Factor Analysis

Psy 427
Cal State Northridge
Andrew Ainsworth PhD

Topics so far...
- Defining Psychometrics and History
- Basic Inferential Stats and Norms
- Correlation and Regression
- Reliability
- Validity

Putting it together
- “Goal” of psychometrics
  - To measure/quantify psychological phenomenon
  - To try and use measurable/quantifiable items (e.g., questionnaires, behavioral observations) to “capture” some metaphysical or at least directly un-measurable concept
Putting it together

To reach that goal we need...
- Items that actually relate to the concept that we are trying to measure (that’s validity)
- And for this we used correlation and prediction to show criterion (concurrent and predictive) and construct (convergent and discriminant) related evidence for validity
- Note: The criteria we use in criterion related validity is not the concept directly either, but another way (e.g. behavioral, clinical) of measuring the concept.
- Content related validity is decided separately

Items that consistently measure the construct across samples and time and that are consistently related to each other (that’s reliability)
- We used correlation (test-retest, parallel forms, split-half) and the variance sum law (coefficient alpha) to measure reliability
- We even talked about ways of calculating the number of items needed to reach a desired reliability

Why do we want consistent items?
- Domain sampling says they should be
- If the items are reliably measuring the same thing they should all be related to each other
- Because we often want to create a single total score for each individual person (scaling)
- How can we do that? What’s the easiest way? Could there be a better way?
Problem #1
- Composite = Item1 + Item2 + Item3 + ... + Itemk
- Calculating a total score for any individual is often just a sum of the item scores which is essentially treating all the items as equally important (it weights them by 1)
- Composite = (1*Item1) + (1*Item2) + (1*Item3) + ... + (1*Itemk), etc.
- Is there a reason to believe that every item would be equal in how well it relates to the intended

Problem #1
- Regression
  - Why not develop a regression model that predicts the concept of interest using the items in the test?
  - \( \hat{Y}_{\text{Depression}} = b_1(\text{item1}) + b_2(\text{item2}) + \cdots + b_k(\text{itemk}) + \alpha \)
  - What does each b represent? \( \alpha \)?
  - What’s wrong with this picture? What’s missing?
Problem #2
- Tests that we use to measure a concept/construct typically have a moderate to large number of items (i.e. domain sampling)
- With this comes a whole mess of relationships (i.e. covariances/correlations)
- Alpha just looks for one consistent pattern, what if there are more patterns? And what if some items relate negatively (reverse coded)?

Correlation Matrix - MAS
Problem #2

- So alpha can give us a single value that illustrates the relationship among the items as long as there is only one consistent pattern.
- If we could measure the concept directly we could do this differently and reduce the entire matrix on the previous page down to a single value as well; a single correlation.

Multiple Correlation

- Remember that:
  \[ Y = b_1X_1 + b_2X_2 + \cdots + b_kX_k + a + e \]
  \[ \hat{Y} = b_1X_1 + b_2X_2 + \cdots + b_kX_k + a \]

  so

  \[ e = Y - \hat{Y}, \text{ or the residual} \]
Multiple Correlation

- So, that means that Y-hat is the part of Y that is related to ALL of the Xs combined
- The multiple correlation is simple the correlation between Y and Y-hat

\[ R_{Y \cdot X_1 X_2 X_3 X_K} = r_{YY} \]

We can even square the value and get the Squared Multiple Correlation (SMC), which will tell us the proportion of Y that is explained by the Xs

So, (importantly) if Y is the concept/criterion we are trying to measure and the Xs are the items of a test this would give us a single measure of how well the items measure the concept.

What to do???

- Same problem, if we can’t measure the concept directly we can’t apply a regression equation to establish the optimal weights for adding items up and we can’t reduce the number of patterns (using R) because we can’t measure the concept directly
- If only there were a way to handle this...
What is Factor Analysis (FA)?

- FA and PCA (principal components analysis) are methods of data reduction
  - Take many variables and explain them with a few “factors” or “components”
  - Correlated variables are grouped together and separated from other variables with low or no correlation

What is FA?

- Patterns of correlations are identified and either used as descriptive (PCA) or as indicative of underlying theory (FA)
- Process of providing an operational definition for latent construct (through a regression like equation)
General Steps to FA
- Step 1: Selecting and Measuring a set of items in a given domain
- Step 2: Data screening in order to prepare the correlation matrix
- Step 3: Factor Extraction
- Step 4: Factor Rotation to increase interpretability
- Step 5: Interpretation
- Step 6: Further Validation and Reliability of measures

Factor Analysis Questions
- Three general goals: data reduction, describe relationships and test theories about relationships (next chapter)
- How many interpretable factors exist in the data? or How many factors are needed to summarize the pattern of correlations?
- What does each factor mean? Interpretation?
- What is the percentage of variance in the data accounted for by the factors?

