# MULTIVARIATE ANALYSIS LECTURE 17: FACTOR ANALYSIS

Zhaoxia Yu | Professor, Department of Statistics

2025-05-27

# REQUIRED PACKAGES

▶ Code

# OUTLINE

- Review of PCA

# MOTIVATING EXAMPLE

## A MOTIVATING EXAMPLE: EXAM SCORES

- Simulated data
- 60 students
- six subjects: Latin, English, History, Arithmetic, Algebra, Geometry

▶ Code

☰

## PAIRWISE CORRELATION

▶ Code

# PCA: VISUALIZE EIGENVALUES

▶ Code

```
   Comp.1    Comp.2    Comp.3    Comp.4    Comp.5    Comp.6 
0.6986579 0.9361760 0.9583366 0.9742750 0.9873617 1.0000000 
```

▶ Code

# PCA LOADINGS

▶ Code

```

Loadings:
           Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
Latin       0.394  0.456  0.176  0.149  0.756  0.107
English     0.426  0.360 -0.127 -0.187 -0.465  0.650
History     0.417  0.393 -0.139        -0.323 -0.740
Arithmetic  0.414 -0.368  0.806        -0.193       
Algebra     0.396 -0.437 -0.456  0.659              
Geometry    0.402 -0.427 -0.275 -0.708  0.265       

               Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6
SS loadings     1.000  1.000  1.000  1.000  1.000  1.000
Proportion Var  0.167  0.167  0.167  0.167  0.167  0.167
Cumulative Var  0.167  0.333  0.500  0.667  0.833  1.000
```

- 1st PC is about average
- 2nd PC is about difference between Latin/English/History and Arithmetic/Algebra/Geometry

## PCA

- 1st PC = 0.39Latin+0.43English+0.42History +0.41Arithmetic+0.40Algebra+0.40Geometry
- 2nd PC = 0.46Latin+0.36English+0.39History -0.37Arithmetic-0.44Algebra-0.43Geometry

▶ Code

```
           Comp.1      Comp.2
 [1,] -0.12510081  0.387885389
 [2,] -0.24313883  0.565843287
 [3,]  0.61706648 -1.169012826
 [4,]  1.57192571 -1.061544977
 [5,] -1.01259642  0.511535779
 [6,]  0.08159618  0.709633832
 [7,] -3.97028171  1.616681077
 [8,]  1.40252349 -1.718247894
 [9,] -0.87166984 -0.676249351
[10,] -2.83899249  1.873213808
[11,]  2.72771614 -0.850526753
[12,] -1.70338268 -1.375732667
[13,]  0.45223712  1.335051512
[14,] -0.73935793 -0.391321467
[15,]  0.36383804 -1.679191292
```

# VISUALIZE PCS (BIPLOT)

▶ Code

---

# INTRODUCTION TO FACTOR ANALYSIS

- Factor analysis (FA) is a statistical model used to identify underlying relationships between variables.
- It tries to understand/identify what (latent factors) leads to the observed correlations among a set of variables.

# PCA VS FA

▶ Code

---

https://livebook.manning.com/book/r-in-action-second-edition/chapter-14/6

## PCA VS FA

- Both reduce dimensionality
- Different goals:
  - PCA: aims to reduce the number of variables while preserving as much variance as possible
  - FA: aims to identify latent factors that explain the correlations among observed variables
- PCA: finds PCs (eg, PC1, PC2). Focuses on total variance.
- FA: find latent factors (eg, F1, F2). Focuses on correlations.

