Phenotypic Multi-group Analysis

Measurement Invariance
Continuous and categorical data

Gitta Lubke [1]   Irene Rebollo[2]

[1]Quantitative Psychology
University of Notre Dame

[2]Biological Psychology
Free University Amsterdam

Twin Workshop, Boulder 2008

Outline

Outline

The general setting

- collected data are scores on items (e.g., questionnaire)
- questionnaire designed to measure a psychiatric disorder
  - or something else
- the grouping variable ‘gender’ defines two groups
  - or some other variable defines some other small number of groups
- interest in investigating differences between the two groups
- genetic decomposition can be added

The general setting

- collected data are scores on items (e.g., questionnaire)
- questionnaire designed to measure a psychiatric disorder
  - or something else
- the grouping variable 'gender' defines two groups
  - or some other variable defines some other small number of groups
- interest in investigating differences between the two groups
- genetic decomposition can be added

The general setting

- collected data are scores on items (e.g., questionnaire)
- questionnaire designed to measure a psychiatric disorder
  - or something else
- the grouping variable ‘gender’ defines two groups
  - or some other variable defines some other small number of groups
- interest in investigating differences between the two groups
- genetic decomposition can be added

The general setting

- collected data are scores on items (e.g., questionnaire)
- questionnaire designed to measure a psychiatric disorder
  - or something else
- the grouping variable 'gender' defines two groups
  - or some other variable defines some other small number of groups
- interest in investigating differences between the two groups
- genetic decomposition can be added

The general setting

- collected data are scores on items (e.g., questionnaire)
- questionnaire designed to measure a psychiatric disorder
    - or something else
- the grouping variable ‘gender’ defines two groups
    - or some other variable defines some other small number of groups
- interest in investigating differences between the two groups
- genetic decomposition can be added

The general setting

- collected data are scores on items (e.g., questionnaire)
- questionnaire designed to measure a psychiatric disorder
  - or something else
- the grouping variable ‘gender’ defines two groups
  - or some other variable defines some other small number of groups
- interest in investigating differences between the two groups
- genetic decomposition can be added

Observed items and theoretical constructs

- questionnaire items designed to measure a theoretical construct
- scores on the questionnaire items are observed
- constructs are usually not observable (e.g., IQ)
  - small number of theoretical construct
  - larger number observed variables
- goal is to compare groups with respect to the constructs

Observed items and theoretical constructs

- questionnaire items designed to measure a theoretical construct
- scores on the questionnaire items are observed
- constructs are usually not observable (e.g., IQ)
  - small number of theoretical construct
  - larger number observed variables
- goal is to compare groups with respect to the constructs

Observed items and theoretical constructs

- questionnaire items designed to measure a theoretical construct
- scores on the questionnaire items are observed
- constructs are usually not observable (e.g., IQ)
    - small number of theoretical construct
    - larger number observed variables
- goal is to compare groups with respect to the constructs

Observed items and theoretical constructs

- questionnaire items designed to measure a theoretical construct
- scores on the questionnaire items are observed
- constructs are usually not observable (e.g., IQ)
  - small number of theoretical construct
  - larger number observed variables
- goal is to compare groups with respect to the constructs

Observed items and theoretical constructs

- questionnaire items designed to measure a theoretical construct
- scores on the questionnaire items are observed
- constructs are usually not observable (e.g., IQ)
  - small number of theoretical construct
  - larger number observed variables
- goal is to compare groups with respect to the constructs

Outline

The general idea

- items that are designed to measure the same construct should have common variance
- common variance is captured by a latent variable
  - items measuring different aspects of math achievement
  - common variance of math items captured by (latent) math achievement
- latent variables (LV's) $\approx$ theoretical constructs
- questionnaire items = observed variables (OV's)
- latent variable models relate OV's to the LV's

The general idea

- items that are designed to measure the same construct should have common variance
- common variance is captured by a latent variable
  - items measuring different aspects of math achievement
  - common variance of math items captured by (latent) math achievement
- latent variables (LV's) $\approx$ theoretical constructs
- questionnaire items = observed variables (OV's)
- latent variable models relate OV's to the LV's

The general idea

- items that are designed to measure the same construct should have common variance
- common variance is captured by a latent variable
  - items measuring different aspects of math achievement
  - common variance of math items captured by (latent) math achievement
- latent variables (LV's) ≈ theoretical constructs
- questionnaire items = observed variables (OV's)
- latent variable models relate OV's to the LV's

The general idea

- items that are designed to measure the same construct should have common variance
- common variance is captured by a latent variable
  - items measuring different aspects of math achievement
  - common variance of math items captured by (latent) math achievement
- latent variables (LV's) $\approx$ theoretical constructs
- questionnaire items = observed variables (OV's)
- latent variable models relate OV's to the LV's

The general idea

- items that are designed to measure the same construct should have common variance
- common variance is captured by a latent variable
  - items measuring different aspects of math achievement
  - common variance of math items captured by (latent) math achievement
- latent variables (LV's) $\approx$ theoretical constructs
- questionnaire items = observed variables (OV's)
- latent variable models relate OV's to the LV's

The general idea

- items that are designed to measure the same construct should have common variance
- common variance is captured by a latent variable
  - items measuring different aspects of math achievement
  - common variance of math items captured by (latent) math achievement
- latent variables (LV's) $\approx$ theoretical constructs
- questionnaire items = observed variables (OV's)
- latent variable models relate OV's to the LV's

The general idea

- items that are designed to measure the same construct should have common variance
- common variance is captured by a latent variable
  - items measuring different aspects of math achievement
  - common variance of math items captured by (latent) math achievement
- latent variables (LV's) $\approx$ theoretical constructs
- questionnaire items = observed variables (OV's)
- latent variable models relate OV's to the LV's

