EDUC 644

Exploratory Factor Analysis

Mark J. Van Ryzin

Base Group Check-In

- How prepared to you feel for today's class?
  - Rate on a scale of 1 (lowest) to 10 (highest)
- What were three things that you recall from last week's class?
- What questions do you have from last week's class?
  - Did you find the answers, or do you still have questions?

"Factor Analysis"

- Encompasses many different methods that have a general goal of identifying 'factors' that account for relationships among observed variables
- Common applications:
  - Data reduction: Principal Components Analysis (PCA)
  - Scale construction: Exploratory Factor Analysis (EFA)
  - Hypothesis testing: Confirmatory Factor Analysis (CFA)
- Terms are often used interchangeably and inconsistently leading to confusion and incorrect application

PCA and EFA

- PCA and EFA are the most generally used methods of data reduction
- These techniques are designed to statistically reduce a larger number of related variables to a smaller number of "components" or "factors"
- Patterns of correlations are identified and grouped in order to either to create linear composites (PCA) or to estimate latent factors (EFA)

EFA

- Unlike PCA, it separates shared variance (or "common variance") from error variance (or "unique variance")
  - EFA generates more reliable solutions that are not sample-specific

- The model includes a residual term, *U*
  $F_1 = w_{11}X_1 + w_{12}X_2 + \ldots + w_{1p}X_p + U$

- The addition of a residual term enables the separation of common and unique variance

EFA

- <u>Communality</u>: proportion of a variables' variance that is explained by common factors
- Variables with high communality share more in common with the rest of the variables
  - If X_j is informative, communality is high (i.e., closer to 1)
  - If X_j is not informative, uniqueness is high (i.e., closer to 0)
  - Generally want communality to be > .20

PCA vs. EFA Extraction

Communalities

	Initial	Extraction
Oral Reading Fluency Fall Score	1.000	.775
Oral Reading Fluency Winter Score	1.000	.761
Oral Reading Fluency Spring Score	1.000	.728
Vocabulary Fall Score	1.000	.637
Vocabulary Spring Score	1.000	.670
Reading Comprehension Fall Score	1.000	.581
Reading Comprehension Winter Score	1.000	.585
Reading Comprehension Spring Score	1.000	.647

Extraction Method: Principal Component Analysis.

Communalities

	Initial	Extraction
Oral Reading Fluency Fall Score	.884	.763
Oral Reading Fluency Winter Score	.886	.745
Oral Reading Fluency Spring Score	.823	.699
Vocabulary Fall Score	.625	.578
Vocabulary Spring Score	.661	.617
Reading Comprehension Fall Score	.529	.514
Reading Comprehension Winter Score	.540	.518
Reading Comprehension Spring Score	.597	.589

Extraction Method: Principal Axis Factoring.

EFA vs. PCA

- By modeling different sources of variance, PCA produces components while EFA estimates factors
  - Components are aggregates or summaries of variables whereas factors are conceived as the “cause” or the mechanism underlying variable scores

- PCA is thus simpler, but EFA tends to more closely represent the intended application in most research contexts (i.e., scale construction)
  - By not accounting for error variance, the estimates of communalities and factor loadings in PCA are inflated (thus the emphasis on the *relative contribution* of each variable)

EFA Extraction Methods

- Principal axis factoring (PAF)
  - Initial communalities are used as starting values for the estimation
  - Iterates until the change between initial and extracted communalities is negligible

- Maximum Likelihood (ML)
  - Computationally intensive method for estimating loadings that maximize the likelihood (probability) of the correlation matrix
  - Assumes multivariate normal distribution and requires large N

Factor Retention

- In EFA, the goal is to discover the number and nature of latent variables that explain the covariation in a set of measured variables

- Should thus use retention criteria that enables the identification of interpretable factors or constructs

- For balance, should also consider Kaiser’s rule (eigenvalues > 1), a scree plot, and the practical significance of the factors

Factor Rotation

- The unrotated solution maximizes the variance explained by each factor, conditional on previous factors
- The result is a factor solution where the majority of items or variables load on/correlate highly with the first factor
- However, if the total explained variation can be more evenly distributed among the factor set, a sharper distinction between factors may emerge

Factor Rotation

- The goal of the rotation is to achieve "simple structure"
  - Simple structure is achieved when variables load highly only one factor and enable a straightforward interpretation of the factor set
- Factor loadings change (and, consequently, the meaning of the factors), but the amount of explained variance does not change
  - The result can be a reduction in complex factor loadings and a more useful factor solution