# FA MODEL

## FACTOR MODEL

- Consider a random vector $\mathbf{X} \in \mathbb{R}^p$
- Let $\boldsymbol{\mu} \in \mathbb{R}^p$ denote the population mean
- Let $F \in \mathbb{R}^m$ denote $m$ factors

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_p \end{pmatrix}, \boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_p \end{pmatrix}, \mathbf{F} = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_m \end{pmatrix}$$

# FACTOR MODEL

- $X_j$, which is observable, is assumed to be a linear function of the unobservable $f_1, \cdots, f_m$ plus specific errors.
  $$X_1 = \mu_1 + l_{11} f_1 + l_{12} f_2 + \cdots + l_{1m} f_m + \epsilon_1$$
  $$X_2 = \mu_2 + l_{21} f_1 + l_{22} f_2 + \cdots + l_{2m} f_m + \epsilon_2$$
  $$\vdots$$
  $$X_p = \mu_p + l_{p1} f_1 + l_{p2} f_2 + \cdots + l_{pm} f_m + \epsilon_p$$
  where $\epsilon_j$ is called the specific factor for feature $j$.
- The means $\mu_1, \cdots, \mu_p$ are parameters
- The coefficients in the factor loading matrix are also parameters

# FACTOR MODEL

- Let $\mathbf{L}$ denote the $p \times m$ matrix of factor loadings
$$\mathbf{L} = \begin{pmatrix} l_{11} & l_{12} & \dots & l_{1m} \\ l_{21} & l_{22} & \dots & l_{2m} \\ \vdots & \vdots & & \vdots \\ l_{p1} & l_{p2} & \dots & l_{pm} \end{pmatrix}$$
- A compact expression of the factor model is
$$\mathbf{X}_{p\times 1} = \boldsymbol{\mu}_{p\times 1} + \mathbf{L}_{p\times m}\mathbf{F}_{m\times 1} + \boldsymbol{\epsilon}_{p\times 1}$$

# EXAMPLE: LOADING

▶ Code

```
Loadings:
           Factor1 Factor2
Latin      0.937   0.195
English    0.905   0.321
History    0.920   0.280
Arithmetic 0.290   0.878
Algebra    0.207   0.923
Geometry   0.220   0.931

               Factor1 Factor2
SS loadings      2.718   2.710
Proportion Var   0.453   0.452
Cumulative Var   0.453   0.905
```

- How to interpret the loadings and the two factors?

## EXAMPLE: SCORES

▶ Code

---

- How to interpret the scores

# EXAMPLE: ORIGINAL SCORES (SORTED)

▶ Code

---

## MATHEMATICAL DETAILS: ASSUMPTIONS

- $\mathbf{F}$ and $\boldsymbol{\epsilon}$ are uncorrelated
- The $m$ common factors are uncorrelated
  $\mathbb{E}(\mathbf{F}) = 0, Cov(\mathbf{F}) = \mathbf{I}_m$
- The specific factors are uncorrelated
  $\mathbb{E}(\boldsymbol{\epsilon}) = 0, Cov(\boldsymbol{\epsilon}) = \boldsymbol{\Psi}$
  where $\Psi$ is a diagonal matrix with non-negative values, i.e.,

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

## THE COVARIANCE

- By the factor model and its assumptions, we have

$$\begin{aligned}\mathbf{\Sigma} &= cov(\mathbf{X}) \\ &= cov(\mathbf{LF} + \boldsymbol{\epsilon}) \\ &= \mathbf{L}Cov(\mathbf{F})\mathbf{L}^T + \mathbf{\Psi} \\ &= \mathbf{LL}^T + \mathbf{\Psi}\end{aligned}$$

The last step is due to our assumption that $cov(\mathbf{F}) = \mathbf{I}$

## THE COVARIANCE

- For $i \neq j$, the covariance between $X_i$ (feature $i$) and $X_j$ (feature $j$) is
  $$\sigma_{ij} = cov(X_i, X_j) = \sum_{k=1}^{m} l_{ik} l_{jk}$$
- The variance of $X_i$ is
  $$\sigma_{ii} = \sum_{k=1}^{m} l_{ik}^2 + \psi_i$$

