

# Item Response Models Part 2: Polytomous Models

## Today's Topics:

- Big Picture Overview
- Ordered Categorical Models → Graded Response
- Maybe Ordered Categorical Models → Partial Credit
- Unordered Categorical Models → Nominal Response

Psyc 948 Class 10b  
1 of 28

# Fundamentals of IRT

- IRT = measurement model in which transformed item responses are predicted using properties of **persons** (Theta) and properties of **items** (difficulty, discrimination)
- “**Rasch models**” are a subset of IRT models with more restrictive slope assumptions...(but don't let Rasch people hear you saying that)
- Items and persons are on the same latent metric: “**conjoint scaling**”
  - Anchor (identify) scale with either persons (z-scored theta) or items
- After controlling for a person's latent trait score (Theta), the item responses should be uncorrelated: “**local independence**”
- Item response models are re-parameterized versions of item factor models (for binary outcomes)
  - Thus, we can now extend IRT to “**polytomous responses**” (3+ options)

Psyc 948 Class 10b  
2 of 28

# Polytomous Items

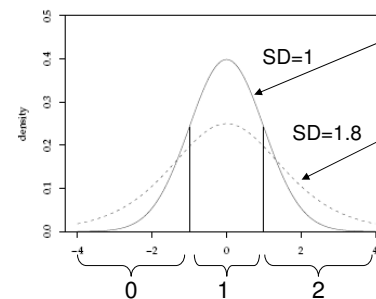
- **Polytomous = more than 2 options**
- Polytomous models are not numbered like binary models, but instead get called different names
  - Most have a 1-PL vs. 2-PL version that go by different names
  - Within those, different constraints on what to do with multiple categories
- **Three main kinds\* of polytomous models:**
  - Outcome is ordered, and I'm sure about that
    - Graded Response or Modified Graded Response Model
  - Outcome could be ordered, but it is a testable hypothesis
    - (Generalized) Partial Credit Model or Rating Scale Model
  - No way is this outcome ordered
    - Nominal Response Model

\* Lots and lots more – these are the major categories

Psyc 948 Class 10b  
3 of 28

# Threshold Concept for Binary and Ordinal Variables

- Each ordinal variable is **really** the chopped-up version of a pretend underlying continuous variable ( $Y^*$ ) with a mean of 0



**Probit** (ogive) model: Pretend variable has a normal distribution (variance = 1)

**Logit** model: Pretend variable has logistic distribution (variance =  $\pi^2/3$ )

Polytomous models will differ in how they make use of multiple (k-1) thresholds per item

# thresholds = # options - 1

Psyc 948 Class 10b  
4 of 28

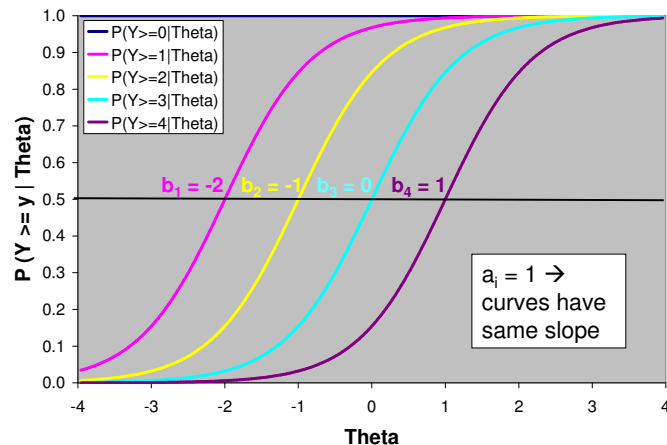
# Graded Response Model (GRM)

- Ideal for items with clear underlying response continuum
- # response options (k) don't have to be same across items
- Is an "indirect" or "difference" model
  - Compute difference between models to get probability of each response
- Estimate 1  $a_i$  per item and k-1 difficulties (4 options → 3 difficulties)
- Models the probability of **any given response category or higher**, so for any given difficulty submodel, it will look like the 2PL
  - Otherwise known as "cumulative logit model"
  - **Like dividing 4-category item into a series of binary items**
  - $0 \text{ vs. } 1,2,3$        $0,1 \text{ vs. } 2,3$        $0,1,2 \text{ vs. } 3$
  - $b_{1i}$                        $b_{2i}$                        $b_{3i}$
  - Each threshold uses all response data in estimation

