Item Response Theory Model for Understanding Item Non-Response in Ghanaian Surveys

Iddrisu Wahab Abdul[*1], Nana Kena Frempong[2] & Martin Owusu Amoamah[1]
[1]Department of Mathematics and Statistics, Accra Polytechnic, Accra, Ghana
[2]Department of Mathematics and Statistics, KNUST, Kumasi, Ghana
[*]E-mail of the corresponding author: perfectwahab@yahoo.com

Abstract

This paper explores four Item Response Theory (IRT) models to determine the most appropriate for understanding item non-response. The selected IRT model was used to identify among five categories of survey questions, the most difficult to answer by respondents and determine the underlying mechanism behind missing data which is defined to include 'don't know' answers. A questionnaire data on Ghana collected in the fifth wave of the World Values Survey was implored. All items were dichotomously scored. Missing or 'don't know' responses were assigned a 0 score whiles answered items were assigned a 1 score. The four IRT models that were explored included both the constrained and unconstrained versions of the Rasch model, the two parameter logistic model (2-PLM), and the three parameter logistic model (3-PLM). The unconstrained Rasch model emerged as the most appropriate model for understanding item non-response. It was observed that, income related questions had the highest difficulty parameter, hence the most difficult category of survey questions to answer. It was also found that, if an individual does not answer a survey question or give a 'don't know' answer, it is not only due to the question's difficulty but also because the respondent doesn't want to answer.

Keywords: Item Non-response, Item Response Theory (IRT), World Values Survey (WVS).

1. Introduction

Surveys are mostly used in conducting research in Ghana. For this reason, Ghanaian researchers are concern about data richness and data quality. Among those concerns, item non-response has caught researchers' special attention. This problem has seriously affected Comparative studies: but what should be done when dealing with a significant quantity of item non-response while interested in drawing meaningful conclusions from the data, since it is not proper to discard it. Proper understanding of the underlying reasons is a huge step in dealing with item non-response.

Ren [8] argued that, any time respondents do not answer a question in a survey, it is possibly due to three reasons: either they do not care, or they do not know, or possibly they do not want to answer. He argued further that don't know is largely due to ignorance, don't care goes to tell the interest of respondents in the item or in the survey in general, whiles don't want to answer is usually associated with fear in the context of politics or desirability in the context of societal norms. However, item non-response in Ghanaian surveys is rarely studied. Even in the absence of problems in relation to sampling, this is a big concern for pollsters in Ghana. The suspicion is that, the increasing rates of item non-response in Ghanaian surveys may be due to the fact that ordinary Ghanaians may not have the cognitive capacity to construct opinions to some particular questions because of low level of education.

According to De Leeuw [4], item non-response may be defined as the inability to get information for an item within a questionnaire. Even though it leads to missing observations for particular questions, it cannot be said that item non-response do not carry information. Rubin [10] differentiated among three kinds of item non-response according to the underlying mechanism: *missing at random* (MAR), *missing completely at random* (MCAR), and *missing not at random* (MNAR). To define the three kinds of item non-response, Rubin [10] distinguished between the observed data Y_{obs} and the missing data Y_{miss}. These constitute the complete data matrix $Y = (Y_{obs}, Y_{miss})$. We adapted this notation to the latent variable framework. Y is the complete data matrix that consists of the observed item responses Y_{obs} and the omitted responses Y_{miss} of the k items Y_1 to Y_k, indexed by i. The values of a latent variable ξ can also be considered to be missing data. The MCAR is the case where the distribution of the item non-response data is independent of the item response data. The MAR holds if the distribution of the missing mechanism is only dependent on the observed data but not dependent on the unobserved values of the missing data. The third type, called MNAR is the opposite of MAR.

In IRT, item parameters and latent traits are related to the probability of a correct answer for both polytomous and dichotomous data. This paper was therefore motivated by the need to use *Item Response Theory (*IRT) to

(a) Identify the most appropriate IRT model for understanding item non-response;
(b) Identify the categories of survey questions that are most difficult to answer by respondents: and,
(c) Find out the reason behind 'don't know' responses and missing data; whether respondents don't really know, or don't want to answer an item.