Types of latent variable models

- different types of latent variable models
  - confirmatory factor analysis model (CFA)
  - item response models (IRT)
- CFA and IRT models are related
- both model types can be used for multiple group analyses
- goal is to compare groups with respect to a few latent variables
  - not with respect to many individual items
- how group comparisons should be done does not depend on the type of LV model

Types of latent variable models

- different types of latent variable models
  - confirmatory factor analysis model (CFA)
  - item response models (IRT)
- CFA and IRT models are related
- both model types can be used for multiple group analyses
- goal is to compare groups with respect to a few latent variables
  - not with respect to many individual items
- how group comparisons should be done does not depend on the type of LV model

Types of latent variable models

- different types of latent variable models
  - confirmatory factor analysis model (CFA)
  - item response models (IRT)
- CFA and IRT models are related
- both model types can be used for multiple group analyses
- goal is to compare groups with respect to a few latent variables
  - not with respect to many individual items
- how group comparisons should be done does not depend on the type of LV model

Types of latent variable models

- different types of latent variable models
  - confirmatory factor analysis model (CFA)
  - item response models (IRT)
- CFA and IRT models are related
- both model types can be used for multiple group analyses
- goal is to compare groups with respect to a few latent variables
  - not with respect to many individual items
- how group comparisons should be done does not depend on the type of LV model

Types of latent variable models

- different types of latent variable models
  - confirmatory factor analysis model (CFA)
  - item response models (IRT)
- CFA and IRT models are related
- both model types can be used for multiple group analyses
- goal is to compare groups with respect to a few latent variables
  - not with respect to many individual items
- how group comparisons should be done does not depend on the type of LV model

Challenge

- prerequisite of group comparisons: show that items are measurement invariant
  - measurement invariance (MI) is absence of bias
- groups can be compared with respect to the latent variables only if MI holds
- to understand why biased items are problematic when comparing groups with respect to latent variables this section will cover
  - an introduction to the multi-group CFA model
  - conceptual approach to bias
  - theory of measurement invariance (MI)
  - how MI relates to the CFA model
  - how MI can be tested using the CFA model
  - MI in the IRT model

Challenge

- prerequisite of group comparisons: show that items are measurement invariant
  - measurement invariance (MI) is absence of bias
- groups can be compared with respect to the latent variables only if MI holds
- to understand why biased items are problematic when comparing groups with respect to latent variables this section will cover
  - an introduction to the multi-group CFA model
  - conceptual approach to bias
  - theory of measurement invariance (MI)
  - how MI relates to the CFA model
  - how MI can be tested using the CFA model
  - MI in the IRT model

Challenge

- prerequisite of group comparisons: show that items are measurement invariant
  - measurement invariance (MI) is absence of bias
- groups can be compared with respect to the latent variables only if MI holds
- to understand why biased items are problematic when comparing groups with respect to latent variables this section will cover
  - an introduction to the multi-group CFA model
  - conceptual approach to bias
  - theory of measurement invariance (MI)
  - how MI relates to the CFA model
  - how MI can be tested using the CFA model
  - MI in the IRT model

Challenge

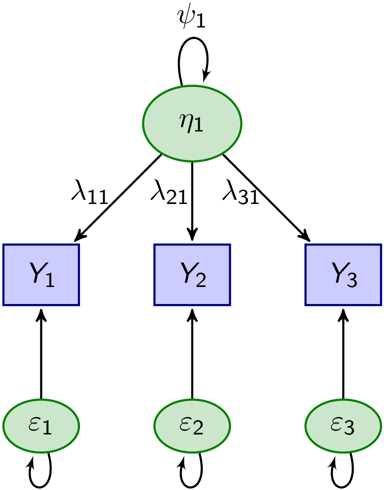
- prerequisite of group comparisons: show that items are measurement invariant
  - measurement invariance (MI) is absence of bias
- groups can be compared with respect to the latent variables only if MI holds
- to understand why biased items are problematic when comparing groups with respect to latent variables this section will cover
  - an introduction to the multi-group CFA model
  - conceptual approach to bias
  - theory of measurement invariance (MI)
  - how MI relates to the CFA model
  - how MI can be tested using the CFA model
  - MI in the IRT model

Challenge

- prerequisite of group comparisons: show that items are measurement invariant
  - measurement invariance (MI) is absence of bias
- groups can be compared with respect to the latent variables only if MI holds
- to understand why biased items are problematic when comparing groups with respect to latent variables this section will cover
  - an introduction to the multi-group CFA model
  - conceptual approach to bias
  - theory of measurement invariance (MI)
  - how MI relates to the CFA model
  - how MI can be tested using the CFA model
  - MI in the IRT model

Outline

The common factor model: Path diagram



The common factor model: Model notation

- the model consists of *random variables* (=differ across subjects) and *model parameters* (=are the same across $i = 1, 2, \ldots, N$ subjects)
- the random variables for subject i are denoted as follows
  - $j = 1, 2, \ldots, J$ observed variables as Y_{ij}
  - $k = 1, 2, \ldots, K$ latent variables (='factors') as η_{ik}
  - measurement error of observed variable Y_j as ε_{ij}
- the model parameters are denoted as
  - regression intercept of Y_j as ν_j
  - regression slope (='loading') of Y_j on η_k as λ_{jk}
- the CFA model is a linear regression model, for subject i we have

