Factor Analysis
Edpsy/Soc 584 & Psych 594

Carolyn J. Anderson
University of Illinois, Urbana-Champaign

April 29, 2009

Overview

- Purpose
- Basic Model Assumptions (one common factor model)
- Generalization to m common factors
- Estimation
- Assessment of Model Fit to Data
- Rotation
- Confirmatory Factor Analysis

Estimation versus Purpose

- Purpose:
  - Explore underlying structure — data driven.
  - Confirm underlying structure — theory driven.
- Estimation method (Factor extraction)
  - Eigen-decomposition based
    - Eigen-decomposition of $\boldsymbol{S}$ which is the "principal components" solution.
    - Iterative eigen-decompositions of $\boldsymbol{S} - \tilde{\boldsymbol{\Psi}}$ which is the "Principal factor" solution.
  - Maximum likelihood estimation — We now must assume that $\boldsymbol{F}$ and $\boldsymbol{\epsilon}$ are multivariate normal. Tends to fit data better & yields scale invariance (ie., use either $\boldsymbol{S}$ or $\boldsymbol{R}$).

The Data ($n = 320$ish)

Data from Espelage, D.L., Holt, M.K., & Henkel, R.R. (2003). Examination of Peer-Group contextual effects on aggression during early adolescence. *Child Development*, *74*, 205–220.

Items have a 5 point response scale: “Never”, “1 to 2 times”, “3 to 4 times”, “5 to 6 times”, “7 or more times”

- Q36: I upset other students for the fun of it.
- Q37: In a group I teased other students.
- Q38: I fought students I could easily beat.
- Q39: Other students picked on me.
- Q40: Other students made fun of me.
- Q41: Other students called me names.
- Q42: I got hit and pushed by other students.
- Q43: I helped harass other students.

The Items

Items have a 5 point response scale: “Never”, “1 to 2 times”, . . . , “7 or more times”

- Q44: I teased other students.
- Q45: I got in a physical fight.
- Q46: I threatened to hurt or hit another student.
- Q47: I got into a physical fight because I was angry.
- Q48: I hit back when someone hit me first.
- Q49: I was mean to someone when I was angry.
- Q50: I spread rumors about other students.
- Q51: I started (instigated) arguments or conflicts.
- Q52: I encouraged people to fight.
- Q53: I excluded other students from my clique (group) of friends

Based on the content,

What are the factors and which items load on which factors?

Item	F_1	F_2	F_3
36			
37			
38			
39			
40			
41			
42			
43			
44			

Item	F_1	F_2	F_3
45			
46			
47			
48			
49			
50			
51			
52			
53			

6 / 52

What They Are Designed To Be

α here is Chronbach's alpha for measuring reliability.

F_1 = Fighting ($\alpha = .88$),
F_2 = Bullying ($\alpha = .88$),
F_3 = Victimization ($\alpha = .87$)

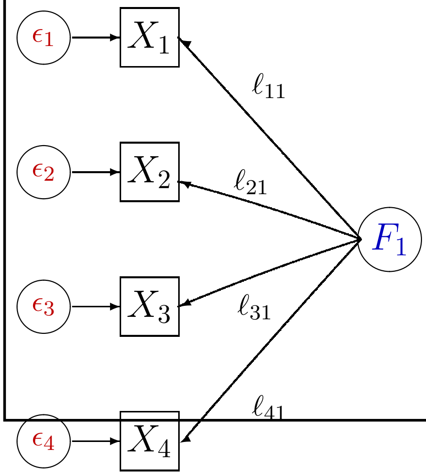
Item	F_1	F_2	F_3
36		X	
37		X	
38	X		
39			X
40			X
41			X
42			X
43		X	
44		X	

Item	F_1	F_2	F_3
45	X		
46	X		
47	X		
48	X		
49		X	
50		X	
51		X	
52		X	
53		X	

7 / 52

Factor Analytic Model

Factor analysis (FA) is a latent variable model.



The One Common Factor Model

$$\begin{aligned} X_1 &= \mu_1 + \ell_{11} F_1 + \epsilon_1 \\ X_2 &= \mu_2 + \ell_{21} F_1 + \epsilon_2 \\ X_3 &= \mu_3 + \ell_{31} F_1 + \epsilon_3 \\ X_4 &= \mu_4 + \ell_{41} F_1 + \epsilon_4 \end{aligned}$$

9 / 52

The One Common Factor Analytic Model

$$X_i = \mu_i + \ell_{i1} F_1 + \epsilon_i$$

- The μ_i's are the means of the X_i's.
- Common Factor: F_1 is an unobserved random variable with mean 0 and variance ϕ.
- The $\ell_{i1}s$ are the Factor Loadings.
- Specificities or Uniquenesses: ϵ_i are independent over i, unobserved random variables with means equal to 0 and variances equal to ψ_i.
- F_1 is independent of the ϵ_i's.

10 / 52

Unique Variables

A little Classical test theory:

$$\text{true score} = \text{observed score} - \text{pure measurement error}$$
$$t_i = X_i - e_i$$

- If a factor model holds for the observed scores X_i, then it should also hold for the true scores t_i.
- The uniqueness in the factor model for the true scores contains specific errors due to the particular variables (items) selected; that is,

$$t_i = \ell_{i1}F_1 + \underbrace{s_i}_{specific}$$

- Solve for X_i:

$$t_i = \ell_{i1}F_1 + s_i = X_i - e_i \Longrightarrow X_i = \ell_{i1}F_i + \underbrace{(s_i + e_i)}_{\epsilon_i}$$

11 / 52

Unique Variables & Common Variables

The unique variables ϵ_i contain

- Pure measurement errors
- Specific errors

The observed variables will be correlated because they all depend (in part) on F_1.

