Statistics in Psychosocial Research
Lecture 8
Factor Analysis I

Lecturer: Elizabeth Garrett-Mayer

Motivating Example: Frailty

- We have a concept of what “frailty” is, but we can’t measure it directly.
- We think it combines strength, weight, speed, agility, balance, and perhaps other “factors”
- We would like to be able to describe the components of frailty with a summary of strength, weight, etc.

Factor Analysis

- Data reduction tool
- Removes redundancy or duplication from a set of correlated variables
- Represents correlated variables with a smaller set of “derived” variables.
- Factors are formed that are relatively independent of one another.
- Two types of “variables”:
  - latent variables: factors
  - observed variables

Frailty Variables

Speed of fast walk (+)
Speed of usual walk (+)
Time to do chair stands (-)
Arm circumference (+)
Body mass index (+)
Tricep skinfold thickness (+)
Shoulder rotation (+)

Upper extremity strength (+)
Pinch strength (+)
Grip strength (+)
Knee extension (+)
Hip extension (+)
Time to do Pegboard test (-)

Other examples

- Diet
- Air pollution
- Personality
- Customer satisfaction
- Depression

Applications of Factor Analysis

1. Identification of Underlying Factors:
   - clusters variables into homogeneous sets
   - creates new variables (i.e. factors)
   - allows us to gain insight to categories
2. Screening of Variables:
   - identifies groupings to allow us to select one variable to represent many
   - useful in regression (recall collinearity)

Applications of Factor Analysis

3. Summary:
   - Allows us to describe many variables using a few factors
4. Sampling of variables:
   - helps select small group of variables of representative variables from larger set
5. Clustering of objects:
   - Helps us to put objects (people) into categories depending on their factor scores

“Perhaps the most widely used (and misused) multivariate [technique] is factor analysis. Few statisticians are neutral about this technique. Proponents feel that factor analysis is the greatest invention since the double bed, while its detractors feel it is a useless procedure that can be used to support nearly any desired interpretation of the data. **The truth, as is usually the case, lies somewhere in between**. Used properly, factor analysis can yield much useful information; when applied blindly, without regard for its limitations, it is about as useful and informative as Tarot cards. **In particular, factor analysis can be used to explore the data for patterns, confirm our hypotheses, or reduce the Many variables to a more manageable number.**

-- Norman Streiner, *PDQ Statistics*

Orthogonal One Factor Model

Classical Test Theory Idea:

Ideal:

$$X_1 = F + e_1 \qquad \mathrm{var}(e_j) = \mathrm{var}(e_k)\ ,\ j \neq k$$
$$X_2 = F + e_2$$
$$\ldots$$
$$X_m = F + e_m$$

Reality:

$$X_1 = \lambda_1 F + e_1 \qquad \mathrm{var}(e_j) \neq \mathrm{var}(e_k)\ ,\ j \neq k$$
$$X_2 = \lambda_2 F + e_2$$
$$\ldots$$
$$X_m = \lambda_m F + e_m$$

(unequal “sensitivity” to change in factor)
(Related to Item Response Theory (IRT))

Key Concepts

- F is latent (i.e.unobserved, underlying) variable
- X's are observed (i.e. manifest) variables
- e_j is measurement error for X_j.
- λ_j is the "loading" for X_j.

Assumptions of Factor Analysis Model

- Measurement error has constant variance and is, on average, 0.

  $Var(e_j) = \sigma_j^2 \qquad E(e_j) = 0$

- No association between the factor and measurement error

  $Cov(F,e_j) = 0$

- No association between errors:

  $Cov(e_j ,e_k) = 0$

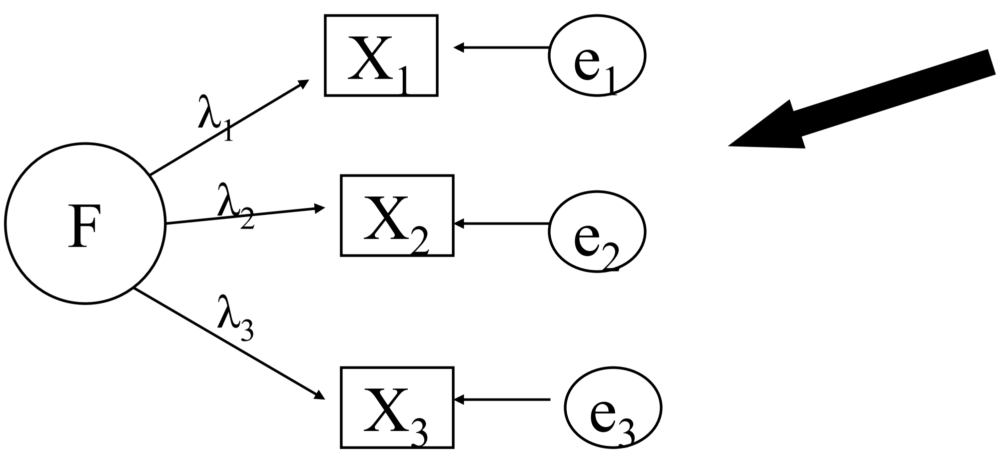
- **Local (i.e. conditional independence): Given the factor, observed variables are independent of one another.**

  $\mathbf{Cov(X_j ,X_k \mid F) = 0}$

Brief Aside in Path Analysis

Local (i.e. conditional independence): Given the factor, observed variables are independent of one another.

$$Cov(X_j ,X_k \mid F) = 0$$



X's are only related to each other through their common relationship with F.

