**ITEM RESPONSE THEORY MODEL SUMMARIES**

Primary reference is:

De Ayala, R.J. (2009). The theory and practice of item response theory. New York, NY: Guilford.

|  |
| --- |
| **IRT MODEL: *Rasch*** |
| *Data & Sample Size Requirements:*  Dichotomously scored items. Item and person parameters can be estimated with at least 100 persons; but analyses involving DIF, fit, and equating require greater precision and stability, with at least 300 responses per item. |
| *Model Specification:* |
| *Model Framework:*  The Rasch model is based on the goal of constructing measures. Discriminations should be similar across items, held constant at 1.0 in the Rasch model; item locations vary. When data do not conform to the model, it is a fundamental data or measurement problem – one that we fix through measure modifications, rather than finding another model that fits the data. It is a parsimonious model, with the summed-score as a sufficient statistic for estimation, and a one-to-one correspondence between number correct and theta. |
| *Item Parameter Interpretations:*  The item parameter estimated with the Rasch model is the item location or item delta (δ). It is the ability required to have a 50% probability of correctly responding to the item. It is the location where the ICC has its inflection point (the point at which the slope is estimated, which in the Rasch model = 1). |
| *Other Considerations:*  The Rasch model is a strong approach to measurement construction. When its requirements are met, the model results in objective measurement that meets the properties of invariance and produces an interval scale of measurement. JMLE is known to produce biased parameter estimates with small numbers of items and small samples of persons. |

|  |
| --- |
| **IRT MODEL: *1PL*** |
| *Data & Sample Size Requirements:*  Dichotomously scored items (any item type), which are unidimensional, and have local independence. As with the Rasch model, sample sizes can be as low as 100, but ideally should be over 300. |
| *Model Specification:* |
| *Model Framework:*  The 1-parameter logistic model is one that is fitted to the data. The model is parsimonious with a one-to-one correspondence between number correct and theta. The item discrimination (α) is estimated, but held constant across items. The 1PL model is equivalent to the Rasch model when α = 1. |
| *Item Parameter Interpretations:*  The item parameter estimated with the 1PL model is the item location or item delta (δ). It is the ability required to have a 50% probability of correct response to the item. It is the location where the ICC has its inflection point (the point at which the slope is estimated, from which the item discrimination is estimated, α).  The item discrimination (α) is a constant that provides the best fit for this data to the model -- for the 1PL model, the value of α is the same for every item. The value of α indicates how well an item can distinguish between people of different abilities, and the maximum discrimination occurs at the item’s location. For example, when α is low, there is little difference in the probability of answering the item correctly between two people who are close in ability; however, if α is high, there is a greater difference in that probability between two people who are close in ability, especially for abilities near the item location. |
| *Other Considerations:*  The item information curve is centered at the item location, which can be useful to create the total test information curve. This allows us to see where the test provides the most information, relative to person ability, and may be useful for constructing a test with regards to a specific purpose or to support precision at a cut score. |

|  |
| --- |
| **IRT MODEL: *2PL*** |
| *Data & Sample Size Requirements:*  Dichotomously scored items (any item type), which are unidimensional, and have local independence. De Ayala recommends a sample size of at least 500 persons and 20 items. |
| *Model Specification:* |
| *Model Framework:*  The 2-parameter logistic model is one that is fitted to the data. The item location (δ) and item discrimination (α) are estimated for each item. The estimation of α for each item is the distinguishing characteristic of the 2PL model from the 1PL model. |
| *Item Parameter Interpretations:*  The item location or item delta (δ) is the ability required to have a 50% probability of correct response to the item. It is the location where the ICC has its inflection point (the point at which the item discrimination, α, is estimated).  The item discrimination (α) is a value that provides the best fit for each item to the model. For the 2PL model, the value of α is estimated for every item. The interpretation of α is the same as for the 1PL model: the value of α indicates how well an item can distinguish between people of different abilities. As α increases, so does the item’s ability to discriminate between two people. The maximum discrimination for an item occurs at its location. Discrimination values of 0.8 to 2.5 are generally considered good (de Ayala, p. 101). |
| *Other Considerations:*  The item information curve is centered at the item location, which can be useful to create a curve of the total information for a test. This allows us to see where the test provides the most information, relative to person ability, and may be useful for constructing a test with regards to a specific purpose or to support precision at a cut score. |