Types of Rotation

- Orthogonal (rigid) rotation: The factors remain uncorrelated (i.e., a right angle between functions is maintained)
  - May not be the best representation of reality, but interpretation is straightforward

- Oblique rotation: The factors are allowed to correlate (i.e., a right angle between functions may or may not be maintained)
  - May provide a better representation of reality, but makes interpretation more complex
  - An oblique rotation is used when there is a theoretical expectation that factors will be correlated

Orthogonal Rotation

- Varimax: most popular ("clean up" the factors)
  - Redistributes the variance from first (largest) factor to secondary factors
  - Total explained variance remains the same, but the percentage of variance explained by specific factors changes
  - Makes large loadings larger and small loadings smaller making interpretation of factors easier
  - Unique variance remains the same

Oblique Rotation

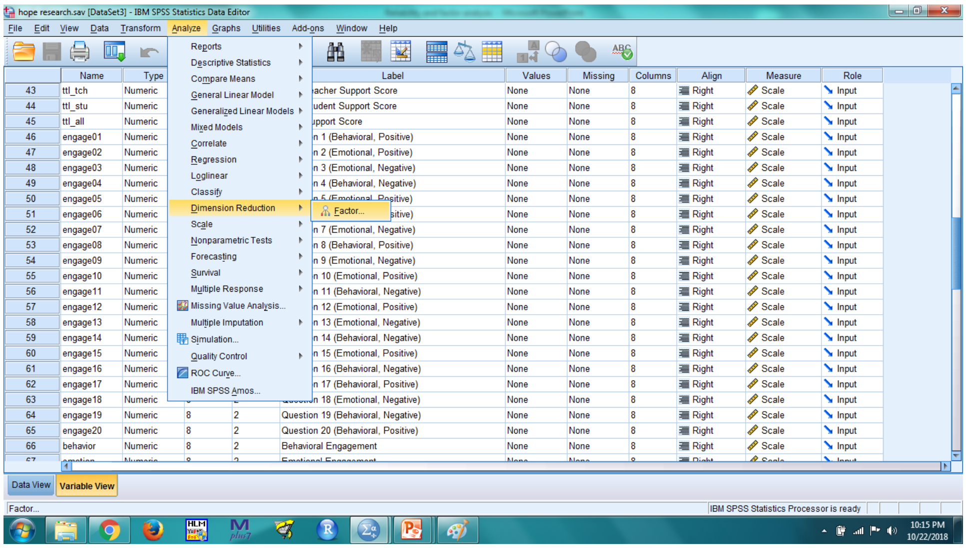
- Promax: most popular
  - An oblique rotation is performed on a solution that has first been orthogonally rotated
  - Redistributes the variance from first (largest) factor to secondary factors
  - Total explained variance remains the same, but the percentage of variance explained by specific factors changes
  - Makes large loadings larger and small loadings smaller making interpretation of factors easier
  - Unique variance remains the same

Oblique Rotation Matrices

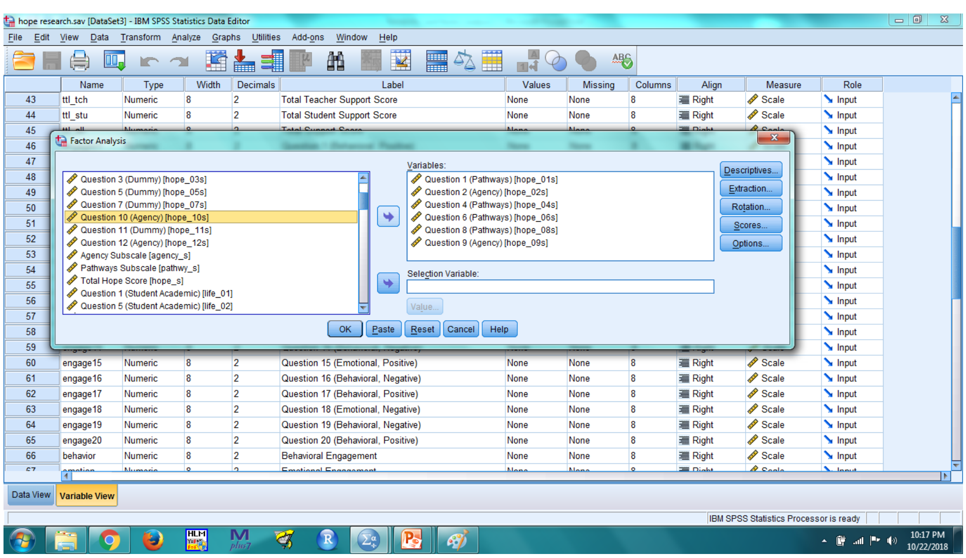
- **Factor Pattern Matrix**: Akin to standardized regression coefficients (β's) in MR: each coefficient indexes the unique relationship between a variable and a particular factor
  - These coefficients are the "factor loadings"

