Item Parameter Interpretation and Model Fit

Jonathan Templin
University of Georgia

National Council on Measurement in Education
Diagnostic Modeling
Training Session
April 13, 2009

Talk Overview

- Here we evaluate the fit of a model
- Item level
  - Item parameter interpretation
- Test level
  - Goodness of fit of a DCM to a test

ITEM PARAMETER INTERPRETATION

Log-linear Cognitive Diagnosis Model

- The LCDM specifies the probability of a correct response as a function of a set of attributes and a Q-matrix:

$$P(X_{ij}=1|\boldsymbol{\alpha}_i)=\frac{e^{\boldsymbol{\lambda}_j^T\mathbf{h}(\mathbf{q}_j,\boldsymbol{\alpha}_i)}}{1+e^{\boldsymbol{\lambda}_j^T\mathbf{h}(\mathbf{q}_j,\boldsymbol{\alpha}_i)}}$$

- For an item, the LCDM has ANOVA-like parameters:

$$\boldsymbol{\lambda}_j^T\mathbf{h}(\mathbf{q}_j,\boldsymbol{\alpha}_i)=\lambda_{j,0}+\sum_{u=1}^{K}\lambda_{j,1,(u)}\left(\alpha_u q_{ju}\right)+\sum_{u=1}^{K}\sum_{v>u}\lambda_{j,2,(u,v)}\left(\alpha_{iu}\alpha_{iv}q_{ju}q_{jv}\right)+\ldots$$

Intercepts

Main Effects

Two-Way Interactions

Higher Interactions

Interpreting LCDM Item Parameters

- Because of the multiple types of item parameters in the LCDM, interpretation of each varies
  - Follows a ANOVA/linear modeling approach
  - Intercepts
  - Main effects (when interactions aren't present)
  - Interactions (when present)
    - Then main effects

Example Item

- To demonstrate parameter interpretation, we include Mplus output from an estimated example item
  - Measured two attributes (of three in Q-matrix)

New/Additional Parameters

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
L6_0	-2.537	0.176	-14.385	0.000
L6_12	2.102	0.226	9.310	0.000
L6_13	2.151	0.218	9.851	0.000
L6_223	1.110	0.376	2.947	0.003

Reading Mplus Output

New/Additional Parameters

	Estimate	S.E.	Est./S.E.	Two-Tailed P-Value
L6_0	-2.537	0.176	-14.385	0.000
L6_12	2.102	0.226	9.310	0.000
L6_13	2.151	0.218	9.851	0.000
L6_223	1.110	0.376	2.947	0.003

- L[i]_[e][a1,…]
  - LCDM parameter
    - i – item number
    - e – type of effect (intercept=0, main effect = 1, two-way interaction = 2, …)
    - a1,… - list of attributes to which effect applies

Reading Mplus Output

```
New/Additional Parameters
                                                           Two-Tailed
                    Estimate       S.E.  Est./S.E.    P-Value
    L6_0              -2.537      0.176    -14.385      0.000
    L6_12              2.102      0.226      9.310      0.000
    L6_13              2.151      0.218      9.851      0.000
    L6_223             1.110      0.376      2.947      0.003
```

- Estimate – LCDM parameter estimate
- S.E. – (Asymptotic) standard error of the estimate
- Est./S.E. – Test statistic for testing null hypothesis parameter is zero
- P-value – Approximate p-value for interaction parameters
  - NOT main effects

LCDM Intercepts

- Estimated Intercept: -2.537 (0.176)
- Indicates the log-odds of a correct response for a non-master of all attributes
  - Here, non-masters have a very low probability of a correct response: exp(-2.537)/1+exp(-2.537) = 0.07
- Hypothesis test is not important
  - Tests whether non-masters have a probability of a correct response of 0.5
- Problematic when very high
  - Difficult to identify other parameters

Higher Order Model Parameters

- Interpretation of main effects and interactions proceeds sequentially:

- If interactions are present:
  - Examine highest level of interaction
    - If significantly different from zero, leave in model
    - If not, term can be omitted

- If interactions are not present:
  - Examine how far main effect is from zero

Examining Interaction Parameters

- 2-way interaction parameter: 1.110 (0.376)

- P-value for parameter was 0.002
  - Indicates parameter is significantly different from zero
  - Candidate to leave in model

- Value indicates that there is an over-additive effect of mastering both attributes
  - Like a bonus for mastery of both attributes

Interpreting Main Effects

- When significant interactions are present, main effects cannot be easily interpreted
  - Need to know *combination* of attributes mastered to fully describe item response function