# COMMUNALITY

## COMMUNALITY AND SPECIFIC VARIANCE

- From last slide
  $$\sigma_{ii} = \sum_{k=1}^{m} l_{ik}^2 + \psi_i$$
- We say that the variance of $X_i$ is partitioned into communality and specific variance where
  - communality is defined as $h_i^2 = \sum_{k=1}^{m} l_{ik}^2$, which is the proportion of variance contributed by common factors
  - specific variance $\psi_i$, which is the specific variance of $X_i$

# EXAMPLE OF COMMUNALITY

▶ Code

```
      Latin     English     History Arithmetic     Algebra    Geometry 
 0.08437518  0.07777364  0.07581204  0.14534743  0.10513511  0.08395685 
```

▶ Code

```
      Latin     English     History Arithmetic     Algebra    Geometry 
  0.9156248   0.9222264   0.9241880   0.8546526   0.8948649   0.9160432 
```

- For Geometry, the communality is ____ and the uniqueness (specific variance) is ____.

# NON-UNIQUENESS

## NON-UNIQUENESS OF FACTOR LOADINGS

- The factor loading coefficient is NOT unique.
- Suppose $\mathbf{X} = \boldsymbol{\mu} + \mathbf{LF} + \boldsymbol{\epsilon}$
- Consider any $m \times m$ orthogonal matrix $\Gamma$, which satisfies $\Gamma\Gamma^T = \Gamma^T\Gamma = \mathbf{I}$.
- Let $\tilde{\mathbf{L}} = \mathbf{L}\Gamma$ The model $\mathbf{X} = \boldsymbol{\mu} + \tilde{\mathbf{L}}\mathbf{F} + \boldsymbol{\epsilon}$

  give the same $\Sigma$ because $cov(\tilde{\mathbf{L}}\mathbf{F}) = \tilde{\mathbf{L}}cov(\mathbf{F})\tilde{\mathbf{L}}^T = \tilde{\mathbf{L}}\tilde{\mathbf{L}}^T = \mathbf{L}\Gamma\Gamma^T\mathbf{L}^T = \mathbf{L}\mathbf{L}^T = cov(\mathbf{LF})$

# NON-UNIQUENESS OF FACTOR LOADINGS

▶ Code

```
               Latin   English   History Arithmetic   Algebra
Geometry
Latin      1.0000002 0.9104285 0.9161363  0.4431684 0.3743267
0.3881987
English    0.9104285 1.0000000 0.9222470  0.5448653 0.4843854
0.4989042
History    0.9161363 0.9222470 1.0000001  0.5128914 0.4493253
0.4636887
Arithmetic 0.4431684 0.5448653 0.5128914  1.0000007 0.8702984
0.8814692
Algebra    0.3743267 0.4843854 0.4493253  0.8702984 0.9999997
0.9053334
Geometry   0.3881987 0.4989042 0.4636887  0.8814692 0.9053334
1.0000002
```

# NON-UNIQUENESS OF FACTOR LOADINGS

- Consider a rotation matrix $R$ and define $\tilde{\mathbf{L}} = LR$

▶ Code

```
                Latin   English   History Arithmetic   Algebra
Geometry
Latin      1.0000002 0.9104285 0.9161363  0.4431684 0.3743267
0.3881987
English    0.9104285 1.0000000 0.9222470  0.5448653 0.4843854
0.4989042
History    0.9161363 0.9222470 1.0000001  0.5128914 0.4493253
0.4636887
Arithmetic 0.4431684 0.5448653 0.5128914  1.0000007 0.8702984
0.8814692
Algebra    0.3743267 0.4843854 0.4493253  0.8702984 0.9999997
0.9053334
Geometry   0.3881987 0.4989042 0.4636887  0.8814692 0.9053334
1.0000002
```

# FACTOR ROTATION

# ROTATION FOR BETTER INTERPRETATION

- Interpretation of final results are easier for some choices of $\mathbf{L}$ than others.
- We often rotate the factors to gain insights or for better interpretation
- This is one advantage of factor analysis
- In practice,
  - Step 1: fit a factor model by imposing conditions that lead to a unique solution
  - Step 2: the loading matrix L is rotated (multiplied by an orthogonal matrix) in a way that gives a good interpretation of the data. Trial and error
- Well know criteria of rotation exist

## TWO MAJOR TYPES OF ROTATION

- An orthogonal rotation
  - maintains the perpendicularity between factors after rotation.
  - assumes that factors are unrelated or independent of each other.
  - Varimax is the most commonly used method of orthogonal rotation.