- **Factor Structure Matrix**: Each coefficient indexes the simple correlation between a variable and a particular factor

EFA in SPSS
(using hope.sav)



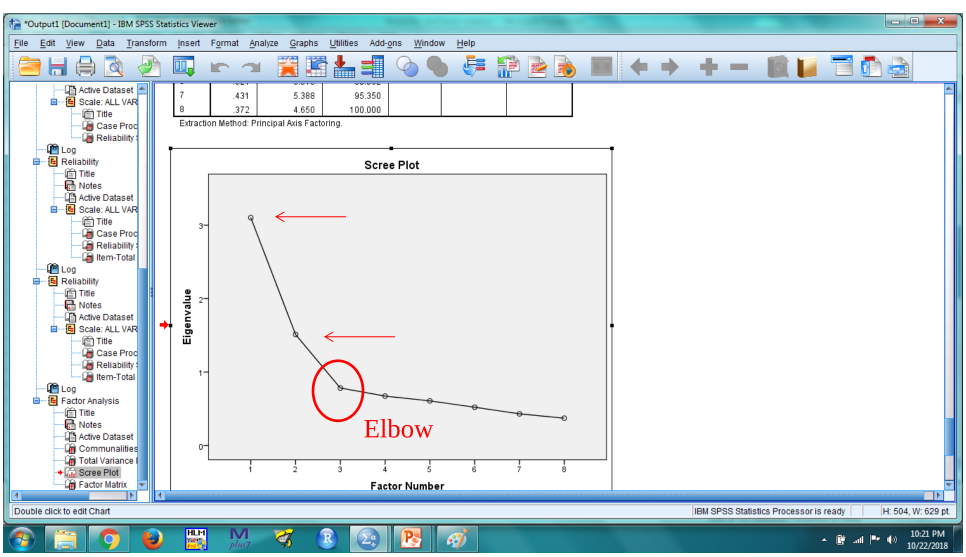
Move Hope Items to the Right



Additional Configurations

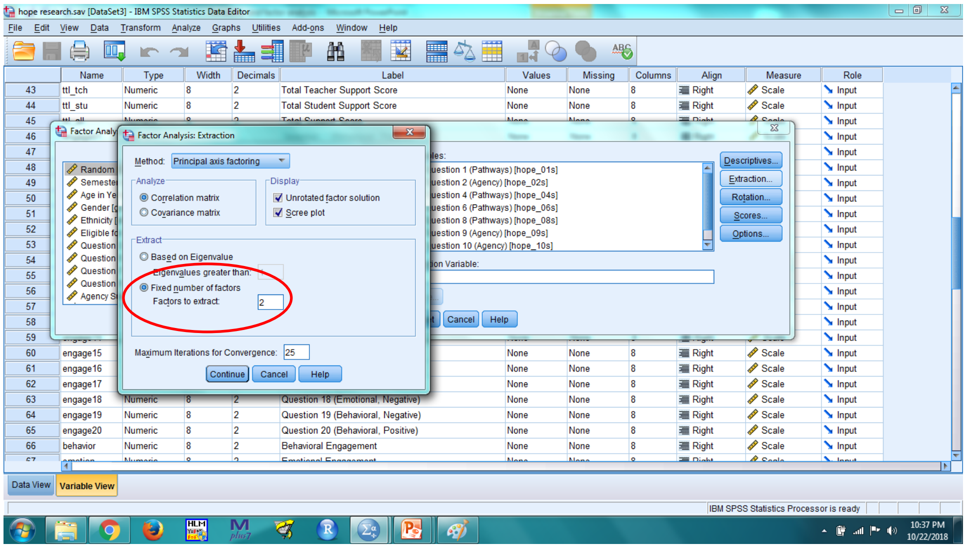
- Extraction button
  - Method: Principal Axis
  - Display: Select “Scree plot”
    - The scree plot will show you how many dimensions or factors you have
    - Find the “elbow”, and count dimensions to the left
    - The “elbow” is the last point on the flat line
- No options for Rotation yet
  - We rotate the factor solution once we have determined how many dimensions we have

