Confirmatory Factor Analysis

Latent Trait Measurement and
Structural Equation Models
Lecture #6

February 13, 2013

Today's Class

- An introduction to confirmatory factor analysis
  - The next piece of the structural equation modeling puzzle
- CFA model:
  - Specification
  - Identification
  - Model fit assessment
  - Interpretation
- Today's data example will use a single factor to familiarize us with the core concepts of CFA
  - Concepts overlap with multiple factors

The Growth of Gambling Access

- In past 25 years:
  - An exponential increase in the accessibility of gambling
  - An increased rate of with problem or pathological gambling (Volberg, 2002, Welte et al., 2009)

- Hence, there is a need to better:
  - Understand the underlying causes of the disorder
  - Reliably identify potential pathological gamblers
  - Provide effective treatment interventions

Pathological Gambling: DSM Definition

- To be diagnosed as a pathological gambler, an individual must meet 5 of 10 defined criteria:

1. Is preoccupied with gambling
2. Needs to gamble with increasing amounts of money in order to achieve the desired excitement
3. Has repeated unsuccessful efforts to control, cut back, or stop gambling
4. Is restless or irritable when attempting to cut down or stop gambling
5. Gambles as a way of escaping from problems or relieving a dysphoric mood
6. After losing money gambling, often returns another day to get even
7. Lies to family members, therapist, or others to conceal the extent of involvement with gambling
8. Has committed illegal acts such as forgery, fraud, theft, or embezzlement to finance gambling
9. Has jeopardized or lost a significant relationship, job, educational, or career opportunity because of gambling
10. Relies on others to provide money to relieve a desperate financial situation caused by gambling

Methods for Assessment of Pathological Gambling

- The most frequently used instrument for assessing pathological gambling is the South Oaks Gambling Screen (Lesieur & Blume, 1987)
  - If you call Gamblers Anonymous, you will be given this screen

- 13 dichotomous (yes/no) items
  - Score of 5 or more indicates a probable pathological gambler

- Example items:
  - Do you feel you have a problem with betting money or gambling?
  - Do you ever gamble more than you intend to?

- Problems with the SOGS in gambling research:
  - Very low variability in item responses (many respond "no")
  - No mapping of SOGS onto DSM criteria

Research on Pathological Gambling

- In order to study the etiology of pathological gambling, more variability in responses was needed

- The Gambling Research Instrument (Feasel, Henson, & Jones, 2002) was created with 41 Likert-type items
  - Items were developed to measure each criterion

- Example items (ratings: Strongly Disagree to Strongly Agree):
  - I worry that I am spending too much money on gambling (C3)
  - There are few things I would rather do than gamble (C1)

- The instrument was used on a sample of experienced gamblers from a riverboat casino in a Flat Midwestern State
  - Casino patrons were solicited after playing roulette

Data For Today's Class

- To demonstrate factor analysis concepts, we will use the first five items of the Gambling Research Instrument:
   1. I would like to cut back on my gambling.
   2. There are few things I would rather do than gamble.
   3. If I lost a lot of money gambling one day, I would be more likely to want to play again the following day.
   4. I enjoy talking with my family and friends about my past gambling experiences.
   5. I find it necessary to gamble with larger amounts of money (than when I first gambled) for gambling to be exciting.

The GRI Items

- The GRI used a 6-point Likert scale
  - 1: Strongly Disagree
  - 2: Disagree
  - 3: Slightly Disagree
  - 4: Slightly Agree
  - 5: Agree
  - 6: Strongly Agree

- To meet the assumptions of factor analysis, we will treat these responses as being continuous
  - This is tenuous at best, but often is the case in factor analysis
  - Later we will discuss how to treat these as categorical items
    - Hint: Item Response Models

The Sample

- Data were collected from two sources:
  - 112 “experienced” gamblers
    - Many from an actual casino
  - 1192 college students from a “rectangular” midwestern state
    - Many never gambled before

- Today, we will combine both samples and treat them as homogenous – one sample of 1304 subjects
  - Later we will test this assumption – measurement invariance (called differential item functioning in item response theory literature)

CONCEPTUAL OVERVIEW OF FACTOR ANALYSIS: LATENT TRAITS

What is a *Latent Trait*?

- Latent trait: An unobservable ability or characteristic
  - e.g., “intelligence”, “extroversion”, or “political idealization”

- A person’s latent trait(s) are estimated (measured) using a ***measurement model***
  - **Measurement model:** A statistical model linking the unobserved latent trait with the observed outcome
    - In social/education research outcomes are generally test items

- Latent traits are measured with multiple observed items
  - Models for item responses end up making implicit assumptions about the covariance among items

A (Very) Brief History of Test Theory

- Modern beginnings date to mid 19th century
  - Measurement of intelligence

- 1904 brought about two seminal papers by Charles Spearman
  - One showed how to estimate the amount of error in test scores
    - Led to field of **Classical Test Theory (CTT)**

  - One showed how measure a single trait from a test
    - Led to field of factor analysis
    - Modern versions feature measurement models under the name of **Confirmatory Factor Analysis (CFA)**

Development of the Field of Test Theory

- Motivated by problems in education and psychology
  - Education: measuring intelligence or achievement
  - Psychology: understanding structure of traits; development of scales for measurement of latent traits

- Early theory developed prior to computers
  - Work prior to the 1960s relied on approximations
  - Most of this was under the heading of “exploratory” factor analysis – more on this after we talk more about the subject

- Mathematicians and statisticians have advanced the field in recent years
  - Brought mathematical rigor and validity to approaches

Measurement Models

- Measurement models can be divided into two families of models based on **response format alone:**
  - Continuous responses - **Confirmatory Factor Models**
  - Categorical responses - **Item Response Models**

- Both of these families fall under a larger framework: **Generalized Linear Latent and Mixed Models**
  - Provide measurement models for other types of responses

- Other relevant families:
  - **Structural Equation Models** - provides estimates of correlations amongst latent variables in measurement models
  - **Path Analysis** - simultaneous regression amongst multiple observed variables

Confirmatory Factor Analysis (CFA) Models

- Main idea of CFA: Build a <u>measurement model</u> for response variables that measure the same trait
  - **CFA = Linear regression model** predicting each continuous observed outcome variable (item, subscale) from a latent trait predictor variable(s)

$$Y_{si} = \mu_i + \lambda_{i1} F_{s1} + e_{si}$$

- i - item; s - subject; μ_i is the item intercept; λ_{i1} is the item slope (factor loading for factor 1); e_{is} is the error for the item and subject; Y_{is} is the item response (***<u>assumed continuous</u>***)

- Differs from exploratory factor analysis:
  - Number and content of factors is decided a priori
  - Alternative models are comparable and testable
  - We will compare and contrast the two approaches after discussing CFA

- Uses of confirmatory factor analysis models:
  - Analyze relationships among subscales that have normal, continuous distributions
  - Provide comparability across persons, items, and occasions

Factor Analysis (Y Observed; F latent)

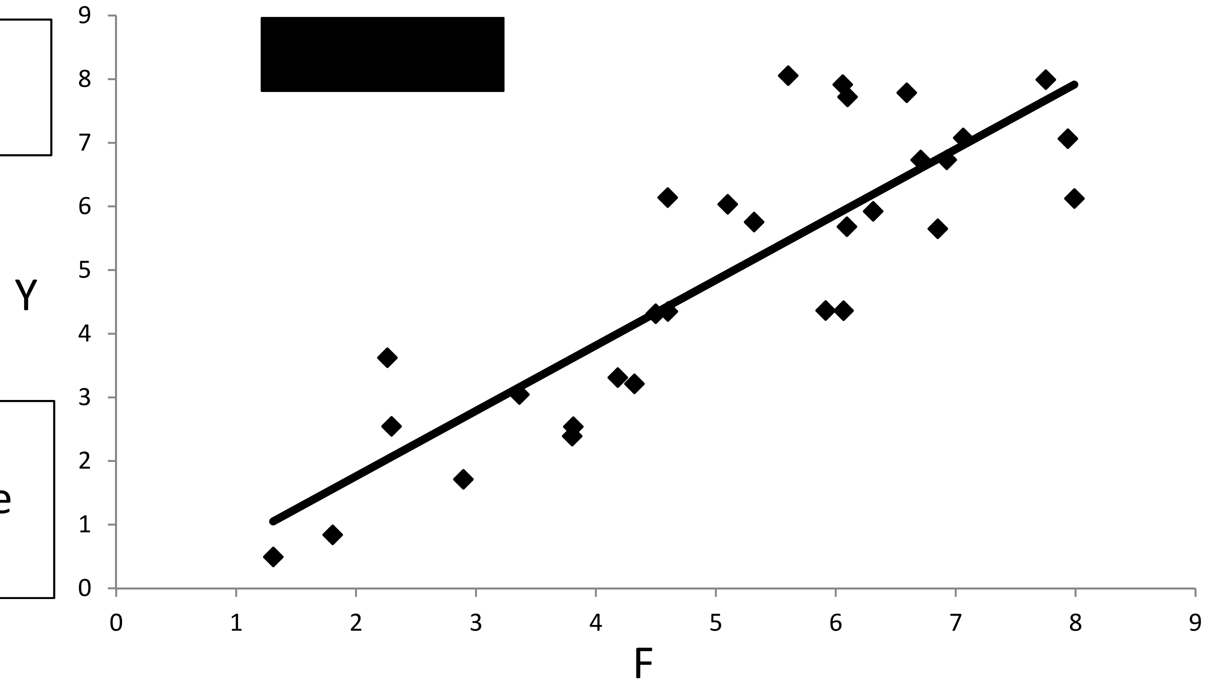
- The prediction of Y is done using a linear regression:

$$Y_{si} = \mu_i + \lambda_{i1} F_{s1} + e_{si}$$

μ is the intercept (where the line crosses the Y axis)

λ is the slope (the increase in Y for a one unit increase in F)

e is the error (or residual), with estimated error variance ψ_i^2



Confirmatory Factor Analysis (CFA)

- Dimensionality is assumed known
  - Local Independence is assumed
    - Errors are independent after controlling for factor(s)

- CFA is a **linear model** - a one-unit change in latent trait/factor F has same increase in expected response Y at all points of Y
  - Implicitly assume that Y is a continuous variable
  - CFA won't work well for binary or categorical data

- Items are allowed to differ from each other in how much they relate to the latent trait, *but a good item is equally good for everybody*
  - This is different from assumptions of classical test theory where all items count the same

A History of "Common Factor Theory" (CFA)

- 1900's: Spearman's G
  - Went looking for single-factor model… and "found" it
  - Led to development of other IQ tests (Stanford-Binet, Wechsler)

- 1930's and 1940's: Thurstone elaborated Spearman's model into a "multiple factor" model
  - Beginnings of exploratory factor analysis to do so
  - Later applied in other personality tests (e.g., MMPI)

- 1940's and 1950's: Guttman's work
  - Factor analysis and test development is about generalizing from measures we have created to more measures of the same kind
  - Thus, need to think about structure *before-hand*

Common Factor Theory, continued

- 1940's: Lawley provided a rigorous foundation for statistical treatment of common factor analysis
  - But had to wait for better computers to be able to implement methods

- 1952: Lawley provided the beginnings of the confirmatory factor model
  - Later extended by Howe and Bargmann (1950's)
  - Further extended by Jöreskog (LISREL – 1970's)

