

Multivariate: Factor analysis

Table of contents

1	Goals	2
1.1	Goals	2
2	PCA vs FA	2
2.1	Statistical measurement	2
2.2	Differences between PCA and FA	4
2.3	Summary	5
3	Data Example	6
3.1	Measure and variables	6
3.2	Output of the analysis	7
4	FA details	12
4.1	Exploratory vs confirmatory	12
4.2	FA model	13
4.3	Extraction methods	15
4.4	Summary	19
5	Number of factors and rotation	19
5.1	How many factors?	19
5.2	Rotation	26
6	Conclusion	27
6.1	Summary of this week	27
6.2	Next week	27

1 Goals

1.1 Goals

1.1.1 Goals of this lecture

- Factor analysis (FA)
 - **Dimension reduction**: reduce number of variables
- A **large set of (potentially correlated) observed variables**
 - Organize the **covariance** in those variables to a **smaller set** of orthogonal (uncorrelated) variables
- Similar to PCA but
 - **Assumptions** of FA are closer to what we expect in psychology

2 PCA vs FA

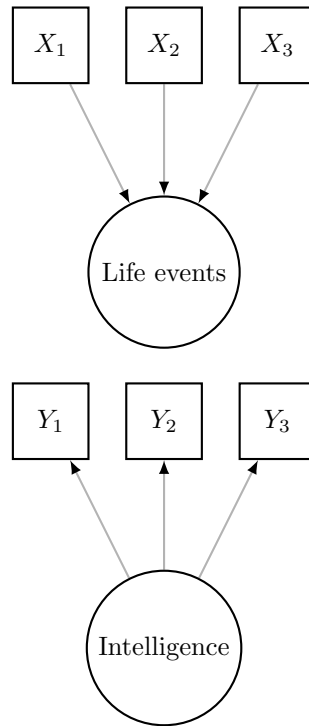
2.1 Statistical measurement

2.1.1 Measuring things is hard

- Psychology: we **cannot directly measure** some constructs
 - No ruler to measure “intelligence” or “introversion”
- We can **indirectly** measure what we really want to measure
 - Want to measure **intelligence**
 - * Math ability, verbal ability, spatial ability, reasoning, general knowledge, etc.
 - Intelligence is a **latent variable**
 - * Not *directly* observed

2.1.2 Formative vs reflective latent variables

- Formative factor
- Reflective factor



2.1.3 Measurement theory

- Psychometric theory: Latent variable is “true score”
- Observed score (Y) is a function of **true score** and **error**
 - True score is “real” score assuming no error (latent variable)
 - Error can be measurement error or random error

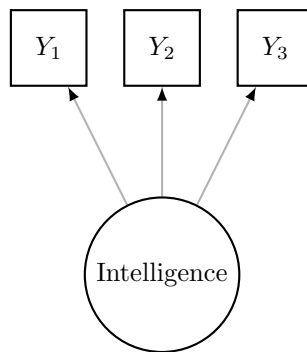
2.1.4 Measurement theory

- Observed score: $Y_i = T_i + e_{ij}$
 - True score (latent variable): T_i
 - Error: e_{ij}
- Variance of score:
 - $Var(Y_i) = Var(T_i) + Var(e_{ij})$
 - (Assuming no covariance between true score and error, which we do assume)

2.1.5 Measurement theory

- $Y_i = T_i + e_{ij}$
- FA partitions variance in each item into
 - **Common portion:** due to latent factors / true scores
 - **Unique portion:** due to error
- Big idea in **factor analysis**: Any correlations between items are due to what they have in common (i.e. a **common latent factor**)

2.1.6 Measurement theory



2.2 Differences between PCA and FA

2.2.1 Similar models, important differences

- Both PCA and FA are **data reduction** methods
 - You have many variables
 - You want to describe them using fewer dimensions
- Both PCA and FA are also **latent variable** methods
 - You have observed variables
 - They're related to unobserved (latent) variables
- Beyond that, several important **theoretical** and **statistical** differences

2.2.2 Common and unique variance

- **PCA: All variance is explained by latent variables**
 - If you retained all components, you'd *perfectly* re-create the observed variables
 - *Initial* communalities = 1.00
- **FA: Variance is divided into common and unique (error)**
 - Even if you retained all factors, there's still **measurement error**
 - Latent variables *never* explain *all* variance in observed
 - *Initial* communalities < 1.00

2.2.3 Causal ordering

- **PCA: Observed variables cause latent variables**
 - Latent variable is linear combination of observed
- **FA: Latent variables cause responses on observed variables**
 - Latent variable is a trait that causes a person to respond to the observed variables in a certain way

2.2.4 Variance or covariance / correlation?

- **PCA: Partitions variance in each variable**
 - Correlation / covariance largely ignored
 - **Variables don't even need to be correlated**
- **FA: Partitions variance, but in the service of splitting into common and unique portion**
 - Correlations between variables define "common variance"
 - **Variables related to the same factor are correlated**

2.3 Summary

2.3.1 Summary

- Factor analysis and PCA are similar
 - FA assumes measurement error
 - FA assumes latent factors cause responses
 - * FA relies on correlations to get at this

