagonal elements of the anti-image correlation matrix: the value should be above the bare minimum of .5 for all variables (and preferably higher). For these data all values are well above .5, which is good news! If you find any variables with values below .5 then you should consider excluding them from the analysis (or run the analysis with and without them and note the difference). Removal of a variable affects the KMO statistics, so if you do remove a variable be sure to re-examine the new anti-image correlation matrix. As for the rest of the anti-image correlation matrix, the off-diagonal elements represent the partial correlations between variables. For a good factor analysis we want these correlations to be very small (the smaller, the better). So, as a final check, you can just look through to see that the off-diagonal elements are small (they should be for these data).

Bartlett's measure tests the null hypothesis that the original correlation matrix is an identity matrix. A significant test tells us that the *R*-matrix is not an identity matrix; therefore, there are some relationships between the variables we hope to include in the analysis. For these data, Bartlett's test is highly significant ($p < .001$); it usually is.

The first part of the factor extraction process is to determine the linear components within the data set (the eigenvectors) by calculating the eigenvalues of the *R*-matrix. We know that there are as many components (eigenvectors) in the *R*-matrix as there are variables, but most will be unimportant. To determine the importance of a particular vector we look at the magnitude of the associated eigenvalue. We can then apply criteria to determine which factors to retain and which to discard. By default SPSS uses Kaiser's criterion of retaining factors with eigenvalues greater than 1 (see Figure 3).

Output lists the eigenvalues associated with each linear component (factor) before extraction, after extraction and after rotation. Before extraction, SPSS has identified 23 linear components within the data set (we know that there should be as many eigenvectors as there are variables and so there will be as many factors as variables). The eigenvalues associated with each factor represent the variance explained by that particular linear component, and SPSS also displays the eigenvalue in terms of the percentage of variance explained (so factor 1 explains 31.696% of total variance). It should be clear that the first few factors explain relatively large amounts of variance (especially factor 1), whereas subsequent factors explain only small amounts of variance. SPSS then extracts all factors with eigenvalues greater than 1, which leaves us with four factors. The eigenvalues associated with these factors are again displayed (and the percentage of variance explained) in the columns labelled *Extraction Sums of Squared Loadings*. The values in this part of the table are the same as the values before extraction, except that the values for the discarded factors are ignored (hence, the table is blank after the fourth factor). In the final part of the table (labelled *Rotation Sums of Squared Loadings*), the eigenvalues of the factors after rotation are displayed. Rotation has the effect of optimizing the factor structure, and one consequence for these data is that the relative importance of the four factors is equalized.

Before rotation, factor 1 accounted for considerably more variance than the remaining three (31.696% compared to 7.560, 5.725 and 5.336%), but after extraction it accounts for only 16.219% of variance (compared to 14.523, 11.099 and 8.475%, respectively).

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	7.290	31.696	31.696	7.290	31.696	31.696	3.730	16.219	16.219
2	1.739	7.560	39.256	1.739	7.560	39.256	3.340	14.523	30.742
3	1.317	5.725	44.981	1.317	5.725	44.981	2.553	11.099	41.842
4	1.227	5.336	50.317	1.227	5.336	50.317	1.949	8.475	50.317
5	.988	4.295	54.612						
6	.895	3.893	58.504						
7	.806	3.502	62.007						
8	.783	3.404	65.410						
9	.751	3.265	68.676						
10	.717	3.117	71.793						
11	.684	2.972	74.765						
12	.670	2.911	77.676						
13	.612	2.661	80.337						
14	.578	2.512	82.849						
15	.549	2.388	85.236						
16	.523	2.275	87.511						
17	.508	2.210	89.721						
18	.456	1.982	91.704						
19	.424	1.843	93.546						
20	.408	1.773	95.319						
21	.379	1.650	96.969						
22	.364	1.583	98.552						
23	.333	1.448	100.000						

Extraction Method: Principal Component Analysis.

Output 4

Output shows the table of communalities before and after extraction. Remember that the communality is the proportion of common variance within a variable. Principal component analysis works on the initial assumption that all variance is common; therefore, before extraction the communalities are all 1 (see the column labelled *Initial*). In effect, all of the variance associated with a variable is assumed to be common variance. Once factors have been extracted, we have a better idea of how much variance is, in reality, common. The communalities in the column labelled *Extraction* reflect this common variance. So, for example, we can say that 43.5% of the variance associated with question 1 is common, or shared, variance. Another way to look at these communalities is in terms of the proportion of variance explained by the underlying factors. Before extraction, there are as many factors as there are variables, so all variance is explained by the factors and communalities are all 1. However, after extraction some of the factors are discarded and so some information is lost. The retained factors cannot explain all of the variance present in the data, but they can explain some. The amount of variance in each variable that can be explained by the retained factors is represented by the communalities after extraction.

Communalities

	Initial	Extraction
Q01	1.000	.435
Q02	1.000	.414
Q03	1.000	.530
Q04	1.000	.469
Q05	1.000	.343
Q06	1.000	.654
Q07	1.000	.545
Q08	1.000	.739
Q09	1.000	.484
Q10	1.000	.335
Q11	1.000	.690
Q12	1.000	.513
Q13	1.000	.536
Q14	1.000	.488
Q15	1.000	.378
Q16	1.000	.487
Q17	1.000	.683
Q18	1.000	.597
Q19	1.000	.343
Q20	1.000	.484
Q21	1.000	.550
Q22	1.000	.464
Q23	1.000	.412

Extraction Method: Principal Component

Component Matrix[a]

	Component			
	1	2	3	4
Q18	.701			
Q07	.685			
Q16	.679			
Q13	.673			
Q12	.669			
Q21	.658			
Q14	.656			
Q11	.652			-.400
Q17	.643			
Q04	.634			
Q03	-.629			
Q15	.593			
Q01	.586			
Q05	.556			
Q08	.549	.401		-.417
Q10	.437			
Q20	.436		-.404	
Q19	-.427			
Q09		.627		
Q02		.548		
Q22		.465		
Q06	.562		.571	
Q23				.507

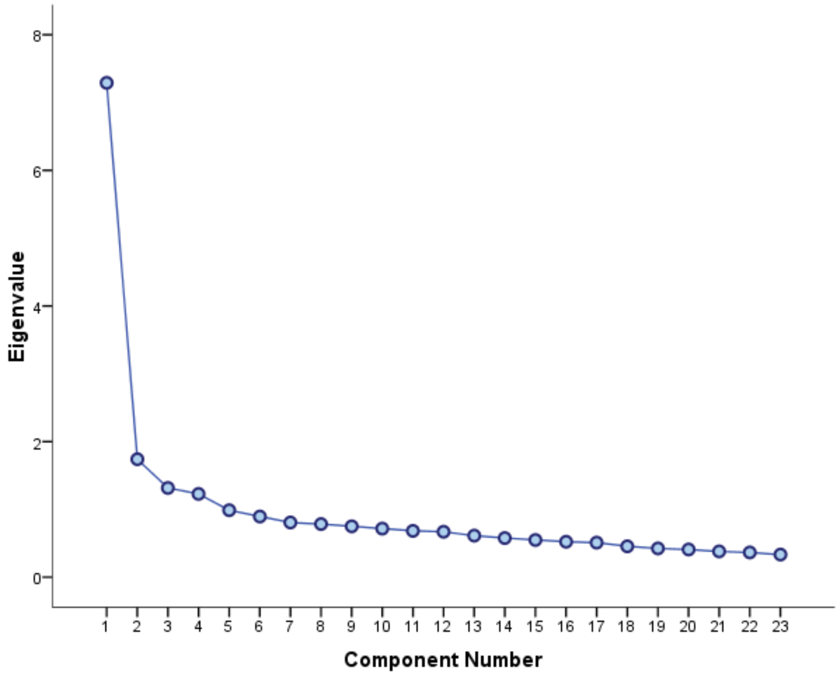
Extraction Method: Principal Component Analysis.

a. 4 components extracted.

