

Chapter 18: Exploratory factor analysis

Labcoat Leni's Real Research

Worldwide addiction?

Problem

Nichols, L. A., & Nicki, R. (2004). *Psychology of Addictive Behaviors*, 18(4), 381-384



In 2007 it was estimated that around 179 million people worldwide used the Internet. From the increasing popularity (and usefulness) of the Internet has emerged a serious and recognized problem of Internet addiction. To research this construct it's helpful to be able to measure it, so Laura Nichols and Richard Nicki developed the Internet Addiction Scale (Nichols & Nicki, 2004). Nichols and Nicki's 36-item questionnaire contains items such as 'I have stayed on the Internet longer than I intended to' and 'My grades/work have suffered because of my Internet use' to which responses are made on a 5-point scale (Never, Rarely, Sometimes, Frequently, Always). (Incidentally, while researching this topic I encountered an Internet addiction recovery website that offered a whole host of resources (e.g., questionnaires, online support groups, videos, podcasts, etc.) that would keep you online for ages. It struck me that this was a like having a heroin addiction recovery centre that had a huge pile of free heroin in the reception area.)

The data from 207 people this study are in the file **Nichols & Nicki (2004).sav**. The authors dropped two items because they had low means and variances, and dropped three others because of relatively low correlations with other items. They performed a principal component analysis on the remaining 31 items. Labcoat Leni wants you to run some descriptive statistics to work out which two items were dropped for having low means/variances, then inspect a correlation matrix to find the three items that were dropped for having low correlations. Finally, he wants you to run a principal component analysis on the data.

Solution

To get the descriptive statistics I would use **Analyze Descriptive Statistics** ▶ **Frequencies...**. Select all of the questionnaire items (Figure 1) but just ask for means and standard deviations (Figure 2) at this stage.

DISCOVERING STATISTICS USING SPSS

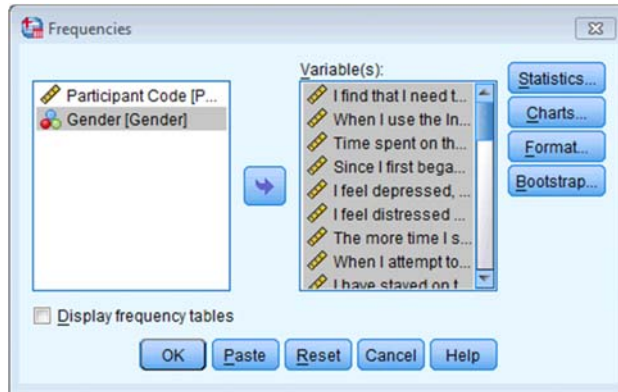


Figure 1

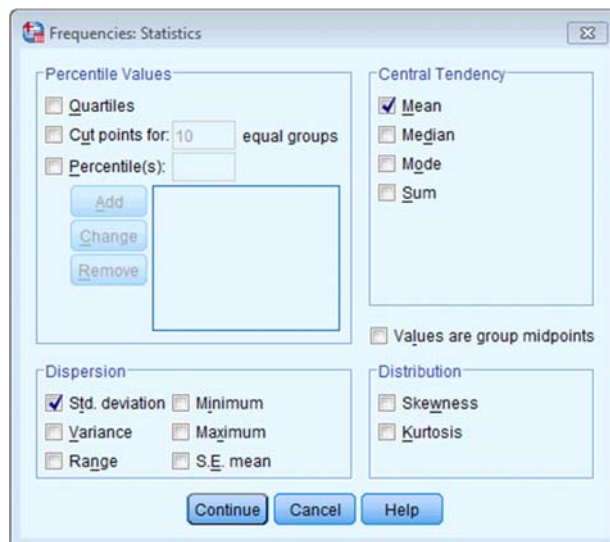


Figure 2

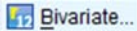
The table of means and standard deviations shows that the items with the lowest values are IAS-23 ('I see my friends less often because of the time that I spend on the Internet') and IAS-34 ('When I use the Internet, I experience a buzz or a high').

