Warsaw Summer School 2023, **OSU Study Abroad Program**

Factor analysis

Factor Analysis

- .
  What is factor analysis?
  What do we need factor analysis for?
  What are the modeling assumptions?
  How to specify, fit, and interpret factor models?
  What is the difference between exploratory and confirmatory factor analysis?
  How to assess model properties

What is factor analysis?

- **Factor analysis is a <u>theory-driven</u> statistical data reduction technique used to explain covariance among observed random variables (<u>indicators</u>) in terms of fewer <u>unobserved</u> variables (<u>factors)</u>**

- **<u>Theory-driven technique</u>: Relationship between factors and indicators must be justified.**
- **<u>Criteria for selecting indicators</u>: (1) face validity, (2) consistency**

Why factor analysis

1. Testing of theory

- **Explain covariation among multiple observed variables**
- **Mapping variables to latent constructs**

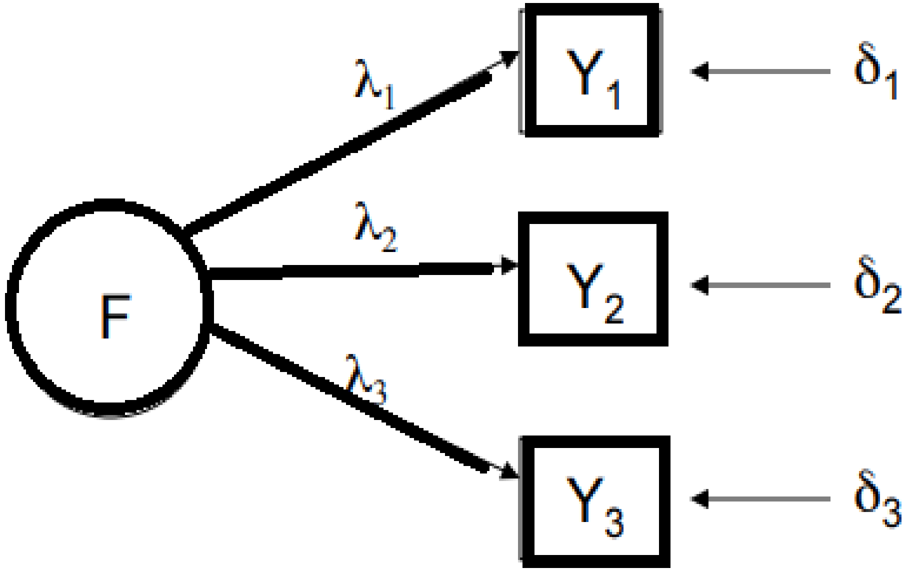
2. Understanding the structure underlying a set of measures

- **Gain insight into an unobserved variable**
- **Construct validation (convergent validity)**

