

---

# **SEM with Non-Normal Data: Generalized Models**

Latent Trait Measurement and  
Structural Equation Models  
Lecture #12  
April 17, 2013

# Today's Class

---

- Data assumptions of SEM
  - Continuous data that are multivariate normal
- Types of data (and similar statistical distributions):
  - Discrete/categorical
  - Continuous
  - Mixture of discrete/categorical with continuous
- SEM with Non-normal Data:
  - Not continuous/not normal: Generalized models
- Issues with generalized models
  - Lack of standard fit tests
  - Estimation time
  - Differing estimators for some models

---

# **DATA ASSUMPTIONS OF SEM**

# Data Assumptions in CFA (to this point)

---

- Up to this point, we have been fixated on a measurement model that looks like this (for all items  $I$  and factors  $F$ ):

$$\mathbf{y}_s = \boldsymbol{\mu}_I + \boldsymbol{\Lambda}\mathbf{F}_s + \mathbf{e}_s$$

- Where:
  - $\mathbf{y}_s$  is the observed data (size  $I \times 1$ )
  - $\boldsymbol{\mu}_I$  is the item intercept vector (size  $I \times 1$ ) – a constant
  - $\boldsymbol{\Lambda}$  is the factor loading matrix (size  $I \times F$ ) – a set of constants, some of which are set to zero (or one)
  - $\mathbf{F}_s$  is the vector of factor scores for subject  $s$  (size  $F \times 1$ )
    - ♦  $\mathbf{F}_s \sim N_F(\mathbf{0}, \boldsymbol{\Phi})$  (factor scores are multivariate normal)
  - $\mathbf{e}_s$  is the vector of residuals for subject  $s$  (size  $I \times 1$ )
    - ♦  $\mathbf{e}_s \sim N_I(0, \boldsymbol{\Psi})$

# Putting the Linear Combination Together

---

- The CFA model is a linear combination of:
  - Constants:  $\boldsymbol{\mu}_I, \boldsymbol{\Lambda}$
  - Random components:  $\mathbf{F}_S, \mathbf{e}_S$
- As the random components are both multivariate normal, the resulting linear combination (and prediction  $\mathbf{y}_S$ ) is also multivariate normal
  - From properties of MVN distributions
- Therefore, the model assumes:
$$\mathbf{y}_S \sim N_I(\boldsymbol{\mu}_I, \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}^T + \boldsymbol{\Psi})$$
  - Our data are assumed to be multivariate normal

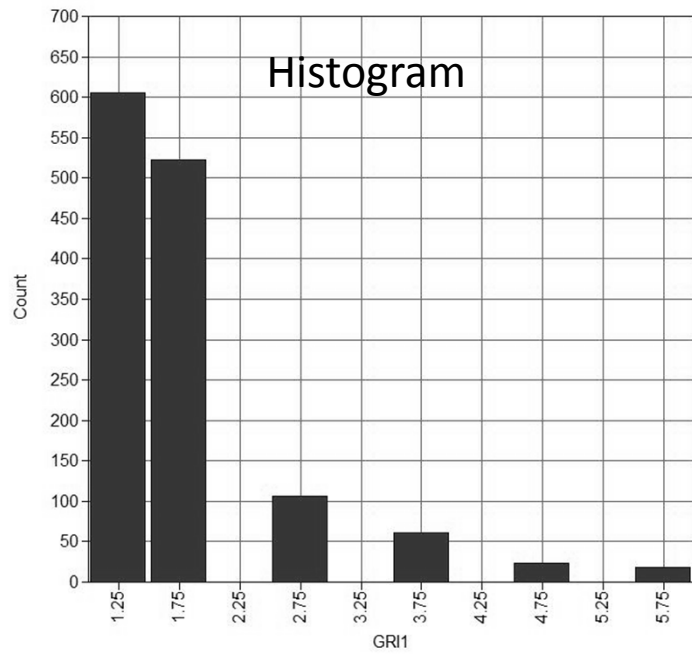
# Multivariate Normal Data

---

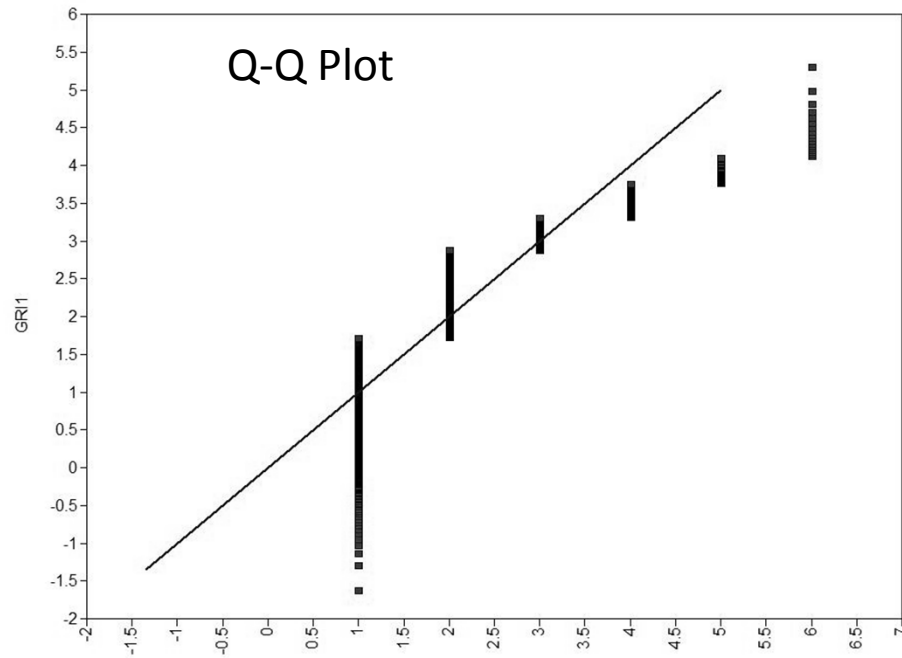
- For data to follow a multivariate normal distribution:
  - All of the items must be univariate normal
  - All of the items together must be multivariate normal
- We discovered how to check the items individually
  - Using Q-Q plots
- Let's re-examine our Gambling Data to see how the individual items look...

# Distributions Example GRI Item: Item 1

- For item 1: “I would like to cut back on my gambling”



illy d



# Impact of Non-normality of Item Responses

---

- Linear model predicting  $\mathbf{y}_s$  from  $\mathbf{F}_s$  may not work well
  - If  $\mathbf{y}_s$  is continuous, the line needs to be shut off at the ends
  - Predicted values of  $\mathbf{y}_s$  will quickly become impossible to observe – such as negative values
- Overall model  $\chi^2$  depends on MVN log-likelihood
  - Wrong log-likelihood means  $\chi^2$  will be incorrect
  - Direction of bias may be positive or may be negative
- Model parameter estimates will have incorrect standard errors
  - Standard errors depend on second derivative of log-likelihood function – if log-likelihood is incorrect, these will be, too
  - Direction of bias may be positive or may be negative



# Estimation/Model Fixes for Non-Normal Data

---

- In the previous years, one would transform the data so as to make them more normal
  - Usually a bad idea – can lose some effects
- Recent advances have delivered two acceptable fixes:
  1. Robust estimation for ML (still multivariate normal, but more realistic estimates)
  2. Generalized models (don't assume multivariate normal)
- The choice of which to use is difficult – so we will start with #1 – and assume our data are still continuous
  - Really – continuous enough

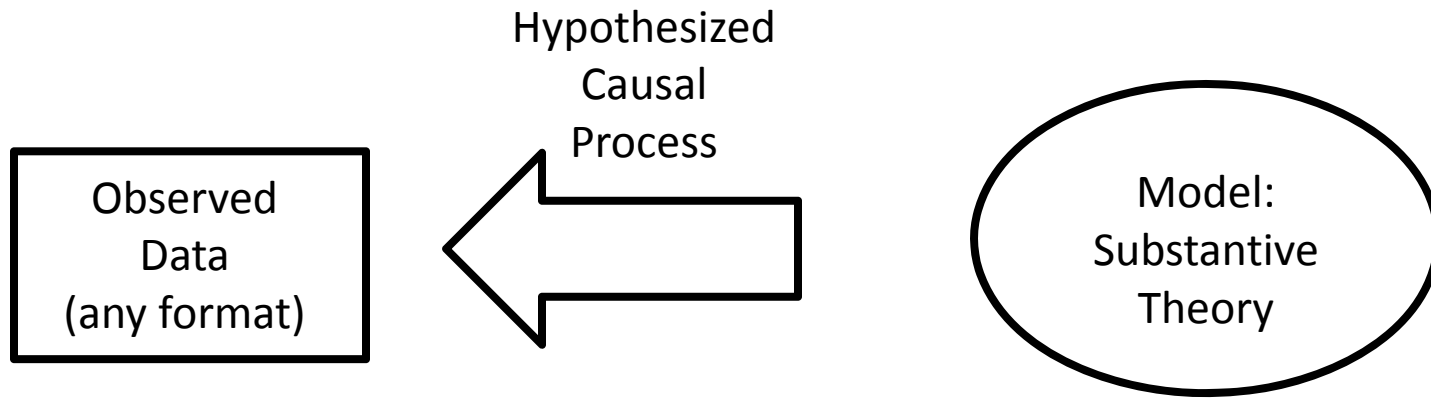
---

# GENERALIZED MODELS

# Generalized Models

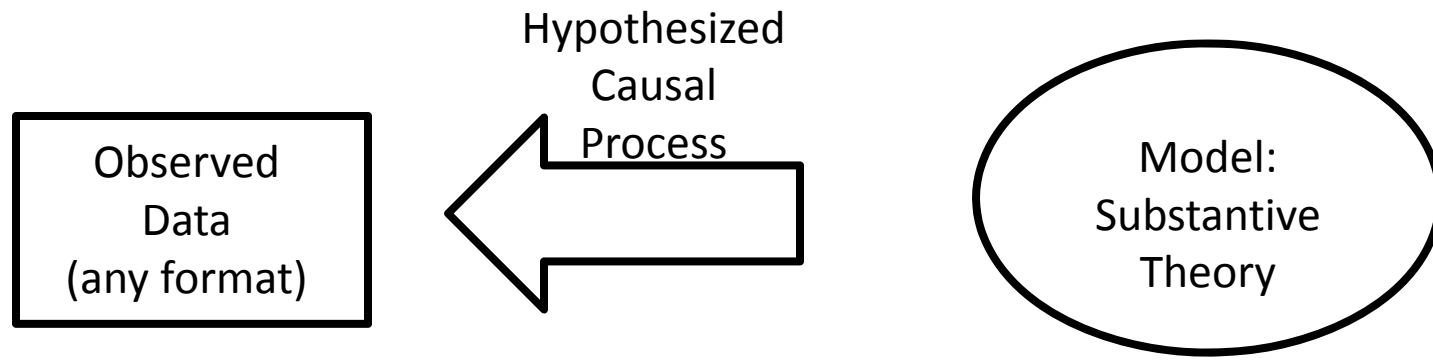
---

- Linear models with random effects (AKA latent variables) incorporates a very general set of statistical tools
  - We have only seen tools for use with continuous data that are multivariate normally distributed
- A bigger picture view of the modeling process sees what we know already as one small part



# Unpacking the Big Picture

---



- Substantive theory: what guides your study
  - Examples: one-factor of gambling tendencies; prediction of endogenous variables in path analysis...
- Hypothetical causal process: what the statistical model is testing when estimated
- Observed data: what you collect and evaluate based on your theory
  - Data can take many forms:
    - ♦ Continuous variables (e.g., time, blood pressure, height)
    - ♦ Categorical variables (e.g., likert-type responses, ordered categories, nominal categories)
    - ♦ Combinations of continuous and categorical (e.g., either 0 or some other continuous number)