Output 5

Output also shows the component matrix before rotation. This matrix contains the loadings of each variable onto each factor. By default SPSS displays all loadings; however, we requested that all loadings less than .4 be suppressed in the output (see Figure 6) and so there are blank spaces for many of the loadings. This matrix is not particularly important for interpretation, but it is interesting to note that before rotation most variables load highly onto the first factor.

At this stage SPSS has extracted four factors. Factor analysis is an exploratory tool and so it should be used to guide the researcher to make various decisions: you shouldn't leave the computer to make them. One important decision is the number of factors to extract. By Kaiser's criterion we should extract four factors and this is what SPSS has done. However, this criterion is accurate when there are less than 30 variables and communalities after extraction are greater than .7 or when the sample size exceeds 250 and the average communality is greater than .6. The communalities are shown in Output , and only one exceeds .7. The average of the communalities can be found by adding them up and dividing by the number of communalities (11.573/23 = .503). So, on both grounds Kaiser's rule may not be accurate. However, you should consider the huge sample that we have, because the research into Kaiser's criterion gives recommendations for much smaller samples. By Jolliffe's criterion (retain factors with eigenvalues greater than 0.7) we should retain 10 factors, but there is little to recommend this criterion over Kaiser's. As a final guide we can use the scree plot which we asked SPSS to produce by using the option in Figure 3. The scree plot is shown in Output . This curve is difficult to interpret because it begins to tail off after three factors, but there is another drop after four factors before a stable plateau is reached. Therefore, we could probably justify retaining either two or four factors. Given the large sample, it is probably safe to assume Kaiser's criterion; however, you might like to rerun the analysis specifying that SPSS extract only two factors (see Figure 3) and compare the results.



Output 8

Output shows an edited version of the reproduced correlation matrix. The top half of this matrix (labelled **Reproduced Correlations**) contains the correlation coefficients between all of the questions based on the factor model. The diagonal of this matrix contains the communalities after extraction for each variable.

Reproduced Correlations

		Question_01	Question_02	Question_03	Question_04	Question_05	Question_19	Question_20	Question_21	Question_22	Question_23
Reproduced Correlation	Question_01	.435[a]	-.112	-.372	.447	.376	-.204	.342	.449	-.025	.045
	Question_02	-.112	.414[a]	.380	-.134	-.122	.357	-.301	-.254	.333	.246
	Question_03	-.372	.380	.530[a]	-.399	-.345	.403	-.440	-.488	.275	.158
	Question_04	.447	-.134	-.399	.469[a]	.399	-.231	.353	.480	-.050	.042
	Question_05	.376	-.122	-.345	.399	.343[a]	-.207	.292	.412	-.060	.028
	Question_06	.218	-.033	-.200	.278	.273	-.147	-.021	.244	-.209	-.082
	Question_07	.366	-.148	-.373	.419	.380	-.254	.219	.430	-.179	-.037
	Question_08	.412	.002	-.270	.390	.312	-.104	.164	.282	-.099	-.136
	Question_09	-.042	.430	.352	-.073	-.080	.363	-.218	-.191	.417	.323
	Question_10	.172	-.061	-.181	.212	.205	-.137	.006	.188	-.197	-.110
	Question_11	.423	-.097	-.357	.419	.348	-.198	.200	.342	-.209	-.202
	Question_12	.402	-.219	-.440	.448	.397	-.302	.354	.503	-.136	.011
	Question_13	.347	-.122	-.342	.395	.360	-.231	.163	.384	-.203	-.079
	Question_14	.362	-.155	-.373	.411	.370	-.254	.241	.431	-.159	-.022
	Question_15	.311	-.158	-.337	.343	.306	-.236	.175	.336	-.230	-.141
	Question_16	.440	-.217	-.458	.466	.400	-.299	.373	.494	-.152	-.049
	Question_17	.439	-.048	-.331	.434	.359	-.162	.196	.347	-.145	-.140
	Question_18	.368	-.149	-.376	.424	.388	-.259	.215	.439	-.183	-.032
	Question_19	-.204	.357	.403	-.231	-.207	.343[a]	-.308	-.324	.294	.196
	Question_20	.342	-.301	-.440	.353	.292	-.308	.484[a]	.457	-.068	.021
	Question_21	.449	-.254	-.488	.480	.412	-.324	.457	.550[a]	-.096	.032
	Question_22	-.025	.333	.275	-.050	-.060	.294	-.068	-.096	.464[a]	.408
	Question_23	.045	.246	.158	.042	.028	.196	.021	.032	.408	.412[a]
Residual[b]	Question_01		.013	.035	-.011	.027	.015	-.128	-.120	-.079	-.049
	Question_02	.013		-.062	.022	.003	-.153	.099	.049	-.102	-.147
	Question_03	.035	-.062		.019	.035	-.061	.115	.071	-.071	-.008
	Question_04	-.011	.022	.019		.002	.045	-.110	-.070	-.049	-.076
	Question_05	.027	.003	.035	.002		.041	-.092	-.078	-.072	-.070
	Question_06	.000	-.041	-.027	.000	-.016	-.020	.122	.029	.043	.013
	Question_07	-.061	-.011	-.009	-.010	-.041	-.015	.002	.053	.010	-.033
	Question_08	-.081	-.052	.011	-.041	-.044	-.056	.011	.014	.020	.086
	Question_09	-.050	-.115	-.052	-.051	-.016	-.114	.060	.055	-.161	-.152
	Question_10	.042	-.023	-.013	.003	.053	.010	.078	.005	.066	.048
	Question_11	-.066	-.046	.006	-.051	-.050	-.002	.056	.005	.047	.116
	Question_12	-.057	.024	.030	-.006	-.050	.036	-.056	-.062	-.031	-.057
	Question_13	.008	-.021	.024	-.051	-.058	.004	.041	-.010	.007	.026
	Question_14	-.024	-.009	.002	-.060	-.055	.000	-.015	-.032	-.011	-.027
	Question_15	-.065	-.007	.025	-.009	-.045	.027	.031	-.036	.062	.079
	Question_16	.059	.050	.039	-.050	-.005	.032	-.108	-.074	-.004	-.032
	Question_17	-.069	-.039	.003	-.052	-.049	-.001	.009	.016	.019	.049
	Question_18	-.020	-.015	.001	-.042	-.066	.003	.020	-.009	.023	-.049
	Question_19	.015	-.153	-.061	.045	.041		.060	.049	-.060	-.073
	Question_20	-.128	.099	.115	-.110	-.092	.060		.010	-.032	-.056
	Question_21	-.120	.049	.071	-.070	-.078	.049	.010		-.033	-.100
	Question_22	-.079	-.102	-.071	-.049	-.072	-.060	-.032	-.033		-.177
	Question_23	-.049	-.147	-.008	-.076	-.070	-.073	-.056	-.100	-.177	

Extraction Method: Principal Component Analysis.

a. Reproduced communalities

b. Residuals are computed between observed and reproduced correlations. There are 91 (35.0%) nonredundant residuals with absolute values greater than 0.05.

