Factor Analysis

Dr. Juliho Castillo

Tecnológico de Monterrey

2025-09-29

Table of contents

Índice

Part I: Factor Analysis Theory

Factor Analysis Theory

Understanding Factor Analysis: The Detective Story

Imagine you're a detective investigating what makes students perform well in school.

You observe their grades in Math, Science, Literature, and History. But you suspect there are **hidden influences** behind these grades - things like «intelligence,» «motivation,» and «study habits» that you can't directly measure.

What Factor Analysis does:

- Looks for these **hidden factors** that explain why variables move together
- Separates what’s **common** (shared patterns) from what’s **unique** (individual noise)
- Tells you how much each hidden factor influences each observed variable

The key insight: If Math and Science grades are highly correlated, there might be a hidden «quantitative ability» factor. If all subjects correlate, there might be a «general intelligence» factor.

Why this matters:

- Helps understand the **underlying structure** of phenomena

- Validates **theoretical models** (like intelligence theory)
- Separates **signal from noise** in measurements

What is Factor Analysis?

- A statistical method for modeling relationships among **observed variables** using **latent factors**.
- It uses a smaller number of *unobserved variables*, known as **common factors**.
- **Key Distinction from PCA**: Explicitly models measurement error and unique variance
- Often used to discover and validate underlying theoretical constructs

Why Factor Analysis Works: Simple Logic

The Factor Model: Each variable = Common factors + Unique part

$$X_i = \lambda_{i1} F_1 + \lambda_{i2} F_2 + \ldots + U_i \tag{1}$$

What this means:

- Common factors (F) affect multiple variables
- Unique parts (U) affect only one variable
- Loadings (λ) tell us how much each factor affects each variable

Key Insight: If two variables share the same factors, they will be correlated!

Simple Example:

- Math and Science both depend on «Analytical Ability» $\rightarrow$ they correlate
- But each also has unique aspects (Math has geometry, Science has memorization)

The Magic Formula:

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi} \tag{2}$$

- $\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top$: correlations due to shared factors
- $\boldsymbol{\Psi}$: unique variance for each variable

Bottom Line: Observable correlations come from shared hidden factors!

Factor Analysis Model: Mathematical Foundation

The General Factor Model for Variable i:

$$X_i = \mu_i + \sum_{j=1}^{k} \lambda_{ij} F_j + U_i \quad \text{for } i = 1, 2, ..., p \tag{3}$$

Centered Factor Model (Standard Practice):

$$X_i - \mu_i = \sum_{j=1}^{k} \lambda_{ij} F_j + U_i \quad \text{for } i = 1, 2, ..., p \tag{4}$$

where:

- X_i: Observed variable i
- μ_i: Mean of variable i (estimated from data: $\mu_i = \overline{x}_i = \frac{1}{n}\sum_{l=1}^{n} X_{il}$)
- F_j: Common factor j with $F_j \sim N(0, 1)$, independent
- λ_{ij}: Factor loading of variable i on factor j
- U_i: Unique factor for variable i with $U_i \sim N(0, \psi_i^2)$
- $k < p$: Number of common factors

Profound Significance of This Model:

- **Theoretical Foundation**: This equation captures the belief that observed phenomena are driven by **underlying latent causes**
- **Parsimony Principle**: A few common factors ($k \ll p$) can explain most relationships among many variables
- **Measurement Theory**: Separates systematic variance (common factors) from random measurement error (unique factors)
- **Scientific Discovery**: Enables identification of unobservable constructs (like intelligence, personality dimensions)
- **Practical Relevance**: Reduces complex multivariate data to interpretable, meaningful dimensions for decision-making

Matrix Form (for centered data):

$$\boldsymbol{X}_c = \boldsymbol{\Lambda F} + \boldsymbol{U} \tag{5}$$

where:

- $\boldsymbol{X}_c = \boldsymbol{X} - \mathbf{1}_n \overline{\boldsymbol{x}}^\top \in \mathbb{R}^{n \times p}$ (centered data matrix)
- $\overline{\boldsymbol{x}} = \left(\overline{x}_1, \overline{x}_2, ..., \overline{x}_p\right)^\top$ (sample mean vector)
- $\boldsymbol{\Lambda} \in \mathbb{R}^{p \times k}$ (factor loadings matrix)
- $\boldsymbol{F} \in \mathbb{R}^{n \times k}$ (factor scores matrix)
- $\boldsymbol{U} \in \mathbb{R}^{n \times p}$ (unique factors matrix)

Why we center: Factor Analysis models the covariance structure, not the means

Variance Decomposition: Communality and Uniqueness

Variance of Observed Variable i (Scalar Form):

$$\mathrm{Var}(X_i) = \mathrm{Var}\left(\sum_{j=1}^{k} \lambda_{ij} F_j\right) + \mathrm{Var}(U_i) \tag{6}$$

Since F_j are independent with unit variance and U_i is independent of all F_j:

$$\mathrm{Var}(X_i) = \sum_{j=1}^{k} \lambda_{ij}^2 + \psi_i^2 = h_i^2 + \psi_i^2 \tag{7}$$

Matrix Form (Variance Structure):

$$\mathrm{Var}(\boldsymbol{X}) = \boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi} \tag{8}$$

where each diagonal element: $\sigma_{ii} = \sum_{j=1}^{k} \lambda_{ij}^2 + \psi_i^2$

Fundamental Insight Behind Variance Decomposition:

- **Total Variance = Systematic Variance + Random Variance**
- This decomposition is the **mathematical foundation of measurement theory**

- **Systematic variance** (h_i^2) represents what the variable shares with underlying constructs
- **Random variance** (ψ_i^2) captures measurement error and variable--specific effects
- **Critical for Quality Assessment**: High communality indicates reliable measurement; high uniqueness suggests measurement problems
- **Model Evaluation**: If most variance is unique, the factor model may be inappropriate for the data