DISCOVERING STATISTICS USING SPSS

Statistics

	N		
	Valid	Mean	Variance
I find that I need to use the Internet more to get the same enjoyment as before.	207	1.49	.678
When I use the Internet now, I do not feel as good as I used to.	207	1.59	.865
Time spent on the Internet now is not as enjoyable as it was when I first started using the Internet.	207	2.68	1.150
Since I first began using the Internet I would say that the amount of time I spend on line has increased but not the satisfaction.	207	2.01	1.150
I feel depressed, moody or nervous when I am off the internet which goes away when I log on.	207	1.51	.717
I feel distressed when I am unable to spend as much time on the Internet as I usually do.	207	1.22	.319
The more time I spend away from the Internet, the more irritable I feel.	207	1.41	.640
When I attempt to cut back of stop using the Internet I find that the irritability that I experience is relieved by going back on the Internet	207	2.09	1.274
I have stayed on the Internet longer than I intended to.	207	1.66	.914
I have said to myself "just a few more minutes on the Internet."	207	1.36	.483
I find myself accessing more information on the Internet that I had planned to.	207	1.48	.571
I find myself doing more things on the Internet than I had intended to	207	1.71	.741
I have felt a persistent desire to cut down or control my use of the Internet.	207	2.03	1.256
I have attempted to spend less time on the Internet but I have been unable to do so.	207	1.33	.416
I have tried unsuccessfully to restrict my Internet use because of previous over use.	207	1.23	.266
I would like to spend less time on the Internet.	207	1.30	.483
I have walked or driven to campus/work specifically to use the Internet at times when I normally would not go to campus/work	207	1.31	.459
After being on the Internet late into the night in sleep late the next morning because of my Internet use.	207	1.33	.466
Once I am on the Internet, I seem to stay on for a long time.	207	2.03	.897
I am on the Internet so much that I have to make up for the lost time.	207	1.32	.412
I have missed class/work so that I would have more time to spend on the Internet.	207	1.58	.924
I have neglected things, which are important and need doing.	207	1.25	.432
I see my friends less often because of the time that I spend on the Internet.	207	1.14	.183
I have given up a particular recreational activity in order that I would have more time on the Internet	207	1.89	.921
At times I have tried to conceal how long I have been on the Internet	207	1.39	.482
My grades/work have suffered because of my Internet use.	207	1.25	.322
I have lost sleep because of my Internet use	207	1.91	.967
The Internet has affected my life in a negative way.	207	1.24	.317
The people I know through the Internet know me better than my friends at university	207	1.23	.322
I prefer socializing on the Internet rather than in person with my friends and family	207	1.51	.639
I feel that life without the Internet would be boring and empty.	207	2.27	1.070
I find myself thinking/longing about when I will go on the Internet again.	207	1.54	.803
When I feel lonely, I use the Internet to talk to others.	207	1.35	.559
When I use the Internet, I experience a buzz or a high (i.e., feeling elated).	207	1.11	.119
I use the Internet as a way of escaping the "real world."	207	1.50	.708
I enjoy the pleasure/excitement of being on the Internet.	207	1.27	.342

Output 1

To get a table of correlations select **Analyze > Correlate**  **Bivariate...**. Select all of the variables and leave the default options as they are (Figure 3).