Output 9

The correlations in the reproduced matrix differ from those in the *R*-matrix because they stem from the model rather than the observed data. If the model were a perfect fit of the data then we would expect the reproduced correlation coefficients to be the same as the original correlation coefficients. Therefore, to assess the fit of the model we can look at the differences between the observed correlations and the correlations based on the model. For example, if we take the correlation between questions 1 and 2, the correlation based on the observed data is −.099. The correlation based on the model is −.112, which is slightly higher. We can calculate the difference as follows:

$$\begin{aligned}\text{residual} &= r_{\text{observed}} - r_{\text{from model}} \\ \text{residual}_{Q_1 Q_2} &= (-0.099) - (-0.112) \\ &= 0.013\end{aligned}$$

You should notice that this difference is the value quoted in the lower half of the reproduced matrix (labelled *Residual*) for questions 1 and 2. Therefore, the lower half of the reproduced matrix contains the differences between the observed correlation coefficients and the ones predicted from the model. For a good model these values will all be small. In fact, we want most values to be less than 0.05. Rather than scan this huge matrix, SPSS provides a footnote summary, which states how many residuals have an absolute value greater than 0.05. For these data there are 91 residuals (35%) that are greater than 0.05. There are no hard and fast rules about what proportion of residuals should be below 0.05; however, if more than 50% are greater than 0.05 you probably have grounds for concern.

Orthogonal rotation (varimax)

Output shows the rotated component matrix (also called the rotated factor matrix in factor analysis), which is a matrix of the factor loadings for each variable onto each factor. This matrix contains the same information as the component matrix, except that it is calculated *after* rotation. There are several things to consider about the format of this matrix. First, factor loadings less than .4 have not been displayed because we asked for these loadings to be suppressed using the option in Figure 6. If you didn't select this option, or didn't adjust the criterion value to .4, then your output will differ. Second, the variables are listed in the order of size of their factor loadings. By default, SPSS orders the variables as they are in the data editor; however, we asked for the output to be *Sorted by size* using the option in Figure 6. If this option was not selected your output will look different. Finally, for all other parts of the output I suppressed the variable labels (for reasons of space), but for this matrix I have allowed the variable labels to be printed to aid interpretation.

The original logic behind suppressing loadings less than .4 was based on Stevens' (2002) suggestion that this cut-off point was appropriate for interpretative purposes (i.e., loadings greater than .4 represent substantive values). However, this means that we have suppressed several loadings that are undoubtedly significant. However, significance itself is not important.

Compare this matrix to the unrotated solution (Output). Before rotation, most variables loaded highly onto the first factor and the remaining factors didn't really get a look in. However, the rotation of the factor structure has clarified things considerably: there are four factors and variables load very highly onto only one factor (with the exception of one question). The suppression of loadings less than .4 and ordering variables by loading size also make interpretation considerably easier (because you don't have to scan the matrix to identify substantive loadings).

The next step is to look at the content of questions that load onto the same factor to try to identify common themes. If the mathematical factor produced by the analysis represents some real-world construct then common themes among highly loading questions can help us identify what the construct might be. The questions that load highly on factor 1 seem to all relate to using computers or SPSS. Therefore we might label this factor *fear of computers*. The questions that load highly on factor 2 all seem to relate to different aspects of statistics; therefore, we might label this factor *fear of statistics*. The three questions that load highly on factor 3 all seem to relate to mathematics; therefore, we might label this factor *fear of mathematics*. Finally, the questions that load highly on factor 4 all contain some component of social evaluation from friends; therefore, we might label this factor *peer evaluation*. This analysis seems to reveal that the initial questionnaire, in reality, is composed of four subscales: fear of computers, fear of statistics, fear of maths and fear of negative peer evaluation. There are two possibilities here. The first is that the SAQ failed to measure what it set out to (namely, SPSS anxiety) but does measure some related constructs. The second is that these four constructs are sub-components of SPSS anxiety. However, the factor analysis does not indicate which of these possibilities is true.

Rotated Component Matrix[a]

	Component			
	1	2	3	4
I have little experience of computers	.800			
SPSS always crashes when I try to use it	.684			
I worry that I will cause irreparable damage because of my incompetenece with computers	.647			
All computers hate me	.638			
Computers have minds of their own and deliberately go wrong whenever I use them	.579			
Computers are useful only for playing games	.550			
Computers are out to get me	.459			
I can't sleep for thoughts of eigen vectors		.677		
I wake up under my duvet thinking that I am trapped under a normal distribtion		.661		
Standard deviations excite me		-.567		
People try to tell you that SPSS makes statistics easier to understand but it doesn't	.473	.523		
I dream that Pearson is attacking me with correlation coefficients		.516		
I weep openly at the mention of central tendency		.514		
Statiscs makes me cry		.496		
I don't understand statistics		.429		
I have never been good at mathematics			.833	
I slip into a coma whenever I see an equation			.747	
I did badly at mathematics at school			.747	
My friends are better at statistics than me				.648
My friends are better at SPSS than I am				.645
If I'm good at statistics my friends will think I'm a nerd				.586
My friends will think I'm stupid for not being able to cope with SPSS				.543
Everybody looks at me when I use SPSS				.427

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 9 iterations.

Component Transformation Matrix

Component	1	2	3	4
1	.635	.585	.443	-.242
2	.137	-.168	.488	.846
3	.758	-.513	-.403	.008
4	.067	.605	-.635	.476

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

Output 8

The final part of the output is the factor transformation matrix. This matrix provides information about the degree to which the factors were rotated to obtain a solution. If no rotation were necessary, this matrix would be an identity matrix. If orthogonal rotation were completely appropriate then we would expect a symmetrical matrix (same values above and below the diagonal). In reality the matrix is not easy to interpret, although very asymmetrical matrices might be taken as a reason to try oblique rotation. For the inexperienced factor analyst you are probably best advised to ignore the factor transformation matrix.

Oblique rotation

Pattern Matrix[a]

	Component			
	1	2	3	4
I can't sleep for thoughts of eigen vectors	.706			
I wake up under my duvet thinking that I am trapped under a normal distribtion	.591			
Standard deviations excite me	-.511			
I dream that Pearson is attacking me with correlation coefficients	.405			
I weep openly at the mention of central tendency	.400			
Statiscs makes me cry				
I don't understand statistics				
My friends are better at SPSS than I am		.643		
My friends are better at statistics than me		.621		
If I'm good at statistics my friends will think I'm a nerd		.615		
My friends will think I'm stupid for not being able to cope with SPSS		.507		
Everybody looks at me when I use SPSS				
I have little experience of computers			.885	
SPSS always crashes when I try to use it			.713	
All computers hate me			.653	
I worry that I will cause irreparable damage because of my incompetenece with computers			.650	
Computers have minds of their own and deliberately go wrong whenever I use them			.588	
Computers are useful only for playing games			.585	
People try to tell you that SPSS makes statistics easier to understand but it doesn't	.412		.462	
Computers are out to get me			.411	
I have never been good at mathematics				-.902
I slip into a coma whenever I see an equation				-.774
I did badly at mathematics at school				-.774

Extraction Method: Principal Component Analysis.
Rotation Method: Oblimin with Kaiser Normalization.
a. Rotation converged in 29 iterations.

Output 9

For the pattern matrix for these data (Output) the same four factors seem to have emerged (although for some variables the factor loadings are too small to be displayed). Factor 1 seems to represent fear of statistics, factor 2 represents fear of peer evaluation, factor 3 represents fear of computers and factor 4 represents fear of mathematics. The structure matrix (Output) differs in that shared variance is not ignored. The picture becomes more complicated because, with the exception of factor 2, several variables load highly onto more than one factor. This has occurred because of the relationship between factors 1 and 3 and factors 3 and 4. This example should highlight why the pattern matrix is preferable for interpretative reasons: because it contains information about the *unique* contribution of a variable to a factor.

