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

Putting it together

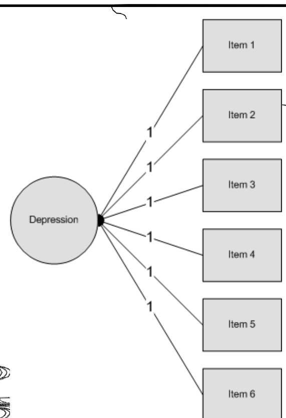
- To reach that goal we need…
  - Items that consistently measure the construct across samples and time and that are consistently related to each other (that's reliability)
  - We used correlation (test-retest, parallel forms, split-half) and the variance sum law (coefficient alpha) to measure reliability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

Putting it together

- 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 concept?

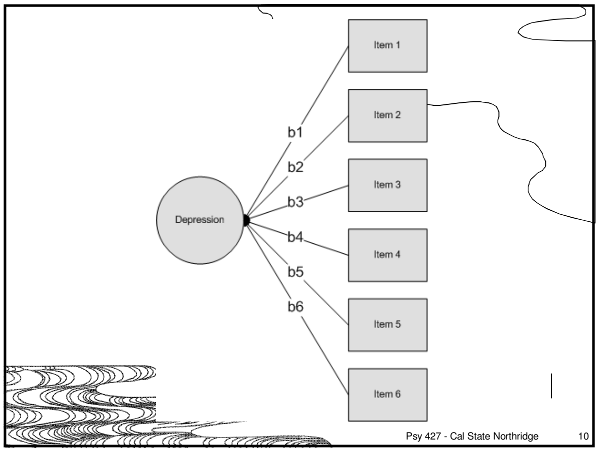


Problem #1

- Regression
  - Why not develop a regression model that predicts the concept of interest using the items in the test?

$$\hat{Y}_{Depression} = b_1(item1) + b_2(item2) + \cdots + b_k(itemk) + a$$

  - What does each b represent? a?
- 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)?