2. Method

2.1 Data

Data for this study was obtained from the World Values Survey (WVS). It is a research exercise that looks into a wide range of topics from values to politics. Data on Ghana from the 5th wave (WVS) was used for this study. It was carried out by the Institute of Social Research (ISR) at the University of Michigan. The Ghanaian survey in this wave was conducted by Markinor Thinking. The survey period for Ghana was from 19th February to 04th April 2007 which included a sample of about 1534 individuals.

2.2 Sample Selection

We grouped all the questions into five categories: Life Related Questions (LRQ), Value Related Questions (VRQ), Political Related Questions (PRQ), Income Related Questions (IRQ), and Democracy Related Questions (DRQ). LRQ relates to questions on life, confidence, morality, religion, and marriage whereas VRQ relates to questions that reflect environmental and personal values. PRQ include questions on politics and its systems. IRQ consist of those relating to family savings and scale of income whiles DRQ consists of those on governance and democracy.
We finally selected one question randomly from each of the categories to be used for the IRT modeling and construction of item characteristics curves. Based on literature study, manual coding scheme was used to code all items. All items were either assigned a value zero or one. Zero was used when an item was not answered or when a 'don't know' answer was provided, and one was assigned to all answered items. More on the sampling design is found in (*http://www.worldvaluessurvey.org*) [12].

2.3 Modeling Approach

A variety of measurement models has been used in this study. The dependent variable in this study, item non-response, is measured as whether or not an item is answered by an individual. This is a dichotomously scored variable: 0 = no answer or 'don't know' answer to the item, 1 = answer to the item. The item non-response variable is extremely skewed; most people answered most items. Such data violate several assumptions of the usual regression methods. For this reason, IRT was used to look into the psychometric characteristics of item non-response and to come out with the latent item non-response trait. The advantages of IRT in comparison with other psychometric methods are numerous. In IRT error estimates specific to the level of trait are provided. Also, estimates for the level of latent trait are neither scale-dependent nor group-dependent. In this study, IRT is used to demonstrate whether an individual's ability score indicates the probability of responding to an item. Among its advantages, IRT provides the relationship between answering or not answering a question, and the individual's latent trait (Question knowledge). The various IRT models explored in this study are described below;

$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}} \qquad (1)$$

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \qquad (2)$$

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \qquad (3)$$

The model in equation (1) was first proposed by Rasch [7]. Birnbaum [2] extended the Model in equation (1) to obtain the model in (2). Finally, Lord [5] extended the model in equation (2) to obtain the model in (3).

In the above models from (1) to (3), θ is a continuous variable (latent Question Knowledge trait) and for $i = 1,2,3,\ldots,n$, $P_i(\theta)$ is the probability of an individual with ability θ responding to item *i*, a_i is the item discrimination parameter for the *ith* item, b_i is the difficulty parameter for the *ith* item, c_i is the item pseudo-chance parameter for the *ith* item, *e* is the natural logarithm constant whose value can be approximated to 2.718, *D* is used to approximate the logistic model to the normal model, and *n* is the size of the respondents. The analysis was done using an *R* package for IRT analyses [6].
Three assumptions must be satisfied when using IRT models. The first assumption referred to as unidimensionality assumption which implies that the probability of responding to a question is a function of just one latent trait. The second assumption referred to as local independence requires no relationship in an individual's responses to different items after taking into account the individual's latent trait level. Finally, the

response of an individual to an item can be modeled by an *item response function* (IRF).

3. Results

Descriptive statistics for the data are produced in Table 1. The output contains among others the proportions for all the possible response categories for each item. We observe from Table 1 that the life related questions seems to have least difficult questions having the highest proportion of about 99% of responses, while the income related questions seems to be the most difficult one having the lowest proportion 90% of responses. The proportion of responses for the politics related question, democracy related question, and value related question were about 93%, 94%, and 95% of the respondents respectively.

Table 1: Proportions for each level of response

Proportions for each level of response:	0(%)	1(%)
Life Related Questions (LRQ)	1.17	98.83
Politics Related Questions (PRQ)	6.98	93.02
Democracy Related Questions (DRQ)	5.67	94.33
Value Related Questions (VRQ)	4.76	95.24
Income Related Questions (IRQ)	10.43	89.57

We have the χ^2 p-values for pairwise associations between the five items, corresponding to the 2×2 contingency tables for all possible pairs. Before an analysis with latent variable models, it is useful to inspect the data for evidence of positive correlations. In this case, the ad hoc checks are performed by constructing the 2×2 contingency tables for all possible pairs of items and examine the chi-squared p-values. Inspection of non-significant results can be used to reveal 'problematic' items [9]. We observe from Table 2 that three pairs of items seem to have weak degree of association, and the life related item is included in all three pairs. The small number of non significant pairwise association poses the data for IRT modeling.

