# LATENT VARIABLE MODELS

## 4: Ordered categorical indicator variables

# Today’s topics

- Binary items
  - least-squares type estimation (factor analysis)
  - ML estimation (IRT)
  - 1PL (Rasch) vs. 2PL model: Same or different loadings between items?
- Generalization to ordered-categorical items

# Ordered categorical indicator variables

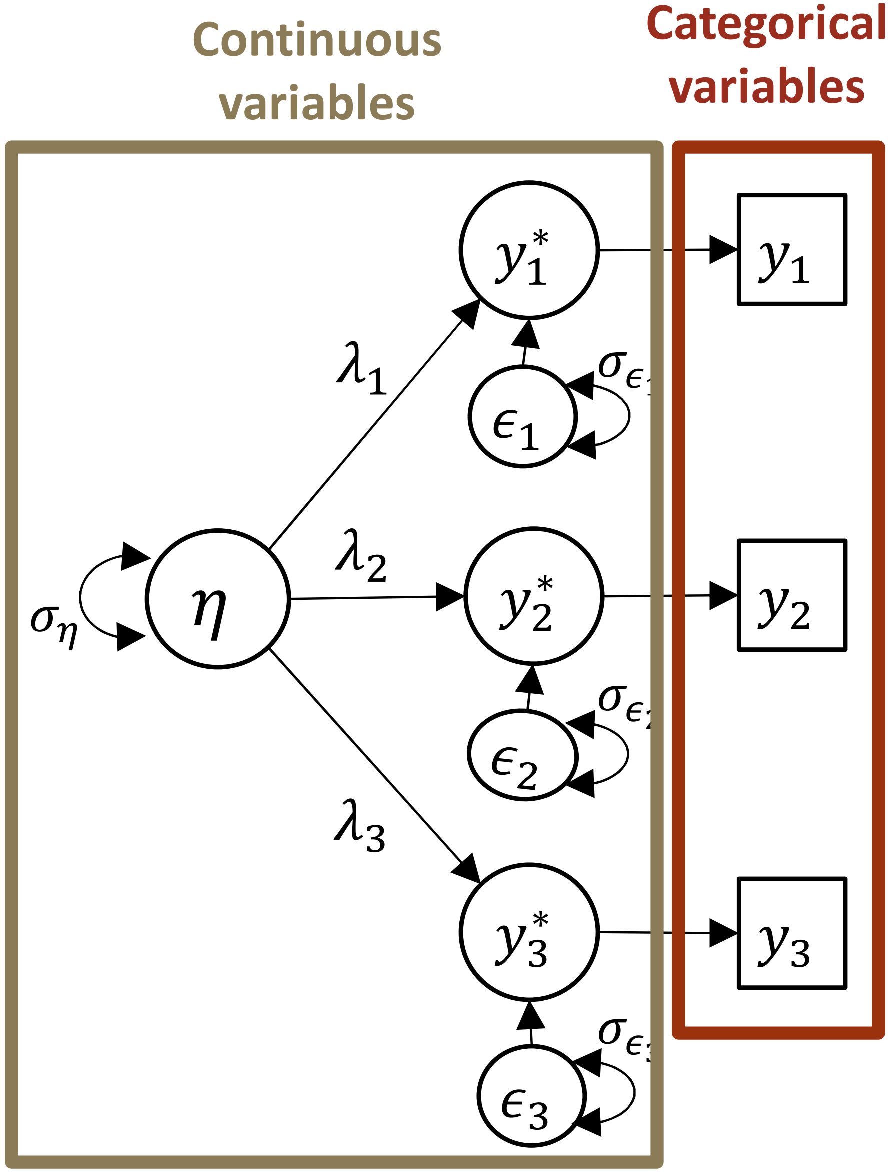
- So far, endogenous variables have always been continuous
- Often variables in psychology are (ordered) categorical
- Need different model
  - Comparable to having binary or ordered categorical response variable in regression: Have to use e.g., logistic or probit regression