Structure Matrix

	Component 1	2	3	4
I wake up under my duvet thinking that I am trapped under a normal distribtion	.695		.477	
I can't sleep for thoughts of eigen vectors	.685			
Standard deviations excite me	-.632		-.407	
I weep openly at the mention of central tendency	.567		.516	-.491
I dream that Pearson is attacking me with correlation coefficients	.548		.487	-.485
Statiscs makes me cry	.520		.413	-.501
I don't understand statistics	.462		.453	
My friends are better at SPSS than I am		.660		
My friends are better at statistics than me		.653		
If I'm good at statistics my friends will think I'm a nerd		.588		
My friends will think I'm stupid for not being able to cope with SPSS		.546		
Everybody looks at me when I use SPSS	-.435	.446		
I have little experience of computers			.777	
SPSS always crashes when I try to use it	.404		.761	
All computers hate me	.401		.723	
I worry that I will cause irreparable damage because of my incompetenece with computers			.723	-.429
Computers have minds of their own and deliberately go wrong whenever I use them	.426		.671	
People try to tell you that SPSS makes statistics easier to understand but it doesn't	.576		.606	
Computers are out to get me			.561	-.441
Computers are useful only for playing games			.556	
I have never been good at mathematics				-.855
I slip into a coma whenever I see an equation			.453	-.822
I did badly at mathematics at school			.451	-.818

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

Output 1

The final part of the output is a correlation matrix between the factors (Output). This matrix contains the correlation coefficients between factors. As predicted from the structure matrix, factor 2 has little or no relationship with any other factors (correlation coefficients are low), but all other factors are interrelated to some degree (notably factors 1 and 3 and factors 3 and 4). The fact that these correlations exist tell us that the constructs measured can be interrelated. If the constructs were independent then we would expect oblique rotation to provide an identical solution to an orthogonal rotation and the component correlation matrix should be an identity matrix (i.e., all factors have correlation coefficients of 0). Therefore, this final matrix gives us a guide to whether it is reasonable to assume independence between factors: for these data it appears that we cannot assume independence. Therefore, the results of the orthogonal rotation should not be trusted: the obliquely rotated solution is probably more meaningful.

On a theoretical level the dependence between our factors does not cause concern; we might expect a fairly strong relationship between fear of maths, fear of statistics and fear of computers. Generally, the less mathematically and technically minded people struggle with statistics. However, we would not expect these constructs to correlate with fear of peer evaluation (because this construct is more socially based). In fact, this factor is the one that correlates fairly badly with all others – so on a theoretical level, things have turned out rather well!

Component Correlation Matrix

Component	1	2	3	4
1	1.000	-.154	.364	-.279
2	-.154	1.000	-.185	8.155E-02
3	.364	-.185	1.000	-.464
4	-.279	8.155E-02	-.464	1.000

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

Output 2

Factor scores

Having reached a suitable solution and rotated that solution, we can look at the factor scores. Output shows the component score matrix *B* from which the factor scores are calculated and the covariance matrix of factor scores. The component score matrix is not particularly useful in itself. It can be useful in understanding how the factor scores have been computed, but with large data sets like this one you are unlikely to want to delve into the mathematics behind the factor scores. However, the covariance matrix of scores is useful. This matrix in effect tells us the relationship between factor scores (it is an unstandardized correlation matrix). If factor scores are uncorrelated then this matrix should be an identity matrix (i.e., diagonal elements will be 1 but all other elements are 0). For these data the covariances are all zero, indicating that the resulting scores are uncorrelated.

Component Score Coefficient Matrix

	Component			
	1	2	3	4
Question_01	-.053	.173	.089	.110
Question_02	.102	-.129	.086	.282
Question_03	.087	-.195	.013	.137
Question_04	-.011	.170	.045	.107
Question_05	.021	.131	.014	.083
Question_06	.383	-.211	-.088	.014
Question_07	.213	.004	-.078	.038
Question_08	-.129	-.074	.460	.013
Question_09	.025	-.029	.108	.354
Question_10	.244	-.161	-.021	-.036
Question_11	-.066	-.087	.379	-.059
Question_12	.097	.161	-.116	.051
Question_13	.224	-.065	-.019	.013
Question_14	.180	.040	-.084	.043
Question_15	.114	-.055	.061	-.058
Question_16	-.015	.146	.046	.014
Question_17	-.057	-.067	.372	.005
Question_18	.242	-.001	-.104	.043
Question_19	.048	-.115	.061	.199
Question_20	-.195	.359	-.061	-.002
Question_21	-.039	.270	-.064	.059
Question_22	-.036	.162	-.048	.382
Question_23	.032	.211	-.162	.379

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

Component Score Covariance Matrix

Component	1	2	3	4
1	1.000	.000	.000	.000
2	.000	1.000	.000	.000
3	.000	.000	1.000	.000
4	.000	.000	.000	1.000

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

Output 3

In the original analysis we asked for scores to be calculated based on the Anderson–Rubin method (hence why they are uncorrelated). You will find these scores in the data editor. There should be four new columns of data (one for each factor) labelled *FAC1_1*, *FAC2_1*, *FAC3_1* and *FAC4_1* respectively. If you asked for factor scores in the oblique rotation then these scores will appear in the data editor in four other columns labelled *FAC2_1* and so on.

Output shows the factor scores for the first 10 participants. It should be pretty clear that participant 9 scored highly on all four factors, and so this person is very anxious about statistics, computing and maths, but less so about peer evaluation (factor 4). Factor scores can be used in this way to assess the relative fear of one person compared to another, or we could add the scores up to obtain a single score for each participant (that we might assume represents SPSS anxiety as a whole). We can also use factor scores in regression when groups of predictors correlate so highly that there is multicollinearity.

Case Summaries[a]

		FAC1_1	FAC2_1	FAC3_1	FAC4_1
1		.10587	-.92816	-1.82760	-.45953
2		-.58279	-.18932	-.04137	.29332
3		-.54761	.02953	.19918	-.97142
4		.74661	.72268	-.68654	-.18547
5		.25169	-.51489	-.63354	.68301
6		1.91614	-.27356	-.68201	-.52709
7		-.26051	-1.40947	-.00431	.91008
8		-.28572	-.91761	-.08947	1.03746
9		1.71750	1.15102	3.15659	.81082
10		-.69151	-.73318	.18376	1.49541
Total	N	10	10	10	10

a. Limited to first 10 cases.

Output 13

Summary

To sum up, the analyses revealed four underlying scales in our questionnaire that may, or may not, relate to genuine sub-components of SPSS anxiety. It also seems as though an obliquely rotated solution was preferred due to the interrelationships between factors. The use of factor analysis is purely exploratory; it should be used only to guide future hypotheses, or to inform researchers about patterns within data sets. A great many decisions are left to the researcher using factor analysis, and I urge you to make informed decisions, rather than basing decisions on the outcomes you would like to get. The next question is whether or not our scale is reliable.

Task 2

*The University of Sussex constantly seeks to employ the best people possible as lecturers. They wanted to revise the 'Teaching of Statistics for Scientific Experiments' (TOSSE) questionnaire, which is based on Bland's theory that says that good research methods lecturers should have: (1) a profound love of statistics; (2) an enthusiasm for experimental design; (3) a love of teaching; and (4) a complete absence of normal interpersonal skills. These characteristics should be related (i.e., correlated). The university revised this questionnaire to become the 'Teaching of Statistics for Scientific Experiments – Revised' (TOSSE-R). They gave this questionnaire to 239 research methods lecturers around the world to see if it supported Bland's theory. The data are in **TOSSE-R.sav**. Conduct principal axis functioning analysis (with appropriate rotation) and interpret the factor structure.*

Teaching of Statistics for Scientific Experiments — Revised (TOSSE-R)