3 Data Example

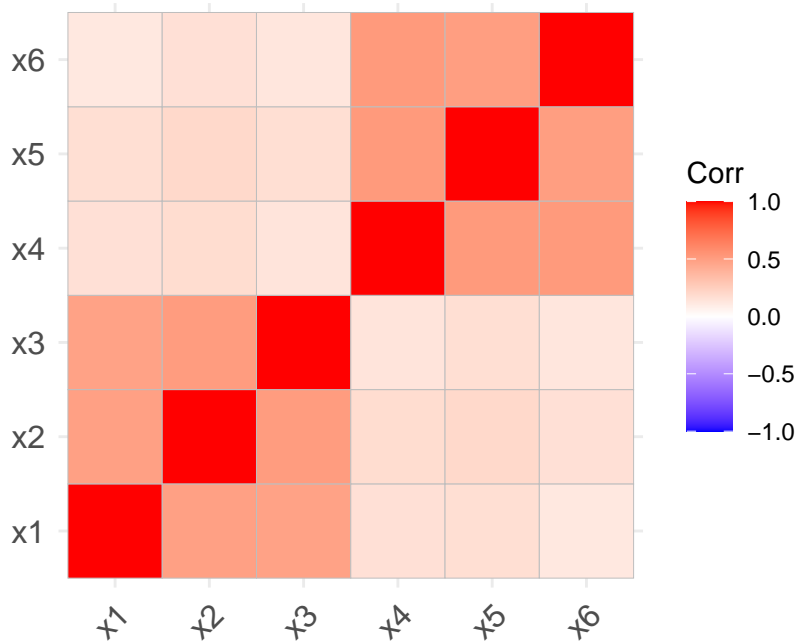
3.1 Measure and variables

3.1.1 Simulated data

- Similar to previous data
 - **1000 subjects**
 - 6 continuous variables
- Covariance matrix

	x1	x2	x3	x4	x5	x6
x1	1.871	0.912	0.944	0.312	0.344	0.226
x2	0.912	1.830	0.994	0.367	0.385	0.315
x3	0.944	0.994	2.059	0.287	0.362	0.267
x4	0.312	0.367	0.287	2.150	1.112	1.091
x5	0.344	0.385	0.362	1.112	2.117	1.041
x6	0.226	0.315	0.267	1.091	1.041	2.016

- Color-coded correlation matrix



3.1.2 Observed and latent variables

- Observed variables
 - 6 variables
 - **These are all Y variables:** they are predicted by the latent variables
- Latent variables
 - These are the X variables
 - They are the **factors**
 - We *create* them in the analysis

3.2 Output of the analysis

3.2.1 Data reduction

- The idea behind FA is to reduce the number of variables
 - Start with **6 items**
 - * Want **fewer** than 6 factors
 - * How many fewer?
- I simulated the data to have 2 “clumps”
 - We talked about this the past few weeks
 - So I’ll show you a **2 factor model** *to start*

3.2.2 FA results

1. Loadings

- Relation between latent factor (X) and observed variable (Y)
 - Matrix with rows = # items, columns = # **factors**
 - High loading = that X is highly related to that Y
- Think: correlation or standardized regression coefficient
 - Range from -1 to 1

3.2.3 Model results: Loadings in R

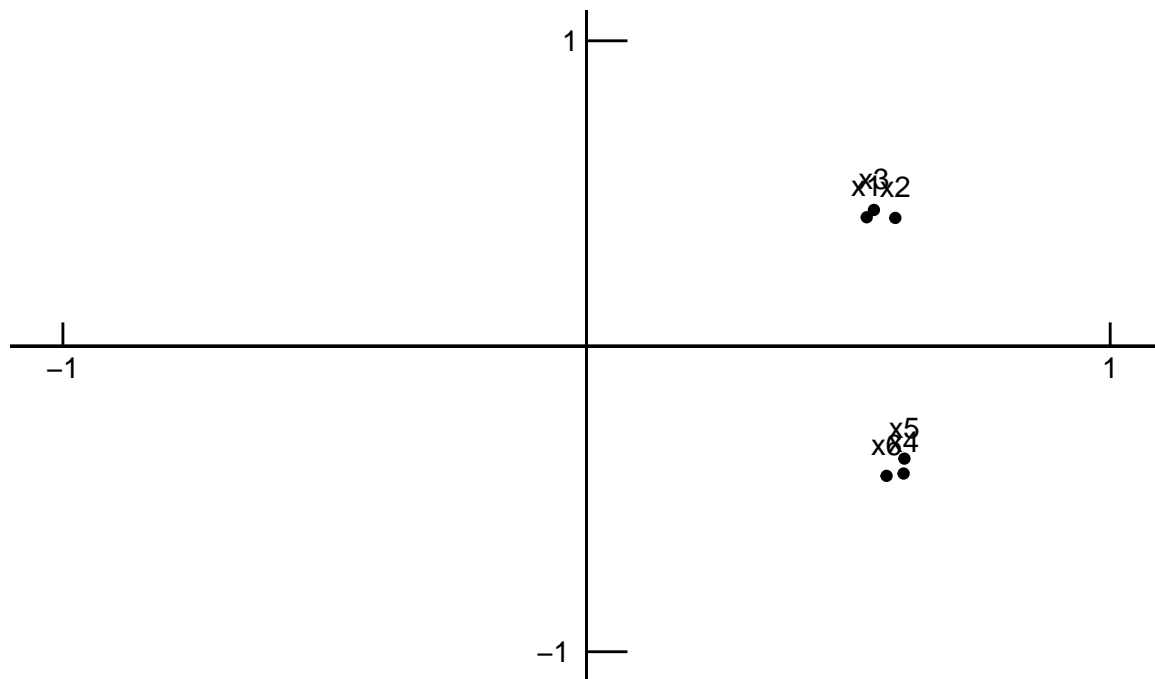
Loadings:

