

Factor Analysis in SPSS

Background

Factor analysis looks at a set of items and attempts to determine the number of constructs (i.e., latent variables) underlying them.

Factor Analysis

The syntax for running factor analysis is:

```
factor
/variables var1 var2 var3 [var4, ...]
[/criteria factors(n)]
/extraction paf
/method=correlation
/rotation varimax
[/plot eigen].
```

For example, if you have five questions (called “q1”, “q2”, ..., in your dataset), you could type:

```
factor
/variables q1 to q5
/extraction paf
/method=correlation
/rotation varimax.
```

It's oftentimes useful to output a scree plot. To do so, simply add the /plot option to your syntax:

```
factor
/variables q1 to q5
/extraction paf
/method=correlation
/rotation varimax
/plot eigen.
```

Finally, it is sometimes useful to dictate how many factors should be extracted. To do so, add the /criteria option to your syntax (the following example dictates that one factor should be extracted).

Note that the /criteria option **must** come before the /extraction and /rotation options.

```
factor
/variables q1 to q5
/criteria factors(1)
/extraction paf
/method=correlation
/rotation varimax
/plot eigen.
```

Reading the Output

There are several tables (and potentially graphs) you'll want to pay attention to in your output. The first table is titled "Total Variance Explained."

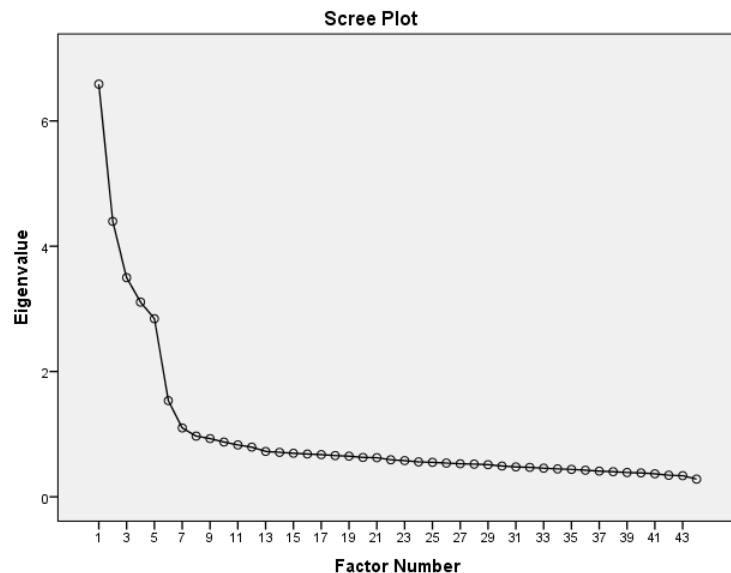
Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	6.587	14.971	14.971	6.054	13.759	13.759	4.118	9.359	9.359
2	4.397	9.993	24.964	3.875	8.807	22.566	3.542	8.050	17.410
3	3.498	7.951	32.915	2.985	6.784	29.350	3.538	8.041	25.450
4	3.108	7.063	39.979	2.533	5.758	35.108	3.128	7.108	32.558
5	2.843	6.460	46.439	2.294	5.213	40.320	2.859	6.498	39.057
6	1.535	3.489	49.928	.974	2.214	42.534	1.224	2.781	41.838
7	1.100	2.501	52.428	.574	1.304	43.838	.880	2.000	43.838
8	.969	2.202	54.630						
9	.929	2.112	56.743						
10	.874	1.987	58.730						

This table gives you information on how many factors (i.e., latent constructs) might underlie your data. The **eigenvalues** (contained in the second column) tell you *how many items* each factor fully explains. For example, in the above example, the first factor fully explains (i.e., gives as much information as) 6.59 items.

By convention, we only extract factors with eigenvalues greater than one (after all, if a factor explains less information than a single item, why do we care about it?).

Scree Plots

If you requested a scree plot, it can help you determine how many factors may be in the data. A scree plot visually demonstrates how much information each factor captures. You can use a scree plot to look for sharp drops in the amount of information factors are providing. Specifically, you want to look for a sharp "leveling" or "elbow" in the scree plot.



The example to the right suggests that there are anywhere between 5-7 factors in the data (notice the sharp drop in eigenvalues after the fifth factor and the severe leveling of the eigenvalues after the seventh factor).

Rotated Factor Matrix

The final table you'll want to look for in your output is the rotated factor matrix.

	Factor						
	1	2	3	4	5	6	7
talkative	.742	.070	-.059	.098	.054	.001	-.139
not_critical	.019	-.235	.028	.466	.013	.031	.210
thorough	.023	-.019	.654	.108	.095	-.024	-.106
depressed	-.272	.501	-.188	-.184	.057	.038	-.069
original	.181	-.076	.056	-.003	.719	.073	.080
not_reserved	.716	-.036	-.070	.017	-.005	.023	.160
helpful	.059	-.028	.168	.543	.139	-.045	-.086

This table tells you the extent to which each item *loads on* (i.e., correlates with) the underlying factor (i.e., construct). For example, in the above table, the item “talkative” loads (i.e., correlates) .74 on the first factor, .07 on the second factor, and so on.