## TWO MAJOR TYPES OF ROTATION

- An oblique rotation
  - allows factors to be correlated and does not maintain a 90 degrees angle.
  - assumes that factors are related or dependent on each other.
  - One popular method is Promax
- Orthogonal rotations are mathematically appealing/convenient
- There is no reason that factors have to be uncorrelated

# VARIMAX AND PROMAX

- https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1745-3984.2006.00003.x
- Consider a rotation matrix with angle $\psi$

$$\begin{pmatrix} cos\psi & -sin\psi \\ sin\psi & cos\psi \end{pmatrix}$$

- The Varimax method looks for $\psi$ that maximizes the Varimax criterion

$$\frac{1}{p}\sum_i \left[\sum_j l_{ij}^4/h_i - (\sum_j (l_{ij}/h_i)^2)^2/p\right]$$

- The Promax is based on Varimax. It basically shrinks small loadings by using a powerful function

# NO ROTATION

▶ Code

```
Principal Components Analysis
Call: principal(r = exam, nfactors = 2, rotate = "none")
Standardized loadings (pattern matrix) based upon correlation
matrix
            PC1   PC2   h2    u2 com
Latin      0.81 -0.54 0.95 0.052 1.8
English    0.87 -0.43 0.95 0.054 1.5
History    0.85 -0.47 0.95 0.052 1.6
Arithmetic 0.85  0.44 0.91 0.090 1.5
Algebra    0.81  0.52 0.93 0.070 1.7
Geometry   0.82  0.51 0.94 0.064 1.7

                       PC1  PC2
SS loadings           4.19 1.43
Proportion Var        0.70 0.24
Cumulative Var        0.70 0.94
```

## NO ROTATION

▶ Code

# FACTOR ROTATION: VARIMAX (AN ORTHOGONAL ROTATION)

▶ Code

```
Principal Components Analysis
Call: principal(r = exam, nfactors = 2)
Standardized loadings (pattern matrix) based upon correlation
matrix
            RC1  RC2   h2    u2 com
Latin      0.96 0.17 0.95 0.052 1.1
English    0.93 0.30 0.95 0.054 1.2
History    0.94 0.25 0.95 0.052 1.1
Arithmetic 0.30 0.90 0.91 0.090 1.2
Algebra    0.22 0.94 0.93 0.070 1.1
Geometry   0.24 0.94 0.94 0.064 1.1

                       RC1  RC2
SS loadings           2.86 2.76
Proportion Var        0.48 0.46
Cumulative Var        0.48 0.94
```

# FACTOR ROTATION: VARIMAX (AN ORTHOGONAL ROTATION)

▶ Code

# FACTOR ROTATION: AN OBLIQUE (NON-ORTHOGONAL) ROTATION

▶ Code

```
Principal Components Analysis
Call: principal(r = exam, nfactors = 2, rotate = "promax")
Standardized loadings (pattern matrix) based upon correlation
matrix
             RC1   RC2   h2    u2 com
Latin       1.01 -0.08 0.95 0.052   1
English     0.93  0.07 0.95 0.054   1
History     0.96  0.02 0.95 0.052   1
Arithmetic  0.07  0.92 0.91 0.090   1
Algebra    -0.03  0.98 0.93 0.070   1
Geometry   -0.02  0.98 0.94 0.064   1

                       RC1  RC2
SS loadings           2.84 2.78
Proportion Var        0.47 0.46
Cumulative Var        0.47 0.94
```