		SD	D	N	A	SA
1.	I once woke up in a vegetable patch hugging a turnip that I'd mistakenly dug up thinking it was Roy's largest root					
2.	If I had a big gun I'd shoot all the students I have to teach					
3.	I memorise probability values for the *F*-distribution					
4.	I worship at the shrine of Pearson					
5.	I still live with my mother and have little personal hygiene					
6.	Teaching others makes me want to swallow a large bottle of bleach because the pain of my burning oesophagus would be light relief in comparison					
7.	Helping others to understand sums of squares is a great feeling					
8.	I like control conditions					
9.	I calculate 3 ANOVAs in my head before getting out of bed					
10.	I could spend all day explaining statistics to people					
11.	I like it when I've helped people to understand factor rotation					
12.	People fall asleep as soon as I open my mouth to speak					
13.	Designing experiments is fun					
14.	I'd rather think about appropriate dependent variables than go to the pub					
15.	I soil my pants with excitement at the mere mention of Factor Analysis					
16.	Thinking about whether to use repeated- or independent-measures thrills me					
17.	I enjoy sitting in the park contemplating whether to use participant observation in my next experiment					
18.	Standing in front of 300 people in no way makes me lose control of my bowels					
19.	I like to help students					
20.	Passing on knowledge is the greatest gift you can bestow an individual					
21.	Thinking about Bonferroni corrections gives me a tingly feeling in my groin					
22.	I quiver with excitement when thinking about designing my next experiment					
22.	I often spend my spare time talking to the pigeons ... and even they die of boredom					
23.	I tried to build myself a time machine so that I could go back to the 1930s and follow Fisher around on my hands and knees licking the floor on which he'd just trodden					
25.	I love teaching					
26.	I spend lots of time helping students					
27.	I love teaching because students have to pretend to like me or they'll get bad marks					
28.	My cat is my only friend					

Figure 7: The TOSSE-R

Multicollinearity. The determinant of the correlation matrix was .00000124, which is smaller than .00001 and, therefore, indicates that multicollinearity could be a problem in these data (although, strictly speaking, because we're using principal component analysis we don't need to worry).

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.894
Bartlett's Test of Sphericity	Approx. Chi-Square	2989.769
	df	378
	Sig.	.000

Output 14

Communalities

	Initial	Extraction
I once woke up in the middle of a vegetable patch hugging a turnip that I'd mistakenly dug up thinking it was Roy's largest root	.611	.601
If I had a big gun I'd shoot all the students I have to teach	.455	.508
I memorize probability values for the F-distribution	.544	.496
I worship at the shrine of Pearson	.535	.481
I still live with my mother and have little personal hygiene	.386	.430
Teaching others makes me want to swallow a large bottle of bleach because the pain of my burning oesophagus would be light relief in comparison	.355	.413
Helping others to understand Sums of Squares is a great feeling	.486	.411
I like control conditions	.686	.665
I calculate 3 ANOVAs in my head before getting out of bed every morning	.615	.607
I could spend all day explaining statistics to people	.373	.299
I like it when people tell me I've helped them to understand factor rotation	.554	.502
People fall asleep as soon as I open my mouth to speak	.171	.107
Designing experiments is fun	.507	.472
I'd rather think about appropriate dependent variables than go to the pub	.601	.631
I soil my pants with excitement at the mere mention of Factor Analysis	.453	.396
Thinking about whether to use repeated or independent measures thrills me	.561	.599
I enjoy sitting in the park contemplating whether to use participant observation in my next experiment	.671	.679
Standing in front of 300 people in no way makes me lose control of my bowels	.399	.405
I like to help students	.342	.367
Passing on knowledge is the greatest gift you can bestow an individual	.399	.342
Thinking about Bonferroni corrections gives me a tingly feeling in my groin	.566	.500
I quiver with excitement when thinking about designing my next experiment	.752	.764
I often spend my spare time talking to the pigeons ... and even they die of boredom	.414	.451
I tried to build myself a time machine so that I could go back to the 1930s and follow Fisher around on my hands and knees licking the floor on which he'd just trodden	.526	.573
I love teaching	.461	.455
I spend lots of time helping students	.511	.526
I love teaching because students have to pretend to like me or they'll get bad marks	.562	.585
My cat is my only friend	.387	.395

Extraction Method: Principal Axis Factoring.

Output 15

Sample size. MacCallum et al. (1999) have demonstrated that when communalities after extraction are above .5 a sample size between 100 and 200 can be adequate, and even when communalities are below .5 a sample size of 500 should be sufficient. We have a sample size of 239 with some communalities below .5, and so the sample size may not be adequate. However, the KMO measure of sampling adequacy is .894, which is above Kaiser's (1974) recommendation of .5. This value is also 'meritorious' (and almost 'marvellous') according to Hutcheson and Sofroniou (1999). As such, the evidence suggests that the sample size is adequate to yield distinct and reliable factors.

Bartlett's test. This tests whether the correlations between questions are sufficiently large for factor analysis to be appropriate (it actually tests whether the correlation matrix is sufficiently different from an identity matrix). In this case it is significant, $\chi^2(378) = 2989.77$, $p < .001$, indicating that the correlations within the *R*-matrix are sufficiently different from zero to warrant factor analysis.

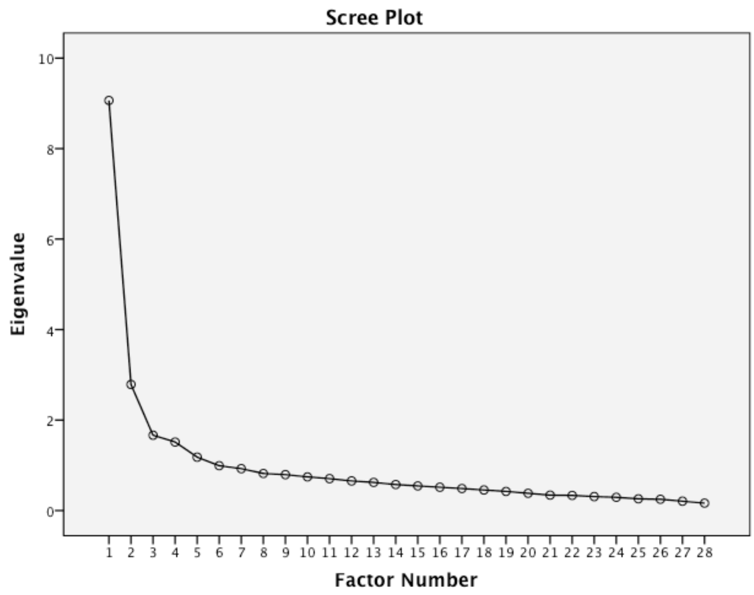
Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total
1	9.064	32.373	32.373	8.608	30.744	30.744	6.485
2	2.787	9.954	42.328	2.252	8.043	38.788	3.861
3	1.664	5.944	48.272	1.104	3.943	42.731	3.278
4	1.515	5.409	53.681	1.069	3.817	46.548	5.120
5	1.180	4.215	57.896	.627	2.239	48.787	3.353
6	.991	3.539	61.435				
7	.925	3.304	64.739				
8	.819	2.924	67.663				
9	.793	2.832	70.495				
10	.744	2.657	73.152				
11	.705	2.518	75.670				
12	.654	2.336	78.005				
13	.623	2.224	80.229				
14	.574	2.051	82.281				
15	.545	1.945	84.225				
16	.516	1.841	86.067				
17	.487	1.740	87.806				
18	.454	1.621	89.427				
19	.423	1.511	90.938				
20	.382	1.363	92.301				
21	.341	1.218	93.519				
22	.334	1.193	94.712				
23	.309	1.102	95.814				
24	.293	1.046	96.860				
25	.260	.928	97.788				
26	.248	.887	98.675				
27	.207	.738	99.414				
28	.164	.586	100.000				

Extraction Method: Principal Axis Factoring.