# Ordered categorical indicator variables

- The regression formula for the continuous item response of person *j* on item *i* is given by:
  $Y_{ij} = \tau_i + \lambda_i \eta_j + \epsilon_{ij}$
- A dichotomous (categorical) response $Y_{ij}$ can only take values 0 or 1
  - Or 0, 1, 2, … for > 2 ordered categorical values

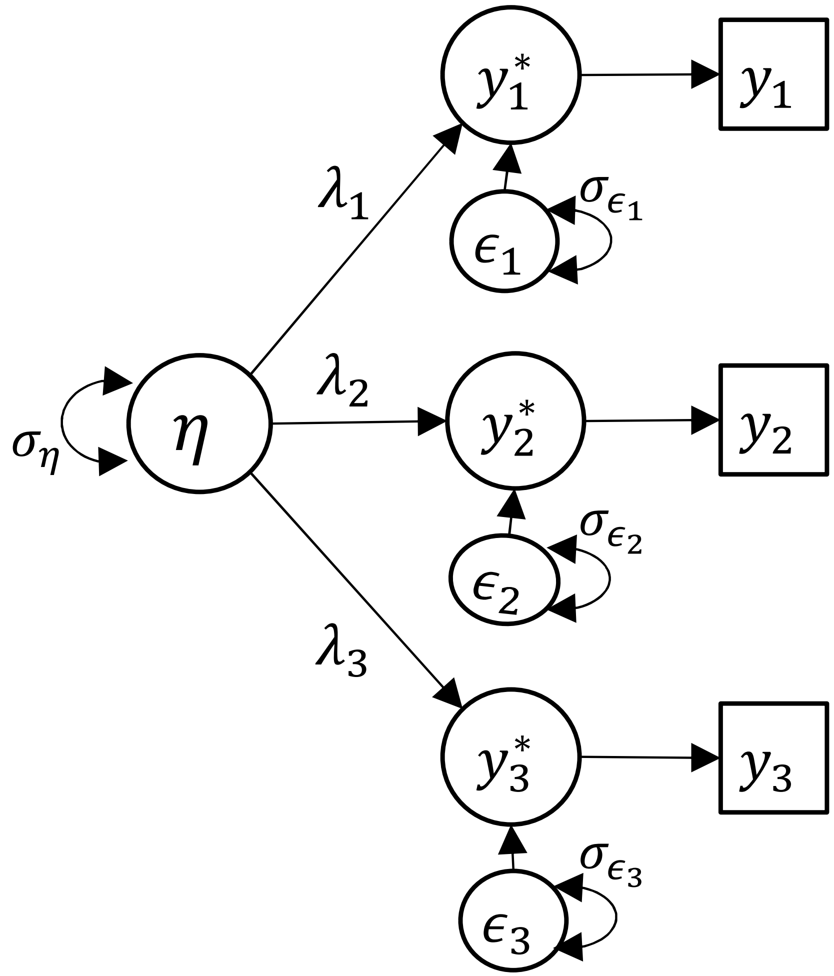
# Ordered categorical indicator variables

- The regression formula for the continuous item response of person *j* on item *i* is given by: $Y_{ij} = \tau_i + \lambda_i \eta_j + \epsilon_{ij}$
- A dichotomous (categorical) response $Y_{ij}$ can only take values 0 or 1
- Solution: we assume a continuous LV $Y^*_{ij}$ underlies categorical item response $Y_{ij}$
- $Y^*_{ij}$ is linearly dependent on $\eta$



# Ordered categorical indicator variables

- We assume a continuous LV $Y_{ij}^*$ underlies categorical item response $Y_{ij}$
- Categorical response $Y_{ij}$ has a threshold $\tau_i$
  - If $Y_{ij}^* < \tau_i$ then $p(Y_{ij} = 1) < .5$
  - If $Y_{ij}^* > \tau_i$ then $p(Y_{ij} = 1) > .5$
- Assumptions:
  - All LVs follow a normale distribution
- In other words: Compute tetra- or polychoric correlation matrix and perform CFA as in the linear / continuous variable case
  - LS (least-squares) approach
  - Very similar to *probit* regression