Communality (Scalar and Matrix Forms):

- **Scalar:** $h_i^2 = \sum_{j=1}^{k} \lambda_{ij}^2$ (variance explained by common factors)
- **Matrix:** $\boldsymbol{h}^2 = \mathrm{diag}(\boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top)$ (communality vector)

Deeper Meaning of Communality:

- **Reliability Indicator**: High h_i^2 (close to 1) means variable is well--explained by common factors
- **Construct Validity**: Variables measuring the same construct should have similar communalities
- **Practical Threshold**: $h_i^2 > 0.5$ suggests variable belongs in the factor structure; $h_i^2 < 0.3$ may indicate poor fit

Uniqueness (Scalar and Matrix Forms):

- **Scalar:** $\psi_i^2 = \mathrm{Var}(U_i)$ (unique variance: specific + measurement error)
- **Matrix:** $\boldsymbol{\Psi} = \mathrm{diag}\left(\psi_1^2, \psi_2^2, ..., \psi_p^2\right)$ (uniqueness matrix)

Deeper Meaning of Uniqueness:

- **Measurement Quality**: High ψ_i^2 may indicate measurement problems or variable doesn't fit the factor structure
- **Specificity vs. Error**: Includes both legitimate specific variance and random measurement error
- **Model Diagnostics**: Variables with $\psi_i^2 > 0.7$ should be examined for potential removal or model revision

Total Variance Decomposition: For standardized variables: $\mathrm{Var}(X_i) = 1 = h_i^2 + \psi_i^2$

Critical Insight: This decomposition enables **quantitative assessment of measurement quality** and guides decisions about variable retention and model adequacy.

Algorithm: Communality and Uniqueness Calculation

Input: Factor loading matrix $\mathbf{\Lambda} \in \mathbb{R}^{p \times k}$

Output: Communalities $\boldsymbol{h}^2$, uniquenesses $\boldsymbol{\psi}^2$

1. **Initialize Arrays**
   - $\boldsymbol{h}^2 =$ zero vector of length p
   - $\boldsymbol{\psi}^2 =$ zero vector of length p
2. **Compute Communalities**

- **for** $i = 1$ to p **do**
    - $h_i^2 = \sum_{j=1}^{k} \lambda_{ij}^2$ (sum of squared loadings for variable i)

3. **Compute Uniquenesses**
    - **for** $i = 1$ to p **do**
        - $\psi_i^2 = 1 - h_i^2$ (for standardized variables)
        - **if** $\psi_i^2 < 0$ **then** $\psi_i^2 = 0.005$ (Heywood case correction)

4. **Validation Check**
    - **for** $i = 1$ to p **do**
        - **assert** $h_i^2 + \psi_i^2 = 1$ (variance decomposition property)
        - **assert** $0 \leq h_i^2 \leq 1$ and $0 \leq \psi_i^2 \leq 1$ (valid proportions)

Factor Extraction Methods: Mathematical Approaches

1. Principal Axis Factoring (PAF):

- Initialize communality estimates: $h_i^2 = R_{ii} - \frac{1}{R_{ii}^{-1}}$ (squared multiple correlation)
- Form reduced correlation matrix: $\boldsymbol{R}^* = \boldsymbol{R} - \mathrm{diag}\left(\psi_1^2, ..., \psi_p^2\right)$
- Eigendecomposition: $\boldsymbol{R}^* = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\top$
- Factor loadings: $\boldsymbol{\Lambda} = \boldsymbol{V}_k\sqrt{\boldsymbol{\Lambda}_k}$ (first k factors)
- Iterate until convergence: Update $h_i^2 = \sum_{j=1}^{k} \lambda_{ij}^2$

2. Maximum Likelihood (ML) Factoring:

- Assumes multivariate normality: $\boldsymbol{X} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$
- Covariance structure: $\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top + \boldsymbol{\Psi}$
- Minimize: $F_{\mathrm{ML}} = \mathrm{tr}\big(\boldsymbol{S}\boldsymbol{\Sigma}^{-1}\big) - \ln|\boldsymbol{\Sigma}^{-1}| - p$
- Provides χ^2 goodness-of-fit test and confidence intervals

Algorithm: Factor Analysis with Principal Axis Factoring

Input: Data matrix $\boldsymbol{X} \in \mathbb{R}^{n \times p}$ (n observations, p variables), number of factors k **Output:** Factor loadings $\boldsymbol{\Lambda}$, uniquenesses $\boldsymbol{\Psi}$, factor scores $\hat{\boldsymbol{F}}$

1. **Data Preprocessing and Suitability Assessment**
   - **Standardize variables:** $\boldsymbol{Z} = \left(\boldsymbol{X} - \mathbf{1}_n \overline{\boldsymbol{x}}^\top\right) \boldsymbol{D}^{-\frac{1}{2}}$ where $\boldsymbol{D} = \operatorname{diag}\left(s_1^2, ..., s_p^2\right)$ (variable variances)
   - **Compute correlation matrix:** $\boldsymbol{R} = \frac{1}{n-1} \boldsymbol{Z}^\top \boldsymbol{Z}$

- **Test data suitability for factoring:**
  - Bartlett's sphericity test: H_0: $\boldsymbol{R} = \boldsymbol{I}$ (variables uncorrelated)
  - Kaiser-Meyer-Olkin (KMO) measure: assess sampling adequacy

2. **Initialize Communality Estimates**
   - **Purpose:** Estimate how much variance each variable shares with factors
   - **for** each variable $i = 1$ to p:
     - **Compute initial estimate:** $h_i^2 = 1 - \frac{1}{(\boldsymbol{R}^{-1})_{ii}}$ (squared multiple correlation with other variables)
     - **Alternative:** Use maximum absolute correlation: $h_i^2 = \max_j |r_{ij}|$ for $j \neq i$

- **Interpretation:** Proportion of variable i's variance explained by common factors