Optional Assumptions

- We will make these to simplify our discussions
- F is “standardized” (think “standard normal”)

  Var(F) = 1   E(F) = 0

- X’s are standardized:
  - In practice, this means that we will deal with “correlations” versus “covariance”
  - This “automatically” happens when we use correlation in factor analysis, so it is not an extra step.

Some math associated with the **ONE FACTOR** model

- λ_j^2 is also called the "communality" of X_j in the one factor case (notation: h_j^2)

- **For standardized X_j , Corr(F, X_j) = λ_j**

- The percentage variability in (standardized) X_j explained by F is λ_j^2. (like an R^2)

- If X_j is N(0,1), then λ_j is equivalent to:
  - the slope in a regression of X_j on F
  - the correlation between F and X_j

- Interpretation of λ_j:
  - standardized regression coefficient (regression)
  - path coefficient (path analysis)
  - factor loading (factor analysis)

Some <u>more</u> math associated with the ONE factor model

- $\text{Corr}(X_j , X_k) = \lambda_j \lambda_k$
- Note that the correlation between X_j and X_k is <u>completely</u> determined by the common factor. Recall $\text{Cov}(e_j,e_k)=0$
- Factor loadings (λ_j) are equivalent to correlation between factors and variables when only a <u>**SINGLE**</u> common factor is involved.

Steps in Exploratory Factor Analysis

(1) Collect and explore data: choose relevant variables.

(2) Extract initial factors (via principal components)

(3) Choose number of factors to retain

(4) Choose estimation method, estimate model

(5) Rotate and interpret

(6) (a) Decide if changes need to be made (e.g. drop item(s), include item(s))

(b) repeat (4)-(5)

(7) Construct scales and use in further analysis

Data Exploration

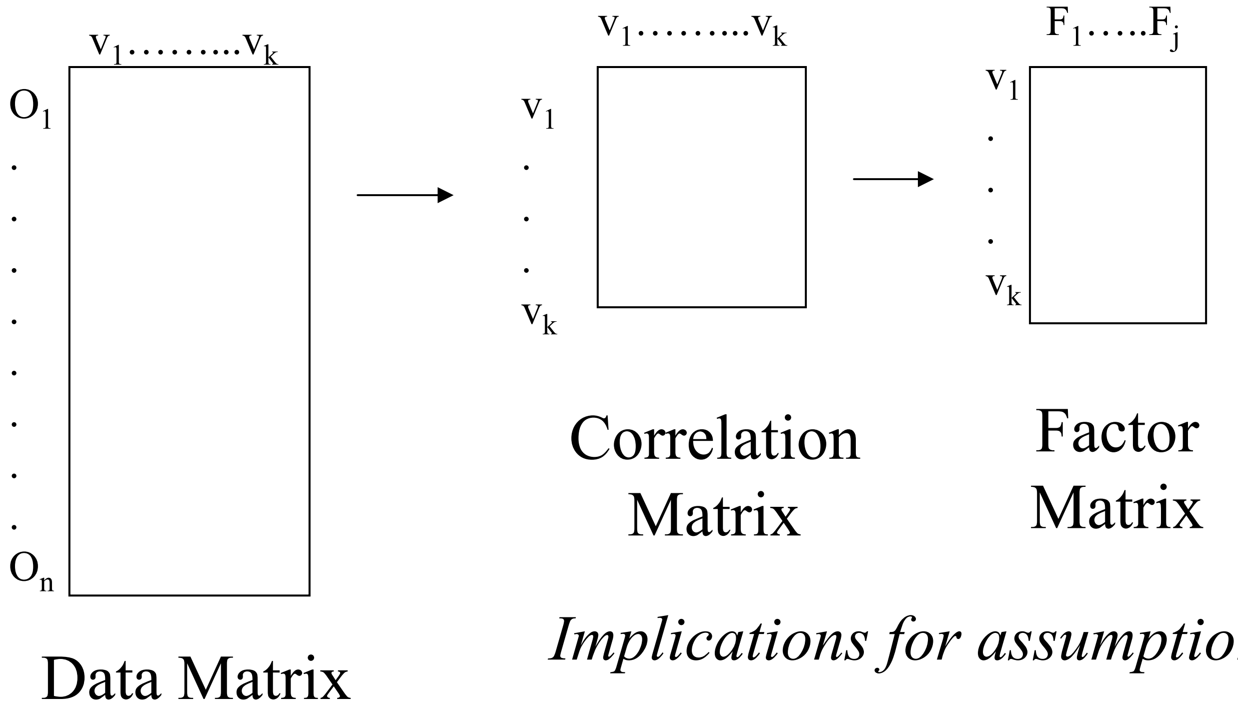
- Histograms
  - normality
  - discreteness
  - outliers
- Covariance and correlations between variables
  - very high or low correlations?
- Same scale
- high = good, low = bad?

Aside: Correlation vs. Covariance

- >90% of Factor Analyses use correlation matrix
- <10% use covariance matrix
- We will focus on correlation matrix because
  - It is less confusing than switching between the two
  - It is much more commonly used and more commonly applicable
- Covariance does have its place (we'll address that next time).