$$Y_{ij} = \nu_j + \sum_{k=1}^{K} \lambda_{jk}\eta_{ik} + \varepsilon_{ij}$$

The common factor model: Model notation

- the model consists of *random variables* (=differ across subjects) and *model parameters* (=are the same across $i = 1, 2, \ldots, N$ subjects)
- the random variables for subject i are denoted as follows
  - $j = 1, 2, \ldots, J$ observed variables as Y_{ij}
  - $k = 1, 2, \ldots, K$ latent variables (='factors') as η_{ik}
  - measurement error of observed variable Y_j as ε_{ij}
- the model parameters are denoted as
  - regression intercept of Y_j as ν_j
  - regression slope (='loading') of Y_j on η_k as λ_{jk}
- the CFA model is a linear regression model, for subject i we have

$$Y_{ij} = \nu_j + \sum_{k=1}^{K} \lambda_{jk}\eta_{ik} + \varepsilon_{ij}$$

The common factor model: Model notation

- the model consists of *random variables* (=differ across subjects) and *model parameters* (=are the same across $i = 1, 2, \ldots, N$ subjects)
- the random variables for subject i are denoted as follows
  - $j = 1, 2, \ldots, J$ observed variables as Y_{ij}
  - $k = 1, 2, \ldots, K$ latent variables (='factors') as η_{ik}
  - measurement error of observed variable Y_j as ε_{ij}
- the model parameters are denoted as
  - regression intercept of Y_j as ν_j
  - regression slope (='loading') of Y_j on η_k as λ_{jk}
- the CFA model is a linear regression model, for subject i we have

$$Y_{ij} = \nu_j + \sum_{k=1}^{K} \lambda_{jk}\eta_{ik} + \varepsilon_{ij}$$

The common factor model: Model notation

- the model consists of *random variables* (=differ across subjects) and *model parameters* (=are the same across $i = 1, 2, \ldots, N$ subjects)
- the random variables for subject i are denoted as follows
  - $j = 1, 2, \ldots, J$ observed variables as Y_{ij}
  - $k = 1, 2, \ldots, K$ latent variables (='factors') as η_{ik}
  - measurement error of observed variable Y_j as ε_{ij}
- the model parameters are denoted as
  - regression intercept of Y_j as ν_j
  - regression slope (='loading') of Y_j on η_k as λ_{jk}
- the CFA model is a linear regression model, for subject i we have

$$Y_{ij} = \nu_j + \sum_{k=1}^{K} \lambda_{jk}\eta_{ik} + \varepsilon_{ij}$$

Constraints to identify parameters

- latent variables have no scale
    - means? variance?
- latent variables η have to be scaled
- set means to zero
- either fix variance to 1 *or* one loading per factor to 1
    - since we are extending to multiple groups and may want to investigate variance differences between groups, we fix one loading to 1

Constraints to identify parameters

- latent variables have no scale
  - means? variance?
- latent variables η have to be scaled
- set means to zero
- either fix variance to 1 *or* one loading per factor to 1
  - since we are extending to multiple groups and may want to investigate variance differences between groups, we fix one loading to 1

Constraints to identify parameters

- latent variables have no scale
  - means? variance?
- latent variables η have to be scaled
- set means to zero
- either fix variance to 1 *or* one loading per factor to 1
  - since we are extending to multiple groups and may want to investigate variance differences between groups, we fix one loading to 1

Constraints to identify parameters

- latent variables have no scale
  - means? variance?
- latent variables η have to be scaled
- set means to zero
- either fix variance to 1 *or* one loading per factor to 1
  - since we are extending to multiple groups and may want to investigate variance differences between groups, we fix one loading to 1

Practical 1: Single group phenotypic CFA model

Irene...

Extension to multiple groups

- the model can be fitted simultaneously in multiple groups
- the model is fitted to the means and covariances of the observed variables in each group g

$$E(\mathbf{Y_g}) = \nu_g + \Lambda_g \alpha_g$$
$$Cov(\mathbf{Y_g}) = \Lambda_g \Psi_g \Lambda_g' + \Theta_g$$

  - α_g: vector containing the factor means
  - Ψ_g: covariance matrix of the factors
  - Θ_g: the covariance matrix of the errors
- some constraints are needed to identify the model parameters, factor means and covariances, and error variances

Extension to multiple groups

- the model can be fitted simultaneously in multiple groups
- the model is fitted to the means and covariances of the observed variables in each group g

$$
\begin{aligned}
E(\mathbf{Y_g}) &= \boldsymbol{\nu}_g + \boldsymbol{\Lambda}_g \boldsymbol{\alpha}_g \\
Cov(\mathbf{Y_g}) &= \boldsymbol{\Lambda}_g \boldsymbol{\Psi}_g \boldsymbol{\Lambda}'_g + \boldsymbol{\Theta}_g
\end{aligned}
$$

  - $\boldsymbol{\alpha}_g$: vector containing the factor means
  - $\boldsymbol{\Psi}_g$: covariance matrix of the factors
  - $\boldsymbol{\Theta}_g$: the covariance matrix of the errors

- some constraints are needed to identify the model parameters, factor means and covariances, and error variances

Extension to multiple groups

- the model can be fitted simultaneously in multiple groups
- the model is fitted to the means and covariances of the observed variables in each group g

$$\begin{aligned} E(\mathbf{Y_g}) &= \boldsymbol{\nu}_g + \boldsymbol{\Lambda}_g \boldsymbol{\alpha}_g \\ Cov(\mathbf{Y_g}) &= \boldsymbol{\Lambda}_g \boldsymbol{\Psi}_g \boldsymbol{\Lambda}_g' + \boldsymbol{\Theta}_g \end{aligned}$$

  - $\boldsymbol{\alpha}_g$: vector containing the factor means
  - $\boldsymbol{\Psi}_g$: covariance matrix of the factors
  - $\boldsymbol{\Theta}_g$: the covariance matrix of the errors

- some constraints are needed to identify the model parameters, factor means and covariances, and error variances