The common factor model makes implications for covariance (and correlation) matrices; therefore, the "data" are the covariance (or correlation) matrix.

12 / 52

Implications for Data

Let's turn it into a linear algebra problem:

$$\boldsymbol{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix} \qquad \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix} \quad \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \dots & \sigma_{1p} \\ \sigma_{12} & \sigma_{22} & \dots & \sigma_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1p} & \sigma_{2p} & \dots & \sigma_{pp} \end{pmatrix}$$

$$\boldsymbol{L} = \begin{pmatrix} \ell_{11} \\ \ell_{21} \\ \vdots \\ \ell_{p1} \end{pmatrix} \quad \boldsymbol{\epsilon} = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_p \end{pmatrix} \qquad \text{var}(F) = \boldsymbol{\Phi} = \phi_{11}$$

$$\boldsymbol{\Psi} = \begin{pmatrix} \psi_1 & 0 & 0 & 0 \\ 0 & \psi_2 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \psi_p \end{pmatrix}$$

So $\boldsymbol{X} - \boldsymbol{\mu} = \boldsymbol{L}F_1 + \boldsymbol{\epsilon}$

Implications for Data (continued)

$$\boldsymbol{X} - \boldsymbol{\mu} = \boldsymbol{L}F_1 + \boldsymbol{\epsilon}$$

- Mean of $\boldsymbol{X} - \boldsymbol{\mu} = \boldsymbol{L}E(F_1) + E(\boldsymbol{\epsilon}) = \boldsymbol{0}$
- Covariance matrix for $\boldsymbol{X}$ or equivalently $\boldsymbol{X} - \boldsymbol{\mu}$:

$$\begin{aligned} \boldsymbol{\Sigma} &= E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'] \\ &= E[(\boldsymbol{L}F_1 + \boldsymbol{\epsilon})(\boldsymbol{L}F_1 - \boldsymbol{\epsilon})'] \\ &= E[(\boldsymbol{L}F_1 + \boldsymbol{\epsilon})(F_1\boldsymbol{L}' - \boldsymbol{\epsilon}')] \\ &= \boldsymbol{L}\underbrace{E[F_1^2]}_{\phi_{11}=1}\boldsymbol{L}' + \boldsymbol{L}\underbrace{E[F_1\boldsymbol{\epsilon}']}_{\boldsymbol{0}} + \underbrace{E[\boldsymbol{\epsilon}F_1]}_{\boldsymbol{0}}\boldsymbol{L}' + \underbrace{E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}']}_{\boldsymbol{\Psi}} \\ &= \boldsymbol{L}\boldsymbol{L}' + \boldsymbol{\Psi} \end{aligned}$$

The "Data" is Covariance Matrix

One common factor model implies:

$$\boldsymbol{\Sigma} = \begin{pmatrix} (\ell_{11}^2 + \psi_1) & \ell_{11}\ell_{21} & \dots & \ell_{11}\ell_{p1} \\ \ell_{11}\ell_{21} & (\ell_{21}^2 + \psi_2) & \dots & \ell_{21}\ell_{p1} \\ \vdots & \vdots & \ddots & \vdots \\ \ell_{11}\ell_{I1} & \ell_{21}\ell_{I1} & \dots & (\ell_{I1}^2 + \psi_p) \end{pmatrix}$$

- $\ell_{i1}^2 = h_i^2$ is the Communality of item i.
- Foreshadowing estimation methods: If $\boldsymbol{\Psi}$ were known,

$$\boldsymbol{\Sigma} - \boldsymbol{\Psi} = \boldsymbol{L}\boldsymbol{L}' = (\boldsymbol{e}_1\sqrt{\lambda_1})(\sqrt{\lambda_1}\boldsymbol{e}_1')$$

or

$$\boldsymbol{\Sigma} = \underbrace{(\boldsymbol{e}_1^*\sqrt{\lambda_1^*})}_{\boldsymbol{L}^{*'}}\underbrace{(\sqrt{\lambda_1^*}\boldsymbol{e}_1^{*'})}_{\boldsymbol{L}^*} \quad \text{and} \quad \tilde{\boldsymbol{\Psi}} = \text{diag}(\boldsymbol{\Sigma} - \boldsymbol{L}^{*'}\boldsymbol{L}^*)$$

15 / 52

Victim Statistics

	Descriptives		Correlations			
Item	Mean	Std Dev	q39	q40	q41	q42
q39	2.11	1.25	1.000	0.858	0.727	0.451
q40	2.08	1.25	0.868	1.000	0.818	0.470
q41	2.00	1.28	0.727	0.818	1.000	0.455
q42	1.47	0.94	0.451	0.470	0.455	1.000

Cronbach Coefficient Alpha with Deleted Variable

	Raw Variables		Std Variables	
Deleted Variable	Correlation with Total	Alpha	Correlation with Total	Alpha
q39: other students picked on me	0.813	0.811	0.799	0.806
q40: students made fun of me	0.875	0.784	0.857	0.781
q41: students called me names	0.792	0.821	0.781	0.813
q42: got hit and pushed	0.492	0.923	0.492	0.923