- Main effects are still increase in log-odds of a correct response for mastery of an attribute
  - Just difficult to parse effect out of context

Other Main Effect Concerns

- Because of lower bound, main effect hypothesis tests from Mplus are invalid near zero
  - So are model standard errors
- For items with one attribute:
  - Must rely on goodness of fit to determine if attribute is influential or sufficient for item fit

General Modeling Tips

- High-level interactions are difficult to estimate in most samples
  - More than 2-way interactions may not be possible

- Modeling strategy:
  - Try all interactions
    - If model does not converge, limit to only 2-way interactions
  - Remove non-significant interactions from model
  - If all interactions and main effects for an attribute are close to zero:
    - Entry for attribute in Q-matrix can be removed
  - Double check with AIC/BIC as hypothesis test is approximate

ASSESSMENT OF MODEL FIT

Assessing Model Fit

- There is no one best way to assess fit in DCMs
- Techniques typically used can put into several general categories:
  - Absolute fit
    - Model based hypothesis tests
    - Entropy
  - Relative fit
    - Information criteria
  - Item fit

Model Chi-Squared test

- For small numbers of items (10-15), the traditional Chi-Squared test of model fit can be used.
  - Test is invalid for too many items – sparse data
- Mplus gives this automatically
  - Omits when data are sparse

```
Chi-Square Test of Model Fit for the Binary
and Ordered Categorical (Ordinal) Outcomes

          Pearson Chi-Square

          Value                                  9.459
          Degrees of Freedom                         6
          P-Value                               0.1494

          Likelihood Ratio Chi-Square

          Value                                  8.966
          Degrees of Freedom                         6
          P-Value                               0.1755
```

(Relative) Entropy

- The entropy of a model is a measure of classification uncertainty.
  - It is an absolute fit statistic

- Mplus reports relative entropy
  - Value of 1.00 means all examinees classified with complete certainty (good fit)
  - Value of 0.00 means all examinees classified with equal probabilities for all classes (poor fit)

Model Comparison: Information Criteria

- Used when comparing between two models, i.e.:
  - Two DCMs (LCDM v. DINA)
  - Two Q-matrices (4 attribute v. 5 attribute)
  - Two different models (IRT v. DCM)
- Mplus reports:
  - AIC
  - BIC
  - Sample size adjusted BIC
- All can be used
  - Smallest value is best

Mplus Model Fit Output

```
TESTS OF MODEL FIT

Loglikelihood

          H0 Value                                  -331.764

Information Criteria

          Number of Free Parameters                        9
          Akaike (AIC)                               681.527
          Bayesian (BIC)                             708.130
          Sample-Size Adjusted BIC                   679.653
            (n* = (n + 2) / 24)
          Entropy                                      0.754
```

Item Fit Statistics

- The TECH10 option reports a degree of misfit for each
  - Item individually (Univariate)
  - Pair of two items (Bivariate)
- Uses Chi-squared test for misfit
  - Values for each item are distributed as Chi-square with 1 df (for binary items)
- Misfitting items can be investigated
  - Q-matrix can be changed
  - Items can be removed

Univariate Fit

UNIVARIATE MODEL FIT INFORMATION

Variable	Estimated Probabilities H1	H0	Standard Residua
U1			
Category 1	0.472	0.472	0.000
Category 2	0.528	0.528	0.000
U2			
Category 1	0.514	0.514	0.000
Category 2	0.486	0.486	0.000
U3			
Category 1	0.739	0.739	0.000
Category 2	0.261	0.261	0.000
U4			
Category 1	0.563	0.563	0.000
Category 2	0.437	0.437	0.000

Bivariate Fit

```
BIVARIATE MODEL FIT INFORMATION

                                          Estimated Probabilities
                                                             Standardized
Variable        Variable                 H1           H0       Residual
                                                               (z-score)
X9              X13
  Category 1      Category 1           0.048        0.038         2.722
  Category 1      Category 2           0.107        0.116        -1.625
  Category 2      Category 1           0.141        0.150        -1.458
  Category 2      Category 2           0.705        0.695         1.134
  Bivariate Pearson Chi-Square                                   11.659
  Bivariate Log-Likelihood Chi-Square                            11.228
```

CONCLUDING REMARKS

Concluding Remarks

- In this section, we discussed
  - Parameter interpretation
  - Modeling strategy for LCDM estimation
  - Model fit and comparison assessment

- More details contained in forthcoming book:
  - Diagnostic Measurement: Theory, Methods, and Applications (Rupp, Templin, & Henson, 2010)