Data Matrix

- Factor analysis is **totally dependent** on correlations between variables.
- Factor analysis summarizes correlation structure



Implications for assumptions about X's?

Frailty Example
(N=571)

	arm	ski	fastw	grip	pincr	upex	knee	hipext	shldr	peg	bmi	usalk
skinfld	**0.71**											
fastwalk	-0.01	0.13										
gripstr	**0.34**	0.26	0.18									
pinchstr	**0.34**	**0.33**	0.16	**0.62**								
upextstr	0.12	0.14	0.26	**0.31**	0.25							
kneeext	0.16	**0.31**	**0.35**	0.28	0.28	0.21						
hipext	0.11	0.28	0.18	0.24	0.24	0.15	**0.56**					
shldrrot	0.03	0.11	0.25	0.18	0.19	**0.36**	**0.30**	0.17				
pegbrd	-0.10	-0.17	**-0.34**	-0.26	-0.13	-0.21	-0.15	-0.11	-0.15			
bmi	**0.88**	**0.64**	-0.09	0.25	0.28	0.08	0.13	0.13	0.01	-0.04		
uslwalk	-0.03	0.09	**0.89**	0.16	0.13	0.27	**0.30**	0.14	0.22	**-0.31**	-0.10	
chrstand	0.01	-0.09	-0.43	-0.12	-0.12	-0.22	-0.27	-0.15	-0.09	0.25	0.03	**-0.42**

One Factor Model

$$X_1 = \lambda_1 F + e_1$$
$$X_2 = \lambda_2 F + e_2$$
$$\ldots$$
$$X_m = \lambda_m F + e_m$$

One Factor Frailty Solution

```
 Variable |   Loadings
----------+-----------
 arm_circ |   0.28
  skinfld |   0.32
 fastwalk |   0.30
  gripstr |   0.32
 pinchstr |   0.31
 upextstr |   0.26
  kneeext |   0.33
   hipext |   0.26
 shldrrot |   0.21
   pegbrd |  -0.23
      bmi |   0.24
  uslwalk |   0.28
 chrstand |  -0.22
```

These numbers represent the correlations between the common factor, F, and the input variables.

Clearly, estimating **F** is **part** of the process

More than One Factor

- m factor orthogonal model
- ORTHOGONAL = INDEPENDENT
- Example: frailty has domains, including strength, flexibility, speed.
- m factors, n observed variables

$$X_1 = \lambda_{11}F_1 + \lambda_{12}F_2 + \ldots + \lambda_{1m}F_m + e_1$$
$$X_2 = \lambda_{21}F_1 + \lambda_{22}F_2 + \ldots + \lambda_{2m}F_m + e_2$$
$$\ldots\ldots\ldots$$
$$X_n = \lambda_{n1}F_1 + \lambda_{n2}F_2 + \ldots + \lambda_{nm}F_m + e_n$$

More than One Factor

- Matrix notation: $\mathbf{X}_{nx1} = \mathbf{\Lambda}_{nxm}\mathbf{F}_{mx1} + \mathbf{e}_{nx1}$
- Same general assumptions as one factor model.
  - $corr(F_s,x_j) = \lambda_{js}$
- Plus:
  - $corr(F_s,F_r) = 0$ for all $s \neq r$ (i.e. orthogonal)
  - this is **forced** independence
  - simplifies covariance structure
  - $corr(x_i,x_j) = \lambda_{i1}\lambda_{j1} + \lambda_{i2}\lambda_{j2} + \lambda_{i3}\lambda_{j3} + \ldots.$
- To see details of <u>dependent</u> factors, see Kim and Mueller.

Matrix notation:

$$\mathbf{X}_{nx1} = \mathbf{\Lambda}_{nxm}\mathbf{F}_{mx1} + \mathbf{e}_{nx1}$$

$$\begin{bmatrix} X_1 \\ \vdots \\ \vdots \\ X_n \end{bmatrix}_{nx1} = \begin{bmatrix} \lambda_{11} & \cdots & \cdots & \lambda_{1m} \\ \vdots & \ddots & & \vdots \\ \vdots & & \ddots & \vdots \\ \lambda_{n1} & \cdots & \cdots & \lambda_{nm} \end{bmatrix}_{nxm} \begin{bmatrix} F_1 \\ \vdots \\ F_m \end{bmatrix}_{mx1} + \begin{bmatrix} e_1 \\ \vdots \\ \vdots \\ e_n \end{bmatrix}_{nx1}$$

Factor Matrix

$$\begin{bmatrix} \lambda_{11} & \lambda_{12} & \cdots & \lambda_{1m} \\ \lambda_{21} & \lambda_{22} & \cdots & \lambda_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{n1} & \lambda_{n1} & \cdots & \lambda_{nm} \end{bmatrix}_{nxm}$$

- Columns represent derived factors
- Rows represent input variables
- Loadings represent degree to which each of the variables "correlates" with each of the factors
- Loadings range from -1 to 1
- Inspection of factor loadings reveals extent to which each of the variables contributes to the meaning of each of the factors.
- High loadings provide meaning and interpretation of factors (~ regression coefficients)