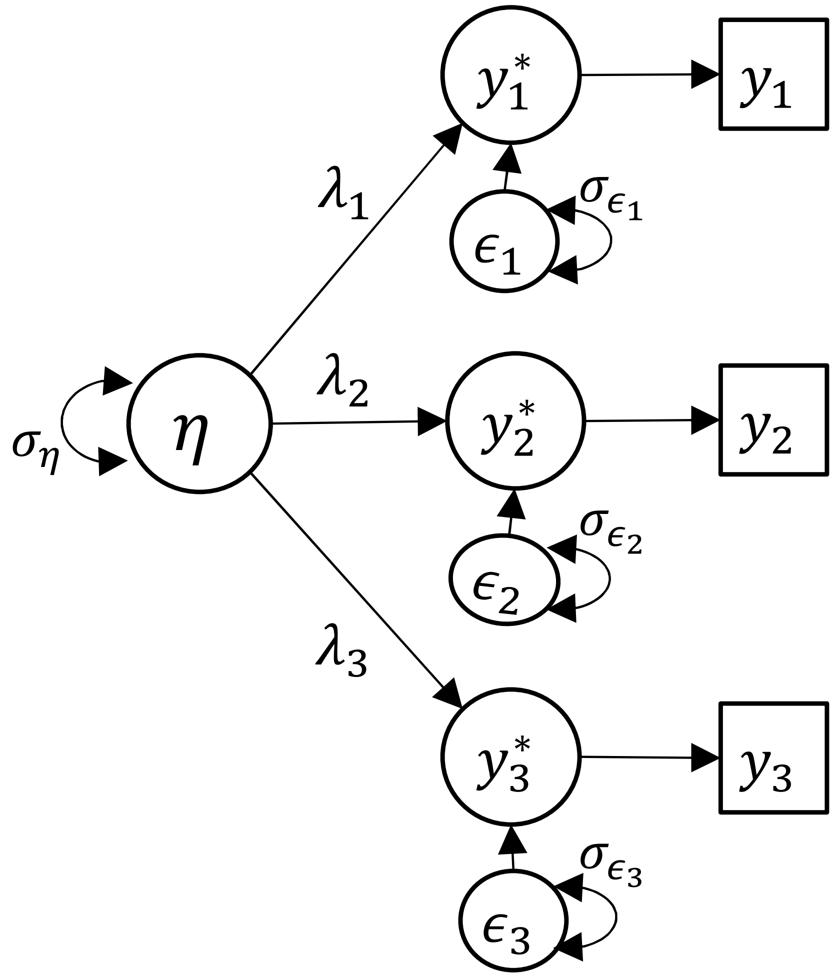


# Ordered categorical indicator variables

- We assume a continuous LV $Y_{ij}^*$ underlies categorical item response $Y_{ij}$
- Categorical response $Y_{ij}$ has a threshold $\tau_i$
  - If $Y_{ij}^* < \tau_i$ then $p(Y_{ij} = 1) < .5$
  - If $Y_{ij}^* > \tau_i$ then $p(Y_{ij} = 1) > .5$
- Assumptions:
  - All LVs follow a normale distribution
- Identification restrictions:
  - All $Y_{ij}^*$ have $\mu = 0$ and $\sigma = 1$, and
  - Standardized LV approach ($\sigma_\eta = 1$)

