

Continue



Spss factor analysis interpretation

Given text: paraphrase this text: This seminar is the first part of a two-part seminar that introduces central concepts in factor analysis. Part 1 focuses on exploratory factor analysis (EFA). Although the implementation is in SPSS, the ideas carry over to any software program. Part 2 introduces confirmatory factor analysis (CFA). Please refer to A Practical Introduction to Factor Analysis: Confirmatory Factor Analysis. 1. Exploratory Factor Analysis Introduction Motivating example: The SAQ Pearson correlation formula Partitioning the variance in factor analysis Extracting factors principal components analysis common factor analysis principal axis factoring maximum likelihood Rotation methods Simple Structure Orthogonal rotation (Varimax) Oblique (Direct Oblimin) Generating factor scores Back to Launch Page Introduction Suppose you are conducting a survey and you want to know whether the items in the survey have similar patterns of responses, do these items "hang together" to create a construct? The basic assumption of factor analysis is that for a collection of observed variables there are a set of underlying variables called factors (smaller than the observed variables), that can explain the interrelationships among those variables. Let's say you conduct a survey and collect responses about people's anxiety about using SPSS. Do all these items actually measure what we call "SPSS Anxiety"? Motivating Example: The SAQ (SPSS Anxiety Questionnaire) Let's proceed with our hypothetical example of the survey which Andy Field terms the SPSS Anxiety Questionnaire. For simplicity, we will use the so-called "SAQ-8" which consists of the first eight items in the SAQ. Click on the preceding hyperlinks to download the SPSS version of both files. The SAQ-8 consists of the following questions: Statistics makes me cry My friends will think I'm stupid for not being able to cope with SPSS Standard deviations excite me I dream that Pearson is attacking me with correlation coefficients I don't understand statistics I have little experience of computers All computers hate me I have never been good at mathematics Pearson Correlation of the SAQ-8 Let's get the table of correlations in SPSS Analyze - Correlate - Bivariate: Correlations 1 2 3 4 5 6 7 8 1 2 -.099** 1 3 -.337** .318** 1 4 .436** -.112** -.380** 1 5 .402** -.119** -.310** .401** 1 6 .217** -.074** -.227** .278** .257** 1 7 .305** -.159** -.382** .409** .339** .514** 1 8 .331** -.050* -.259** .349** .269** .223** .297** 1 **. Correlation is significant at the 0.01 level (2-tailed). *. Correlation is significant at the 0.05 level (2-tailed). From this table we can see that most items have some correlation with each other ranging from $V(r=-0.382)$ for Items 3 and 7 to $V(r=.514)$ for Items 6 and 7. Due to relatively high correlations among items, this would be a good candidate for factor analysis. Recall that the goal of factor analysis is to model the interrelationships between items with fewer (latent) variables. These interrelationships can be broken up into multiple components Partitioning the variance in factor analysis Since the goal of factor analysis is to model the interrelationships among items, we focus primarily on the variance and Factor analysis assumes that variance can be split into two types: shared and individual. Shared variance is what items have in common; similar questions will share a lot of it. Communality measures this shared variance, ranging from 0 to 1, with higher numbers indicating more explained variance for an item. Individual variance includes specific and error variances. Specific variance is unique to one item, like anxiety about computers or SPSS. Error variance comes from measurement errors or unexplained factors. The total variance is made up of common and individual variances, where individual variance includes specific and error variances. This concept helps in understanding the SAQ-8 factor model, where SPSS Anxiety explains shared variance across items, but each item also has unique specific variance and error variance. For example, while SPSS Anxiety might explain part of the variance for "Computers are useful only for playing games," there could be systematic or non-systemic factors like technophobia or getting a speeding ticket that aren't explained by SPSS Anxiety or technophobia. Now that we understand how to partition variance, we can move on to performing our first factor analysis. This process involves reducing variables and interpreting results through two steps: factor extraction and rotation. Factor extraction decides the type of model and number of factors to extract, while factor rotation aims for a simpler structure to improve interpretation. There are two main approaches to factor extraction based on variance partitioning: principal components analysis (PCA) and common factor analysis. Unlike factor analysis, PCA assumes there's no unique variance; all variance is shared. This means the total variance equals the shared variance, making it equal to communality if the total variance is 1. PCA can be run by considering this assumption and aiming for a model where every variable contributes equally in explaining the overall variance. To replicate the correlation matrix using a set of components that are fewer in number and linear combinations of the original set of items, we will extract as many components as possible from the SPSS data. Firstly, go to Analyze - Dimension Reduction - Factor. Move all observed variables into the Variables box to be analyzed. Under Extraction - Method, select Principal components and ensure that the Correlation matrix is analyzed. Also, request the Unrotated factor solution and Scree plot. Under Extract, choose Fixed number of factors, and enter 8 as the number of factors to extract. Increase the Maximum Iterations of Convergence to 100 for better results. The equivalent SPSS syntax is: FACTOR /VARIABLES q01 q02 q03 q04 q05 q06 q07 q08 /MISSING LISTWISE /ANALYSIS q01 q02 q03 q04 q05 