3. Scale development

- **Factor scores**

Simple model

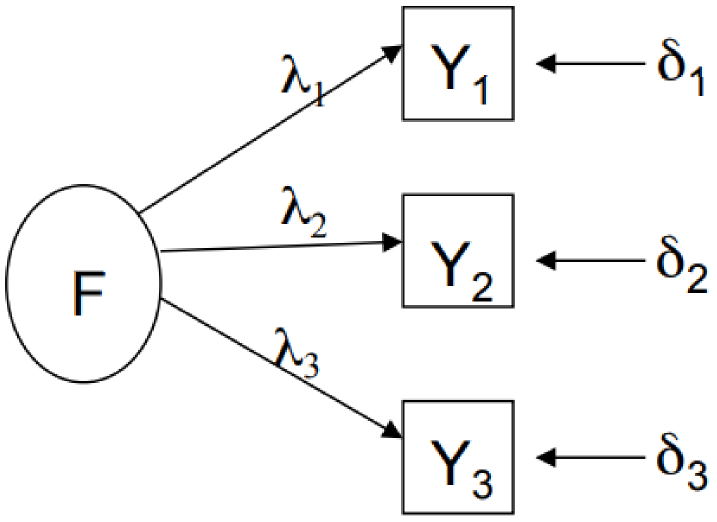


$$Y_1 = \lambda_1 F + \delta_1$$

$$Y_2 = \lambda_2 F + \delta_2$$

$$Y_3 = \lambda_3 F + \delta_3$$

Five properties



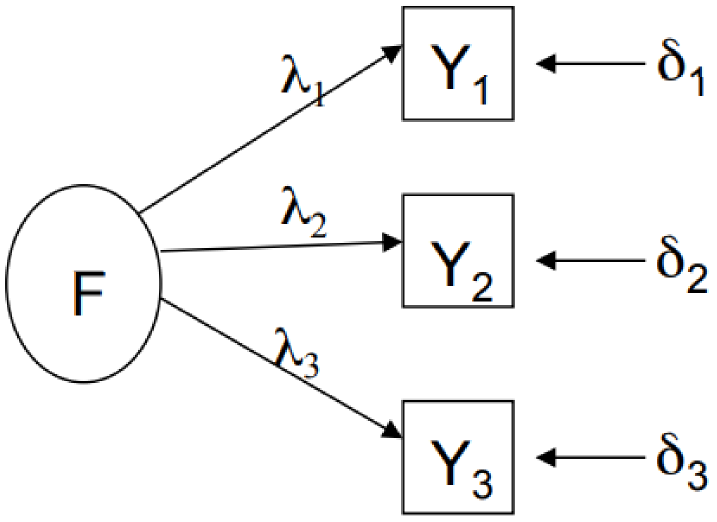
- **Factor F is not observed**
- **Y_1, Y_2, Y_3 are observed**
- **λ_i (lambda) = relations of Y_i and F**
- **δ_i (delta) = variability in the Y_i not explained by F**
- **Y_i is a linear function of F and δ_i**

$$Y_1 = \lambda_1 F + \delta_1$$

$$Y_2 = \lambda_2 F + \delta_2$$

$$Y_3 = \lambda_3 F + \delta_3$$

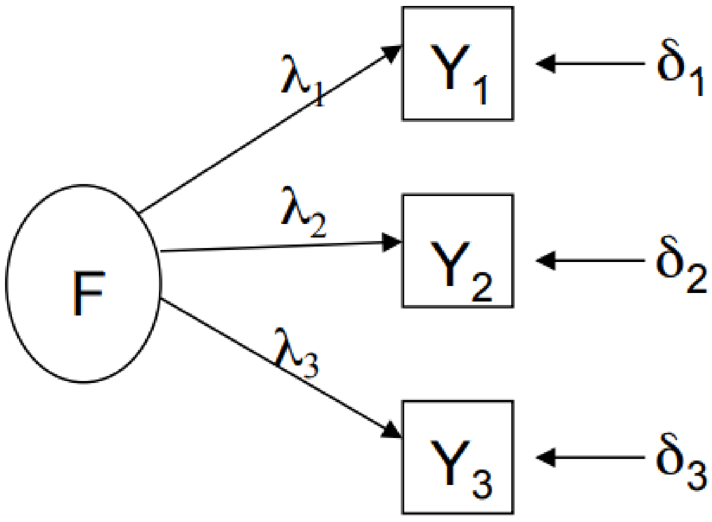
What is assumed and realized



$$Y_1 = \lambda_1 F + \delta_1$$
$$Y_2 = \lambda_2 F + \delta_2$$
$$Y_3 = \lambda_3 F + \delta_3$$

- **Factorial causation**
- **F is independent of δ_i , cov(F, δ_i) = 0**
- **δ_i and δ_j are independent for $i \neq j$, cov(δ_i ,δ_j) = 0**
- **Conditional independence: Given the factor, observed variables are independent of one another, cov(Y_i ,Y_j | F) = 0**

Terminology



$$Y_1 = \lambda_1 F + \delta_1$$
$$Y_2 = \lambda_2 F + \delta_2$$
$$Y_3 = \lambda_3 F + \delta_3$$

Factor loadings λ_i : λ_i = corr(Yi ,F)

