EXPLORATORY FACTOR ANALYSIS IN SPSS

Daniel Boduszek
d.boduszek@hud.ac.uk

University of
HUDDERSFIELD

Outline

- Theoretical Introduction to Exploratory Factor Analysis (EFA)

- How to run EFA in SPSS

- Interpreting Output of EFA in SPSS

Factor Analysis

- Researchers use factor analysis for two main purposes:

  - Development of psychometric measures (Exploratory Factor Analysis – EFA)

  - Validation of psychometric measures (Confirmatory Factor Analysis – CFA – cannot be done in SPSS, you have to use e.g., Amos or Mplus).

- This presentation will explain EFA in a straightforward, non-technical manner, and provide detailed instructions on how to carry out an EFA using the SPSS

© Can Stock Photo – csp3121687

Factor Analysis

- Researcher may be interested in a particular phenomenon – Criminal Social Identity (CSI).

- If you are interested in measuring this construct, you might develop a series of questions that reflect the features of CSI.

- These questions will likely be developed based upon your theoretical knowledge of the construct.

- But if you select a large group of items to tap into CSI, you may be left with an unwieldy amount of data.

Factor Anlaysis

- Factor analysis can be used to find meaningful patterns within a large amount of data.

- It's possible that you will find that a certain group of questions seem to cluster together.

- You might then infer that the first set of questions is tapping into one particular aspect of CSI (Centrality), while the other set of questions is tapping into a distinct aspect of CSI (e.g., In Group Affect).

- This process simplifies your data and allows for the development of a more parsimonious presentation of the data.

3 Stages

- The process of conducting an EFA involves three stages:

  - 1. Extraction

  - 2. Rotation

  - 3. Interpretation

- We will consider each one of these separately.

Extraction

□ Extraction simply refers to the process of determining how many factors best explains the observed covariation matrix within the data set.

□ Because we are scientists, we are concerned with issues of *parsimony*:

  ◘ We want the fewest number of factors that explains the largest amount of variation (among the observed variables).
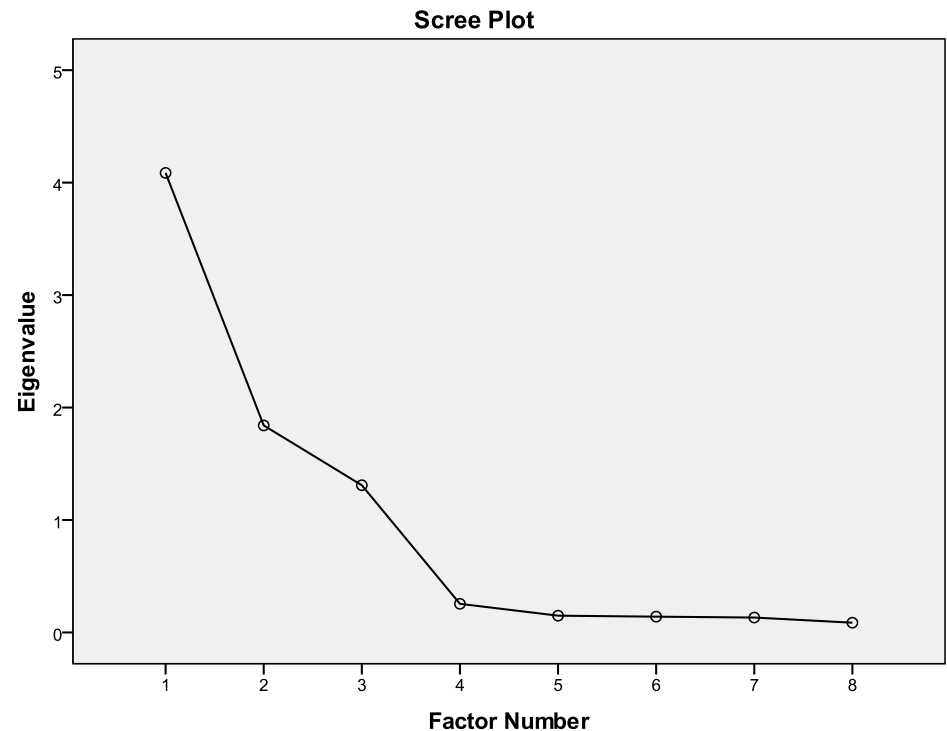
Extraction

- SPSS will extract as many factors as there are items in the data (in this case 8).

- By what method then can we determine the appropriate number of factors to Extract?

  - Eigenvalues

  - Scree plot

Scree Plot

Eigenvalue

- Remember, higher factor loadings suggest that more of the variance in that observed variable is attributable to the latent variable.