DISCOVERING STATISTICS USING SPSS

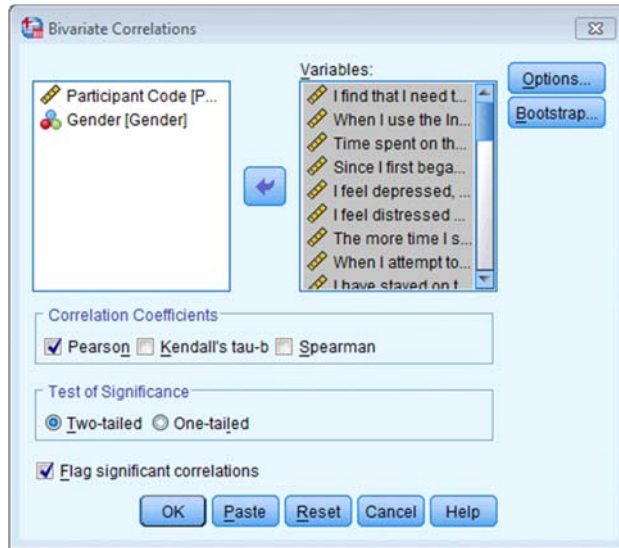


Figure 3

We know that the authors eliminated three items for having low correlations. My table of correlations has the average correlation (Output 2). The lowest average correlations are for items IAS-13 ('I have felt a persistent desire to cut down or control my use of the Internet'), IAS-22 ('I have neglected things which are important and need doing'), and IAS-32 ('I find myself thinking/longing about when I will go on the Internet again.'). As such these variables will also be excluded from the factor analysis.

To do the principal component analysis select **Analyze > Dimension Reduction > Factor...**. Choose all of the variables except for the five that we have excluded (Figure 4).