Constraining the model

- constraints to identify parameters
  - scale the latent variables in both groups
  - identify the mean model
- the resulting model permits groups to differ with respect to most model parameters
  - intercepts, loadings, error variances, factor means and variances
- question is whether we can compare groups with respect to the latent variables if all these parameters differ across groups

Constraining the model

- constraints to identify parameters
  - scale the latent variables in both groups
  - identify the mean model
- the resulting model permits groups to differ with respect to most model parameters
  - intercepts, loadings, error variances, factor means and variances
- question is whether we can compare groups with respect to the latent variables if all these parameters differ across groups

Constraining the model

- constraints to identify parameters
  - scale the latent variables in both groups
  - identify the mean model
- the resulting model permits groups to differ with respect to most model parameters
  - intercepts, loadings, error variances, factor means and variances
- question is whether we can compare groups with respect to the latent variables if all these parameters differ across groups

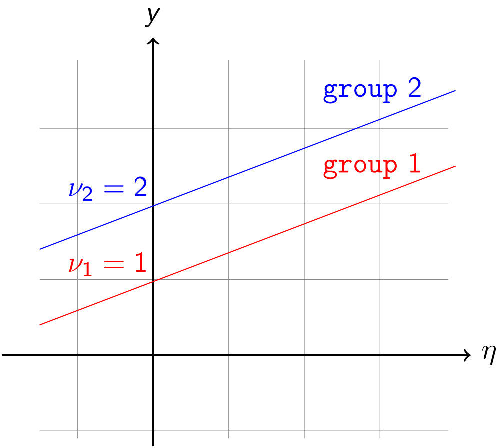
Constraining the model

- constraints to identify parameters
  - scale the latent variables in both groups
  - identify the mean model
- the resulting model permits groups to differ with respect to most model parameters
  - intercepts, loadings, error variances, factor means and variances
- question is whether we can compare groups with respect to the latent variables if all these parameters differ across groups

Practical 2: Two-group phenotypic CFA model

Irene...

Illustration: Group-specific intercepts



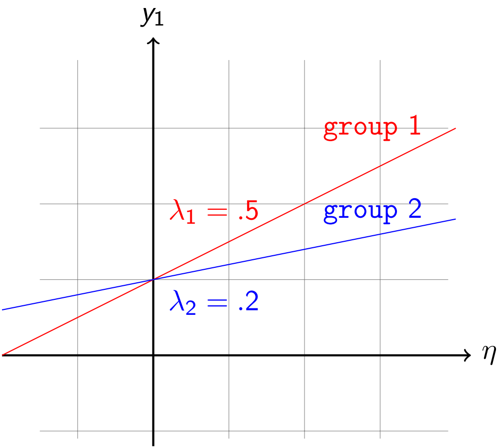
The groups differ with respect to the intercept ν.

$$E(Y_1) = 1 + \lambda\alpha$$
$$E(Y_2) = 2 + \lambda\alpha$$

Given the same score on η group 2 scores on average 1 unit higher on Y.

Consequence: We can't use Y to compare the groups with respect to η if the groups have different intercepts ν.

Illustration: Group-specific loadings



The groups differ with respect to the loading λ.

$$E(Y_1) = \nu + .5\alpha$$
$$E(Y_2) = \nu + .2\alpha$$

Given the same score on η group 2 scores on average increasingly higher (for positive η.

Consequence: We can't use Y to compare the groups with respect to η if the groups have different loadings λ.

Group-specific residual variances

- residual variance is the variability of the observed score Y at a given level of the latent variable η
- if groups differ, then one group has more variability on Y (i.e., more extreme scores) at each given level of η
- if we use Y to estimate η, the precision in that group is less
  - say we use Y to select subjects above a given level of η, then we make more mistakes in that group compared to the group with less residual variance
- it is sometimes argued that differences in residual variance are less problematic than differences in intercepts and loadings when comparing groups with respect to η

Group-specific residual variances

- residual variance is the variability of the observed score Y at a given level of the latent variable η
- if groups differ, then one group has more variability on Y (i.e., more extreme scores) at each given level of η
- if we use Y to estimate η, the precision in that group is less
  - say we use Y to select subjects above a given level of η, then we make more mistakes in that group compared to the group with less residual variance
- it is sometimes argued that differences in residual variance are less problematic than differences in intercepts and loadings when comparing groups with respect to η

Group-specific residual variances

- residual variance is the variability of the observed score Y at a given level of the latent variable η
- if groups differ, then one group has more variability on Y (i.e., more extreme scores) at each given level of η
- if we use Y to estimate η, the precision in that group is less
  - say we use Y to select subjects above a given level of η, then we make more mistakes in that group compared to the group with less residual variance
- it is sometimes argued that differences in residual variance are less problematic than differences in intercepts and loadings when comparing groups with respect to η

Group-specific residual variances

- residual variance is the variability of the observed score Y at a given level of the latent variable η
- if groups differ, then one group has more variability on Y (i.e., more extreme scores) at each given level of η
- if we use Y to estimate η, the precision in that group is less
  - say we use Y to select subjects above a given level of η, then we make more mistakes in that group compared to the group with less residual variance
- it is sometimes argued that differences in residual variance are less problematic than differences in intercepts and loadings when comparing groups with respect to η

Group-specific residual variances

- residual variance is the variability of the observed score Y at a given level of the latent variable η
- if groups differ, then one group has more variability on Y (i.e., more extreme scores) at each given level of η
- if we use Y to estimate η, the precision in that group is less
  - say we use Y to select subjects above a given level of η, then we make more mistakes in that group compared to the group with less residual variance
- it is sometimes argued that differences in residual variance are less problematic than differences in intercepts and loadings when comparing groups with respect to η

Group-specific factor means and factor (co)variances

- these are the differences we are interested in when comparing groups!
- recall, the latent variables represent the theoretical constructs
- if groups don't differ with respect to intercepts, loadings, and residual variances, then observed scores are not biased, and factor mean and (co)variance differences can be investigated