# The Goal of Generalized Models

---

- Generalized models map the substantive theory onto the space of the observed data
  - Space = type/range/outcomes that are possible
  - Often called sample space in statistics
- The general idea is that the statistical model will not approximate the data well if the assumed distribution is not a good fit to the sample space of the data
- The key to making all of this work is the use of differing statistical distributions

# The Basics of Statistical Distributions

---

- Statistical distributions are functions that describe the probability of a random variable taking certain values
  - In the case of generalized models, we are finding the “right” distribution for our data (the random variables of interest)
- Statistical distributions can be categorized into three classes:
  - Completely continuous
  - Completely categorical (also called discrete)
  - Mixtures of continuous and categorical
- Distributions are defined on a sample space – the range of values random variables can take
  - Univariate normal distribution:  $(-\infty, \infty)$  – all real numbers
  - Chi-squared distribution:  $[0, \infty)$  – all positive numbers
  - Bernoulli distribution:  $\{0,1\}$  – binary digits

# More on Distributions

---

- A statistical distribution has the property that the sum (for categorical) or integral (for continuous) of the distribution equals one across the sample space
  - Subsets of these define the probability of values occurring
- An infinite number of distributions exist – and almost any can be used in generalized models
  - You may have to build your own estimator, though
- More commonly, generalized models allow you to pick from a handful of families of distributions
  - We will stick with what Mplus gives us
- In modern statistical analysis, multiple distributions can be used for different items/variables in an analysis
  - Not every item or variable has to follow one distribution

# Link Functions: How Generalized Models Work

---

- Generalized models work by providing a mapping of the theoretical portion of the model (the right hand side of the equation) to the sample space of the data (the left hand side of the equation)
  - The mapping is done by a feature called a link function
- The link function is a non-linear function that takes the linear model predictors, random/latent terms, and constants and puts them onto the space of the outcome observed variables
- Link functions are typically expressed for the mean of the outcome variable (I will only focus on that)
  - In generalized models, the variance is often a function of the mean



# Link Functions in Practice

---

- The link function expresses the value of the mean of the outcome  $E(y_{si}) = \mu_y$  (E stands for expectation)...
- ...through a (typically) non-linear function  $g(\cdot)$  (when used on the mean; or its inverse  $g^{-1}(\cdot)$  when used on the predictors...
- ...of the observed predictors (and their regression weights)  $\mathbf{X}_s\boldsymbol{\beta}$  ...
- ...and of the random/latent predictors (and their observed or estimated weights – think factor loadings)  $\mathbf{Z}_s\boldsymbol{\Gamma}_s$ ...

$$E(y_{si}) = \mu_y = g^{-1}(\mathbf{X}_s\boldsymbol{\beta} + \mathbf{Z}_s\boldsymbol{\Gamma}_s)$$

- The term  $\mathbf{X}_s\boldsymbol{\beta} + \mathbf{Z}_s\boldsymbol{\Gamma}_s$  is called the **linear predictor**
  - Within the function, the values are linear combinations

# CFA in a Generalized Model Context

---

- Our familiar CFA model is a member of the generalized linear model family
  - The link function is called the identity – it is what it is!
- We knew from before that the expected value of an item from the CFA model is given by:

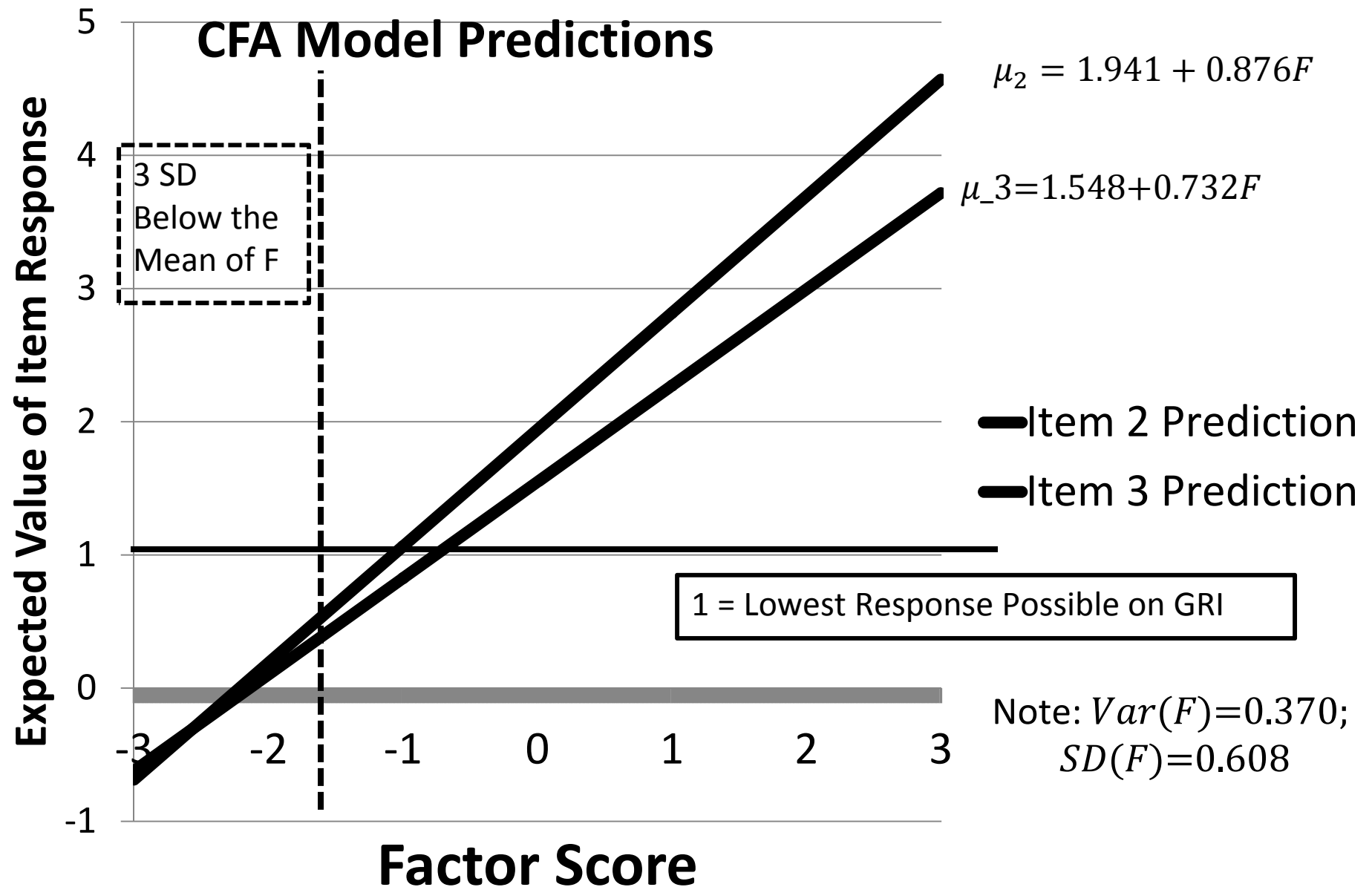
$$E(y_{si}) = \mu_y = g^{-1}(\mu_{I_i} + \Lambda_i \mathbf{F}_s) = \mu_{I_i} + \Lambda_i \mathbf{F}_s$$

- Here, the inverse link function is the identity

$$g^{-1}(\cdot) = I(\cdot)$$

- The identity does not alter the predicted values – they can be any real number
- This matches the sample space of the normal distribution

## CFA Model Mean Response Predictions – Using Estimates from 24 Item Analysis



## CFA: What About the Variance of the Item?

---

- The variance of the item (observed outcome data) in a CFA model is given by the estimated unique variance in the model  $\psi_i^2$
- In generalized models, the variance term is often found as a function of the mean (more on that shortly)
- But in this case, we can say:

$$Var(y_{si}) = V(\mu_y) = V\left(g^{-1}(\mu_{I_i} + \mathbf{\Lambda}_i \mathbf{F}_s)\right) = \sigma_y^2 = \psi_i^2$$

# Putting the Expected Mean/Variance Into a Distribution

---

- In CFA, we assume our observed data are normally distributed, which means the statistical distribution (conditional on the factor score) for the item is given by:

$$f(y_{si}|\mathbf{F}_s) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(y_{si} - \mu_y)^2}{2\sigma_y^2}\right)$$

- Plugging in our model expression for the mean and variance gives:

$$f(y_{si}|\mathbf{F}_s) = \frac{1}{\sqrt{2\pi\psi_i^2}} \exp\left(-\frac{(y_{si} - (\mu_{I_i} + \mathbf{\Lambda}_i\mathbf{F}_s))^2}{2\psi_i^2}\right)$$

# Where This Is Going...

---

- In order to explain several key concepts about generalized models, we are going to work through them using our CFA model (identity link; normally distributed data)
- Of importance in comparing GLMM to what we know:
  - Estimation is more complicated (not quite impossible)
  - Evaluation of model fit is more complicated (virtually impossible)
- With CFA (and an identity link), we have a normal distribution assumed for items
  - The normal distribution has two parameters:  $\mu_y, \sigma_y^2$
  - The CFA model makes predictions about these parameters
- In the rest of the generalized models a similar process holds
  - Each statistical distribution has a set of parameters
  - The model makes predictions about the parameters

# Generalized Linear Mixed Models

---

- The overarching name used for linear models with differing types of outcomes (data) and different types of predictors (observed and random/latent variables) is generalized linear mixed models
- This comes from the progression of statistics:
  - Linear models: regression (continuous predictors) w/o random/latent predictors for predicting continuous outcome
  - General linear models: ANOVA (categorical predictors) and regression (continuous predictors) w/o random/latent predictors for predicting continuous outcome
  - Generalized linear models: ANOVA (categorical predictors) and regression (continuous predictors) w/o random/latent predictors for predicting different types of outcomes
  - Generalized linear mixed models: ANOVA (categorical predictors) and regression (continuous predictors) **with** random/latent predictors for predicting different types of outcomes

---

# **MARGINAL ML ESTIMATION OF GENERALIZED LINEAR MIXED MODELS**



# Moving from Marginal (One Item) Distributions to Joint (All Items)

---

- In order to estimate the model parameters, we need the joint distribution of all of our observed data  $f(\mathbf{y}_s)$ 
  - This joint distribution cannot have any random/latent terms
  - It is just for **all** of the observed data
- At the item level, we have the **conditional** distribution of an item response given our random/latent term (the factor score):  $f(y_{si}|\mathbf{F}_s)$
- To get to the joint distribution of the observed data we must go through a series of steps (these are common across GLMMs)
  1. We must first aggregate across all conditional distributions of items to form the joint conditional distribution of all the data  $f(\mathbf{y}_s|\mathbf{F}_s)$ 
    - ◆ Still conditional on the random/latent terms
  2. We must then **marginalize** (remove) the random/latent term from the conditional distribution in order to get to the joint distribution of the data  $f(\mathbf{y}_s)$

## Step #1: The Joint Conditional Distribution

---

The joint conditional distribution comes from the individual distributions of all of the item responses:

$$f(\mathbf{y}_s | \mathbf{F}_s) = \prod_{i=1}^I f(y_{si} | \mathbf{F}_s)$$

This is built from the assumption of item responses being independent given the factor scores (conditional independence) – and gives us the product

Specifically for our data (with a normal distribution) this is:

$$f(\mathbf{y}_s | \mathbf{F}_s) = \prod_{i=1}^I \frac{1}{\sqrt{2\pi\psi_i^2}} \exp\left(-\frac{(y_{si} - (\mu_{I_i} + \mathbf{\Lambda}_i \mathbf{F}_s))^2}{2\psi_i^2}\right)$$