3. **Principal Axis Factoring Iteration**
   - **Iterative refinement:** Improve communality estimates until convergence
   - **repeat**:
     - **Form reduced correlation matrix:** $\boldsymbol{R}^* = \boldsymbol{R} - \mathrm{diag}\left(1 - h_1^2, ..., 1 - h_p^2\right)$ (Replace diagonal with communalities instead of $1'$s)
     - **Eigenvalue decomposition:** $\boldsymbol{R}^* = \boldsymbol{V}\boldsymbol{\Lambda}\boldsymbol{V}^\top$
     - **Extract k factors:** $\boldsymbol{L} = \boldsymbol{V}_k\boldsymbol{\Lambda}_k^{\frac{1}{2}}$ (first k columns and eigenvalues)

- **Update communalities:** $h_i^2 = \sum_{j=1}^{k} l_{ij}^2$ for each $i = 1, ..., p$ (Sum of squared loadings for variable i)

- **until** $\max_i |h_i^{2_{\text{new}}} - h_i^{2_{\text{old}}}| < \varepsilon$ (convergence achieved)

4. **Factor Rotation** (Optional but Recommended)
   - **Purpose:** Achieve simpler, more interpretable factor structure
   - **Apply rotation method:**
     - Varimax (orthogonal): maintains factor independence, $\boldsymbol{\Lambda}^* = \boldsymbol{L}\boldsymbol{T}$
     - Promax (oblique): allows correlated factors for more flexibility
   - **Result:** Rotated loadings $\boldsymbol{\Lambda}^*$ with cleaner interpretation

5. **Factor Score Estimation**

- **Estimate individual factor scores for each observation**
- **Regression method:** $\hat{\boldsymbol{F}} = \boldsymbol{Z}\boldsymbol{\Lambda}\left(\boldsymbol{\Lambda}^\top\boldsymbol{\Lambda}\right)^{-1}$ (Simple, but scores may be correlated even with orthogonal factors)
- **Bartlett method:** $\hat{\boldsymbol{F}} = \boldsymbol{Z}\boldsymbol{\Lambda}\left(\boldsymbol{\Lambda}^\top\boldsymbol{\Psi}^{-1}\boldsymbol{\Lambda}\right)^{-1}\boldsymbol{\Lambda}^\top\boldsymbol{\Psi}^{-1}$ (More complex, but preserves factor orthogonality if assumed)

Factor Analysis: Simple Numerical Example

Given Data: 3 variables, 1 factor model

$$\boldsymbol{R} = \begin{pmatrix} 1.00 & 0.60 & 0.48 \\ 0.60 & 1.00 & 0.72 \\ 0.48 & 0.72 & 1.00 \end{pmatrix} \tag{9}$$

Step 1: Initial communality estimates (SMC)

- Compute $\boldsymbol{R}^{-1}$ and extract diagonal elements

- $h_1^2 = 1 - \frac{1}{(\boldsymbol{R}^{-1})_{11}} = 1 - \frac{1}{2.78} = 0.64$
- $h_2^2 = 1 - \frac{1}{(\boldsymbol{R}^{-1})_{22}} = 1 - \frac{1}{3.57} = 0.72$
- $h_3^2 = 1 - \frac{1}{(\boldsymbol{R}^{-1})_{33}} = 1 - \frac{1}{2.17} = 0.54$

Step 2: Form reduced correlation matrix

$$\boldsymbol{R}^* = \begin{pmatrix} 0.64 & 0.60 & 0.48 \\ 0.60 & 0.72 & 0.72 \\ 0.48 & 0.72 & 0.54 \end{pmatrix} \tag{10}$$

Step 3: Eigenvalue decomposition

- Largest eigenvalue: $\lambda_1 = 1.84$

- Corresponding eigenvector: $\boldsymbol{v}_1 = \begin{pmatrix} 0.53 \\ 0.67 \\ 0.52 \end{pmatrix}$

Step 4: Factor loadings

$$\boldsymbol{L} = \boldsymbol{v}_1 \sqrt{\lambda_1} = \begin{pmatrix} 0.53 \\ 0.67 \\ 0.52 \end{pmatrix} \times 1.36 = \begin{pmatrix} 0.72 \\ 0.91 \\ 0.71 \end{pmatrix} \tag{11}$$

Pointwise Form:

$$\lambda_{i1} = v_{i1} \sqrt{\lambda_1} \quad \text{for } i = 1, 2, 3 \tag{12}$$

Step 5: Updated communalities

$$h_1^2 = 0.72^2 = 0.52, \quad h_2^2 = 0.91^2 = 0.83, \quad h_3^2 = 0.71^2 = 0.50 \tag{13}$$

Final Model: $\boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^\top + \boldsymbol{\Psi}$

$$= \begin{pmatrix} 0.72 \\ 0.91 \\ 0.71 \end{pmatrix} \begin{pmatrix} 0.72 & 0.91 & 0.71 \end{pmatrix} + \begin{pmatrix} 0.48 & 0 & 0 \\ 0 & 0.17 & 0 \\ 0 & 0 & 0.50 \end{pmatrix} \tag{14}$$

Pointwise Form:

$$\sigma_{ij} = \lambda_{i1}\lambda_{j1} + \psi_i^2 \delta_{ij} \tag{15}$$

Python Implementation: Factor Analysis Example

```
import numpy as np
from factor_analyzer import FactorAnalyzer
from sklearn.datasets import make_spd_matrix

# Create correlation matrix (our example)
R = np.array([[1.00, 0.60, 0.48],
              [0.60, 1.00, 0.72],
              [0.48, 0.72, 1.00]])
```

```python
# For real data, you'd start with raw data:
# X = your_data  # shape (n_samples, n_features)
# R = np.corrcoef(X.T)  # correlation matrix

# Perform Factor Analysis
fa = FactorAnalyzer(n_factors=1, rotation=None)
fa.fit(R)  # For correlation matrix
# fa.fit(X)  # For raw data

# Get results
loadings = fa.loadings_
communalities = fa.get_communalities()
uniqueness = fa.get_uniquenesses()

print(f"Factor loadings:\n{loadings}")
```

```python
print(f"Communalities: {communalities}")
print(f"Uniqueness: {uniqueness}")
```

Note: Install with: pip install factor_analyzer

From Detective Work to Algorithm: Factor Analysis

Now let's turn our detective story into a systematic investigation!

