3/1/98

MULTIVARIATE DATA ANALYSIS w/Readings 3rd Ed.
HAIR, ANDERSON, TATHAM, BLACK 1992 MacMillan

p. 225 FACTOR ANALYSES

"BROADLY SPEAKING, IT [FACTOR ANALYSIS] ADDRESSES ITSELF TO THE PROBLEM OF ANALYZING THE INTER-RELATIONSHIPS AMONG A LARGE NUMBER OF VARIABLES (e.g., TEST SCORES, TEST ITEMS, QUESTIONNAIRE RESPONSES) AND THEN EXPLAINING THE VARIABLES IN TERMS OF THEIR COMMON UNDERLYING FACTORS."

"BY USING FACTOR ANALYSIS, THE ANALYST CAN IDENTIFY THE SEPARATE DIMENSIONS BEING MEASURED BY THE [A] SURVEY AND DETERMINE A FACTOR LOADING FOR EACH VARIABLE (TEST ITEM) ON EACH FACTOR."

"IN A SENSE, EACH OF THE OBSERVED [ORIGINAL] VARIABLES IS CONSIDERED AS A DEPENDENT VARIABLE THAT IS A FUNCTION OF SOME UNDERLYING, LATENT, AND HYPOTHETICAL SET OF FACTORS (DIMENSIONS). CONVERSELY, ONE CAN LOOK AT EACH FACTOR AS A DEPENDENT VARIABLE THAT IS A FUNCTION OF THE ORIGINALLY OBSERVED VARIABLES."

Purposes:

1. Identify a set of Dimensions that are latent (not easily observed) in a large set of variables; this is referred to as R factor analysis.
2. Devise a method of combining or condensing large numbers of people into distinctly different groups within a larger population; this is referred to as Q factor analysis
3. Identify the appropriate variables for subsequent regression, correlation, or discriminant analysis from a much larger set of variables.
4. Create an entirely new set of variables (smaller) to partially or completely replace the original set of variables in subsequent analyses.

FACTOR LOADING = SUMMARY OF ORIGINAL VARIABLES
FACTOR SCORES = ESTIMATES OF FACTORS USED IN SUBSEQUENT ANALYSES.

Exploratory FA: searching for structure among a set of variables (data reduction).
- often used (most frequent purpose of FA).

Confirmatory FA: testing hypotheses as to which variables group together based on theoretical assumptions or prior research.
- the assessment of the degree to which the data meet the expected structure of the analyst.

Component Analysis - this model is used when the objective is to summarize most of the original information (variance) in a minimum number of factors for prediction purposes.

Common FA - this model is used primarily to identify underlying factors or dimensions not easily recognized.

- Factor Extraction -

orthogonal = solution in which the factors are extracted in such a way that the factor axes are maintained at 90 degrees, meaning that each factor is independent of all other factors.
• correlation between factors arbitrarily determined to be zero.

Oblique = more complex than the orthogonal solution.
- subject to controversy
• the factor solution is computed so that the extracted factors are correlated.

p.228 "Oblique solutions assume that the original variables are correlated to some extent; therefore the underlying factors must be similarly correlated."

① correlation matrix → Q or R
② FACTOR MODEL → COMPONENT or COMMON
③ EXTRACTION → ORTHOGONAL or OBLIQUE.

EXTRACTION OF INITIAL "UNROTATED" FACTORS ALLOWS THE ANALYST TO EXPLORE THE DATA REDUCTION POSSIBILITIES FOR A SET OF VARIABLES AND OBTAIN A PRELIMINARY ESTIMATE OF THE NUMBER OF FACTORS TO EXTRACT.

EACH FACTOR HAS A DIFFERENT FACTOR LOADING - WHICH IF YOU ARE LOOKING TO USE THE DATA FOR ANOTHER ANALYSIS SHOULD BE EXAMINED, THE FACTOR w/ THE HIGHEST LOADING IS CHOSEN AS A SURROGATE REPRESENTATIVE OF THE DATA.

- Correlation Matrix derivation -

R - factor Analysis — a result that reflects underlying pattern of variables
Q - Factor Analysis - a result that reflects underlying pattern of the case
i.e, recognizes similar individuals.

- Types of Variance -

Common = variance in variable shared w/ all other variables in the analysis
Specific = variance associated only w/ a specific variable.
Error = variance due to unreliability in the data gathering process.

PCA → total variance considered
→ unities are inserted in the diagonal of the correlation matrix

CFA → Factors derived only w/ common variance
→ communalities inserted in the diagonal of the correlation matrix.

Unities = full variance brought into factor matrix
Communalities = result or factor solution dealing only w/ common variance.