## Pre-Step #2...Mathematical Statistics

---

- To get to the joint distribution of just the data, we must **marginalize** across the random/latent terms
  - Before we do that, a primer on statistics is in order
- The joint (bivariate) distribution is written as  $f(X, Y)$
- The marginal distributions are written as  $f(X)$  and  $f(Y)$
- Depending on the type of random variable (continuous or discrete) marginal distribution comes from integrating or summing the joint distribution across the sample space of the other variable:

$$f(X) = \int_Y f(X, Y) dY \text{ - continuous}$$

$$f(X) = \sum_Y f(X, Y) \text{ - discrete}$$

# Conditional Distributions

---

- For two random variables  $X$  and  $Y$ , a conditional distribution is written as:  $f(X|Y)$ 
  - The distribution of  $X$  given  $Y$
- The conditional distribution is also equal to the joint distribution divided by the marginal distribution of the conditioning random variable

$$f(X|Y) = \frac{f(X, Y)}{f(Y)}$$

- To get to the marginal (where we need to go) from the conditional (what we have), we have to first get to the joint distribution:

$$f(X) = \int_Y f(X|Y)f(Y) dY = \int_Y \frac{f(X, Y)}{f(Y)} f(Y) dY = \int_Y f(X, Y) dY$$

- This is what we will use to get the distribution we are after

## Step #2: Marginalizing Across the Random/Latent Terms

---

The joint marginal distribution of the data  $f(\mathbf{y}_s)$  is derived from the same process detailed on the two previous slides:

$$f(\mathbf{y}_s) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(\mathbf{y}_s | \mathbf{F}_s) f(\mathbf{F}_s) dF_F \dots dF_2 dF_1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left[ \prod_{i=1}^I f(y_{si} | \mathbf{F}_s) \right] f(\mathbf{F}_s) dF_F \dots dF_2 dF_1$$

*Note: if there is more than one random/latent term, there is more than one integral...one for every random/latent term  $F_1, F_2, \dots F_F$*

Regardless of the type of item – this marginalization is the same in a GLMM with continuous random/latent terms  
We used it in CFA...as we will see

# “Marginal” ML Estimation

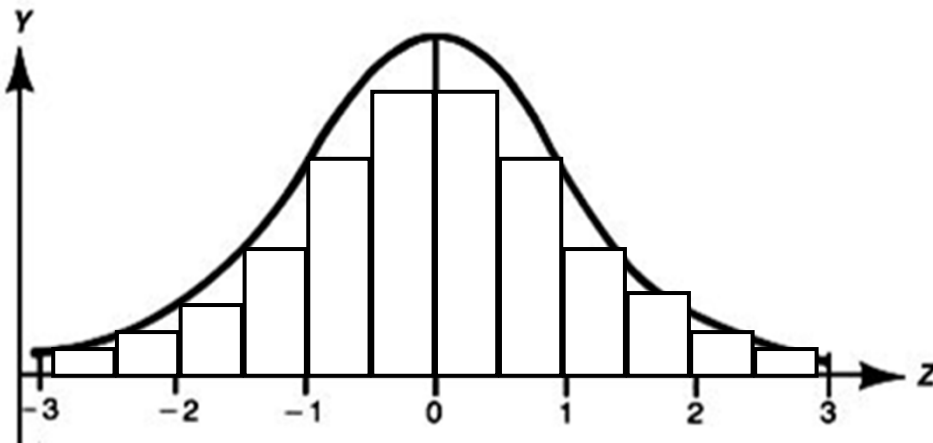
---

- How integration works, computationally:

Divide the distribution into rectangles

“Gaussian Quadrature” (# rectangles = # “quadrature points”)

- You can either divide the whole distribution into rectangles, or take the most likely section for each person and rectangle that
  - ♦ This is “adaptive quadrature” and is computationally more demanding, but gives more accurate results with fewer rectangles



The likelihood of each person’s observed data at each value of the random/latent term rectangle is then weighted by that rectangle’s probability of being observed (as given by the normal distribution). The weighted likelihoods are then added together across all rectangles.

# Distribution of Random/Latent Terms

---

- Central to the marginalization is the distribution of random/latent terms  $f(\mathbf{F}_s)$ 
  - These are typically assumed to be continuous and normally distributed
- In most GLMMs, these follow a MVN distribution
  - Latent class models and Diagnostic Classification Models use different (categorical) distributions
- The mean of the random/latent terms ( $\boldsymbol{\mu}_F$ ) is usually set to zero, and the covariance matrix ( $\boldsymbol{\Phi}$ ) is estimated:

$$f(\mathbf{F}_s) = \frac{1}{(2\pi)^{\frac{p}{2}} |\boldsymbol{\Phi}|^{\frac{1}{2}}} \exp \left[ -\frac{(\mathbf{F}_s^T - \boldsymbol{\mu}_F)^T \boldsymbol{\Phi}^{-1} (\mathbf{F}_s^T - \boldsymbol{\mu}_F)}{2} \right]$$

## Putting CFA Into The GLMM Context

---

From previous slides, we found the conditional distribution in CFA to be:

$$f(\mathbf{y}_s | \mathbf{F}_s) = \prod_{i=1}^I \frac{1}{\sqrt{2\pi\psi_i^2}} \exp\left(-\frac{(y_{si} - (\mu_{I_i} + \mathbf{\Lambda}_i \mathbf{F}_s))^2}{2\psi_i^2}\right)$$

We also found the distribution of the latent factors to be:

$$f(\mathbf{F}_s) = \frac{1}{(2\pi)^{\frac{p}{2}} |\mathbf{\Phi}|^{\frac{1}{2}}} \exp\left[-\frac{(\mathbf{F}_s^T - \boldsymbol{\mu}_F)^T \mathbf{\Phi}^{-1} (\mathbf{F}_s^T - \boldsymbol{\mu}_F)}{2}\right]$$



# Putting CFA Into The GLMM Context

---

Putting these together, we get:

$$\begin{aligned}
 f(\mathbf{y}_S) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(\mathbf{y}_S | \mathbf{F}_S) f(\mathbf{F}_S) dF_F \dots dF_2 dF_1 = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left[ \prod_{i=1}^I f(y_{si} | \mathbf{F}_S) \right] f(\mathbf{F}_S) dF_F \dots dF_2 dF_1 \\
 &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \left[ \prod_{i=1}^I \frac{1}{\sqrt{2\pi\psi_i^2}} \exp \left( -\frac{(y_{si} - (\mu_{I_i} + \mathbf{\Lambda}_i \mathbf{F}_S))^2}{2\psi_i^2} \right) \right] \times \\
 &\quad \frac{1}{(2\pi)^{\frac{p}{2}} |\mathbf{\Phi}|^{\frac{1}{2}}} \exp \left[ -\frac{(\mathbf{F}_S^T - \boldsymbol{\mu}_F)^T \mathbf{\Phi}^{-1} (\mathbf{F}_S^T - \boldsymbol{\mu}_F)}{2} \right] dF_F \dots dF_2 dF_1
 \end{aligned}$$

OMFG!

# CFA Relies on MVN Properties to Simplify

---

- The monstrous equation from the last slide has an easier version – all due to properties of MVN distributions
  - Conditional distributions of MVN are also MVN
  - Marginal distributions of MVNs are also MVN
- Therefore, we can show that for CFA (under identification where the factor mean is zero), the last slide becomes:

$$f(\mathbf{y}_s) = \frac{1}{(2\pi)^{\frac{p}{2}} |\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}^T + \mathbf{\Psi}|^{\frac{1}{2}}} \exp \left[ -\frac{(\mathbf{y}_s^T - \boldsymbol{\mu}_I)^T (\mathbf{\Lambda}\mathbf{\Phi}\mathbf{\Lambda}^T + \mathbf{\Psi})^{-1} (\mathbf{y}_s^T - \boldsymbol{\mu}_I)}{2} \right]$$

# What All This Means

---

- The integrals in the non-specific GLMM are difficult to estimate computationally
  - They take a long time – and get approximated
  - CFA doesn't have them because of the MVN distribution
- Model fit is based on the joint distribution of the data  $f(\mathbf{y}_s)$ , across all subjects  $s$ , or  $f(\mathbf{Y})$ 
  - In general, this is difficult to impossible to figure out for differing distributions in the GLMM
  - CFA doesn't have this problem as the joint distribution is MVN
- Therefore, two fundamental aspects of CFA don't map well onto GLMMs
  - Easy estimation
  - Relatively easy model fit determination

---

## **TYPES OF GLMMS – SEM FOR DIFFERENT RESPONSE TYPES**

# Links/Distributions (from Wikipedia)

Common distributions with typical uses and canonical link functions

Distribution	Support of distribution	Typical uses	Link name	Link function	Mean function
Normal	real: $(-\infty, +\infty)$	Linear-response data	Identity	$\mathbf{X}\beta = \mu$	$\mu = \mathbf{X}\beta$
Exponential	real: $(0, +\infty)$	Exponential-response data, scale parameters	Inverse	$\mathbf{X}\beta = \mu^{-1}$	$\mu = (\mathbf{X}\beta)^{-1}$
Gamma					
Inverse Gaussian			Inverse squared	$\mathbf{X}\beta = \mu^{-2}$	$\mu = (\mathbf{X}\beta)^{-1/2}$
Poisson	integer: $[0, +\infty)$	count of occurrences in fixed amount of time/space	Log	$\mathbf{X}\beta = \ln(\mu)$	$\mu = \exp(\mathbf{X}\beta)$
Bernoulli	integer: $[0, 1]$	outcome of single yes/no occurrence	Logit	$\mathbf{X}\beta = \ln\left(\frac{\mu}{1-\mu}\right)$	$\mu = \frac{\exp(\mathbf{X}\beta)}{1 + \exp(\mathbf{X}\beta)} = \frac{1}{1 + \exp(-\mathbf{X}\beta)}$
Binomial	integer: $[0, N]$	count of # of "yes" occurrences out of N yes/no occurrences			
Categorical	integer: $[0, K)$	outcome of single K-way occurrence			
	K-vector of integer: $[0, 1]$ , where exactly one element in the vector has the value 1				
Multinomial	K-vector of integer: $[0, N]$	count of occurrences of different types (1 .. K) out of N total K-way occurrences			

# Which Distribution/Link Do I Choose?

---

- The choice of distribution and link function can be difficult to decide – but there are some general guidelines
- If you have discrete (categorical) variables and they have:
  - 2-3 options: use a categorical distribution/link (usually logit – stay tuned)
  - 4-5 options: consider a categorical distribution/link OR a continuous distribution/link (typically CFA)
  - More than 5 options: use a continuous distribution/link (CFA)
- If you have censored data – use a censored model
- If you have count data that are bunched toward zero – use a model for counts
- In general, the normal distribution/identity link is a good approximation to most cases
  - Except those above

# Why Review Generalized Models?

---

- We will begin our GLMM discussion with the case of items that are categorical (discrete) and only have a few options
- These items treat their options as being ordered
  - Higher response = more of the trait measured
  - Question is how big is the difference between one category and the next?