CONFIRMATORY FACTOR MODELS

Confirmatory Factor Models

- To describe CFA models, we will revisit the gambling data
  - The goal: measurement of a “gambling tendency” factor
  - Higher levels of the factor = higher ratings on GRI items
- CFA models work at the **item** level
  - Other variants exist at the scale level – called parceling
    - More on this later
- Our five items will be described with five equations
  - Similar to path analysis, except with a latent factor now involved in the equation for each item

One-Factor Model of Five GRI Items

- The CFA model for the five GRI items:

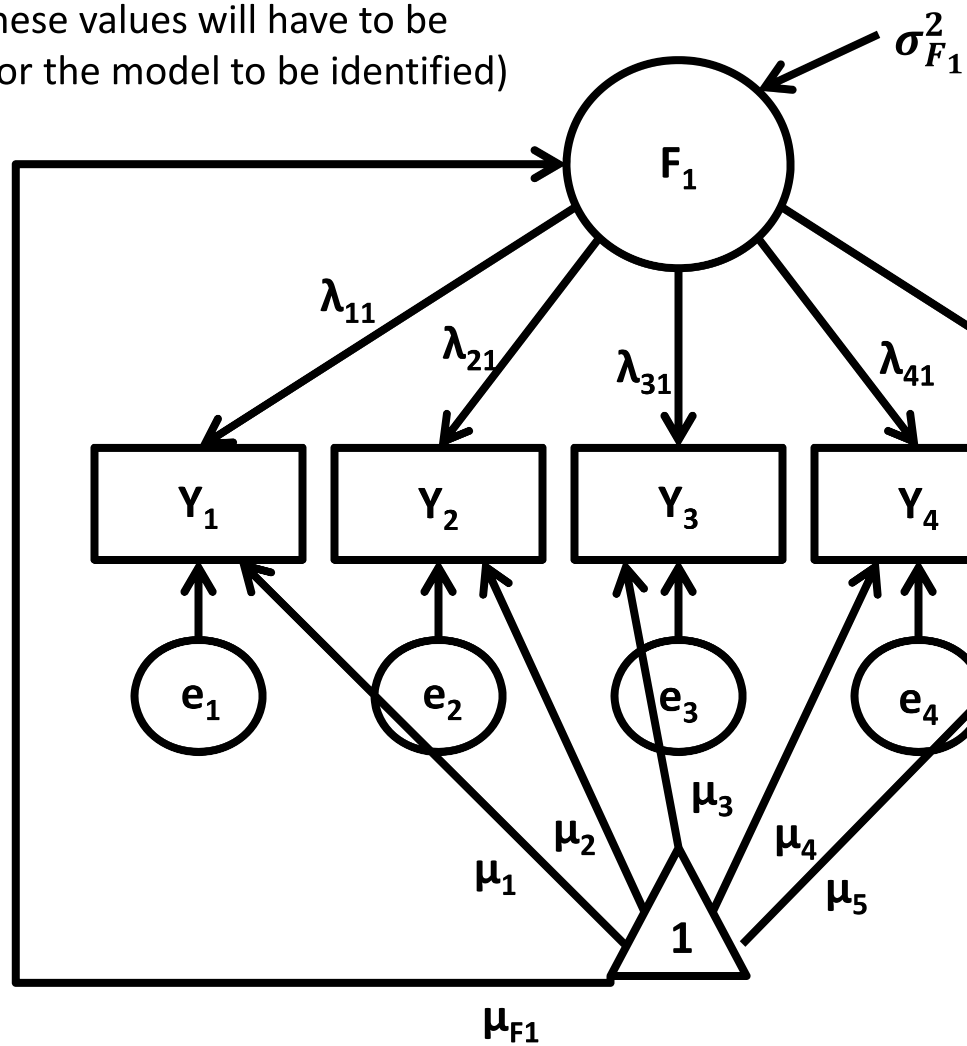
$$Y_{s1} = \mu_1 + \lambda_{11}F_{s1} + e_{s1}$$
$$Y_{s2} = \mu_2 + \lambda_{21}F_{s1} + e_{s2}$$
$$Y_{s3} = \mu_3 + \lambda_{31}F_{s1} + e_{s3}$$
$$Y_{s4} = \mu_4 + \lambda_{41}F_{s1} + e_{s4}$$
$$Y_{s5} = \mu_5 + \lambda_{51}F_{s1} + e_{s5}$$

- Here:
  - Y_{si} - response of subject s on item i
  - μ_i - intercept of item i (listed as a mean as this is typically what it becomes)
  - λ_{i1} - factor loading of item i on factor 1 (only one factor today)
  - F_{s1} - latent "factor score" for subject s (same for all items) to factor 1 (only one today)
  - e_{si} - regression-like residual for subject s on item i
    - We assume $e_{si} \sim N(0, \psi_i^2)$; ψ_i^2 is called the **unique variance** of item i
    - We also assume e_{si} and F_{s1} are independent

- Also, we will assume $F_{s1} \sim N(\mu_{F_1}, \sigma_{F_1}^2)$
  - Typically $\mu_{F_1} = 0$ (but not always)
  - Factor variance can be estimated or fixed (more on both in identification)

Our CFA Model Path Diagram

(Some of these values will have to be restricted for the model to be identified)



Measurement Model:

λ's = factor loadings

e's = error variances

μ's = item intercepts

Structural Model:

$\sigma^2_{F_1}$ = factor variance

μ_{F1} = factor mean

Parsing the CFA Model: What Each Term Does

- The path diagram for the CFA model is a picture that implies a specific model
  - The weights of the paths multiply the quantities from which the paths emanate
  - The resulting linear combinations are the predictions for the variables where the paths terminate

- As with all other models we have used to this point, we will assume a multivariate normal distribution for the data
  - The various model terms combine to predict the elements of the mean vector and covariance matrix of the items
  - Understanding this is the key to understanding how the model works – and which terms have to be fixed

Model Predicted Item Mean

- Using the algebra of expectations, the mean for an item under the our CFA model with a single factor is:

$$E(Y_{si}) = E(\mu_i + \lambda_{i1}F_{s1} + e_{si})$$

$$= E(\mu_i) + E(\lambda_{i1}F_{s1}) + E(e_{si})$$

$$= \mu_i + \lambda_{i1}E(F_{s1}) + E(e_{si})$$

Is zero model specification

$$= \mu_i + \lambda_{i1}\mu_{F_1}$$

Model Predicted Mean Vector

- Combining across all items, the mean vector for the items is given by:

$$\boldsymbol{\mu}_Y = \boldsymbol{\mu}_I + \boldsymbol{\Lambda}\boldsymbol{\mu}_{\boldsymbol{F}}$$

$$\begin{bmatrix} \mu_{Y_1} \\ \mu_{Y_2} \\ \mu_{Y_3} \\ \mu_{Y_4} \\ \mu_{Y_5} \end{bmatrix} = \begin{bmatrix} \mu_{I_1} \\ \mu_{I_2} \\ \mu_{I_3} \\ \mu_{I_4} \\ \mu_{I_5} \end{bmatrix} + \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \\ \lambda_{41} \\ \lambda_{51} \end{bmatrix} \begin{bmatrix} \mu_{F_1} \end{bmatrix} = \begin{bmatrix} \mu_{I_1} + \lambda_{11}\mu_{F_1} \\ \mu_{I_2} + \lambda_{21}\mu_{F_1} \\ \mu_{I_3} + \lambda_{31}\mu_{F_1} \\ \mu_{I_4} + \lambda_{41}\mu_{F_1} \\ \mu_{I_5} + \lambda_{51}\mu_{F_1} \end{bmatrix}$$

Model Predicted Item Variance

- Using the algebra of expectations, the variance for an item under the our CFA model with a single factor is:

$$
\begin{aligned}
Var(Y_{si}) &= Var(\mu_i + \lambda_{i1}F_{s1} + e_{si}) \\
&= Var(\lambda_{i1}F_{s1} + e_{si}) \\
&= Var(\lambda_{i1}F_{s1}) + Var(e_{si}) + 2Cov(F_{s1}, e_{si}) \\
&= \lambda_{i1}^2 Var(F_{s1}) + Var(e_{si}) \\
&= \lambda_{i1}^2 \sigma_{F_1}^2 + \psi_i^2
\end{aligned}
$$

Is zero by independence

We define the variance of e to be the unique variance of the item

Model Predicted Item Covariances

- Using the algebra of expectations, the covariance for a pair of items i and j under the single factor model is:

$$Cov\left(Y_{si}, Y_{sj}\right) = Cov\left(\mu_i + \lambda_{i1}F_{s1} + e_{si}, \mu_j + \lambda_{j1}F_{s1} + e_{sj}\right)$$

$$= Cov\left(\lambda_{i1}F_{s1} + e_{si}, \lambda_{j1}F_{s1} + e_{sj}\right)$$

$$= Cov\left(\lambda_{i1}F_{s1}, \lambda_{j1}F_{s1}\right) + \cancel{Cov\left(\lambda_{i1}F_{s1}, e_{sj}\right)} + \cancel{Cov\left(\lambda_{j1}F_{s1}, \mathrm{e_{sj}}\right)} + \cancel{Cov\left(e_{si}, e_{sj}\right)}$$

$$= \lambda_{i1}\lambda_{j1}Cov(F_{s1}, F_{s1})$$

$$= \lambda_{i1}\lambda_{j1}\sigma^2_{F_1}$$

The covariance of a variable with itself is its variance

Model Implied Covariance Matrix

- Combining across all items, the covariance matrix for the items is given by:

$$\mathbf{\Sigma}_Y = \mathbf{\Lambda\Phi\Lambda}^T + \mathbf{\Psi}$$

  - Get used to seeing this – although you already have (see the path a slides)

$$\begin{bmatrix} \sigma^2_{Y_1} & \sigma_{Y_1,Y_2} & \sigma_{Y_1,Y_3} & \sigma_{Y_1,Y_4} & \sigma_{Y_1,Y_5} \\ \sigma_{Y_1,Y_2} & \sigma^2_{Y_2} & \sigma_{Y_2,Y_3} & \sigma_{Y_2,Y_4} & \sigma_{Y_2,Y_5} \\ \sigma_{Y_1,Y_3} & \sigma_{Y_2,Y_3} & \sigma^2_{Y_3} & \sigma_{Y_3,Y_4} & \sigma_{Y_3,Y_5} \\ \sigma_{Y_1,Y_4} & \sigma_{Y_2,Y_4} & \sigma_{Y_3,Y_4} & \sigma^2_{Y_4} & \sigma_{Y_4,Y_5} \\ \sigma_{Y_1,Y_5} & \sigma_{Y_2,Y_5} & \sigma_{Y_3,Y_5} & \sigma_{Y_4,Y_5} & \sigma^2_{Y_5} \end{bmatrix}$$

$$= \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \\ \lambda_{41} \\ \lambda_{51} \end{bmatrix} \left[\sigma^2_{F_1}\right] \begin{bmatrix} \lambda_{11} & \lambda_{21} & \lambda_{31} & \lambda_{41} & \lambda_{51} \end{bmatrix} + \begin{bmatrix} \psi^2_1 & 0 & 0 & 0 & 0 \\ 0 & \psi^2_2 & 0 & 0 & 0 \\ 0 & 0 & \psi^2_3 & 0 & 0 \\ 0 & 0 & 0 & \psi^2_4 & 0 \\ 0 & 0 & 0 & 0 & \psi^2_5 \end{bmatrix} =$$