Remember, Factor Analysis is like being a detective looking for hidden influences. The algorithm is our systematic method for uncovering these hidden factors.

What makes FA algorithm different from PCA?

- **Iterative process**: Like refining a theory through multiple rounds of evidence
- **Two unknowns**: We don’t know the factors OR how they influence variables
- **Modeling errors**: Accounts for measurement noise (like witness reliability)
- **Maximum Likelihood**: Finds the most probable explanation for what we observe

The detective’s toolkit:

1. **Start with a guess** (initial hypothesis)
2. **Estimate hidden factors** (what would the suspects be doing?)
3. **Update our theory** (revise how factors influence variables)

4. **Repeat until story makes sense** (convergence)
5. **Test the theory** (does it explain the data well?)

This is more complex than PCA, but gives us **richer insights** about underlying structure!

Algorithm: Maximum Likelihood Factor Analysis

Input: Data matrix $\boldsymbol{X} \in \mathbb{R}^{n \times p}$, number of factors k, tolerance ε

Output: ML factor loadings $\boldsymbol{\Lambda}$, uniquenesses $\boldsymbol{\Psi}$, model fit statistics

1. **Start with Initial Guess**
   - Make first estimate of factor loadings
   - Use simple method (like PCA) as starting point
2. **Iterative Improvement** (EM Algorithm)

- Step E: Estimate what the hidden factors would be
- Step M: Update loadings based on those estimates
- Repeat until no more improvement
- **Why this works:** Each step makes the model fit better

3. **Check How Well It Fits**
   - Does our model explain the correlations well?
   - Statistical tests tell us if the fit is good enough
   - **Good fit:** Model explains most observed relationships

4. **Get Final Results**
   - Factor loadings (how factors affect variables)
   - Confidence intervals (how certain are we?)
   - **Software does the complex calculations automatically**

Making Sense of the Results: Factor Rotation

You’ve found the hidden factors, but they’re still confusing! Now what?

Imagine you’ve identified two hidden factors in student performance, but:

- Factor 1 loads on Math (0.6), Science (0.7), Literature (0.5), History (0.4)

- Factor 2 loads on Math (0.4), Science (0.3), Literature (0.6), History (0.7)

This is messy! What do these factors actually represent?

Factor Rotation is like adjusting the camera angle after taking the photo:

- We want **simple structure**: each factor should clearly represent something
- Goal: High loadings for relevant variables, low for irrelevant ones
- Like rotating a map so north points up - same information, clearer presentation

Two types of rotation:

- **Orthogonal** (factors stay independent): Clean, simple interpretation
- **Oblique** (factors can correlate): More realistic, factors can be related

This is where art meets science in factor analysis!

Factor Rotation: Mathematical Transformation

Purpose: Transform initial factor loadings $\boldsymbol{\Lambda}$ to rotated loadings $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda T}$ where $\boldsymbol{T}$ is transformation matrix.

Profound Significance of Factor Loadings (λ_{ij}):

- **Conceptual Meaning**: Each loading represents the **strength of relationship** between observed variable i and latent factor j
- **Statistical Interpretation**: Correlation between variable and factor (in standardized form)

- **Practical Meaning**: How much a 1-unit change in the factor affects the observed variable
- **Measurement Perspective**: Quality indicator - high loadings suggest variable is a good indicator of the construct

Orthogonal Rotation (Varimax):

- Constraint: $\boldsymbol{T}^\top \boldsymbol{T} = \boldsymbol{I}$ (orthogonal transformation)
- Objective: Maximize variance of squared loadings within each factor
- Varimax criterion: $V = \frac{1}{p} \sum_{j=1}^{k} \left[\sum_{i=1}^{p} \left(\lambda_{ij}^* \right)^4 - \frac{1}{p} \left(\sum_{i=1}^{p} \left(\lambda_{ij}^* \right)^2 \right)^2 \right]$
- Seeks «simple structure»: high loadings for few variables, low for others

Why Rotation Matters: Initial loadings may be mathematically optimal but practically uninterpretable. Rotation preserves statistical properties while achieving **psychological/theoretical interpretability**.

Oblique Rotation (Promax):

- Allows factor correlations: $\boldsymbol{\Phi} = \mathrm{Corr}(\boldsymbol{F})$ may be non-diagonal
- Pattern matrix $\boldsymbol{P}$: direct effects (regression coefficients)
- Structure matrix $\boldsymbol{S} = \boldsymbol{P}\boldsymbol{\Phi}$: correlations with factors
- Relationship: $\boldsymbol{\Sigma} = \boldsymbol{P}\boldsymbol{\Phi}\boldsymbol{P}^\top + \boldsymbol{\Psi}^2$

Algorithm: Varimax Rotation for Simple Structure

Input: Initial factor loadings $\boldsymbol{\Lambda} \in \mathbb{R}^{p \times k}$, convergence tolerance ε

Output: Rotated loadings $\boldsymbol{\Lambda}^*$, rotation matrix $\boldsymbol{T}$

1. **Initialize Rotation Matrix**
   - $\boldsymbol{T} = \boldsymbol{I}_k$ (identity matrix)
   - $\boldsymbol{\Lambda}^* = \boldsymbol{\Lambda}$ (initial loadings)
2. **Rotate for Simple Structure**

- **Goal:** Make each factor load highly on few variables
- Algorithm finds the best rotation angle automatically
- **Process:**
  - Compare pairs of factors
  - Rotate them to maximize simple structure
  - Repeat until no more improvement