Group-specific factor means and factor (co)variances

- these are the differences we are interested in when comparing groups!
- recall, the latent variables represent the theoretical constructs
- if groups don't differ with respect to intercepts, loadings, and residual variances, then observed scores are not biased, and factor mean and (co)variance differences can be investigated

Group-specific factor means and factor (co)variances

- these are the differences we are interested in when comparing groups!
- recall, the latent variables represent the theoretical constructs
- if groups don't differ with respect to intercepts, loadings, and residual variances, then observed scores are not biased, and factor mean and (co)variance differences can be investigated

Outline

Definition of Measurement Invariance

$$f(y|\eta, s) = f(y|\eta)$$

- the distribution of the observed variable Y given the latent variables η and a grouping variable s equals the distribution of the observed variable Y given the latent variables η alone
  - in other words, the distribution of Y conditional on the latent variables does not differ across groups

- since the definition concerns the distribution of Y conditional on η, the distribution of η may differ across groups

Definition of Measurement Invariance

$$\boxed{f(y|\boldsymbol{\eta}, s) = f(y|\boldsymbol{\eta})}$$

- the distribution of the observed variable Y given the latent variables $\boldsymbol{\eta}$ and a grouping variable s equals the distribution of the observed variable Y given the latent variables $\boldsymbol{\eta}$ alone
  - in other words, the distribution of Y conditional on the latent variables does not differ across groups
- since the definition concerns the distribution of Y conditional on η, the distribution of η may differ across groups

Summary: MI and the CFA model

- the definition of MI requires the distribution of $Y|\eta$ to be the same for all groups
- Y is assumed to be normally distributed
  - therefore means and covariance matrix of Y are sufficient statistics
- we impose the CFA model on the means and covariance matrix of Y in each group

$$\begin{aligned} E(\mathbf{Y_g}) &= \nu_g + \mathbf{\Lambda}_g E(\alpha_g) \\ Cov(\mathbf{Y_g}) &= \mathbf{\Lambda}_g \mathbf{\Psi}_g \mathbf{\Lambda}_g' + \mathbf{\Theta}_g \end{aligned}$$

- parameters that need to be equal across groups are
  - intercepts ν, loadings $\mathbf{\Lambda}$, residual variances $\mathbf{\Theta}$
- this confirms our more conceptual prior considerations

Summary: MI and the CFA model

- the definition of MI requires the distribution of $Y|\eta$ to be the same for all groups
- Y is assumed to be normally distributed
  - therefore means and covariance matrix of Y are sufficient statistics
- we impose the CFA model on the means and covariance matrix of Y in each group

$$\begin{aligned} E(\mathbf{Y_g}) &= \nu_g + \mathbf{\Lambda}_g E(\alpha_g) \\ Cov(\mathbf{Y_g}) &= \mathbf{\Lambda}_g \mathbf{\Psi}_g \mathbf{\Lambda}_g' + \mathbf{\Theta}_g \end{aligned}$$

- parameters that need to be equal across groups are
  - intercepts ν, loadings $\mathbf{\Lambda}$, residual variances $\mathbf{\Theta}$
- this confirms our more conceptual prior considerations

Summary: MI and the CFA model

- the definition of MI requires the distribution of $Y|\eta$ to be the same for all groups
- Y is assumed to be normally distributed
  - therefore means and covariance matrix of Y are sufficient statistics
- we impose the CFA model on the means and covariance matrix of Y in each group

$$\begin{aligned} E(\mathbf{Y_g}) &= \boldsymbol{\nu}_g + \boldsymbol{\Lambda}_g E(\boldsymbol{\alpha}_g) \\ Cov(\mathbf{Y_g}) &= \boldsymbol{\Lambda}_g \boldsymbol{\Psi}_g \boldsymbol{\Lambda}'_g + \boldsymbol{\Theta}_g \end{aligned}$$

- parameters that need to be equal across groups are
  - intercepts ν, loadings $\boldsymbol{\Lambda}$, residual variances $\boldsymbol{\Theta}$
- this confirms our more conceptual prior considerations

Summary: MI and the CFA model

- the definition of MI requires the distribution of $Y|\eta$ to be the same for all groups
- Y is assumed to be normally distributed
  - therefore means and covariance matrix of Y are sufficient statistics
- we impose the CFA model on the means and covariance matrix of Y in each group

$$\begin{aligned} E(\mathbf{Y_g}) &= \boldsymbol{\nu}_g + \boldsymbol{\Lambda}_g E(\boldsymbol{\alpha}_g) \\ Cov(\mathbf{Y_g}) &= \boldsymbol{\Lambda}_g \boldsymbol{\Psi}_g \boldsymbol{\Lambda}_g' + \boldsymbol{\Theta}_g \end{aligned}$$

- parameters that need to be equal across groups are
  - intercepts $\boldsymbol{\nu}$, loadings $\boldsymbol{\Lambda}$, residual variances $\boldsymbol{\Theta}$
- this confirms our more conceptual prior considerations

Summary: MI and the CFA model

- the definition of MI requires the distribution of $Y|\eta$ to be the same for all groups
- Y is assumed to be normally distributed
  - therefore means and covariance matrix of Y are sufficient statistics
- we impose the CFA model on the means and covariance matrix of Y in each group

$$
\begin{aligned}
E(\mathbf{Y_g}) &= \boldsymbol{\nu}_g + \boldsymbol{\Lambda}_g E(\boldsymbol{\alpha}_g) \\
Cov(\mathbf{Y_g}) &= \boldsymbol{\Lambda}_g \boldsymbol{\Psi}_g \boldsymbol{\Lambda}_g' + \boldsymbol{\Theta}_g
\end{aligned}
$$

