Survey Sampling Weights and Item Response Parameter Estimation

Spring 2014 Survey Methodology Colloquium Series

Simmons School of Education and Human Development
Center on Research & Evaluation

Paul Yovanoff, Ph.D.
Department of Teaching & Learning
pyovanoff@smu.edu
(541) 346-1495
Topics

• Measurement modeling of survey questionnaire responses is completed for various purposes, e.g., estimate respondent latent traits, estimate item parameters, or to improve survey instrumentation.

• Survey respondents are often sampled such that sampling weights are necessary analytically for statistical representation of sampled subpopulations.

• Are sampling weights important when modeling survey question item parameters?
Survey Item Response Modeling (briefly)

Why should we model survey item responses?

• By item design, an item response is generated by an underlying hypothetical process.

• Responses to items are used to estimate respondent characteristics.

• Therefore, item psychometrics should be known.
Survey Item Response Modeling (briefly)

The Dichotomous Response Model

\[
P(X_{ij} | \theta_j, \beta_i, \alpha_i, \gamma_i) = \gamma_i + (1 - \gamma_i) \frac{\exp[\alpha_i(\theta_j - \beta_i)]}{1 + \exp[\alpha_i(\theta_j - \beta_i)]}
\]

where,
- \(X\) is the response to item \(i\) by respondent \(j\)
- \(\theta\) is the respondent \(j\) latent trait level
- \(\beta\) is the item \(i\) difficulty
- \(\alpha\) is the item \(i\) discrimination (constrained to 1 for these items)
- \(\gamma\) is the item \(i\) guessing (constrained to 0)
Survey Item Response Modeling (briefly)

World Values Survey (dichotomous response)

People use different sources to learn what is going on in their country and the world. For each of the following sources, please indicate whether you used it last week or did not use it last week to obtain information.

<table>
<thead>
<tr>
<th>Source</th>
<th>No</th>
<th>Yes</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. Daily newspaper</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>b. News broadcasts on radio or TV</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>c. Printed magazines</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>d. In depth reports on radio or TV</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>e. Books</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>f. Internet, Email</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>g. Talk with friends or colleagues</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
Survey Item Response Modeling (briefly)
Survey Item Response Modeling (briefly)

Polytomous Response Model

Graded Response Model (Samejima, 1997)

\[ P_{x_{ik}}^*(\theta_j) = P \left( X_{ik} \geq x_{ik} \mid \theta_j \right) = \frac{e^{Da_i(0_j-b_{xik})}}{1 + e^{Da_i(0_j-b_{xik})}} \]

where,

\[ P \left( X_{ik} = x_{ik} \mid \theta_j \right) = P_{x_{ij}}^*(\theta_i) - P_{x_{ij+1}}^*(\theta_i) \]

for person \( j \)

responding in category \( k \) of item \( i \)
Survey Item Response Modeling (briefly)

World Values Survey (polytomous response)

In your opinion, how important should the following be as requirements for somebody seeking citizenship of your country? Specify for each requirement if you consider it as very important, rather important or not important

<table>
<thead>
<tr>
<th>Requirement</th>
<th>Very Important</th>
<th>Rather Important</th>
<th>Not Important</th>
</tr>
</thead>
<tbody>
<tr>
<td>Having ancestors from my country</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Being born on my country’s soil</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Adopting the customs of my country</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Abiding by my country’s laws</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
Survey Item Response Modeling (briefly)

Having ancestors from my country

- Very Important
- Rather Important
- Not Important

leniency

probability
Survey Item Response Modeling (briefly)

Being born on my country’s soil

- Blue line: Very Important
- Red line: Rather Important
- Green line: Not Important

Probability vs. Leniency graph
Survey Item Response Modeling (briefly)

Adopting the customs of my country

- Very Important
- Rather Important
- Not Important

Leniency vs. Probability graph
Survey Item Response Modeling (briefly)

Abiding by my country’s laws

Probability

Leniency

-4 -3 -2 -1 0 1 2 3 4

Very Important
Rather Important
Not Important
Survey Item Response Modeling (briefly)

Sample Invariant Parameter Estimates

• Item response model parameters are expected to be sample invariant with respect to trait level.

• Samples that vary with respect to trait level are expected to provide equivalent parameter estimates.

**Question.** Are parameter estimates invariant with respect to group identity, e.g., sex, ethnicity?
Survey Sampling Weights (briefly)

What are sampling weights?

• Survey data are often collected using a nonrandom sampling procedure, e.g., cluster sampling, stratified sampling, multistage sampling; subpopulations may have been over- or under-sampled.

• With known probability of being included (inclusion probabilities), sampling weights can be computed.