3. **Check the Results**
   - Verify factors are still independent (orthogonal)
   - Loadings should be clearer: high or low, not medium

Varimax Rotation: Simple Numerical Example

Given Factor Loadings: 3 variables, 2 factors (before rotation)

$$\mathbf{\Lambda} = \begin{pmatrix} 0.71 & 0.45 \\ 0.89 & 0.32 \\ 0.67 & -0.58 \end{pmatrix} \tag{16}$$

Before Rotation: Mixed loadings (hard to interpret)

- Variable 1: loads on both factors (0.71, 0.45)

- Variable 2: loads on both factors (0.89, 0.32)
- Variable 3: loads on both factors (0.67, −0.58)

After Rotation: Clearer structure

$$\mathbf{\Lambda}^* = \begin{pmatrix} 0.78 & 0.33 \\ 0.93 & 0.16 \\ 0.56 & -0.67 \end{pmatrix} \tag{17}$$

- Variable 1: mainly Factor 1 (0.78 vs 0.33)
- Variable 2: mainly Factor 1 (0.93 vs 0.16)
- Variable 3: mainly Factor 2 (0.56 vs −0.67)

Interpretation:

- Factor 1: Variables 1 & 2 (maybe «Verbal ability»)
- Factor 2: Variable 3 (maybe «Spatial ability»)

Software handles the complex rotation calculations automatically!

Python Implementation: Factor Rotation

```python
import numpy as np
from factor_analyzer import FactorAnalyzer

# Simulate data that would produce our example loadings
np.random.seed(42)
X = np.random.randn(100, 3)

# Apply Factor Analysis with rotation
fa_no_rotation = FactorAnalyzer(n_factors=2, rotation=None)
fa_varimax = FactorAnalyzer(n_factors=2, rotation='varimax')
```

```python
fa_no_rotation.fit(X)
fa_varimax.fit(X)

# Compare loadings before and after rotation
loadings_before = fa_no_rotation.loadings_
loadings_after = fa_varimax.loadings_

print("Before rotation:")
print(loadings_before)
print("\nAfter Varimax rotation:")
print(loadings_after)

# Other rotation options: 'promax', 'oblimin', 'quartimax'
fa_promax = FactorAnalyzer(n_factors=2, rotation='promax')
```

```python
fa_promax.fit(X)
print("\nAfter Promax rotation:")
print(fa_promax.loadings_)
```

Part II: Theoretical Comparison

Theoretical Comparison: PCA vs Factor Analysis

Conceptual Differences

Principal Component Analysis	**Factor Analysis**
Dimensionality reduction	Latent variable modeling
Components are linear combinations of all variables	Factors are hypothetical constructs
Explains total variance	Explains common variance only
No measurement error model	Explicitly models unique variance

Descriptive technique	Statistical model with assumptions

Mathematical Perspective

PCA Approach:

- Components are exact linear combinations: $\mathrm{PC}_j = \sum_{i=1}^{p} w_{ij} X_i$
- Maximizes variance explained: $\max \mathrm{Var}(\mathrm{PC}_j)$ subject to orthogonality
- All variance (including noise) is retained in the full solution

Factor Analysis Approach:

- Models observed variables: $X_i = \sum_{j=1}^{k} \lambda_{ij} F_j + U_i$
- Separates common factors from unique variance

- Estimates factor loadings and unique variances simultaneously

When Theory Guides Method Selection

Choose PCA when:

- Goal is data compression or visualization
- Want to capture maximum variance with minimal components
- No strong assumptions about underlying constructs

Choose Factor Analysis when:

- Testing specific theories about latent constructs
- Need to separate measurement error from true factors
- Interpretability of factors is primary concern

- Building models for prediction or explanation

Quick Decision Guide: PCA vs Factor Analysis

Use PCA when:

- You want to reduce dimensions for visualization
- You want to compress data (remove redundancy)
- You don’t have specific theory about hidden factors
- You want the mathematically optimal solution

Use Factor Analysis when:

- You want to test a theory (like «intelligence» causes test performance)
- You want to understand underlying causes
- You want to separate «signal» from «noise»
- You care more about interpretation than data compression

Quick Rules:

- Psychology/Education → Usually Factor Analysis
- Data compression/Machine learning → Usually PCA
- Exploratory analysis → Try both and compare!

Practical Checklist for Your Analysis

Before You Start:

1. Do variables measure similar things? (similar units?)
2. Do you have enough data? (at least 100 observations)
3. Are variables correlated? (if not, both methods will fail)

Choosing the Method:

4. Do you have a theory about hidden factors? → Factor Analysis
5. Just want to reduce dimensions? → PCA
6. Want to understand causes? → Factor Analysis

Interpreting Results:

7. PCA: How much variance explained by first few components?
8. FA: Do factor loadings make sense theoretically?
9. Rotation: Does it make interpretation clearer?

Red Flags:

- Eigenvalues all very similar → No clear structure
- Factors don’t make theoretical sense → Reconsider approach
- Too many factors needed → Maybe not suitable for these methods