Frailty Example

Factors

Variable	size	speed	Hand strength	Leg strength	Uniqueness
arm_circ	**0.97**	-0.01	0.16	0.01	**0.02**
skinfld	**0.71**	0.10	0.09	0.26	0.40
fastwalk	-0.01	**0.94**	0.08	0.12	0.08
gripstr	0.19	0.10	**0.93**	0.10	0.07
pinchstr	0.26	0.09	**0.57**	0.19	0.54
upextstr	0.08	0.25	0.27	0.14	**0.82**
kneeext	0.13	0.26	0.16	**0.72**	0.35
hipext	0.09	0.09	0.14	**0.68**	0.48
shldrrot	0.01	0.22	0.14	0.26	0.85
pegbrd	-0.07	-0.33	-0.22	-0.06	0.83
bmi	**0.89**	-0.09	0.09	0.04	0.18
uslwalk	-0.03	**0.92**	0.07	0.07	0.12
chrstand	0.02	-0.43	-0.07	-0.18	0.77

Communalities

- The communality of X_j is the proportion of the variance of X_j explained by the *m* common factors:

$$Comm(X_j) = \sum_{i=1}^{m} \lambda_{ij}^2$$

- Recall one factor model: What was the interpretation of λ_j^2?

$$Comm(X_j) = \lambda_j^2$$

- In other words, it can be thought of as the sum of squared multiple-correlation coefficients between the X_j and the factors.

- ***Uniqueness(X_j) = 1 - Comm(X_j)***

Communality of X_j

- "Common" part of variance
  - covariance between X_j and the part of X_j due to the underlying factors
  - For standardized X_j:
    - 1 = communality + uniqueness
    - uniqueness = 1 – communality
  - Can think of uniqueness = var(e_j)

➔ If X_j is informative, communality is high
➔ If X_j is not informative, uniqueness is high

- Intuitively: variables with high communality share more in common with the rest of the variables.

Communalities

- *Unstandardized X’s:*
  - *Var(X) = Var(F) + Var(e)*
  - *Var(X) = Communality + Uniqueness*
  - *Communality ≈ Var(F)*
  - *Uniqueness ≈ Var(e)*

- *How can Var(X)=Var(F)= 1 when using standardized variables? That implies that Var(e)=0.*
  - *After Var(F) is derived, then F is ‘standardized’ to have variance of 1. Two step procedure.*
  - Actual variances are “irrelevant” when using correlations and/or standardized X’s.

How many factors?

- Intuitively: The number of uncorrelated constructs that are jointly measured by the X's.
- Only useful if number of factors is less than number of X's (recall "data reduction").
- Identifiability: Is there enough information in the data to estimate all of the parameters in the factor analysis? May be constrained to a certain number of factors.

Choosing Number of Factors

Use "principal components" to help decide

- type of factor analysis
- number of factors is equivalent to number of variables
- each factor is a weighted combination of the input variables:

$$\mathbf{F_1 = a_{11}X_1 + a_{12}X_2 + \ldots.}$$

- Recall: For a factor analysis, <u>generally</u>,

$$\mathbf{X_1 = a_{11}F_1 + a_{12}F_2 + \ldots}$$

Estimating Principal Components

- **The first PC is the linear combination with maximum variance**
- That is, it finds vector a_1 to maximize

$$Var(F_1) = Var(a_1^T X) = a_1^T Cov(X) a_1$$

- (Can use correlation instead, equation is more complicated looking)
- Constrained such that $\Sigma a_1^2 = 1$
- First PC: linear combination $a_1 X$ that maximizes $Var(a_1^T X)$ such that $\Sigma a_1^2 = 1$
- Second PC: linear combination $a_2 X$ that maximizes $Var(a_2^T X)$ such that $\Sigma a_2^2 = 1$ AND $Corr(a_1^T X, a_2^T X)=0$.
- And so on…..