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

Output 16



Output 17

Extraction: SPSS has extracted five factors based on Kaiser's criterion of retaining factors with eigenvalues greater than 1. Is this warranted? Kaiser's criterion is accurate when there are less than 30 variables and the communalities after extraction are greater than .7, or when the sample size exceeds 250 and the average communality is greater than .6. For these data the sample size is 239, there are 28 variables, and the mean communality is .488, so extracting five factors is not really warranted. The scree plot shows clear inflexions at 3 and 5 factors and so using the scree plot you could justify extracting 3 or 5 factors.

Pattern Matrix[a]

	Factor				
	1	2	3	4	5
Thinking about whether to use repeated or independent measures thrills me	.757				
I'd rather think about appropriate dependent variables than go to the pub	.756				
I quiver with excitement when thinking about designing my next experiment	.727				
I enjoy sitting in the park contemplating whether to use participant observation in my next experiment	.668				
Designing experiments is fun	.496				
I like control conditions	.485				
I could spend all day explaining statistics to people					
I like to help students		.615			
I love teaching		.601			
I love teaching because students have to pretend to like me or they'll get bad marks		.553			
Passing on knowledge is the greatest gift you can bestow an individual		.524			
Helping others to understand Sums of Squares is a great feeling		.467			
I spend lots of time helping students		.449			
I like it when people tell me I've helped them to understand factor rotation					
I often spend my spare time talking to the pigeons ... and even they die of boredom			.59		
I still live with my mother and have little personal hygiene			.57		
My cat is my only friend			.57		
People fall asleep as soon as I open my mouth to speak					
I tried to build myself a time machine so that I could go back to the 1930s and follow Fisher around on my hands and knees licking the floor on which he'd just trodden				.693	
I memorize probability values for the F-distribution				.544	
I worship at the shrine of Pearson				.476	
I once woke up in the middle of a vegetable patch hugging a turnip that I'd mistakenly dug up thinking it was Roy's largest root				.448	
I soil my pants with excitement at the mere mention of Factor Analysis				.425	
Thinking about Bonferroni corrections gives me a tingly feeling in my groin	.406			.411	
If I had a big gun I'd shoot all the students I have to teach					.70
Teaching others makes me want to swallow a large bottle of bleach because the pain of my burning oesophagus would be light relief in comparison					.67
Standing in front of 300 people in no way makes me lose control of my bowels					.46
I calculate 3 ANOVAs in my head before getting out of bed every morning					

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

a. Rotation converged in 15 iterations.

Output 18

Rotation: You should choose an oblique rotation because the question says that the constructs we're measuring are related.

Looking at the pattern matrix (and using loadings greater than .4 as recommended by Stevens) we see the following pattern:

Factor 1:

- Q 16. Thinking about whether to use repeated or independent measures thrills me
- Q 14. I'd rather think about appropriate dependent variables than go to the pub
- Q 22. I quiver with excitement when thinking about designing my next experiment
- Q 17. I enjoy sitting in the park contemplating whether to use participant observation in my next experiment
- Q 13. Designing experiments is fun
- Q 8. I like control conditions
- Q 10. I could spend all day explaining statistics to people

Factor 2:

- Q 19. I like to help students
- Q 20. Passing on knowledge is the greatest gift you can bestow an individual
- Q 25. I love teaching

- Q 27. I love teaching because students have to pretend to like me or they'll get bad marks
- Q 7. Helping others to understand sums of squares is a great feeling
- Q 26. I spend lots of time helping students

Factor 3:

- Q 23. I often spend my spare time talking to the pigeons ... and even they die of boredom
- Q 28. My cat is my only friend
- Q 5. I still live with my mother and have little personal hygiene
- Q 12. People fall asleep as soon as I open my mouth to speak

Factor 4:

- Q 24. I tried to build myself a time machine so that I could go back to the 1930s and follow Fisher around on my hands and knees licking the floor on which he'd just trodden
- Q 3. I memorize probability values for the *F*-distribution
- Q 4. I worship at the shrine of Pearson
- Q 15. I soil my pants with excitement at the mere mention of factor analysis
- Q 21. Thinking about Bonferroni corrections gives me a tingly feeling in my groin
- Q 1. I once woke up in the middle of a vegetable patch hugging a turnip that I'd mistakenly dug up thinking it was Roy's largest root

Factor 5:

- Q 6. Teaching others makes me want to swallow a large bottle of bleach because the pain of my burning oesophagus would be light relief in comparison
- Q 2. If I had a big gun I'd shoot all the students I have to teach
- Q 18. Standing in front of 300 people in no way makes me lose control of my bowels

No factor:

- Q 9. I calculate three ANOVAs in my head before getting out of bed every morning
- Q 11. I like it when people tell me I've helped them to understand factor rotation

Factor 1 seems to relate to research methods, factor 2 to teaching, factor 3 to general social skills, factor 4 to statistics and factor 5 to, well, err, teaching again. All in all, this isn't particularly satisfying and doesn't really support the four-factor model. We saw earlier that the extraction of five factors probably wasn't justified. In fact the scree plot seems to indicate three. Let's rerun the analysis but asking SPSS for three factors. Let's see how this changes the pattern matrix:

Pattern Matrix[a]

	Factor 1	Factor 2	Factor 3
I quiver with excitement when thinking about designing my next experiment	.893		
I enjoy sitting in the park contemplating whether to use participant observation in my next experiment	.844		
I like control conditions	.781		
Designing experiments is fun	.714		
I calculate 3 ANOVAs in my head before getting out of bed every morning	.702		
Thinking about Bonferroni corrections gives me a tingly feeling in my groin	.699		
I once woke up in the middle of a vegetable patch hugging a turnip that I'd mistakenly dug up thinking it was Roy's largest root	.657		
Thinking about whether to use repeated or independent measures thrills me	.603		
I soil my pants with excitement at the mere mention of Factor Analysis	.535		
I memorize probability values for the F-distribution	.519		
I worship at the shrine of Pearson	.508		
I'd rather think about appropriate dependent variables than go to the pub	.506		
I tried to build myself a time machine so that I could go back to the 1930s and follow Fisher around on my hands and knees licking the floor on which he'd just trodden	.504		
I could spend all day explaining statistics to people	.461		
Helping others to understand Sums of Squares is a great feeling	.442	.398	
I like it when people tell me I've helped them to understand factor rotation	.438	.314	
I like to help students		.547	
I spend lots of time helping students	.318	.516	
I love teaching		.448	
If I had a big gun I'd shoot all the students I have to teach	.323	-.427	
Standing in front of 300 people in no way makes me lose control of my bowels	.303	-.419	
Teaching others makes me want to swallow a large bottle of bleach because the pain of my burning oesophagus would be light relief in comparison	.314	-.382	
Passing on knowledge is the greatest gift you can bestow an individual		.369	
I often spend my spare time talking to the pigeons ... and even they die of boredom			.636
I still live with my mother and have little personal hygiene			.566
My cat is my only friend			.551
I love teaching because students have to pretend to like me or they'll get bad marks		.435	.497
People fall asleep as soon as I open my mouth to speak			.331

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

a. Rotation converged in 14 iterations.