• Sampling weights are the inverse of the inclusion probability of being selected given the sampling design.
Survey Sampling Weights (briefly)

Why use sampling weights?

• They provide unbiased parameter estimates
• They provide relatively accurate standard errors
Survey Sampling Weights (briefly)

Why don’t we use sampling weights?.

• Weights are often unavailable.
• Not all computing software enables use of weights.
• Complex methods may be needed to properly use survey weights, e.g., multilevel regression modeling, computation of normalized weights rather than raw weights.
Item Response Modeling and Sampling Weights

2007 National Household Education Surveys Program
Parent and Family Involvement in Education Survey
(http://nces.ed.gov/nhes/surveytopics.asp)

Complex Sampling Design
Subsampling of telephone numbers
2-phase stratification by minority
Phase 1—over sampling areas with high % of Black or Hispanic
Phase 2—within minority stratum mailable
Item Response Modeling and Sampling Weights

2007 National Household Education Surveys Program
Parent and Family Involvement in Education Survey
(http://nces.ed.gov/nhes/surveytopics.asp)

Case Level Sampling Weights
80 replicates (80 subsamples)
Full sample weight (average of the 80 replicates)
Item Response Modeling and Sampling Weights

Methods

• 10,681 Children Attending Public or Private School
  • Unweighted (13.3% in private school, N=1,687)
  • Weighted (11.1% in private school)

• Item Response Modeling. Satisfaction with School Features; Graded response model parameter estimation

• Test of Measurement Invariance using SEM with unweighted and weighted samples
Item Response Modeling and Sampling Weights

2007 Parent and Family Involvement in Education Survey

Would you say that you are very satisfied, somewhat satisfied, somewhat dissatisfied, or very dissatisfied with . . .

<table>
<thead>
<tr>
<th>Item</th>
<th>Very Dissatisfied</th>
<th>Somewhat Dissatisfied</th>
<th>Somewhat Satisfied</th>
<th>Very Satisfied</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. School</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>b. Teachers</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>c. Academic Standards</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>d. Order and Discipline</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>e. Staff/Parent Interaction</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>
Item Response Modeling and Sampling Weights

Odds That Respondents from Private Schools are Satisfied Compared to Respondents from Public Schools

<table>
<thead>
<tr>
<th>School Feature</th>
<th>Unweighted</th>
<th>Weighted</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Odds Ratio</td>
<td>CI</td>
</tr>
<tr>
<td>School</td>
<td>5.13</td>
<td>1.75, 8.79</td>
</tr>
<tr>
<td>Teacher</td>
<td>2.80</td>
<td>1.75, 4.47</td>
</tr>
<tr>
<td>Standards</td>
<td>4.94</td>
<td>2.83, 8.64</td>
</tr>
<tr>
<td>Discipline</td>
<td>3.91</td>
<td>2.53, 6.04</td>
</tr>
<tr>
<td>Communication</td>
<td>3.57</td>
<td>2.39, 5.33</td>
</tr>
</tbody>
</table>

Clearly, respondents from private schools are more satisfied.
## Item Response Modeling and Sampling Weights

### Unweighted Data

Model Information Fit Statistics for Model Comparison

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>BIC</th>
<th>A-BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null</td>
<td>42754.512</td>
<td>42865.820</td>
<td>42811.800</td>
</tr>
<tr>
<td>School</td>
<td>42740.661</td>
<td>42871.611</td>
<td>42808.058</td>
</tr>
<tr>
<td>Teacher</td>
<td>42734.264</td>
<td>42865.214</td>
<td>42801.661</td>
</tr>
<tr>
<td>Standards</td>
<td>42731.756</td>
<td>42862.707</td>
<td>42799.153</td>
</tr>
<tr>
<td>Discipline</td>
<td>42746.041</td>
<td>42876.991</td>
<td>42813.438</td>
</tr>
<tr>
<td>Communication</td>
<td>42752.954</td>
<td>42883.904</td>
<td>42820.351</td>
</tr>
</tbody>
</table>

The **Standards** model has the best fit among the models compared.
Item Response Modeling and Sampling Weights

Weighted Data

Model Information Fit Statistics for Model Comparison

<table>
<thead>
<tr>
<th>Model</th>
<th>AIC</th>
<th>BIC</th>
<th>A-BIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null</td>
<td>43111.932</td>
<td>43223.240</td>
<td>43169.220</td>
</tr>
<tr>
<td>School</td>
<td>43102.391</td>
<td>43233.342</td>
<td>43169.789</td>
</tr>
<tr>
<td>Teacher</td>
<td>43088.161</td>
<td>43219.111</td>
<td>43155.558</td>
</tr>
<tr>
<td>Standards</td>
<td>43085.894</td>
<td>43216.844</td>
<td>43153.291</td>
</tr>
<tr>
<td>Discipline</td>
<td>43098.215</td>
<td>43229.166</td>
<td>43165.612</td>
</tr>
<tr>
<td>Communication</td>
<td>43113.756</td>
<td>43244.706</td>
<td>43181.153</td>
</tr>
</tbody>
</table>