Psy 427 - Cal State Northridge 11

Correlation Matrix - MAS

MAS	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	1	0.696	0.641	0.669	0.641	0.745	0.631	0.238	0.416	0.441	0.701	0.402	0.470	0.243	0.365	0.556	0.358	0.550	0.256	0.407	0.460	0.469	0.578	0.634
2	0.696	1	0.542	0.547	0.551	0.728	0.536	0.381	0.494	0.462	0.675	0.371	0.494	0.382	0.425	0.411	0.339	0.646	0.378	0.603	0.485	0.396	0.680	0.603
3	0.641	0.542	1	0.568	0.610	0.692	0.685	0.351	0.646	0.521	0.758	0.362	0.475	0.391	0.412	0.591	0.302	0.617	0.255	0.488	0.401	0.519	0.694	0.646
4	0.669	0.547	0.568	1	0.533	0.521	0.697	0.534	0.414	0.347	0.614	0.490	0.701	0.251	0.328	0.828	0.434	0.492	0.550	0.347	0.586	0.600	0.585	0.581
5	0.641	0.551	0.610	0.533	1	0.676	0.570	0.215	0.592	0.353	0.578	0.316	0.442	0.374	0.326	0.290	0.462	0.540	0.278	0.456	0.536	0.425	0.635	0.531
6	0.745	0.728	0.692	0.521	0.676	1	0.644	0.300	0.594	0.642	0.810	0.425	0.461	0.378	0.395	0.409	0.443	0.750	0.351	0.656	0.543	0.570	0.691	0.745
7	0.631	0.536	0.685	0.697	0.570	0.644	1	0.331	0.547	0.482	0.739	0.238	0.623	0.299	0.243	0.597	0.312	0.566	0.408	0.277	0.523	0.637	0.676	0.665
8	0.238	0.381	0.351	0.534	0.215	0.300	0.331	1	0.235	0.117	0.454	0.424	0.571	0.180	0.419	0.593	0.187	0.449	0.334	0.363	0.377	0.435	0.472	0.204
9	0.416	0.494	0.646	0.414	0.592	0.594	0.547	0.235	1	0.579	0.662	0.278	0.402	0.545	0.309	0.382	0.462	0.717	0.340	0.453	0.614	0.490	0.697	0.609
10	0.441	0.462	0.521	0.347	0.353	0.642	0.482	0.117	0.579	1	0.686	0.389	0.379	0.624	0.458	0.369	0.565	0.648	0.322	0.451	0.640	0.570	0.611	0.717
11	0.701	0.675	0.758	0.614	0.578	0.810	0.739	0.454	0.662	0.686	1	0.405	0.684	0.404	0.413	0.561	0.409	0.772	0.466	0.639	0.588	0.679	0.740	0.732
12	0.402	0.371	0.362	0.490	0.316	0.425	0.238	0.424	0.278	0.389	0.405	1	0.396	0.266	0.402	0.442	0.366	0.445	0.462	0.543	0.550	0.594	0.382	0.255
13	0.470	0.494	0.475	0.701	0.442	0.461	0.623	0.571	0.402	0.379	0.684	0.396	1	0.360	0.258	0.553	0.428	0.535	0.680	0.474	0.521	0.643	0.637	0.484
14	0.243	0.382	0.391	0.251	0.374	0.378	0.299	0.180	0.545	0.624	0.404	0.266	0.360	1	0.495	0.316	0.684	0.535	0.230	0.324	0.577	0.306	0.573	0.499
15	0.365	0.425	0.412	0.328	0.326	0.395	0.243	0.419	0.309	0.458	0.413	0.402	0.258	0.495	1	0.339	0.454	0.414	0.220	0.407	0.450	0.444	0.388	0.391
16	0.556	0.411	0.591	0.828	0.290	0.409	0.597	0.593	0.382	0.369	0.561	0.442	0.553	0.316	0.339	1	0.405	0.492	0.323	0.327	0.524	0.557	0.501	0.515
17	0.358	0.339	0.302	0.434	0.462	0.443	0.312	0.187	0.462	0.565	0.409	0.366	0.428	0.684	0.454	0.405	1	0.491	0.418	0.450	0.711	0.521	0.429	0.499
18	0.550	0.646	0.617	0.492	0.540	0.750	0.566	0.449	0.717	0.648	0.772	0.445	0.535	0.535	0.414	0.492	0.491	1	0.485	0.655	0.699	0.561	0.753	0.678
19	0.256	0.378	0.255	0.550	0.278	0.351	0.408	0.334	0.340	0.322	0.466	0.462	0.680	0.230	0.220	0.323	0.418	0.485	1	0.448	0.476	0.608	0.390	0.431
20	0.407	0.603	0.488	0.347	0.456	0.656	0.277	0.363	0.453	0.451	0.639	0.543	0.474	0.324	0.407	0.327	0.450	0.655	0.448	1	0.479	0.521	0.439	0.517
21	0.460	0.485	0.401	0.586	0.536	0.543	0.523	0.377	0.614	0.640	0.588	0.550	0.521	0.577	0.450	0.524	0.711	0.699	0.476	0.479	1	0.675	0.624	0.530
22	0.469	0.396	0.519	0.600	0.425	0.570	0.637	0.435	0.490	0.570	0.679	0.594	0.643	0.306	0.444	0.557	0.521	0.561	0.608	0.521	0.675	1	0.534	0.586
23	0.578	0.680	0.694	0.585	0.635	0.691	0.676	0.472	0.697	0.611	0.740	0.382	0.637	0.573	0.388	0.501	0.429	0.753	0.390	0.439	0.624	0.534	1	0.592
24	0.634	0.603	0.646	0.581	0.531	0.745	0.665	0.204	0.609	0.717	0.732	0.255	0.484	0.499	0.391	0.515	0.499	0.678	0.431	0.517	0.530	0.586	0.592	1