Table 2: Pairwise associations

	Item i	Item j	p. value
1	LRQ	VRQ	1.000*
2	LRQ	PRQ	0.638*
3	LRQ	DRQ	0.071*
4	LRQ	IRQ	0.002
5	DRQ	VRQ	1e-03
6	VRQ	IRQ	1e-04
7	PRQ	VRQ	1e-06
8	PRQ	DRQ	1e-12
9	PRQ	IRQ	2e-13
10	DRQ	IRQ	2e-16

3.1 The Constrained Rasch Model

We start by fitting the original form of the Rasch model that assumes known discrimination parameter fixed at value one. In this Model, a respondent is characterized by a level on a latent trait (Question knowledge), and an item is characterized by a degree of difficulty. The larger the value of the difficulty parameter implies the more difficult the question. Table 3 presents results for the constrained Rasch model parameter estimates. The results of the descriptive analysis above are also validated by the model fit in Table 3, where the income related questions and the life related questions are the most difficult and the least easy, respectively. A transformation of the parameter estimates into probability estimates results is computed. The probability of responding to an item is seen as a function of the ratio of a respondent's level on the trait (Question Knowledge) to the item difficulty. The column $P(X = 1|Z = 0)$ denotes the probability of responding to the *ith* item for the average individual. These probabilities were sorted according to the difficulty estimates as shown in Table 3. We observe that the probability of the average individual responding to the life related question is higher than responding to the other related question.

Table 3: Difficulty and probability estimates under the constrained Rasch model

	Difficulty	Discrimination	P(x=1|z=0)
LRQ	-4.9433	1	0.9929
PRQ	-3.4492	1	0.9692
DRQ	-3.2511	1	0.9627
VRQ	-3.0124	1	0.9531
IRQ	-2.5282	1	0.9261

3.2 The One-Parameter Logistic Model (1-PLM), Unconstrained Rasch Model

Both the constrained and the unconstrained Rasch models have similar features and are mathematically equivalent except that, where the constrained Rasch model had a fixed slope of unity for all items, the unconstrained Rasch model only requires the slope to be equal for all items. The discrimination parameter estimated from this model is 1.602 which is different from one. Comparing the difficulty parameters under this model, similar results shows that the income related question and the life related question are most difficult and easiest respectively. The probability of an average individual under the unconstrained Rasch model responding to the life related question is higher than responding to the value related question. Similarly, the probability of responding to the democracy related question is higher than responding to the politics related question for the average individual under the unconstrained Rasch model.

Table 4: Results for the unconstrained Rasch model

	Difficulty	Discrimination	P(x=1|z=0)
LRQ	-3.4995	1.6016	0.9963
VRQ	-2.5059	1.6016	0.9822
DRQ	-2.3699	1.6016	0.9780
PRQ	-2.2046	1.6016	0.9716
IRQ	-1.8644	1.6016	0.9519

3.3 The Two-Parameter Logistic Model (2-PLM)

Here, we explore how the two-parameter logistic fits the data. Whereas the Rasch models constrain the discrimination parameter to be equal, the two-parameter logistic model allows the slope or discrimination parameter to vary across items. Discrimination is deemed high if its value is greater than 1.35 [1]. Table 5 presents results for the *2-PLM* estimates which show that the discrimination parameter estimates is not the same for all items. Comparing the difficulty parameters under this model, we observe in Table 5 that the income related question and the life related question are most difficult and easiest respectively. In terms of discrimination, we observe that, all the questions have high discrimination, especially the democracy related question. The probability of an average individual under the two-parameter logistic model responding to the life related question is higher than responding to all other related questions. Similarly, the probability of responding to the democracy related question is higher than responding to the politics related question for the average individual under the two-parameter logistic model.