- parameters that need to be equal across groups are
  - intercepts $\boldsymbol{\nu}$, loadings $\boldsymbol{\Lambda}$, residual variances $\boldsymbol{\Theta}$
- this confirms our more conceptual prior considerations

Outline

Testing MI in the CFA model (1)

- three consecutive tests using model comparisons
  - the three tests correspond to testing invariance of loadings, intercepts, and residual variances
- starting with the most lenient multi-group model (M_1), a series of four increasingly restrictive models can be fitted (M_1-M_4)
  - each of the three tests compare model M_k to the more lenient model M_{k-1} using a likelihood ratio test
- if the more constrained model is tenable, the next model in the series is fitted and compared to the previous model

Testing MI in the CFA model (1)

- three consecutive tests using model comparisons
  - the three tests correspond to testing invariance of loadings, intercepts, and residual variances
- starting with the most lenient multi-group model (M_1), a series of four increasingly restrictive models can be fitted (M_1-M_4)
  - each of the three tests compare model M_k to the more lenient model M_{k-1} using a likelihood ratio test
- if the more constrained model is tenable, the next model in the series is fitted and compared to the previous model

Testing MI in the CFA model (1)

- three consecutive tests using model comparisons
  - the three tests correspond to testing invariance of loadings, intercepts, and residual variances
- starting with the most lenient multi-group model (M_1), a series of four increasingly restrictive models can be fitted (M_1-M_4)
  - each of the three tests compare model M_k to the more lenient model M_{k-1} using a likelihood ratio test
- if the more constrained model is tenable, the next model in the series is fitted and compared to the previous model

Testing MI in the CFA model (1)

- three consecutive tests using model comparisons
  - the three tests correspond to testing invariance of loadings, intercepts, and residual variances
- starting with the most lenient multi-group model (M_1), a series of four increasingly restrictive models can be fitted (M_1-M_4)
  - each of the three tests compare model M_k to the more lenient model M_{k-1} using a likelihood ratio test
- if the more constrained model is tenable, the next model in the series is fitted and compared to the previous model

Testing MI in the CFA model (1)

- three consecutive tests using model comparisons
  - the three tests correspond to testing invariance of loadings, intercepts, and residual variances
- starting with the most lenient multi-group model (M_1), a series of four increasingly restrictive models can be fitted (M_1-M_4)
  - each of the three tests compare model M_k to the more lenient model M_{k-1} using a likelihood ratio test
- if the more constrained model is tenable, the next model in the series is fitted and compared to the previous model

Testing MI in the CFA model (2)

- the models are
  - M_1: intercepts, loadings, and residual variances are group specific, factor means are fixed to zero for identification
  - M_2: loadings are fixed to be equal across groups, the rest remains as in M_1
  - M_3 intercepts are fixed to be equal across groups, factor means are fixed to zero in one group and estimated in all other groups, rest as in M_2
  - M_4 residual variances are fixed to be equal across groups, rest as in M_3

Testing MI in the CFA model (2)

- the models are
  - M_1: intercepts, loadings, and residual variances are group specific, factor means are fixed to zero for identification
  - M_2: loadings are fixed to be equal across groups, the rest remains as in M_1
  - M_3 intercepts are fixed to be equal across groups, factor means are fixed to zero in one group and estimated in all other groups, rest as in M_2
  - M_4 residual variances are fixed to be equal across groups, rest as in M_3

Testing MI in the CFA model (2)

- the models are
  - M_1: intercepts, loadings, and residual variances are group specific, factor means are fixed to zero for identification
  - M_2: loadings are fixed to be equal across groups, the rest remains as in M_1
  - M_3 intercepts are fixed to be equal across groups, factor means are fixed to zero in one group and estimated in all other groups, rest as in M_2
  - M_4 residual variances are fixed to be equal across groups, rest as in M_3

Testing MI in the CFA model (2)

- the models are
  - M_1: intercepts, loadings, and residual variances are group specific, factor means are fixed to zero for identification
  - M_2: loadings are fixed to be equal across groups, the rest remains as in M_1
  - M_3 intercepts are fixed to be equal across groups, factor means are fixed to zero in one group and estimated in all other groups, rest as in M_2
  - M_4 residual variances are fixed to be equal across groups, rest as in M_3

Practical 3: Testing MI in the phenotypic CFA model

Irene...

Outline

Categorical data

- so far we assumed that **Y** is normally distributed
- more common: categorical outcomes
  - binary outcomes: endorsing a symptom, checklists
  - Likert data
- two approaches to model categorical data
  - latent response variable (or Y^*) approach (LRV)
  - item response theory approach
- the two approaches are equivalent for proportional odds model

Categorical data

- so far we assumed that **Y** is normally distributed
- more common: categorical outcomes
  - binary outcomes: endorsing a symptom, checklists
  - Likert data
- two approaches to model categorical data
  - latent response variable (or Y^*) approach (LRV)
  - item response theory approach
- the two approaches are equivalent for proportional odds model

Categorical data

- so far we assumed that **Y** is normally distributed
- more common: categorical outcomes
  - binary outcomes: endorsing a symptom, checklists
  - Likert data
- two approaches to model categorical data
  - latent response variable (or Y^*) approach (LRV)
  - item response theory approach
- the two approaches are equivalent for proportional odds model

Categorical data

- so far we assumed that **Y** is normally distributed
- more common: categorical outcomes
  - binary outcomes: endorsing a symptom, checklists
  - Likert data
- two approaches to model categorical data
  - latent response variable (or Y^*) approach (LRV)
  - item response theory approach
- the two approaches are equivalent for proportional odds model