<b>Outcome Type → <i>Model Family</i></b>	<b>Observed X</b>	<b>Latent X</b>
<b>Continuous Y → <i>“General Linear Model”</i></b>	Linear Regression	Confirmatory Factor Models
<b>Discrete Y → <i>“Generalized Linear Model”</i></b>	Logistic Regression	Item Response Models

# Means and Variances by Item Type

---

- Means:

- Quantitative item mean  $\rightarrow \sum_{p=1}^N \frac{Y_{ip}}{N} = \bar{Y}_i = \mu_{Y_i}$
- Binary item mean  $\rightarrow \sum_{p=1}^N \frac{Y_{ip}}{N} = p_{Y_i}$

- Variances:

- Quantitative item:  $\text{Var}(Y) = \frac{\sum_{i=1}^N (Y_i - \bar{Y}_i)^2}{N}$
- Binary item:  $\text{Var}(Y) = p_{Y_i}(1 - p_{Y_i}) = p_{Y_i}q_{Y_i} = \sigma_{Y_i}^2$
- With 2 options, the variance IS determined by the mean ( $p_y$ )**

TABLE 3.2  
Binary Item Variance and Difficulty

p	.0	.1	.2	.3	.4	.5	.6	.7	.8	.9	1.0
variance	.0	.09	.16	.21	.24	.25	.24	.21	.16	.09	.0



# A General Linear Model Predicting Binary Outcomes?

---

- If Y is a binary (0 or 1) outcome...
  - Expected mean is proportion of people who have a 1 (or “p”, the probability of Y=1)
    - ♦ The probability of having a 1 is what we’re trying to predict for each person, given the values on the predictors
    - ♦ General linear model:  $Y_s = \beta_0 + \beta_1 x_s + \beta_2 z_s + e_s$ 
      - $\beta_0$  = expected probability when all predictors are 0
      - $\beta$ ’s = expected change in probability for a one-unit change in the predictor
      - $e_s$  = difference between observed and predicted values
  - Model becomes  $Y_s = (\text{predicted probability of 1}) + e_s$

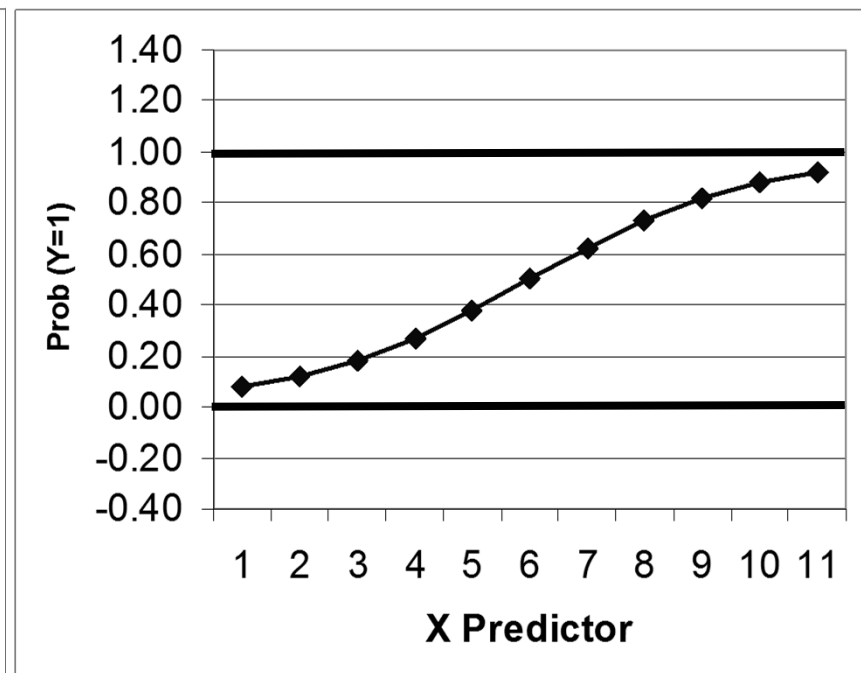
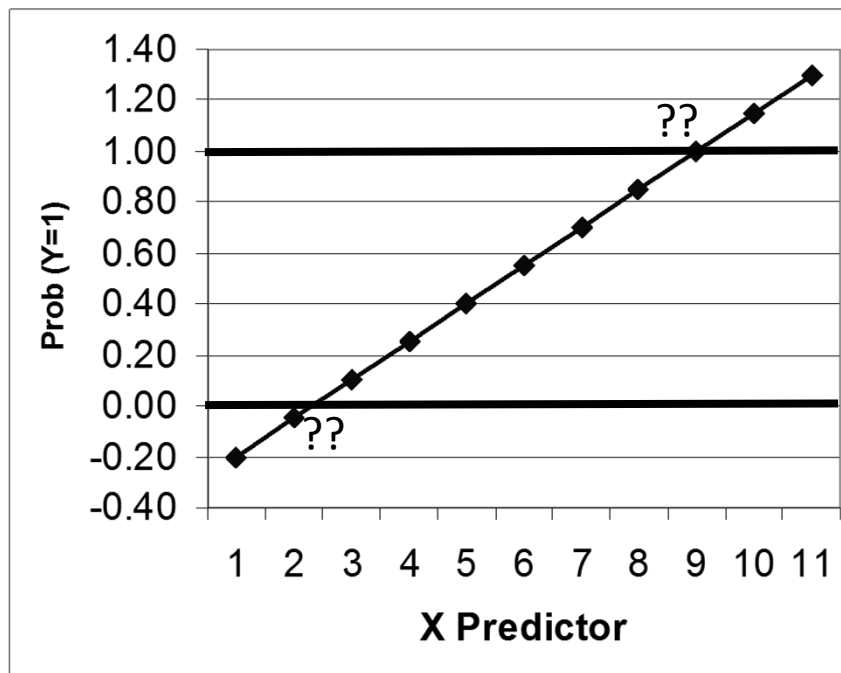
## A General Linear Model Predicting Binary Outcomes?

---

- But if  $Y_s$  is binary, then  $e_s$  can only be 2 things:
  - $e_s = \text{Observed } Y_s \text{ minus Predicted } Y_s$ 
    - ♦ If  $Y_s = 0$  then  $e = (0 - \text{predicted probability})$
    - ♦ If  $Y_s = 1$  then  $e = (1 - \text{predicted probability})$
  - Mean of errors would still be 0...
  - But variance of errors can't possibly be constant over levels of  $X$  like we assume in general linear models
    - ♦ The mean and variance of a binary outcome are **dependent!**
    - ♦ This means that because the conditional mean of  $Y$  ( $p$ , the predicted probability  $Y=1$ ) is dependent on  $X$ ,  
*then so is the error variance*

# A General Linear Model With Binary Outcomes?

- How can we have a linear relationship between X & Y?
- Probability of a 1 is bounded between 0 and 1, but predicted probabilities from a linear model aren't bounded
  - Impossible values
- Linear relationship needs to 'shut off' somehow → made nonlinear



### 3 Problems with General\* Linear Models Predicting Binary Outcomes

---

- \*General = model for continuous, normal outcome
- Restricted range (e.g., 0 to 1 for binary item)
  - Predictors should not be linearly related to observed outcome  
→ Effects of predictors need to be 'shut off' at some point to keep predicted values of binary outcome within range
- Variance is dependent on the mean, and not estimated
  - Fixed (→predicted value) and random (error) parts are related  
→ So residuals can't have constant variance
- Residuals have a limited number of possible values
  - Predicted values can each only be off in two ways  
→ So residuals can't be normally distributed

# Generalized vs. General Models

---

- Generalized Linear Models → General Linear Models with funky error terms and transformed Ys to obtain some kind of continuous outcome to work with
  - Many kinds of non-normally distributed outcomes have some kind of generalized linear model to go with them (and all of these are available within Mplus very easily):
    - Binary (dichotomous)
    - Unordered categorical (nominal)
    - Ordered categorical (ordinal)
    - Counts (discrete, positive values)
    - Censored (piled up and cut off at one end – left or right)
    - Zero-inflated (pile of 0's, then some distribution after)
- These two are often called “multinomial” inconsistently

# 3 Parts of a Generalized Linear Model

---

- Link Function (main difference from GLM):
  - How a non-normal outcome gets transformed into something we can predict that is more continuous (unbounded)
  - For outcomes that are already normal, general linear models are just a special case with an “identity” link function ( $Y * 1$ )
- Model for the Means (“Structural Model”):
  - How predictors linearly relate to the transformed outcome
  - New transformed  $Y_s = \beta_0 + \beta_1 x_s + \beta_2 z_s$
- Model for the Variance (“Sampling/Stochastic Model”):
  - If the errors aren’t normal and homoscedastic, what are they?
  - Family of alternative distributions at our disposal that map onto what the distribution of errors could possibly look like

# The Binary Case: Bernoulli Distribution

---

For items that are binary (dichotomous/two options), a frequent distribution chosen is the Bernoulli distribution:

(note:  $s$  is subscript for subject;  $i$  is subscript for item)

**Notation:**  $y_{si} \sim B(p_{si})$  (where  $p_{si}$  is the probability of a 1)

**Probability distribution function:**

$$f(y_{si}) = P(y_{si} = c) = (p_{si})^{y_{si}}(1 - p_{si})^{1-y_{si}}$$

**Expected value (mean) of  $y$ :**  $E(y_{si}) = \mu_{y_{si}} = p_{si}$

**Variance of  $y$ :**  $Var(y_{si}) = \sigma_{y_{si}}^2 = p_{si}(1 - p_{si})$

Note:  $p_i$  is the only parameter – so we only need to provide a link function for it...

# Generalized Models for Binary Outcomes

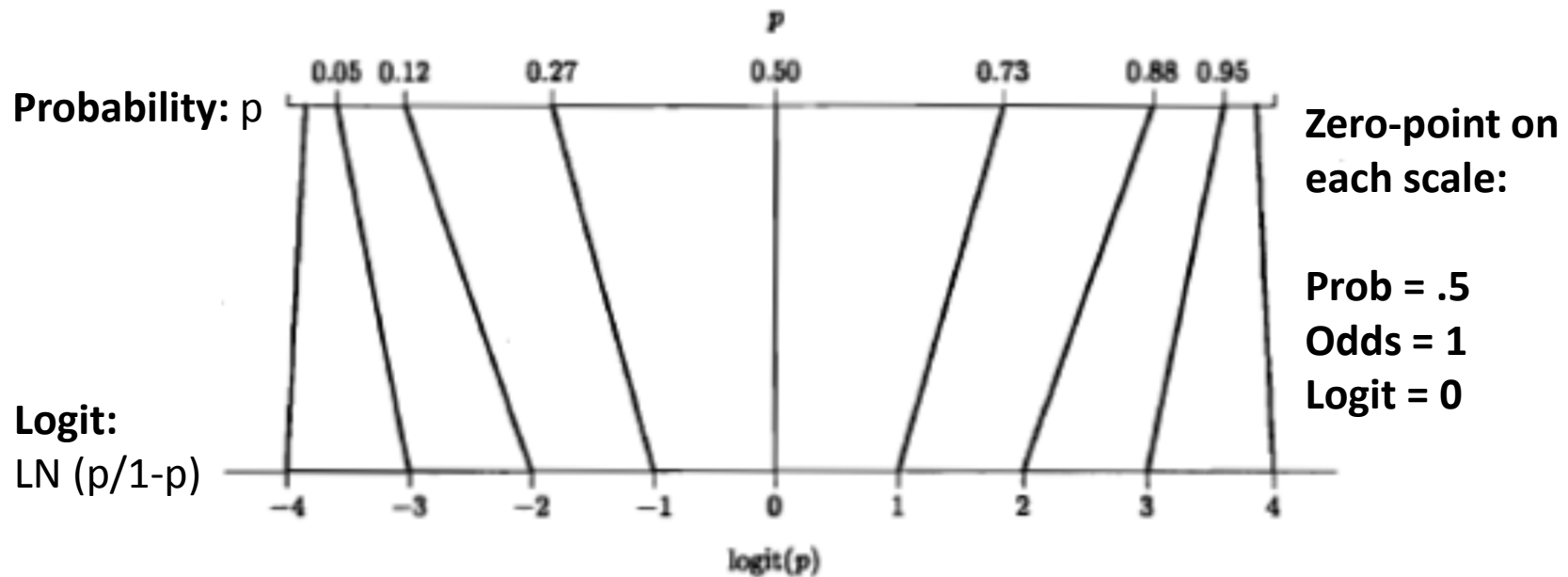
---

- Rather than modeling the probability of a 1 directly, we need to transform it into a more continuous outcome with a link function, for example:
  - We could transform probability into an odds ratio:
    - ♦ Odds ratio:  $(p / 1-p) \rightarrow \text{prob}(1) / \text{prob}(0)$
    - ♦ If  $p = .7$ , then  $\text{Odds}(1) = 2.33$ ;  $\text{Odds}(0) = .429$
    - ♦ Odds scale is way skewed, asymmetric, and ranges from 0 to  $+\infty$ 
      - Nope, that's not helpful
  - Take natural log of odds ratio  $\rightarrow$  called “logit” link
    - ♦  $\text{LN}(p / 1-p) \rightarrow \text{Natural log of } (\text{prob}(1) / \text{prob}(0))$
    - ♦ If  $p = .7$ , then  $\text{LN}(\text{Odds}(1)) = .846$ ;  $\text{LN}(\text{Odds}(0)) = -.846$
    - ♦ Logit scale is now symmetric about 0  $\rightarrow$  DING