Deciding on Number of Factors to Extract and Interpreting the Factors

Deciding on the number of factors to extract and how to interpret them is as much an art as it is a science. Generally, we never extract factors with eigenvalues less than or equal to one. Conversely, you would generally not exclude a factor well above a clear “elbow” in a scree plot. However, as in the above example, there is oftentimes ambiguity in the number of factors that should be extracted (e.g., the eigenvalues and scree plot for the above example suggest the presence of 5-7 underlying factors).

So, how do we decide whether a 5-, 6-, or 7- factor solution is most appropriate? The answer is simply *to try them all* and see what the results look like. As a scientist, you are then free to pick what you believe is the most interpretable solution. In a scientific paper, you would likely report the results of all three factor analyses.

Tips for Probing Factor Analyses

It can be very helpful to copy and paste your rotated factor matrix into Excel.

	A	B	C	D	E	F	G	H
1	item	1	2	3	4	5	6	7
2	talkative	0.74	0.07	-0.06	0.10	0.05	0.00	-0.14
3	not_critical	0.02	-0.24	0.03	0.47	0.01	0.03	0.21
4	thorough	0.02	-0.02	0.65	0.11	0.10	-0.02	-0.11
5	depressed	-0.27	0.50	-0.19	-0.18	0.06	0.04	-0.07
6	original	0.18	-0.08	0.06	0.00	0.72	0.07	0.08
7	not_reserved	0.72	-0.04	-0.07	0.02	-0.01	0.02	0.16

Once your factor matrix is in Excel, you can sort by the loadings on each factor (one at a time). It's oftentimes helpful to highlight cells with high loadings ($\geq \sim .40$) on given factors:

	A	B	C	D	E	F	G	H
1	item	1	2	3	4	5	6	7
2	not_quiet	0.82	-0.05	-0.02	-0.01	-0.02	0.02	0.11
3	outgoing	0.77	-0.11	0.03	0.20	0.04	0.03	-0.09
4	talkative	0.74	0.07	-0.06	0.10	0.05	0.00	-0.14
5	not_reserved	0.72	-0.04	-0.07	0.02	-0.01	0.02	0.16
6	not_shy	0.68	-0.19	0.06	-0.07	0.01	0.03	0.19
7	enthusiastic	0.60	-0.05	0.06	0.28	0.18	0.02	-0.11
8	energetic	0.55	-0.13	0.13	0.21	0.16	-0.04	-0.10
9	assertive	0.48	-0.10	0.17	-0.26	0.17	-0.04	-0.15
10	not_cold	0.29	-0.06	0.08	0.50	-0.08	0.02	0.28
11	original	0.18	-0.08	0.06	0.00	0.72	0.07	0.08
12	cooperative	0.16	-0.05	0.11	0.51	-0.01	0.02	-0.10
13	trusting	0.15	-0.09	0.03	0.41	-0.02	-0.04	-0.11
14	determined	0.12	-0.04	0.55	0.06	-0.02	-0.02	-0.06

In the above example, we can see a clear first factor that has to do with extraversion.

	A	B	C	D	E	F	G	H
1	item	1	2	3	4	5	6	7
2	not_rude	-0.08	-0.11	0.13	0.60	-0.01	0.06	0.29
3	not_cold	0.29	-0.06	0.08	0.50	-0.08	0.02	0.28
4	not_careless	-0.09	-0.03	0.56	0.12	-0.09	0.05	0.26
5	not_lazy	0.10	-0.10	0.60	0.06	0.02	-0.01	0.23
6	not_distractabl	-0.10	-0.21	0.50	-0.05	-0.08	0.03	0.23
7	not_critical	0.02	-0.24	0.03	0.47	0.01	0.03	0.21
8	not_relaxed	-0.01	0.74	-0.09	-0.14	-0.09	0.03	0.20
9	not_calm	0.05	0.58	-0.12	-0.09	-0.13	-0.01	0.19

In contrast, notice that the seventh factor has no items loading strongly on it. This clearly shows that a seven-factor solution is not appropriate. Thus, we would discard the seven-factor solution and try a six-factor one.

In addition to empty factors (as above), you will occasionally encounter "junk factors" that have random items that don't seem to conceptually go together loading on them. This suggests that you should reduce to a fewer number of factors, as well.

The "proper" solution to a factor analysis is subjective, but generally occurs when you (a) have clean factors (i.e., high loadings on all factors and low cross-loadings) and (b) you can successfully label and interpret all factors.