IRT workshop

Christopher David Desjardins

19 March 2014

Overview

Review

- Last time: Polytomous IRT models
  - Partial credit (and Generalized)
  - Graded response
  - Nominal response

Proposed topics

Differential Item Functioning
Multidimensional models (2/4)

Linking/Equating & Presentations (16/4)

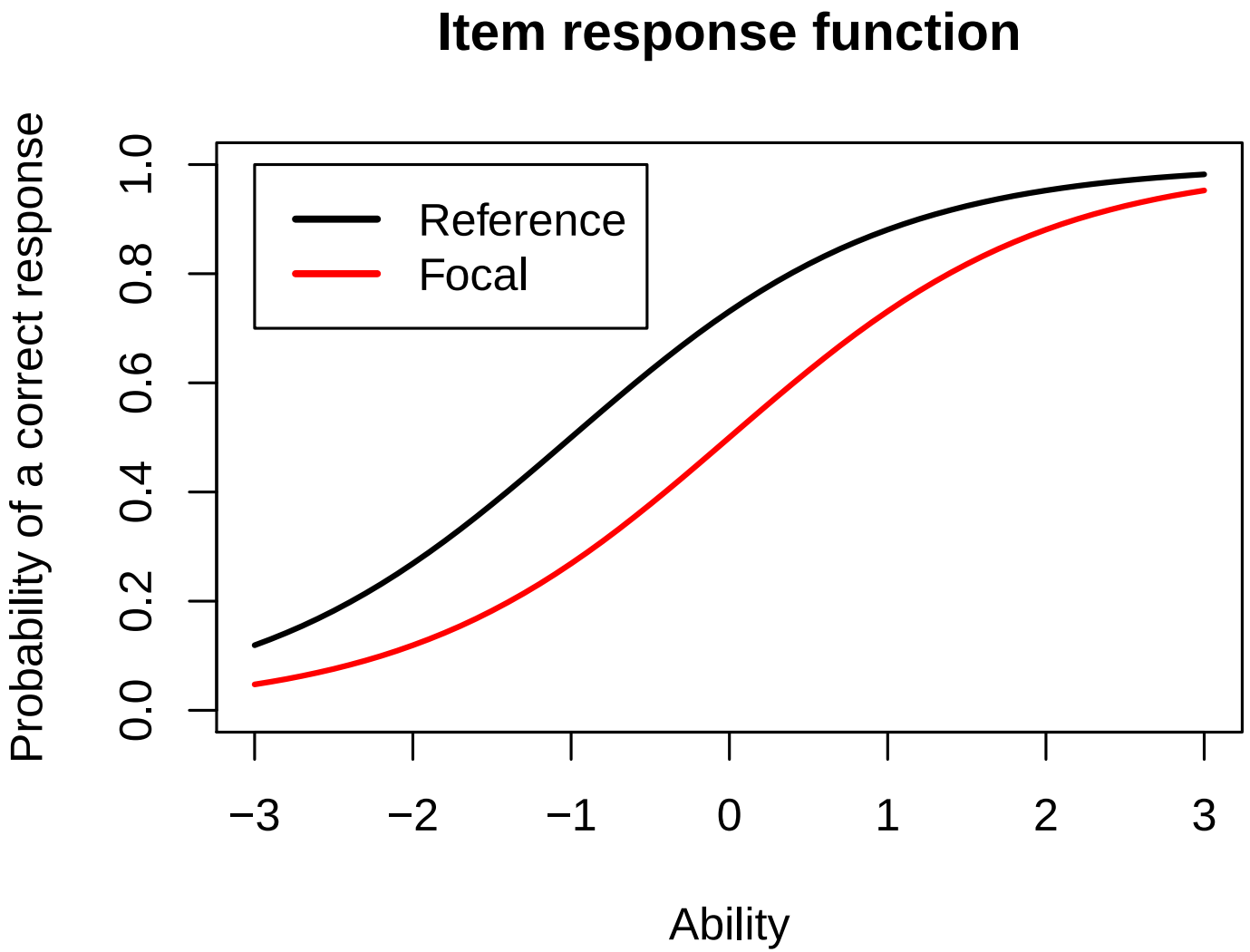
Differential item functioning

- Differential item functioning (DIF) occurs when an item is functioning differently across manifest groups of individuals.
- DIF is related to the notion of bias, unfairness.
- An item that is biased has an adverse effect on a group.
- If an item has DIF, it needs to be evaluated for rewording or removal.
- Items with DIF are typically reviewed by experts to determine if the differential performance is relevant or irrelevant to the measured construct.