or

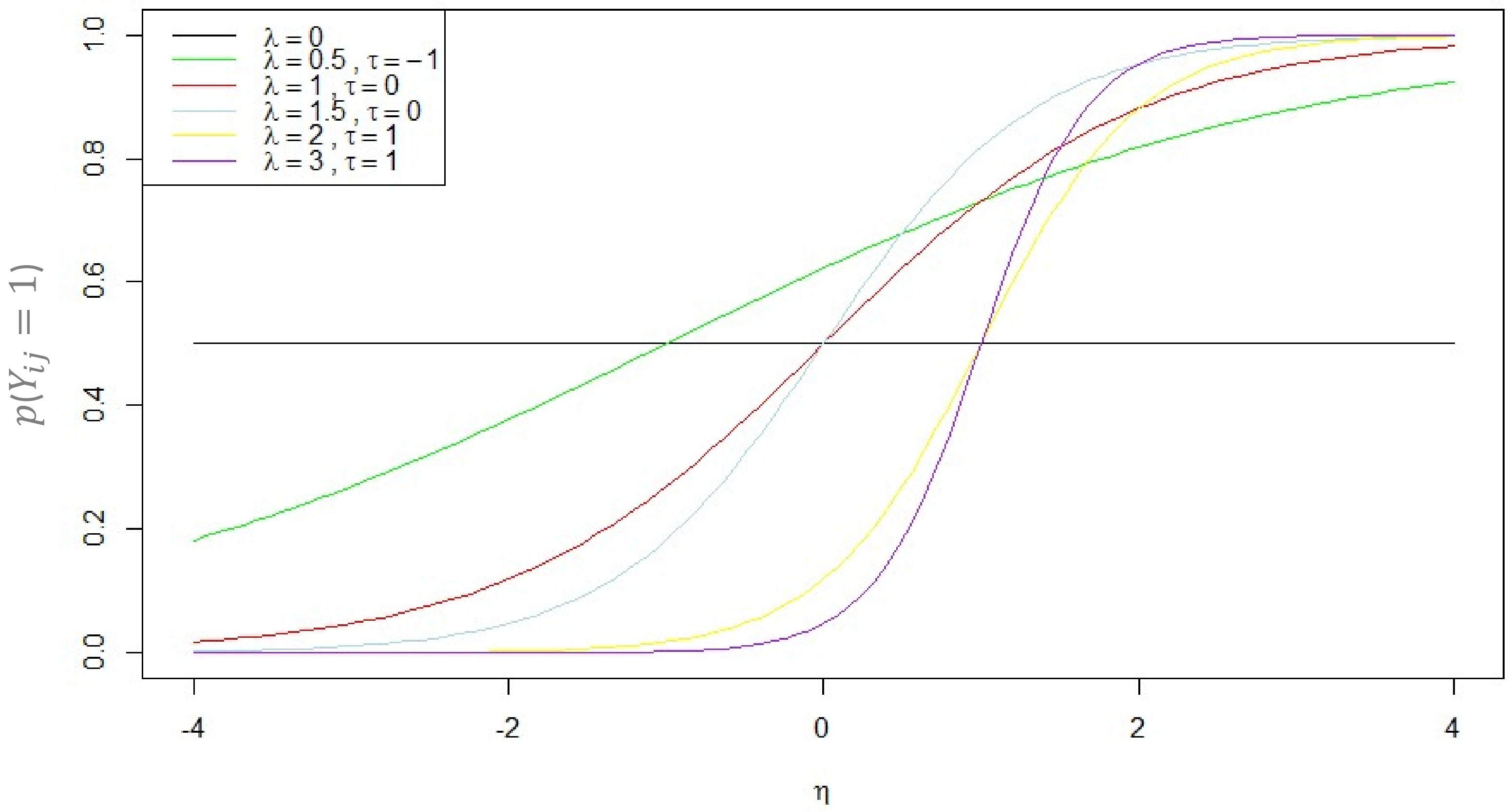
Marker var. approach ($\lambda_1 = 1$)



# Item characteristic curves (ICCs)

To get from latent $\eta$ to observed dichotomous response $Y_{ij}$:

$$p(Y_{ij} = 1|\eta) = \frac{e^{\lambda_i(\eta_j - \tau_i)}}{1 + e^{\lambda_i(\eta_j - \tau_i)}}$$