|  |
| --- |
| **IRT MODEL: *3PL*** |
| *Data & Sample Size Requirements:*  Dichotomously scored items (any item type), which are unidimensional, and have local independence. De Ayala recommends a sample size of at least 1,000 people and 20 items. Larger samples will provide a better basis of estimating abilities at the low-end of the ability continuum to estimate the lower asymptote (χ-parameter) with sufficient precision. |
| *Model Specification:* |
| *Model Framework:*  The 3-PL model is fitted to the data. The item location (δ), item discrimination (α), and lower asymptote (χ) are estimated for each item. The addition of the lower asymptote (χ) is the distinguishing characteristic of the 3PL model from the 2PL model. |
| *Item Parameter Interpretations:*  The interpretation of α is the same for 3PL as for 2PL.  The item location or item delta (δ) is the ability required to have a [(1.0 + χ ) / 2]% probability of correct response. It is the location where the ICC has its inflection point (the point at which the item discrimination, α, is estimated).  The lower asymptote χ is the probability of a correct response for test-takers at the lowest level of the ability continuum. The value can range from 0 to 1, and represents a floor for the lowest probability of a correct answer for any student. It does not necessarily represent the probability of guessing the correct answer, because we can account for small amounts of student ability beyond guessing (e.g., test-taking skills, partial knowledge). |
| *Other Considerations:*  The 2PL model is equivalent to a 3PL model where χ = 0. Smaller values of χ are more desirable than greater values; items with large values of χ should be closely examined for possible flaws in the item, or issues with the model-fit for the 3PL model with this set of data.  When comparing 2PL and 3PL models for the same set of data, the 3PL model re-estimates every parameter from the 2PL model (i.e., it does not simply “add a parameter” to the estimated 2PL values). The 3PL model will often report the items as being more difficult and less discriminating. The 3PL model will usually yield less information -- than the 2PL model).  The 3PL model requires a sufficient number of people at the very low end of the ability continuum, in order to accurately estimate the lower asymptote values. If the sample does not include enough people at the low end of the ability continuum, the estimate of χ could be inaccurate and misleading; poor estimation of one parameter affects precision of all others.  For the 3PL model, the item information curves are not symmetrical, and thus not centered at the item location. They cannot be used the same way that the 1PL and 2PL item information curves can be used. |

|  |
| --- |
| **IRT MODEL: *Partial Credit (PC)*** |
| *Data & Sample Size Requirements:*  Polytomously scored ordinal item responses, which are unidimensional. A sample size should be 2-5 times larger than the number of parameters, with a lower bound of around 250. De Ayala points out that there is likely a large enough sample size after which accuracy in estimation is not improved much – maybe around 1200 (p. 199). |
| *Model Specification:* |
| *Model Framework:*  Master’s Partial Credit model estimates transition locations for adjacent ordered categories, where distances between transitions vary across items. This is an extension of the Rasch model where item discriminations equal 1. The number of category options can vary by item, making the Partial Credit model useful for tests with multiple format items. The number of transitions for an item is one less than the item’s number of category options. The counts for each category are sufficient statistics for determining transition locations. A test-taker’s observed score is sufficient for estimating ability.  The total number of parameters estimated is the product of the number of items and the number of transition locations. A maximum likelihood function is used to estimate parameters using JMLE or MMLE. |
| *Item Parameter Interpretations:*  Threshold is the location of the transition point *h* for item *j* and can be interpreted as a difficulty parameter. The transition location is interpreted as the ability required to have equal probability of choosing one of two adjacent category options for item *j*. |
| *Other Considerations:*  Option response functions (ORFs) show the probabilities of the response options being selected as a function of theta. The transition locations are where consecutive ORFs intersect. The ORF for the first category is nonincreasing and for the last category is nondecreasing. ORFs for intermediate categories are unimodal. ORFs are also known as category probability curves, category response functions, operating characteristic curves, and option characteristic curves.  Although the option categories must be ordered in the Partial Credit model, the transition locations can be disordered.  Items contain more information if polytomously scored rather than dichotomously, and in fact, items with more categories give more information than items with fewer categories (p. 201).  The Partial Credit model simplifies to the Rasch model if there are only two options for all items (dichotomous). |

|  |
| --- |
| **IRT MODEL: *Rating Scale (RS)*** |
| *Data & Sample Size Requirements:*  Polytomously scored ordinal item responses. A sample size should be 2-5 times larger than the number of parameters, with a lower bound of around 250. De Ayala points out that there is likely a large enough sample size after which accuracy in estimation is not improved much – maybe around 1200 (p. 199). |
| *Model Specification:* |
| *Model Framework:*  This is a constrained version of the PC model, where the distances between the thresholds are held constant across items. Item locations vary. The number of category options must be the same for all items. The number of thresholds for an item is one less than the item’s number of category options. The counts for each category are sufficient statistics for determining thresholds. A person’s observed score is sufficient for estimating a person’s location.  The total number of parameters estimated is the sum of the number of items and the number of thresholds.  A maximum likelihood function is used to estimate parameters using JMLE or MMLE. |
| *Item Parameter Interpretations:*  δ*j* is the location of item *j* with *mj* thresholds. *xj* is the number of thresholds “passed” due to the ordering of the responses. τ*h* is the constant difference between thresholds (Andrich thresholds). Relating this back to the Partial Credit model: δ*jh*=δ*j*+τ*h* is the location of the threshold *h* for item *j*. |
| *Other Considerations:*  ORFs for the Rating Scale model are similar to those for the more general Partial Credit model, although the intersections of consecutive ORFs (thresholds) are equidistant.  Unlike the Partial Credit model, the thresholds are estimated to remain ordered in the Rating Scale model.  The Rating Scale model simplifies to the Rasch model if there are only two options for all items (dichotomous). |