Factor Analysis Questions
- Which factors account for the most variance?
- How well does the factor structure fit a given theory?
- What would each subject’s score be if they could be measured directly on the factors?
Types of FA

- Exploratory FA
  - Summarizing data by grouping correlated variables
  - Investigating sets of measured variables related to theoretical constructs
  - Usually done near the onset of research
  - The type we are talking about in this lecture

Types of FA

- Confirmatory FA
  - More advanced technique
  - When factor structure is known or at least theorized
  - Testing generalization of factor structure to new data, etc.
  - This is often tested through Structural Equation Model methods (beyond this course)

Remembering CTT

- Assumes that every person has a true score on an item or a scale if we can only measure it directly without error
- CTT analyses assumes that a person's test score is comprised of their "true" score plus some measurement error.
- This is the common true score model

\[ X = T + E \]
The common factor model is like the true score model where

\[ X_k = T + E \]

Except let's think of it at the level of variance for a second

\[ \sigma^2 = \sigma^2_T + \sigma^2_E \]

Since we don't know T let's replace that with what is called the "common variance" or the variance that this item shares with other items in the test

This is called communality and is indicated by h-squared

\[ \sigma^2 = h^2 + \sigma^2_E \]

Instead of thinking about E as "error" we can think of it as the variance that is NOT shared with other items in the test or that is "unique" to this item

The unique variance (u-squared) is made up of variance that is specific to this item and error (but we can't pull them apart)

\[ \sigma^2 = h^2 + u^2 \]
Common Factor Model

- The common factor model assumes that the commonalities represent variance that is due to the concept (i.e. factor) you are trying to measure.
- That's great but how do we calculate communalities?

Let's rethink the regression approach:
- The multiple regression equation from before:
  $$Y_{Factor} = b_1(item_1) + b_2(item_2) + \cdots + b_k(item_k) + a + e$$
- Or it's more general form:
  $$Y_{Factor} = \sum(b_kx_k) + a + e$$
- Let's think about this more theoretically.
Common Factor Model

- Still rethinking regression
  - So, theoretically items don’t make up a factor (e.g., depression), the factor should predict scores on the item
  - Example: if you know someone is “depressed” then you should be able to predict how they will respond to each item on the CES-D

Regression Model Flipped Around

Let’s predict the item from the Factor(s)

\[ x_k = \sum (\psi_{jk}F_j) + \varepsilon_k \]

- Where \( x_k \) is the item on a scale
- \( \psi_{jk} \) is the relationship (slope) b/t factor and item
- \( F_j \) is the Factor
- \( \varepsilon_k \) is the error (residual) predicting the item

Notice the change in the direction of the arrows to indicate the flow of theoretical influence.
Common Factor Model

- **Communality**
  - The communality is a measure of how much each item is explained by the Factor(s) and is therefore also a measure of how much each item is related to other items.
  - The communality for each item is calculated by:
    \[ h_i^2 = \sum_j \psi_{jk}^2 \]
  - Whatever is left in an item is the uniqueness.

The big burning question

- How do we predict items with factors we can’t measure directly?
- This is where the mathematics comes in.
- Long story short, we use a mathematical procedure to piece together “super variables” that we use as a fill-in for the factor in order to estimate the previous formula.

Factors come from geometric decomposition

- **Eigenvalue/Eigenvector Decomposition**
  - Sometimes called Singular Value Decomposition.
- A correlation matrix is broken down into smaller “chunks”, where each “chunk” is a projection into a cluster of data points (eigenvectors).
- Each vector (chunk) is created to explain the maximum amount of the correlation matrix (the maximum amount explained is the eigenvalue).
Common Factor Model

- Factors come from geometric decomposition
  - Each eigenvector is created to maximize the relationships among the variables (communality)
  - Each vector “stands in” for a factor and then we can measure how well each item is predicted by (related to) the factor (i.e. the common factor model)

Factor Analysis Terms

- Observed Correlation Matrix – is the matrix of correlations between all of your items
- Reproduced Correlation Matrix – the correlation that is “reproduced” by the factor model
- Residual Correlation Matrix – the difference between the Observed and Reproduced correlation matrices

Factor Analysis Terms

- Extraction – refers to 2 steps in the process
  - Method of extraction (there are dozens)
    - PCA is one method
    - FA refers to a whole mess of them
  - Number of factors to “extract”
- Loading – is a measure of relationship (analogous to correlation) between each item and the factor(s); the Ψ’s in the common factor model
Matrices