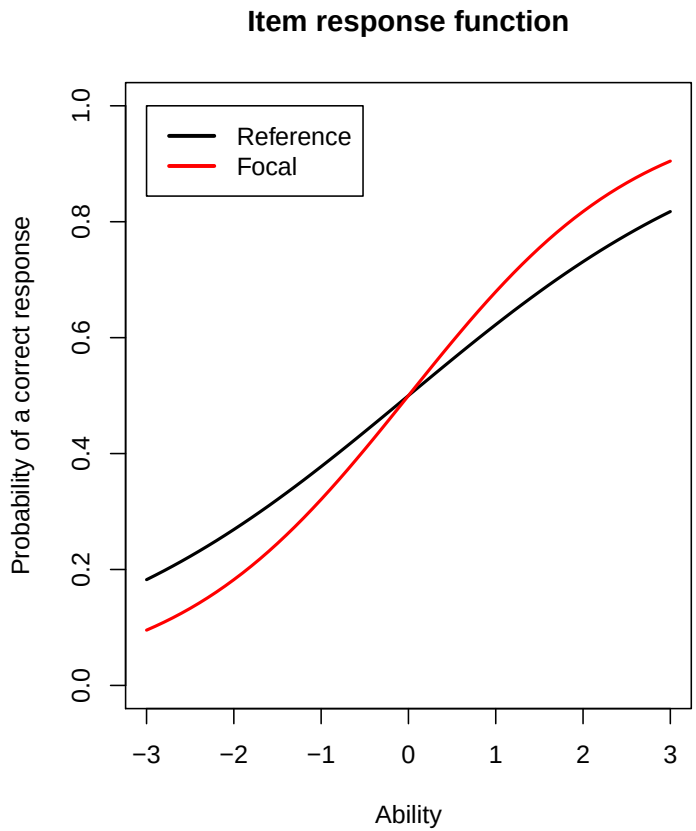
DIF

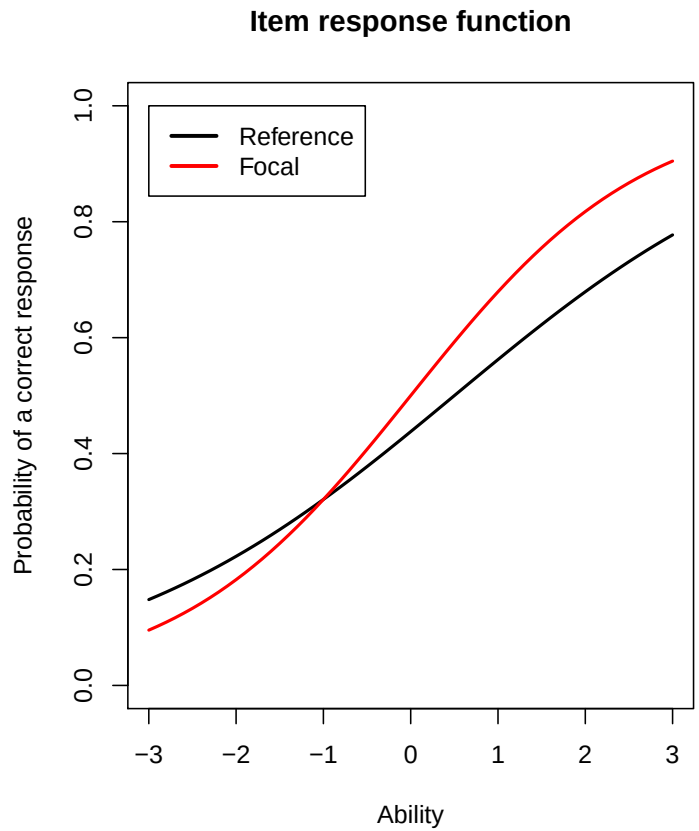
- DIF is an item that displays different statistical properties for different manifest groups after the groups have been matched on a proficiency measure (i.e. after conditioning on the construct of interest).
- In DIF, typically talk about the reference (majority) group and the focal (minority) group.
- The focal group is the group we are interested in (i.e. are they adversely affected by the item).
- There are two types of DIF: Uniform and non-uniform.

Uniform DIF


Item response function
Probability of a correct response
0.0
0.2
0.4
0.6
0.8
1.0
Reference
Focal
−3
−2
−1
0
1
2
3
Ability

Non-uniform DIF





Uniform and non-uniform

- Uniform - item difficulties (locations) differ by group.
- Non-uniform - item discriminations (slopes) differ by group (locations may differ too).
- DIF is plausibly created by multidimensionality of the instrument (need to investigate with multidimensional IRT).
- Three issues with DIF
  - Need to investigate wording of item and plausible cause and consider removing item.
  - How much DIF is there?
  - How much does the item with DIF affect the DIF analysis?
- DIF in IRT means that the items parameter estimates are NOT invariant across the groups.