	PA1	PA2
x1	0.535	0.422
x2	0.590	0.420
x3	0.549	0.446
x4	0.605	-0.417
x5	0.607	-0.368
x6	0.573	-0.424

	PA1	PA2
SS loadings	1.999	1.043
Proportion Var	0.333	0.174
Cumulative Var	0.333	0.507

3.2.4 Model results: Loadings in SPSS

3.2.5 Loadings



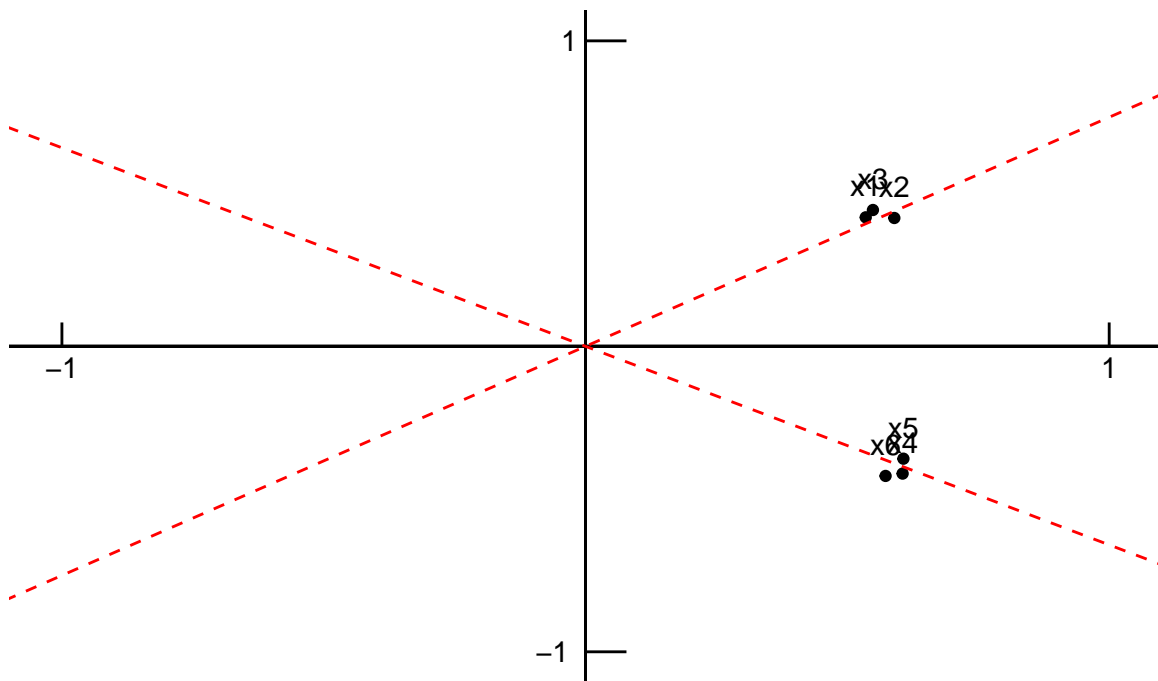
Factor Matrix^a		
	Factor	
	1	2
x1	.535	.422
x2	.590	.420
x3	.549	.446
x4	.605	-.417
x5	.607	-.368
x6	.573	-.424
Extraction Method: Principal Axis Factoring.		
a. 2 factors extracted. 7 iterations required.		

Total Variance Explained						
Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.490	41.506	41.506	1.999	33.309	33.309
2	1.537	25.612	67.118	1.043	17.383	50.691
3	.528	8.806	75.924			
4	.499	8.315	84.239			
5	.482	8.027	92.267			
6	.464	7.733	100.000			
Extraction Method: Principal Axis Factoring.						

3.2.6 Simple structure and rotation

- Solution has **simple structure** if each item has **high loadings** on only one factor and **near zero loadings** on all other factor
 - i.e., points are near the axes
 - Easier to interpret: items only relate to one axis
- **Rotated solution** rotates the axes to get closer to *simple structure*
 - We'll look at some different ways to rotate the solution
 - * Conceptual version now
 - Easier to **interpret** a solution that has simple structure

3.2.7 Loadings on rotated axes



3.2.8 FA results

2. Communalities

- Remember that we don't retain all the factors
- Communalities are the proportion of variance in Y that's explained by the factors (X) that you do retain

- Think: $R^2_{multiple}$ for X s predicting Y s
 - This is the normal order (unlike PCA): X predicts Y

3.2.9 Model results: Communalities in R

x1	x2	x3	x4	x5	x6
0.4635790	0.5250197	0.5010631	0.5416307	0.5028439	0.5079958

3.2.10 Model results: Communalities in SPSS

Communalities		
	Initial	Extraction
x1	.318	.465
x2	.352	.524
x3	.334	.500
x4	.368	.540
x5	.355	.504
x6	.349	.508
Extraction Method: Principal Axis Factoring.		