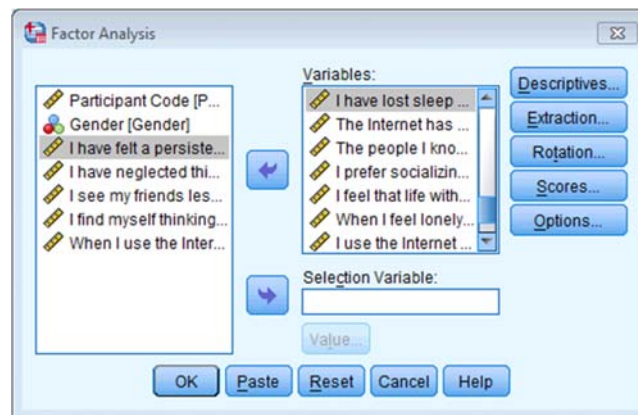


Figure 4

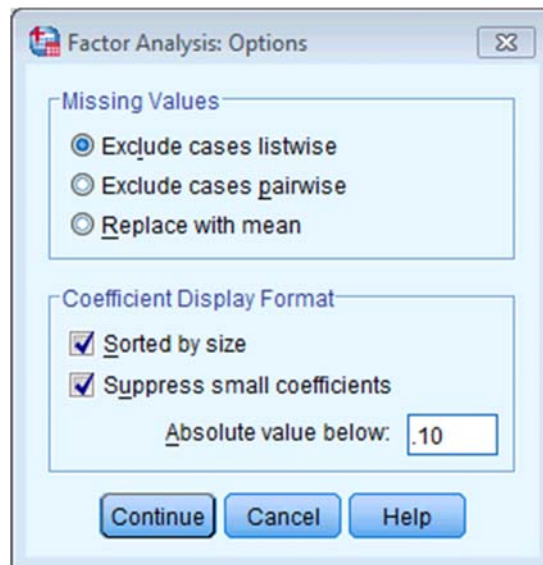


Figure 5

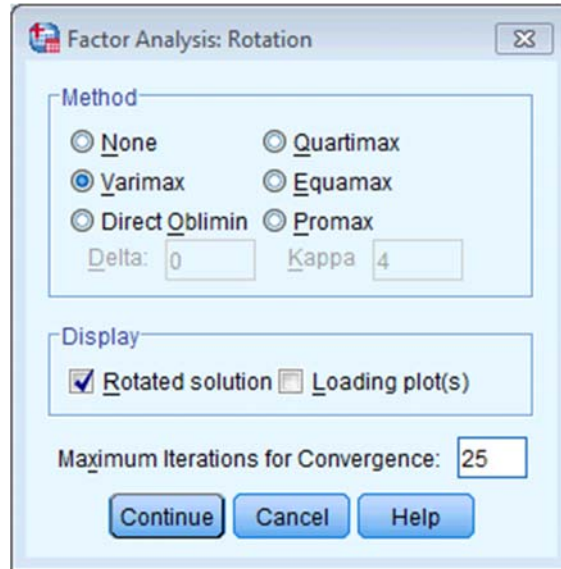


Figure 6

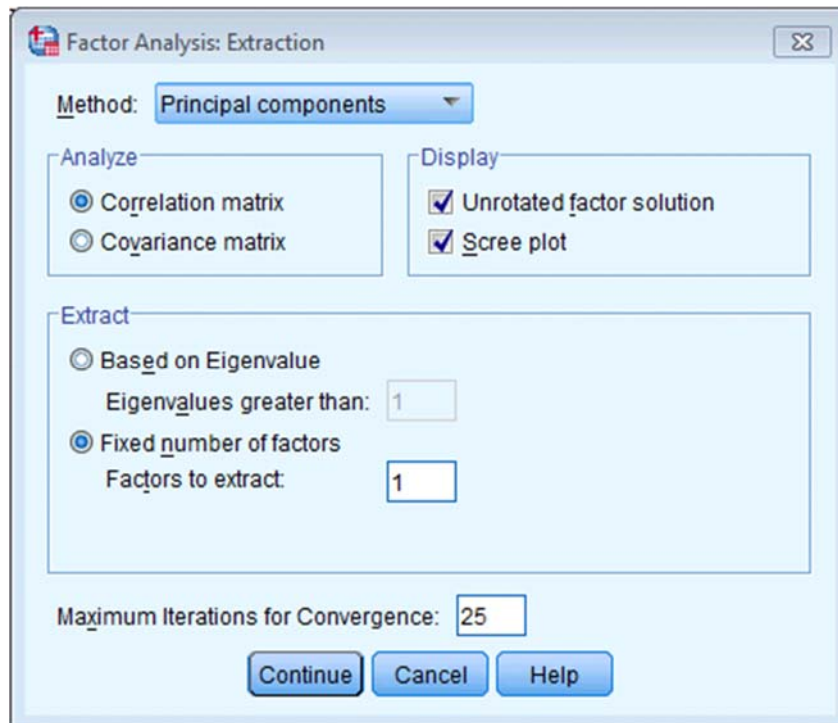


Figure 7

The output should look like this:

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.942
Bartlett's Test of Sphericity	Approx. Chi-Square	4238.976
	df	465.000
	Sig.	.000

Output 3

Sample size: 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 (MacCallum, Widaman, Zhang, & Hong, 1999). We have a sample size of 207 with only one communality below .5, and so the sample size should be adequate. However, the KMO measure of sampling adequacy is .942, which is above Kaiser's (1974) recommendation of .5. This value is also '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, $\chi^2(465) = 4238.98$, $p < .001$, indicating that the correlations within the *R*-matrix are sufficiently different from zero to warrant factor analysis.

