Overview of Item Response Theory

Ron D. Hays November 14, 2012 (8:10-8:30am) Geriatrics Society of America (GSA) Pre-Conference Workshop on Patient-Reported Outcome Item Banks San Diego Convention Center (Room 14-A)

Latent Trait and Item Responses



Item Responses and Trait Levels



Item Response Theory (IRT)

IRT models the relationship between a person's response Y_i to the question (i) and his or her level of the latent construct θ being measured by positing

$$\Pr(Y_i \ge k) = \frac{1}{1 + \exp(-a_i\theta + b_{ik})}$$

 b_{ik} estimates how difficult it is for the item (i) to have a score of k or more and the discrimination parameter a_i estimates the discriminatory power of the item.

If for one group versus another at the same level θ we observe systematically different probabilities of scoring k or above then we will say that item i displays DIF

Some Nice IRT Features

- Category response curves (CRCs)
- Computer-adaptive testing (CAT)
- Assessing differential item functioning

Posttraumatic Growth Inventory

Indicate for each of the statements below the degree to which this change occurred in your life as a result of your crisis. (*Appreciating each day*)

(0) I did not experience this change as result of my crisis

- (1) I experienced this change to a <u>very small degree</u> as a result of my crisis
- (2) I experienced this change to a <u>small degree</u> as a result of my crisis
- (3) I experienced this change to a moderate degree as a result of my crisis
- (4) I experienced this change to a great degree as a result of my crisis
- (5) I experienced this change to a <u>very great degree</u> as a result of my crisis

Category Response Curves



Drop Response Options?

Indicate for each of the statements below the degree to which this change occurred in your life as a result of your crisis. (*Appreciating each day*)

(0) I <u>did not</u> experience this change as result of my crisis

- (1) I experienced this change to a moderate degree as a result of my crisis
- (2) I experienced this change to a great degree as a result of my crisis
- (3) I experienced this change to a <u>very great degree</u> as a result of my crisis

Reword?

 Might be challenging to determine what alternative wording to use so that the replacements are more likely to be endorsed.

Keep as is?

- CAHPS global rating items
 - 0 = worst possible
 - 10 = best possible
- 11 response categories capture about 3 levels of information.

-10/9/8-0 or 10-9/8/7-0

 Scale is administered as is and then collapsed in analysis

Response Burden vs. Standard Error (SE)

- 3-5 items per minute rule of thumb for paper survey
 - -8 items per minute for dichotomous items
- Lowering SE means adding or replacing existing items with more informative ones at the target range of the continuum.

Computer Administration

- Polimetrix panel sample
 - –12-13 items per minute (automatic advance)
 - -8-9 items per minute (next button)
- Scleroderma patients at UCLA
 - -6 items per minute

CAT

- Only as much response burden as needed for target level of reliability
- For z-scores (mean = 0 and SD = 1):
- Reliability = $1 SE^2 = 0.90$ (when SE = 0.32)
- Information = $1/SE^2 = 10$ (when SE = 0.32)
- Reliability = 1 1/information
- CATs for patient-reported outcomes yield 0.90 reliability with about 5 items

Differential Item Functioning (DIF)

- Probability of choosing each response category should be the same for those who have the same estimated scale score, regardless of other characteristics
- Evaluation of DIF

 Different subgroups
 Mode differences

DIF (2-parameter model)