3.2.11 FA overview

- **Loadings** tell us *how items are correlated with factors*
 - Simple structure makes loadings more interpretable
 - Use rotation to try to get simple structure
- **Communalities** tell us how much *variance* in the items is *explained* by the factors we kept

4 FA details

4.1 Exploratory vs confirmatory

4.1.1 Exploratory vs confirmatory FA

- Two kinds of factor analysis
 - Obviously related, but also different models
- 1. **Exploratory factor analysis (EFA)** in this course
 - It is a “classic multivariate technique”
- 2. **Confirmatory factor analysis (CFA)** discussed briefly
 - CFA falls under structural equation modeling (SEM)

4.1.2 Exploratory factor analysis (EFA)

- EFA **explores** the factor structure of the variables
 - Largely *atheoretical*
 - *Discover* how many factors may be present
 - Few (if any) pre-conceptions about *which items may have high or low loadings on which factors*

4.1.3 Confirmatory factor analysis (CFA)

- CFA **confirms** a pre-existing factor structure
 - Requires *theory* to construct
 - Hypothesize a *specific number of factors*
 - Allow each item to *load on only one factor*
 - * All other loadings are 0

4.1.4 EFA vs CFA: matrix of loadings

- EFA

Item	F1	F2	F3
1	0.618	0.094	-0.049
2	0.440	-0.075	0.065
3	0.671	0.037	0.041
4	0.031	0.731	-0.079
5	0.126	0.705	0.053
6	0.265	0.296	0.603
⋮	⋮	⋮	⋮

- CFA

Item	F1	F2	F3
1	0.620	0	0
2	0.450	0	0
3	0.665	0	0
4	0	0.725	0
5	0	0.689	0
6	0	0	0.613
⋮	⋮	⋮	⋮

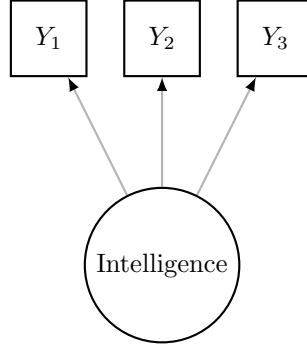
- Zeroes are “fixed”: we specify that **those loadings are 0** so don’t estimate them

4.2 FA model

4.2.1 FA model

$$Y_i = T_i + e_{ij}$$

- FA partitions variance in each item into
 - **Common portion:** due to latent factors / true scores
 - **Unique portion:** due to error
- Big idea: Any correlations between items are due to what they have in common (i.e. a **common latent factor**)



4.2.2 FA model

4.2.3 FA model

- Partition variance of each item into **common** and **unique** portions
 - **Common portion:** due to latent factors
 - * Correlations between variables due to common latent factor
 - **Unique portion:** Error (measurement or otherwise)

4.2.4 FA model

- There are p variables (items) and m factors
- $\mathbf{R}_{YY} = \mathbf{A}\mathbf{R}_F\mathbf{A}' + \mathbf{D}^2$
 - $\mathbf{R}_{YY} = p \times p$ matrix of **observed item correlations**
 - $\mathbf{A} = p \times m$ matrix of **loadings**
 - $\mathbf{R}_F = m \times m$ matrix of **correlations between factors**
 - $\mathbf{D}^2 = p \times p$ matrix of **unique variances**

4.2.5 Common variance portion

$$\mathbf{A} \mathbf{R}_F \mathbf{A}' =$$

$$\begin{bmatrix} a_{1,1} & a_{1,2} \\ a_{2,1} & a_{2,2} \\ a_{3,1} & a_{3,2} \\ a_{4,1} & a_{4,2} \\ a_{5,1} & a_{5,2} \\ a_{6,1} & a_{6,2} \end{bmatrix} \begin{bmatrix} \sigma_{F_1}^2 & \sigma_{F_1 F_2} \\ \sigma_{F_2 F_1} & \sigma_{F_2}^2 \end{bmatrix} \begin{bmatrix} a_{1,1} & a_{2,1} & a_{3,1} & a_{4,1} & a_{5,1} & a_{6,1} \\ a_{1,2} & a_{2,2} & a_{3,2} & a_{4,2} & a_{5,2} & a_{6,2} \end{bmatrix}$$

- The common (shared) portion of the variance involves the **correlations among factors** (\mathbf{R}_F) and the **loadings** (\mathbf{A})

4.2.6 Unique variance portion

- The matrix of “uniquenesses” is a **diagonal** matrix
 - Items are related only through **common factor**
 - There are *no correlations between uniquenesses across variables*

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 & 0 & 0 & 0 & 0 \\ 0 & d_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & d_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & d_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & d_5 & 0 \\ 0 & 0 & 0 & 0 & 0 & d_6 \end{bmatrix}$$

- d_1 is the “uniqueness” for item 1
 - d_1^2 is the unique **variance** for item 1

4.3 Extraction methods

4.3.1 Extraction methods

- Two kinds of **extraction** (estimation) for EFA
 1. **Principal axis factoring** (PAF), also called principal factor analysis
 2. **Maximum likelihood factor analysis** (MLFA)
- Either method is fine, but use PAF if items are *not normally distributed*
 - Fabrigar, Wegener, MacCallum, & Strahan (1999)
 - Osborne & Costello (2005)

4.3.2 Principal axis factoring

- Uses “reduced” correlation matrix ($\mathbf{R}_{reduced}$)
 - Diagonal of 1s replaced with communalities
 - * Why? More in a minute
- Perform PCA on reduced correlation matrix

- Iterate between loadings and correlation matrix until **observed correlations** and **model-implied correlations** are sufficiently close
 - What? More in a minute

4.3.3 Maximum likelihood factor analysis

- Uses a **weighted** version of $\mathbf{R}_{reduced}$
 - Diagonal elements are divided by that item's uniqueness (d_i)
 - * Why? More in a minute
- Perform PCA on weighted reduced correlation matrix
- Iterate between loadings and correlation matrix until **observed correlations** and **model-implied correlations** are sufficiently close
 - What? More in a minute