Communalities

	Initial	Extraction
I find that I need to use the Internet more to get the same enjoyment as before.	1.000	.630
When I use the Internet now, I do not feel as good as I used to.	1.000	.688
Time spent on the Internet now is not as enjoyable as it was when I first started using the Internet.	1.000	.693
Since I first began using the Internet I would say that the amount of time I spend on line has increased but not the satisfaction.	1.000	.682
I feel depressed, moody or nervous when I am off the internet which goes away when I log on.	1.000	.574
I feel distressed when I am unable to spend as much time on the Internet as I usually do.	1.000	.657
The more time I spend away from the Internet, the more irritable I feel.	1.000	.621
When I attempt to cut back or stop using the Internet I find that the irritability that I experience is relieved by going back on the Internet	1.000	.614
I have stayed on the Internet longer than I intended to.	1.000	.566
I have said to myself "just a few more minutes on the Internet."	1.000	.682
I find myself accessing more information on the Internet than I had planned to.	1.000	.537
I find myself doing more things on the Internet than I had intended to	1.000	.657
I have attempted to spend less time on the Internet but I have been unable to do so.	1.000	.482
I have tried unsuccessfully to restrict my Internet use because of previous over use.	1.000	.639
I would like to spend less time on the Internet.	1.000	.555
I have walked or driven to campus/work specifically to use the Internet at times when I normally would not go to campus/work	1.000	.612
After being on the Internet late into the night in sleep late the next morning because of my Internet use.	1.000	.680
Once I am on the Internet, I seem to stay on for a long time.	1.000	.659
I am on the Internet so much that I have to make up for the lost time.	1.000	.734
I have missed class/work so that I would have more time to spend on the Internet.	1.000	.664
I have given up a particular recreational activity in order that I would have more time on the Internet	1.000	.704
At times I have tried to conceal how long I have been on the Internet	1.000	.640
My grades/work have suffered because of my Internet use.	1.000	.779
I have lost sleep because of my Internet use	1.000	.507
The Internet has affected my life in a negative way.	1.000	.801
The people I know through the Internet know me better than my friends at university	1.000	.694
I prefer socializing on the Internet rather than in person with my friends and family	1.000	.450
I feel that life without the Internet would be boring and empty.	1.000	.637
When I feel lonely, I use the Internet to talk to others.	1.000	.684
I use the Internet as a way of escaping the "real world."	1.000	.640
I enjoy the pleasure/excitement of being on the Internet.	1.000	.703

Extraction Method: Principal Component Analysis.

Output 4

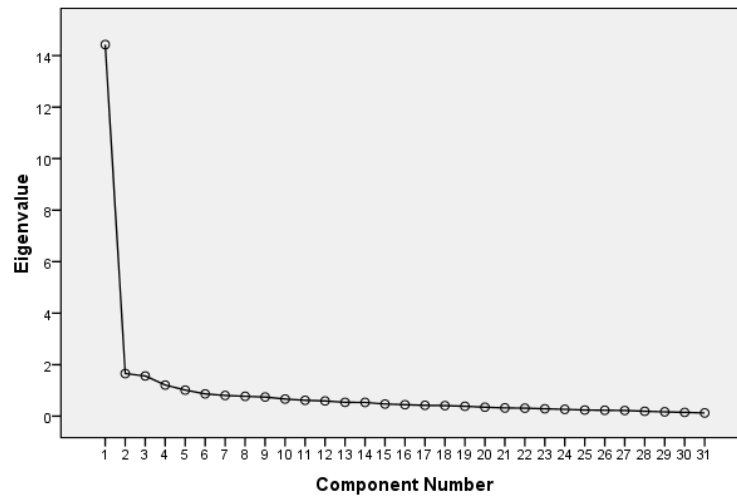
Total Variance Explained

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	14.427	46.539	46.539	14.427	46.539	46.539	6.079	19.609	19.609
2	1.654	5.337	51.875	1.654	5.337	51.875	4.283	13.816	33.425
3	1.562	5.037	56.913	1.562	5.037	56.913	3.633	11.719	45.144
4	1.209	3.901	60.814	1.209	3.901	60.814	3.427	11.056	56.200
5	1.014	3.272	64.086	1.014	3.272	64.086	2.445	7.887	64.086
6	.867	2.797	66.883						
7	.805	2.597	69.480						
8	.772	2.492	71.972						
9	.744	2.399	74.370						
10	.665	2.144	76.515						
11	.617	1.992	78.506						
12	.592	1.910	80.416						
13	.539	1.739	82.155						
14	.535	1.726	83.881						
15	.472	1.524	85.404						
16	.445	1.436	86.841						
17	.420	1.356	88.197						
18	.410	1.323	89.519						
19	.385	1.241	90.760						
20	.350	1.129	91.889						
21	.322	1.038	92.928						
22	.311	1.004	93.931						
23	.290	.934	94.865						
24	.267	.861	95.726						
25	.241	.776	96.503						
26	.230	.742	97.245						
27	.222	.716	97.960						
28	.190	.614	98.574						
29	.168	.543	99.117						
30	.146	.472	99.589						
31	.127	.411	100.000						

Extraction Method: Principal Component Analysis.