|  |
| --- |
| **IRT MODEL: *Generalized Partial Credit (GPC)*** |
| *Data & Sample Size Requirements:*  Polytomously scored ordered item responses. Requires a larger sample because of the large number of parameters. With MMLE, symmetric distribution, and reasonable distribution of responses across response categories, a sample of at least 500 may be sufficient; more than 1200 will not improve estimation precision much. |
| *Model Specification:*  the *Generalized Rating Scale (GRS)* Model: |
| *Model Framework:*  The Generalized Partial Credit model is based on the assumption that the probability of choosing the *kj*th category over the (*kj*–1)th category is governed by the dichotomous response model. This generalization of the Partial Credit model (developed by Muraki) relaxes the equal-discrimination restriction on all items from the Partial Credit model. This GPC model estimates item discrimination for each item. This is a 2PL model rather than a Rasch model. In this model, δ*j*1 is defined to always be 0. The number of categories can vary across items for the GPC model but is constant for the GRS model.  The number of parameters estimated per item is one more than the number of transitions.  A maximum likelihood function is used to estimate parameters using JMLE or MMLE. |
| *Item Parameter Interpretations:*  The model specifies the probability of providing a response x*jk* in category *k* to item *j*. The item parameters are interpreted the same as for the Partial Credit model (and Rating Scale model), with the addition of the item discrimination (α*j*) as in the 2PL model.  When used to estimate the Rating Scale model, Andrich thresholds are estimated and held constant across items. |
| *Other Considerations:*  The ORFs for the GPC model reflect both item location and discrimination. As with the Partial Credit Model, the intersection of consecutive ORFs are the transition locations.  The GPC model simplifies to the 2PL model if there are only two options for all items (dichotomous). |

|  |
| --- |
| **IRT MODEL: *Graded Response*** |
| *Data & Sample Size Requirements:*  Polytomously scored ordered item responses. This model can be applied to partial-credit and rating-scale data. Requires a larger sample because of the large number of parameters. With MMLE, symmetric distribution, and reasonable distribution of responses across response categories, a sample of at least 500 may be sufficient; more than 1200 will not improve estimation precision much. |
| *Model Specification:* |
| *Model Framework:*  This is a 2PL parameterization of the partial credit model, estimating category boundaries in a cumulative manner, successive 2PL models. This is a cumulative probability model, making its approach to defining categories different than the other polytomous models. The model estimates one discrimination parameter and m-1 category boundaries; there are m + 1 parameters per item. The number of categories may vary across items. |
| *Item Parameter Interpretations:*  The GR model estimates the probability of a person responding with a given category score xj or higher, rather than a lower category. δ*xj* is the “category boundary” location. |
| *Other Considerations:*  This does not give us the probability of obtaining a specific category score – to get those values, we must take the difference between cumulative probabilities for adjacent categories. Because we obtain sequentially ordered δxj values, we cannot review the δxj values to see which categories are not likely to be obtained or chosen. The ORFs must be plotted to see this. Because this is a cumulative probability model, considerations for why to choose this model include:   * When the underlying response continuum or trait is clear * Cognitive items can be graded * Rating scale items that we want to treat as having an underlying continuous trait * Forces category boundary locations to be ordered * Requires a larger sample |

|  |
| --- |
| **IRT MODEL: *Nominal Response*** |
| *Data & Sample Size Requirements:*  Categorical item responses, such that each category is nominal, not ordered. Applies to MC items. With MMLE, symmetric distribution, and reasonable distribution of responses across response categories, a sample of at least 600 may be sufficient; more than 1500 will not improve estimation precision much. |
| *Model Specification:*  where and  or  where |
| *Model Framework:*  Category slopes and locations vary within items, and between items, using a slope-intercept model to estimate parameters for each category. Categories are seen as being independent. It is an extension of the 2PL model, estimating one discrimination and one location parameter for each category. αjk are the slope parameters and γjk are the intercept parameters for mj response categories. |
| *Item Parameter Interpretations:*  The sum of the αk and γk equals zero within each item. For item j, γjk is the individual’s propensity to use response category k, and αjk is the option’s discrimination. |
| *Other Considerations:*  ORFs can be created to illustrate the probability of response to each category as a function of ability (θ). The ORFs’ intersection point is found by setting the corresponding category logits equal and solving for θ.  For dichotomous items, the 2PL and NR models are equivalent, with the transition point δ. The PC, RS, and GPC are special cases of the NR model. When the αjk are forced to increase in steps of one, the NR model is equivalent to PC/GPC models. So the NR model can be applied to ordinal polytomous data. |