Model Implied Covariance Matrix

$$\begin{bmatrix} \lambda_{11}^2\sigma_{F_1}^2 + \psi_1^2 & \lambda_{11}\lambda_{21}\sigma_{F_1}^2 & \lambda_{11}\lambda_{31}\sigma_{F_1}^2 & \lambda_{11}\lambda_{41}\sigma_{F_1}^2 & \lambda_{11}\lambda_{51}\sigma_{F_1}^2 \\ \lambda_{11}\lambda_{21}\sigma_{F_1}^2 & \lambda_{21}^2\sigma_{F_1}^2 + \psi_2^2 & \lambda_{21}\lambda_{31}\sigma_{F_1}^2 & \lambda_{21}\lambda_{41}\sigma_{F_1}^2 & \lambda_{21}\lambda_{51}\sigma_{F_1}^2 \\ \lambda_{11}\lambda_{31}\sigma_{F_1}^2 & \lambda_{21}\lambda_{31}\sigma_{F_1}^2 & \lambda_{31}^2\sigma_{F_1}^2 + \psi_3^2 & \lambda_{31}\lambda_{41}\sigma_{F_1}^2 & \lambda_{31}\lambda_{51}\sigma_{F_1}^2 \\ \lambda_{11}\lambda_{41}\sigma_{F_1}^2 & \lambda_{21}\lambda_{41}\sigma_{F_1}^2 & \lambda_{31}\lambda_{41}\sigma_{F_1}^2 & \lambda_{41}^2\sigma_{F_1}^2 + \psi_4^2 & \lambda_{41}\lambda_{51}\sigma_{F_1}^2 \\ \lambda_{11}\lambda_{51}\sigma_{F_1}^2 & \lambda_{21}\lambda_{51}\sigma_{F_1}^2 & \lambda_{31}\lambda_{51}\sigma_{F_1}^2 & \lambda_{41}\lambda_{51}\sigma_{F_1}^2 & \lambda_{51}^2\sigma_{F_1}^2 + \psi_5^2 \end{bmatrix}$$

Counting Parameters

- As in path analysis (and any MVN-assumed data model), the total number of parameters cannot exceed the number of parameters in a saturated model

- With 5 variables, our saturated model has 20 parameters:
  - 5 means
  - 5 variances
  - 10 covariances

- The 1-factor CFA model we have built has 17 parameters:
  - 5 factor loadings
  - 5 unique variances
  - 5 item intercepts
  - 1 factor mean
  - 1 factor variance

- However, as it is stated, the model is not identified
  - "Local" identification is an issue (too many parameters within sub-equations)

Local Identification Issue #1: Mean Vector

- Examining the mean vector we discover our first issue with local identification:
  - We have 5 means in our saturated mean vector (one for each item): $\mu_{Y_1}, \mu_{Y_2}, \mu_{Y_3}, \mu_{Y_4}, \mu_{Y_5}$
  - Disregarding the factor loadings (for a moment – these are used in the covariance matrix), we have 6 mean parameters:
    - 5 item indicators: $\mu_{I_1}, \mu_{I_2}, \mu_{I_3}, \mu_{I_4}, \mu_{I_5}$
    - 1 factor mean: μ_{F_1}

- This is the source of the mean vector portion of our local independence issue – too many parameters for our model

Local Identification Issue #2: Covariance Matrix

- Examining the covariance matrix we discover our second issue with local identification:
  - We have 5 variances in our saturated covariance matrix (one for each item): $\sigma^2_{Y_1}, \sigma^2_{Y_2}, \sigma^2_{Y_3}, \sigma^2_{Y_4}, \sigma^2_{Y_5}$
  - We have 10 covariance in our saturated covariance matrix (one for each pair of items): σ_{Y_1,Y_2}, σ_{Y_1,Y_3}, σ_{Y_1,Y_4}, σ_{Y_1,Y_5}, σ_{Y_2,Y_3}, σ_{Y_2,Y_4}, σ_{Y_2,Y_5}, σ_{Y_3,Y_4}, σ_{Y_3,Y_5}, σ_{Y_4,Y_5}

- Our factor model has 11 parameters for the covariance matrix:
  - 5 factor loadings: $\lambda_{11}, \lambda_{21}, \lambda_{31}, \lambda_{41}, \lambda_{51}$
  - 5 unique variances: $\psi^2_1, \psi^2_2, \psi^2_3, \psi^2_4, \psi^2_5$
  - 1 factor variance: $\sigma^2_{F_1}$

- The issue, though, is that for every:
  - Pair of saturated parameters – 3 factor model parameters
  - Triple of saturated parameters – 4 factor model parameters

- The CFA model we specified is <u>**unidentified**</u>

CFA MODEL IDENTIFICATION

CFA Model Identification

- The CFA model was unidentified as specified

- The issue comes from the nature of the latent variable
  - ***<u>In reality it does not exist</u>***

- Because it does not exist, it cannot have its own scale
  - Meaning, it does not have a mean and variance by itself

- Therefore, we must identify our model by picking the scale of the factor
  - The choice is arbitrary – factors do not exist!

Options for Selecting Factor Scale and Location

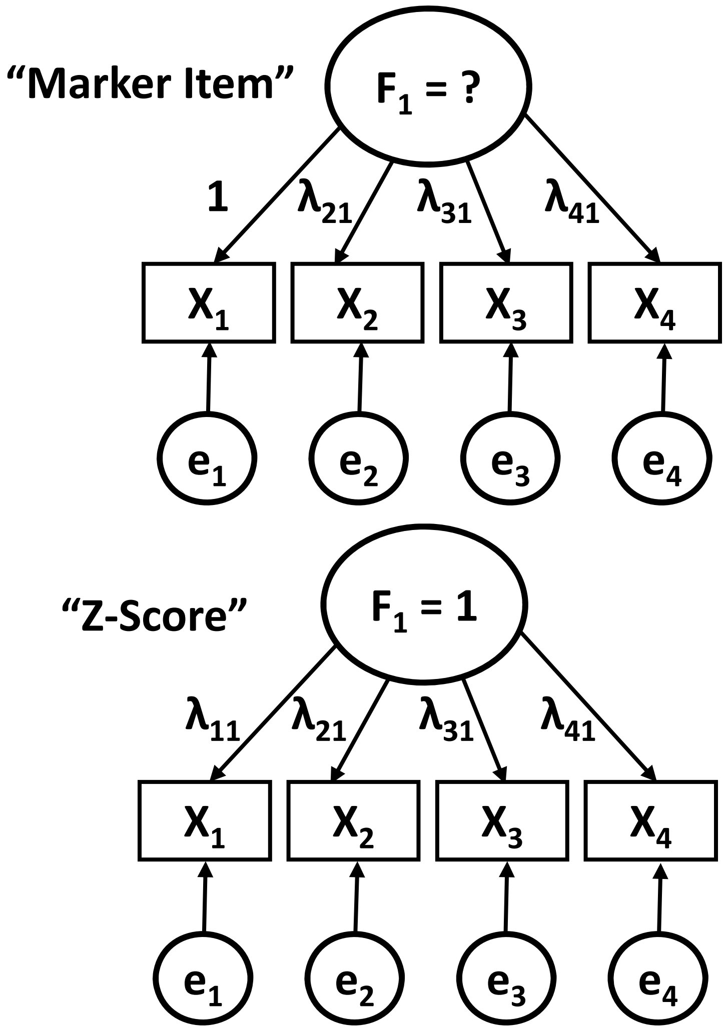
- For both the factor scale (factor variance) and location (factor mean) there are two options in widespread use:
  - **<u>Any combination will be equivalent for the CFA model</u>**

	Estimated Factor Variance (marker item)	**Standardized Factor Variance (fixed variance)**
Estimated Factor Mean (marker item)	Set one item's factor loading to 1 Set one item's intercept to 0	Set the factor variance to 1 Set one item's intercept to 0
Standardized Factor Mean (fixed mean)	Set one item's factor loading to 1 Set the factor mean to 0	Set the factor variance to 1 Set the factor mean to 0

Mplus Default

Common in Scale Building

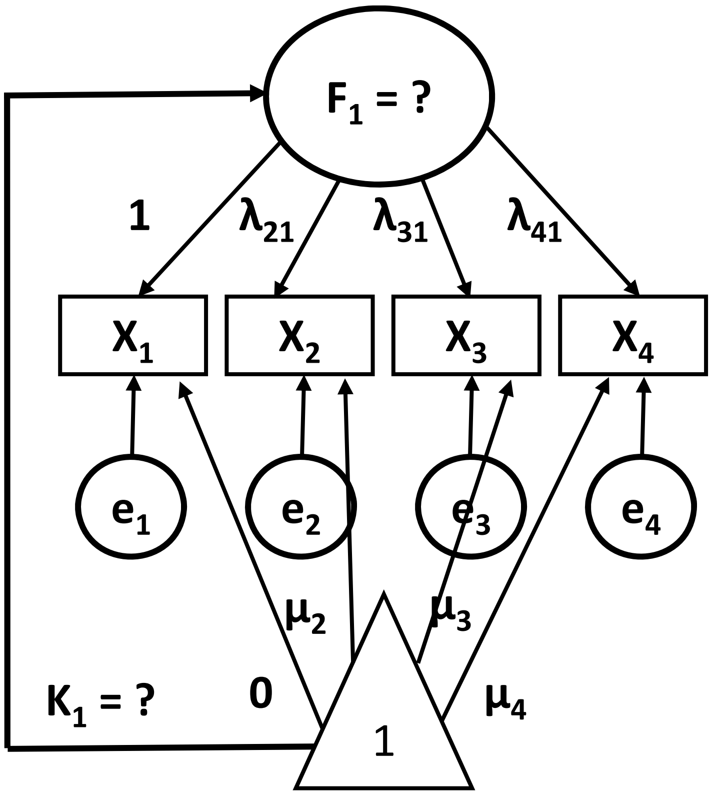
CFA Model Identification:
Create a Scale for the Latent Variable

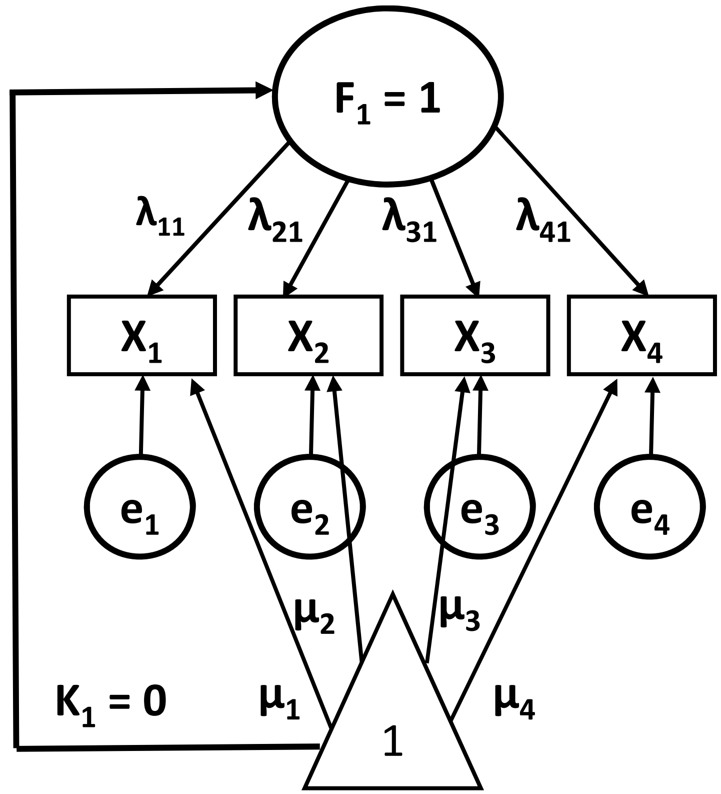


- The factor doesn't exist, so **it needs a scale** (a mean and variance):
- Two **equivalent** options to do so