Psy 427 - Cal State Northridge 12

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

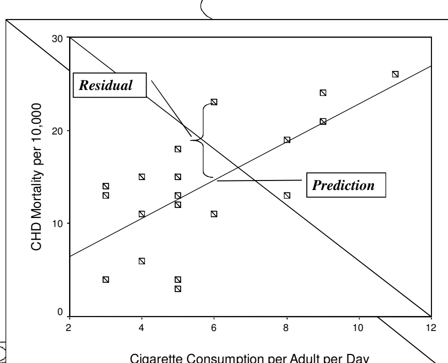
- Remember that:

$$Y = b_1X_1 + b_2X_2 + \cdots + b_kX_k + a + e$$

$$\hat{Y} = b_1X_1 + b_2X_2 + \cdots + b_kX_k + a$$

$$so$$

$$e = Y - \hat{Y},\ \text{or the residual}$$



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

$$R_{Y \cdot X_1 X_2 X_3 X_K} = r_{Y\hat{Y}}$$

- Let's demonstrate

Psy 427 - Cal State Northridge 16

Multiple Correlation

- 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

Psy 427 - Cal State Northridge 17

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…

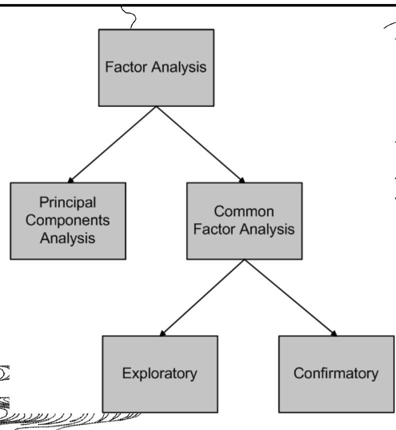
Psy 427 - Cal State Northridge 18

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 the measures

Psy 427 - Cal State Northridge 22

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?

Psy 427 - Cal State Northridge 23

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?

Psy 427 - Cal State Northridge 24

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

$$X = T + E$$

Common Factor Model

- 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

$$\sigma^2_{X_k} = \sigma^2_T + \sigma^2_E$$

Psy 427 - Cal State Northridge 28

Common Factor Model

- 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

$$\sigma^2_{X_k} = h^2 + \sigma^2_E$$

Psy 427 - Cal State Northridge 29

Common Factor Model

- 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)

$$\sigma^2_{X_k} = h^2 + u^2$$

Psy 427 - Cal State Northridge 30

Common Factor Model

Variance of an item	=	The common variance (variance shared with the other items)	+	The variance specific to the item	+	Random Error

Variance of an item	=	The common variance (variance shared with the other items)	+	The Unique variance of the item

Variance of an item	=	Communality	+	Uniqueness

$$\sigma^2_{x_k} = h^2 + u^2$$

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?

Common Factor Model

- 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:

$$Y_{Factor} = \sum(b_k x_k) + a + e$$

  - Now, let's think about this more theoretically

Common Factor Model

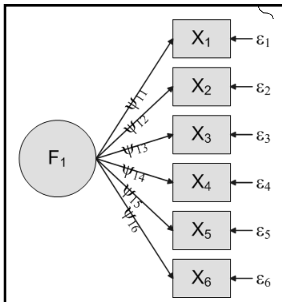
- Still rethinking regression
  - So, theoretically items don't make up a factor (e.g. depression), the factor should predict scores on the item
  - Example: if you know someone is "depressed" then you should be able to predict how they will respond to each item on the CES-D

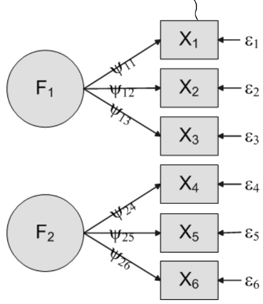
Common Factor Model

- Regression Model Flipped Around
  - Let's predict the item from the Factor(s)

$$x_k = \sum \left(\psi_{jk} F_j \right) + \varepsilon_k$$

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





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

$$h_k^2 = \sum \psi_{jk}^2$$

  - Whatever is left in an item is the uniqueness

Psy 427 - Cal State Northridge 37

Common Factor Model

- 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

Psy 427 - Cal State Northridge 38

Common Factor Model

- 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 amount variability explained is the eigenvalue)