Complete Python Workflow Example

```
import numpy as np
import pandas as pd
from sklearn.decomposition import PCA
from sklearn.preprocessing import StandardScaler
from factor_analyzer import FactorAnalyzer
from factor_analyzer.factor_analyzer import calculate_kmo,
calculate_bartlett_sphericity
import matplotlib.pyplot as plt

# 1. Load and prepare data
```

```python
# X = pd.read_csv('your_data.csv')  # Your actual data
X = np.random.randn(100, 5)  # Simulated data for demo

# 2. Standardize if needed
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)

# 3. Check suitability for factor analysis
kmo_all, kmo_model = calculate_kmo(X_scaled)
chi_square_value, p_value =
calculate_bartlett_sphericity(X_scaled)

print(f"KMO: {kmo_model:.3f} (>0.6 is good)")
print(f"Bartlett's test p-value: {p_value:.3f} (<0.05 is
good)")
```

```
# 4. Determine number of factors
pca = PCA()
pca.fit(X_scaled)
eigenvalues = pca.explained_variance_
n_factors = sum(eigenvalues > 1)  # Kaiser criterion

print(f"Suggested factors: {n_factors}")

# 5. Perform Factor Analysis with rotation
fa = FactorAnalyzer(n_factors=n_factors, rotation='varimax')
fa.fit(X_scaled)

# 6. Get and interpret results
loadings = fa.loadings_
```

```python
communalities = fa.get_communalities()
variance_explained = fa.get_factor_variance()

print(f"Factor loadings:\n{loadings}")
print(f"Communalities: {communalities}")
print(f"Variance explained: {variance_explained[1]}")  #
Proportional variance

# 7. Compare with PCA
pca_result = pca.transform(X_scaled)
fa_scores = fa.transform(X_scaled)

print("\nPCA vs FA comparison completed!")
```

This workflow covers the complete analysis pipeline!

Covariance Structure and Model Identification

Implied Covariance Matrix: $\mathbf{\Sigma} = \mathbf{\Lambda}\mathbf{\Lambda}^\top + \mathbf{\Psi}$

where $\mathbf{\Psi} = \mathrm{diag}\left(\psi_1^2, \psi_2^2, ..., \psi_p^2\right)$ is the uniqueness matrix.

Pointwise Form: $\sigma_{ij} = \sum_{l=1}^{k} \lambda_{il}\lambda_{jl} + \psi_i^2\delta_{ij}$ where $\delta_{ij} = 1$ if $i = j$, 0 otherwise.

Model Parameters:

- Factor loadings: $k \times p$ parameters in $\mathbf{\Lambda}$

- Unique variances: p parameters in $\mathbf{\Psi}$
- Total: $kp + p$ parameters to estimate

Identification Constraints:

- Sample covariance matrix has $\frac{p(p+1)}{2}$ unique elements
- For identification: $kp + p \leq \frac{p(p+1)}{2}$
- Simplifies to: $k \leq \frac{\frac{p(p-1)}{2} - p}{p} = \frac{p-1}{2} - 1$
- Additional constraint: Fix factor scale (unit variance) or loading scale

Degrees of Freedom: $\mathrm{df} = \frac{p(p+1)}{2} - (kp + p - k)$

Algorithm: Factor Analysis Suitability Tests

Input: Correlation matrix $\boldsymbol{R} \in \mathbb{R}^{p \times p}$, significance level α

Output: KMO measure, Bartlett test statistic, suitability decision

1. **Kaiser-Meyer-Olkin (KMO) Test**
   - Compute anti-image correlation matrix: $\boldsymbol{A} = -\text{diag}(\boldsymbol{R}^{-1})^{-1}\boldsymbol{R}^{-1}\text{diag}(\boldsymbol{R}^{-1})^{-1}$
   - **for** $i, j = 1$ to p **do**

- $a_{ij} = \frac{A_{ij}}{\sqrt{A_{ii}A_{jj}}}$ (off-diagonal anti-image correlations)

- Compute KMO: $\mathrm{KMO} = \frac{\sum_{i \neq j} R_{ij}^2}{\sum_{i \neq j} R_{ij}^2 + \sum_{i \neq j} a_{ij}^2}$

2. **Interpret KMO Value**
   - **if** KMO ≥ 0.9 **then** «marvelous»
   - **else if** KMO ≥ 0.8 **then** «meritorious»
   - **else if** KMO ≥ 0.7 **then** «middling»
   - **else if** KMO ≥ 0.6 **then** «mediocre»
   - **else if** KMO ≥ 0.5 **then** «miserable»
   - **else** «unacceptable for FA»

3. **Bartlett's Test of Sphericity**
   - $\chi^2 = -\left(n - 1 - \frac{2p+5}{6}\right) \ln(|\boldsymbol{R}|)$

- Degrees of freedom: $\mathrm{df} = \frac{p(p-1)}{2}$
- **if** $\chi^2 >$ critical value at α **then** reject sphericity (good for FA)

4. **Final Decision**
   - **if** KMO ≥ 0.6 **and** Bartlett significant **then** «proceed with FA»
   - **else** «reconsider data or variables»

FA vs PCA — Quick comparison

- PCA: descriptive, decomposes total variance via $S = V\Lambda V^T$, components are orthogonal
- FA: model-based, explains common variance; explicitly models uniqueness (U_i)
- When to use: PCA for compression/visualization; FA when modeling latent constructs & measurement error

Part III: Factor Analysis Examples

Example 1B: Educational Assessment
Factor Analysis

Educational Assessment: Dataset Overview

Research Question: What are the underlying cognitive abilities measured by standardized tests?

Dataset: 500 high school students, 8 standardized test scores

- Math Reasoning (MR), Math Computation (MC)
- Reading Comprehension (RC), Vocabulary (V)
- Spatial Visualization (SV), Pattern Recognition (PR)
- Writing Skills (WS), Logical Reasoning (LR)

Theoretical Expectation: Three latent abilities

- **Quantitative Ability**: Math tests should load highly
- **Verbal Ability**: Language tests should load highly
- **Spatial Ability**: Visual-spatial tests should load highly

Factor Analysis Focus:

- Model latent cognitive constructs
- Separate true ability from measurement error
- Test theoretical structure of intelligence

FA Model Setup and Suitability Tests

Data Suitability Assessment:

- KMO Measure: 0.84 (meritorious)
- Bartlett's Test: $\chi^2 = 1247.3$, $p < 0.001$ (sphericity rejected)
- All individual KMO values > 0.7 (adequate sampling)

Factor Analysis Model:

```
X_i = λ_i1 × F_1 + λ_i2 × F_2 + λ_i3 × F_3 + U_i
```

- 3 common factors (based on Kaiser criterion and theory)
- Principal Axis Factoring with Varimax rotation