# FACTOR ROTATION: AN OBLIQUE (NON-ORTHOGONAL) ROTATION

▶ Code

## FACTOR ROTATION

- The following articles provide nice descriptions of the two major types of rotations:

  https://scholarworks.umass.edu/cgi/viewcontent.cgi?article=1251&context=pare
  https://www.theanalysisfactor.com/rotations-factor-analysis/

# COMPUTATION

## METHOD 1: USE PCA

- By the spectral decomposition of $\Sigma$ we have
  $\Sigma = \Gamma\Lambda\Gamma^T$
  where $\Gamma = (\gamma_1, \cdots, \gamma_p)$ is an orthogonal matrix and
  $\Lambda = diag(\lambda_1, \cdots, \lambda_p)$ be the diagonal matrix of eigenvalues.
- The spectral decomposition can be rewritten to
  $$\boldsymbol{\Sigma} = \sum_{i=1}^{p} \lambda_i \gamma_i \gamma_i^T = \sum_{i=1}^{p} (\sqrt{\lambda_i}\gamma_i)(\sqrt{\lambda_i}\gamma_i)^T$$

# METHOD 1: USE PCA

- Suppose that $\lambda_m, \lambda_{m+1}, \cdots, \lambda_p$ are small. Then

$$\mathbf{\Sigma} \approx \sum_{i=1}^{m} (\sqrt{\lambda_i}\gamma_i)(\sqrt{\lambda_i}\gamma_i)^T$$

- Let $\mathbf{L} = \left(\sqrt{\lambda_1}\gamma_1, \cdots, \sqrt{\lambda_m}\gamma_m\right)$
- Let $\mathbf{\Psi} = \mathbf{\Sigma} - \mathbf{L}\mathbf{L}^T$

# R CODE

▶ Code

```
Principal Components Analysis
Call: principal(r = exam, nfactors = 2, rotate = "none")
Standardized loadings (pattern matrix) based upon correlation
matrix
            PC1   PC2   h2    u2 com
Latin      0.81 -0.54 0.95 0.052 1.8
English    0.87 -0.43 0.95 0.054 1.5
History    0.85 -0.47 0.95 0.052 1.6
Arithmetic 0.85  0.44 0.91 0.090 1.5
Algebra    0.81  0.52 0.93 0.070 1.7
Geometry   0.82  0.51 0.94 0.064 1.7

                       PC1  PC2
SS loadings           4.19 1.43
Proportion Var        0.70 0.24
Cumulative Var        0.70 0.94
```

▶ Code

```
Principal Components Analysis
Call: principal(r = exam, nfactors = 2)
Standardized loadings (pattern matrix) based upon correlation
matrix
            RC1  RC2   h2    u2 com
Latin      0.96 0.17 0.95 0.052 1.1
English    0.93 0.30 0.95 0.054 1.2
History    0.94 0.25 0.95 0.052 1.1
Arithmetic 0.30 0.90 0.91 0.090 1.2
Algebra    0.22 0.94 0.93 0.070 1.1
Geometry   0.24 0.94 0.94 0.064 1.1

                       RC1  RC2
SS loadings           2.86 2.76
Proportion Var        0.48 0.46
Cumulative Var        0.48 0.94
```

## METHOD 2: MLE

- We impose multivariate normality on the common and specific factors

  $\mathbf{F} \sim N(\mathbf{0}, \mathbf{I}), \epsilon \sim N(\mathbf{0}, \Psi)$

- The log-likelihood is

$$l(\mu, \mathbf{L}, \Psi) = -\frac{np}{2}\log 2\pi - \frac{n}{2}\log|\mathbf{L}\mathbf{L}^T + \Psi| - \frac{1}{2}\sum_{i=1}^{n}(\mathbf{X}_i - \mu)^T(\mathbf{L}\mathbf{L}^T + \Psi)^{-1}(\mathbf{X}_i - \mu)$$

where $\mathbf{X}_i \in \mathbb{R}^p$ denotes the $i$ observation (not the $i$th feature).