Psy 427 - Cal State Northridge 39

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)

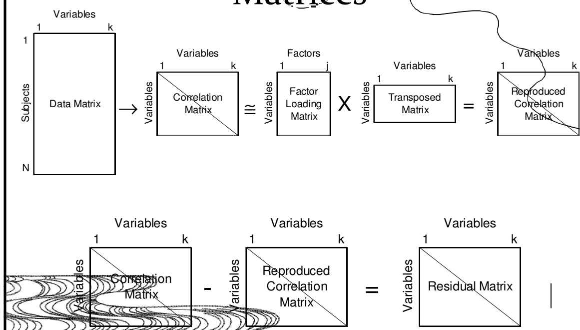
Factor Analysis Terms

- 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

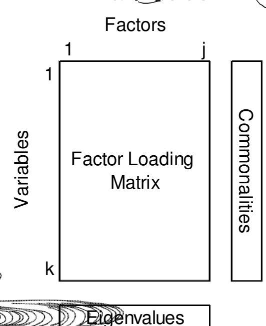
- Extraction – refers to 2 steps in the process
  - Method of extraction (there are dozens)
    - PCA is one method
    - FA refers to a whole mess of them
  - Number of factors to "extract"
- Loading – is a measure of relationship (analogous to correlation) between each item and the factor(s); the Ψ's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

Matrices



Psy 427 - Cal State Northridge 43

Matrices



Psy 427 - Cal State Northridge 44

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)

Psy 427 - Cal State Northridge 45

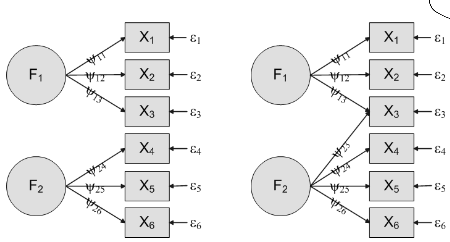
Factor Analysis Terms

- 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 (i.e. the relationship between an item an a factor that is not shared by other factors); **this is the matrix you interpret**

Factor Analysis Terms

- 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

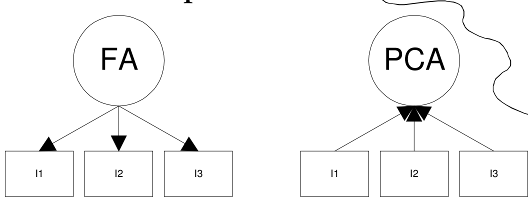
Simple vs. Complex Loading



FA vs. PCA

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

Conceptual FA vs. PCA



FA 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 well

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 correlations 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

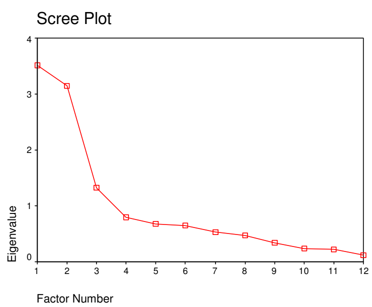
Total Variance Explained

Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.513	29.276	29.276	3.296	27.467	27.467	3.251	27.094	27.094
2	3.141	26.171	55.447	2.681	22.338	49.805	1.509	12.573	39.666
3	1.321	11.008	66.455	.843	7.023	56.828	1.495	12.455	52.121
4	.801	6.676	73.132	.329	2.745	59.573	.894	7.452	59.573
5	.675	5.623	78.755						
6	.645	5.375	84.131						
7	.527	4.391	88.522						
8	.471	3.921	92.443						
9	.342	2.851	95.294						
10	.232	1.936	97.231						
11	.221	1.841	99.072						
12	.111	.928	100.000						

Extraction Method: Principal Axis Factoring.

Psy 427 - Cal State Northridge 55

Scree Plot



Psy 427 - Cal State Northridge 56

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 come in.

Psy 427 - Cal State Northridge 57