Scree Plot

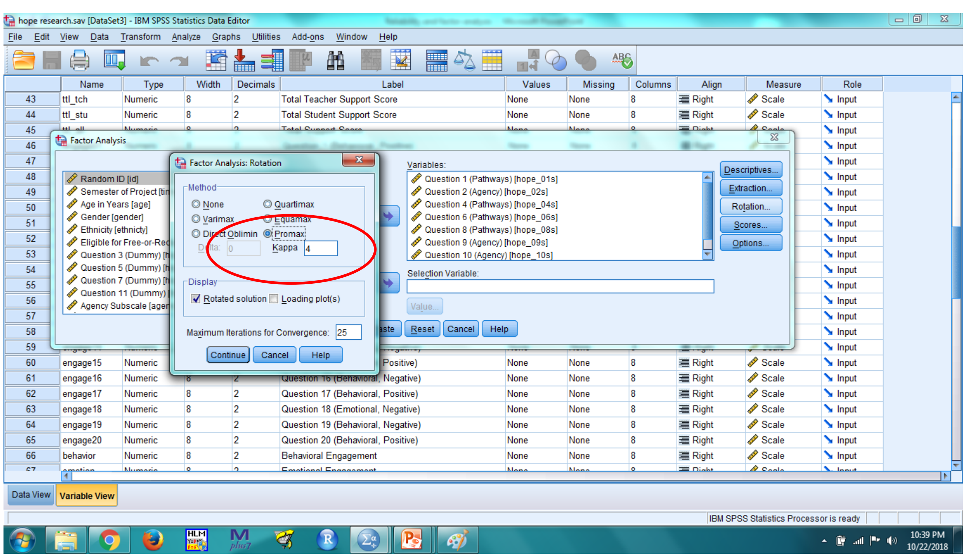


This scale has two independent dimensions

Extract Two Dimensions

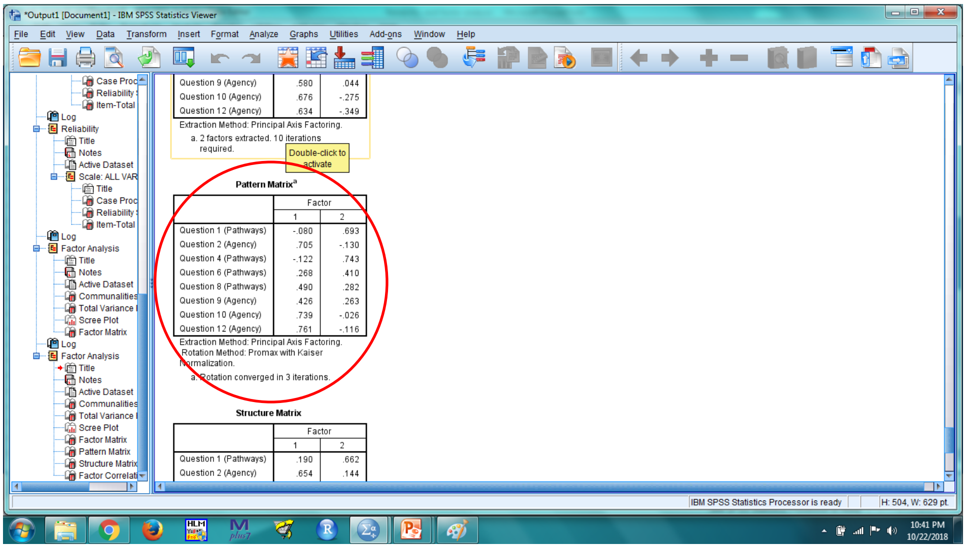


Use Promax Rotation



Promax rotation allows the dimensions to be correlated

Use the Pattern Matrix



Do these items belong to dimensions as we would expect?