- An eigenvalue is simply the *sum of the squared factor loadings for a given factor.*

- So, because we have 8 indicators, we would check each indicator's factor loading for a given factor, square this value, and then add them all up.

- This gives us our eigenvalue for that factor.

Eigenvalue

- The eigenvalue for each factor tells us something about how much variance in the observed indicators is being explained by that latent factor.

- Because we are interested in explaining as much variance in observed indicators as possible, with the fewest latent factors, we can decide to retain only those latent factors with sufficiently high eigenvalues.

- The common practice is to retain factors that have eigenvalues above 1

>1

GREATER THAN ONE

Eigenvalue

- Is this an arbitrary figure? yes it is!

- Important therefore not to hold rigidly to this criterion.

- What if we were to find one factor with an eigenvalue of 1.005 and another with an eigenvalue of 0.999998?

- A rigid adherence to the "only eigenvalues above 1" rule may lead to some nonsensical decision.

- As always, your decisions should be theoretically determined, not statistically determined!
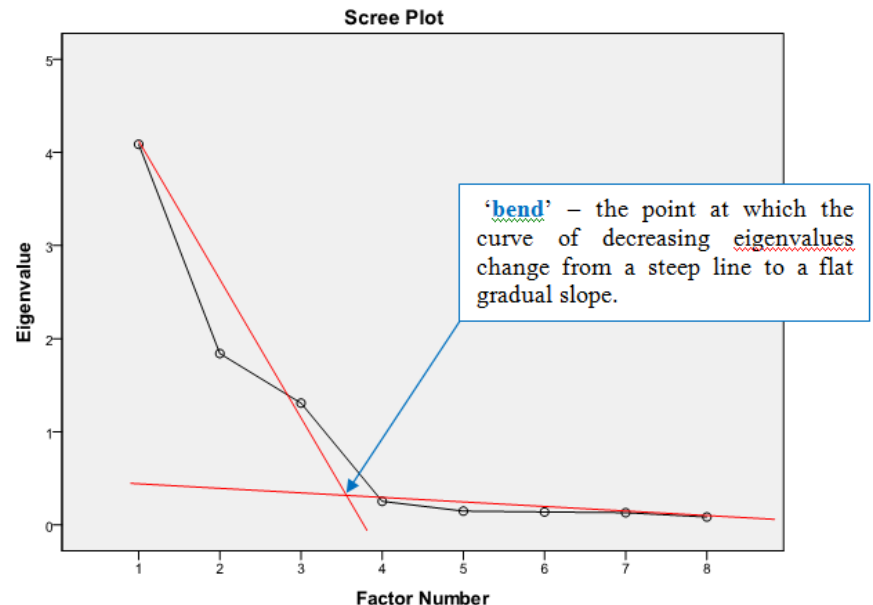
Scree Plot

- An alternative method of determining the appropriate number of factors to retain is to consider the relative size of the eigenvalues rather than the absolute size.

- One way of doing this is by inspecting a scree plot.

- Welcome to the realms of fuzzy science!

- Determining the appropriate number of factors to retain by inspecting a scree plot is subjective and open to different interpretations.
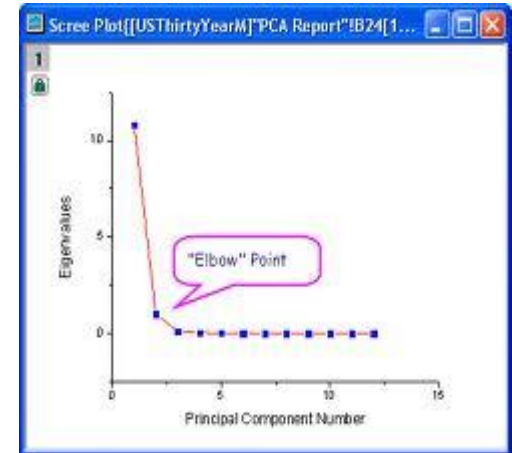
Scree Plot

- When we inspect a **scree plot** we fit a line through the all the "scree factors" and a line through the "mountain-side factors".

- This creates an elbow.

- We ignore the scree factors – meaningless as each factor explains far too little variance in the observed indicators to be of any explanatory value

- Focus only on the factors above the elbow point – factors which explain a sufficient degree of variance in the observed indicators.

Scree Plot

'**bend**' – the point at which the curve of decreasing eigenvalues change from a steep line to a flat gradual slope.

Scree Plot

- If you only extract one factor than your data set is best explained by a unidimensional solution and you don't need to worry about rotation.