Output 19

Looking at the pattern matrix (and using loadings greater than .4 as recommended by Stevens) we see the following pattern:

Factor 1:

- Q 22. I quiver with excitement when thinking about designing my next experiment
- Q 8. I like control conditions
- Q 17. I enjoy sitting in the park contemplating whether to use participant observation in my next experiment
- Q 21. Thinking about Bonferroni corrections gives me a tingly feeling in my groin
- Q 13. Designing experiments is fun
- Q 9. I calculate three ANOVAs in my head before getting out of bed every morning
- Q 3. I memorize probability values for the *F*-distribution
- Q 1. I once woke up in the middle of a vegetable patch hugging a turnip that I'd mistakenly dug up thinking it was Roy's largest root
- Q 24. I tried to build myself a time machine so that I could go back to the 1930s and follow Fisher around on my hands and knees licking the floor on which he'd just trodden
- Q 4. I worship at the shrine of Pearson
- Q 16. Thinking about whether to use repeated or independent measures thrills me
- Q 7. Helping others to understand sums of squares is a great feeling
- Q 15. I soil my pants with excitement at the mere mention of factor analysis
- Q 11. I like it when people tell me I've helped them to understand factor rotation
- Q 10. I could spend all day explaining statistics to people

- Q 14. I'd rather think about appropriate dependent variables than go to the pub

Factor 2:

- Q 19. I like to help students
- Q 2. If I had a big gun I'd shoot all the students I have to teach (note negative weight)
- Q 6. Teaching others makes me want to swallow a large bottle of bleach because the pain of my burning oesophagus would be light relief in comparison (note negative weight)
- Q 18. Standing in front of 300 people in no way makes me lose control of my bowels (note negative weight)
- Q 26. I spend lots of time helping students
- Q 25. I love teaching
- Q 20. Passing on knowledge is the greatest gift you can bestow an individual

Factor 3:

- Q 5. I still live with my mother and have little personal hygiene
- Q 23. I often spend my spare time talking to the pigeons ... and even they die of boredom
- Q 28. My cat is my only friend
- Q 12. People fall asleep as soon as I open my mouth to speak
- Q 27. I love teaching because students have to pretend to like me or they'll get bad marks

This factor is a lot clearer-cut: factor 1 relates to a love of methods and statistics, factor 2 to a love of teaching, and factor 3 to an absence of normal social skills. This doesn't support the original four-factor model suggested because the data indicate that love of methods and statistics can't be separated (if you love one you love the other).

Task 3

*Dr Sian Williams (University of Brighton) devised a questionnaire to measure organizational ability. She predicted five factors to do with organizational ability: (1) preference for organization; (2) goal achievement; (3) planning approach; (4) acceptance of delays; and (5) preference for routine. These dimensions are theoretically independent. Williams' questionnaire contains 28 items using a 7-point Likert scale (1 = strongly disagree, 4 = neither, 7 = strongly agree). She gave it to 239 people. Run a principal component analysis on the data in **Williams.sav**.*

1	I like to have a plan to work to in everyday life
2	I feel frustrated when things don't go to plan
3	I get most things done in a day that I want to
4	I stick to a plan once I have made it
5	I enjoy spontaneity and uncertainty
6	I feel frustrated if I can't find something I need

7	I find it difficult to follow a plan through
8	I am an organized person
9	I like to know what I have to do in a day
10	Disorganized people annoy me
11	I leave things to the last minute
12	I have many different plans relating to the same goal
13	I like to have my documents filed and in order
14	I find it easy to work in a disorganized environment
15	I make 'to do' lists and achieve most of the things on it
16	My workspace is messy and disorganized
17	I like to be organized
18	Interruptions to my daily routine annoy me
19	I feel that I am wasting my time
20	I forget the plans I have made
21	I prioritize the things I have to do
22	I like to work in an organized environment
23	I feel relaxed when I don't have a routine
24	I set deadlines for myself and achieve them
25	I change rather aimlessly from one activity to another during the day
26	I have trouble organizing the things I have to do
27	I put tasks off to another day
28	I feel restricted by schedules and plans

Correlation Matrix[a]

Correlation
i like to have a plan to work to in everyday life

a. Determinant = 1.240E-06

Output 4

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.894
Bartlett's Test of Sphericity	Approx. Chi-Square	2989.769
	df	378
	Sig.	.000

Output 5

Communalities

	Initial	Extraction
i like to have a plan to work to in everyday life	1.000	.646
i feel frustrated when things don't go to plan	1.000	.624
i get most thigs done in a day that i want to	1.000	.591
i stick to a plan once i have made it	1.000	.589
i enjoy spontaneity and uncertainty	1.000	.545
i feel frustrated if i can't find something i need	1.000	.621
i find it difficult to follow a plan through	1.000	.486
i am an organised person	1.000	.683
i ike to know what i have to do in a day	1.000	.638
disorganised people annoy me	1.000	.417
i leace things to the last minute	1.000	.539
i have many different plans relating to th esame goal	1.000	.297
i like to have my documents filed and in order	1.000	.531
i find it easy to work in a disorganised environment	1.000	.709
i make 'to do' lists and acheive most of the things on it	1.000	.511
my workspace is messy and disorganised	1.000	.681
i like to be organised	1.000	.705
interruptions to my daily routine annoy me	1.000	.514
i feel that i am wasting my time	1.000	.536
i forget the plans i have made	1.000	.477
i prioritise the things i have to do	1.000	.566
i like to work in an organised environment	1.000	.766
i feel relaxed when i don't have a routine	1.000	.587
i set deadlines for myself and acheive them	1.000	.649
i change rather aimlessly from one activity to another during the day	1.000	.550
i have trouble organising the things i have to do	1.000	.599
i put tasks off to another day	1.000	.619
i feel restristed by schedules and plans	1.000	.538

Extraction Method: Principal Component Analysis.

Output 22

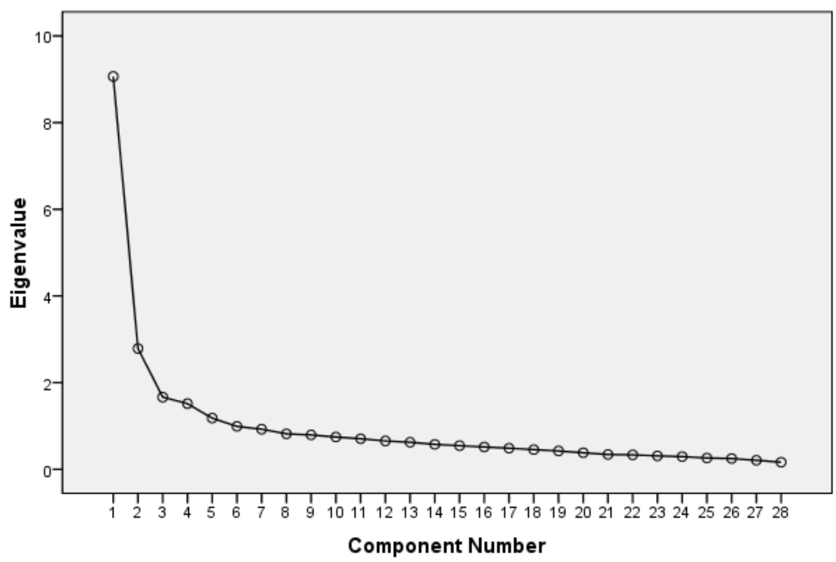
Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	9.064	32.373	32.373	9.064	32.373	32.373	4.558	16.279	16.279
2	2.787	9.954	42.328	2.787	9.954	42.328	3.460	12.356	28.635
3	1.664	5.944	48.272	1.664	5.944	48.272	3.239	11.568	40.203
4	1.515	5.409	53.681	1.515	5.409	53.681	2.631	9.397	49.600
5	1.180	4.215	57.896	1.180	4.215	57.896	2.323	8.296	57.896
6	.991	3.539	61.435						
7	.925	3.304	64.739						
8	.819	2.924	67.663						
9	.793	2.832	70.495						
10	.744	2.657	73.152						
11	.705	2.518	75.670						
12	.654	2.336	78.005						
13	.623	2.224	80.229						
14	.574	2.051	82.281						
15	.545	1.945	84.225						
16	.516	1.841	86.067						
17	.487	1.740	87.806						
18	.454	1.621	89.427						
19	.423	1.511	90.938						
20	.382	1.363	92.301						
21	.341	1.218	93.519						
22	.334	1.193	94.712						
23	.309	1.102	95.814						
24	.293	1.046	96.860						
25	.260	.928	97.788						
26	.248	.887	98.675						
27	.207	.738	99.414						
28	.164	.586	100.000						

Extraction Method: Principal Component Analysis.