- 8 unique factors (measurement error + specific abilities)

Why FA over PCA here:

- Testing specific theory about cognitive structure
- Need to model measurement error in test scores
- Want interpretable factors representing abilities
- Focus on common variance shared across tests

Factor Loadings Matrix (After Varimax Rotation)

Test	Factor 1	Factor 2	Factor 3	Communality
Math Reasoning	.82	.31	.15	.78
Math Computation	.89	.22	.08	.84
Reading Comprehension	.24	.91	.19	.92
Vocabulary	.19	.85	.23	.81

Spatial Visualization	.18	.21	.87	.83
Pattern Recognition	.25	.15	.79	.71
Writing Skills	.31	.73	.28	.71
Logical Reasoning	.64	.45	.33	.72

Factor Interpretation:

- **Factor 1**: Quantitative Ability (Math tests + Logic)
- **Factor 2**: Verbal Ability (Language tests)
- **Factor 3**: Spatial Ability (Visual-spatial tests)

Key FA Insights:

- High communalities (0.71-0.92) indicate good factor representation

- Clear simple structure after rotation
- Logical Reasoning loads on both Quantitative and Verbal (as expected)

Uniqueness and Model Diagnostics

Uniqueness Analysis:

Test	Uniqueness	Interpretation
Math Reasoning	.22	Low measurement error
Math Computation	.16	Very reliable

Reading Comprehension	.08	Excellent reliability
Vocabulary	.19	Good reliability
Spatial Visualization	.17	Good reliability
Pattern Recognition	.29	Some specific variance
Writing Skills	.29	Some measurement error
Logical Reasoning	.28	Complex construct

Model Fit Assessment:

- Total variance explained: 76.3% (common factors only)
- Average communality: 0.79 (excellent)
- All uniqueness values < 0.3 (acceptable measurement quality)
- Residual correlations mostly < 0.05 (good model fit)

FA vs PCA Comparison:

- FA explains less total variance (76% vs 85%) but focuses on reliable variance
- FA provides cleaner factor interpretation due to rotation
- FA separates measurement error, PCA includes it in components

Example 2B: European Stock Markets
Factor Analysis

Stock Markets: Dataset and Research Question

Research Question: What are the underlying economic factors driving European stock market movements?

Dataset: Daily returns (252 trading days) for 8 major indices

- FTSE 100 (UK), DAX (Germany), CAC 40 (France), IBEX 35 (Spain)
- AEX (Netherlands), BEL 20 (Belgium), ATX (Austria), PSI 20 (Portugal)

Economic Theory: Common factors might represent

- **Market-wide sentiment** (systematic risk affecting all markets)
- **Regional economic integration** (EU economic policies)
- **Currency effects** (Euro vs non-Euro countries)

Factor Analysis Approach:

- Model common economic forces across markets
- Separate systematic risk from country-specific effects
- Identify diversification opportunities (unique variances)
- Test economic theory about market integration

FA Results: Market Factor Structure

Model Specifications:

- Principal Axis Factoring, 2 factors retained (Kaiser + Scree test)
- Promax rotation (allows correlated factors - more realistic for economics)

Rotated Factor Pattern Matrix:

Market Index	Factor 1	Factor 2	Communality
FTSE 100 (UK)	.73	.21	.58

DAX (Germany)	.91	-.08	.84
CAC 40 (France)	.88	.15	.80
IBEX 35 (Spain)	.67	.52	.72
AEX (Netherlands)	.85	.22	.77
BEL 20 (Belgium)	.79	.31	.72
ATX (Austria)	.58	.48	.57
PSI 20 (Portugal)	.43	.71	.69

Factor Correlation: $r = 0.34$ (moderate positive correlation)

Economic Interpretation:

- **Factor 1**: Core EU Market Integration (Germany, France, Netherlands)
- **Factor 2**: Peripheral Market Dynamics (Spain, Portugal, Austria)

Economic Insights from FA Model

Unique Variance Analysis (Risk Diversification):

- UK (FTSE): 42% unique variance (Brexit effects, Sterling currency)
- Austria (ATX): 43% unique variance (smaller economy, local factors)
- Germany (DAX): 16% unique variance (highly integrated with EU)

Practical Investment Implications:

1. **Systematic Risk**: 65% average communality means most risk is systematic

2. **Diversification**: UK and Austria offer best diversification benefits
3. **Core vs Periphery**: Two-factor structure confirms economic theory
4. **Currency Effects**: UK's unique variance partly reflects non-Euro status

Model Validation:

- Factor scores correlate with EU economic indicators
- Crisis periods show increased factor loadings (contagion effect)
- Model explains 69% of total market variance
- Residuals show minimal autocorrelation (good model fit)

FA vs PCA for Finance:

- FA focuses on systematic risk (relevant for portfolio theory)

- PCA would mix systematic and idiosyncratic risks
- FA provides better economic interpretation of market structure

Example 3B: Kuiper Belt Objects
Factor Analysis

Kuiper Belt Objects: Astronomical Dataset

Research Question: What are the fundamental physical processes that shaped Kuiper Belt Object (KBO) characteristics?