## THE NUMBER OF COMMON FACTORS

- The $p \times p$ covariance matrix $\Sigma$ is symmetric. As a result, there are $\frac{p(p+1)}{2}$ parameters.
- A factor mode imposes a structure on $\Sigma$
- For a FA model with $m$ common factors
- A FA model a small number common factors, i.e., when $m$ is small, the model uses fewer parameters
  - the model is more parsimonious
  - the model might not be adequate is $m$ is too small
- One can test whether an $m$ is large enough

☰

## CHOOSE $m$ OF FACTORS COMPUTED USING PCA

- Cumulatative variance explained is should be reasonably large, such as >$80\%$
- Look for elbow from the scree plot

# A GOODNESS OF FIT TEST FOR THE ADEQUACY OF THE NUMBER OF COMMON FACTORS

$$H_0: \quad \mathbf{\Sigma}_{(p\times p)} = \mathbf{L}_{(p\times m)}\,\mathbf{L}'_{(m\times p)} + \mathbf{\Psi}_{(p\times p)}$$

$$-2\ln\Lambda = -2\ln\left[\frac{\text{maximized likelihood under } H_0}{\text{maximized likelihood}}\right]$$

$$= -2\ln\left(\frac{|\hat{\mathbf{\Sigma}}|}{|\mathbf{S}_n|}\right)^{-n/2} + n\left[\operatorname{tr}(\hat{\mathbf{\Sigma}}^{-1}\mathbf{S}_n) - p\right]$$

$$\text{It can be shown that } \operatorname{tr}(\hat{\mathbf{\Sigma}}^{-1}\mathbf{S}_n) - p = 0$$

$$= n\ln\left(\frac{|\hat{\mathbf{\Sigma}}|}{|\mathbf{S}_n|}\right)$$

# TEST FOR THE ADEQUACY OF THE NUMBER OF COMMON FACTORS

- The number of parameters for covariance in the full model is $\frac{p(p+1)}{2}$
- The number of parameters for covariance in the reduced model is $mp + p - \frac{m(m-1)}{2}$

  Note: $-\frac{m(m-1)}{2}$ is due to the nonuniqueness of $\mathbf{L}$.

## DEGREES OF FREEDOM

- The difference in numbers of parameters between the two models is

$$\begin{aligned} df &= \frac{p(p+1)}{2} - [mp + p - \frac{m(m-1)}{2}] \\ &= \frac{1}{2}[(p-m)^2 - p - m] \end{aligned}$$

- Under the null hypothesis (adequate), the test statistics follows a chi-squared distribution when the sample size is large enough.

## TEST FOR THE ADEQUACY OF THE NUMBER OF COMMON FACTORS

- The result indicates that 1 factor is not adequate as the p-value is small.
- The p-value is about whether the correlation structure specified in the proposed model is significantly different from that of the full model

▶ Code

```
[1] "p-value for 1 factor"
```

▶ Code

```
   objective 
2.264759e-30 
```

# TEST FOR THE ADEQUACY OF THE NUMBER OF COMMON FACTORS

- The result indicates that 2 factors is adequate because the fit is not substantially from the full model.

▶ Code

```
[1] "p-value for 2 factors"
```

▶ Code

```
objective
 0.930055
```

# CFA VS EFA

# EXPLORATORY FACTOR ANALYSIS

- The FA approach we have discussed is exploratory in nature.
- In fact, we can perform EFA and identify latent factors by using only correlations, not the data
- The purpose of EFA is to explore the possible underlying structure that can explain the observed pattern of correlations
- EFA is used when researchers do not have a specific idea about the underlying structure of data
- EFA tries to identify the factor configuration (model)
- EFA is hypothesis-generating

## EXPLORATORY OR CONFIRMATORY FACTOR ANALYSIS

- Confirmatory Factory Analysis (CFA) is used when a researcher has specific hypotheses or theories about the factor structure of their data.
- It is a “theory-driven” approach.
- In CFA, the researcher specifies the number of factors and which variables load onto which factors.
- CFA is typically used in later stages of research to test or confirm the factor structure suggested by EFA
- CFA is hypothesis testing. A pre-specified model is required

## EXPLORATORY OR CONFIRMATORY FACTOR ANALYSIS

- Use EFA when:
  - You are unsure about the underlying structure.
  - You aim to uncover complex patterns.
  - You need to form hypotheses and develop theory.
- Use CFA when:
  - You have a predetermined theory or model.
  - You aim to test the hypothesis about the factor structure.
  - You need to confirm or disconfirm theories.