Output 23

Scree Plot



Output 24

Component Matrix[a]

	Component				
	1	2	3	4	5
i like to have a plan to work to in everyday life	.684				
i feel frustrated when things don't go to plan		-.543			
i get most thigs done in a day that i want to	.584				
i stick to a plan once i have made it	.600			.452	
i enjoy spontaneity and uncertainty	.446		.524		
i feel frustrated if i can't find something i need		-.501			.453
i find it difficult to follow a plan through	.528				
i am an organised person	.803				
i ike to know what i have to do in a day	.723				
disorganised people annoy me	.502				
i leace things to the last minute	.675				
i have many different plans relating to th esame goal			.519		
i like to have my documents filed and in order	.673				
i find it easy to work in a disorganised environment	.614			-.517	
i make 'to do' lists and acheive most of the things on it	.559				
my workspace is messy and disorganised	.650			-.497	
i like to be organised	.768				
interruptions to my daily routine annoy me	.421	-.523			
i feel that i am wasting my time		.620			
i forget the plans i have made	.456				
i prioritise the things i have to do	.674				
i like to work in an organised environment	.791				
i feel relaxed when i don't have a routine	.432		.518		
i set deadlines for myself and acheive them	.614				
i change rather aimlessly from one activity to another during the day	.501	.444			
i have trouble organising the things i have to do	.533	.502			
i put tasks off to another day	.580				
i feel restristed by schedules and plans	.458		.520		

Extraction Method: Principal Component Analysis.

a. 5 components extracted.

Output 25

Rotated Component Matrix[a]

	Component				
	1	2	3	4	5
i like to have a plan to work to in everyday life	.409	.545			
i feel frustrated when things don't go to plan				.765	
i get most thigs done in a day that i want to		.666			
i stick to a plan once i have made it		.619			
i enjoy spontaneity and uncertainty					.666
i feel frustrated if i can't find something i need				.781	
i find it difficult to follow a plan through			.535		
i am an organised person	.587				
i ike to know what i have to do in a day	.432	.470		.447	
disorganised people annoy me	.440			.450	
i leace things to the last minute			.435		
i have many different plans relating to th esame goal					.506
i like to have my documents filed and in order	.593				
i find it easy to work in a disorganised environment	.764				
i make 'to do' lists and acheive most of the things on it	.447	.509			
my workspace is messy and disorganised	.775				
i like to be organised	.714				
interruptions to my daily routine annoy me				.586	
i feel that i am wasting my time			.712		
i forget the plans i have made			.649		
i prioritise the things i have to do	.505	.523			
i like to work in an organised environment	.748				
i feel relaxed when i don't have a routine					.672
i set deadlines for myself and acheive them		.744			
i change rather aimlessly from one activity to another during the day			.688		
i have trouble organising the things i have to do	.407		.568		
i put tasks off to another day			.613		.411
i feel restristed by schedules and plans					.673

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

a. Rotation converged in 7 iterations.

Output 26

Component Transformation Matrix

Component	1	2	3	4	5
1	.633	.520	.384	.302	.301
2	-.118	.050	.738	-.650	-.129
3	-.188	-.346	.106	-.053	.911
4	-.742	.503	.201	.393	.038
5	.025	-.595	.506	.574	-.246

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

Output 27

Extraction. SPSS has extracted five factors based on Kaiser's criterion of retaining factors with eigenvalues greater than 1. Is this warranted? Kaiser's criterion is accurate when there are less than 30 variables and the communalities after extraction are greater than .7, or when the sample size exceeds 250 and the average communality is greater than .6. For these data the sample size is 239 and the mean communality is .579, so extracting five factors is not really warranted. The scree plot shows clear inflexions at 3 and 5 factors, and so using the scree plot you could justify extracting 3 or 5 factors.

Looking at the rotated component matrix (and using loadings greater than .4 as recommended by Stevens) we see the following pattern:

Factor 1: preference for organization

- Q8: I am an organized person
- Q13: I like to have my documents filed and in order
- Q14: I find it easy to work in a disorganized environment
- Q 16: My workspace is messy and disorganized
- Q17: I like to be organized
- Q22: I like to work in an organized environment

Note: It's odd that none of these have reverse loadings.

Factor 2: plan approach

- Q1: I like to have a plan to work to in everyday life
- Q3: I get most things done in a day that I want to
- Q4: I stick to a plan once I have made it
- Q9: I like to know what I have to do in a day
- Q15: I make 'to do' lists and achieve most of the things on it
- Q 21: I prioritize the things I have to do
- Q24: I set deadlines for myself and achieve them

Factor 3: goal achievement

- Q7: I find it difficult to follow a plan through
- Q11: I leave things to the last minute
- Q19: I feel that I am wasting my time
- Q20: I forget the plans I have made
- Q25: I change rather aimlessly from one activity to another during the day
- Q26: I have trouble organizing the things I have to do
- Q27: I put tasks off to another day

Factor 4: acceptance of delays

- Q2: I feel frustrated when things don't go to plan
- Q6: I feel frustrated if I can't find something I need
- Q10: Disorganized people annoy me
- Q18: Interruptions to my daily routine annoy me

Factor 5: preference for routine

- Q5: I enjoy spontaneity and uncertainty
- Q12: I have many different plans relating to the same goal
- Q23: I feel relaxed when I don't have a routine

- Q28: I feel restricted by schedules and plans

Therefore, it seems as though there is some factorial validity to the structure.

Task 4

*Zibarras, Port, and Woods (2008) looked at the relationship between personality and creativity. They used the Hogan Development Survey (HDS), which measures 11 dysfunctional dispositions of employed adults: being **volatile**, **mistrustful**, **cautious**, **detached**, **passive-aggressive**, **arrogant**, **manipulative**, **dramatic**, **eccentric**, **perfectionist**, and **dependent**. Zibarras et al. wanted to reduce these 11 traits and, based on parallel analysis, found that they could be reduced to three components. They ran a principal component analysis with varimax rotation. Repeat this analysis (**Zibarras et al. (2008)**) to see which personality dimensions clustered together (see p. 210 of the original paper)**.***

Rotated Component Matrix[a]

	Component		
	1	2	3
Dramatic	.833		
Manipulative	.785		
Arrogant	.677		
Cautious	-.659	.500	
Eccentric	.549		
Perfectionist	-.319		
Volatile		.790	
Mistrustful		.681	
Detached			.765
Dependent	-.380	.335	-.639
Passive-Agressive			.617

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

a. Rotation converged in 5 iterations.

Output 28

Output shows the results the principal component analysis with varimax rotation – I requested that SPSS extracts three factors to match Zibarras et al. (2008). We can see that the following pattern:

Factor 1:

Dramatic

Manipulative

Arrogant

Cautious (negative weight)

Eccentric

Perfectionist (negative weight)

Factor 2:

Volatile

Mistrustful

Factor 3:

Detached

Dependent (negative weight)

Passive-aggressive

If we compare our results to those of Zibarras et al. (Figure 8), we can see that they are the same.

TABLE 4. Factor loadings of HDS primary dimensions on four extracted factors from principal components analysis.

Primary Scale	Factor I	Factor II	Factor III
Dramatic	**.83**		
Manipulative	**.79**		
Arrogant	**.68**		
Cautious	**−.66**	.50	
Eccentric	**.55**		
Perfectionist	**−.32**		
Volatile		**.79**	
Mistrustful		**.68**	
Detached			**.77**
Dependent	−.38		**−.64**
Passive-aggressive			**.62**

Note. Primary factor loadings shown in bold. Absolute factor loadings under 0.30 not reported.

Figure 8