- If you extract two or more factors, you must decide whether or not your factors are correlated.

- This requires a good deal of psychological knowledge.

- You will need to figure out, based on theory, whether your factors should be related or unrelated to each other.

Rotation

- Rotation then is a method that allows for the creation of a simple structure.

- Note that all the items in this example load onto all three factors – cross factor loadings.

- Rotation simply maximises the factor loadings for the items that best measure their respective factor.
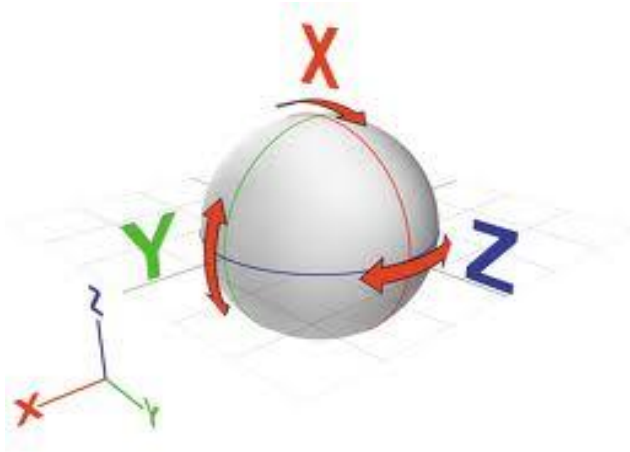
Rotation

- If you decide that your factors should be correlated then you select an *oblique solution* (e.g. direct oblimin) – Mplus will produce factor correlation values.

- If you decide that your factors should not be correlated then you select an *orthogonal solution* (e.g. varimax).

Limitations

- EFA can provide an infinite number of possible solutions.

- The method of determining the appropriate number of factors to retain is very subjective.

- EFA is also a highly data-driven rather than a theory-driven method of investigation.

Conclusion

- Carrying out EFA involves three stages:

  - **1. Extraction** – determination of the appropriate number of factors. Concerned with parsimonious solutions. The number of factors to be retained can be based on eigenvalues and/or the inspection of a scree-plot.

  - **2. Rotation** – Specifying the nature of relationship between the factors. Factors can be deemed to be correlated (oblique) or uncorrelated (orthogonal).

  - **3. Interpretation** – Naming the factors. Using your psychological knowledge to provide a meaningful understanding of the common feature among the relevant items.

SPSS PROCEDURE

Data Used

- In this analysis we are using the responses from 312 prisoners on the **Measure of Criminal Social Identity (Boduszek et al., 2012).** The scale is comprised of 8 items designed to measure social identity as criminal. There have been different factor analytic solutions reported for social identification. Some authors suggest that the scale measures one factor, some suggested two factor solution, whereas others have stated that it measures three corrected factors, which are cognitive centrality, in-group affect and in-group ties.
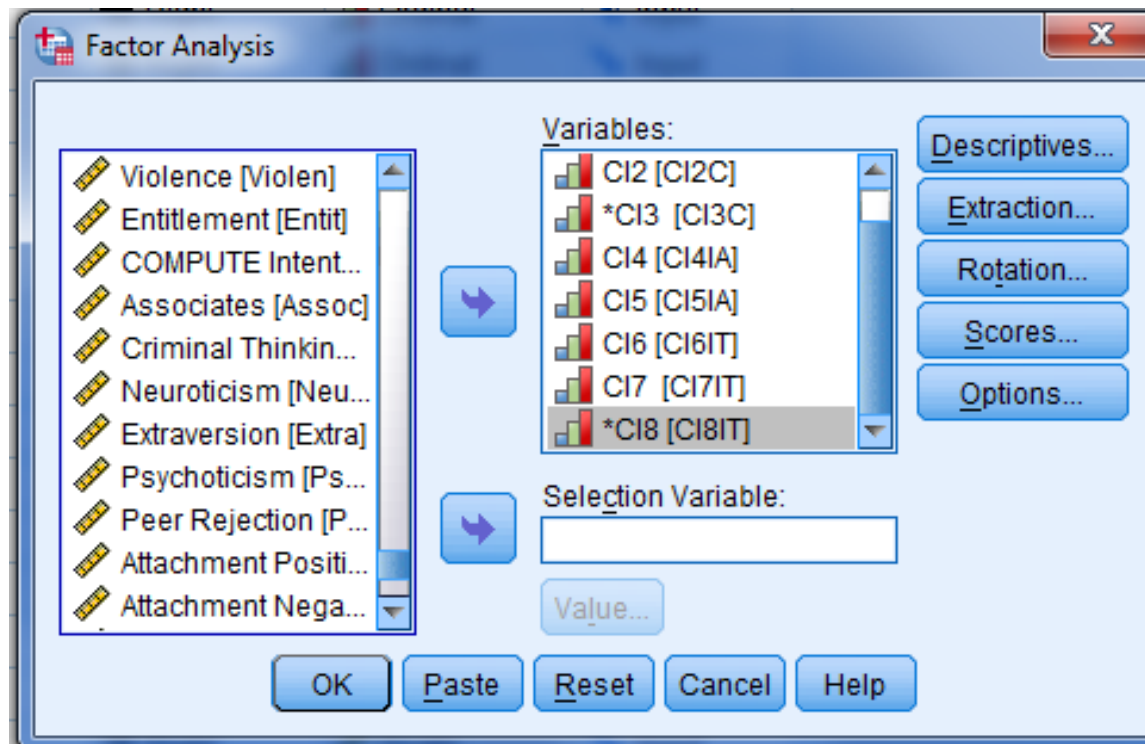
SPSS Procedure