4.3.4 Correlation matrix

$$\mathbf{R}_{YY} = \begin{bmatrix} 1 & r_{12} & r_{13} & r_{14} & r_{15} & r_{16} \\ r_{21} & 1 & r_{23} & r_{24} & r_{25} & r_{26} \\ r_{31} & r_{32} & 1 & r_{34} & r_{35} & r_{36} \\ r_{41} & r_{42} & r_{43} & 1 & r_{45} & r_{46} \\ r_{51} & r_{52} & r_{53} & r_{54} & 1 & r_{56} \\ r_{61} & r_{62} & r_{63} & r_{64} & r_{65} & 1 \end{bmatrix}$$

$$\mathbf{S}_{YY} = \begin{bmatrix} s_1^2 & s_{12} & s_{13} & s_{14} & s_{15} & s_{16} \\ s_{21} & s_2^2 & s_{23} & s_{24} & s_{25} & s_{26} \\ s_{31} & s_{32} & s_3^2 & s_{34} & s_{35} & s_{36} \\ s_{41} & s_{42} & s_{43} & s_4^2 & s_{45} & s_{46} \\ s_{51} & s_{52} & s_{53} & s_{54} & s_5^2 & s_{56} \\ s_{61} & s_{62} & s_{63} & s_{64} & s_{65} & s_6^2 \end{bmatrix}$$

- **Off-diagonal** elements are **common only**
 - Variables are related by what they have in **common**
- **Diagonal** involve **both common and unique**
 - FA only cares about common factors
 - Modify diagonal to make it **common only**

4.3.5 Principal axis factoring

$$\mathbf{R}_{reduced} = \begin{bmatrix} h_1^2 & r_{12} & r_{13} & r_{14} & r_{15} & r_{16} \\ r_{21} & h_2^2 & r_{23} & r_{24} & r_{25} & r_{26} \\ r_{31} & r_{32} & h_3^2 & r_{34} & r_{35} & r_{36} \\ r_{41} & r_{42} & r_{43} & h_4^2 & r_{45} & r_{46} \\ r_{51} & r_{52} & r_{53} & r_{54} & h_5^2 & r_{56} \\ r_{61} & r_{62} & r_{63} & r_{64} & r_{65} & h_6^2 \end{bmatrix}$$

4.3.6 Principal axis factoring

- Elements on diagonal are **initial communalities**
 - What each variable has in **common** with all other variables
 - Squared multiple correlation (SMC) of that variable *predicted by all other variables*
 - * h_1^2 is the $R_{multiple}^2$ for: $\hat{Y}_1 = b_0 + b_1Y_2 + b_2Y_3 + b_3Y_4 + b_4Y_5 + b_5Y_6$

4.3.7 Principal axis factoring

- Remember PCA?
 - No measurement error
 - Initial communalities = 1.00
 - * Using correlation matrix in PCA: 1s on diagonal
 - So actually the same idea
 - * Diagonal is common variance only
 - * In PCA, **everything** is common variance

4.3.8 Principal axis factoring

- Perform a PCA on the **reduced correlation matrix**
- Pattern of eigenvalues will be similar for PCA and PAF
 - Diagonal is reduced from all 1s to initial communalities
 - Eigenvalues are similarly reduced
 - Scree plot is the same shape, just shifted down for PAF

4.3.9 ML factor analysis

- Uses a **weighted** version of $\mathbf{R}_{reduced}$
 - Weights each initial communality by the inverse of its uniqueness
 - h_1^2 in PAF $\rightarrow \frac{h_1^2}{d_1^2}$ in MLFA
 - Increases values of main diagonal
 - Eigenvalues are larger compared to PAF or PCA
 - Scree plot can be different shape from PAF and PCA

4.3.10 Iterations in PAF and MLFA

- $\mathbf{R}_{reduced}$ to estimated loadings: $\hat{\mathbf{A}}_1$
 - Loadings to estimated correlation matrix: $\mathbf{R}_{estimated1} = \hat{\mathbf{A}}_1 \hat{\mathbf{A}}_1'$
- $\mathbf{R}_{estimated1}$ to estimated loadings: $\hat{\mathbf{A}}_2$
 - Loadings to estimated correlation matrix: $\mathbf{R}_{estimated2} = \hat{\mathbf{A}}_2 \hat{\mathbf{A}}_2'$
- $\mathbf{R}_{estimated2}$ to estimated loadings: $\hat{\mathbf{A}}_3$
 - Loadings to estimated correlation matrix: $\mathbf{R}_{estimated3} = \hat{\mathbf{A}}_3 \hat{\mathbf{A}}_3'$
- Repeat until the difference between estimated and observed correlation matrices is “small enough”

4.3.11 Heywood cases

- The iterative process sometimes causes problems
 - Heywood case = communality > 1 or loading > 1
- Causes: too few cases, bad start values, too many factors, too few factors, non-linear relationships between factors
- Some solutions:
 - Too few cases: drop items or add cases
 - Bad start values: use highest correlation of item with a **single other item** instead of SMC for initial communality

4.4 Summary

4.4.1 Summary

- EFA divides variance into
 - **Common** (latent factors)
 - **Unique** (error)
- Two approaches to estimating
 - Principal axis factoring (PAF)
 - Maximum likelihood (ML)