- Create a scale for the VARIANCE:
  - **1) Scale using a marker item**
    - Fix one loading to 1; factor is scaled as "reliable" variance of marker item
    - Could fix to any value except 0 (but 1 makes interpretation easy)
    - Loading = .9, variance =16? $Var(F_1) = (.9^2)*16 = 12.96$
    - Good for models where factors are predicted (factor variance can change based on predictors)
  - **2) Fix factor variance to 1**
    - Factor is interpreted as z-score
    - Good for scale building – although same results can be had with #1

CFA Model Identification:
Two Options for Scaling the Factor Mean





"Marker Item" → Fix 1 item intercept to 0; estimate factor mean

Item intercept is expected outcome when factor = 0 (when item = 0)

"Z-Score" → Fix factor mean to 0, estimate all item intercepts

Item intercept is expected outcome when factor = 0 (when item = mean)

Unpacking Marker Item Identification with Equations

- Using the marker item identification for the factor scale and the zero factor mean identification for the factor location, the CFA model for the five GRI items is:

$$\begin{aligned} Y_{s1} &= \mu_1 + (1)F_{s1} + e_{s1} \\ Y_{s2} &= \mu_2 + \lambda_{21}F_{s1} + e_{s2} \\ Y_{s3} &= \mu_3 + \lambda_{31}F_{s1} + e_{s3} \\ Y_{s4} &= \mu_4 + \lambda_{41}F_{s1} + e_{s4} \\ Y_{s5} &= \mu_5 + \lambda_{51}F_{s1} + e_{s5} \end{aligned}$$

Choice of marker item is arbitrary (and equivalent) so long as the item measures factor

- Here:
  - Y_{si} - response of subject s on item i
  - μ_i - intercept of item i
  - λ_{i1} - factor loading of item i
  - F_{s1} - latent "factor score" for subject s (same for all items)
  - e_{si} - regression-like residual for subject s on item i
    - We assume $e_{si} \sim N(0, \psi_i^2)$; ψ_i^2 is called the **unique variance** of item i
    - We also assume e_{si} and F_{s1} are independent

- Because of identification, we will assume $F_{s1} \sim N(0, \sigma_{F_1}^2)$

Fixed Factor Mean

Estimated Factor Variance

Fixed Factor Mean: Model Predicted Mean Vector

- Combining across all items, the mean vector for the items is given by:

$$\boldsymbol{\mu}_Y = \boldsymbol{\mu}_I + \boldsymbol{\Lambda}\mathbf{0}$$

Factor Mean is fixed at 0

$$\begin{bmatrix}\mu_{Y_1}\\ \mu_{Y_2}\\ \mu_{Y_3}\\ \mu_{Y_4}\\ \mu_{Y_5}\end{bmatrix} = \begin{bmatrix}\mu_{I_1}\\ \mu_{I_2}\\ \mu_{I_3}\\ \mu_{I_4}\\ \mu_{I_5}\end{bmatrix} + \begin{bmatrix}1\\ \lambda_{21}\\ \lambda_{31}\\ \lambda_{41}\\ \lambda_{51}\end{bmatrix}[0] = \begin{bmatrix}\mu_{I_1} + 1^*(0)\\ \mu_{I_2} + \lambda_{21}(0)\\ \mu_{I_3} + \lambda_{31}(0)\\ \mu_{I_4} + \lambda_{41}(0)\\ \mu_{I_5} + \lambda_{51}(0)\end{bmatrix} = \begin{bmatrix}\mu_{I_1}\\ \mu_{I_2}\\ \mu_{I_3}\\ \mu_{I_4}\\ \mu_{I_5}\end{bmatrix}$$

Marker item factor loading shown as a 1 (but it does not contribute here as the factor mean is fixed at zero)

All Item Intercepts Estimated – And all equal item mean

Model Implied Covariance Matrix

- Combining across all items, the covariance matrix for the items is given by:

$$\boldsymbol{\Sigma}_Y = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi}$$

Marker Item Factor Loading is fixed at 1

$$\begin{bmatrix} \sigma^2_{Y_1} & \sigma_{Y_1,Y_2} & \sigma_{Y_1,Y_3} & \sigma_{Y_1,Y_4} & \sigma_{Y_1,Y_5} \\ \sigma_{Y_1,Y_2} & \sigma^2_{Y_2} & \sigma_{Y_2,Y_3} & \sigma_{Y_2,Y_4} & \sigma_{Y_2,Y_5} \\ \sigma_{Y_1,Y_3} & \sigma_{Y_2,Y_3} & \sigma^2_{Y_3} & \sigma_{Y_3,Y_4} & \sigma_{Y_3,Y_5} \\ \sigma_{Y_1,Y_4} & \sigma_{Y_2,Y_4} & \sigma_{Y_3,Y_4} & \sigma^2_{Y_4} & \sigma_{Y_4,Y_5} \\ \sigma_{Y_1,Y_5} & \sigma_{Y_2,Y_5} & \sigma_{Y_3,Y_5} & \sigma_{Y_4,Y_5} & \sigma^2_{Y_5} \end{bmatrix}$$

$$= \begin{bmatrix} 1 \\ \lambda_{21} \\ \lambda_{31} \\ \lambda_{41} \\ \lambda_{51} \end{bmatrix} \begin{bmatrix} \sigma^2_{F_1} \end{bmatrix} \begin{bmatrix} 1 & \lambda_{21} & \lambda_{31} & \lambda_{41} & \lambda_{51} \end{bmatrix} + \begin{bmatrix} \psi^2_1 & 0 & 0 & 0 & 0 \\ 0 & \psi^2_2 & 0 & 0 & 0 \\ 0 & 0 & \psi^2_3 & 0 & 0 \\ 0 & 0 & 0 & \psi^2_4 & 0 \\ 0 & 0 & 0 & 0 & \psi^2_5 \end{bmatrix} =$$

Marker Item Model Implied Covariance Matrix

$$\begin{bmatrix} (1)\sigma_{F_1}^2 + \psi_1^2 & (1)\lambda_{21}\sigma_{F_1}^2 & (1)\lambda_{31}\sigma_{F_1}^2 & (1)\lambda_{41}\sigma_{F_1}^2 & (1)\lambda_{51}\sigma_{F_1}^2 \\ (1)\lambda_{21}\sigma_{F_1}^2 & \lambda_{21}^2\sigma_{F_1}^2 + \psi_2^2 & \lambda_{21}\lambda_{31}\sigma_{F_1}^2 & \lambda_{21}\lambda_{41}\sigma_{F_1}^2 & \lambda_{21}\lambda_{51}\sigma_{F_1}^2 \\ (1)\lambda_{31}\sigma_{F_1}^2 & \lambda_{21}\lambda_{31}\sigma_{F_1}^2 & \lambda_{31}^2\sigma_{F_1}^2 + \psi_3^2 & \lambda_{31}\lambda_{41}\sigma_{F_1}^2 & \lambda_{31}\lambda_{51}\sigma_{F_1}^2 \\ (1)\lambda_{41}\sigma_{F_1}^2 & \lambda_{21}\lambda_{41}\sigma_{F_1}^2 & \lambda_{31}\lambda_{41}\sigma_{F_1}^2 & \lambda_{41}^2\sigma_{F_1}^2 + \psi_4^2 & \lambda_{41}\lambda_{51}\sigma_{F_1}^2 \\ (1)\lambda_{51}\sigma_{F_1}^2 & \lambda_{21}\lambda_{51}\sigma_{F_1}^2 & \lambda_{31}\lambda_{51}\sigma_{F_1}^2 & \lambda_{41}\lambda_{51}\sigma_{F_1}^2 & \lambda_{51}^2\sigma_{F_1}^2 + \psi_5^2 \end{bmatrix}$$

Marker Item Variance is Partitioned into variance due to factor and variance due to error

Factor Variance becomes proxy for marker item's factor loading (appears in covariance with other items)

Our marker-item-fixed-factor-mean CFA Model in Mplus

- Mplus syntax:
  - The “BY” statement is for factor analysis
    - factorname BY item list

```
TITLE:
    Gambling Research Instrument Items
    Data from 1192 College Students/112 Gamblers
    41 Likert Items (1-6): GRI1-GRI41
    ========================================
    Identification: Zero Factor Mean, Marker Item
    ========================================

DATA:
    FILE = alldata_gri.csv;

VARIABLE:
    NAMES = GRI1-GRI41;
    USEVARIABLES = GRI1-GRI5;

    MISSING = ALL(99);

ANALYSIS:
    ESTIMATOR = MLR;

MODEL:
    GAMBLING BY GRI1-GRI5;

OUTPUT:
    SAMPSTAT STANDARDIZED MODINDICES(ALL 0) RESIDUAL;

SAVEDATA:
    SAVE = FSCORES;            !saves latent trait estimates
    FILE = alldata_gri_person.dat; !puts latent trait estimates into file named *.dat
```

Mplus Estimates

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 GAMBLING BY
    GRI1               1.000      0.000    999.000    999.000
    GRI2               1.293      0.121     10.652      0.000
    GRI3               0.765      0.077      9.959      0.000
    GRI4               1.297      0.138      9.379      0.000
    GRI5               1.040      0.094     11.042      0.000

 Intercepts
    GRI1               1.804      0.028     64.252      0.000
    GRI2               1.932      0.045     43.195      0.000
    GRI3               1.551      0.024     64.701      0.000
    GRI4               1.926      0.038     51.046      0.000
    GRI5               1.577      0.026     59.592      0.000