□ Go to the **Analysis**, then select **Dimension Reduction**, and then **Factor** as shown in the graphic below.

SPSS Procedure

☐ To identify which items you want to use in the analysis, highlight the name of each item from the list on the left-hand side and use the button 🔹 to move it across to the **Variables box**

SPSS Procedure

- When you click on **Extraction** a new window will appear. This allows you to decide on how the factors are to be extracted and how many factors to retain.

- Select
  - **Method** (Maximum likelihood)
  - **Display** (Unrotated factor solution and Scree plot)
  - **Extract**
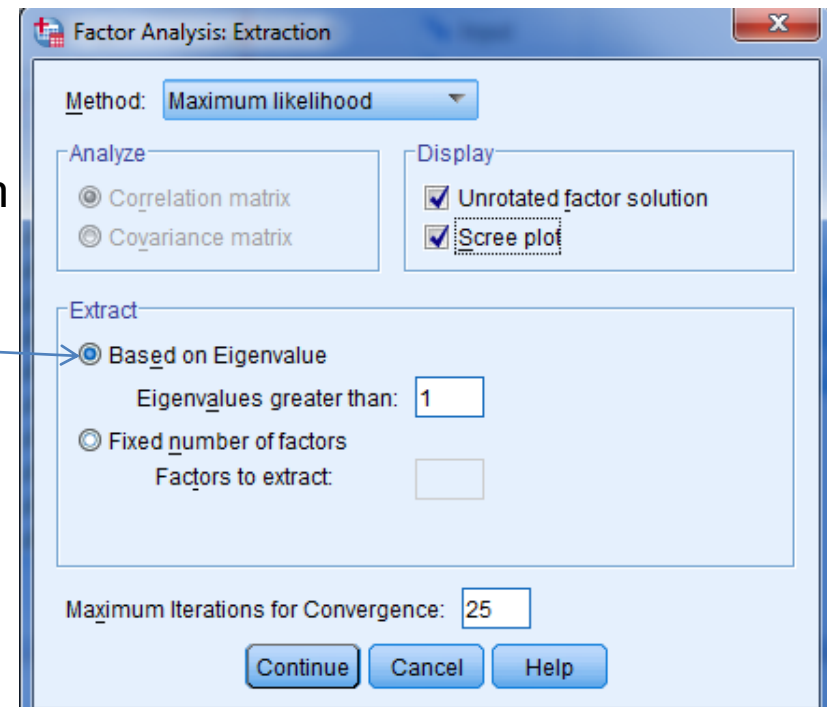
SPSS Procedure

- **Maximum likelihood** allows the test of a solution using a chi-square statistic.

- **Extract** – you can either use a statistical criterion by allowing factors with eigenvalues greater than 1 to be retained (this is the default option and this option is used in the current example), or you can use previous research and theory to guide your decision on how many factors there are.

  - For example, if you were factoring the EPQ you would click the Number of factors circle and specify 3 (or 4 if the Lie Scale is included)

SPSS Procedure

- Now you want to specify whether the factors that are extracted are correlated (oblique solution) or uncorrelated (orthogonal). This is done using the **Rotation button.**