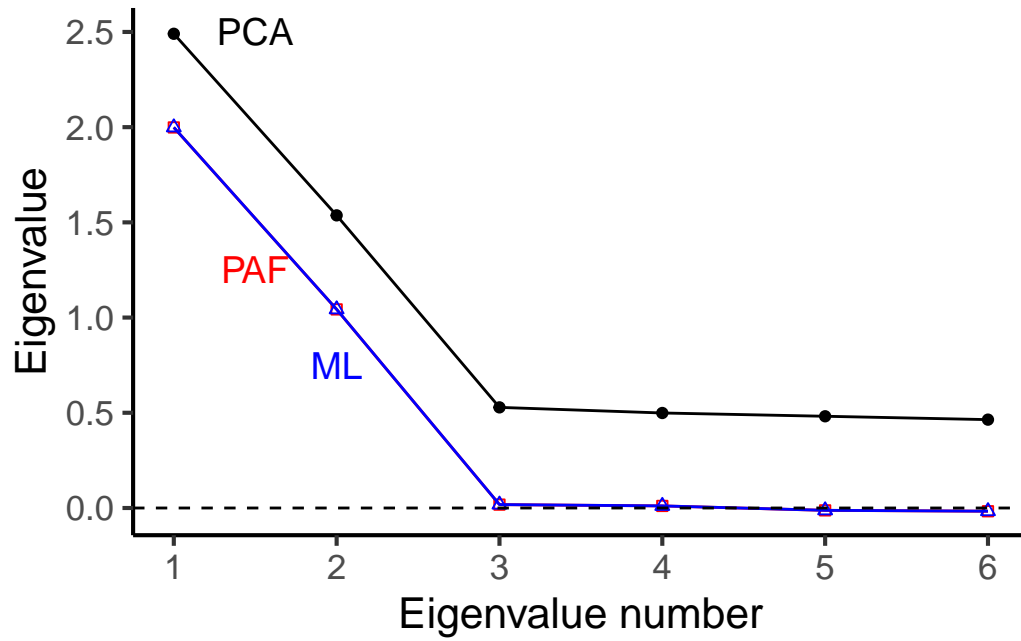
5 Number of factors and rotation

5.1 How many factors?

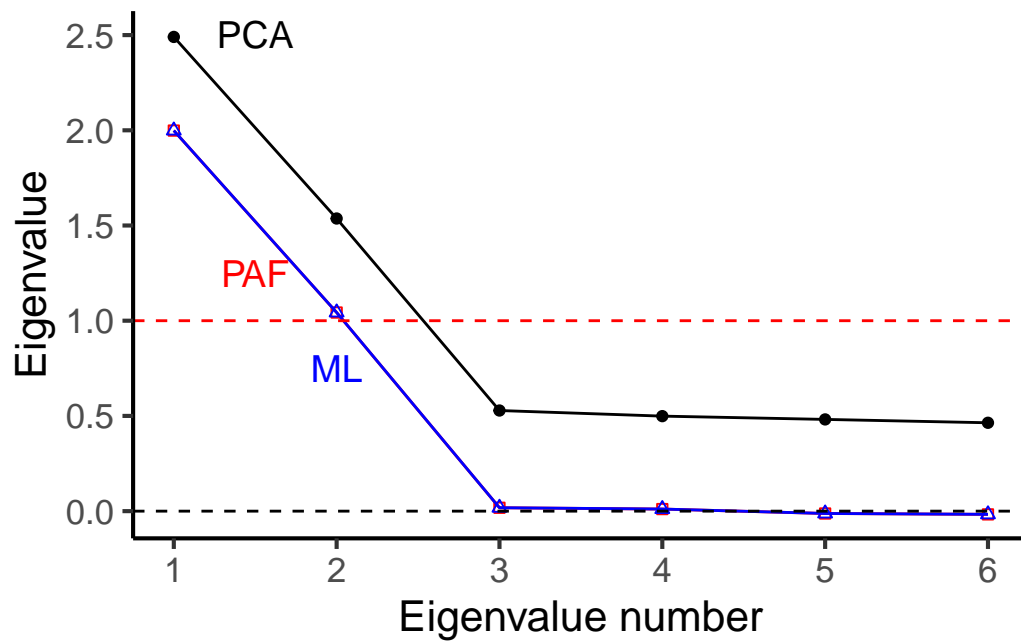
5.1.1 How many factors?

- Same options to pick number of factors as PCA
 - Bad: Kaiser criteria
 - Ok: Scree plot, proportion of variance accounted for
 - Good: Parallel analysis, MAP test
 - Also: Solution makes sense / theory
 - For MLFA: chi-square test