Table 5: Difficulty, Discrimination and Probability estimates under the 2-PLM

	Difficulty	Discrimination	P(x=1|z=0)
LRQ	-4.7096	1.0534	0.9930
VRQ	-2.7544	1.3655	0.9773
PRQ	-2.2738	1.5143	0.9690
DRQ	-2.0177	2.2710	0.9899
IRQ	-1.8684	1.5945	0.9516

3.4 The Three-Parameter Logistic Model (3-PLM)

Here, we explore how the three-parameter logistic model fits the data. Whereas the Rasch models constrain the discrimination parameter to be equal, the three-parameter logistic model allows the slope or discrimination parameter to vary across items and also incorporates a guessing parameter. This model is usually employed to handle the phenomenon of non-random guessing in the case of difficult items. Comparing the difficulty parameters under this model in Table 6, we observe that the value related question and the life related question are most difficult and easiest respectively. In terms of discrimination, we observe from Table 6 that, all the questions have very high discrimination. It is important to mention that under the three-parameter model, the values of the guessing parameter are not apparent since difficulty values are less than zero and discrimination values are greater than one [1].

Unlike the *1-PLM* and the *2-PLM*, it is shown in Table 6 that the probability of an average individual under the *3-PLM* responding to the life related question is lower than responding to the value related question. Also, the probabilities of responding to the democracy related question, the politics related question, and the income related question for the average individual under the three-parameter logistic model is certain.

Table 6: Guessing, Difficulty, Discrimination and Probability estimates under the 3-PLM

	Guessing	Difficulty	Discrimination	P(x=1|z=0)
LRQ	0.0547	-4.1906	1.2071	0.9940
IRQ	0.2908	-1.1578	56.9407	1.0000
DRQ	0.7933	-0.6688	42.6518	1.0000
PRQ	0.7869	-0.6422	47.5954	1.0000
VRQ	0.8678	-0.3507	36.9218	0.9999

3.5 Model Selection

To determine which of the four IRT models fitted above is the most appropriate for the data, the goodness of fit indicators which compares the unconstrained version of the Rasch model, the constrained Rasch model, the two-parameter logistic model, and the three parameter logistic models are used. The estimated goodness of fit indicators in Table 7 shows that the unconstrained Rasch model has the smallest AIC value (3104.63) and BIC value (3136.64), hence the more suitable for the data [3].

Table 7: Likelihood ratio test for the unconstrained Rasch, Constrained Rasch, 2-PLM and the 3-PLM

Likelihood Ratio Table	AIC	BIC	log.Lik	LRT	df	p.value
Unconstrained Rasch	3104.63	3136.64	-1546.31			
Constrained Rasch	3136.83	3163.51	-1563.41	34.2	1	<0.001
2-PLM	3106.11	3159.46	-1543.05	6.52	4	0.163
3-PLM	3109.89	3189.92	-1539.95	12.74	9	0.175

Adopting the unconstrained Rasch model as the most appropriate for our data, we produce results for the estimated Item Characteristic, the Item Information and the Test Information Curves. The Item Characteristic Curve (ICC) is the basic building block in IRT. The ICC models the relationship between a person's probability of responding to an item category and the level on the construct measured by the scale [1]. The properties of the ICC needed to describe the item's characteristics are its location and the steepness. The steepness of the ICC reflects the discrimination property of an item whereas the difficulty parameter which is represented by location is the point on the ability scale at which the probability of responding to the item is 0.5. We observe from Figure 1 that the life related question and the income related question are the easiest and the most difficult respectively.