- If correlated

- If not correlated

- In **Display**

- Click **Continue**
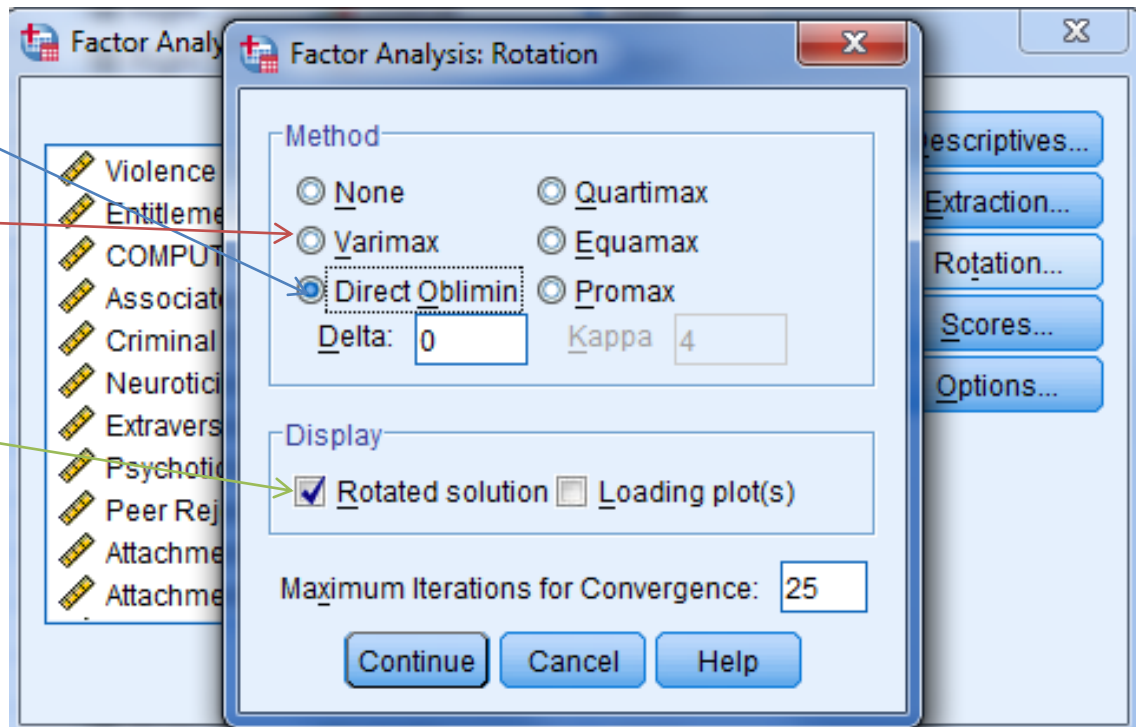
SPSS Procedure

- Click on the **Descriptives**

- **Select**

  - **Univariate descriptives**
  - **Initial solution**
  - **Coefficients** and **Significance levels** (which will provide you with the matrix of correlations)

  - **Continue**
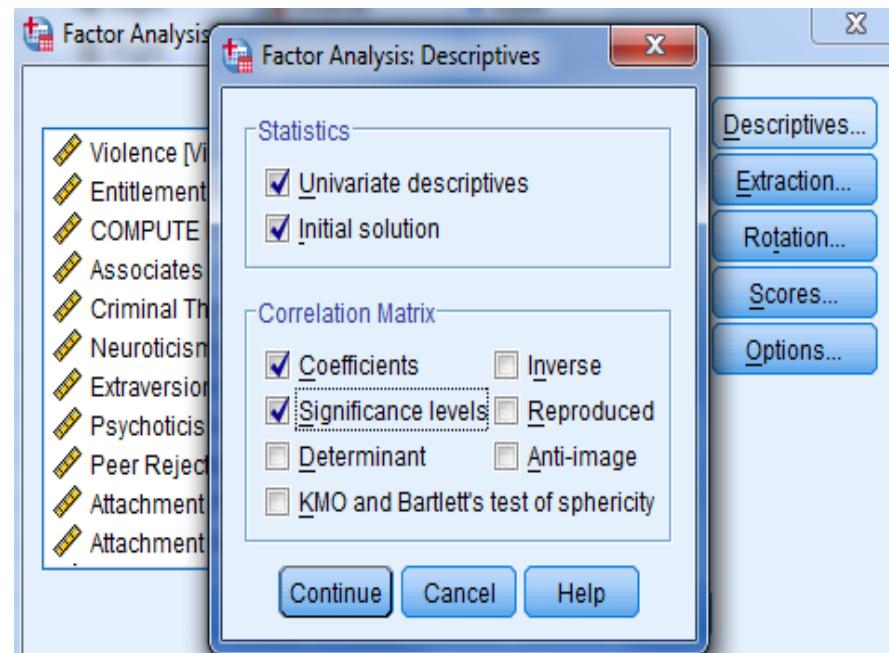
SPSS Procedure

- **Options** button allows you to tell SPSS how to deal with missing values and also how to structure the output

In the **Missing Values** select **Exclude cases listwise**

The **Exclude case pairwise** allows the calculation of each correlation with the maximum number of cases possible. Pairwise deletion and mean substitution (**Replace with mean**) are not recommended.