Eigenvalues

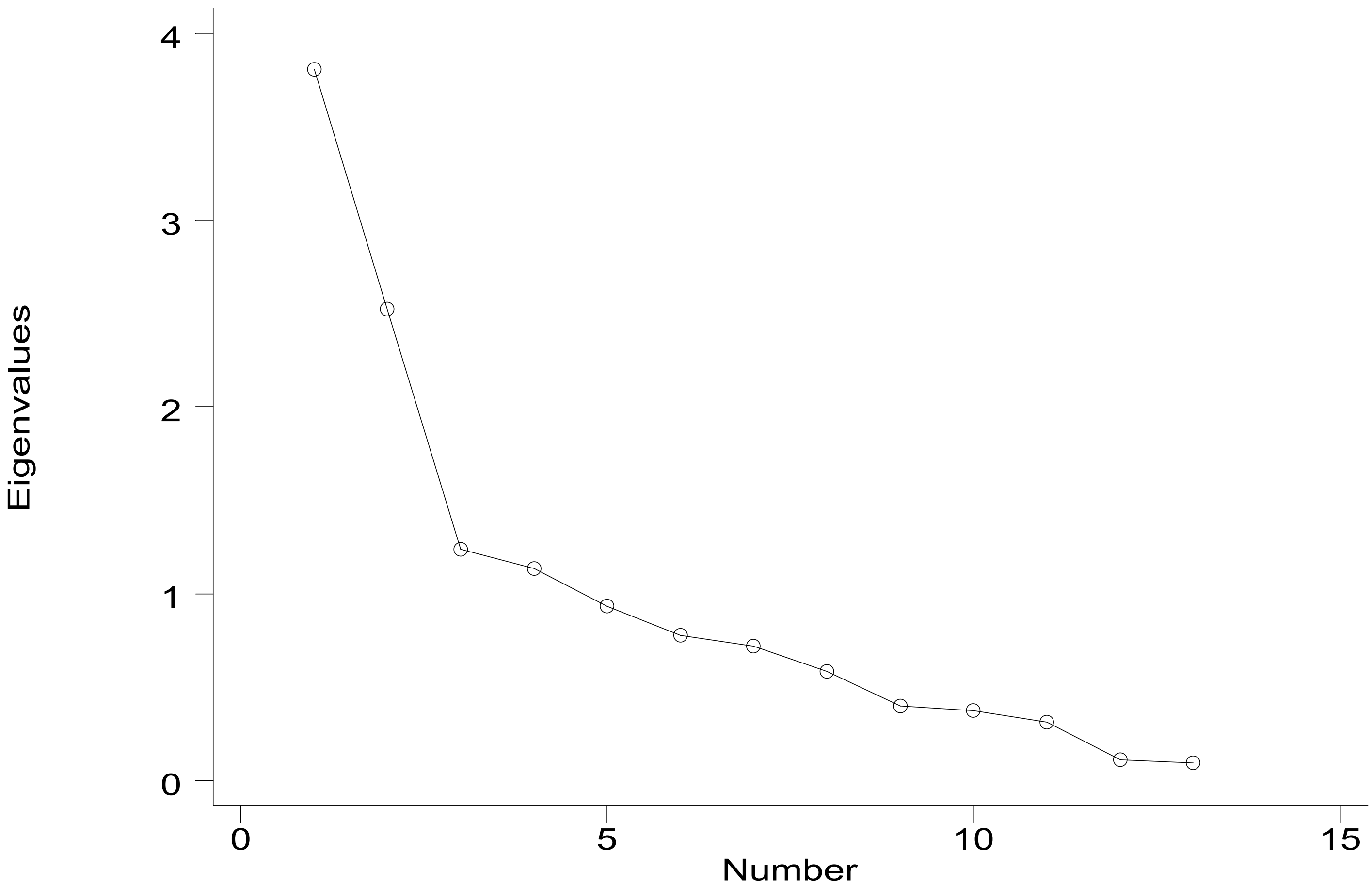
- To select how many factors to use, consider eigenvalues from a principal components analysis
- Two interpretations:
  - eigenvalue $\cong$ equivalent number of variables which the factor represents
  - eigenvalue $\cong$ amount of variance in the data described by the factor.
- Rules to go by:
  - number of eigenvalues > 1
  - scree plot
  - % variance explained
  - comprehensibility

Frailty Example

(principal components; 13 components retained)

Component	Eigenvalue	Difference	Proportion	Cumulative
1	3.80792	1.28489	0.2929	0.2929
2	2.52303	1.28633	0.1941	0.4870
3	1.23669	0.10300	0.0951	0.5821
4	1.13370	0.19964	0.0872	0.6693
5	0.93406	0.15572	0.0719	0.7412
6	0.77834	0.05959	0.0599	0.8011
7	0.71875	0.13765	0.0553	0.8563
8	0.58110	0.18244	0.0447	0.9010
9	0.39866	0.02716	0.0307	0.9317
10	0.37149	0.06131	0.0286	0.9603
11	0.31018	0.19962	0.0239	0.9841
12	0.11056	0.01504	0.0085	0.9927
13	0.09552	.	0.0073	1.0000

Scree Plot for Frailty Example



First 6 factors in principal components

Eigenvectors

Variable	1	2	3	4	5	6
arm_circ	0.28486	0.44788	-0.26770	-0.00884	0.11395	0.06012
skinfld	0.32495	0.31889	-0.20402	0.19147	0.13642	-0.03465
fastwalk	0.29734	-0.39078	-0.30053	0.05651	0.01173	0.26724
gripstr	0.32295	0.08761	0.24818	-0.37992	-0.41679	0.05057
pinchstr	0.31598	0.12799	0.27284	-0.29200	-0.38819	0.27536
upextstr	0.25737	-0.11702	0.17057	-0.38920	0.37099	-0.03115
kneeext	0.32585	-0.09121	0.30073	0.45229	0.00941	-0.02102
hipext	0.26007	-0.01740	0.39827	0.52709	-0.11473	-0.20850
shldrrot	0.21372	-0.14109	0.33434	-0.16968	0.65061	-0.01115
pegbrd	-0.22909	0.15047	0.22396	0.23034	0.11674	0.84094
bmi	0.24306	0.47156	-0.24395	0.04826	0.14009	0.02907
uslwalk	0.27617	-0.40093	-0.32341	0.02945	0.01188	0.29727
chrstand	-0.21713	0.27013	0.23698	-0.10748	0.19050	0.06312

At this stage….

- Don't worry about interpretation of factors!
- Main concern: whether a smaller number of factors can account for variability
- Researcher (i.e. YOU) needs to:
  - provide number of common factors to be extracted OR
  - provide objective criterion for choosing number of factors (e.g. scree plot, % variability, etc.)