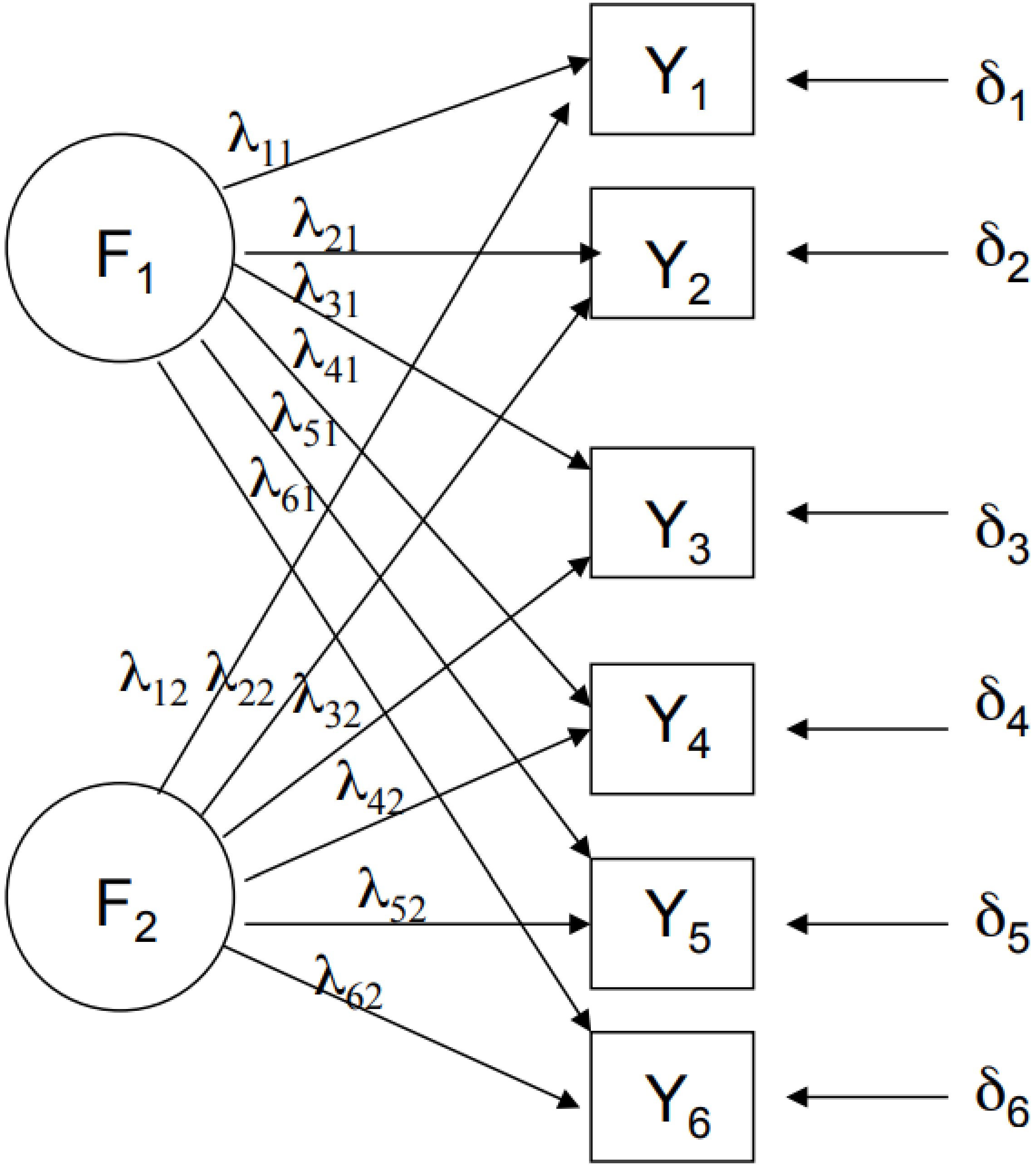
Communality of Y_i : $h_i^2 = \lambda_i^2 = [corr(Y_i, F)]^2$

=% variance of Y_i explained by F

Uniqueness of Y_i : $1-h_i^2$

= residual variance of Y_i

Degree of factorial determination: $=\Sigma \lambda i^2/n$, where n=numer of observed variables Y