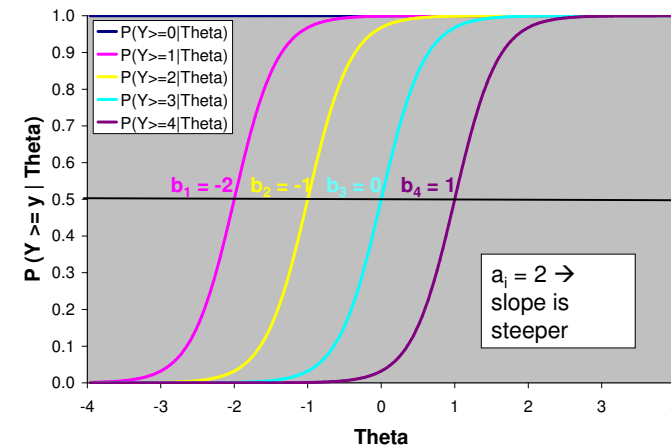
# Example GRM for 4 Options (0-3): 3 Submodels with common a

- Prob of 0 vs 123 →  $P_{i1}(y \geq 1) = \frac{\exp(1.7a_i(\theta_s - b_{i1}))}{1 + \exp(1.7a_i(\theta_s - b_{i1}))}$
  - Prob of 01 vs 23 →  $P_{i2}(y \geq 2) = \frac{\exp(1.7a_i(\theta_s - b_{i2}))}{1 + \exp(1.7a_i(\theta_s - b_{i2}))}$
  - Prob of 012 vs 3 →  $P_{i3}(y \geq 3) = \frac{\exp(1.7a_i(\theta_s - b_{i3}))}{1 + \exp(1.7a_i(\theta_s - b_{i3}))}$
  - Prob of 0 →  $1 - P_{i1}$
  - Prob of 1 →  $P_{i1} - P_{i2}$
  - Prob of 2 →  $P_{i2} - P_{i3}$
  - Prob of 3 →  $P_{i3} - 0$
- Note  $a_i$  is the same across thresholds → only one slope per item
- $b_{ik}$  = trait level needed to have a 50% probability of responding in that category or higher

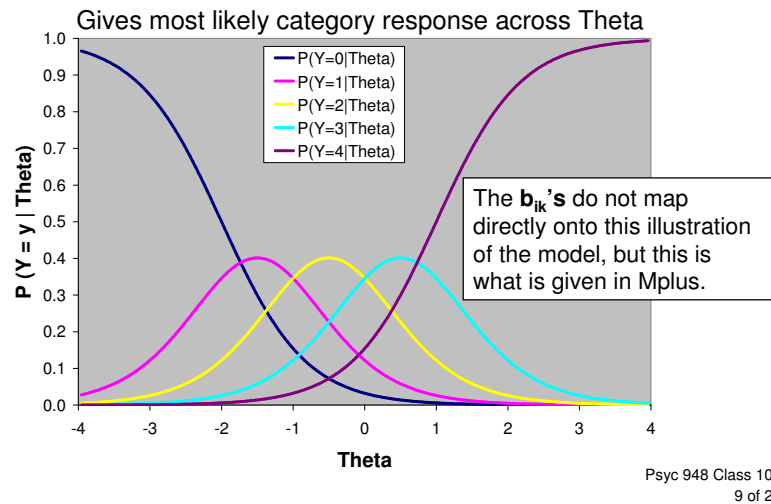
# Cumulative Item Response Curves (GRM for 5-Category Item, $a_i = 1$ )