 Variances
    GAMBLING           0.363      0.052      6.969      0.000

 Residual Variances
    GRI1               0.666      0.057     11.716      0.000
    GRI2               2.001      0.138     14.504      0.000
    GRI3               0.537      0.040     13.402      0.000
    GRI4               1.245      0.094     13.291      0.000
    GRI5               0.520      0.047     11.050      0.000
```

Factor loadings estimates are located under the name of the factor

The first factor loading is listed as 1.000, with a standard error 0.000

This indicates the value is fixed

Mplus Estimates

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GAMBLING BY				
GRI1	1.000	0.000	999.000	999.000
GRI2	1.293	0.121	10.652	0.000
GRI3	0.765	0.077	9.959	0.000
GRI4	1.297	0.138	9.379	0.000
GRI5	1.040	0.094	11.042	0.000
Intercepts				
GRI1	1.804	0.028	64.252	0.000
GRI2	1.932	0.045	43.195	0.000
GRI3	1.551	0.024	64.701	0.000
GRI4	1.926	0.038	51.046	0.000
GRI5	1.577	0.026	59.592	0.000
Variances				
GAMBLING	0.363	0.052	6.969	0.000
Residual Variances				
GRI1	0.666	0.057	11.716	0.000
GRI2	2.001	0.138	14.504	0.000
GRI3	0.537	0.040	13.402	0.000
GRI4	1.245	0.094	13.291	0.000
GRI5	0.520	0.047	11.050	0.000

$$\widehat{\boldsymbol{\Lambda}} = \begin{bmatrix} 1 \\ 1.293 \\ 0.765 \\ 1.297 \\ 1.040 \end{bmatrix}$$

$$\widehat{\boldsymbol{\mu}}_I = \begin{bmatrix} 1.804 \\ 1.932 \\ 1.551 \\ 1.926 \\ 1.577 \end{bmatrix}$$

$$\hat{\sigma}_F^2 = 0.363$$

$$\widehat{\boldsymbol{\Psi}} = \begin{bmatrix} 0.666 & 0 & 0 & 0 & 0 \\ 0 & 2.001 & 0 & 0 & 0 \\ 0 & 0 & 0.537 & 0 & 0 \\ 0 & 0 & 0 & 1.245 & 0 \\ 0 & 0 & 0 & 0 & 0.520 \end{bmatrix}$$

MODEL FIT INFORMATION

Number of Free Parameters	15
Loglikelihood	
H0 Value	-9564.556
H0 Scaling Correction Factor for MLR	1.8591
H1 Value	-9547.913
H1 Scaling Correction Factor for MLR	1.8825

Mplus Model Estimated Mean Vector

- From Mplus:

Model Estimated Means/Intercepts/Thresholds

	GRI1	GRI2	GRI3	GRI4	GRI5
1	1.804	1.932	1.551	1.926	1.577

- From the results:

$$\boxed{\hat{\boldsymbol{\mu}}_I = \begin{bmatrix} 1.804 \\ 1.932 \\ 1.551 \\ 1.926 \\ 1.577 \end{bmatrix}}$$

- Note: this is a saturated means model
  - Historically, this is the case for factor analysis
  - This will change in later methods (hence the inclusion here)

Mplus Model Estimated Covariance Matrix

- From Mplus:

Model Estimated Covariances/Correlations/Residual Correlations

	GRI1	GRI2	GRI3	GRI4	GRI5
GRI1	1.028				
GRI2	0.469	2.608			
GRI3	0.277	0.359	0.749		
GRI4	0.471	0.609	0.360	1.856	
GRI5	0.377	0.488	0.289	0.489	0.913

- Using the estimates:

$$\widehat{\boldsymbol{\Lambda}} = \begin{bmatrix} 1 \\ 1.293 \\ 0.765 \\ 1.297 \\ 1.040 \end{bmatrix}$$

$$\hat{\sigma}_F^2 = 0.363$$

$$\widehat{\boldsymbol{\Psi}} = \begin{bmatrix} 0.666 & 0 & 0 & 0 & 0 \\ 0 & 2.001 & 0 & 0 & 0 \\ 0 & 0 & 0.537 & 0 & 0 \\ 0 & 0 & 0 & 1.245 & 0 \\ 0 & 0 & 0 & 0 & 0.520 \end{bmatrix}$$

$$\begin{bmatrix} 0.363 + 0.666 & 1.293^{*}0.363 & 0.765^{*}0.363 & 1.297^{*}0.363 & 1.040^{*}0.363 \\ 1.293^{*}0.363 & 1.293^{2^{*}}0.363 + 2.001 & 1.293^{*}0.765^{*}0.363 & 1.293^{*}1.297^{*}0.363 & 1.293^{*}1.040^{*}0.363 \\ 0.765^{*}0.363 & 1.293^{*}0.765^{*}0.363 & 0.765^{2^{*}}0.363 + 0.537 & 0.765^{*}1.297^{*}0.363 & 0.765^{*}1.040^{*}0.363 \\ 1.297^{*}0.363 & 1.293^{*}1.297^{*}0.363 & 0.765^{*}1.297^{*}0.363 & 1.297^{2^{*}}0.363 + 1.245 & 1.297^{*}1.040^{*}0.363 \\ 1.040^{*}0.363 & 1.293^{*}1.040^{*}0.363 & 0.765^{*}1.040^{*}0.363 & 1.297^{*}1.040^{*}0.363 & 1.040^{2^{*}}0.363 + 0.520 \end{bmatrix}$$

Unpacking Standardized Factor Identification with Equations

- Using the unit variance identification for the factor scale and the zero factor mean identification for the factor location, the CFA model for the five GRI items is:

$$
\begin{aligned}
Y_{s1} &= \mu_1 + \lambda_{11}F_{s1} + e_{s1} \\
Y_{s2} &= \mu_2 + \lambda_{21}F_{s1} + e_{s2} \\
Y_{s3} &= \mu_3 + \lambda_{31}F_{s1} + e_{s3} \\
Y_{s4} &= \mu_4 + \lambda_{41}F_{s1} + e_{s4} \\
Y_{s5} &= \mu_5 + \lambda_{51}F_{s1} + e_{s5}
\end{aligned}
$$

All factor loadings estimated

- Here:
  - Y_{si} - response of subject s on item i
  - μ_i - intercept of item i
  - λ_{i1} - factor loading of item i
  - F_{s1} - latent "factor score" for subject s (same for all items)
  - e_{si} - regression-like residual for subject s on item i
    - We assume $e_{si} \sim N(0, \psi_i^2)$; ψ_i^2 is called the **unique variance** of item i
    - We also assume e_{si} and F_{s1} are independent

- Because of identification, we will assume $F_{s1} \sim N(0,1)$

Fixed Factor Mean

Fixed Factor Variance

Fixed Factor Mean: Model Predicted Mean Vector

- Combining across all items, the mean vector for the items is given by:

$$\boldsymbol{\mu}_Y = \boldsymbol{\mu}_I + \boldsymbol{\Lambda}\mathbf{0}$$

Factor Mean is fixed at 0

$$\begin{bmatrix} \mu_{Y_1} \\ \mu_{Y_2} \\ \mu_{Y_3} \\ \mu_{Y_4} \\ \mu_{Y_5} \end{bmatrix} = \begin{bmatrix} \mu_{I_1} \\ \mu_{I_2} \\ \mu_{I_3} \\ \mu_{I_4} \\ \mu_{I_5} \end{bmatrix} + \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \\ \lambda_{41} \\ \lambda_{51} \end{bmatrix} [0] = \begin{bmatrix} \mu_{I_1} + \lambda_{11}(0) \\ \mu_{I_2} + \lambda_{21}(0) \\ \mu_{I_3} + \lambda_{31}(0) \\ \mu_{I_4} + \lambda_{41}(0) \\ \mu_{I_5} + \lambda_{51}(0) \end{bmatrix} = \begin{bmatrix} \mu_{I_1} \\ \mu_{I_2} \\ \mu_{I_3} \\ \mu_{I_4} \\ \mu_{I_5} \end{bmatrix}$$

This is the same as the previous model – the mean vector consists of the item intercepts when the factor mean is fixed at zero.

All Item Intercepts Estimated – And all equal item mean

Model Implied Covariance Matrix

- Combining across all items, the covariance matrix for the items is given by:

$$\mathbf{\Sigma}_Y = \mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}^T + \mathbf{\Psi}$$

$$\begin{bmatrix} \sigma^2_{Y_1} & \sigma_{Y_1,Y_2} & \sigma_{Y_1,Y_3} & \sigma_{Y_1,Y_4} & \sigma_{Y_1,Y_5} \\ \sigma_{Y_1,Y_2} & \sigma^2_{Y_2} & \sigma_{Y_2,Y_3} & \sigma_{Y_2,Y_4} & \sigma_{Y_2,Y_5} \\ \sigma_{Y_1,Y_3} & \sigma_{Y_2,Y_3} & \sigma^2_{Y_3} & \sigma_{Y_3,Y_4} & \sigma_{Y_3,Y_5} \\ \sigma_{Y_1,Y_4} & \sigma_{Y_2,Y_4} & \sigma_{Y_3,Y_4} & \sigma^2_{Y_4} & \sigma_{Y_4,Y_5} \\ \sigma_{Y_1,Y_5} & \sigma_{Y_2,Y_5} & \sigma_{Y_3,Y_5} & \sigma_{Y_4,Y_5} & \sigma^2_{Y_5} \end{bmatrix}$$

All factor loadings are estimated

$$= \begin{bmatrix} \lambda_{11} \\ \lambda_{21} \\ \lambda_{31} \\ \lambda_{41} \\ \lambda_{51} \end{bmatrix} [1] [\lambda_{11} \quad \lambda_{21} \quad \lambda_{31} \quad \lambda_{41} \quad \lambda_{51}] + \begin{bmatrix} \psi^2_1 & 0 & 0 & 0 & 0 \\ 0 & \psi^2_2 & 0 & 0 & 0 \\ 0 & 0 & \psi^2_3 & 0 & 0 \\ 0 & 0 & 0 & \psi^2_4 & 0 \\ 0 & 0 & 0 & 0 & \psi^2_5 \end{bmatrix} =$$

Factor variance fixed to 1

Marker Item Model Implied Covariance Matrix

$$\begin{bmatrix} \lambda_{11}^2(1)+\psi_1^2 & \lambda_{11}\lambda_{21}(1) & \lambda_{11}\lambda_{31}(1) & \lambda_{11}\lambda_{41}(1) & \lambda_{11}\lambda_{51}(1) \\ \lambda_{11}\lambda_{21}(1) & \lambda_{21}^2(1)+\psi_2^2 & \lambda_{21}\lambda_{31}(1) & \lambda_{21}\lambda_{41}(1) & \lambda_{21}\lambda_{51}(1) \\ \lambda_{11}\lambda_{31}(1) & \lambda_{21}\lambda_{31}(1) & \lambda_{31}^2(1)+\psi_3^2 & \lambda_{31}\lambda_{41}(1) & \lambda_{31}\lambda_{51}(1) \\ \lambda_{11}\lambda_{41}(1) & \lambda_{21}\lambda_{41}(1) & \lambda_{31}\lambda_{41}(1) & \lambda_{41}^2(1)+\psi_4^2 & \lambda_{41}\lambda_{51}(1) \\ \lambda_{11}\lambda_{51}(1) & \lambda_{21}\lambda_{51}(1) & \lambda_{31}\lambda_{51}(1) & \lambda_{41}\lambda_{51}(1) & \lambda_{51}^2(1)+\psi_5^2 \end{bmatrix}$$

Each item's variance is partitioned into portion due to factor (loading squared) and portion due to error

Factor variance term now disappears from each part of covariance matrix (1)

Our Standardized Factor CFA Model in Mplus

- Mplus syntax:
  - The “BY” statement is for factor analysis
    - factorname BY item list

```
TITLE:
    Gambling Research Instrument Items
    Data from 1192 College Students/112 Gamblers
    41 Likert Items (1-6): GRI1-GRI41
    ========================================
    Identification: Zero Factor Mean, One Factor Variance
    ========================================

DATA:
    FILE = alldata_gri.csv;

VARIABLE:
    NAMES = GRI1-GRI41;
    USEVARIABLES = GRI1-GRI5;
    MISSING = ALL(99);

ANALYSIS:
    ESTIMATOR = MLR;

MODEL:
    GAMBLING by GRI1* GRI2-GRI5;
    GAMBLING@1;

OUTPUT:
    STANDARDIZED MODINDICES(ALL 0) RESIDUAL;