# CFA: EXAMPLE

## R PACKAGES FOR CFA

- Conduct CFA in R
- R packages:
  - sem
  - OpenMx
  - lavaan

## AN EXAMPLE USING "LAVAAN"

- CFA can be performed using the latent variable analysis ("lavaan") package in r

▶ Code

```
lavaan 0.6-19 ended normally after 51 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                        13

  Number of observations                            60

Model Test User Model:

  Test statistic                                 8.586
  Degrees of freedom                                 8
  P-value (Chi-square)                           0.378

Parameter Estimates:
```

# THE USER MODEL

▶ Code

## THE USER MODEL

- The model fits well. This suggests that the model does not significantly deviate from the observed data.

```
Model Test User Model:

  Test statistic                                         8.586
  Degrees of freedom                                         8
  P-value (Chi-square)                                   0.378
```

## UNDERSTAND THE BASELINE MODEL

▶ Code

## UNDERSTAND THE BASELINE MODEL

- The null model assumes no relationships between the variables. The output indicates that the baseline model is a poor fit, which is expected since it does not account for any relationships.

▶ Code

```
lavaan 0.6-19 ended normally after 8 iterations

  Estimator                                         ML
  Optimization method                           NLMINB
  Number of model parameters                         6

  Number of observations                            60

Model Test User Model:

  Test statistic                               462.072
  Degrees of freedom                                15
  P-value (Chi-square)                           0.000

Parameter Estimates:
```

## MODEL FIT STATISTICS

- We often need to compare a FA (reduced) model to the full model $H_0 : \mathbf{\Sigma}(\mathbf{L}, \Psi) = \mathbf{\Sigma}$
- Chi-square test. Derived from the likelihood ratio test. Depends on sample sizes.
- RMSEA: root mean square error of approximation compares the sample correlation matrix and the model correlation matrix. <0.06 is good.
  $RMSEA = \sqrt{\frac{\delta}{df(N-1)}}$
  where $\delta = \chi^2 - df$.

## MODEL FIT STATISTICS

- CFI: comparative fit index that measures the relative difference between two models; is not affected by sample size too much; between 0 and 1; the larger the better. >0.9 is good

$$CFI = \frac{\delta(\text{Baseline}) - \delta(\text{User})}{\delta(\text{Baseline})}$$

- More can be found in wikipedia and
  - https://doi.org/10.1016/B978-0-444-53737-9.50010-4
  - Bentler PM. Comparative fit indexes in structural models. Psychological bulletin. 1990 Mar;107(2):238.
  - http://www.davidakenny.net/cm/fit.htm

## SEM RESULTS: CFI AND TLI

```
CFI: 0.999
TLI: 0.998
```

- The Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) are both close to 1, indicating a good fit of the model to the data. For the two-factor model:
  - acceptable: >0.9
  - excellent: >0.95

# CFI AND TLI FOR USER MODEL AND BASELINE MODEL

▶ Code

```
  chisq     cfi   rmsea 
462.072   0.000   0.705 
```

▶ Code

```
chisq   cfi rmsea 
8.586 0.999 0.035 
```

# COMPUTE FACTOR SCORES

▶ Code

## HELPFUL RESOURCES

- https://quantdev.ssri.psu.edu/tutorials/intro-basic-confirmatory-factor-analysis
- https://lavaan.ugent.be/tutorial/tutorial.pdf
- https://stats.oarc.ucla.edu/r/seminars/rcfa/

▶ Code