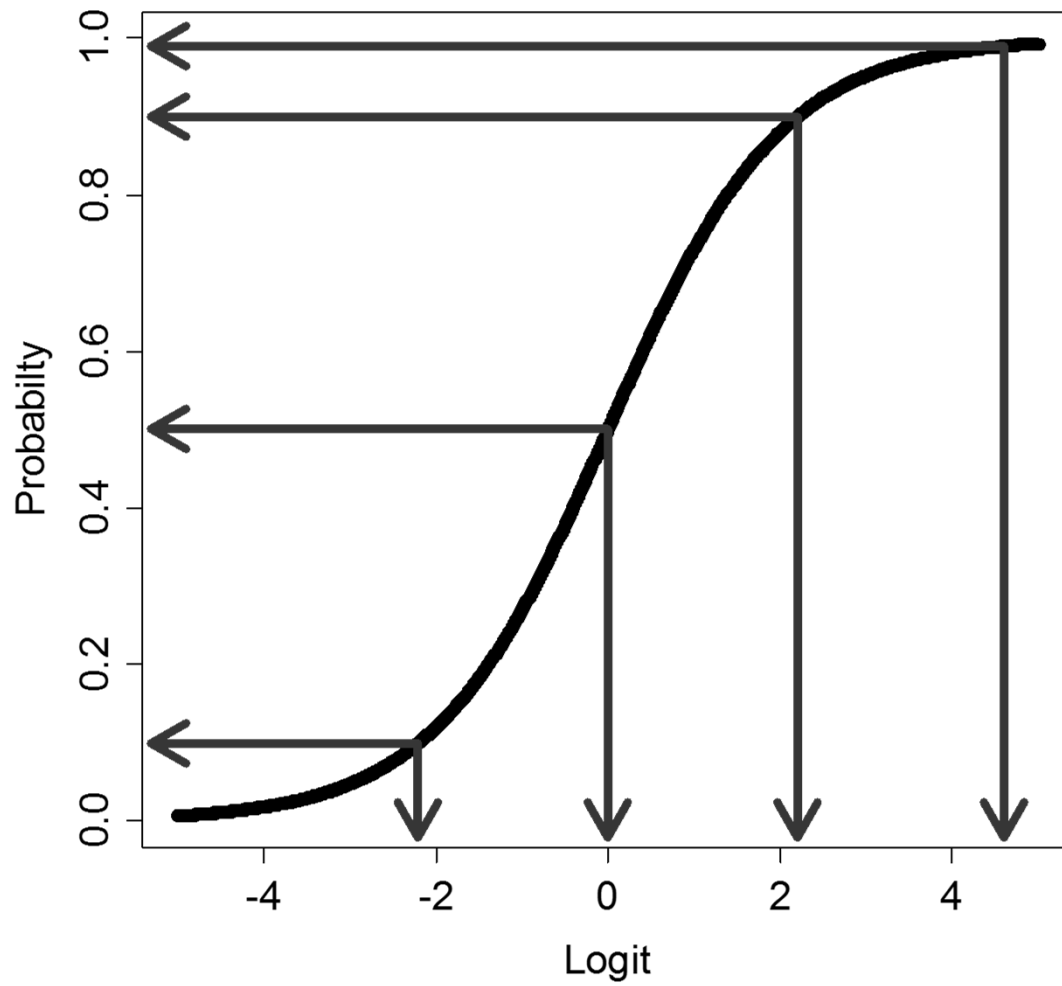


# Turning Probability into Logits

- Logit is a non-linear transformation of probability:
  - Equal intervals in logits are NOT equal in probability
  - The logit goes from  $\pm\infty$  and is symmetric about  $p = .5$  (logit = 0)
  - This solves the problem of using a linear model
    - ♦ The model will be linear with respect to the logit, which translates into nonlinear with respect to probability (i.e., it shuts off as needed)



# Transforming Probabilities to Logits



Probability	Logit
0.99	4.6
0.90	2.2
0.50	0.0
0.10	-2.2

Can you guess what a probability of .01 would be on the logit scale?

# Our New Model using LN(Odds)

---

- Outcome is log odds (logit) of probability instead of probability → symmetric, unbounded outcome
  - Assume linear relationship between X's and log odds (logit)
  - This allows an overall nonlinear (S-shaped) relationship between X's and probability of Y=1
- Don't assume errors are normal with constant variance
  - Note that 'e' was missing – residual variance is NOT estimated
  - Errors are assumed to follow a logistic distribution with a known residual variance of  $\frac{\pi^2}{3} = 3.29$

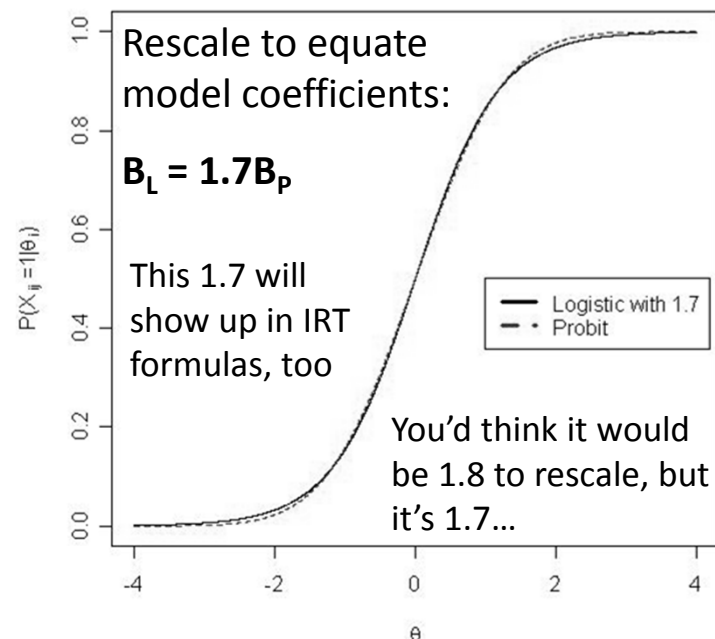
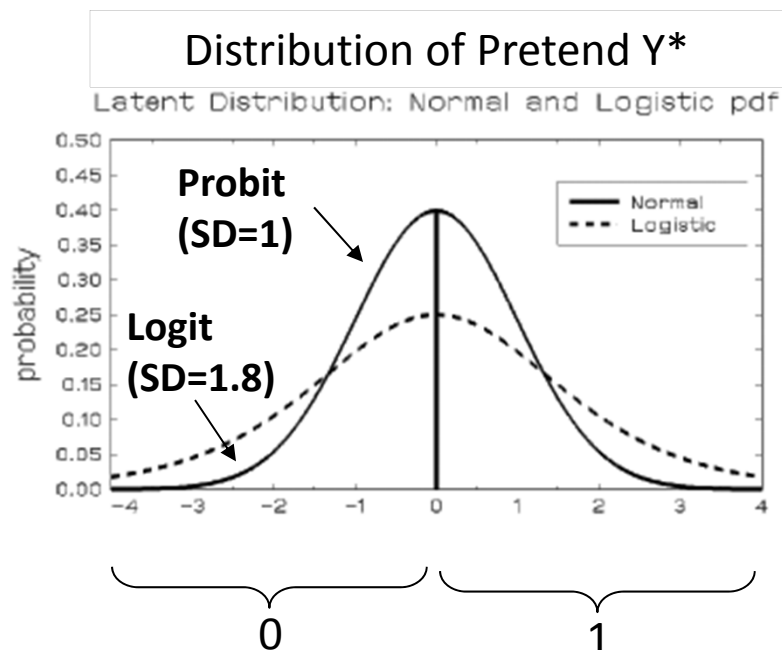
# Model Parts for Binary Outcomes:

## 2 Choices → Logit vs. Probit

---

- 2 Alternative Link Functions
  - Logit link: binary  $Y = \ln(p/1-p) \rightarrow$  logit is new transformed  $Y$ 
    - ♦  $Y$  is 0/1, but  $\text{logit}(Y)$  goes from  $-\infty$  to  $+\infty$
  - Probit link: binary  $Y = \Phi(Y)$ 
    - ♦ Observed probability replaced by value of standard normal curve below which observed proportion is found  
 $\rightarrow$  Z-score is new transformed  $Y$
    - ♦  $Y$  is 0/1, but  $\text{probit}(Y)$  goes from  $-\infty$  to  $+\infty$
- Same Model for the Means:
  - Main effects and interactions of predictors as desired...
  - No analog to odds coefficients in probit, however
- 2 Alternative Models for the Variances
  - Logit:  $e's \sim$  logistic distributed with known variance of  $\pi^2/3$ , or 3.29
  - Probit:  $e's \sim$  normally distributed with known variance of 1

# “Threshold Model” Concept



Another way these models are explained is with the “threshold concept”

Underlying the observed 0/1 response is really a pretend continuous variable called  $y^*$ , such that: if  $y^* < 0$  then  $y = 0$  and if  $y^* \geq 0$  then  $y = 1$

Accordingly, the difference between logit and probit is that the continuous underlying variable  $Y^*$  has a variance of 3.29 (SD = 1.8, logit) or 1.0 (probit)

# Summary: Generalized Models

---

- “General” models assume a continuous outcome with normally distributed residual errors – won’t work (well) for discrete outcomes
- “Generalized” models for other kinds of outcomes have 3 parts:
  - Link function
    - ♦ How Y gets transformed into something more continuous we can use
  - Model for the means
    - ♦ Regular old additive linear regression model on transformed Y
  - Model for the variances
    - ♦ Because the residuals errors cannot be normally distributed with constant variance, we change the assumption we make about them (i.e., to Bernoulli)
    - ♦ The residual variance is dependent on the mean, so it’s not estimated – it is fixed at some constant (i.e., at 3.29 in logit models, at 1.0 in probit models)
- As we will see next, IRT is a generalized mixed model
  - Link function to transform binary or categorical responses
  - But predictor is a latent trait (this is the same as a random intercept)

---

GLMM EXAMPLE:

# **ONE FACTOR MODEL FOR BINARY ITEMS**

# One-Factor Model for Binary Items

---

- Our gambling data set has data on the South Oaks Gambling Screen – a 12 item test used to classify probable pathological gamblers
  - Most items were dichotomous – but I will dichotomize all of the items to make things easy here
- We will use this test to demonstrate how GLMMs work
  - Again, we will use a one-factor model
- The goal of this demonstration is to get you used to how generalized models differ from our well-known CFA/SEM using normal distributions and identity links
  - But first...the math



# A One-Factor Model for the SOGS

---

- The one factor model assumes a single latent trait is measured by every item of the test
  - Because the items are dichotomous, we will use the Bernoulli distribution with a logit link
  - The model is then for the mean/variance of each item  $p_i$

Using the logit link (where  $g(\cdot)$  is the notation from GLMM):

$$g(E(y_{si})) = g(p_{si}) = \log\left(\frac{p_{si}}{1 - p_{si}}\right) = \mu_{I_i} + \lambda_{i,1}F_s + e_{si}$$

- $e_{is}$  is typically omitted because it has a constant variance – no unique variances get estimated
- $\mu_{I_i}$  - the item intercept, now has more meaning – it is no longer omitted casually from conversation

# Re-organizing the Link Function

---

- The logit link function  $g(p_{si})$  takes the right-hand side of the equation as a linear model (that is our CFA model)
- The Bernoulli distribution is expressed in terms of  $p_{si}$ , so we must solve for this to put the model into the equation