In the **Coefficient Display Format** select **Sorted by size** and **Suppress small coefficients** - factor loadings below a specified magnitude (default is .10) will not be printed

Continue
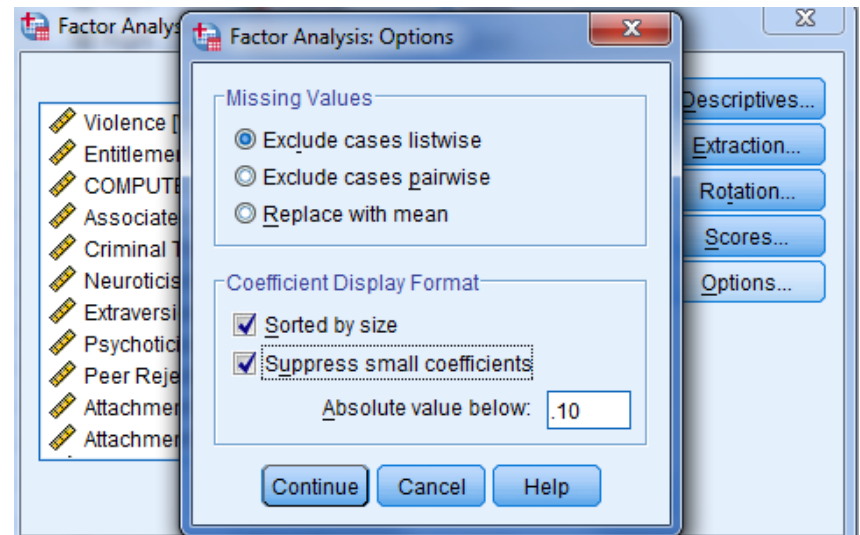
SPSS Procedure

- **Scores** option

- You can calculate and save factor score(s) for each case by checking the **Save as variable** box.

- You can print the coefficients by which variables are multiplied to obtain factor scores show the correlations between factor scores by checking the **Display factor score coefficient matrix**.

- **Continue** then click **OK** to run factor analysis
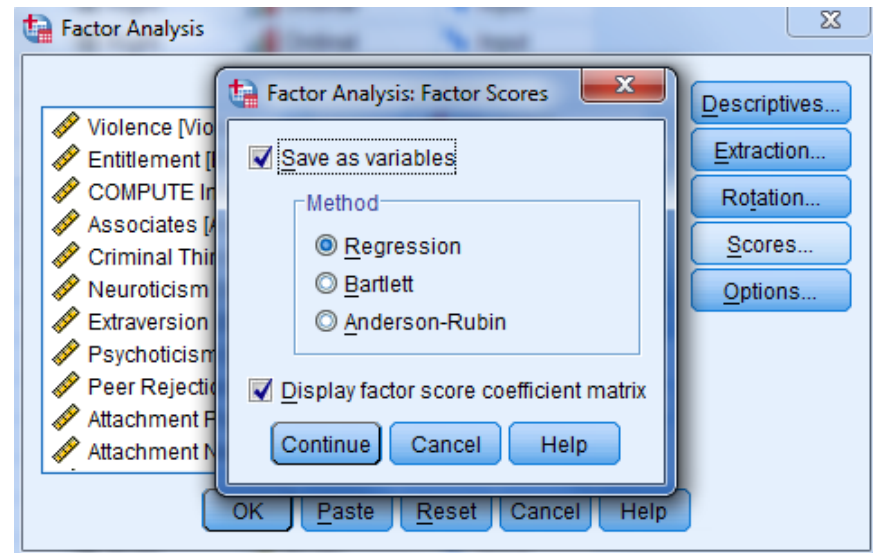
Things to Check

Descriptives – We have requested the **univariate descriptives** to check for any irregularities in the data. We have also asked for the **initial solution** to be reported. We have not asked for the correlation matrix and significance level to be printed (although this information is useful it occupies a lot of space).

Extraction – We have chosen **maximum likelihood** as the method of extraction as it has many desirable statistical properties. As there is no agreement in the literature about how many factors the scale measures, we have asked for **factors with eigenvalues greater than 1** to be retained, rather than specifying the number ourselves. We have asked for a **Scree plot** to be provided to help us determine the number of meaningful factors.