Variables

Data Matrix →

Factors

Covariance Matrix

Factor Loading Matrix

Redundant Correlation Matrix

Commonalities

Observables

Factors

Commonalities

Factor Loading Matrix

Commonalities

Observables

Factor Analysis Terms

- Factor Scores – the factor model is used to generate a combination of the items to generate a single score for the factor
- Factor Coefficient matrix – coefficients used to calculate factor scores (like regression coefficients)
Factor Analysis Terms

- Rotation – used to mathematically convert the factors so they are easier to interpret
  - Orthogonal – keeps factors independent
    - There is only one matrix and it is rotated
    - Interpret the rotated loading matrix
  - Oblique – allows factors to correlate
    - Factor Correlation Matrix – correlation between the factors
    - Structure Matrix – correlation between factors and variables
    - Pattern Matrix – unique relationship between each factor and an item uncontaminated by overlap between the factors
      - The relationship between an item and a factor that is not shared by other factors; this is the matrix you interpret

Simple Structure – refers to the ease of interpretability of the factors (what they mean).
- Achieved when an item only loads highly on a single factor when multiple factors exist (previous slide)
- Lack of complex loadings (items load highly on multiple factors simultaneously)

Simple vs. Complex Loading
FA vs. PCA

- FA produces factors; PCA produces components
- Factors cause variables; components are aggregates of the variables

Conceptual FA vs. PCA

FA

I1  I2  I3

PCA

I1  I2  I3

FA vs. PCA

- FA analyzes only the variance shared among the variables (common variance without unique variance)
  - PCA analyzes all of the variance
- FA: "What are the underlying processes that could produce these correlations?"
  - PCA: Just summarize empirical associations, very data driven
FA vs. PCA

- PCA vs. FA (family)
  - PCA begins with 1s in the diagonal of the correlation matrix
  - All variance extracted
  - Each variable giving equal weight initially
  - Commonalities are estimated as the output of the model and are typically inflated
  - Can often lead to an over extraction of factors as

FA vs. PCA

- PCA vs. FA (family)
  - FA begins by trying to only use the common variance
  - This is done by estimating the communality values (e.g. SMC) and placing them in the diagonal of the correlation matrix
  - Analyzes only common variance
  - Outputs a more realistic (often smaller) communality estimate
  - Usually results in far fewer factors overall

What else?

- How many factors do you extract?
  - How many do you expect?
  - One convention is to extract all factors with eigenvalues greater than 1 (Kaiser Criteria)
  - Another is to extract all factors with non-negative eigenvalues
  - Yet another is to look at the scree plot
  - Try multiple numbers and see what gives best interpretation.
Eigenvalues greater than 1

<table>
<thead>
<tr>
<th>Factor</th>
<th>Total Variance Explained</th>
<th>Fixed Sum of Squared Loadings</th>
<th>Rotation Sum of Squared Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>3.513</td>
<td>0.377</td>
<td>0.377</td>
</tr>
<tr>
<td>1</td>
<td>2.681</td>
<td>0.377</td>
<td>0.377</td>
</tr>
<tr>
<td>2</td>
<td>2.681</td>
<td>0.377</td>
<td>0.377</td>
</tr>
<tr>
<td>3</td>
<td>2.681</td>
<td>0.377</td>
<td>0.377</td>
</tr>
<tr>
<td>4</td>
<td>2.681</td>
<td>0.377</td>
<td>0.377</td>
</tr>
<tr>
<td>5</td>
<td>2.681</td>
<td>0.377</td>
<td>0.377</td>
</tr>
<tr>
<td>6</td>
<td>2.681</td>
<td>0.377</td>
<td>0.377</td>
</tr>
<tr>
<td>7</td>
<td>2.681</td>
<td>0.377</td>
<td>0.377</td>
</tr>
<tr>
<td>8</td>
<td>2.681</td>
<td>0.377</td>
<td>0.377</td>
</tr>
<tr>
<td>9</td>
<td>2.681</td>
<td>0.377</td>
<td>0.377</td>
</tr>
<tr>
<td>10</td>
<td>2.681</td>
<td>0.377</td>
<td>0.377</td>
</tr>
<tr>
<td>11</td>
<td>2.681</td>
<td>0.377</td>
<td>0.377</td>
</tr>
<tr>
<td>12</td>
<td>2.681</td>
<td>0.377</td>
<td>0.377</td>
</tr>
</tbody>
</table>

Extraction Method: Principal Axis Factoring.

Scree Plot

What else?

- How do you know when the factor structure is good?
  - When it makes sense and has a (relatively) simple structure.
  - When it is the most useful.

- How do you interpret factors?
  - Good question, that is where the true art of this