➤ This is called the inverse link function  $g^{-1}(\cdot)$

$$\begin{aligned} p_{si} = P(Y = 1|F_{s1}) &= g^{-1}(\mu_{I_i} + \lambda_{i,1}F_s) \\ &= \frac{\exp(\mu_{I_i} + \lambda_{i,1}F_s)}{1 + \exp(\mu_{I_i} + \lambda_{i,1}F_s)} \end{aligned}$$

- In other contexts, this is called the Two-Parameter Logistic Item Response Model
  - Item difficulty is  $b_i = -\frac{\mu_{I_i}}{\lambda_{i,1}}$
  - Item discrimination is  $a_i = \lambda_{i,1}$

## Conditional Distribution for an Item

---

- Recall the likelihood function for a given subject was composed of two parts:
  - $f(y_{si}|\mathbf{F}_s)$  - the conditional distribution of the data given the latent trait
  - $f(\mathbf{F}_s)$  – the distribution of the latent trait
- For the one-factor model the conditional distribution for one item given the value of the latent trait is:

$$\begin{aligned} f(y_{si}|\mathbf{F}_s) &= (p_i)^{y_{si}}(1 - p_i)^{1-y_{si}} \\ &= \left( \frac{\exp(\mu_{I_i} + \lambda_{i,1}F_s)}{1 + \exp(\mu_{I_i} + \lambda_{i,1}F_s)} \right)^{y_{si}} \left( 1 - \frac{\exp(\mu_{I_i} + \lambda_{i,1}F_s)}{1 + \exp(\mu_{I_i} + \lambda_{i,1}F_s)} \right)^{1-y_{si}} \end{aligned}$$

# Distribution of the Latent Trait

---

- The distribution of the latent trait is still what it was under CFA: normal with a mean  $\mu_{F_1}$  and variance  $\sigma_{F_1}^2$

$$f(F_{s1}) = \frac{1}{\sqrt{2\pi\sigma_{F_1}^2}} \exp\left(-\frac{(F_{s1} - \mu_{F_1})^2}{2\sigma_{F_1}^2}\right)$$

- In IRT models, the standardized factor identification method is predominantly used – but it doesn't have to be

# Putting the Two Distributions Together

---

Putting these together, we get:

$$\begin{aligned} f(\mathbf{y}_s) &= \int_{-\infty}^{\infty} f(\mathbf{y}_s | F_{s1}) f(F_{s1}) dF_{s1} \\ &= \int_{-\infty}^{\infty} \left[ \prod_{i=1}^I f(y_{si} | F_{s1}) \right] f(F_{s1}) dF_{s1} \\ &= \int_{-\infty}^{\infty} \left[ \prod_{i=1}^I \left( \frac{\exp(\mu_{I_i} + \lambda_{i,1} F_s)}{1 + \exp(\mu_{I_i} + \lambda_{i,1} F_s)} \right)^{y_{si}} \left( 1 - \frac{\exp(\mu_{I_i} + \lambda_{i,1} F_s)}{1 + \exp(\mu_{I_i} + \lambda_{i,1} F_s)} \right)^{1-y_{si}} \right] \times \\ &\quad \frac{1}{\sqrt{2\pi\sigma_{F_1}^2}} \exp\left(-\frac{(F_{s1} - \mu_{F_1})^2}{2\sigma_{F_1}^2}\right) dF_{s1} \end{aligned}$$

SERIOUSLY?!?

# Ramifications for Estimation of Bernoulli GLMMs

---

- Unlike the normal distribution/identity link, the Bernoulli distribution/logit link does not have a concise way of being summarized
- Therefore, we must approximate the integral if we are to find marginal ML parameter estimates
  - Approximation takes time – although for one factor it works pretty quickly in Mplus
- For more than one factor, it can be impossible
  - Mplus has a non-ML estimator that gets used frequently (called WLSMV) – I don't use it here as I want to focus on ML

# Ramifications for Fit of Bernoulli GLMMs

---

- Unlike the normal distribution/identity link, the Bernoulli distribution/logit link does not have a concise way of being summarized
- Under normal/identity, we used the mean vector/covariance matrix
- Under Bernoulli/logit – the sufficient statistic is the probabilities of any given response pattern
  - Therefore, for a test with  $I$  binary items, there are  $2^I - 1$  of these terms
  - We will see that makes fit difficult

# Mplus Code

---

- All of the math to this point leads to the Mplus code for categorical items:

```
VARIABLE:
  NAMES = SOGS4-SOGS15 SOGS4r SOGS5r SOGS6r student ID;
  USEVARIABLES = SOGS4r SOGS5r SOGS6r SOGS7-SOGS15; !use only binary items
  IDVARIABLE = ID; !label subjects
  CATEGORICAL = SOGS4r SOGS5r SOGS6r SOGS7-SOGS15; !categorical option for all
  MISSING = ALL(99); !missing data = 99 (MCAR)

ANALYSIS:
  ESTIMATOR = MLR; !set the estimator to (marginal) maximum likelihood
  PROCESSORS = 8; !use multiple processors (if available)

MODEL:

  GAMBLING by SOGS4r SOGS5r SOGS6r SOGS7-SOGS15;

PLOT:
  TYPE = PLOT1 PLOT2 PLOT3; !indicates we wish to have IRT graphics created

SAVEDATA:
  SAVE = FSCORES; !saves latent trait estimates
  FILE = alldata_sogsr_person2PL.dat;

OUTPUT:
  STANDARDIZED TECH1 TECH5 TECH8 TECH10; !displays model estimation and fit information
```

The line  
“CATEGORICAL”  
makes it all work



# Mplus Output: Model Fit

---

- The model fit information omits the typical CFA fit indices under ML
  - This is due to the covariance matrix not being a summary for the data

```
MODEL FIT INFORMATION

Number of Free Parameters                24

Loglikelihood

      H0 Value                        -3297.399
      H0 Scaling Correction Factor      1.0220
      for MLR

Information Criteria

      Akaike (AIC)                     6642.798
      Bayesian (BIC)                   6766.954
      Sample-Size Adjusted BIC         6690.718
      (n* = (n + 2) / 24)
```

- Where count of parameters comes from:
  - Each item has: 1 threshold (think intercept\*-1) and 1 factor loading (so, 2 x 12 items)
    - ◆ Note: no residual/unique variance is estimated (it doesn't exist in the model)

## Mplus Output: Model Fit Part 2

---

- There is a test of model fit – this uses the observed response pattern frequencies compared with the frequencies expected under the model

```
Chi-Square Test of Model Fit for the Binary and Ordered Categorical  
(Ordinal) Outcomes**
```

```
Pearson Chi-Square
```

Value	2016.578
Degrees of Freedom	4054
P-Value	1.0000

```
Likelihood Ratio Chi-Square
```

Value	565.122
Degrees of Freedom	4054
P-Value	1.0000

- If this were to work, this would be our gold standard for absolute model fit
  - The target distribution for dichotomous items is all possible response patterns
- This test is not useful most of the time (violates assumptions for anything less than about 5 items or so)
  - Lots of empty cells in the table – makes the  $\chi^2$  distribution not work correctly

## Model Fit, Part #3: Looking at Pair-wise Item Tables

---

- In this model, we do not have an overall omnibus goodness of fit test...but we do have some quirky ways of examining model data fit through expected/observed tables
  - Examine contingency tables for all pairs of items
- Under the OUTPUT section, if you put the word TECH10 Mplus will give you some sense of model fit (buried deep within the output):

### BIVARIATE MODEL FIT INFORMATION

Variable	Variable	Estimated Probabilities		Standardized Residual (z-score)
		H1	H0	
SOGS4R	SOGS5R			
Category 1	Category 1	0.811	0.813	-0.165
Category 1	Category 2	0.095	0.094	0.224
Category 2	Category 1	0.078	0.078	0.093
Category 2	Category 2	0.015	0.016	-0.207
Bivariate Pearson Chi-Square				0.101
Bivariate Log-Likelihood Chi-Square				0.101

- Listed here: H1 model → observed probability of sample responding for given pair  
H0 model → probability expected under estimated model
  - Chi-square statistic for any given item pair works like you would expect (1 DF)
  - Reject H0 (your current model fit) for pair if > 3.84 (for .05 Type I error rate)
- We can see how many pairs had significant model misfit
  - We can also see which items are causing the biggest issues with misfit

# Model Fit Assessment: Overall Summary for Our Data

---

- With 12 items, we have  $12 \times (12-1) / 2 = 66$  pairs of items

Overall Bivariate Pearson Chi-Square	106.742
Overall Bivariate Log-Likelihood Chi-Square	99.475

- Sum of 66 (1 df) Chi-square should have a mean equal to sum of df (so 66 is a rough guide as to where “good” fit would be)
  - We have a observed larger than the mean
  - Is it significant? Not able to assess using traditional Chi-Square distribution as sums are not independent
    - ♦ See FlexMIRT for a p-value
- Instead, we will look at the pairs of items that have Chi-Squares > 3.84
- We have 8 item pairs with significant misfit – three of them involved item 13
  - Item 4 and 6
  - Item 4 and 10
  - Item 5 and 13
  - Item 6 and 15
  - Item 7 and 15
  - Item 10 and 13
  - Item 11 and 13
  - Item 13 and 15

# Making the Model Fit Better: Removing Misfitting Items

---

- As item 13 was the one that seemed to have the most problems, we will examine it:
  - 13. (if yes to 12) Have money arguments ever centered on your gambling? (yes/no)
- Item 13 actually relates to item 12 (so more dependency than a model would need):
  - 12. Have you ever argued with people you love over how you handle money? (yes/no)
- I think this is evidence enough to remove the item and re-check the fit
  - Typically, these types of items cause issues in analysis as there are item-level dependencies
  - Sometimes these dependencies can be accounted for in analysis (see testlet models)...but with only two items, this won't work
- Our analysis now has 11 items (so a “good” fit would be a total Chi-Square of 55)
  - Our model was pretty close – so we will call this good enough

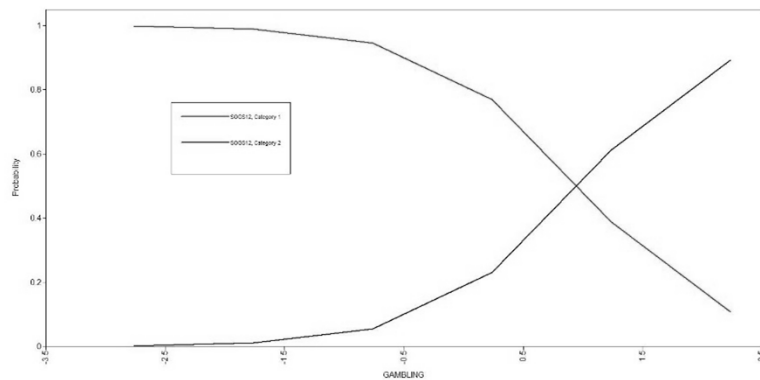
Overall Bivariate Pearson Chi-Square	61.067
Overall Bivariate Log-Likelihood Chi-Square	58.360
- Now only 4 items had significant Chi-Square values:
  - Item 4 and 6
  - Item 4 and 10
  - Item 6 and 15
  - Item 7 and 10