Rotation

- In principal components, the first factor describes most of variability.
- After choosing number of factors to retain, we want to <u>spread variability</u> more evenly among factors.
- To do this we "rotate" factors:
  - redefine factors such that loadings on various factors tend to be very high (-1 or 1) or very low (0)
  - intuitively, it makes sharper distinctions in the meanings of the factors
  - **We use "factor analysis" for rotation NOT principal components!**

5 Factors, **Unrotated**

Factor Loadings

Variable	1	2	3	4	5	Uniqueness
arm_circ	**0.59934**	**0.67427**	-0.26580	-0.04146	0.02383	**0.11321**
skinfld	**0.62122**	**0.41768**	-0.13568	0.16493	0.01069	0.39391
fastwalk	**0.57983**	**-0.64697**	**-0.30834**	-0.00134	-0.05584	**0.14705**
gripstr	**0.57362**	0.08508	**0.31497**	**-0.33229**	-0.13918	0.43473
pinchstr	**0.55884**	0.13477	**0.30612**	-0.25698	-0.15520	0.48570
upextstr	**0.41860**	-0.15413	0.14411	-0.17610	0.26851	0.67714
kneeext	**0.56905**	-0.14977	0.26877	**0.36304**	-0.01108	0.44959
hipext	**0.44167**	-0.04549	**0.31590**	**0.37823**	-0.07072	0.55500
shldrrot	**0.34102**	-0.17981	0.19285	-0.02008	**0.31486**	0.71464
pegbrd	-**0.37068**	0.19063	0.04339	0.12546	-0.03857	0.80715
bmi	**0.51172**	**0.70802**	-0.24579	0.03593	0.04290	**0.17330**
uslwalk	**0.53682**	**-0.65795**	**-0.33565**	-0.03688	-0.05196	**0.16220**
chrstand	-**0.35387**	**0.33874**	0.07315	-0.03452	0.03548	0.75223

5 Factors, Rotated

(varimax rotation)

Rotated Factor Loadings

Variable	1	2	3	4	5	Uniqueness
arm_circ	-0.00702	**0.93063**	0.14300	0.00212	0.01487	**0.11321**
skinfld	0.11289	**0.71998**	0.09319	0.25655	0.02183	0.39391
fastwalk	**0.91214**	-0.01357	0.07068	0.11794	0.04312	**0.14705**
gripstr	0.13683	0.24745	**0.67895**	0.13331	0.08110	0.43473
pinchstr	0.09672	0.28091	**0.62678**	0.17672	0.04419	0.48570
upextstr	0.25803	0.08340	0.28257	0.10024	**0.39928**	0.67714
kneeext	0.27842	0.13825	0.16664	**0.64575**	0.09499	0.44959
hipext	0.11823	0.11857	0.15140	**0.62756**	0.01438	0.55500
shldrrot	0.20012	0.01241	0.16392	0.21342	**0.41562**	0.71464
pegbrd	**-0.35849**	-0.09024	-0.19444	-0.03842	-0.13004	0.80715
bmi	-0.09260	**0.90163**	0.06343	0.03358	0.00567	**0.17330**
uslwalk	**0.90977**	-0.03758	0.05757	0.06106	0.04081	**0.16220**
chrstand	**-0.46335**	0.01015	-0.08856	-0.15399	-0.03762	0.75223

2 Factors, Unrotated

	Factor Loadings		
Variable	1	2	Uniqueness
arm_circ	0.62007	0.66839	0.16876
skinfld	0.63571	0.40640	0.43071
fastwalk	0.56131	-0.64152	0.27339
gripstr	0.55227	0.06116	0.69126
pinchstr	0.54376	0.11056	0.69210
upextstr	0.41508	-0.16690	0.79985
kneeext	0.55123	-0.16068	0.67032
hipext	0.42076	-0.05615	0.81981
shldrrot	0.33427	-0.18772	0.85303
pegbrd	-0.37040	0.20234	0.82187
bmi	0.52567	0.69239	0.24427
uslwalk	0.51204	-0.63845	0.33020
chrstand	-0.35278	0.35290	0.75101

2 Factors, Rotated

(varimax rotation)

Variable	Rotated Factor Loadings 1	2	Uniqueness
arm_circ	-0.04259	**0.91073**	0.16876
skinfld	0.15533	**0.73835**	0.43071
fastwalk	**0.85101**	-0.04885	0.27339
gripstr	0.34324	**0.43695**	0.69126
pinchstr	0.30203	**0.46549**	0.69210
upextstr	**0.40988**	0.17929	0.79985
kneeext	**0.50082**	0.28081	0.67032
hipext	0.33483	0.26093	0.81981
shldrrot	**0.36813**	0.10703	0.85303
pegbrd	-0.40387	-0.12258	0.82187
bmi	-0.12585	**0.86017**	0.24427
uslwalk	**0.81431**	-0.08185	0.33020
chrstand	-**0.49897**	-0.00453	0.75101

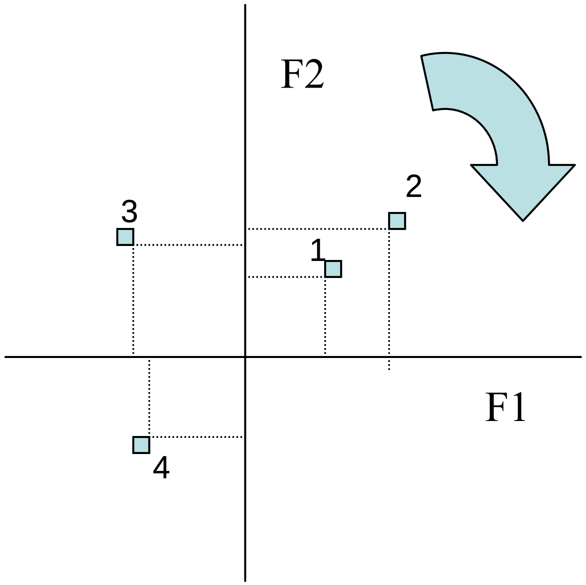
Unique Solution?