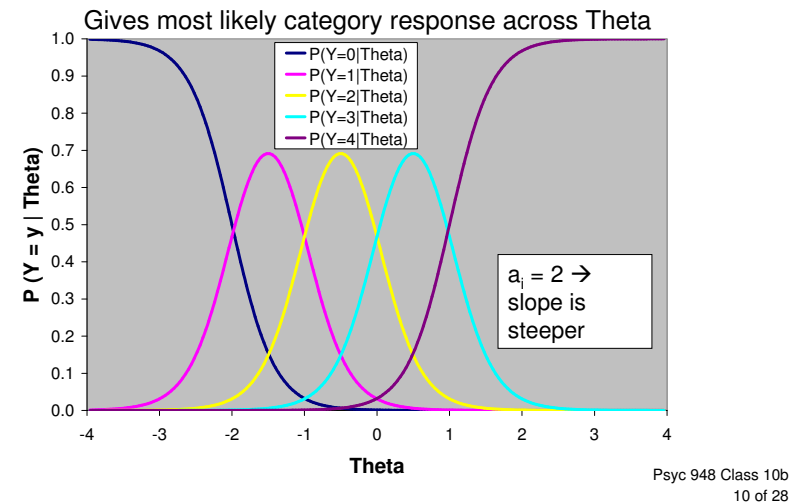
# Cumulative Item Response Curves (GRM for 5-Category Item, $a_i = 2$ )



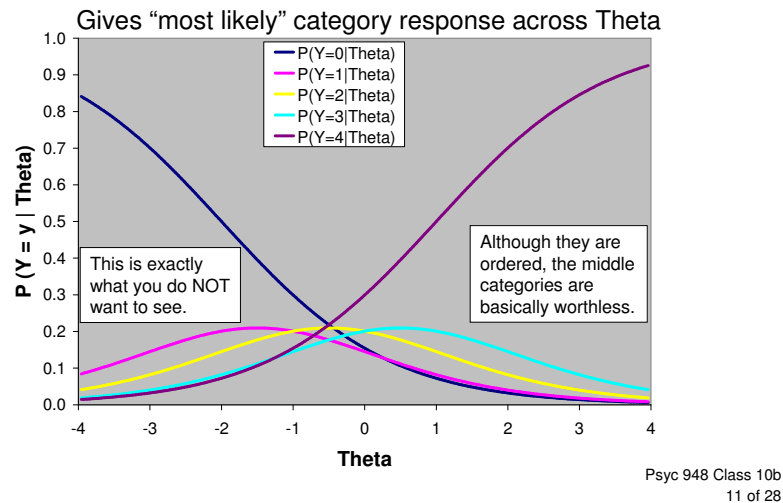
## Category Response Curves (GRM for 5-Category Item, $a_i = 1$ )



## Category Response Curves (GRM for 5-Category Item, $a_i = 2$ )



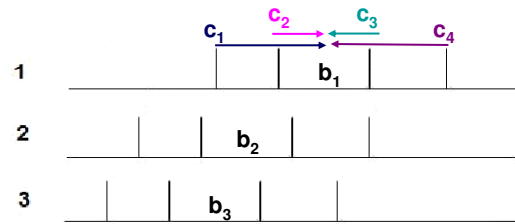
## Category Response Curves (GRM 5-Category Item, $a_i = .5$ )



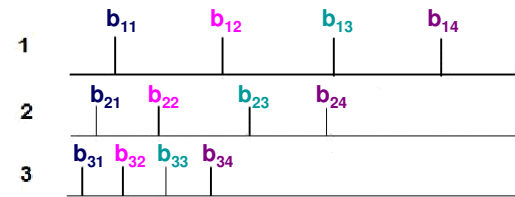
## "Modified" ("Rating Scale") Graded Response Model

- Is more parsimonious version of graded response model
- Designed for items with **same response format**
- In GRM, there are  $(\#options-1) * (\#items)$  thresholds estimated + one slope per item
- In MGRM, each item gets own slope and own 'location' parameter, but the differences between categories around that location are constrained equal across items (get a "c" shift for each threshold)
  - Items differ in overall location, but spread of categories within is equal
  - So, different  $a_i$  and  $b_i$  per item, but same  $c_1$ ,  $c_2$ , and  $c_3$  across items
- Prob of 0 vs 123  $\rightarrow P_{i1}(y \geq 1) = \frac{\exp(1.7a_i(\theta_s - (b_i + c_1)))}{1 + \exp(1.7a_i(\theta_s - (b_i + c_1)))}$   
(and so forth for  $c_2$  and  $c_3$ )
  - Not same 'c' as guessing parameter – sorry, they reuse letters...
  - Not directly available within Mplus, but pry could be using constraints