# Mplus Unstandardized Parameter Results

- Example for item 12:
- Threshold estimate: 1.599
- Loading estimate: 1.661

What this means:

$$\text{logit}(P(X_{s,12} = 0|\theta_s)) \\ = 1.599 + 1.661\theta_s$$



## MODEL RESULTS

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
<b>GAMBLING BY</b>				
SOGS4R	1.000	0.000	999.000	999.000
SOGS5R	4.064	1.543	2.634	0.008
SOGS6R	8.934	3.283	2.721	0.007
SOGS7	5.942	2.275	2.611	0.009
SOGS8	6.969	2.634	2.645	0.008
SOGS9	5.582	2.126	2.626	0.009
SOGS10	10.372	4.003	2.591	0.010
SOGS11	9.272	3.742	2.478	0.013
SOGS12	1.661	0.646	2.571	0.010
SOGS14	4.711	1.762	2.673	0.008
SOGS15	5.823	2.295	2.537	0.011
<b>Thresholds</b>				
SOGS4R\$1	2.355	0.114	20.636	0.000
SOGS5R\$1	3.153	0.208	15.137	0.000
SOGS6R\$1	7.250	0.797	9.097	0.000
SOGS7\$1	2.344	0.222	10.574	0.000
SOGS8\$1	5.167	0.477	10.830	0.000
SOGS9\$1	4.564	0.384	11.875	0.000
SOGS10\$1	9.005	1.314	6.853	0.000
SOGS11\$1	8.445	1.274	6.630	0.000
SOGS12\$1	1.599	0.088	18.220	0.000
SOGS14\$1	5.364	0.426	12.596	0.000
SOGS15\$1	5.723	0.510	11.232	0.000
<b>Variances</b>				
GAMBLING	0.212	0.153	1.385	0.166

# Mplus Standardized Parameter Results

- The STD standardization as that gives us the parameters if the factor was standardized for identification
- This mirrors most IRT-type identification methods
  - IRT is typically used for scoring and calibrating items so setting the factor variance is typical

## STDYX Standardization

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
<b>GAMBLING BY</b>				
SOGS4R	0.246	0.083	2.948	0.003
SOGS5R	0.718	0.035	20.318	0.000
SOGS6R	0.915	0.019	47.732	0.000
SOGS7	0.833	0.028	29.429	0.000
SOGS8	0.871	0.024	36.358	0.000
SOGS9	0.817	0.031	26.600	0.000
SOGS10	0.935	0.020	47.544	0.000
SOGS11	0.920	0.025	37.535	0.000
SOGS12	0.388	0.046	8.433	0.000
SOGS14	0.767	0.040	19.241	0.000
SOGS15	0.828	0.033	25.280	0.000
<b>Thresholds</b>				
SOGS4R\$1	1.259	0.052	24.215	0.000
SOGS5R\$1	1.210	0.046	26.081	0.000
SOGS6R\$1	1.613	0.055	29.197	0.000
SOGS7\$1	0.714	0.038	18.764	0.000
SOGS8\$1	1.402	0.050	28.033	0.000
SOGS9\$1	1.451	0.052	27.909	0.000
SOGS10\$1	1.763	0.060	29.583	0.000
SOGS11\$1	1.821	0.062	29.363	0.000
SOGS12\$1	0.813	0.039	20.740	0.000
SOGS14\$1	1.897	0.072	26.203	0.000
SOGS15\$1	1.768	0.063	28.091	0.000
<b>Variances</b>				
GAMBLING	1.000	0.000	999.000	999.000

# Mplus Additional Parameters: IRT Equivalents

- For **binary items** with **one factor**, Mplus also gives the parameters as commonly used IRT terms

- IRT Parameterization:  

$$\text{logit}(P(X_{si} = 1|\theta_s)) = 1.7a_i(\theta_s - b_i)$$

- Comparison:

Mplus threshold:  $\tau_i = -1.7a_ib_i$

Mplus loading:  $\lambda_i = 1.7a_i$

IRT difficulty:  $b_i = -\frac{\tau_i}{1.7a_i}$

IRT discrimination:  $a_i = \frac{\lambda_i}{1.7}$

IRT PARAMETERIZATION IN TWO-PARAMETER LOGISTIC METRIC  
 WHERE THE LOGIT IS DISCRIMINATION\*(THETA - DIFFICULTY)

## Item Discriminations

### GAMBLING BY

SOGS4R	0.460	0.166	2.770	0.006
SOGS5R	1.871	0.190	9.843	0.000
SOGS6R	4.113	0.529	7.770	0.000
SOGS7	2.736	0.304	8.987	0.000
SOGS8	3.208	0.364	8.806	0.000
SOGS9	2.570	0.291	8.843	0.000
SOGS10	4.775	0.797	5.994	0.000
SOGS11	4.269	0.744	5.740	0.000
SOGS12	0.765	0.107	7.161	0.000
SOGS14	2.169	0.274	7.918	0.000
SOGS15	2.681	0.338	7.937	0.000

## Item Difficulties

SOGS4R\$1	5.116	1.729	2.959	0.003
SOGS5R\$1	1.685	0.103	16.328	0.000
SOGS6R\$1	1.763	0.074	23.919	0.000
SOGS7\$1	0.857	0.054	15.936	0.000
SOGS8\$1	1.610	0.073	21.939	0.000
SOGS9\$1	1.776	0.091	19.451	0.000
SOGS10\$1	1.886	0.079	23.944	0.000
SOGS11\$1	1.978	0.086	22.925	0.000
SOGS12\$1	2.092	0.256	8.161	0.000
SOGS14\$1	2.473	0.168	14.694	0.000
SOGS15\$1	2.135	0.121	17.668	0.000