Latent Response Variable approach

- latent response variable (LRV) formulation
  - see Agresti, Categorical Data analysis
- assumes that the 'true' response to an item is normally distributed
  - the 'true' response is not observed but latent: LRV
  - we observe a categorized version of the LRV
- threshold structure is imposed on the the LRV
  - denote the LRV as Y^*, then for $C = 1, \ldots, c$ response categories we have $c + 1$ thresholds τ, where $\tau_1 = -\infty$ and $\tau_{c+1} = \infty$ and

$$y = c \quad \text{if} \quad \tau_{c-1} < y^* \leq \tau_{c+1}$$

Latent Response Variable approach

- latent response variable (LRV) formulation
  - see Agresti, Categorical Data analysis
- assumes that the ‘true’ response to an item is normally distributed
  - the ‘true’ response is not observed but latent: LRV
  - we observe a categorized version of the LRV
- threshold structure is imposed on the the LRV
  - denote the LRV as Y^*, then for $C = 1, \ldots, c$ response categories we have $c + 1$ thresholds τ, where $\tau_1 = -\infty$ and $\tau_{c+1} = \infty$ and

$$y = c \quad \text{if} \quad \tau_{c-1} < y^* \leq \tau_{c+1}$$

Latent Response Variable approach

- latent response variable (LRV) formulation
  - see Agresti, Categorical Data analysis
- assumes that the 'true' response to an item is normally distributed
  - the 'true' response is not observed but latent: LRV
  - we observe a categorized version of the LRV
- threshold structure is imposed on the the LRV
  - denote the LRV as Y^*, then for $C = 1, \ldots, c$ response categories we have $c + 1$ thresholds τ, where $\tau_1 = -\infty$ and $\tau_{c+1} = \infty$ and

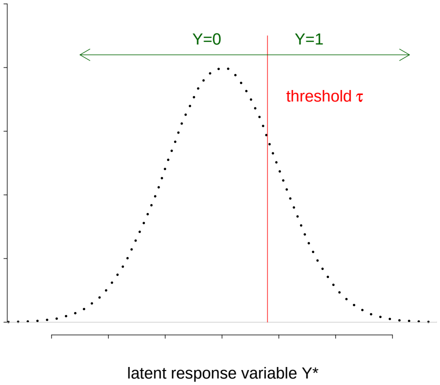
$$y = c \quad \text{if} \quad \tau_{c-1} < y^* \leq \tau_{c+1}$$

Latent Response Variable approach

- latent response variable (LRV) formulation
  - see Agresti, Categorical Data analysis
- assumes that the 'true' response to an item is normally distributed
  - the 'true' response is not observed but latent: LRV
  - we observe a categorized version of the LRV
- threshold structure is imposed on the the LRV
  - denote the LRV as Y^*, then for $C = 1, \ldots, c$ response categories we have $c + 1$ thresholds τ, where $\tau_1 = -\infty$ and $\tau_{c+1} = \infty$ and

$$\boxed{y = c \quad \texttt{if} \quad \tau_{c-1} < y^* \leq \tau_{c+1}}$$

Y^*, τ, and categorical Y



Outline

Item Response Theory model

- the probability $P(Y=1) \equiv P(Y^* > \tau)$ is modeled directly
- rather than using linear regression to relate Y^* to the underlying factor, IRT uses logistic regression to model the probabilities
  - predicted value of a linear regression ranges from $-\infty$ to ∞
  - logistic regression stays between 0 and 1
  - convenient when modeling probabilities
- the equation for a 2-p IRT model for binary data is

$$P(Y=1|\theta) = \frac{1}{1+exp[-b(\eta-a)]}$$

  - a is the discrimination parameter
  - b is the difficulty parameter
- parameters a and b are related to thresholds, loadings, and residual variances of the LRV model
  - see Webnote 4, Muthen & Asparouhov, www.statmodel.com

Item Response Theory model

- the probability $P(Y=1) \equiv P(Y^* > \tau)$ is modeled directly
- rather than using linear regression to relate Y^* to the underlying factor, IRT uses logistic regression to model the probabilities
  - predicted value of a linear regression ranges from $-\infty$ to ∞
  - logistic regression stays between 0 and 1
  - convenient when modeling probabilities
- the equation for a 2-p IRT model for binary data is

$$P(Y=1|\theta) = \frac{1}{1+exp[-b(\eta-a)]}$$

  - a is the discrimination parameter
  - b is the difficulty parameter
- parameters a and b are related to thresholds, loadings, and residual variances of the LRV model
  - see Webnote 4, Muthen & Asparouhov, www.statmodel.com

Item Response Theory model

- the probability $P(Y=1) \equiv P(Y^* > \tau)$ is modeled directly
- rather than using linear regression to relate Y^* to the underlying factor, IRT uses logistic regression to model the probabilities
  - predicted value of a linear regression ranges from $-\infty$ to ∞
  - logistic regression stays between 0 and 1
  - convenient when modeling probabilities
- the equation for a 2-p IRT model for binary data is

$$P(Y=1|\theta) = \frac{1}{1+exp[-b(\eta - a)]}$$

  - a is the discrimination parameter
  - b is the difficulty parameter
- parameters a and b are related to thresholds, loadings, and residual variances of the LRV model
  - see Webnote 4, Muthen & Asparouhov, www.statmodel.com

Item Response Theory model

- the probability $P(Y = 1) \equiv P(Y^* > \tau)$ is modeled directly
- rather than using linear regression to relate Y^* to the underlying factor, IRT uses logistic regression to model the probabilities
  - predicted value of a linear regression ranges from $-\infty$ to ∞
  - logistic regression stays between 0 and 1
  - convenient when modeling probabilities
- the equation for a 2-p IRT model for binary data is

$$P(Y = 1|\theta) = \frac{1}{1 + exp[-b(\eta - a)]}$$

  - a is the discrimination parameter
  - b is the difficulty parameter
- parameters a and b are related to thresholds, loadings, and residual variances of the LRV model
  - see Webnote 4, Muthen & Asparouhov, www.statmodel.com