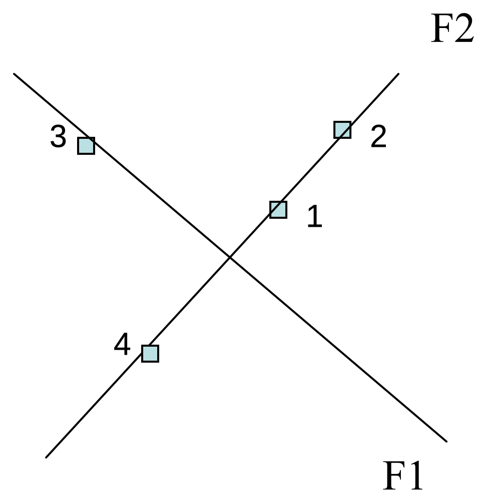
- The factor analysis solution is NOT unique!
- More than one solution will yield the same "result."
- We will understand this better by the end of the lecture…..

Rotation (continued)

- Uses “ambiguity” or non-uniqueness of solution to make interpretation more simple
- Where does ambiguity come in?
  - Unrotated solution is based on the idea that each factor tries to maximize variance explained, conditional on previous factors
  - What if we take that away?
  - Then, there is not one “best” solution.
- All solutions are <u>relatively</u> the same.
- Goal is simple structure
- Most construct validation assumes simple (typically rotated) structure.
- **Rotation does NOT improve fit!**

Rotating Factors (Intuitively)





	Factor 1	Factor 2
x1	0.5	0.5
x2	0.8	0.8
x3	-0.7	0.7
x4	-0.5	-0.5

	Factor 1	Factor 2
x1	0	0.6
x2	0	0.9
x3	-0.9	0
x4	0	-0.9

Orthogonal vs. Oblique Rotation

- Orthogonal: Factors are independent
  - **varimax**: maximize squared loading variance **across variables** (sum over factors)
  - quartimax: maximize squared loading variance **across factors** (sum over variables)
  - **Intuition: from previous picture, there is a right angle between axes**
- **Note: “Uniquenesses” remain the same!**

Orthogonal vs. Oblique Rotation

- Oblique: Factors **not** independent. Change in "angle."
  - oblimin: minimize squared loading covariance between factors.
  - **promax**: simplify orthogonal rotation by making small loadings even closer to zero.
  - Target matrix: choose "simple structure" a priori. (see Kim and Mueller)
  - **Intuition: from previous picture, angle between axes is not necessarily a right angle.**
- **Note: "Uniquenesses" remain the same!**

Promax Rotation: 5 Factors

(promax rotation)

Rotated Factor Loadings

Variable	1	2	3	4	5	Uniqueness
arm_circ	0.01528	**0.94103**	0.05905	-0.09177	-0.00256	0.11321
skinfld	0.06938	**0.69169**	-0.03647	0.22035	-0.00552	0.39391
fastwalk	**0.93445**	-0.00370	-0.02397	0.02170	-0.02240	0.14705
gripstr	-0.01683	0.00876	**0.74753**	-0.00365	0.01291	0.43473
pinchstr	-0.04492	0.04831	**0.69161**	0.06697	-0.03207	0.48570
upextstr	0.02421	0.02409	0.10835	-0.05299	0.50653	**0.67714**
kneeext	0.06454	-0.01491	0.00733	**0.67987**	0.06323	0.44959
hipext	-0.06597	-0.04487	0.04645	**0.69804**	-0.03602	**0.55500**
shldrrot	-0.06370	-0.03314	-0.05589	0.10885	**0.54427**	**0.71464**
pegbrd	-0.29465	-0.05360	-0.13357	0.06129	-0.13064	**0.80715**
bmi	-0.07198	**0.92642**	-0.03169	-0.02784	-0.00042	0.17330
uslwalk	**0.94920**	-0.01360	-0.02596	-0.04136	-0.02118	0.16220
chrstand	**-0.43302**	0.04150	-0.02964	-0.11109	-0.00024	**0.75223**

Promax Rotation: 2 Factors

(promax rotation)

Rotated Factor Loadings

Variable	1	2	Uniqueness
arm_circ	-0.21249	**0.96331**	0.16876
skinfld	0.02708	**0.74470**	0.43071
fastwalk	**0.90259**	-0.21386	0.27339
gripstr	0.27992	**0.39268**	0.69126
pinchstr	0.23139	**0.43048**	0.69210
upextstr	**0.39736**	0.10971	0.79985
kneeext	**0.47415**	0.19880	0.67032
hipext	0.30351	0.20967	0.81981
shldrrot	**0.36683**	0.04190	0.85303
pegbrd	**-0.40149**	-0.05138	0.82187
bmi	-0.29060	**0.92620**	0.24427
uslwalk	**0.87013**	-0.24147	0.33020
chrstand	**-0.52310**	0.09060	0.75101

.