q06 q07 q08 /PRINT INITIAL EXTRACTION /PLOT EIGEN /CRITERIA FACTORS(8) ITERATE(100) /EXTRACTION PC /ROTATION NOROTATE /METHOD=CORRELATION. Eigenvalues and Eigenvectors: Eigenvalues represent the total variance explained by a principal component. They can be positive or negative, but in practice, they explain only positive variance. Eigenvector times the square root of eigenvalue gives component loadings, which are correlations between each item and the principal component. For example, Item 1 has an eigenvector of 0.377 and eigenvalue of 3.057, resulting in a correlation of 0.659 with the first component. **Understanding PCA** Principal Component Analysis (PCA) is a technique used to reduce the dimensionality of a dataset while retaining most of its information. The process involves extracting components, which are linear combinations of the original variables, and then selecting the optimal number of components based on certain criteria. **Variance Explained** In a PCA, each component explains a portion of the total variance in the data. The sum of the squared loadings for all components equals 1 or 100%, known as the communality. By adding up the squared loadings cumulatively, we can see how much variance is explained by each component. **Calculating Eigenvalues** To calculate the eigenvalue (total variance explained) for a component, we sum the squares of its loadings. For example, the first eigenvalue is calculated as: $0.659^2 + (-0.300)^2 + (-0.653)^2 + \dots = 3.057$. **Choosing the Number of Components** The goal of PCA is to reduce the number of variables while retaining most of the information. One criterion for selecting the optimal number of components is to choose those with eigenvalues greater than 1. This can be confirmed using a Scree Plot, which plots the eigenvalue against the component number. **Interpreting Results** The Total Variance Explained table shows how much variance each component explains. The first component typically explains the most variance, and subsequent components explain less. By selecting only the top components with eigenvalues greater than 1, we can reduce the dimensionality of our dataset while retaining most of its information. **Common Terminology** **Component**: a linear combination of the original variables * **Variance explained**: the amount of variance in the data explained by each component * **Eigenvalue**: a measure of the total variance explained by a component (equal to 1 or 100% when all components are summed) * **Communality**: the sum of the squared loadings for all components (equals 1 or 100%) * **Scree Plot**: a plot that helps select the optimal number of components based on their eigenvalues. Given article text here The final answer to the user's question about PCA component extraction is as follows: To determine the optimal number of components for a Principal Component Analysis (PCA), one should examine the Scree Plot, which displays the total variance explained by each component. The first component typically has the highest total variance and the last component has the least. However, the key is to identify where the largest drop occurs, often referred to as the "elbow" joint. The interpretation of the scree plot can vary, but a common guideline is to select the number of components to the left of the elbow point. This may result in selecting only one component, or choosing between 1 and 4 components based on additional evidence. Some criteria suggest that the total variance explained should be between 70% to 80%. In this case, let's assume a two-component solution is acceptable, as suggested by the principal investigator. To implement this, simply enter "2" under Fixed number of factors - Factors to extract in SPSS. The main differences between the eight and two-component solutions are: * The total variance explained percentage * The cumulative percent variance * The Component Matrix, which displays correlations between items and components In a two-component solution, the Component Matrix shows only two columns, whereas an eight-component solution has eight columns. The Component Matrix in PCA can be seen as correlations, while the Total Variance Explained table represents (R^2) . 1. T, 2. F (sum of squared loadings), 3. T Communality of the 2-component PCA The communality is the sum of the squared component loadings up to the number of components extracted. In the SPSS output, a table of communalities shows initial extraction values. Communalities Initial Extraction 1 1.000 0.453 2 1.000 0.840 3 1.000 0.594 4 1.000 0.532 5 1.000 0.431 6 1.000 0.361 7 1.000 0.517 8 1.000 0.394 Extraction Method: Principal Component Analysis. Since PCA is an iterative process, it starts with 1 as an initial estimate of the communality, which represents total variance across all components. The extraction method proceeds until a final communality is extracted. Notice that the extraction column in the communalities table has fewer values than the initial column because only two components were extracted. When dealing with a dataset comprising smaller components, Principal Component Analysis (PCA) is often the preferred approach. However, if you suspect that there exists a latent construct underlying the interrelationships among items, factor analysis might be more suitable. In this case, we assume that a construct called SPSS Anxiety explains why certain correlations exist among items on the SAQ-8. We acknowledge that SPSS Anxiety cannot fully account for all shared variance, so we model unique variance as well. Based on PCA results, we will begin with a two-factor extraction. To run factor analysis, follow the same steps as PCA (Analyze - Dimension Reduction - Factor), but under Method, select Principal Axis Factoring instead of PCA. Note that we set Maximum Iterations for Convergence at 100 and will discuss this later. The syntax for factor analysis is: FACTOR /VARIABLES q01 q02 q03 q04 q05 q06 q07 q08 /MISSING LISTWISE /ANALYSIS q01 q02 q03 q04 q05 q06 q07 q08 /PRINT INITIAL EXTRACTION /PLOT EIGEN /CRITERIA FACTORS(2) ITERATE(100) /EXTRACTION PAF /ROTATION NOROTATE /METHOD=CORRELATION. The output includes three tables: Communalities, Total Variance Explained, and Factor Matrix. Let's review each of these and compare them to PCA output. Communalities in the 2-factor PAF show that the initial extraction is no longer one, unlike PCA. Instead, Principal Axis Factoring uses the squared multiple correlation coefficient (R^2) as the initial guess. To demonstrate this, run a linear regression where Item 1 is the dependent variable and Items 2-8 are independent variables. The output from this analysis shows that the R Square value matches the initial communality estimate for Item 1. Factor analysis also uses an iterative estimation process to obtain final estimates under the Extraction column. Finally, summing all rows in the extraction column yields a total common variance of 3.00. The analysis continues with a focus on Total Variance Explained, highlighting similarities and differences between the two-factor solution and Principal Component Analysis (PCA). Notably, both the initial eigenvalues and factor matrix reveal eight rows for each "factor," indicating that SPSS relies on PCA's results. The primary distinction lies in extraction sums of squares loadings, where corresponding rows display lower values, reflecting a separation into common and unique variance. The percentages of explained variance differ significantly between factors, with Factor 1 accounting for approximately 31.38% of total variance and Factor 2 contributing around 6.24%. This mirrors the trend observed in PCA, where more extracted factors result in less variance explanation per successive factor. The correlation between each item and its corresponding factor is represented by loadings, similar to PCA. The total variance explained by each factor can be calculated by squaring each loading and summing down the rows. This process yields the elements of the Total Variance Explained table. Additionally, communality estimates for each item can be obtained by summing the squared loadings across factors. This result matches the Communalities table. The relationship between the Factor Matrix, Communalities table, and Total Variance Explained table can be visualized as follows: Squaring elements in the Factor Matrix gives squared loadings, which represent the proportion of variance explained by each factor for each item. Summing these across factors yields communality estimates. Summing down the rows of the Factor Matrix provides Sums of Squared Loadings or eigenvalues, which are used to fill the Total Variance Explained table. Finally, summing down all items in the Communalities table also results in the total common variance explained, matching the outcome from the Total Variance Explained table, is equivalent to summing eigenvalues or Sums of Squared Loadings down all factors under the Extraction column of the Total Variance Explained table. True or False (assuming a two-factor Principal Axis Factor solution with 8 items) The elements in the Factor Matrix represent correlations of each item with a factor, where each squared element of Item 1 represents communality. Summing down all 8 items within Factor 1 equals the first Sums of Squared Loading under the Extraction column of Total Variance Explained table. Summing down all 8 items in the Communalities table gives us the total common variance explained by both factors. The total common variance explained is obtained by summing all eigenvalues from the Initial column of the Total Variance Explained table. The total Sums of Squared Loadings in the Extraction column under the Total Variance Explained table represents the total variance, which consists of total common variance plus unique variance. In common factor analysis, the sum of squared loadings is the eigenvalue. Answers: 1. True, 2. False, 3. True, 4. True, 5. False, 6. False, 7. False. Maximum Likelihood Estimation (2-factor ML) This non-technical introduction to factor analysis doesn't delve into the differences between Principal Axis Factoring and Maximum Likelihood. The main concept is that Maximum Likelihood also assumes common factor analysis using R^2 to obtain initial estimates of communalities, but uses a different iterative process to achieve extraction solution. To run factor analysis using maximum likelihood estimation under Analyze - Dimension Reduction - Factor - Extraction - Method choose Maximum Likelihood. Although the initial communalities are the same between PAF and ML, the final extraction loadings will differ, resulting in distinct Communalities, Total Variance Explained, and Factor Matrix tables (although Initial columns will overlap). The main difference is that you obtain a Goodness-of-fit Test table, giving you an absolute test of model fit. Non-significant values suggest a good-fitting model. 1. No solution means there's no viable answer, while N/A indicates not applicable. 2. In SPSS, extracting 7 to 9 factors leads to an error due to negative degrees of freedom (unacceptable in statistics). Eight factors aren't even a possibility because it would trigger a warning stating that you can't request more than the available variables; instead, the number of extracted factors is reduced by one. 3. After understanding the table, we looked for the threshold at which the absolute fit becomes non-significant. This happens when the p-value drops to insignificance at three factors, differing from eigenvalues greater than 1 criteria, which chose two factors based on percentage variance explained - a preference for four or five factors wasn't reached. 4. The Principal Investigator was consulted, and at this point, two-factor solution is preferred. It's understood that there isn't one 'right' answer in selecting the best factor model; it all depends on your theory. 5. We will explore interpreting the factor loadings when discussing factor rotation for further guidance on choosing the correct number of factors. 6. Factors: - Chi-square Df p-value Iterations needed 1: 553.08 20