**EPSY 5221: Principles of Educational & Psychological Measurement**

Interpreting Model Fit in IRT (the Rasch Model)

Model-data fit is always an important aspect of the data to evaluate whenever doing statistical analysis. This essentially amounts to checking the assumptions of the model. In IRT, there are a number of ways to evaluate model-data fit. We are focusing on two.

**Item Fit**

First is the item fit statistic. We do this with a weighted mean squares statistic generated by the model (WMS). If the data fit the model for items, then the data are consistent as the model predicts.

*Model Prediction: The model predicts the probability of correct response to the item based on person ability.*

For an item, we expect people with lower abilities to have lower probabilities of correct responses and people with higher abilities to have higher probabilities of correct responses. If we look at groups of people with various levels of ability across the measure, then smaller proportions with lower scores get the item correct and larger proportions with higher scores get the item correct. If this is what we observe in the data, then "the item fits the model."

If an item produces responses where lower-ability people have a higher than predicted probability of correctly responding and/or people with higher abilities have a lower than predicted probability of correctly responding, then the item doesn't fit the model. When this occurs, the WMS value increases (bigger differences between what we observe in the data and what the model predicts).

When WMS values go beyond 1.5, we are concerned about the quality of the item. If the item doesn't fit the model, there is a good chance that the item isn't measuring the same thing as the overall test - or that the item has serious flaws. Some high-stakes testing programs use more conservative criteria for making fit decisions, such as when WMS values are larger than 1.2 or 1.3, the item is reviewed for misfit.



For item 54 above, the WMS value is 1.17, not bad, just a little over 1.0 which is ideal.

We can see that the observed proportion of correct response (y-axis) does not map exactly to the model-predicted curve.

At some points around lower abilities (particularly -2.0 to -1.0), a larger proportion of people answer the item correctly than the model predicts – and in areas of moderate ability (particularly 1.0 to 3.0), a smaller proportion answer the item correctly than the model predicts. It appears that this item might be less discriminating (have a lower slope) than the model predicts.



For item 43 above, the WMS is 1.01. This item fits the model well, since the observed data fit the model predictions.

**Person Fit**

Second is the person fit statistic. We can also interpret fit for persons in the same way, also using a WMS statistic for persons. The model predicts the probability of correctly responding to the items on the test based on person ability. If a person responds to items in a way that is consistent with their estimated ability, then the person is responding in a way that is consistent with the model (as the model predicts response to an item as a function of their ability).

*Model Prediction: The model predicts the probability of correct response to items on the test based on person ability.*

This is difficult to graph. But conceptually, it goes like this:

If a person correctly answers items that are easier than their ability level and incorrectly responds to items that are more difficult than their ability level, then the person is responding in ways that are consistent with their ability. The model predicts that the only thing that determines the probability of correctly responding to an item is a person’s ability. So, knowing their ability, we should be able to predict which items they will answer correctly and which they will answer incorrectly.

If a person answers items in a way that is consistent with their ability, the person will fit the model – have WMS values near 1.0.

If a person answers items in a way that is not consistent with their ability, the person will not fit the model – have WMS values that are larger. When the values go above 1.5, we are concerned that this person is not paying attention, putting forth their best effort, or has information about correct responses to some items (for example, in some cases, person fit is used to detect cheating).