Item Response Theory model

- the probability $P(Y=1) \equiv P(Y^* > \tau)$ is modeled directly
- rather than using linear regression to relate Y^* to the underlying factor, IRT uses logistic regression to model the probabilities
  - predicted value of a linear regression ranges from $-\infty$ to ∞
  - logistic regression stays between 0 and 1
  - convenient when modeling probabilities
- the equation for a 2-p IRT model for binary data is

$$P(Y=1|\theta) = \frac{1}{1+exp[-b(\eta - a)]}$$

  - a is the discrimination parameter
  - b is the difficulty parameter
- parameters a and b are related to thresholds, loadings, and residual variances of the LRV model
  - see Webnote 4, Muthen & Asparouhov, www.statmodel.com

MI in CFA and IRT models
Lubke, Rebollo
Motivation
Comparing groups
Latent Variable Models
CFA and MI
Multi-group CFA
Theory of MI
Testing MI
Categorical data: LRV and IRT
LRV
IRT
Summary

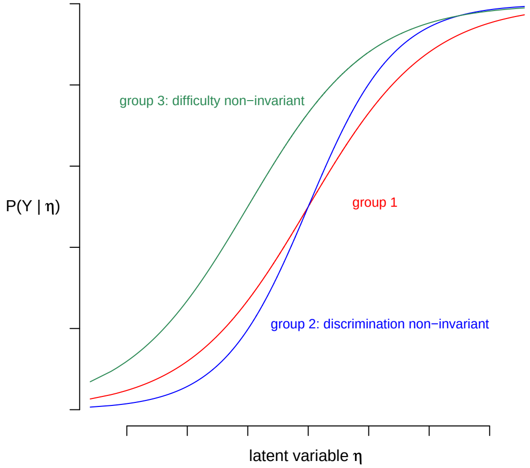
Item Response Theory model

- the probability $P(Y=1) \equiv P(Y^* > \tau)$ is modeled directly
- rather than using linear regression to relate Y^* to the underlying factor, IRT uses logistic regression to model the probabilities
  - predicted value of a linear regression ranges from $-\infty$ to ∞
  - logistic regression stays between 0 and 1
  - convenient when modeling probabilities
- the equation for a 2-p IRT model for binary data is

$$P(Y=1|\theta) = \frac{1}{1 + exp[-b(\eta - a)]}$$

  - a is the discrimination parameter
  - b is the difficulty parameter
- parameters a and b are related to thresholds, loadings, and residual variances of the LRV model
  - see Webnote 4, Muthen & Asparouhov, www.statmodel.com

Group differences in $P(Y|\eta)$



Summary

- Latent variable models structure the relations between observed variables and underlying latent variables
- the latent variables represent the theoretical constructs
- group comparisons with respect to the latent variables is possible `iff` measurement invariance holds
- MI can be easily tested for continuous data and IRT
  - a bit more thought needed in case of LRV approach
- tests do not depend on the type of measurement model

- Outlook
  - extension to genetic decomposition models (ACE type) is straightforward
  - either decompose the factors η (common pathway model)
  - or decompose all observed variables $\mathbf{Y}$ (independent pathway model)

Summary

MI in CFA and IRT models
Lubke, Rebollo
Motivation
Comparing groups
Latent Variable Models
CFA and MI
Multi-group CFA
Theory of MI
Testing MI
Categorical data: LRV and IRT
LRV
IRT
Summary

- Latent variable models structure the relations between observed variables and underlying latent variables
- the latent variables represent the theoretical constructs
- group comparisons with respect to the latent variables is possible `iff` measurement invariance holds
- MI can be easily tested for continuous data and IRT
  - a bit more thought needed in case of LRV approach
- tests do not depend on the type of measurement model

- Outlook
  - extension to genetic decomposition models (ACE type) is straightforward
  - either decompose the factors η (common pathway model)
  - or decompose all observed variables $\mathbf{Y}$ (independent pathway model)

Summary

- Latent variable models structure the relations between observed variables and underlying latent variables
- the latent variables represent the theoretical constructs
- group comparisons with respect to the latent variables is possible `iff` measurement invariance holds
- MI can be easily tested for continuous data and IRT
  - a bit more thought needed in case of LRV approach
- tests do not depend on the type of measurement model

- Outlook
  - extension to genetic decomposition models (ACE type) is straightforward
  - either decompose the factors η (common pathway model)
  - or decompose all observed variables $\mathbf{Y}$ (independent pathway model)

Summary

- Latent variable models structure the relations between observed variables and underlying latent variables
- the latent variables represent the theoretical constructs
- group comparisons with respect to the latent variables is possible `iff` measurement invariance holds
- MI can be easily tested for continuous data and IRT
    - a bit more thought needed in case of LRV approach
- tests do not depend on the type of measurement model

- Outlook
    - extension to genetic decomposition models (ACE type) is straightforward
    - either decompose the factors η (common pathway model)
    - or decompose all observed variables $\mathbf{Y}$ (independent pathway model)

Summary

- Latent variable models structure the relations between observed variables and underlying latent variables
- the latent variables represent the theoretical constructs
- group comparisons with respect to the latent variables is possible `iff` measurement invariance holds
- MI can be easily tested for continuous data and IRT
  - a bit more thought needed in case of LRV approach
- tests do not depend on the type of measurement model

- Outlook
  - extension to genetic decomposition models (ACE type) is straightforward
  - either decompose the factors η (common pathway model)
  - or decompose all observed variables $\mathbf{Y}$ (independent pathway model)