Output 5

Scree Plot



Output 6

Extraction: SPSS has extracted five factors based on Kaiser's criterion of retaining factors with eigenvalues greater than 1 (Output 5). 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 207, there are 31 variables and the mean communality is .64, so extracting five factors is probably not warranted. The scree plot (Output 6) shows a clear one-factor solution. This is the solution that the authors adopted.

Because we are retaining only one factor we can ignore the rotated factor solution and just look at the unrotated component matrix (Output 7). This shows that all items have a high loading on factor 1.

Component Matrix ^a					
	Component				
	1	2	3	4	5
I enjoy the pleasure/excitement of being on the Internet.	.802				
I am on the Internet so much that I have to make up for the lost time.	.787				
My grades/work have suffered because of my Internet use.	.773				
The people I know through the Internet know me better than my friends at university	.755				
Once I am on the Internet, I seem to stay on for a long time.	.747				
I have missed class/work so that I would have more time to spend on the Internet.	.743				
When I attempt to cut back or stop using the Internet I find that the irritability that I experience is relieved by going back on the Internet	.739				
The Internet has affected my life in a negative way.	.737		.428		
I have said to myself "just a few more minutes on the Internet."	.727				
I have stayed on the Internet longer than I intended to.	.720				
I have given up a particular recreational activity in order that I would have more time on the Internet	.720				
I find myself doing more things on the Internet than I had intended to	.720				
After being on the Internet late into the night in sleep late the next morning because of my Internet use.	.715				
When I feel lonely, I use the Internet to talk to others.	.714				
I find that I need to use the Internet more to get the same enjoyment as before.	.701				
The more time I spend away from the Internet, the more irritable I feel.	.700				
I feel depressed, moody or nervous when I am off the internet which goes away when I log on.	.694				
I have tried unsuccessfully to restrict my Internet use because of previous over use.	.683				
Time spent on the Internet now is not as enjoyable as it was when I first started using the Internet.	.680				
I feel distressed when I am unable to spend as much time on the Internet as I usually do.	.670				
I find myself accessing more information on the Internet that I had planned to.	.669				
I have attempted to spend less time on the Internet but I have been unable to do so.	.667				
At times I have tried to conceal how long I have been on the Internet	.662				
I have walked or driven to campus/work specifically to use the Internet at times when I normally would not go to campus/work	.653				
I prefer socializing on the Internet rather than in person with my friends and family	.644				
I use the Internet as a way of escaping the "real world."	.558		.416		
Since I first began using the Internet I would say that the amount of time I spend on line has increased but not the satisfaction.	.557				
I would like to spend less time on the Internet.	.548				
I have lost sleep because of my Internet use	.529				
I feel that life without the Internet would be boring and empty.	.497	.438		-.422	
When I use the Internet now, I do not feel as good as I used to.	.477			.428	

Extraction Method: Principal Component Analysis.

a. 5 components extracted.

Output 7

The authors reported their analysis as follows:

We conducted principal-components analyses on the log transformed scores of the IAS (see above). On the basis of the scree test (Cattell, 1978) and the percentage of variance accounted for by each factor, we judged a one-factor solution to be most appropriate. This component accounted for a total of 46.50% of the variance. A value for loadings of .30 (Floyd & Widaman, 1995) was used as a cut-off for items that did not relate to a component.

All 31 items loaded on this component, which was interpreted to represent aspects of a general factor relating to Internet addiction reflecting the negative consequences of excessive Internet use. (p. 382)