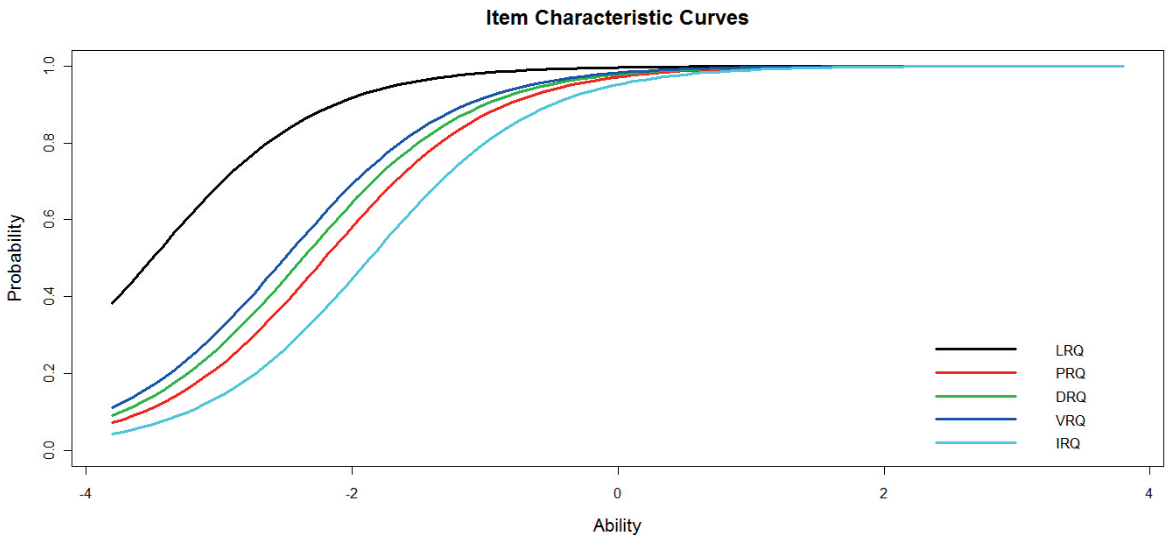

Figure 1: Estimated Item Characteristic Curve obtained from the data

Item information is the amount of information based upon a single item. It can be computed at any ability level. Because only a single item is involved, the amount of information at any point on the ability scale is going to be rather small. An item measures ability with greatest precision at the ability level corresponding to the item's difficulty parameter [1]. We observe from Figure 2 that the amount of item information for each item decreases as the ability level departs from the item difficulty and approaches zero at the extremes of the ability scale.

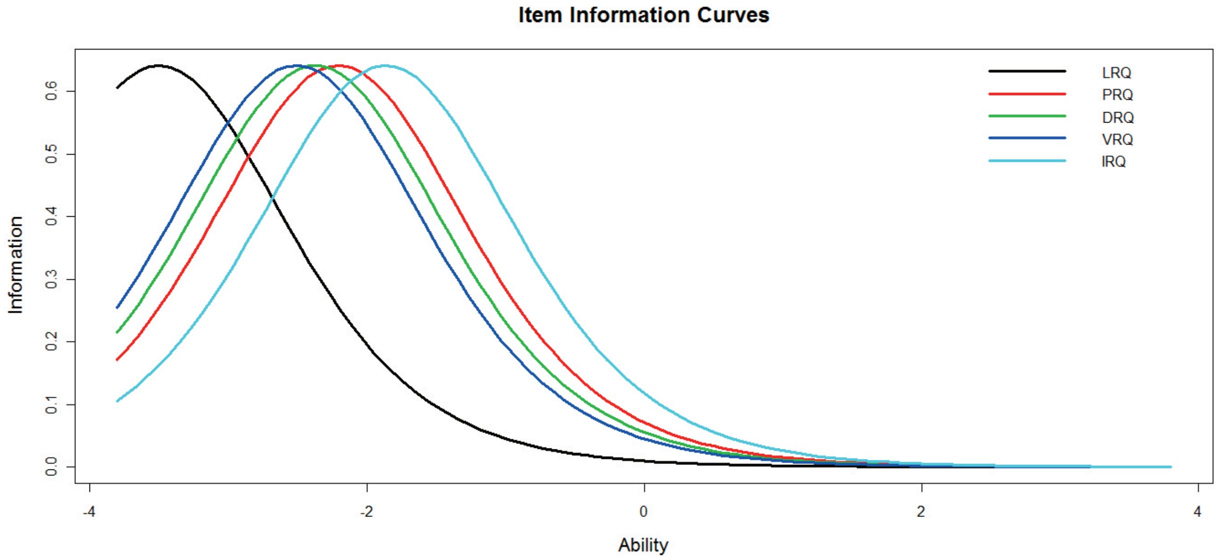

Figure 2: Estimated Item Information Curve obtained from the data

Since a test is used to estimate the ability of an individual, the amount of information yielded by the test at any ability level can also be obtained. A test is a set of items; therefore, the test information at a given ability level is simply the sum of the item information at that level. The general level of the test information function will be much higher than that for a single item information function. Thus, a test measures ability more precisely than does a single item [1]. We observe from Figure 3 that the maximum value of the test information function is at ability level -2. However, as the ability level increases, the amount of test information decreases significantly. This indicates that the items asked in our data mainly provide information for respondents with low ability. In particular, the amount of test information for ability levels in the interval (−4, 0) is almost 90%.