# Summary of Models for Ordered Categorical Responses



**Modified GRM** →  
**1 Location, k-1 c's**  
 All category distances are same across items



**Original GRM** →  
**k-1 locations**  
 All category distances are allowed to differ across items

Item Difficulty / Latent Ability

Available in Mplus with "CATEGORICAL ARE" option	Difficulty Per Item Only (category distances equal)	Difficulty Per Category Per Item
<b>Equal discrimination across items (1-PLish)?</b>	(possible, but no special name)	(possible, but no special name)
<b>Unequal discriminations (2-PLish)?</b>	"Modified GRM" or "Rating Scale GRM" (same response options)	"Graded Response Model" "Cumulative Logit"

- GRM and Modified GRM are reliable models for ordered categorical data
  - Commonly used in real-world test situations; most stable to use in practice
  - Least amount of data demands because all data get used in estimating each threshold
  - Only *major* deviations from the model will end up causing problems

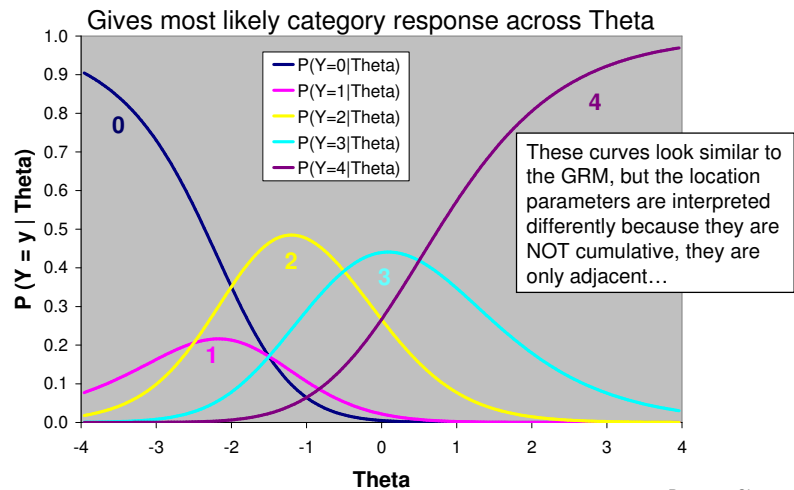
## Partial Credit Model (PCM)

- Ideal for items for which you want to **test** an assumption of an ordered underlying continuum
- # response options doesn't have to be same across items
- Is a "direct, divide-by-total" model (prob of response given directly)
- Estimate k-1 thresholds (so 4 options → 3 thresholds)
- Models the probability of **adjacent response categories**:
  - Otherwise known as "adjacent category logit model"
  - Divide item into a series of binary items, but without order constraints beyond adjacent categories because it only uses those 2 categories:
    - 0 vs. 1      1 vs. 2      2 vs. 3
    - $\delta_{1i}$        $\delta_{2i}$        $\delta_{3i}$
  - No guarantee that any category will be most likely at some point