# Examples and exercises

- Example 6.2 - part I
- Exercise 6.1

# Identifying scale of underlying latent variable

- By definition we have

$$\sigma^2_{y_i} = \lambda_i^2 \sigma_\eta^2 + \sigma^2_{\epsilon_i}$$

$$\Delta_i = \frac{1}{\sigma^2_{y\ i}}$$

- 'Delta', or marginal, parameterization assumes $\sigma^2_{y\ i}$=1

  and thus $\Delta_i$=1 and $\sigma^2_{\epsilon_i} = 1 - \lambda_i^2 \sigma_\eta^2$
- 'Theta', or conditional, parameterization assumes $\sigma^2_{\epsilon_i} = 1$

- Delta parameterization is more natural from FA viewpoint, theta parameterization is more natural from IRT viewpoint

# Categorical FA vs. IRT

Correspondence: Aim is to model association between LV and observed item responses

Historical differences:

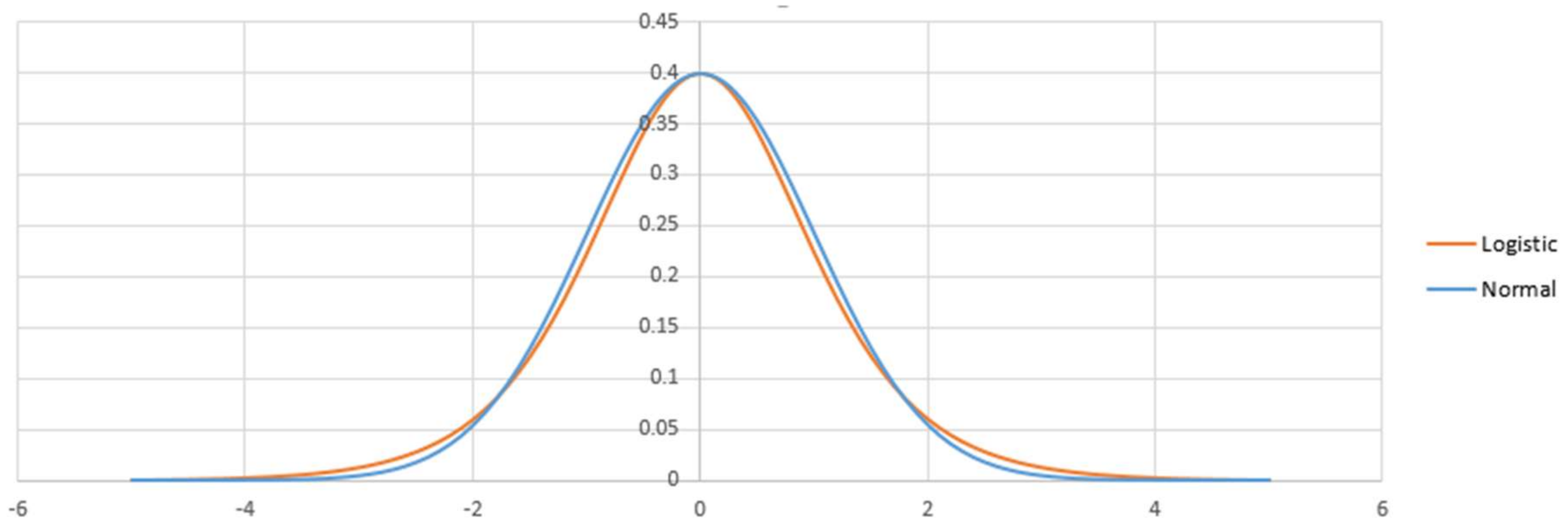
- Estimation:
  - IRT: maximum likelihood (ML)
    - Estimates model parameters in one step
    - Not available for ordered categorical indicators in lavaan
    - Similar to a logistic regression
  - FA: diagonally weighted least squares (DWLS)
    - Estimates tetra- or polychoric correlation matrix, performs continuous variable CFA on that matrix
    - Only option for ordered categorical indicators in lavaan
    - Similar to a probit regression
- Parameterization:
  - Delta parameterization in FA, theta parameterization in IRT
  - In IRT, latent trait often scaled by assuming mean 0 and variance 1
  - In CFA, latent trait often scaled by setting loading of first item to 1
  - Loadings and thresholds in CFA are discrimination and difficulty in IRT