5.1.2 Scree plots

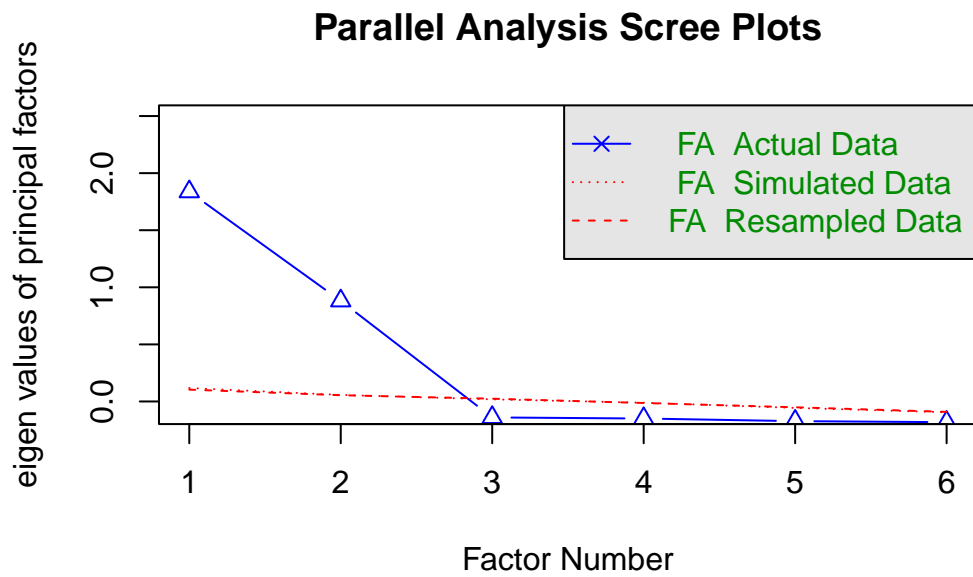


5.1.3 Kaiser criteria



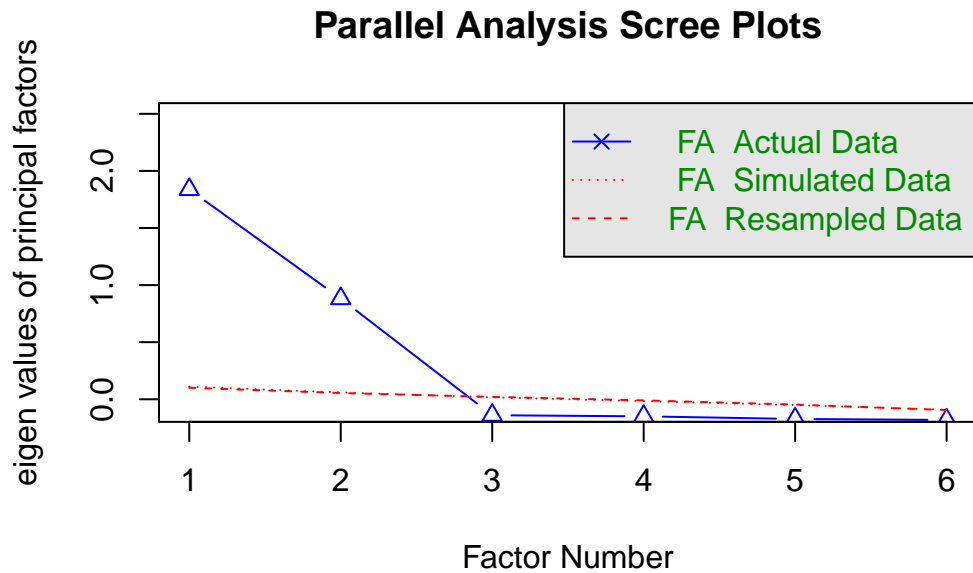
5.1.4 Parallel analysis in R

- PAF



Parallel analysis suggests that the number of factors = 2 and the number of components = 1

- MLFA



Parallel analysis suggests that the number of factors = 2 and the number of components = 1

5.1.5 Parallel analysis in SPSS

Run MATRIX procedure:

PARALLEL ANALYSIS:

Principal Axis / Common Factor Analysis

Specifications for this Run:

Ncases 1000
Nvars 6
Ndatsets 1000
Percent 95

Random Data Eigenvalues

	Root	Means	<u>Prcentyle</u>
1.000000	.112143	.157845	
2.000000	.059266	.092894	
3.000000	.020263	.044896	
4.000000	-.014555	.005881	
5.000000	-.051602	-.026956	
6.000000	-.095223	-.064847	

- SPSS gives you the eigenvalues for the original correlation matrix, not the reduced one, so...

5.1.6 MAP test in R

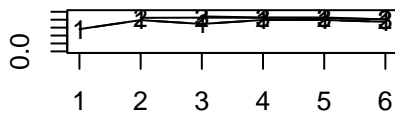
- PAF
 - Error: “imaginary eigen value”
 - No idea why
- MLFA

Total Variance Explained						
Factor	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.490	41.506	41.506	1.999	33.309	33.309
2	1.537	25.612	67.118	1.043	17.383	50.691
3	.528	8.806	75.924			
4	.499	8.315	84.239			
5	.482	8.027	92.267			
6	.464	7.733	100.000			

Extraction Method: Principal Axis Factoring.

Very Simple Structure Fit

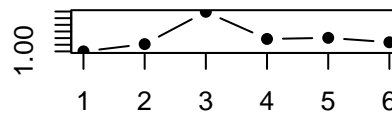
Very Simple Structure



Number of Factors

Complexity

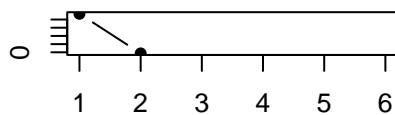
Complexity



Number of factors

Empirical BIC

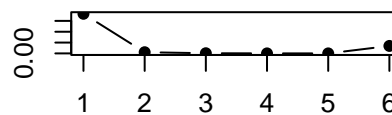
Empirical BIC



Number of factors

Root Mean Residual

SRMR



Number of factors

Number of factors

Call: vss(x = x, n = n, rotate = rotate, diagonal = diagonal, fm = fm,

n.obs = n.obs, plot = FALSE, title = title, use = use, cor = cor)

