Assessment of Model Fit

Session 6
Session Overview

• Session 6 will provide an overview of model fit for DCMS

• Model fit is used to help:
  • Determine if a model fits the data well enough in an absolute sense to use the examinee estimates
  • Select best model among competing models

• ECPE data will be used to illustrate model fit in practice
Session 6: Assessment of Model Fit

ASSESSMENT OF MODEL FIT
Assessing Model Fit

• There is no one best way to assess fit in DCMs

• Techniques typically used can be put into several general categories:
  ➢ Absolute fit
    • Model based hypothesis tests (if available)
    • Entropy
  ➢ Relative fit
    • Information criteria
  ➢ Item fit

• Topics discussed here will mainly focus on fit statistics available in Mplus (also discussed in Session 5)
Overall Model Fit: 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
  - Shown for 28 item ECPE

- Mplus gives this automatically
  - Omits when data are sparse
  - Can omit extreme cells from an analysis
    - Misleading

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

<table>
<thead>
<tr>
<th></th>
<th>Value</th>
<th>Degrees of Freedom</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson Chi-Square</td>
<td>18967.274</td>
<td>268433324</td>
<td>1.0000</td>
</tr>
<tr>
<td>Likelihood Ratio Chi-Square</td>
<td>3765.577</td>
<td>268433324</td>
<td>1.0000</td>
</tr>
</tbody>
</table>

** Of the 268435456 cells in the latent class indicator table, 2050 were deleted in the calculation of chi-square due to extreme values.
Overall Model Fit: (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 respondents classified with complete certainty (good fit)
  ➢ Value of 0.00 means all respondents classified with equal probabilities for all classes (poor fit)

• ECPE (relative) entropy: 0.672
  ➢ Hard to interpret by itself
Relative Model Fit: Information Criteria

- Used when comparing between two models:
  - Two DCMs (LCDM v. DINA)
  - Two Q-matrices (4 v. 5 attributes)
  - Two different models (IRT v. DCM)

- Mplus reports:
  - AIC and BIC
  - Sample size adjusted BIC

- All can be used
  - Smallest value is best

Information Criteria

<table>
<thead>
<tr>
<th>Number of Free Parameters</th>
<th>81</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akaike (AIC)</td>
<td>85641.425</td>
</tr>
<tr>
<td>Bayesian (BIC)</td>
<td>86125.807</td>
</tr>
<tr>
<td>Sample-Size Adjusted BIC</td>
<td>85868.440</td>
</tr>
<tr>
<td>(n* = (n + 2) / 24)</td>
<td></td>
</tr>
</tbody>
</table>
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
Item Fit Statistics: Univariate Fit

- Univariate fit attempts to determine if the model fits each item marginally
  - A limited information statistic

- Not useful in DCMs
  - Model is for probability
  - Will always fit perfectly
Bivariate fit is an index of fit for a pair of items.

Compared observed data with frequency expected under DCM.

- Produces a 1-df Chi-Squared test.

Can help identify items that do not fit model.

Rough approximation.
ECPE Results: Model Fit

• Overall model fit
  - Chi-square not computable
  - AIC: 85641.425; BIC: 86125.807
    - Used to reduce model (e.g., take out higher order interactions and see if AIC/BIC is reduced)

• Bivariate model fit (Session 4)
  - Compares model predicted and observed frequencies of responses for all pairs of items
  - Of 378 item pairs 45 had p-values less than 0.01
  - Items most indicated
    - Item 13 (9 of 45 pairs)
    - Item 4 (6 of 45 pairs)
    - Item 5 (6 of 45 pairs)

• Indicates some items are not fit well by model
  - We will ignore this and continue with analysis as example
• We can evaluate the full LCDM vs the HDCM shown in the previous session using:
  - Overall Fit using information indices
    • Can be used for non-nested models
  - Deviance or Likelihood Ratio Test
    • Used with nested models only
Overall Fit using information indices

- Lower values indicate better fit
- Found to select correct models for attribute hierarchies

Results: HDCM fits better

Conclusion: Attribute hierarchy is present...but let’s see if the deviance test agrees
Deviance Test

• The deviance test can be applied for comparing fit for nested models
  - $H_0$: The nested model fits better than the non-nested model

• Deviance = -2 (Log-likelihood)
  - Is distributed Chi-Square with degrees of freedom equal to the difference in the number of parameters in the two models
  - Is not distributed Chi-Square if the parameter set to zero in the nested model is against its boundary
    - p-value will be incorrect, yielding conservative hypothesis test
    - True p-value can be determined through simulation
Deviance Test: HDCM vs LCDM in ECPE Data

• For the ECPE data, the HDCM with 4 classes is nested within the full LCDM with 8 classes
  ➢ $H_0$: The attribute hierarchy exists (HDCM fits better)

• Deviance = 23.06; df = 13; $\chi^2_{13} (23.06) = .039$
  ➢ Main effects in HDCM constrained to equal boundary of 0
  ➢ Results: Conservative test rejects at .05 and not at .01
  ➢ Conclusion: Depends on researcher’s criterion

• p-value determined through simulation
  ➢ Simulated data with attribute hierarchy present
  ➢ 1 of 76 replications rejects null: p-value = 1/76 = 0.013
  ➢ Results: Correct test fails to reject at .01 or .05
  ➢ Conclusion: Attribute hierarchy exists
CONCLUDING REMARKS

Session 6: Assessment of Model Fit
Concluding Remarks: Model Fit

• Assessment of model fit in DCMs is currently a difficult task
  ➢ Easily accessible options are limited
  ➢ Can quickly find options that take longer to assess fit than to estimate model
  ➢ Mplus options are adequate for initial screening

• DCMs share this problem with IRT models
  ➢ General categorical data analyses

• Other model fit options are available and forthcoming
  ➢ Based on limited information (i.e., Templin, 2007)
  ➢ Need further testing