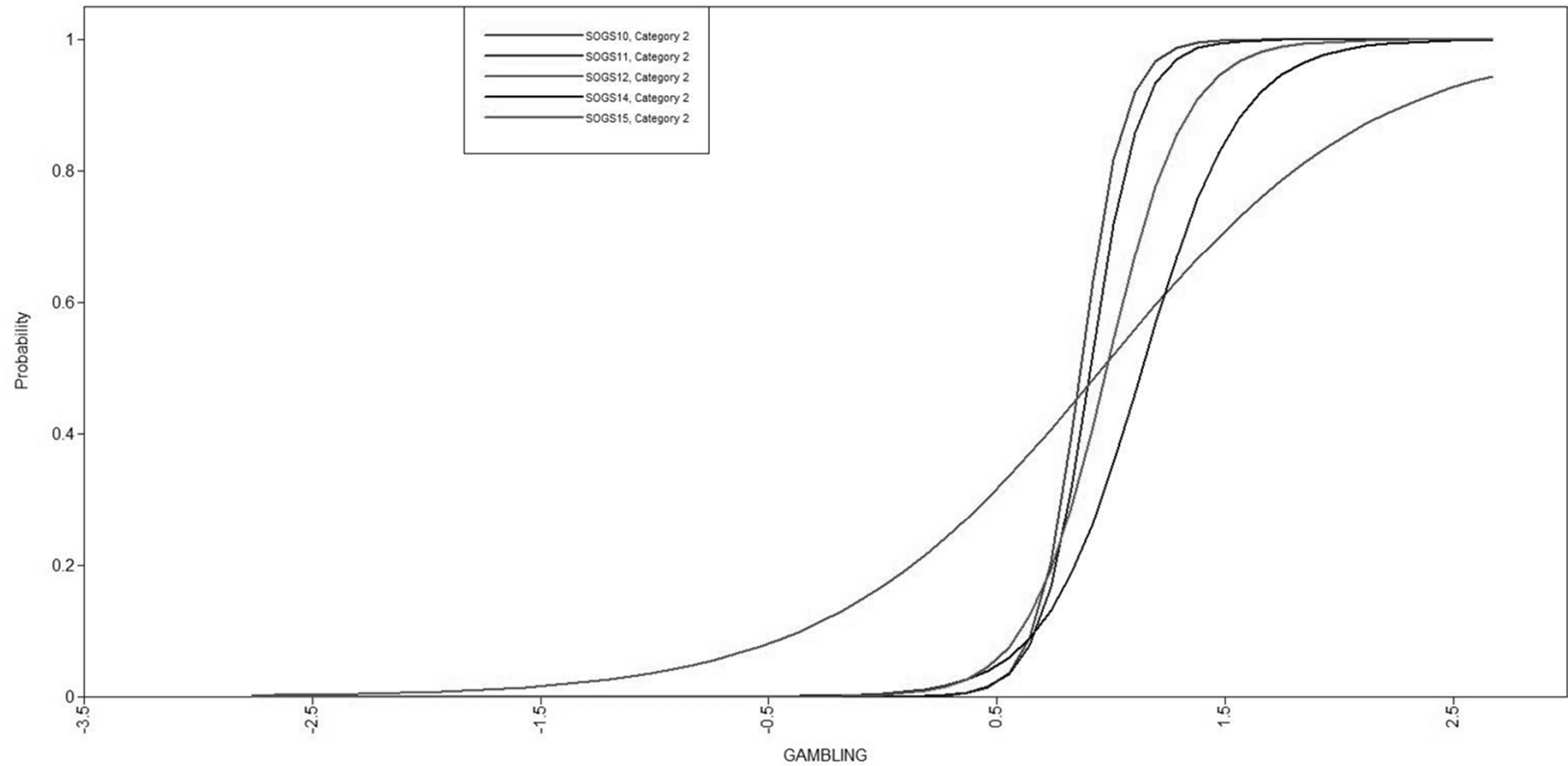
## Variances

GAMBLING	1.000	0.000	0.000	1.000
----------	-------	-------	-------	-------



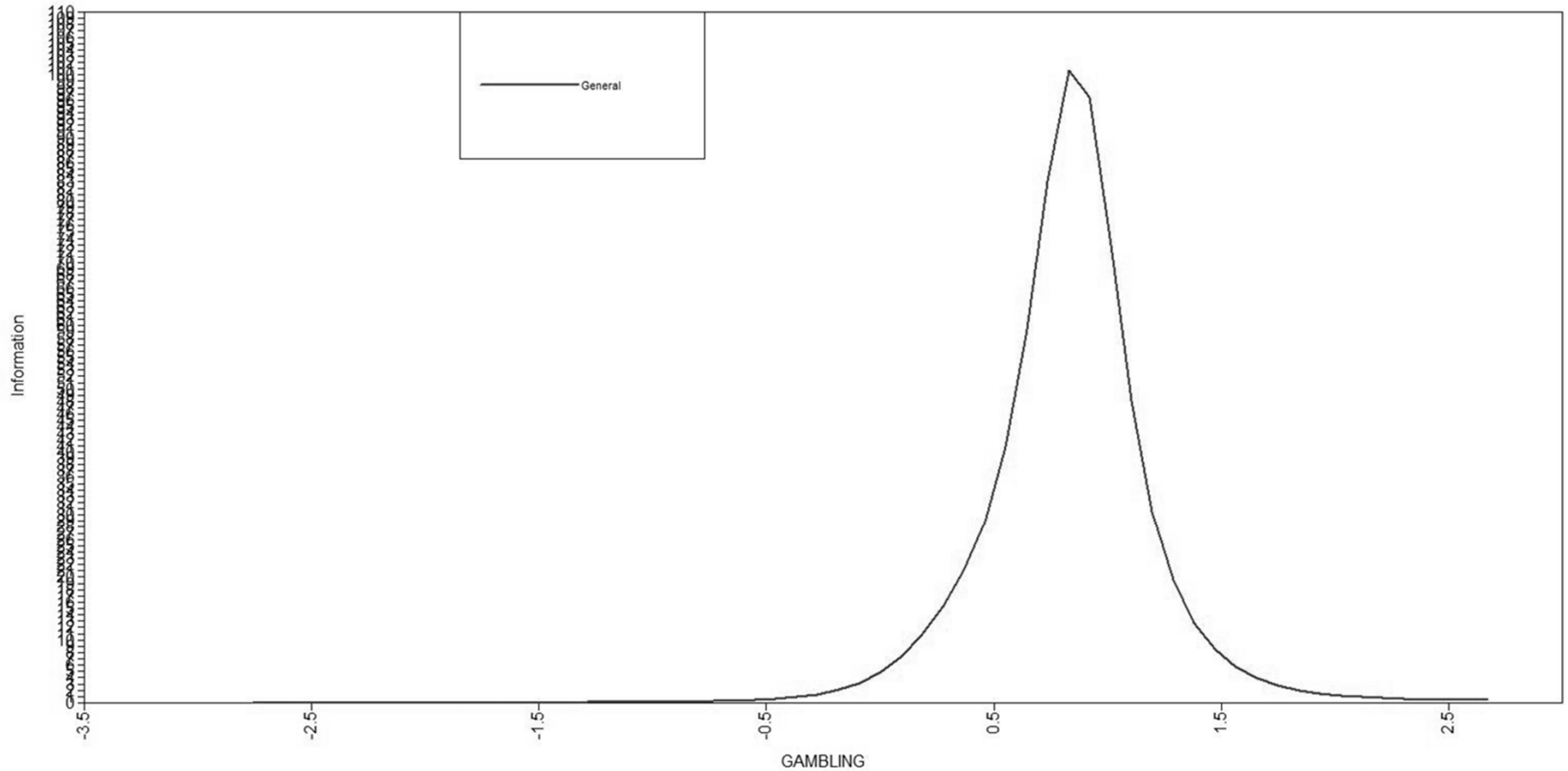
# Item Results - Plots

---



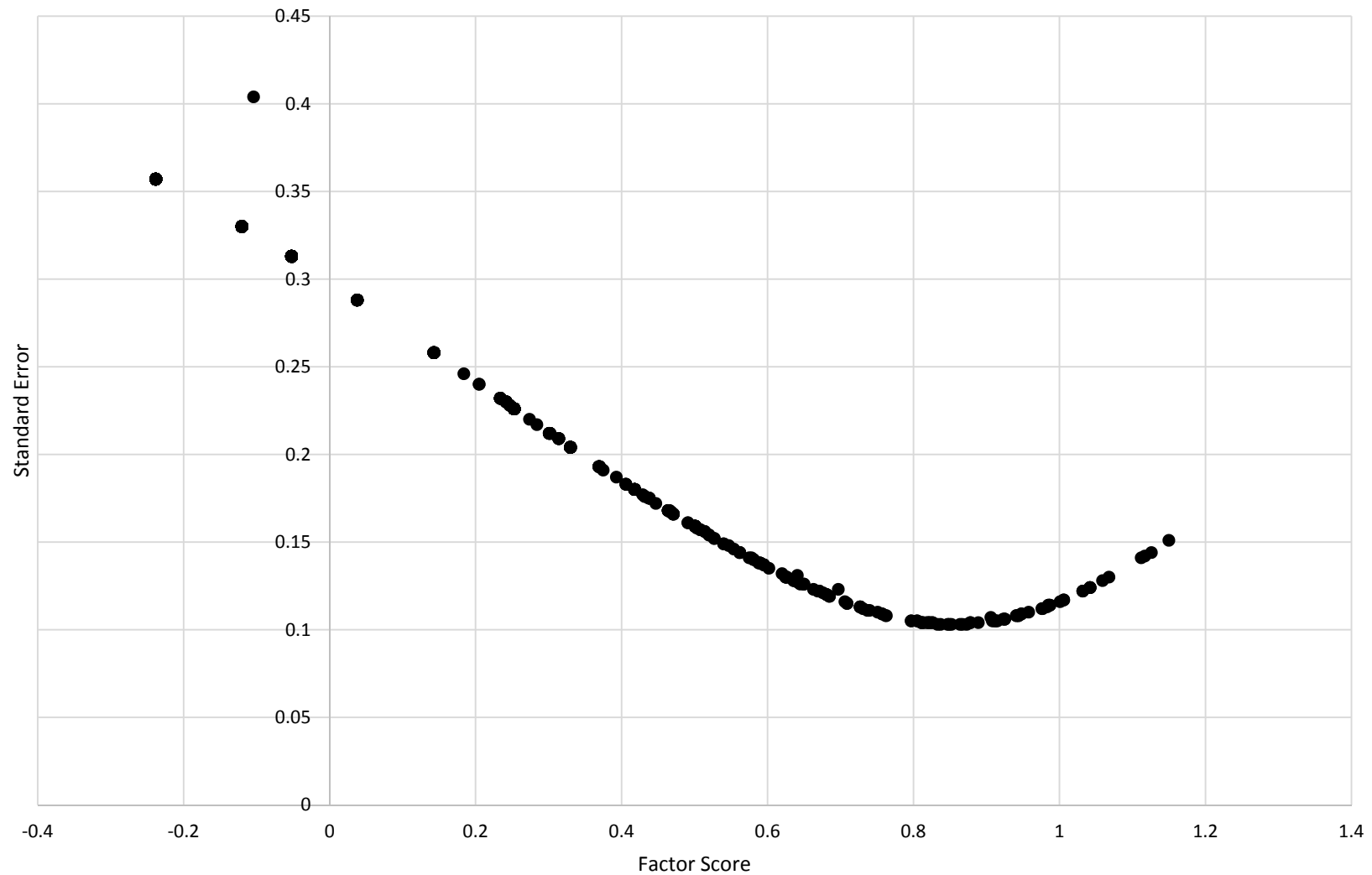
# Reliability and Test Information: Not Constant Outside of CFA

---



# Comparison of Factor Score by Standard Error of Factor Score:

---



# Bernoulli Example Wrap Up

---

- For dichotomous items, the Bernoulli/logit GLMM is a common choice
- Under ML – it is hard to assess fit and even harder to estimate
  - Lots of data needed...and lots of time needed
- The modeling process and set up are the same, though
  - The right-hand side of the equation (the theory) still is the same as in CFA
  - Only the data have changed
- Although all of our items were Bernoulli, they didn't have to be
  - Could have just omitted some from the CATEGORICAL line
  - The likelihood function still applies (although the easy form of CFA being MVN won't happen)

---

**ADDITIONAL GLMMS AVAILABLE  
IN MPLUS**

# Ordered Categorical Outcomes

---

- One option: **Cumulative Logit Model**
  - Called “graded response model” in IRT
  - Assumes ordinal categories
  - Model logit of category vs. all lower/higher via submodels
    - ♦ 3 categories → 2 models: 0 vs. 1 or 2, 0 or 1 vs. 2
  - Get separate threshold (-intercept) for each submodel
  - Effects of predictors are assumed the same across submodels → “Proportional odds assumption”
    - ♦ Is testable in some software (e.g., Mplus, NLMIXED)
  - In Mplus, can do this with the **CATEGORICAL ARE** option

## Unordered Categorical Outcomes: “Nominal Model”

---

- Compare each category against a reference category using a binary logit model → referred to as “baseline category logit”
- End up with multiple logistic submodels up to #categories – 1  
(2 submodels for 3 categories, 3 for 4 categories, etc)
- Intercept/thresholds and slopes for effects of predictors (factor loadings) are estimated separately within each binary submodel
  - Can get effects for missing contrast via subtraction
  - Effects are interpreted as “given that it’s one of these two categories, which has the higher probability”?
- Model comparisons proceed as in logistic regression
  - Can also test whether outcome categories can be collapsed
- In Mplus, can do this with the **NOMINAL ARE** option

# Censored (“Tobit”) Outcomes

---

- For outcomes with ceiling or floor effects
  - Can be “Right censored” and/or “left censored”
  - Also “inflated” or not → inflation = binary variable in which  
1 = censored, 0 = not censored
- Model assumes unobserved continuous distribution instead for the part it is missing
- In Mplus, can do with various **CENSORED ARE** (with options):
  - CENSORED ARE y1 (a) y2 (b) y3 (ai) y4 (bi);
    - ♦ y1 is censored from **a**bove (right); y2 is censored from **b**elow (left)
    - ♦ y3 is censored from **a**bove (right) and has inflation variable (inflated: y3#1)
    - ♦ y4 is censored from **a**bove (below) and has inflation variable (inflated: y4#1)
  - So, can predict distribution of y1-y4, as well as whether or not y3 and y4 are censored (“inflation”) as separate outcomes
    - ♦ y3 ON x; → x predicts value of Y if at censoring point or above
    - ♦ y3#1 ON x; → x predicts whether Y is censored (1) or not (0)



# A Family of Options in Mplus for Count Outcomes (COUNT ARE)

---

- Counts: non-negative integer unbounded responses
  - e.g., how many cigarettes did you smoke this week?
- Poisson and negative binomial models
  - **Same Link: count  $Y = \ln(Y)$**  (makes the count stay positive)
  - **$\text{LN}(Y_{is}) = \mu_i + \lambda_i F_s + e_{is}$**  (model has intercepts and loadings)
  - Residuals follow 1 of 2 distributions:
    - ♦ Poisson distribution in which  $k = \text{Mean} = \text{Variance}$
    - ♦ Negative binomial distribution that includes a new  $\alpha$  “scaling” or “over-dispersion” parameter that allows the variance to be bigger than the mean  $\rightarrow \text{variance} = k(1 + k\alpha)$
    - ♦ Poisson is nested within negative binomial (can test of  $\alpha \neq 0$ )
    - ♦ COUNT ARE y1 (p) y2 (nb);  $\rightarrow$  y1 is Poisson; y2 is neg. binomial

# Issues with Zeros in Count Data

---

- No zeros → zero-truncated negative binomial
  - e.g., how many days were you in the hospital? (has to be  $>0$ )
  - COUNT ARE y1 (nbt);
- Too many zeros → zero-inflated poisson or negative binomial
  - e.g., # cigarettes smoked when asked in non-smokers too
  - COUNT ARE y2 (pi) y3 (nbi);
    - ◆ Refer to “inflation” variable as y2#1 or y3#1
  - Tries to distinguish 2 kinds of zeros
    - ◆ “Structural zeros” – would never do it
      - Inflation is predicted as logit of being a structural zero
    - ◆ “Expected zeros” – could do it, just didn’t (part of regular count)
      - Count with expected zeros predicted by poisson or negative binomial
  - Poisson or neg binomial without inflation is nested within models with inflation (and poisson is nested within negative binomial)

# Issues with Zeros in Count Data

---

- Other more direct ways of dealing with too many zeros: split distribution into (0 or not) and (if not 0, how much)?
  - Negative binomial “hurdle” (or “zero-altered” negative binomial)
    - ♦ COUNT ARE y1 (nbh);
    - ♦ 0 or not: predicted by logit of being a 0 (“0” is the higher category)
    - ♦ How much: predicted by zero-truncated negative binomial
  - Two-part model uses Mplus DATA TWOPART: command
    - ♦ NAMES ARE y1-y4; → list outcomes to be split into 2 parts
    - ♦ CUTPOINT IS 0; → where to split observed outcomes
    - ♦ BINARY ARE b1-b4; → create names for “0 or not” part
    - ♦ CONTINUOUS ARE c1-c4; → create names for “how much” part
    - ♦ TRANSFORM IS LOG; → transformation of continuous part
    - ♦ 0 or not: predicted by logit of being NOT 0 (“something” is the 1)
    - ♦ How much: predicted by transformed normal distribution (like log)

---

## **CONCLUDING REMARKS**

# Wrapping Up...

---

- When fitting latent factor models (or when just predicting observed outcomes from observed predictors instead), you have many options to fit non-normal distributions
  - **CFA:** Continuous outcomes with normal residuals,  $X \rightarrow Y$  is linear
    - ♦ If residuals may not be normal *but a linear  $X \rightarrow Y$  relationship is still plausible*, you can use MLR estimation instead of ML to control for that
  - **IRT and IFA:** Categorical or ordinal outcomes with Bernoulli/multinomial residuals,  $X \rightarrow$  transformed  $Y$  is linear;  $X \rightarrow$  original  $Y$  is nonlinear
    - ♦ Full information MML traditionally paired with IRT version of model; limited information WLSMV traditionally paired with IFA version of model instead
  - **Censored:** Continuous outcomes that shut off,  $X \rightarrow Y$  is linear
    - ♦ Model tries to predict what would happen if  $Y$  kept going instead
  - **Count family:** Non-negative integer outcomes,  $X \rightarrow \text{LN}(Y)$  is linear
    - ♦ Residuals can be Poisson (where mean = variance) or negative binomial (where variance > mean); either can be zero-inflated or zero-truncated
    - ♦ Hurdle or two-part may be more direct way to predict/interpret excess zeros (predict zero or not and how much rather than two kinds of zeros)