VSS complexity 1 achieves a maximum of Although the vss.max shows 5 factors, it is probab

VSS complexity 2 achieves a maximum of 0.85 with 3 factors

The Velicer MAP achieves a minimum of 0.1 with 2 factors

Empirical BIC achieves a minimum of -26.76 with 2 factors

Sample Size adjusted BIC achieves a minimum of -13.16 with 2 factors

Statistics by number of factors

	vss1	vss2	map	dof	chisq	prob	sqresid	fit	RMSEA	BIC	SABIC	complex
1	0.56	0.00	0.12	9	5.8e+02	3.1e-118	4.2	0.56	0.25	514	542	1.0
2	0.79	0.85	0.10	4	1.8e+00	7.8e-01	1.5	0.85	0.00	-26	-13	1.1
3	0.69	0.85	0.22	0	2.8e-02	NA	1.1	0.88	NA	NA	NA	1.3
4	0.79	0.85	0.42	-3	1.6e-09	NA	1.4	0.86	NA	NA	NA	1.1
5	0.79	0.84	1.00	-5	0.0e+00	NA	1.4	0.86	NA	NA	NA	1.1
6	0.75	0.81	NA	-6	2.6e+01	NA	1.8	0.81	NA	NA	NA	1.1

	eChisq	SRMR	eCRMS	eBIC
1	1.0e+03	1.8e-01	0.24	940
2	8.7e-01	5.4e-03	0.01	-27
3	1.7e-02	7.5e-04	NA	NA
4	8.5e-10	1.7e-07	NA	NA
5	2.7e-16	9.5e-11	NA	NA
6	3.6e+01	3.4e-02	NA	NA

5.1.7 MAP test in SPSS

5.1.8 Chi-square test: ML only

- Null hypothesis: This number of factors is sufficient
- Alternative hypothesis: Need more factors
- R

“The total number of observations was 1000 with Likelihood Chi Square = 1.77 with prob < 0.78”

Factor Analysis using method = ml

Call: fa(r = FA_data, nfactors = 2, rotate = "none", SMC = TRUE, warnings = TRUE, fm = "ml")

Standardized loadings (pattern matrix) based upon correlation matrix

	ML1	ML2	h2	u2	com
x1	0.52	0.44	0.46	0.54	1.9
x2	0.58	0.44	0.53	0.47	1.9
x3	0.53	0.46	0.50	0.50	2.0
x4	0.62	-0.40	0.54	0.46	1.7
x5	0.62	-0.35	0.50	0.50	1.6
x6	0.59	-0.41	0.51	0.49	1.8

	ML1	ML2
SS loadings	2.00	1.04
Proportion Var	0.33	0.17

Run MATRIX procedure:

Velicer's Minimum Average Partial (MAP) Test:

Eigenvalues

2.4904
1.5367
.5284
.4989
.4816
.4640

Average Partial Correlations

	squared	power4
.0000	.1180	.0268
1.0000	.1246	.0223
2.0000	.1003	.0250
3.0000	.2156	.1155
4.0000	.4226	.2868
5.0000	1.0000	1.0000

The smallest average squared partial correlation is
.1003

The smallest average 4rth power partial correlation is
.0223

The Number of Components According to the Original (1976) MAP Test is
2

The Number of Components According to the Revised (2000) MAP Test is
1

----- END MATRIX -----

Cumulative Var 0.33 0.51
 Proportion Explained 0.66 0.34
 Cumulative Proportion 0.66 1.00

Mean item complexity = 1.8

Test of the hypothesis that 2 factors are sufficient.

df null model = 15 with the objective function = 1.49 with Chi Square = 1483.93
 df of the model are 4 and the objective function was 0

The root mean square of the residuals (RMSR) is 0.01

The df corrected root mean square of the residuals is 0.01

The harmonic n.obs is 1000 with the empirical chi square 0.87 with prob < 0.93

The total n.obs was 1000 with Likelihood Chi Square = 1.77 with prob < 0.78

Tucker Lewis Index of factoring reliability = 1.006

RMSEA index = 0 and the 90 % confidence intervals are 0 0.032

BIC = -25.86

Fit based upon off diagonal values = 1

Measures of factor score adequacy

	ML1	ML2
Correlation of (regression) scores with factors	0.90	0.82
Multiple R square of scores with factors	0.80	0.68
Minimum correlation of possible factor scores	0.61	0.36

- SPSS

Goodness-of-fit Test		
Chi-Square	df	Sig.
1.769	4	.778

5.2 Rotation

5.2.1 Rotated solutions

- Same purpose for rotation

- Make the solution more interpretable and clean
- Same options for rotation in EFA as in PCA
 - Orthogonal rotation: varimax
 - Oblique rotation: oblimin, promax

6 Conclusion

6.1 Summary of this week

6.1.1 Summary of this week

- Factor analysis (FA)
 - Reduce # of variables (from p variables to $< p$ **factors**)
 - Loadings relate items to **factors**
 - Communalities are how much variance in each item is explained by latent factors
 - Focus on **common** variance due to latent factor
 - * Also measurement error

6.2 Next week

6.2.1 Next week

- A latent factor in FA is *continuous*
 - Continuous score on intelligence, depression, etc.
- What if you instead expect **discrete** groups? LCA / LPA
 - Distinct “classes” or “profiles” based on observed variables
 - Dimension reduction: Fewer classes than items
 - Model: **Almost completely different** from FA
 - * Conceptually: **Very similar** from FA