Rotation – If there is more than one factor, we suspect that they will be correlated. Previous research suggests that the scale measures the factors of 3 different aspects of identity. It is unlikely that these will be independent. For this reason we have selected the **Direct oblimin** (oblique) rotation that allows the factors to be correlated. We have also asked the **rotated solution** to be reported.

Scores – We have asked for factor scores to be calculated and saved.

Options – To help interpretation we have asked the factor loadings to be ordered by size and factor loadings less that 0.10 to be omitted from the output.

Output for EFA

The means for each of the items appear to be reasonable as each of the items is measured on a 5-point Likert scale. No values are above 5 or below 1. The standard deviations are all similar suggesting that there are no outliers for any of the items.

The 'Analysis N' shows the number of valid cases. Here there are 9 missing values because the entire sample included 312 prisoners.

Descriptive Statistics

	Mean	Std. Deviation	Analysis N
*CI1	2.9307	1.14733	303
CI2	2.8218	1.17419	303
*CI3	2.9439	1.18775	303
CI4	1.9835	1.12603	303
CI5	2.0627	1.09455	303
CI6	2.9769	1.20819	303
CI7	2.6832	1.10625	303
*CI8	3.0099	1.03733	303

Output for EFA

Communalities can be thought of as the R^2 for each of the variables that have been included in the analysis using the factors as IV's and the item as a DV.

It represents the proportion of variance of each item that is explained by the factors. This is calculated of the initial solution and then after extraction.

These are reported in the **Initial** and

Extraction

Communalities

	Initial	Extraction
*CI1	.806	.843
CI2	.871	.937
*CI3	.832	.872
CI4	.741	.927
CI5	.748	.800
CI6	.765	.835
CI7	.783	.885
*CI8	.645	.696

Extraction Method: Maximum Likelihood.

Output for EFA

Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	4.086	51.080	51.080	3.839	47.987	47.987
2	1.841	23.010	74.090	1.654	20.671	68.659
3	1.309	16.365	90.455	1.302	16.276	84.935
4	.255	3.182	93.637			
5	.149	1.865	95.502			
6	.140	1.755	97.257			
7	.133	1.658	98.915			
8	.087	1.085	100.000			

Extraction Method: Maximum Likelihood.
a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.

The **Initial Eigenvalues -** first 3 factors are meaningful as they have **Eigenvalues > 1.** Factors 1, 2 and 3 explain 51.08%, 23.01%, and 16.37% of the variance respectively – a cumulative total of **90.46%** (total acceptable). The **Extraction Sums of Squared Loadings** provides similar information based only on the extracted factors

Output for EFA

- This plot shows that there are three relatively high (factors 1, 2, and 3) eigenvalues. Retain factors that are above the 'bend' – the point at which the curve of decreasing eigenvalues change from a steep line to a flat gradual slope

Scree Plot

'bend' – the point at which the curve of decreasing eigenvalues change from a steep line to a flat gradual slope.
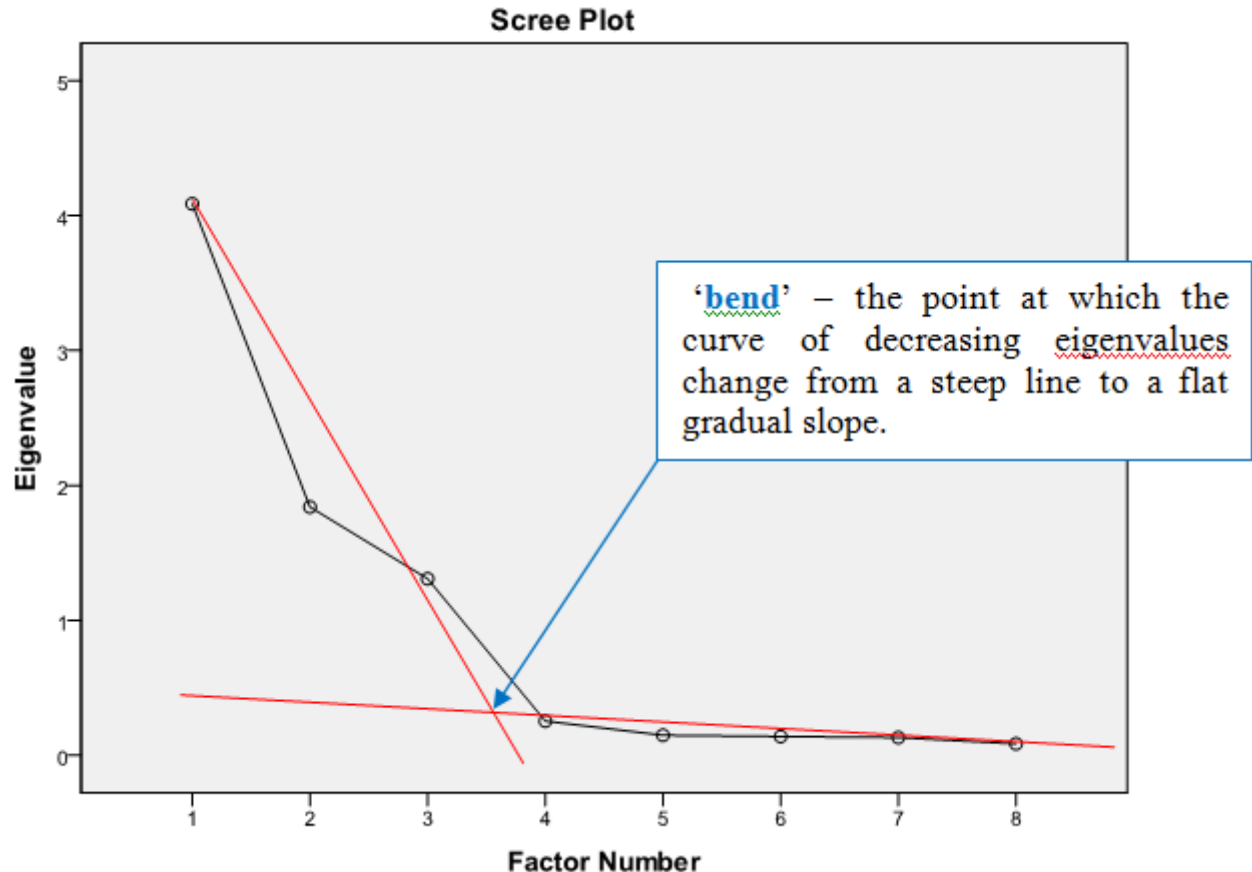
Output for EFA