Items and Dimensions

- To belong to a dimension, an items must load > .30
  - If one item has a loading > .30 on more than one dimension, we say the item is "cross-loading" and such an item must be removed
- An item that doesn't load > .30 on any dimension should also be removed

Items and Dimensions

Pattern Matrix[a]

What is wrong here?

	Factor	
	1	2
Question 1 (Pathways)	-.080	.693
Question 2 (Agency)	.705	-.130
Question 4 (Pathways)	-.122	.743
Question 6 (Pathways)	.268	.410
Question 8 (Pathways)	.490	.282
Question 9 (Agency)	.426	.263
Question 10 (Agency)	.739	-.026
Question 12 (Agency)	.761	-.116

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

a. Rotation converged in 3 iterations.

Problematic Items

- If the factor analytic results reveal that some items have low communalities or cross load on two or more factors, it may be necessary to rerun the analysis with the problematic items removed from the analysis
  - A reanalysis may extract a different number of factors and redistribute the factor loadings (i.e., items may load on different factors and have different weights)
  - If items are dropped, it is necessary to provide a rationale and detail all steps of the reanalysis
  - Changes to the factor structure may be relatively minor or quite substantive

CFA

- CFA is an application of factor analysis design to test theories regarding item communalities
  - Unlike EFA, a specific relationship between items and factors is specified *a priori*
- CFA is used to confirm a hypothesized factor structure
  - Number of factors specified *a priori*
  - Factors specified as correlated or not *a priori*
  - Variables specified to load only on specific factors
- CFA is applicable only when theory and/or a strong empirical base exists

EFA vs. CFA

- E (exploratory) FA
  - Theory generating: data driven, when theory or previous results are absent
  - Identify # of factors
  - Identify whether factors are intercorrelated
  - Variables/items free to load on all factors
- C (confirmatory) FA
  - Theory testing: strong literature base
  - Factor structure specified
  - Variables/items are hypothesized to load on certain factors and not on others
  - Model testing

CFA Model "Fit" in SEM

- After identifying and estimating the model, "fit" is evaluated by examining the residual difference between the observed covariance matrix and the covariance matrix implied by the model
  - Residuals should be small if the model "fits the data"

- A discrepancy function is used to evaluate model fit (i.e., χ^2 test)
  - As the χ^2 test is sample size dependent, trivial differences will be significant when *N* is large
  - Numerous other fit statistics have been developed to overcome the limitations of the χ^2 test (e.g., CFI, TLI, NFI, RMSR, RMSEA, etc.)

In-Class Exercise

Turn to your neighbor and…

- What is the scale structure for the support items (Student Academic, Student Personal, etc.)?
  – How many factors do you think you will find? How many do you actually find?
- What is the scale structure for the Goal Orientation items?
  – Do you see two distinct factors?
  – If so, how highly are the factors correlated?

Pattern Matrix[a]

	Factor	
	1	2
Item 1 - Task Goal Orientation	-.276	.331
Item 2 - Performance Goal Orientation	.687	.053
Item 3 - Task Goal Orientation	.078	.577
Item 4 - Performance Goal Orientation	.577	-.082
Item 5 - Task Goal Orientation	.066	.631
Item 6 - Performance Goal Orientation	.776	.066
Item 7 - Task Goal Orientation	.077	.624
Item 8 - Performance Goal Orientation	.613	-.088
Item 9 - Task Goal Orientation	-.082	.497
Item 10 - Performance Goal Orientation	.524	.082
Item 11 - Task Goal Orientation	-.124	.459

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

a. Rotation converged in 3 iterations.

Writing up EFA

- Discuss/defend your choices for extraction, rotation (if any)
  - Discuss your decision-making regarding the number of factors to extract
  - Make sure to document any items that are dropped and the results from the subsequent analysis
- Present the data from the Pattern Matrix
  - Discuss what each factor represents

Homework #4

Using tchsurvey18.sav

- Conduct an EFA for survey items d46 – d84
- Determine the correct number of factors
- Extract factors and look for unloaded or cross-loaded items that should be removed
  - If any items removed, re-analyze the data
- Consider items text in your interpretation of each factor
- Document every step in the process and include the final Pattern Matrix from SPSS

Base Group Check-Out

- What was your level of engagement in today's class and why?
  - Rate on a scale of 1 (lowest) to 10 (highest)
- What were three things that you learned from today's class?
- What questions do you have from today's class?
  - Try to find the answers before next week and report back to the group