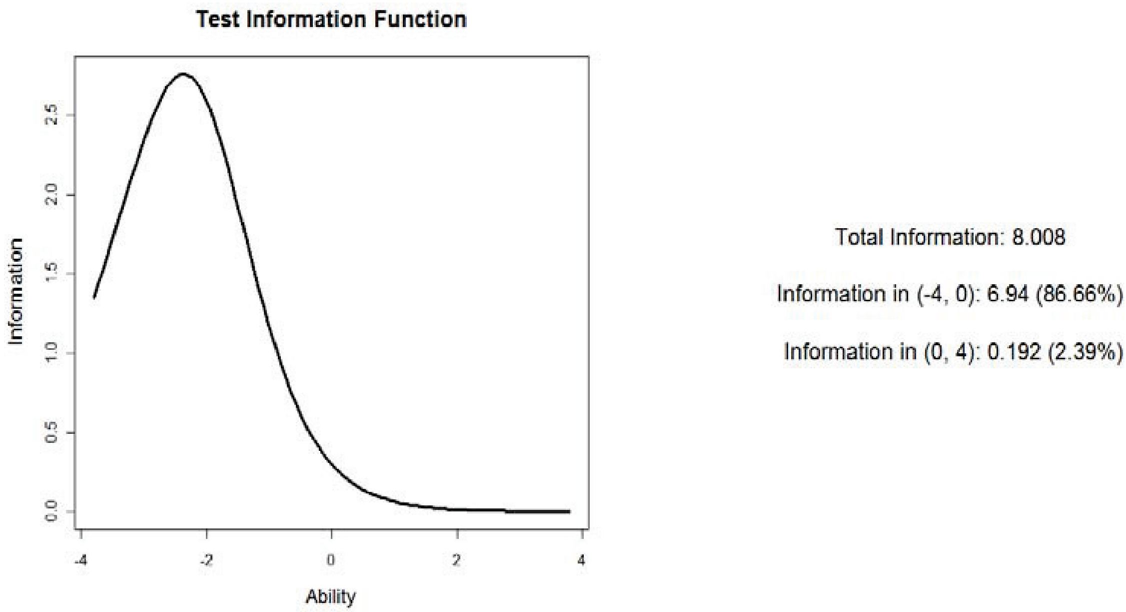

Figure 3: Estimated Test Information Curve

3.6 Ability Estimates

Finally, the ability estimates for respondents are obtained. The primary purpose for using IRT in this study is to locate respondents on the ability scale. Since this will help us evaluate respondents in terms of how much underlying ability (Question knowledge) they possess. Factor scores or ability estimates are summary measures of the posterior distribution $P(Z/X)$, where Z denotes the vector of latent variables and X the vector of manifest variables. By default factor scores produces ability estimates for the observed response patterns. The items asked in the data mainly provide information for respondents with low ability (i.e., below 0). That is, most of the items in the dataset are relatively easy for the average respondent to answer. Figure 4 is a Plot of a Kernel Density Estimation of the distribution of the factor scores (i.e., person parameters). Kernel density estimation is a non-parametric way of estimating the probability density function of a random variable. Kernel density estimation is a fundamental data smoothing problem where inferences about the population are made, based on a finite data sample [11]. It includes in the plot the item difficulty parameters (similar to the Item Person Maps). The plot confirms the fact that the data is extremely skewed.