Which to use?

- Choice is generally not critical
- Interpretation with orthogonal is “simple” because factors are independent: Loadings are correlations.
- Structure may appear more simple in oblique, but correlation of factors can be difficult to reconcile (deal with interactions, etc.)
- Theory? Are the conceptual meanings of the factors associated?
- <u>Oblique</u>:
  - Loading is no longer interpretable as covariance or correlation between object and factor
  - 2 matrices: pattern matrix (loadings) and structure matrix (correlations)
- Stata: varimax, promax

Steps in Exploratory Factor Analysis

(1) Collect data: choose relevant variables.

(2) Extract initial factors (via principal components)

(3) Choose number of factors to retain

(4) Choose estimation method, estimate model

(5) Rotate and interpret

(6) (a) Decide if changes need to be made (e.g. drop item(s), include item(s))

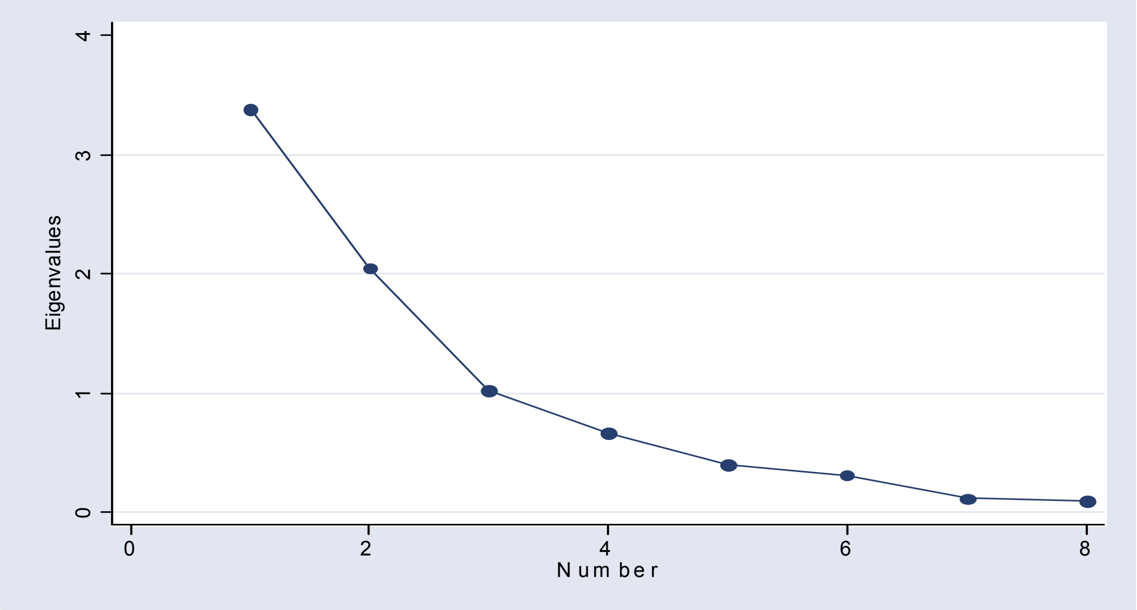
(b) repeat (4)-(5)

(7) Construct scales and use in further analysis

Drop variables with Uniqueness>0.50 in 5 factor model

```
. pca  arm_circ skinfld fastwalk gripstr pinchstr kneeext bmi uslwalk
(obs=782)

            (principal components; 8 components retained)
Component     Eigenvalue     Difference     Proportion     Cumulative
------------------------------------------------------------------
     1          3.37554        1.32772        0.4219           0.4219
     2          2.04782        1.03338        0.2560           0.6779
     3          1.01444        0.35212        0.1268           0.8047
     4          0.66232        0.26131        0.0828           0.8875
     5          0.40101        0.09655        0.0501           0.9376
     6          0.30446        0.19361        0.0381           0.9757
     7          0.11085        0.02726        0.0139           0.9896
     8          0.08358              .        0.0104           1.0000
```



3 Factor, Varimax Rotated

(varimax rotation)

Rotated Factor Loadings

Variable	weight	Leg agility.	hand str	Uniqueness
arm_circ	**0.93225**	0.00911	-0.19238	0.09381
skinfld	**0.84253**	0.17583	-0.17748	0.22773
fastwalk	0.01214	**0.95616**	-0.11423	0.07256
gripstr	0.19156	0.13194	**-0.86476**	0.19809
pinchstr	0.20674	0.13761	**-0.85214**	0.21218
kneeext	0.22656	**0.52045**	-0.36434	0.54505
bmi	**0.92530**	-0.07678	-0.11021	0.12579
uslwalk	-0.00155	**0.95111**	-0.09161	0.08700

2 Factor, Varimax Rotated

(varimax rotation)

Rotated Factor Loadings

Variable	weight	speed	Uniqueness
arm_circ	**0.94411**	0.01522	0.10843
skinfld	**0.76461**	0.16695	0.38751
fastwalk	0.01257	**0.94691**	0.10320
gripstr	0.43430	0.33299	0.70050
pinchstr	0.44095	0.33515	0.69324
kneeext	0.29158	0.45803	0.70519
bmi	**0.85920**	-0.07678	0.25589
uslwalk	-0.00163	**0.89829**	0.19308