16 / 52

One Factor Model Solution

Item		Factor Pattern Factor1 ℓ_{i1}	Communalities h_i^2	Specific var: ψ_i
q39	other students picked on me	0.88	0.77	0.23
q40	students made fun of me	0.97	0.95	0.05
q41	students called me names	0.84	0.70	0.30
q42	got hit and pushed	0.49	0.24	0.76

(maximum likelihood estimation using the correlation matrix)

Residual Correlation Model with Uniqueness ($\hat{\psi}_i$) on the diagonal:

	q39	q40	q41	q42
q39	.2260	.0010	−.0110	.0178
q40	.0010	.0502	.0008	−.0097
q41	−.0110	.0008	.2962	.0419
q42	.0178	−.0097	.0419	.7576

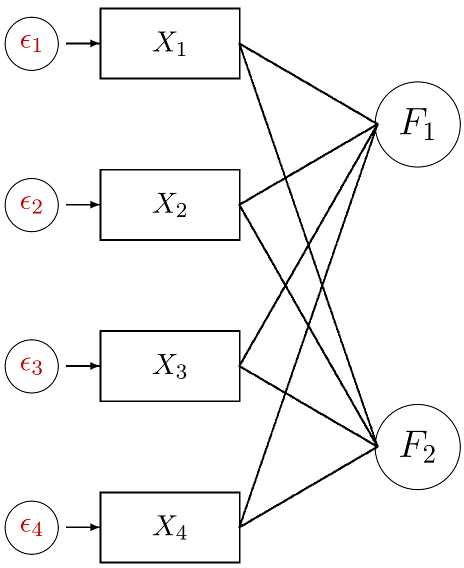
Root Mean Square Off-Diagonal Residuals: Overall $= 0.0196$

17 / 52

Generalization to m Common Factors

18 / 52

Generalization to m Common Factors



Picture of this for $p = 4$ and $m = 2$.

19 / 52

Generalization to m Common Factors

$$
\begin{aligned}
X_1 &= \mu_1 + \ell_{11}F_1 + \ell_{12}F_2 + \ldots + \ell_{1m}F_m + \epsilon_1 \\
X_2 &= \mu_2 + \ell_{21}F_1 + \ell_{22}F_2 + \ldots + \ell_{2m}F_m + \epsilon_2 \\
\vdots & \quad \vdots \\
X_p &= \mu_p + \ell_{p1}F_1 + \ell_{p2}F_2 + \ldots + \ell_{pm}F_m + \epsilon_p \\
\\
\boldsymbol{X} &= \boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{F} + \boldsymbol{\epsilon}
\end{aligned}
$$

- $E(\boldsymbol{F}) = \mathbf{0}$ and cov$(\boldsymbol{F}) = \boldsymbol{\Phi} = \boldsymbol{I}$... for now
- $E(\boldsymbol{\epsilon}) = \mathbf{0}$ and cov$(\boldsymbol{\epsilon}) = \boldsymbol{\Psi} =$ diag(ψ_i)
- cov$(\boldsymbol{F}, \boldsymbol{\epsilon}) = \mathbf{0}$
- $\boldsymbol{L}$ is matrix of Factor Loadings
- $\sum_{q=1}^{m} \ell_{iq}^2 = h_i^2 =$Communality of item i.
- ψ_i is Specific variance of item i.

Implication of m Model for Data

Using the model for $i = 1, \ldots, I$ observed variables and m common factors, the covariance matrix is

$$\boldsymbol{X} - \boldsymbol{\mu} = \boldsymbol{L}\boldsymbol{F} + \boldsymbol{\epsilon}$$

- Mean of $\boldsymbol{X} - \boldsymbol{\mu} = E(\boldsymbol{L}\boldsymbol{F} + \boldsymbol{\epsilon}) = \boldsymbol{L}E(\boldsymbol{F}) + E(\boldsymbol{\epsilon}) = \mathbf{0}$
- Covariance matrix for $\boldsymbol{X}$ or equivalently $\boldsymbol{X} - \boldsymbol{\mu}$:

$$
\begin{aligned}
\boldsymbol{\Sigma} &= E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{X} - \boldsymbol{\mu})'] \\
&= E[(\boldsymbol{L}\boldsymbol{F} + \boldsymbol{\epsilon})(\boldsymbol{L}\boldsymbol{F} - \boldsymbol{\epsilon})'] \\
&= E[(\boldsymbol{L}\boldsymbol{F} + \boldsymbol{\epsilon})(\boldsymbol{F}'\boldsymbol{L}' - \boldsymbol{\epsilon}')] \\
&= \boldsymbol{L}\underbrace{E[\boldsymbol{F}\boldsymbol{F}']}_{\boldsymbol{I}}\boldsymbol{L}' + \boldsymbol{L}\underbrace{E[\boldsymbol{F}1\boldsymbol{\epsilon}']}_{\mathbf{0}} + \underbrace{E[\boldsymbol{\epsilon}\boldsymbol{F}']}_{\boldsymbol{I}}\boldsymbol{L}' + \underbrace{E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}']}_{\boldsymbol{\Psi}} \\
&= \boldsymbol{L}\boldsymbol{L}' + \boldsymbol{\Psi}
\end{aligned}
$$