# Scale of common factor: CFA (LS) vs. IRT (ML)

- In IRT, the default is to specify the common factor to have mean 0 and variance 1
  - Other approaches possible
  - IRT parameters: $\alpha_i$ (discrimination) and $\beta_i$ (difficulty)
- In CFA with binary items, we often take the same approach (std.lv = TRUE in lavaan)
  - Other approaches possible
  - CFA parameters: $\lambda_i$ (loading), $\tau_i$ (threshold) and $\sigma_\epsilon^2$ (measurement error variance; which is a function of $\lambda_i$)

# ML (IRT) vs LS (FA) - Logistic vs probit

- Logistic model (employed in ML estimation) assumes binomial error distribution
- Probit model (employed in LS-type estimation) assumes Gaussian error distribution
- Their probability distributions have very similar shapes:



# Logistic vs probit

- To get the logit and the probit to align, the logit's slope must be ≈1.7 times the slope value for the probit
- Philosophical differences:
  - Logistic model assumes the common factor to be *directly* connected to the probability of a correct response
    - Note: in line with ML estimation, where model is estimated in one step
  - Probit model assumes the dichotomous response resulted from a dichotomization of an underlying normally distributed variable
    - Note: in line with LS-type estimation, where we estimate tetrachoric (polychoric) correlation matrix, and then fit a CFA for continuous variables to that correlation matrix

# Confusing?

Many scalings, but interpretation is all alike:

- Values of slope, factor loading, discrimination parameters (λ, a, α) increase together
  - Higher values: better discrimination, stronger indicator, less measurement error
- Values of threshold, difficulty (τ, b, β) increase together
  - Higher values: more difficult (need higher value of latent trait for correct (or affirmative) response)
- Can compute any parameterization from any other parameterization (but may lead to headache)
- Most important:
  - Be aware of existence of different parameterizations
  - Do not directly compare results from different estimators and parameterizations when interpreting models

# Examples and exercises

- Example 6.2 – Part II

- Exercise 6.2 a-d

# IRT models

- Binary items:
  - 1PL, or Rasch model (loadings equal, thresholds free)
  - 2PL (loadings free, thresholds free)
  - …
- Polytomous items
  - Partial credit mo esholds free)
  - Graded response thresholds free)
  - …


Georg Rasch (1901-1980)