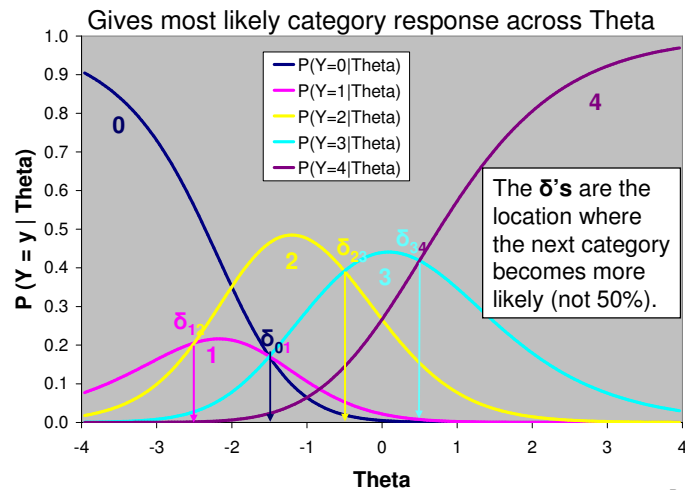
## Partial Credit Model

- With different slopes ( $a_i$ ) per item, then it's "generalized partial credit model"; otherwise 1-PLish version is Partial Credit Model
- Still 3 submodels for 4 options, but set up differently:**
  - Given 0 or 1, prob of 1 →  $P_{11}(\theta_s) = \frac{\exp(1.7a_i(\theta_s - \delta_{1i}))}{1 + \exp(1.7a_i(\theta_s - \delta_{1i}))}$
  - Given 1 or 2, prob of 2 →  $P_{12}(\theta_s) = \frac{\exp(1.7a_i(\theta_s - \delta_{2i}))}{1 + \exp(1.7a_i(\theta_s - \delta_{2i}))}$
  - Given 2 or 3, prob of 3 →  $P_{13}(\theta_s) = \frac{\exp(1.7a_i(\theta_s - \delta_{3i}))}{1 + \exp(1.7a_i(\theta_s - \delta_{3i}))}$
- $\delta$  is the 'step parameter' → latent trait where the next category becomes *more likely* – not necessarily 50%
- Other parameterizations also used – check the program manuals
- Currently not directly available in Mplus

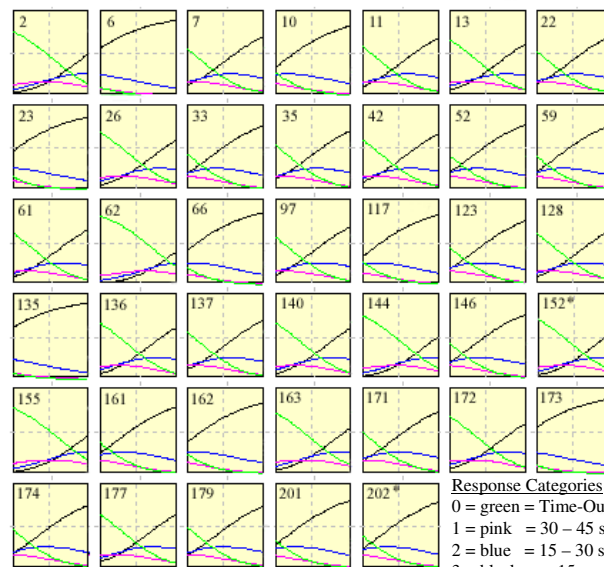
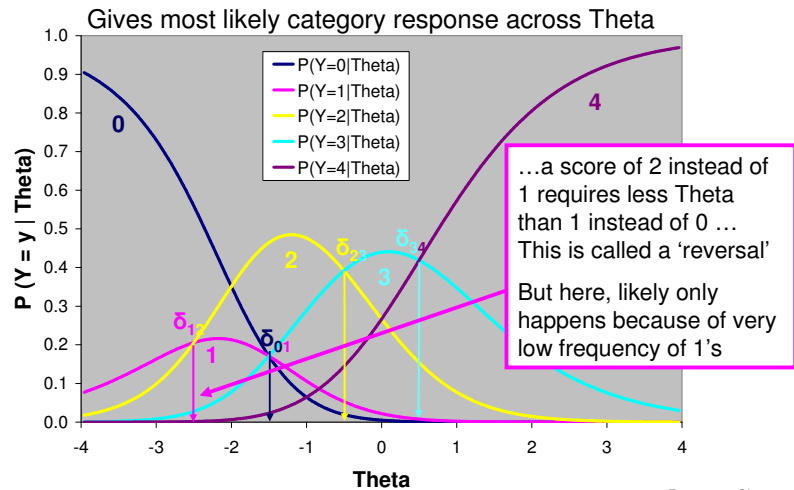
# Category Response Curves (PCM for 5-Category Item, $a_i = 1$ )



# Category Response Curves (PCM for 5-Category Item, $a_i = 1$ )

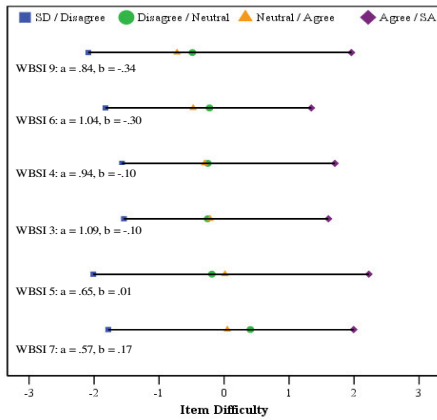


# Category Response Curves (PCM for 5-Category Item, $a_i = 1$ )



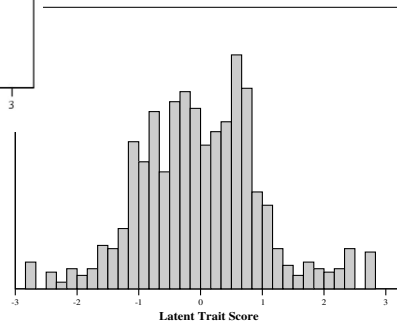
More of what you don't want to see... category response curves from a PCM where reversals are a plenty...  
...and the middle categories are fairly useless.