SAVEDATA:
    SAVE = FSCORES;            !saves latent trait estimates
    FILE = alldata_gri_person.dat; !puts latent trait estimates into file named *.dat
```

GRI1* frees the default marker item constraint in Mplus

GAMBLING@1 fixes the factor variance to 1

Mplus Estimates

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 GAMBLING BY
    GRI1               0.602      0.043     13.939      0.000
    GRI2               0.779      0.055     14.086      0.000
    GRI3               0.461      0.040     11.459      0.000
    GRI4               0.781      0.050     15.529      0.000
    GRI5               0.626      0.044     14.166      0.000

 Intercepts
    GRI1               1.804      0.028     64.251      0.000
    GRI2               1.932      0.045     43.195      0.000
    GRI3               1.551      0.024     64.699      0.000
    GRI4               1.926      0.038     51.046      0.000
    GRI5               1.577      0.026     59.590      0.000

 Variances
    GAMBLING           1.000      0.000    999.000    999.000

 Residual Variances
    GRI1               0.666      0.057     11.716      0.000
    GRI2               2.001      0.138     14.504      0.000
    GRI3               0.537      0.040     13.402      0.000
    GRI4               1.246      0.094     13.292      0.000
    GRI5               0.520      0.047     11.050      0.000
```

Factor loadings estimates are located under the name of the factor

The factor variance is listed as 1.000, with a standard error 0.000

This indicates the value is fixed

Mplus Estimates

```
MODEL FIT INFORMATION

Number of Free Parameters                       15

Loglikelihood

          H0 Value                       -9564.556
          H0 Scaling Correction Factor      1.8591
            for MLR
          H1 Value                       -9547.913
          H1 Scaling Correction Factor      1.8825
            for MLR
```

```
MODEL RESULTS

                                                    Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value

 GAMBLING BY
    GRI1               0.602      0.043     13.939      0.000
    GRI2               0.779      0.055     14.086      0.000
    GRI3               0.461      0.040     11.459      0.000
    GRI4               0.781      0.050     15.529      0.000
    GRI5               0.626      0.044     14.166      0.000

 Intercepts
    GRI1               1.804      0.028     64.251      0.000
    GRI2               1.932      0.045     43.195      0.000
    GRI3               1.551      0.024     64.699      0.000
    GRI4               1.926      0.038     51.046      0.000
    GRI5               1.577      0.026     59.590      0.000

 Variances
    GAMBLING           1.000      0.000    999.000    999.000

 Residual Variances
    GRI1               0.666      0.057     11.716      0.000
    GRI2               2.001      0.138     14.504      0.000
    GRI3               0.537      0.040     13.402      0.000
    GRI4               1.246      0.094     13.292      0.000
    GRI5               0.520      0.047     11.050      0.000
```

$$\widehat{\boldsymbol{\Lambda}} = \begin{bmatrix} .602 \\ .779 \\ .461 \\ .781 \\ .626 \end{bmatrix}$$

$$\widehat{\boldsymbol{\mu}}_I = \begin{bmatrix} 1.804 \\ 1.932 \\ 1.551 \\ 1.926 \\ 1.577 \end{bmatrix}$$

$$\hat{\sigma}_F^2 = 1.000$$

$$\widehat{\boldsymbol{\Psi}} = \begin{bmatrix} 0.666 & 0 & 0 & 0 & 0 \\ 0 & 2.001 & 0 & 0 & 0 \\ 0 & 0 & 0.537 & 0 & 0 \\ 0 & 0 & 0 & 1.245 & 0 \\ 0 & 0 & 0 & 0 & 0.520 \end{bmatrix}$$

The item intercepts and unique variances are the same from the marker item analysis

Mplus Model Estimated Mean Vector

- From Mplus:

```
Model Estimated Means/Intercepts/Thresholds
        GRI1          GRI2          GRI3          GRI4          GRI5
     ________      ________      ________      ________      ________
 1      1.804         1.932         1.551         1.926         1.577
```

- From the results:

$$\boxed{\hat{\boldsymbol{\mu}}_I = \begin{bmatrix} 1.804 \\ 1.932 \\ 1.551 \\ 1.926 \\ 1.577 \end{bmatrix}}$$

- Note: this is a saturated means model
  - Historically, this is the case for factor analysis
  - This will change in later methods (hence the inclusion here)

Mplus Model Estimated Covariance Matrix

- From Mplus:

Model Estimated Covariances/Correlations/Residual Correlations

	GRI1	GRI2	GRI3	GRI4	GRI5
GRI1	1.028				
GRI2	0.469	2.608			
GRI3	0.277	0.359	0.749		
GRI4	0.471	0.609	0.360	1.856	
GRI5	0.377	0.488	0.289	0.489	0.913

- Using the estimates:

$$\widehat{\boldsymbol{\Lambda}} = \begin{bmatrix} .602 \\ .779 \\ .461 \\ .781 \\ .626 \end{bmatrix}$$

$$\hat{\sigma}_F^2 = 1.000$$

$$\widehat{\boldsymbol{\Psi}} = \begin{bmatrix} 0.666 & 0 & 0 & 0 & 0 \\ 0 & 2.001 & 0 & 0 & 0 \\ 0 & 0 & 0.537 & 0 & 0 \\ 0 & 0 & 0 & 1.245 & 0 \\ 0 & 0 & 0 & 0 & 0.520 \end{bmatrix}$$

$$\begin{bmatrix} .602^2 + 0.666 & .602^* .779 & .602^* .461 & .602^* .781 & .602^* .626 \\ .602^* .779 & .779^2 + 2.001 & .779^* .461 & .779^* .781 & .779^* .626 \\ .602^* .461 & .779^* .461 & .461^2 + 0.537 & .461^* .781 & .461^* .626 \\ .602^* .781 & .779^* .781 & .461^* .781 & .781^2 + 1.245 & .781^* .626 \\ .602^* .626 & .779^* .626 & .461^* .626 & .781^* .626 & .626^2 + 0.520 \end{bmatrix}$$

Other Identification Methods

- We could have picked models from the other two cells of our identification table
  - Standardized factor variance, estimated factor mean
    - All loadings estimated – one item intercept fixed to 0
  - Estimated factor variance, estimated factor mean
    - One item factor loading fixed to 1, one item intercept fixed to 0
- These methods are not typically used (but are equivalent)
- In fact, I did just that…

Standardized Factor Variance
Estimated Factor Mean

```
MODEL:
    GAMBLING by GRI1* GRI2-GRI5;
    [GRI1@0];
    [GAMBLING];
    GAMBLING@1;
```

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GAMBLING BY				
GRI1	1.000	0.000	999.000	999.000
GRI2	1.293	0.121	10.653	0.000
GRI3	0.765	0.077	9.959	0.000
GRI4	1.297	0.138	9.380	0.000
GRI5	1.040	0.094	11.042	0.000
Means				
GAMBLING	1.804	0.028	64.252	0.000
Intercepts				
GRI1	0.000	0.000	999.000	999.000
GRI2	-0.402	0.213	-1.888	0.059
GRI3	0.171	0.127	1.350	0.177
GRI4	-0.414	0.239	-1.730	0.084
GRI5	-0.299	0.157	-1.909	0.056
Variances				
GAMBLING	0.363	0.052	6.970	0.000
Residual Variances				
GRI1	0.666	0.057	11.716	0.000
GRI2	2.001	0.138	14.505	0.000
GRI3	0.537	0.040	13.402	0.000
GRI4	1.245	0.094	13.292	0.000
GRI5	0.520	0.047	11.050	0.000

MODEL FIT INFORMATION

Number of Free Parameters	15
Loglikelihood	
H0 Value	-9564.556
H0 Scaling Correction Factor for MLR	1.8591
H1 Value	-9547.913
H1 Scaling Correction Factor for MLR	1.8825

Model Estimated Means/Intercepts/Thresholds

GRI1	GRI2	GRI3	GRI4	GRI5
1.804	1.932	1.551	1.926	1.577

Model Estimated Covariances/Correlations/Residual Correlations

	GRI1	GRI2	GRI3	GRI4	GRI5
RI1	1.028				
RI2	0.469	2.608			
RI3	0.277	0.359	0.749		
RI4	0.471	0.609	0.360	1.856	
RI5	0.377	0.488	0.289	0.489	0.913

Estimated Factor Variance
Estimated Factor Mean

```
MODEL:
    GAMBLING by GRI1-GRI5;
    [GRI1@0];
    [GAMBLING];
```

MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GAMBLING BY				
GRI1	1.000	0.000	999.000	999.000
GRI2	1.293	0.121	10.653	0.000
GRI3	0.765	0.077	9.959	0.000
GRI4	1.297	0.138	9.380	0.000
GRI5	1.040	0.094	11.042	0.000
Means				
GAMBLING	1.804	0.028	64.252	0.000
Intercepts				
GRI1	0.000	0.000	999.000	999.000
GRI2	-0.402	0.213	-1.888	0.059
GRI3	0.171	0.127	1.350	0.177
GRI4	-0.414	0.239	-1.730	0.084
GRI5	-0.299	0.157	-1.909	0.056
Variances				
GAMBLING	0.363	0.052	6.970	0.000
Residual Variances				
GRI1	0.666	0.057	11.716	0.000
GRI2	2.001	0.138	14.505	0.000
GRI3	0.537	0.040	13.402	0.000
GRI4	1.245	0.094	13.292	0.000
GRI5	0.520	0.047	11.050	0.000

```
MODEL FIT INFORMATION

Number of Free Parameters                        15

Loglikelihood

          H0 Value                        -9564.556
          H0 Scaling Correction Factor       1.8591
            for MLR
          H1 Value                        -9547.913
          H1 Scaling Correction Factor       1.8825
            for MLR
```

Model Estimated Means/Intercepts/Thresholds

GRI1	GRI2	GRI3	GRI4	GRI5
1.804	1.932	1.551	1.926	1.577

Model Estimated Covariances/Correlations/Residual Correlations

GRI1	GRI2	GRI3	GRI4	GRI5
1.028				
0.469	2.608			
0.277	0.359	0.749		
0.471	0.609	0.360	1.856	
0.377	0.488	0.289	0.489	0.913

Picking an Identification Method

- The identification methods presented provide a mechanism to estimate the CFA model
  - Four alternatives (from table on previous slide)

- **All four methods provide identical fit – the log-likelihood and fit statistics are all identical!**
  - Choice of method depends upon analysis goals
  - We will stick with marker item technique

- Interpretation will be identical provided you use standardized estimates
  - The key in our future analysis

Factor Model Identification

- Goal: Reproduce observed covariance matrix among items with as few estimated parameters as possible
  - Maximum likelihood usually used to estimate model parameters
    - **Measurement Model**: Factor loadings, item intercepts, error variances
    - **Structural Model**: Factor variances and covariances, factor means
  - Global model fit is evaluated as difference between model-predicted matrix and observed matrix (but only the covariances really contribute)

- How many possible parameters can you estimate (total DF)?
  - **Total DF depends on # ITEMS →** p (NOT on # people)
  - Total number of 'unique elements' in covariance matrix
    - Unique elements = each variance, each covariance, each mean
    - Total unique elements = (p(p+1) / 2) + p → if 5 items, then ((5*6)/2) + 5 = 20
- Model degrees of freedom (df)
  - Model df = # possible parameters − # estimated parameters

CFA Model Identification: Two Options for Scaling the Factor

- Summary: two options for giving the factor a scale:
  - Marker item: Borrow a scale from one of the items
    - Fix that item's factor loading to 1 and its intercept to 0
    - Factor variance is based on part of that item that relates to factor (as opposed to part that is measurement error)
    - Will cause problems when marker item is unrelated to factor
  - Z-score: Put factor on scale of mean=0 and variance=1
    - All item factor loadings and all item intercepts are estimated
    - Can't be used in higher-order factor models or in models

- Most common approach is a hybrid:
    - Fix factor mean to 0, estimate all item intercepts → "z-score"
    - Estimate factor variance, fix first item factor loading to 1 → "marker"

- In reality, all methods of scaling the factor will fit equivalently well, so long as the marker item loads at all

Additional Factor Identification Issues

- Beyond setting the factor scale and location, one must also consider the minimum number of items that can be used for measurement of a factor

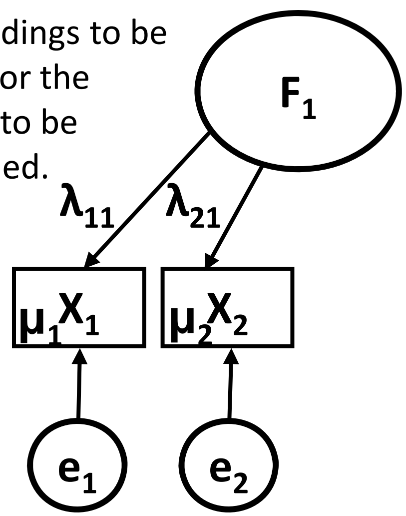
- The short story: factors need 3 or more items to be identified (but sometimes two will work)
  - 2 items – need additional factors/items in a model
  - 3 items – just identified one-factor model, perfect fit
  - 4 items – over-identified one-factor model, not perfect fit

- Next class we will discuss scale building with factor analysis – you will see that more items are usually better!