# No Rasch, no good?

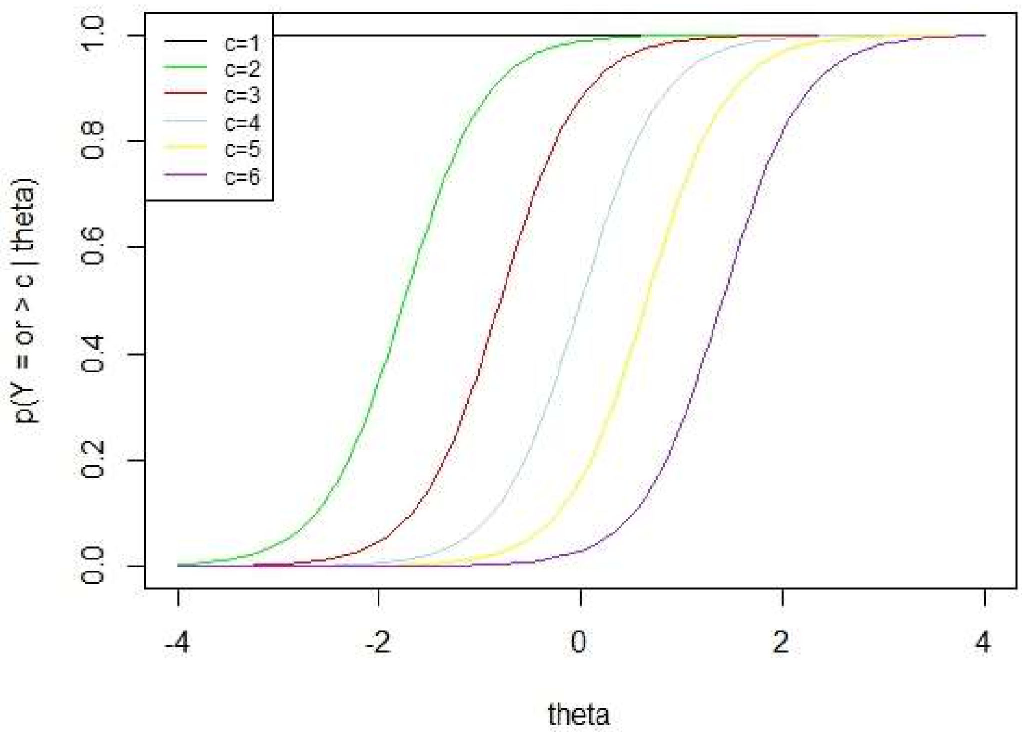
- Often in psychology, we want to use the test score: the (unweighted) sum of item scores
  - Easy to calculate, you need no IRT or SEM software to estimate it
- In the Rasch model, all item loadings are equal, so all item scores contribute equally to estimation of the latent trait
  - Test score is ‘sufficient statistic’ for eta (latent trait)
    - "no other statistic that can be calculated from the same sample provides any additional information as to the value of the parameter"
  - Well-fitting Rasch model: test score contains all information about latent trait

# Examples and exercises

- Example 6.2 – Part III
- Exercise 6.2 e

# Ordinal responses

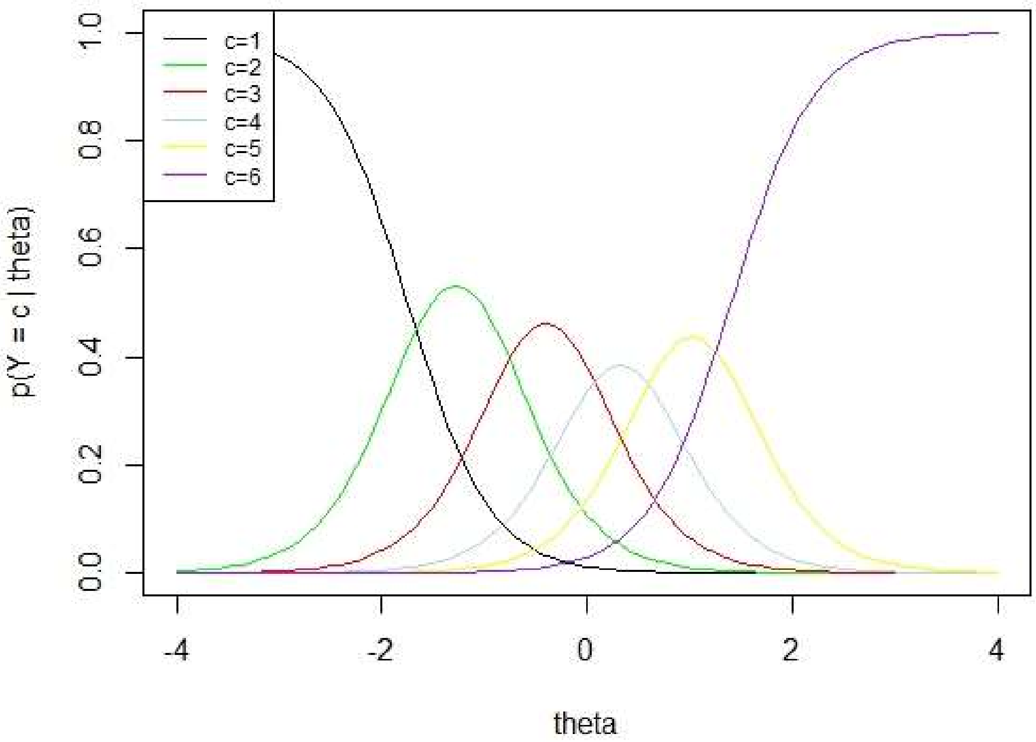
- If we have items with ordered response options: e.g., *a* < *b* < *c* < *d*
- We can model the following probabilities:
  - $p(Y_i \geq a|\eta) = 1$
  - $p(Y_i \geq b|\eta) = \frac{e^{\lambda_i \eta - \tau_{i,b}}}{1+e^{\lambda_i \eta - \tau_{i,b}}}$
  - $p(Y_i \geq c|\eta) = \frac{e^{\lambda_i \eta - \tau_{i,c}}}{1+e^{\lambda_i \eta - \tau_{i,c}}}$
  - $p(Y_i \geq d|\eta) = \frac{e^{\lambda_i \eta - \tau_{i,d}}}{1+e^{\lambda_i \eta - \tau_{i,d}}}$