Factor Matrix^a

	Factor		
	1	2	3
CI2	.876	-.408	
*CI3	.823	-.441	
*CI1	.800	-.450	
CI5	.617	.533	-.367
CI4	.617	.578	-.461
CI7	.611	.421	.578
CI6	.602	.418	.545
*CI8	.506	.348	.565

Extraction Method: Maximum Likelihood.
a. 3 factors extracted. 6 iterations required.

The **Factor Matrix** represents information from initial unrotated solution. The values are weights that relate the item (or variable) to the respective factor. All the items have high(ish) positive weights with the first factor.

At this stage the solution has not taken into consideration the correlation between the three factors. Subsequent information is more readily interpretable.

Output for EFA

Goodness-of-fit Test

Chi-Square	df	Sig.
5.241	7	.631

The **Goodness-of-fit Test** determines if the sample data (correlations) are likely to arisen from three correlated factors. In this situation we want the probability value of the Chi-Square statistic to be greater than the chosen alpha (generally 0.05). Based on our results the three factor model is a good description of the data.

Output for EFA

Pattern Matrix^a

	Factor		
	1	2	3
CI2	.950		
*CI3	.938		
*CI1	.928		
CI4		.979	
CI5		.872	
CI7			.930
CI6			.893
*CI8			.855

Extraction Method: Maximum Likelihood.
Rotation Method: Oblimin with Kaiser Normalization.
a. Rotation converged in 5 iterations.

The **Pattern Matrix** shows the factor loadings for the rotated solution. Factor loadings are similar to regression weights (or slopes) and indicate the strength of the association between the variables and the factors. The solution has been rotated to achieve an interpretable structure.

When the factors are uncorrelated the Pattern Matrix and the Structure Matrix should be the same.

Output for EFA

Structure Matrix

	Factor		
	1	2	3
CI2	.967	.361	.343
*CI3	.934	.291	.314
*CI1	.918	.264	.306
CI4	.293	.962	.354
CI5	.311	.893	.395
CI7	.329	.396	.940
CI6	.324	.404	.913
*CI8	.270	.289	.833

Extraction Method: Maximum Likelihood.
Rotation Method: Oblimin with Kaiser Normalization.

The **Structure Matrix** shows the correlations between the factors and the items for the rotated solution.

Output for EFA

Factor Correlation Matrix

Factor	1	2	3
1	1.000	.323	.341
2	.323	1.000	.400
3	.341	.400	1.000

Extraction Method: Maximum Likelihood.
Rotation Method: Oblimin with Kaiser Normalization.

The **Factor Correlation Matrix** shows that factor 1, 2 and 3 are statistically correlated.

THANK YOU FOR YOUR TIME!

d.boduszek@hud.ac.uk

University of
HUDDERSFIELD