Under-Identified Factor: 2 Items

- Model is <u>under-identified</u> when there are more unknowns then pieces of information with which to estimate them
  - Cannot be solved because there are an infinite number of different parameter estimates that would result in perfect fit
  - Example: Solve x + y = 7 ??

You'd have to set the loadings to be equal for the model to be identified.



Total possible df = unique elements = **5**

0 factor variances		**1 factor variance**
0 factor means		**1 factor mean**
2 loadings	**OR**	**1 item loading**
2 item intercepts		**1 item intercept**
2 error variances		**2 error variances**

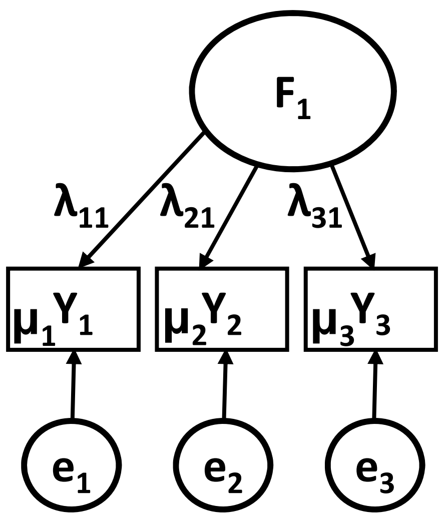
df = 5 – 6 = -1

If $r_{y1,y2}$ = .64, then:

λ_{11} = .800, λ_{21} = .800 ??
λ_{11} = .900, λ_{21} = .711 ??
λ_{11} = .750, λ_{21} = .853 ??

Just-Identified Factor: 3 Items

- Model is just-identified when there are as many unknowns as pieces of information with which to estimate them
  - Parameter estimates have a unique solution that will perfectly reproduce the observed matrix
  - Example: Solve x + y = 7, 3x – y = 1



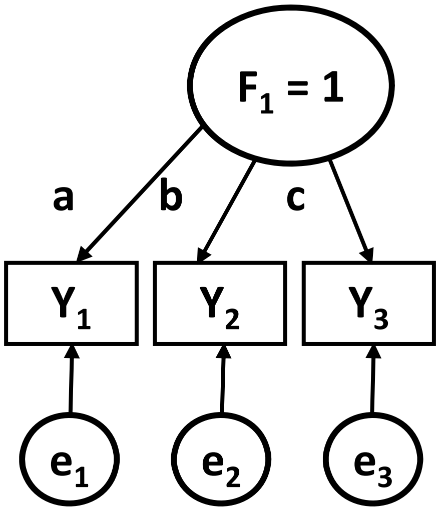
Total possible df = unique elements = **9**

0 factor variances		**1 factor variance**
0 factor means		**1 factor mean**
3 loadings	**OR**	**2 item loadings**
3 item intercepts		**2 item intercepts**
3 error variances		**3 error variances**

df = 9 – 9 = 0

Not really a model – more like a description

Solving a Just-Identified Model

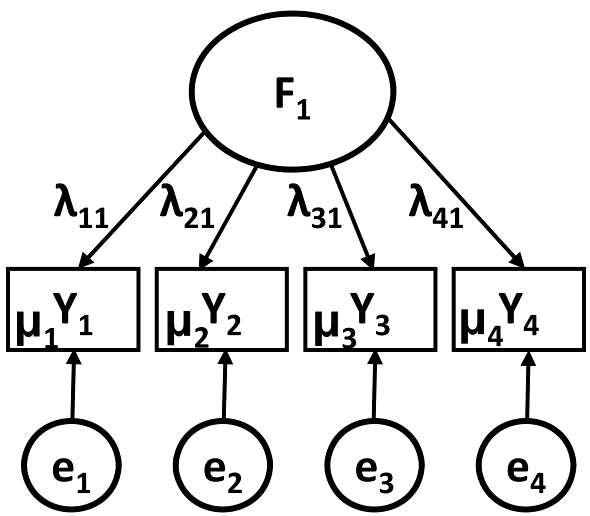


	Y_1	Y_2	Y_3
Y_1	1.00		
Y_2	.595	1.00	
Y_3	.448	.544	1.00

- Step 1: $ab = .595$
  $ac = .448$
  $bc = .544$
- Step 2: $b = .595/a$
  $c = .488/a$
  $(.595/a)(.448/a) = .544$
- Step 3: $.26656/a^2 = .544$
  $a = .70$
- Step 4: $.70b = .595 \quad b = .85$
  $.70c = .448 \quad c = .64$
- Step 5: $Var(e_1) = 1 - a^2 = .51$

Over-Identified Factor: 4+ Items

- Model is <u>over-identified</u> when there are fewer unknowns than pieces of information with which to estimate them
  - Parameter estimates have a unique solution that will NOT perfectly reproduce the observed matrix
  - **NOW we can test model fit**



Total possible df = unique elements = **14**

0 factor variances		**1 factor variance**
0 factor means		**1 factor mean**
4 loadings	**OR**	**3 item loadings**
4 item intercepts		**3 item intercepts**
4 error variances		**4 error variances**

df = 14 – 12 = 2

Did we do a 'good enough' job reproducing the matrix with 2 fewer parameters than was possible to use?

ASSESSMENT OF MODEL FIT IN CFA

Assessment of Model Fit in CFA

- As with path analysis (and general multivariate statistics using the multivariate normal distribution), if the saturated model covariance matrix is not closely approximated, inferences about the model will be biased (and likely, wrong)

- As the CFA model is covariance-specific, we must make sure a model adequately fits the data before reporting results and making inferences
  - Good news: model fit statistics are the same as in path analysis

- We discuss model fit here – after model specification – as it comes next in the use of CFA
  - We will use our marker item (zero factor mean) CFA model to show how these are calculated and interpreted

Indices of Global Model Fit

- Primary: obtained model **χ^2** (from Chi-Square Test of Model Fit section of Mplus output) – here we use the MLR rescaled **χ^2**
  - χ^2 is evaluated based on model df (difference in parameters between your CFA model and the saturated model)
  - Tests null hypothesis that your CFA model (H_0) fits equally to saturated model (H_1) so significance is undesirable (smaller χ^2, bigger p-value is better)
  - Just using χ^2 is insufficient, however:
    - Distribution doesn't behave like a true χ^2 if sample sizes are small or if items are non-normally distributed
    - Obtained χ^2 depends largely on sample size
    - Some mention this is an unreasonable null hypothesis (perfect fit??)
      - I believe it is not unreasonable
- Because of these issues, alternative measures of fit are usually used in conjunction with the χ^2 test of model fit
  - Absolute Fit Indices (besides χ^2)
  - Parsimony-Corrected; Comparative (Incremental) Fit Indices

Global Chi-Square from Our CFA Model

- From the Mplus output:

```
MODEL FIT INFORMATION

Number of Free Parameters                        15

Loglikelihood

          H0 Value                        -9564.556
          H0 Scaling Correction Factor       1.8591
            for MLR
          H1 Value                        -9547.913
          H1 Scaling Correction Factor       1.8825
            for MLR
```

- Here, we reject our model and state that the saturated model fits better

- Calculation:
  - 15 parameters in our model; 20 in saturated model
  - Scaling correction factor:

$$c_{LR} = \left|\frac{(q_{restricted})(c_{restricted}) - (q_{full})(c_{full})}{(q_{restricted} - q_{full})}\right| = \left|\frac{(15 \times 1.8591) - (20 \times 1.8825)}{15 - 20}\right|$$
$$= 1.9526$$

  - $\chi^2 = -2^*(-9564.556 - -9547.913) = \frac{33.285}{1.9526} = 17.047$
  - DF = 5

```
Chi-Square Test of Model Fit

          Value                              17.047*
          Degrees of Freedom                      5
          P-Value                            0.0044
          Scaling Correction Factor          1.9526
            for MLR
```

Indices of Global Model Fit

- Absolute Fit: $\boldsymbol{\chi^2}$ (from Chi-Square Test of Model Fit section of Mplus output)
  - Does not use 'ratio rules' like χ^2/df > 2 or χ^2/df > 3 (some do – you don't)
- Absolute Fit: **SRMR**
  - **Standardized Root Mean Square Residual**
  - Get difference of saturated covariance matrix and CFA model covariance matrix → residual matrix
  - Standardize the residual matrix (divide by product of standard deviations of respective variables)
  - Sum the squared residuals in matrix, divide by number of residuals summed
  - Ranges from 0 to 1: smaller is better
  - ".08 or less" → good fit
- See also: **RMR** (**Root Mean Square Residual**)

Indices of Global Model Fit

Parsimony-Corrected: **RMSEA**

- **Root Mean Square Error of Approximation**
- Uses comparison with CFA model and saturated model
  - χ^2 listed here is from Chi-Square Test of Model Fit section of Mplus output
- Relies on a non-centrality parameter (NCP)
  - Indexes how far off your model is → χ^2 distribution shoved over
  - NCP → $d = (\chi^2 - df) / (N-1)$ Then, RMSEA = SQRT(d/df)
    - df is difference between # parameters in CFA model and saturated model
  - RMSEA ranges from 0 to 1; smaller is better
  - < .05 or .06 = "good", .05 to .08 = "acceptable", .08 to .10 = "mediocre", and >.10 = "unacceptable"
  - In addition to point estimate, get 90% confidence interval
  - RMSEA penalizes for model complexity – it's discrepancy in fit per df left in model (but not sensitive to N, although CI can be)
  - Test of "close fit": null hypothesis that RMSEA ≤ .05

RMSEA from Our Example

- From Mplus:

```
MODEL FIT INFORMATION

Number of Free Parameters                        15

Loglikelihood

          H0 Value                        -9564.556
          H0 Scaling Correction Factor       1.8591
            for MLR
          H1 Value                        -9547.913
          H1 Scaling Correction Factor       1.8825
            for MLR
```