Implication of m Model for Data

The Common factor model implies :

$$\boldsymbol{\Sigma} = \begin{pmatrix} (\sum_q \ell_{1q}^2 + \psi_1) & \sum_q \ell_{1q}\ell_{2q} & \ldots & \sum_q \ell_{1q}\ell_{pq} \\ \sum_q \ell_{1q}\ell_{2q} & (\sum_q \ell_{2q}^2 + \psi_2) & \ldots & \sum_q \ell_{2q}\ell_{pq} \\ \vdots & \vdots & \ddots & \vdots \\ \sum_q \ell_{1q}\ell_{pq} & \sum_q \ell_{2q}\ell_{pq} & \ldots & (\sum_q \ell_{pq}^2 + \psi_I) \end{pmatrix}$$

- Variance for X_i:

$$\sigma_{ii} = \underbrace{\sum_{q=1}^{m} \ell_{im}^2}_{h_i^2} + \psi_i$$

- Covariance between X_i and X_k:

$$\sigma_{ik} = \sum_{q=1}^{m} \ell_{iq}\ell_{kq}$$

22 / 52

A Closer Look at Implied Covariance Matrix

- Covariance between $\boldsymbol{X}$ and $\boldsymbol{F}$:

$$\begin{aligned} \text{cov}(\boldsymbol{X}, \boldsymbol{F}) &= \text{cov}((\boldsymbol{X} - \boldsymbol{\mu}), \boldsymbol{F}) = E[(\boldsymbol{X} - \boldsymbol{\mu})(\boldsymbol{F} - \mathbf{0})'] \\ &= E[(\boldsymbol{LF} + \boldsymbol{\epsilon})(\boldsymbol{F})'] \\ &= \boldsymbol{L}\underbrace{E(\boldsymbol{FF}')}_{\boldsymbol{I}} + \underbrace{E(\boldsymbol{\epsilon F}')}_{\mathbf{0}} \\ &= \boldsymbol{L} = \{\ell_{iq}\}_{p \times q} \end{aligned}$$

- The correlation between observed variables and the common factors equal

$$\frac{\text{cov}(X_i, F_q)}{\sqrt{h_i^2 + \psi_i}\sqrt{1}} = \frac{\ell_{iq}}{\sqrt{h_i^2 + \psi_i}}$$

  The correlations $\rho(X_i, F_q)$ are called Structure Coefficients.

23 / 52

Summary of Terminology and Model Components

- F_q are Common factors (latent variables).
- ℓ_{iq} are Factor Loadings.
- ϵ_i are latent variables specific or Unique to item i.
- ψ_i are the unique variance or Specifcities.
- $h_i^2 = \sum_{q=1}^{m} \ell_{iq}$ are the common variance or Communalities.
- Correlation between X_i and F_q are Structure Coefficients:

$$\rho(X_i, F_q) = \frac{\ell_{iq}}{\sqrt{\sum_q^m \ell_{iq}^2 + \psi_i}} = \frac{\ell_{iq}}{\sqrt{h_i^2 + \psi_i}}$$

Example: Bully Scale

Basic Descriptive Statistics with a little item analysis:

	Variable	Mean	Std Dev	Corr w/ Total	Alpha
q36	upset students for fun	1.793	1.062	.75	.85
q37	in group teased students	1.881	0.997	.75	.85
q43	helped harass students	1.315	0.801	.56	.87
q44	teased other students	1.850	1.066	.77	.85
q49	mean to someone when angry	1.940	0.966	.52	.87
q50	spread rumors	1.381	0.844	.55	.87
q51	started arguments or conflicts	1.587	0.936	.58	.87
q52	encouraged people to fight	1.300	0.843	.55	.87
q53	excluded students from clique	1.696	0.978	.58	.87

Example: Bully Scale

Correlations:

	q36	q37	q43	q44	q49	q50	q51	q52	q53
q36	1.00								
q37	0.73	1.00							
q43	0.58	0.52	1.00						
q44	0.79	0.74	0.55	1.00					
q49	0.41	0.42	0.21	0.45	1.00				
q50	0.41	0.45	0.32	0.41	0.38	1.00			
q51	0.40	0.43	0.30	0.48	0.42	0.46	1.00		
q52	0.40	0.38	0.42	0.45	0.35	0.39	0.44	1.00	
q53	0.45	0.52	0.32	0.45	0.41	0.40	0.42	0.39	1.00

26 / 52

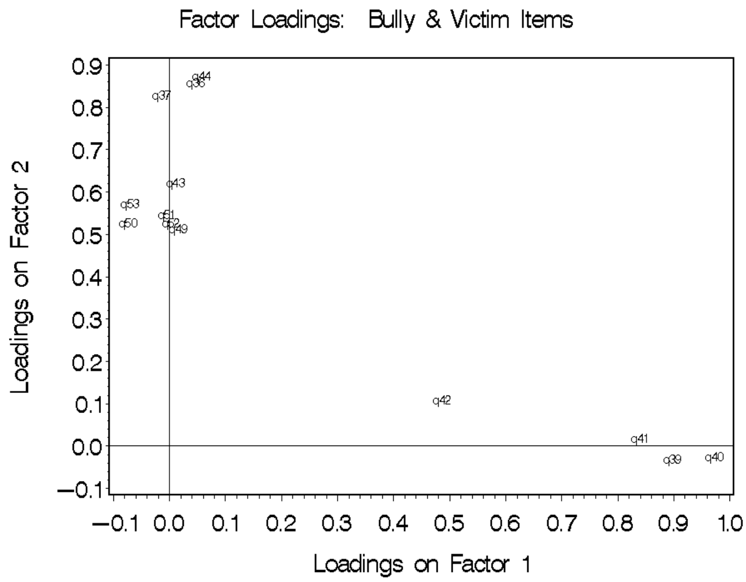
Example: Bully & Victim Scales

Correlations:

q36	q37	q43	q44	q49	q50	q51	q52	q53	q39	q40	q41	q42
1.0												
.73	1.0											
.58	.52	1.0										
.79	.74	.55	1.0									
.41	.42	.21	.45	1.0								
.41	.45	.32	.41	.38	1.0							
.40	.43	.30	.48	.42	.46	1.0						
.40	.38	.42	.45	.35	.39	.44	1.0					
.45	.52	.32	.45	.41	.40	.42	.39	1.0				
.03	-.03	.00	.02	-.02	-.07	.01	-.05	-.09	1.0			
.02	-.01	.01	.04	.00	-.08	-.02	.01	-.07	.85	1.0		
.06	.01	.01	.04	.08	-.04	.04	-.02	-.03	.72	.81	1.0	
.12	.10	.07	.10	.02	.00	.04	.06	.04	.45	.47	.45	1.0

What do you notice and what does this imply?

27 / 52



Factor Loadings

Item		Factor 1	Factor 2
Bully items			
q36	upset students for fun	.048	.866
q37	in group teased students	−.014	.837
q43	helped harass students	.011	.629
q44	teased other students	.058	.882
q49	mean to someone when angry	.016	.522
q50	spread rumors	−.073	.535
q51	started arguments or conflicts	−.005	.555
q52	encouraged people to fight	.004	.536
q53	excluded students from clique	−.070	.581
Victim items			
q39	other students picked on me	.898	−.024
q40	students made fun of me	.973	−.017
q41	students called me names	.839	.025
q42	got hit and pushed	.487	.118

Residual Correlation Matrix w/ Uniqueness on Diagonal

q36	q37	q43	q44	q49	q50	q51	q52	q53	q39	q40	q41	q42
0.25												
0.01	0.30											
0.04	−0.01	0.60										
0.03	0.00	−0.01	0.22									
−0.04	−0.02	−0.12	−0.01	0.73								
−0.05	0.01	−0.01	−0.05	0.11	0.71							
−0.08	−0.03	−0.04	−0.00	0.13	0.17	0.69						
−0.06	−0.06	0.10	−0.01	0.07	0.11	0.15	0.71					
−0.04	0.04	−0.04	−0.05	0.11	0.09	0.10	0.08	0.66				
0.01	0.01	0.00	−0.01	−0.02	0.01	0.03	−0.04	−0.02	0.19			
−0.01	−0.00	0.00	0.01	−0.00	−0.00	−0.01	0.02	0.00	0.00	0.05		
0.00	−0.00	−0.02	−0.02	0.07	0.01	0.04	−0.04	0.02	−0.01	0.00	0.29	
0.01	0.01	−0.01	−0.01	−0.03	−0.01	−0.01	−0.00	0.01	0.04	−0.01	0.03	0.75

	q41	q42
q41	0.29	0.03
q42	0.03	0.75

Root Mean Square Off-Diagonal Residuals: Overall = 0.052

Example Two: Fight Scale

Deleted Variable	Correlation with Total	Alpha	Label
q38	0.559236	0.875625	fought students could beat
q45	0.798226	0.823482	got in physical fight
q46	0.674541	0.852222	threatened to hurt or hit student
q47	0.786967	0.825238	physical fight because angry
q48	0.717408	0.844596	hit back when hit first

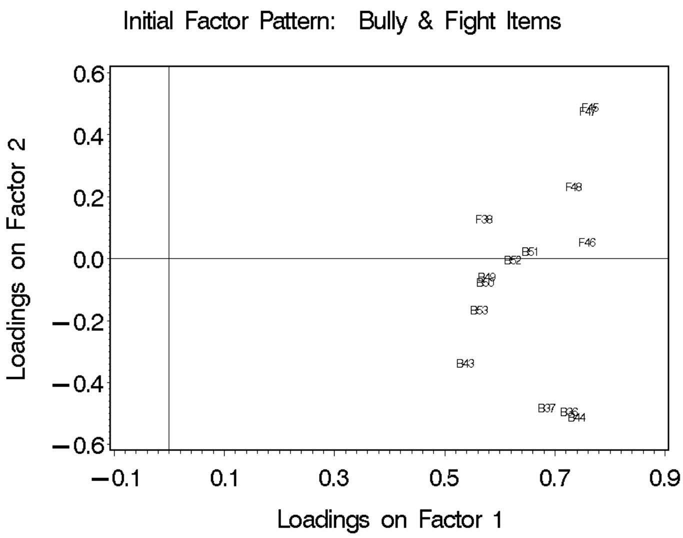
	q38	q45	q46	q47	q48
q38 fought students could beat	1.000	0.528	0.436	0.515	0.455
q45 got in physical fight	0.528	1.000	0.580	0.821	0.668
q46 threatened to hurt or hit student	0.436	0.580	1.000	0.609	0.615
q47 physical fight because angry	0.515	0.821	0.609	1.000	0.627
q48 hit back when hit first	0.455	0.668	0.615	0.627	1.000