$$Y_1 = \lambda_{11}F_1 + \lambda_{12}F_2 + \delta_1$$

$$Y_2 = \lambda_{21}F_1 + \lambda_{22}F_2 + \delta_2$$

$$Y_3 = \lambda_{31}F_1 + \lambda_{32}F_2 + \delta_3$$

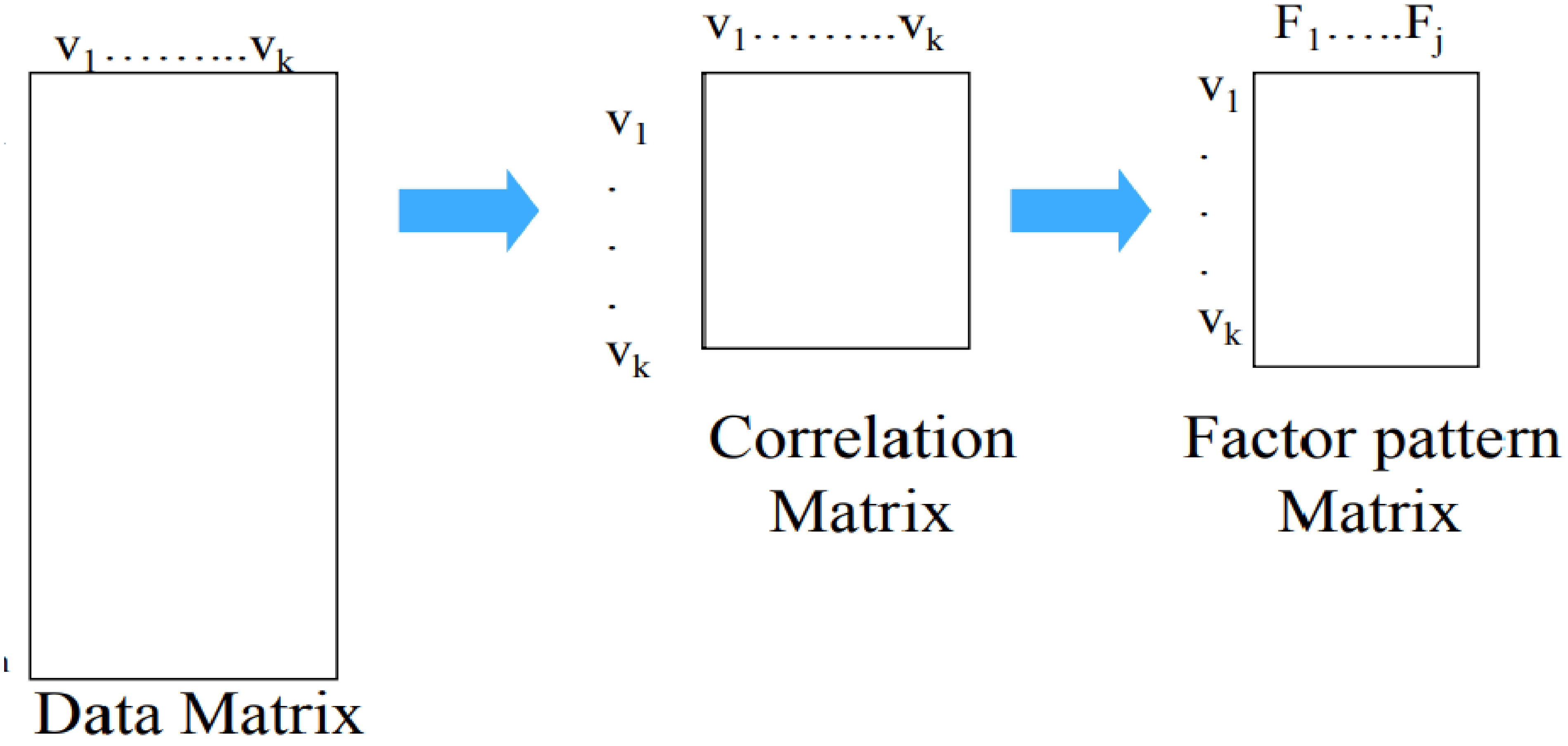
$$Y_4 = \lambda_{41}F_1 + \lambda_{42}F_2 + \delta_4$$