Response Categories  
0 = green = Time-Out  
1 = pink = 30 - 45 s  
2 = blue = 15 - 30 s  
3 = black = < 15 s  
\*Misfit ( $p < .05$ )

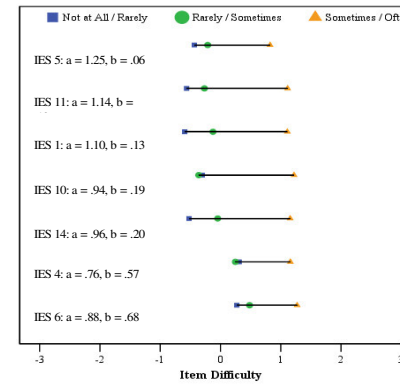


### PCM Example: General Intrusive Thoughts (5 options)

Note that the 4 thresholds **cover a wide range** of the latent trait, and what the distribution of Theta looks like as a result...  
But the middle 3 categories are used infrequently &/or are not differentiable

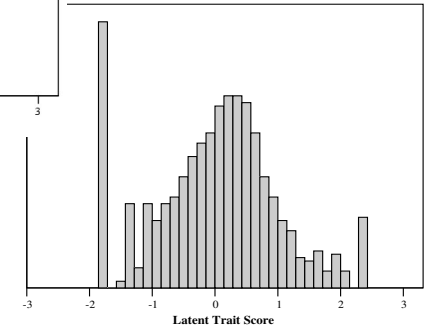


Psyc 948 Class 10b  
21 of 28



### Partial Credit Model Example: Event- Specific Intrusive Thoughts (4 options)

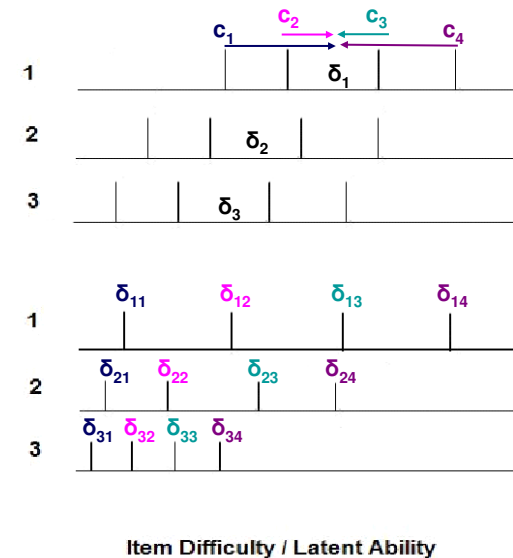
Note that the 3 thresholds **do not cover a wide range** of the latent trait, and what the distribution of theta looks like as a result...



Psyc 948 Class 10b  
22 of 28

## Rating Scale Model

- “Rating Scale” is to PCM what “Modified GRM” is to GRM
- Is more parsimonious version of partial credit model
- Designed for items with **same response format**
- In PCM, there are (#options-1)\*(#items) step parameters estimated (+ one slope per item in generalized PCM version)
- In RSM, each item gets own slope and own ‘location’ parameter, but the differences between categories around that location are constrained equal across items
  - Items differ in overall location, but spread of categories within is equal
  - So, different  $\delta_i$  per item, but same  $c_1, c_2,$  and  $c_3$  across items
- If 0 or 1, prob of 1  $\rightarrow P_{i1}(\theta_s) = \frac{\exp(1.7a_i(\theta_s - (\delta_i + c_1)))}{1 + \exp(1.7a_i(\theta_s - (\delta_i + c_1)))}$  (and so forth for  $\delta_2$  and  $\delta_3$ )
  - $\delta_i$  is a ‘location’ parameter, and  $c$  is the step parameter as before
  - Constrains curves to look same across items, just shifted by  $\delta_i$