Correlations between Fight and Bully

	Fight				
Bully	q38	q45	q46	q47	q48
q36	.331	.339	.469	.289	.428
q37	.315	.289	.458	.316	.347
q43	.335	.261	.375	.228	.260
q44	.359	.298	.509	.312	.422
q49	.323	.400	.461	.353	.462
q50	.305	.379	.504	.357	.413
q51	.302	.479	.498	.502	.506
q52	.315	.470	.583	.449	.478
q53	.207	.349	.407	.316	.358

Test of $H_o : \mathbf{\Sigma}_{bully,fight} = \mathbf{0}$ versus $H_o : \mathbf{\Sigma}_{bully,fight} \neq \mathbf{0}$
$F(45, 1367) = 8.72$, $p < .01$, canonical correlation $= 0.77$

32 / 52



33 / 52

Bully/Fight Factor Pattern

Scale	Item	Factor 1	Factor 2
Bully	q36	.726	−.485
Bully	q37	.686	−.473
Bully	q43	.538	−.330
Bully	q44	.740	−.504
Bully	q49	.577	−.049
Bully	q50	.574	−.068
Bully	q51	.655	.032
Bully	q52	.625	.006
Bully	q53	.563	−.158
Fight	q38	.572	.136
Fight	q45	.765	.498
Fight	q46	.759	.062
Fight	q47	.760	.485
Fight	q48	.734	.241

Root Mean Square Off-Diagonal Residuals: Overall $= 0.040$

34 / 52

Rotation

35 / 52

Factor Pattern not Unique

Two kinds:

- Orthogonal
- Oblique

Orthogonal: Let $\boldsymbol{T}$ be an orthogonal matrix such that

$$TT' = T'T = \boldsymbol{I}$$

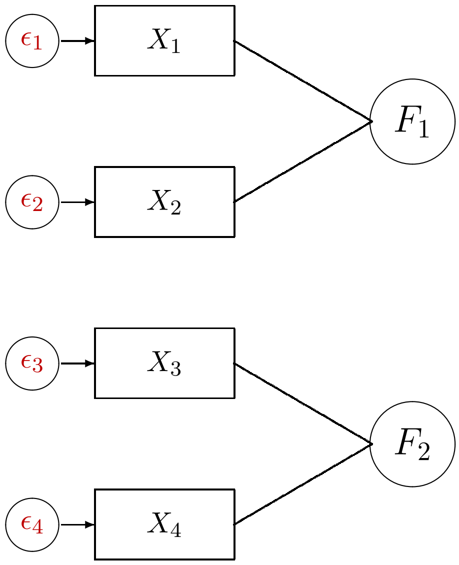
Therefore,

$$\begin{aligned} \boldsymbol{\Sigma} &= \boldsymbol{L}\boldsymbol{L}' + \boldsymbol{\Psi} \\ &= \boldsymbol{L}\underbrace{\boldsymbol{T}\boldsymbol{T}'}_{\boldsymbol{I}}\boldsymbol{L}' + \boldsymbol{\Psi} \\ &= \boldsymbol{L}^*\boldsymbol{L}^{*'} + \boldsymbol{\Psi} \end{aligned}$$

Still keeping $\text{cov}(\boldsymbol{F}) = \boldsymbol{I}$ and the model fit to data remains the same.

36 / 52

Desirable Structure



for easier interpretation:

Orthogonal Rotations

- Change in coordinates corresponding to a rigid rotation of the axses.
- For interpretation, it's easier if there is a "Simple Structure" or "Simple Solution" such that each factor has either large or small (near zero) loadings.
- There are 10 methods for orthogonal rotation in SAS PROC FACTOR.
- The most commonly used one is VARIMAX that is due to Henry Kaiser (1958).
  - Maximize the sum of the variances of the vectors of loadings.
  - Formally

$$\mathcal{V} = \sum_{i=1}^{p} (\ell_{iq}^2 - \bar{\ell}_{iq}^2)^2$$

- If you have a target matrix, then use PROCRUSTES.

Varimax for Victim & Bully Items

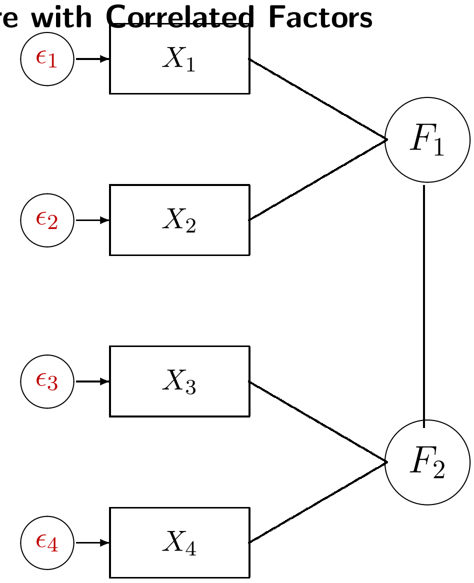
	Item	Initial Solution Fac 1	Initial Solution Fac 2	VARIMAX Fac 1	VARIMAX Fac 2
Bully items					
q36	upset students for fun	.048	.866	.864	.075
q37	in group teased students	−.014	.837	.837	.013
q43	helped harass students	.011	.629	.629	.031
q44	teased other students	.058	.882	.880	.086
q49	mean to someone when angry	.016	.522	.521	.032
q50	spread rumors	−.073	.535	.537	−.056
q51	started arguments or conflicts	−.005	.555	.555	.012
q52	encouraged people to fight	.004	.536	.535	.021
q53	excluded students from clique	−.070	.581	.583	−.052
Victim items					
q39	other students picked on me	.898	−.024	−.052	.898
q40	students made fun of me	.973	−.017	−.047	.973
q41	students called me names	.839	.025	−.001	.840
q42	got hit and pushed	.487	.118	.102	.491

About the same.