The best fitting model is the **Standards** model.
# Item Response Modeling and Sampling Weights

## Satisfaction with Teachers Parameters by School Type

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Public</th>
<th>Private</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est</td>
<td>SE</td>
</tr>
<tr>
<td>unweighted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category 1 – 2+</td>
<td>-6.931</td>
<td>0.150</td>
</tr>
<tr>
<td>Category 2 – 3+</td>
<td>-4.762</td>
<td>0.094</td>
</tr>
<tr>
<td>Category 3 - 4</td>
<td>-1.130</td>
<td>0.060</td>
</tr>
<tr>
<td>weighted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category 1 – 2+</td>
<td>-6.959</td>
<td>0.203</td>
</tr>
<tr>
<td>Category 2 – 3+</td>
<td>-4.731</td>
<td>0.135</td>
</tr>
<tr>
<td>Category 3 - 4</td>
<td>-1.096</td>
<td>0.086</td>
</tr>
</tbody>
</table>
## Item Response Modeling and Sampling Weights

**Satisfaction w/ Standards Parameters by School Type**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Public</th>
<th></th>
<th></th>
<th>Private</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Est</td>
<td>SE</td>
<td></td>
<td>Est</td>
<td>SE</td>
</tr>
<tr>
<td>unweighted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category 1 – 2+</td>
<td>-6.276</td>
<td>0.130</td>
<td></td>
<td>-6.927</td>
<td>0.615</td>
</tr>
<tr>
<td>Category 2 – 3+</td>
<td>-4.460</td>
<td>0.089</td>
<td></td>
<td>-5.140</td>
<td>0.323</td>
</tr>
<tr>
<td>Category 3 - 4</td>
<td>-0.984</td>
<td>0.059</td>
<td></td>
<td>-1.731</td>
<td>0.147</td>
</tr>
<tr>
<td>weighted</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Category 1 – 2+</td>
<td>-6.287</td>
<td>0.171</td>
<td></td>
<td>-7.755</td>
<td>0.629</td>
</tr>
<tr>
<td>Category 2 – 3+</td>
<td>-4.360</td>
<td>0.121</td>
<td></td>
<td>-4.816</td>
<td>0.464</td>
</tr>
<tr>
<td>Category 3 - 4</td>
<td>-0.887</td>
<td>0.091</td>
<td></td>
<td>-1.754</td>
<td>0.189</td>
</tr>
</tbody>
</table>
Item Response Modeling and Sampling Weights

Satisfaction with Teachers (unweighted)

![Graph showing satisfaction distribution for public and private sectors]
Item Response Modeling and Sampling Weights

Satisfaction with Teachers (weighted)
Item Response Modeling and Sampling Weights

Satisfaction with Standards (unweighted)

- Public:
  - Very dissatisfied (blue)
  - Dissatisfied (red)
  - Satisfied (green)
  - Very satisfied (purple)

- Private:
  - Very dissatisfied (blue)
  - Dissatisfied (red)
  - Satisfied (green)
  - Very satisfied (purple)
Item Response Modeling and Sampling Weights

Satisfaction with Standards (weighted)

- Public:
  - Very dissatisfied: blue dashed line
  - Dissatisfied: red dashed line
  - Satisfied: green line
  - Very satisfied: purple line

- Private:
  - Very dissatisfied: blue dashed line
  - Dissatisfied: red dashed line
  - Satisfied: green line
  - Very satisfied: purple line

Spring 2014 CORE Survey Methodology
Colloquium Series
Item Response Modeling and Sampling Weights

**Question.** Are survey sampling weights relevant when estimating item response parameters?

- Parameter estimates are affected by weights, they will theoretically be less biased.
- Parameter standard errors tend to be affected by weights; they tend to increase.
- Whether estimating parameters or testing for measurement invariance, use sampling weights when possible.
Item Response Modeling and Sampling Weights

**Question.** Are survey sampling weights relevant when estimating item response parameters?

- If the weights change representation of subsamples and if the items are biased with respect to subsamples, then the estimates are biased in both cases. The weighted data provides a better basis for testing for sample invariance.

- Be sure to use software that accommodates weights.
References and Resources


References and Resources