$$Y_5 = \lambda_{51}F_1 + \lambda_{52}F_2 + \delta_5$$

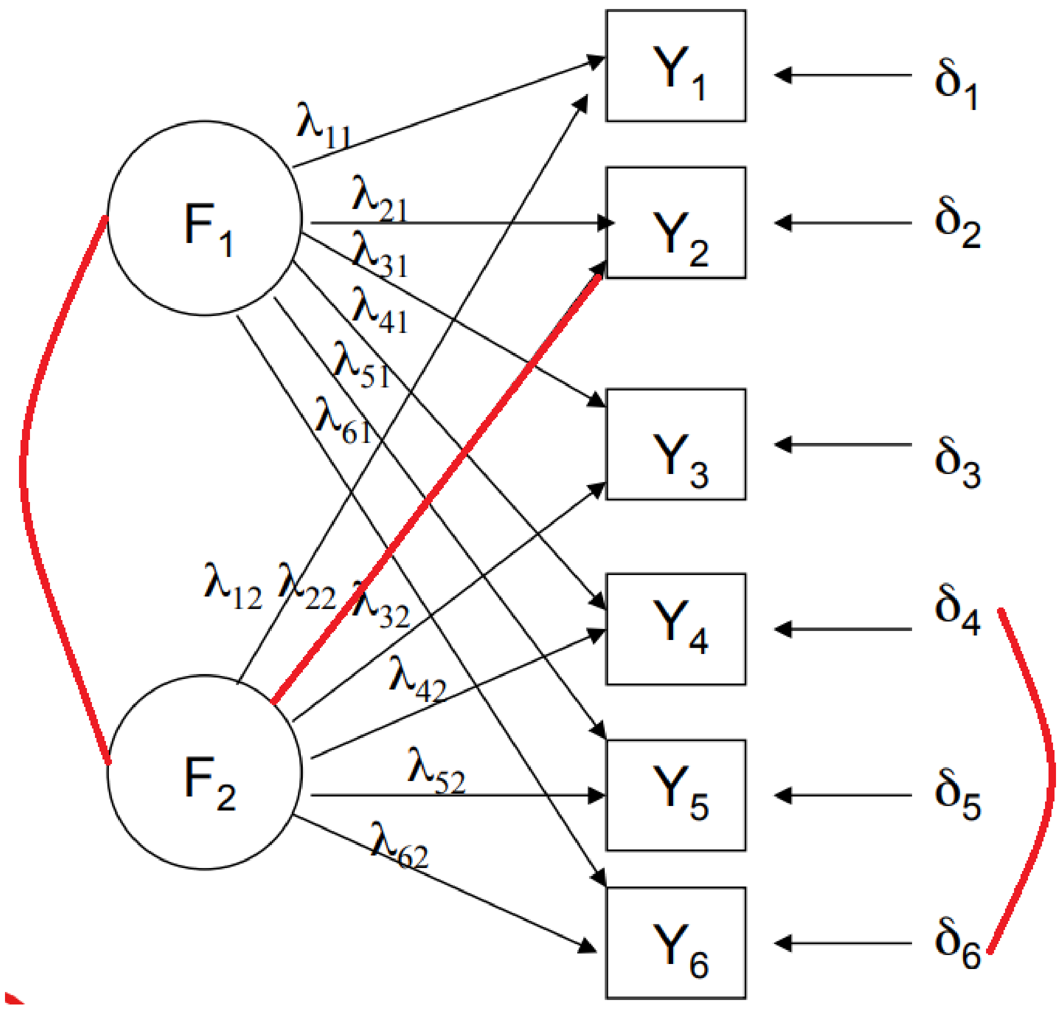
$$Y_6 = \lambda_{61}F_1 + \lambda_{62}F_2 + \delta_6$$

	Factor1	Factor2	Factor3
V1	0.97472	0.07264	0.04130
V2	0.91105	-0.03646	0.00203
V3	0.71305	0.18920	0.20310
V4	0.29647	0.20598	0.11790
V5	-0.06282	0.94342	-0.04330
V6	-0.09157	0.92812	-0.10349
V7	0.07547	0.38164	0.09303
V8	0.04973	-0.44975	-0.13030
V9	-0.42488	0.31303	0.64579
V10	-0.38193	0.15755	0.62447

General schema for factor analysis



Exploratory and confirmatory factor analysis



Methods of estimation

- **Least-squares method (LS)**

(For a small number of indicators)

- **Maximum likelihood method (ML)**

(More universal method, especially for CFA)

Least-squares method (LS)

- **Goal: minimize the sum of squared differences between observed and estimated correlation matrices**
- **Fitting steps:**

a) Obtain initial estimates of communalities (h^2)

b) Solve objective function: $\det(R-\eta I)=0$, where R is the correlation matrix with h^2 in the main diagonal, η is an eigenvalue

c) Re-estimate h^2

d) Repeat b) and c) until no improvement can be made

Factor scores

- Each person gets a factor score for each factor:

- The factors themselves are variables

- Factor score is weighted combination of values on input variables

F = WY where W is the weighted matrix

- These weights are NOT the factor loadings

- Different approaches exist for estimating (use regression method)

- Using factors scores instead of factor indicators can reduce measurement error

Rules

1) **At least one factor**
2) **at least three indicators per factor**
3) **non-correlated errors**