Varimax Bully/Fight Factor Pattern

Scale	Item	Initial Fac 1	Initial Fac 2	VARIMAX Fac 1	VARIMAX Fac 2
Bully	q36	.726	−.485	.228	.843
Bully	q37	.686	−.473	.206	.808
Bully	q43	.538	−.330	.189	.602
Bully	q44	.740	−.504	.226	.866
Bully	q49	.577	−.049	.403	.416
Bully	q50	.574	−.068	.388	.429
Bully	q51	.655	.032	.515	.406
Bully	q52	.625	.006	.475	.406
Bully	q53	.563	−.158	.321	.489
Fight	q38	.572	.136	.521	.274
Fight	q45	.765	.498	.904	.127
Fight	q46	.759	.062	.613	.452
Fight	q47	.760	.485	.891	.134
Fight	q48	.734	.241	.712	.301

Desirable (Oblique) Structure with Correlated Factors



for easier interpretation:

Oblique Rotation: Bully/Fight Factor Pattern

		Initial		VARIMAX		OBLIMIN	
Scale	Item	Fac 1	Fac 2	Fac 1	Fac 2	Fac 1	Fac 2
Bully	q36	−.485	.726	.228	.843	−.040	.895
Bully	q37	−.473	.686	.206	.808	−.053	.862
Bully	q43	−.330	.538	.189	.602	.002	.630
Bully	q44	−.504	.740	.226	.866	−.050	.922
Bully	q49	−.049	.577	.403	.416	.316	.339
Bully	q50	−.068	.574	.388	.429	.294	.359
Bully	q51	.032	.655	.515	.406	.449	.288
Bully	q52	.006	.625	.475	.406	.403	.302
Bully	q53	−.158	.563	.321	.489	.195	.452
Fight	q38	.136	.572	.521	.274	.503	.132
Fight	q45	.498	.765	.904	.127	.999	−.173
Fight	q46	.062	.759	.613	.452	.546	.306
Fight	q47	.485	.760	.891	.134	.982	−.161
Fight	q48	.241	.734	.712	.301	.715	.096

From OBLIMIN, $\hat{\rho}(F_1, F_2) = 0.564$

Estimation Methods

Estimation method (Factor extraction)

- Eigen-decomposition based
  - Eigen-decomposition of $\boldsymbol{S}$ or the "principal components" solution of the factor model.
  - Iterative eigen-decompositions of $\boldsymbol{S} - \tilde{\boldsymbol{\Psi}}$ or the "Principal factor" or "Principal Axis" solution.
- Maximum likelihood estimation — We now must assume that $\boldsymbol{F}$ and $\boldsymbol{\epsilon}$ are multivariate normal. Tends to fit data better & yields scale invariance (ie., use either $\boldsymbol{S}$ or $\boldsymbol{R}$).

Principal Components Solution of Factor Model

Let $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \ldots \geq \hat{\lambda}_p$ be the eigenvalues of the sample covariance matrix $\boldsymbol{S}$ and $\hat{\boldsymbol{e}}_1, \hat{\boldsymbol{e}}_2, \ldots, \hat{\boldsymbol{e}}_p$ be corresponding eigenvectors. Let $m < p$ (ie, the number of factors be less than the number of X variables), then the matrix of factor loadings equals

$$\tilde{\boldsymbol{L}} = \left(\sqrt{\hat{\lambda}_1}\hat{\boldsymbol{e}}_1, \sqrt{\hat{\lambda}_2}\hat{\boldsymbol{e}}_2, \ldots, \sqrt{\hat{\lambda}_m}\hat{\boldsymbol{e}}_m\right)$$

The estimate specific variances are provided by the diagonal elements of the matrix $\boldsymbol{S} - \tilde{\boldsymbol{L}}\tilde{\boldsymbol{L}}'$, So

$$\tilde{\boldsymbol{\Psi}} = \begin{pmatrix} \tilde{\psi}_1 & 0 & \ldots & 0 \\ 0 & \tilde{\psi}_2 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \tilde{\psi}_m \end{pmatrix} \quad \text{where} \quad \tilde{\psi}_i = s_{ii} - \sum_{q=1}^{m} \tilde{\ell}_{iq}^2$$

And estimates of communalities are $\tilde{h}_i^2 = \sum_{q=1}^m \tilde{\ell}_{iq}^2$

Alternatively, use $\boldsymbol{R}$ instead of $\boldsymbol{S}$ (will get a different result).

Principal Factor or Axis Solution

1. Get an initial estimate of $\boldsymbol{\Psi}$. Most common choice is the square of the multiple correlation (SMC) coefficient between X_i and the other $p-1$ variables. The SMC is the diagonal element of $\boldsymbol{R}^{-1}$ and is our initial estimate of the ψ_i's; that is,

$$\psi_i^* = \text{ the } i^{th}\text{diagonal element of } \boldsymbol{R}^{-1} = \text{SMC}$$

2. Find the eigenvalues λ_q^* and eigenvectors $\boldsymbol{e}_i^*$ of $\boldsymbol{R} - \text{diag}(\psi_i^*)$ and set $\boldsymbol{L}^* = (\sqrt{\lambda}_1^* \boldsymbol{e}_1^*, \sqrt{\lambda}_2^* \boldsymbol{e}_2^*, \ldots, \sqrt{\lambda}_m^* \boldsymbol{e}_m^*)$.