# Ordinal responses

- If we have items with ordered response options: e.g., *a* < *b* < *c* < *d*
- We can model the following probabilities:
  - $p(Y_i \geq a|\eta) = 1$
  - $p(Y_i \geq b|\eta) = \frac{e^{\lambda_i \eta - \tau_{i,b}}}{1+e^{\lambda_i \eta - \tau_{i,b}}}$
  - $p(Y_i \geq c|\eta) = \frac{e^{\lambda_i \eta - \tau_{i,c}}}{1+e^{\lambda_i \eta - \tau_{i,c}}}$
  - $p(Y_i \geq d|\eta) = \frac{e^{\lambda_i \eta - \tau_{i,d}}}{1+e^{\lambda_i \eta - \tau_{i,d}}}$
- This gives us the probabilities for each response option as follows:
  - $p(Y_i = a|\eta) = p(Y_i \geq a|\eta) - p(Y_i \geq b|\eta)$
  - $p(Y_i = b|\eta) = p(Y_i \geq b|\eta) - p(Y_i \geq c|\eta)$
  - $p(Y_i = c|\eta) = p(Y_i \geq c|\eta) - p(Y_i \geq d|\eta)$
  - $p(Y_i = d|\eta) = p(Y_i \geq d|\eta)$

# Ordinal responses

- If we have ite 

$a < b < c < d$
- We can mode
  - $p(Y_i \geq a|\eta$
  - $p(Y_i \geq b|\eta$
  - $p(Y_i \geq c|\eta)$
  - $p(Y_i \geq d|\eta$
- This gives us the probabilities for each response option as follows:
  - $p(Y_i = a|\eta) = p(Y_i \geq a|\eta) - p(Y_i \geq b|\eta)$
  - $p(Y_i = b|\eta) = p(Y_i \geq b|\eta) - p(Y_i \geq c|\eta)$
  - $p(Y_i = c|\eta) = p(Y_i \geq c|\eta) - p(Y_i \geq d|\eta)$
  - $p(Y_i = d|\eta) = p(Y_i \geq d|\eta)$

# Ordinal responses

- For every item with *k* ordered categories, we need to estimate one loading, and *k-1* thresholds
- In lavaan, we use the same approach as with dichotomous data: use ‘ordered = ….’ argument
  - For every item declared ordered, lavaan checks number of categories, and estimates *k*-1 thresholds

# Ordered-categorical responses

- Partial credit model is the Rasch (1PL) model generalized to polytomous items
  - Same loadings for all items
  - Freely estimates thresholds for all categories and items
- Graded response model is the 2PL model generalized to polytomous items
  - Freely estimates loadings for all items
  - Freely estimates thresholds for all categories and items
- Note: Unlike in Rasch model, in PCM test score is not a sufficient statistic (does not contain all information about) for the latent trait (eta)

# Examples and exercises

- Example 6.2 – Part IV
- Additional Exercise: HADS