**Rating Scale**  $\rightarrow$   
**1 Location, k-1 c’s**  
All category distances are same across items

**Original PCM**  $\rightarrow$   
**k-1 locations**  
All category distances are allowed to differ across items

Psyc 948 Class 10b  
23 of 28

Psyc 948 Class 10b  
24 of 28

# Summary of Models for Maybe-Ordered Categorical Responses

Not directly available in Mplus	Difficulty Per Item Only (category distances equal)	Difficulty Per Category Per Item
<b>Equal discrimination across items (1-PLish)?</b>	“Rating Scale PCM”	“Partial Credit Model”
<b>Unequal discriminations (2-PLish)?</b>	“Generalized Rating Scale PCM”?? (same response options)	“Generalized PCM” “Adjacent Category Logit”

- Partial Credit Models test the assumption of ordered categories
  - This can be useful for item screening, but perhaps not for actual analysis
- These models have additional data demands relative to GRM
  - Only data from that threshold get used (i.e., for 1 vs. 2, 0 and 3 don't contribute)
  - So larger sample sizes are needed to identify all model parameters
  - Sometimes categories have to be consolidated to get the model to not blow up

# Nominal Response Model

- Ideal for items with no ordering of any kind (e.g., *dog, cat, bird*)
  - # response options don't have to be same across items
  - Is a “direct” model (probability of response given directly)
  - Models the probability of one response category against all others
    - Still like dividing item into a series of binary items, but now each option is really considered as a separate item (“Baseline category logit”)
    - $0 \text{ vs. } 1,2,3$   $1 \text{ vs. } 0,2,3$   $2 \text{ vs. } 0,1,3$ 

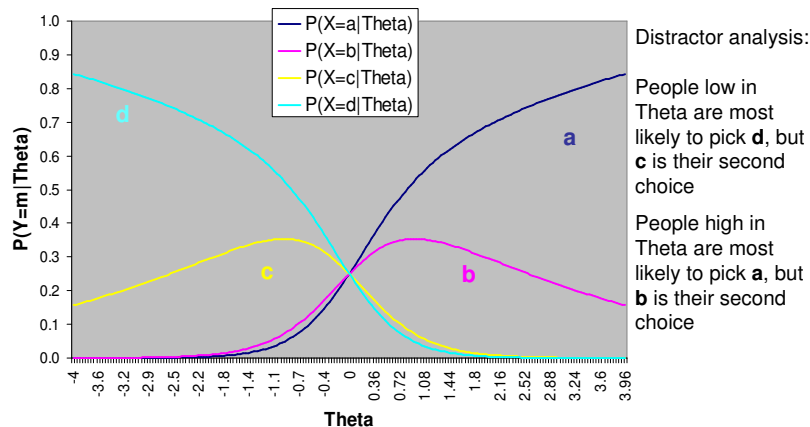
$c_{1i}$

$c_{2i}$

$c_{3i}$
- Estimate one slope ( $a_i$ ) and one “intercept” ( $c_i$ ) parameter *per item, per threshold*, such that  $\sum(a's)=0$ ,  $\sum(c's)=0$  (so  $a$  and  $c$  are only relatively meaningful within a single item)
- $$P(y=1) = \frac{\exp(1.7a_{1i}(\theta_s + c_{1i}))}{\sum_{y=0}^3 \exp(1.7a_{iy}(\theta_s + c_{iy}))}$$
- Available in Mplus with NOMINAL ARE option
  - Can be useful to examine ‘distractors’ in multiple choice tests

# Category Response Curves (NRM for 5-Category Item)

Nominal Response Item Response Function



# Summary: Polytomous Models

Many kinds of polytomous IRT models...

- Some **assume order** of response options... (done in Mplus)
  - **Graded Response Model Family** → “cumulative logit model”
    - Model cumulative change in categories using all data for each
- Some allow you to **test order** of response options... (no Mplus)
  - **Partial Credit Model Family** → “adjacent category logit model”
    - Model adjacent category thresholds only, so they allow you to see reversals (empirical mis-ordering of your response options with respect to Theta)
    - PCM useful for identifying separability and adequacy of categories
- Some assume **no order** of response options... (done in Mplus)
  - **Nominal Model** → “baseline category logit model”
    - Useful to examine probability of each response option
    - Is very unparsimonious and thus can be hard to estimate