Selection of PCA v. CFA:

(1) Based on goal of the research
(2) Based on the amount of prior knowledge about the variance in the variables.

p. 231 PCA "When the analyst is primarily concerned about prediction and about determining the minimum number of factors needed to account for the maximum portion of the variance represented in the original set of variables, and has prior knowledge suggesting that specific and error variance represent a relatively small ~~portion~~ proportion of the total variance, the appropriate model to select is the component analysis [PCA] model."

CFA "In contrast, when the primary objective is to identify the latent dimensions or constructs represented in the original variables, and the researcher has little knowledge about the amount of unique or error variance and therefore wishes to eliminate this variance, the appropriate model to select is the common factor model." [CFA]

- computers derived good approximations of communalities through repeated calculations."

ROTATION OF FACTORS:

① Compute Unrotated Factor Matrix to assess the number of Factors to extract (a preliminary indication).

- simply interested in the best linear combinations

② of variables. In the sense that one linear combination accounts for more of the variance in the data than any other combination of the original variables.

a) The First Factor may be regarded as the single best summary of linear relationships in the data. The Second Factor is the second-best linear combination of the variables subject to the constraint that F2 is Orthogonal (at right angle to) to the F1.

b) To be orthogonal to the ~~F~~ F1, F2 must be extracted from the proportion of the variance remaining after the F1 has been removed.

p.237 • "Thus the second factor [F2] may be defined as the linear combination of variables that accounts for the most residual variance after the effect of the first factor [F1] has been removed from the data."

• all following factors are defined similarly until the variance in the data is exhausted.

c) unrotated Factor solutions are useful in reducing the data — but is the result in an interpretable form? will it provide info. that offers the most adequate interpretation of the variables under examination?

(2) Rotation: → the reason to rotate is to achieve simpler and theoretically more meaningful factor solutions.

→ factor rotation reduces some ambiguities in the unrotated factor solutions.

• if the unrotated factors are expected to be meaningful then rotation is not needed.

• but rotation is usually desirable for two reasons

i) it simplifies the factor structure

ii) it is usually difficult determine whether unrotated factors will be meaningful or not.

* "The ultimate effect of rotating the factor matrix is to redistribute the variance from earlier factors to later ones to achieve a simpler, theoretically more meaningful, factor pattern."

• Variables might fall into same pattern (e.g., clusteing) with rotation → the relations (pattern) observed simply becomes more clear as the Factor axes are more alligned with the variable distributions after rotation.

• the relative distribution or configuration of the variables does not change.

Oblique rotation is more flexible because the Factor axes do not have to remain independent (or at right angles).

p. 234 • "It also is more realistic because the theoretically important underlying dimensions are not assumed to be uncorrelated with each other."

• The oblique correlation provides information about the extent to which the factors are correlated with each other.

* When the objective is to utilize factor results in further analyses orthogonal rotation should always be used because orthogonal factors eliminate collinearity.

* But if the goal is to simply obtain theoretically meaningful constructs or dimensions, oblique rotation is more desirable because it is more flexible — can be made to fit the variable distribution better — and hence it is theoretically and empirically more realistic.

FACTOR MATRIX (UNDERSTANDING IT!)

1. columns represent Factors
2. Rows are variable's loadings on the factor.
   a) Simplifying the rows (making as many values in the row as close to zero as possible - which maximizes a variable's load on a single or few factors)
   b.) Simplifying the columns (making as many values in each column as close to zero as possible - make the number of "high" loadings as few as possible).

Orthogonal Approaches to simplification

Quartimax Rotation - simplifies the Rows of a factor matrix
- focuses on rotating the unrotated factor matrix so that a variable loads high on one factor, but as low as possible on ~~another~~ all other factors.

Varimax Rotation - simplifies the columns of the matrix

p.235 "Note that in Quatimax approaches many variables can load high or near high on the same factor because the technique centers on simplifying the rows. With the Varimax rotational approach, the maximum possible simplification is reached if there are only 1's and Ø's in a single column."

Equimax Rotation - a compromise between Quatimax and Varimax rotation.
- tries to accomplish some simplification of rows and columns.
- used infrequently.

"The Quartimax method has not proved very successful in producing simpler structures"
- tends to create a large general factor which is not in line w/ the goals of rotation
  - i) to simplify factor structure
  - ii) to make unrotated matrix meaningful.
  - i.e., to achieve simpler and more meaningful factor solutions to reduce ambiguities in unrotated solution.

Varimax has proved successful in producing simpler structures "[it] maximizes the sum of variance of required loadings of the factor matrix."

p.235-236 "With the varimax rotational approach, there tends to be some high loadings (i.e., close to -1 or +1) and some loadings near zero in each column of the matrix. The logic is that interpretation is easiest when the variable-factor correlations are close to either +1 or -1, thus indicating a clear positive or negative association between the variable and the factor, or close to Ø, indicating a clear lack of association."

- choose kind of rotation based on research problem - but Varimax has been shown to offer desired simpler structure in many cases.*