Detecting DIF - Mantel-Haenszel Chi-Square

- Note: Before DIF can be performed the groups metrics must be linked (i.e. put on a common scale)!
- The Mantel-Haenszel (MH) determines whether two variables are independent after conditioning on a third variable (typically the observed response).
- For DIF, the MH consists of a series of 2 x 2 contingency tables based on persons with the same observed test score.
- For a single dichotomous test,

	Item Response		
	0	1	Total
Reference	B_t	A_t	n_{Rt}
Focal	D_t	C_t	n_{Ft}
Total	m_{0t}	m_{1t}	T_t

MH χ^2 Statistic

For a dichotomous item,

$$MH_{\chi^2} = \frac{\{\sum_{t=1}^{L-1} |A_t - \frac{n_{Rt} m_{1t}}{T_t}| - 0.5\}^2}{\sum_{t=1}^{L-1} \frac{n_{Rt} n_{Ft} m_{1t} m_{0t}}{T_t^2 (T_t - 1)}}$$

- The 0.5 is Yate's correction for continuity and df = 1.
- $\sum_{t=1}^{L-1}$ sum over the length of the test (L) - 1. Will have L - 1 tables.
- H_0: Odds_{ref} with a 1 = Odds_{foc} with a 1
- Reject the null when frequency of the reference group receiving a 1 is greater (i.e. uniform DIF favors ref) or less (i.e. uniform DIF favors foc) than would be expected.
- Generally, useful just for uniform DIF but could detect non-uniform DIF if A_t differs by table.

MH effect size estimate

To calculate the effect size, just calculate the odds ratio.

$$\hat{\alpha}_{MH} = \frac{\sum_{t=1}^{L-1} A_t D_t}{\sum_{t=1}^{L-1} B_t C_t}$$

- Greater than 1, favors the reference group
- Can also calculate the logit (by taking log) and getting a confidecne interval for this estimate.
- ETS, $deltaMH = 2.35 * log(Odds)$

Detecting DIF - TSW LRT

- The Thissen, Steinberg, and Wainer DIF is based on comparing two IRT models using a likelihood ratio test.
- If there is no DIF, then item locations and discrimination should be the same across the two IRT models.
- This is a 3 step procedure.
  1. Fit IRT with all item parameter estimates constrained, expect the item you are interested in investigating. Let the item location (1-PL) and/or slope (2-PL) differ by group for just this one item.
  2. Fit a IRT with all item parameter estimates constrained.
  3. Calculate the difference in log-likelihood between these two models. This will be distributed, typically, with 1 df (if only item location varies) or 2 df (if both location and slope vary).
- Significant, LRT signifies DIF!

Detecting DIF - Logistic Regression

- Another approach to detecting DIF involves using logistic regression (LR).
- In LR, we regress the response to an item on predictor(s) of interest.
- $\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_i + \beta_2 \text{group}_i + \beta_3 X_i * \text{group}_i$ (M1)
  - Compare M1 to $\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_i + \beta_2 \text{group}_i$ (M2) via a LRT.
    - Significant LRT, i.e. $\beta_3 \neq 0$, indicates non-uniform DIF
  - If not significant, compare M2 to $\log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 X_i$
    - Significant LRT, i.e. $\beta_2 \neq 0$, indicates uniform DIF
  - Can look at odds (exponentiate the estimated parameters) to express magnitude or look at change in R^2

MH in R

```
> library(difR)
> data(verbal)
> head(verbal)
> # Excluding the "Anger" variable
> verbal <- verbal[colnames(verbal)!="Anger"]
> r <- difMH(verbal, group=25, focal.name=1)
> r
```

Logistic Regression in R

```
> data(verbal)
> # Look just at DoScold
> m1 <- glm(S3DoScold ~ Anger + Gender + Anger:Gender,
+           data = verbal, family = "binomial")
> m2 <- update(m1,.~. -Anger:Gender)
> # No evidence of non-uniform DIF
> 1-pchisq(anova(m2,m1)$Dev[2],df=1)
> m3 <- update(m2,.~. -Gender)
> # Evidence of uniform DIF
> 1-pchisq(anova(m3,m2)$Dev[2],df=1)
> # Report the change in R Squared
> m2.r2 <- cor(fitted(m2),verbal$S3DoScold)^2
> m3.r2 <- cor(fitted(m3),verbal$S3DoScold)^2
> m2.r2 - m3.r2
```

DIF without binary data

- For ordered data, use ordinal regression model.
- For nominal data, use multinomial logistic regression model.
- Fit both of these models in a similar manner to how the logistic regression model is fit!