Dataset: 347 Kuiper Belt Objects with 7 orbital/physical parameters

- Semi-major axis (a), Eccentricity (e), Inclination (i)
- Perihelion distance (q), Aphelion distance (Q)
- Absolute magnitude (H), Estimated diameter (D)

Astronomical Theory: Two main formation processes

- **Primordial Disk Structure**: Original solar nebula properties

- **Dynamical Evolution**: Gravitational perturbations over 4.5 billion years

Factor Analysis Goals:

- Identify latent physical processes from observable parameters
- Separate primordial signals from evolutionary effects
- Model measurement uncertainties in astronomical observations
- Test theoretical models of outer solar system formation

FA Model: Cosmic Factor Structure

Preprocessing and Suitability:

- Log-transformed skewed variables (a, D)
- Standardized all parameters (different units/scales)
- KMO = 0.73 (middling), Bartlett $p < 0.001$ (suitable for FA)

Factor Analysis Results (Varimax Rotation):

Orbital Parameter	Factor 1	Factor 2	Communality
Semi-major axis (log a)	.89	.12	.81

Aphelion distance (log Q)	.94	.08	.89
Perihelion distance (q)	.76	-.35	.70
Eccentricity (e)	.23	.87	.81
Inclination (i)	-.08	.82	.68
Absolute magnitude (H)	-.67	.31	.55
Diameter (log D)	.71	-.28	.58

Astrophysical Interpretation:

- **Factor 1**: Distance/Size Relationship (primordial disk structure)
- **Factor 2**: Dynamical Excitation (scattering events, orbital evolution)

Astronomical Insights and Model Implications

Factor 1 - Primordial Disk Structure (62% of common variance):

- Larger objects found at greater distances (mass segregation)
- Reflects original solar nebula density gradient
- Size-distance correlation preserved over 4.5 Gyr

Factor 2 - Dynamical Evolution (31% of common variance):

- High eccentricity and inclination cluster together
- Indicates gravitational scattering by giant planets

- Separates «cold» vs «hot» KBO populations

Unique Variance Analysis:

- Absolute magnitude (H): 45% unique (observational bias effects)
- Perihelion distance: 30% unique (specific orbital resonances)
- Most orbital elements well-explained by two factors

Scientific Validation:

- Factor structure matches theoretical predictions
- Cold Classical KBOs score low on Factor 2
- Scattered Disk Objects score high on both factors
- Results consistent with Nice Model of solar system evolution

Why FA over PCA for Astronomy:

- Models observational uncertainties explicitly
- Tests specific theoretical hypotheses about formation
- Separates physical processes from measurement noise
- Provides meaningful astrophysical factor interpretation

Example 4B: Hospital Health Outcomes Factor Analysis

Hospital Quality: Healthcare Dataset

Research Question: What are the underlying dimensions of hospital quality that affect patient outcomes?

Dataset: 285 hospitals with 9 quality indicators

- 30-day mortality rates (Heart Attack, Heart Failure, Pneumonia)
- 30-day readmission rates (same 3 conditions)
- Patient satisfaction scores (Communication, Responsiveness, Pain Management)

Healthcare Theory: Quality dimensions might include

- **Clinical Excellence**: Medical competency, protocols, outcomes
- **Patient Experience**: Communication, comfort, service quality
- **System Efficiency**: Care coordination, discharge planning

Factor Analysis Applications:

- Identify core quality constructs for hospital evaluation
- Separate clinical performance from patient satisfaction
- Model measurement error in quality metrics
- Guide quality improvement initiatives and resource allocation

FA Results: Healthcare Quality Factors

Model Configuration:

- Maximum Likelihood estimation (for significance tests)
- 2 factors retained (eigenvalue > 1, theoretical fit)
- Promax rotation (quality dimensions may correlate)

Factor Pattern Matrix:

Quality Indicator	Factor 1	Factor 2	Communality
Heart Attack Mortality	.78	-.12	.62

Heart Failure Mortality	.71	.08	.51
Pneumonia Mortality	.69	.15	.50
Heart Attack Readmission	.83	-.05	.69
Heart Failure Readmission	.77	.22	.64
Pneumonia Readmission	.65	.31	.52
Communication Score	-.21	.89	.84
Responsiveness Score	-.15	.82	.69
Pain Management Score	.08	.74	.55

Factor Correlation: $r = -0.41$ (better clinical outcomes associated with better patient experience)

Healthcare Quality Interpretation:

- **Factor 1**: Clinical Performance (mortality and readmission rates)
- **Factor 2**: Patient Experience (satisfaction and communication)

Healthcare Policy Implications

Clinical Performance Factor Analysis:

- High communalities (0.50-0.69) suggest reliable quality measurement
- Hospitals strong in one clinical area tend to excel in others
- Readmission rates more reliable indicators than mortality (higher loadings)

Patient Experience Insights:

- Communication most important ($\lambda = 0.89$) for patient satisfaction

- All satisfaction measures highly correlated (common service quality)
- 55-84% of satisfaction variance explained by common factor

Unique Variance Components:

- Heart Failure Mortality: 49% unique (condition-specific factors)
- Pain Management: 45% unique (specialty-specific protocols)
- Other indicators: 31-48% unique variance

Quality Improvement Strategies:

1. **Systemic Approach**: Factors are correlated - improve both simultaneously
2. **Clinical Focus**: Target readmission reduction (high factor loadings)

3. **Patient Experience**: Prioritize communication training
4. **Measurement**: Consider composite scores based on factor weights

FA Advantages for Healthcare:

- Models measurement error in quality metrics
- Identifies underlying quality constructs
- Supports evidence-based quality improvement
- Enables fair hospital comparisons accounting for measurement uncertainty

Summary: Factor Analysis Across Domains

Common FA Benefits Across All Examples:

1. **Theoretical Validation**: Tests specific theories about latent constructs
   - Education: Cognitive ability structure
   - Finance: Economic integration theory
   - Astronomy: Solar system formation models
   - Healthcare: Quality dimension theory

2. **Measurement Error Modeling**: Separates signal from noise
   - Explicit uniqueness estimates
   - Better reliability assessment
   - More accurate factor interpretation

3. **Interpretable Results**: Rotation achieves simple structure
   - Clear factor loadings patterns
   - Meaningful construct identification
   - Actionable insights for practitioners

When Factor Analysis Excels:

- Theory-driven research questions
- Need to model measurement error
- Focus on common variance and latent constructs

- Interpretation more important than data compression

Key Insight: Factor Analysis transforms correlation patterns into theoretically meaningful latent constructs, making it invaluable for scientific understanding and practical decision-making across diverse fields.