Find # of parameters in saturated model, calculate difference

Make -2*(difference in log-likelihoods)/c_{LR}

```
Chi-Square Test of Model Fit

          Value                              17.047*
          Degrees of Freedom                      5
          P-Value                            0.0044
          Scaling Correction Factor          1.9526
            for MLR
```

For RMSEA calculation: $\chi^2 = 17.047; df = 5$

```
SUMMARY OF ANALYSIS

Number of groups                                                 1
Number of observations                                        1304
```

Create Non-Centrality Parameter $d = \frac{\chi^2 - df}{N-1} = \frac{17.047-5}{1304-1} = 0.009$

```
RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.043
          90 Percent C.I.                    0.022  0.066
          Probability RMSEA <= .05           0.657
```

Calculate RMSEA: $\text{RMSEA} = \sqrt{\frac{d}{df}} = \sqrt{\frac{0.009}{5}} = 0.043$

Indices of Global Model Fit

Comparative (Incremental) Fit Indices

- Fit evaluated relative to a 'null' model (of 0 covariances)
- Relative to that, your model should be great!
- **CFI: Comparative Fit Index**
  - Also based on idea of NCP (χ^2 – df)
  - $CFI = 1 - \frac{\max(\chi^2_T - df_T, 0)}{\max(\chi^2_T - df_T, \chi^2_N - df_N, 0)}$
  - From 0 to 1: bigger is better, > .90 = "acceptable", > .95 = "good"

T = target model
N = null model

- **TLI: Tucker-Lewis Index** (= **Non-Normed Fit Index**)
  - $TLI = \frac{\frac{\chi^2_N}{df_N} - \frac{\chi^2_T}{df_T}}{\frac{\chi^2_N}{df_N} - 1}$
  - From <0 to >1, bigger is better, >.95 = "good"

Comparative Fit Index Calculation

The estimated CFA model (the target)

```
Number of Free Parameters                       15

Chi-Square Test of Model Fit

          Value                              17.047*
          Degrees of Freedom                      5
          P-Value                            0.0044
          Scaling Correction Factor          1.9526
            for MLR
```

The independence model (the null model)

```
Number of Free Parameters                       10

Chi-Square Test of Model Fit

          Value                             511.548*
          Degrees of Freedom                     10
          P-Value                            0.0000
          Scaling Correction Factor          1.9629
            for MLR
```

Compute numerator:

$$\max(\chi^2_T - df_T, 0)$$
$$= \max(17.047 - 5{,}0) = 12.047$$

Compute denominator:

$$\max(\chi^2_T - df_T, \chi^2_N - df_N, 0)$$
$$= \max(33.285 - 5{,}511.548 - 10{,}0)$$
$$= 501.549$$

Compute CFI:

$$CFI = 1 - \frac{\max(\chi^2_T - df_T, 0)}{\max(\chi^2_T - df_T, \chi^2_N - df_N, 0)}$$
$$= 1 - \frac{12.047}{501.549} = 0.976$$

```
CFI/TLI

          CFI                                 0.976
          TLI                                 0.952
```

Tucker-Lewis Index Calculation

The estimated CFA model (the target)

```
Number of Free Parameters                     15

Chi-Square Test of Model Fit

          Value                               17.047*
          Degrees of Freedom                       5
          P-Value                             0.0044
          Scaling Correction Factor           1.9526
            for MLR
```

The independence model (the null model)

```
Number of Free Parameters                     10

Chi-Square Test of Model Fit

          Value                              511.548*
          Degrees of Freedom                      10
          P-Value                             0.0000
          Scaling Correction Factor           1.9629
            for MLR
```

Compute Target Model ratio:

$$\frac{\chi_T^2}{df_T} = \frac{17.047}{5} = 3.4094$$

Compute Null Model Ratio:

$$\frac{\chi_N^2}{df_N} = \frac{511.548}{10} = 51.1548$$

Compute TLI:

$$TLI = \frac{\frac{\chi_N^2}{df_N} - \frac{\chi_T^2}{df_T}}{\frac{\chi_N^2}{df_N} - 1} = \frac{51.1548 - 3.4094}{51.1548 - 1} = 0.952$$

```
CFI/TLI

          CFI                                  0.976
          TLI                                  0.952
```

Building a Case for Model Fit

- Model fit statistics can vary widely: some will look like your model fits while others will not
  - In fact, the 1980s was the model-fit decade in SEM
  - Lots more indices of model fit are available!
- The best course of action is to ensure your model fits well under most of these indices
  - The more that are favorable – the more you can believe in your result – the better chance you will have to publish your findings
- Be sure to report all fit statistics – it gives the reader a better picture of your model

Evaluating Model Fit for our CFA model

- From Mplus:

```
Chi-Square Test of Model Fit

          Value                              17.047*
          Degrees of Freedom                      5
          P-Value                            0.0044
          Scaling Correction Factor          1.9526
            for MLR
```

Our model is rejected when compared with saturated model – no surprise

```
RMSEA (Root Mean Square Error Of Approximation)

          Estimate                           0.043
          90 Percent C.I.                    0.022  0.066
          Probability RMSEA <= .05           0.657
```

RMSEA indicates "good" model fit (<.05)

```
CFI/TLI

          CFI                                0.976
          TLI                                0.952
```

CFI indicates good model fit (> .95)
TLI indicates acceptable model fit (>.90)

```
SRMR (Standardized Root Mean Square Residual)

          Value                              0.024
```

Standardized root mean squared residuals (average "miss" for correlations) indicates good model fit

Based on these results, we will claim our model fit to be acceptable and move on to model interpretation

CFA MODEL PARAMETER INTERPRETATION

CFA Model Parameter Interpretation

- Give that (1) our model was identified and (2) our model fit well, we can now move onto (3) model interpretation

- In CFA, there are several sets of parameters:
  - Factor mean – usually set to zero – **not interpreted**
  - Factor variance – can be fixed or estimated, but in general is arbitrary – **not interpreted**
  - Item Intercepts – usually saturated (equal to item means) – **not interpreted**
  - Unique variances – depend on the scale of the item – **not interpreted directly**
    - but will be indirectly through other measures – i.e., item information and variance accounted for
  - Factor loadings – key statistics of the analysis – **heavily interpreted**
    - Typically through standardized coefficients

Interpretation of Factor Loadings

- Unstandardized item factor loadings are interpreted as you would regression slopes:

$$Y_{si} = \mu_i + \lambda_{i1} F_{s1} + e_{si}$$

- A one-unit increase in F_{s1} brings about a λ_{i1} increase in the predicted response Y_{si}
  - The ***covariance*** between the item and the factor

- The issue: the scale of the factor is arbitrary and depends on the identification method
  - The marker-item method of identification will produce different results from the standardized factor method of identification

Comparison of Results

The marker-item CFA model

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GAMBLING BY				
GRI1	1.000	0.000	999.000	999.000
GRI2	1.293	0.121	10.652	0.000
GRI3	0.765	0.077	9.959	0.000
GRI4	1.297	0.138	9.379	0.000
GRI5	1.040	0.094	11.042	0.000

The marker item $\hat{\lambda}_{11}$ indicates that a response to item 1 increases by 1 for every one-unit increase in F (but F has a variance of 0.363 – so a one-unit increase in F is nearly 1.5 SD)

The standardized factor CFA model

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GAMBLING BY				
GRI1	0.602	0.043	13.939	0.000
GRI2	0.779	0.055	14.086	0.000
GRI3	0.461	0.040	11.459	0.000
GRI4	0.781	0.050	15.529	0.000
GRI5	0.626	0.044	14.166	0.000

The standardized factor $\hat{\lambda}_{11}$ indicates that a response to item 1 increases by .602 for every one-unit increase in F

(F has a variance of 1– so a one-unit of F is nearly 1 SD)

Standardized Factor Loadings

- As in regression, the solution to comparing, interpreting, and reporting factor loadings is to use the standardized loadings
  - Standardized loading = $\lambda_i \frac{SD(F)}{SD(Y_i)}$
  - Represents the **correlation** between item and factor
  - Makes it so items with scales (variances) can be comparable
  - Essentially gives results that would be obtained by using z-scored variables
  - Standardized error variance = 1 – standardized λ_i^2 = "variance due to *not* factor"
  - R^2 for item = standardized λ_i^2 = "variance due to the factor"

More on Standardized Factor Loadings

- You will note that Mplus provides more than one version of standardization:
  - STDYX – use these for standardized coefficients
    - Here Y is your item and X is the factor
  - STDY – don't use (only standardizes by Y – good for dichotomous X)
  - STD – only standardizes by F – these are equivalent to the results from the standardized factor identification method
- The key is to **be clear** as to which standardized coefficients you are reporting because many exist
  - We will use the STDYX standardization

Interpreting Standardized Factor Loadings

The marker-item CFA model

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GAMBLING BY				
GRI1	0.594	0.035	17.067	0.000
GRI2	0.482	0.034	14.286	0.000
GRI3	0.532	0.036	14.895	0.000
GRI4	0.573	0.035	16.418	0.000
GRI5	0.656	0.033	19.877	0.000

The marker item $\hat{\lambda}_{11}$ indicates that a response to item 1 increases by .594 SD (of item 1) for every one-SD increase in F

The correlation between item 1 and the gambling factor is .594

The standardized factor CFA model

STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GAMBLING BY				
GRI1	0.594	0.035	17.070	0.000
GRI2	0.482	0.034	14.288	0.000
GRI3	0.532	0.036	14.896	0.000
GRI4	0.573	0.035	16.417	0.000
GRI5	0.656	0.033	19.879	0.000

The standardized factor $\hat{\lambda}_{11}$ indicates that a response to item 1 increases by .594 SD (of item 1) for every one-SD increase in F

The correlation between item 1 and the gambling factor is .594

The results are identical for both identification methods!

Item R^2 - Variance "Accounted For"

- With the standardized solution, we can also calculate the variance of each item accounted for by each factor
  - Take caution – factors do not exist!
  - The R^2 reports partitioned item variance
    - This is not EXPLAINED like in regression – only the part that the factor has in common with the item
  - The R^2 is equal to the squared standardized loading

R-SQUARE

Observed Variable	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
GRI1	0.353	0.041	8.535	0.000
GRI2	0.233	0.033	7.144	0.000
GRI3	0.283	0.038	7.448	0.000
GRI4	0.329	0.040	8.208	0.000
GRI5	0.430	0.043	9.939	0.000

CONCLUDING REMARKS

Wrapping Up

- The measurement model portion of structural equation modeling covers factor analysis
  - Each SEM with latent variables has a measurement model

- Today's class covered the details of factor analysis
  - How it works
  - Where terms come from

- As in path analysis we had:
  - A series of simultaneous regression equations
  - A need to establish an identified model
  - A need to make sure a model fit before interpreting results

Factor Analysis: Big Picture

- Unlike path analysis, our factor analysis regression equations contained latent (unobserved) variables
  - It is important to remember these **factors do not exist**
  - Don't factorpomorphize!

- This point is overlooked in measurement, psychometrics, and SEM analyses
  - There is an arbitrary nature to constructs/factors

- Items that do not "fit" are discarded
  - Sometimes they are bad items (poorly written)
  - Sometimes they do not measure the factor
  - Sometimes the factor model does not work

- What does exist are the items
  - The factors are culled from the covariances between items
  - CFA models are re-expressions of covariances into meaningful terms (factor loadings, unique variances)

- The methods of CFA are limited by the items – and their covariance matrix
  - Limited by the laws of statistics (sample size) and variability of estimates of covariances