3. New estimates of ψ

$$\psi_i^* = 1 - \sum_{q=1}^{m} \ell_{iq}^{*2} = 1 - h_i^{*2}.$$

4. Repeat steps 2 through 3 until convergence.

46 / 52

A Modern Solution: Maximum Likelihood Estimation

- With principal component solution, extraction of an additional factor does not effect the values already extracted; this is not the case with the MLE or the principal factor method.
- Both MLE and principal factor methods can run into problems (Heywood cases, improper solutions) where the ψs want to be negative.
- You get different results if you use $\boldsymbol{S}$ or $\boldsymbol{R}$ (ie, not scale invariant); however, this is not the case for MLE:

  If ℓ_{iq} is a loading from a (population) correlation matrix, then $\ell_{iq}\sigma_{ii}$ is the corresponding loading from the (population) covariance matrix.
- You can get a better fit via MLE.
- Statistical tests become possible.
- With MLE it becomes possible to set values for ℓ_{iq} (and/or ψ_i).... a confirmatory analysis...

47 / 52

Maximum Likelihood Estimation

For MLE, we need to add the assumption that

$$\boldsymbol{X} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}) \quad i.i.d \quad \text{where } \boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}' + \boldsymbol{\Psi}.$$

(or $\boldsymbol{X} - \boldsymbol{\mu}$).
Note that

$$\begin{aligned}
\boldsymbol{\Sigma} &= \boldsymbol{L}\boldsymbol{\Phi}\boldsymbol{L}' + \boldsymbol{\Psi} \\
\underbrace{\boldsymbol{\Psi}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Psi}^{-1/2}}_{\boldsymbol{\Sigma}^*} &= \underbrace{\boldsymbol{\Psi}^{-1/2}\boldsymbol{L}\boldsymbol{\Phi}^{1/2}}_{\boldsymbol{L}^*}\underbrace{\boldsymbol{\Phi}^{1/2}\boldsymbol{L}'\boldsymbol{\Psi}^{-1/2}}_{\boldsymbol{L}^{*'}} + \boldsymbol{I} \\
\boldsymbol{\Sigma}^* &= \boldsymbol{L}^*\boldsymbol{L}^{*'} + \boldsymbol{I}
\end{aligned}$$

Given $\boldsymbol{\Psi}$, we can get the $\boldsymbol{L}^*$'s:

$$\begin{aligned}
\boldsymbol{\Sigma}^* - \boldsymbol{I} = \boldsymbol{\Psi}^{-1/2}\boldsymbol{\Sigma}\boldsymbol{\Psi}^{-1/2} - \boldsymbol{I} &= \boldsymbol{P}\boldsymbol{\Lambda}\boldsymbol{P}' \\
&= \underbrace{\boldsymbol{P}\boldsymbol{\Lambda}^{1/2}}_{\boldsymbol{L}^*}\underbrace{\boldsymbol{\Lambda}^{1/2}\boldsymbol{P}'}_{\boldsymbol{L}^{*'}}
\end{aligned}$$

48 / 52

MLE continued

Starting with a initial "guess" for $\boldsymbol{\Psi}$ this gives us an initial estimate of $\boldsymbol{L}$.

Using an optimization algorithm, iteratively up-date estimates by maximizing

$$\mathcal{L}(\boldsymbol{L}, \boldsymbol{\Psi}) = -\frac{n}{2}(\ln(|\boldsymbol{\Sigma}|) + \text{Tr}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1})) + \text{ constant}$$

or equivalently minimize

$$\text{F}(\boldsymbol{L}, \boldsymbol{\Psi}) = \ln(|\boldsymbol{\sigma}|) + \text{Tr}(\boldsymbol{S}\boldsymbol{\Sigma}^{-1})) - \ln(|\boldsymbol{S}|) - p$$

49 / 52

Rotation & Indeterminacy

- To improve interpretation, we can rotate factor loadings using either an orthogonal or oblique method.
- With oblique,

$$\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{\Phi}\boldsymbol{L}'$$

The $m \times m$ matrix $\boldsymbol{\Phi}$ is the Factor Pattern Matrix that contains correlations between the factors.

- The $p \times m$ matrix $\boldsymbol{L}\boldsymbol{\Phi}$ is the Factor Structure Matrix that contains the covariances (or correlations) between the $\boldsymbol{X}_{p\times 1}$ observed variables and the $\boldsymbol{F}_{m\times 1}$ latent variables.
- Due to the indeterminacy, computing factor scores is not a reasonable thing to do event though many computer programs will give them to you if asked.
- With MLE, you can fix parameters (e.g., set some loadings equal to 0) and this leads to "confirmatory factor analysis."

50 / 52

Assessing Model Fit to Data

51 / 52

Or How Many Factors Needed?

Data points: There are p variances and $p(p-1)/2$ covariances.

Parameters to be estimated: There are pm Factor loadings and p unique variances.

- Residual correlation matrix
- Root mean squared errors
- Statistical tests for number of factors
- Proportion of total sample variance due to the q^{th} factor
- Proportion of total sample variance accounted for by the m factors.
- Others.

52 / 52