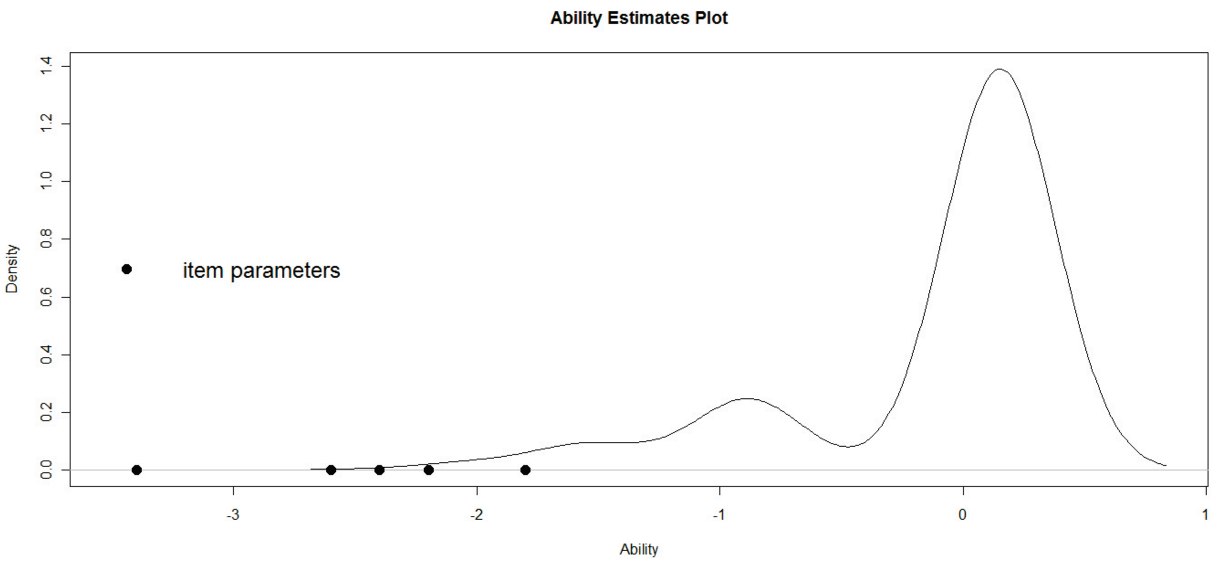

Figure 4: Ability Estimates Plot

4. Conclusion

IRT models have been extensively developed and used in educational and psychological measurement. However, use of IRT models outside of these areas is limited probably due to the fact that these methods have not been well understood by non psychometricians. With the increasing rates of item non-response in surveys, formulation of IRT models to provide in-depth understanding of this problem can help researchers overcome these obstacles. IRT modeling provides a useful method for addressing questions about patterning of behavior beyond mere frequency reports.

The implications from this study are quite clear. First, we investigated to identify the most appropriate IRT model for understanding item non-response by exploring the four IRT models for dichotomous data which include the constrained Rasch model, the unconstrained Rasch model, the two-parameter logistic model, and the three-parameter logistic model. From the likelihood ratio test, we observed by looking at the AIC and BIC values that, the unconstrained Rasch model had the smallest AIC and BIC values. Hence, the most appropriate model for the data. Furthermore, we investigated to identify the categories of survey questions that are most difficult to answer by respondents. As indicated from the results of the unconstrained model, the income related question recorded the highest difficulty parameter. In terms of probability estimates, we observe from that, the probability of responding to the income related question by the average individual as compared to the other categories of questions is the smallest. Therefore, the income related questions are the most difficult category of survey questions to answer by respondents.

Finally, we analyzed the reason behind don't know responses and missing data; whether respondents don't really know, don't care, or don't want to answer. From the selected model, the difficulty of a question explains whether or not an individual will respond to that question. We also observe from the ability estimates and the test information curve that, almost 90% of the total test information for ability levels lies in the interval (-4, 0). This means that most of the questions in the dataset were easy questions. Also, because the difficulty values are less than 0 and discrimination values are greater than 1, 'don't care' which is usually associated with guessing is not apparent in the dataset. Therefore, if an individual does not answer a survey question or give a 'don't know' answer, it is not only because of the question's difficulty but also because the individual doesn't want to answer.

References

[1] Baker, F.B. (2001). *The Basics of Item Response Theory.* USA: ERIC Clearinghouse on Assessment and Evaluation.

[2] Birnbaum, A. (1968). *Some latent trait models.* In F. M. Lord & M. R. Novick (Eds), Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

[3] Cavanaugh, J.E. (2009). *171:290 Model Selection, Lecture VI: The Bayesian Information Criterion,* Department of Biostatistics, Department of Statistics and Actuarial Science, University of Iowa

[4] De Leeuw, E.D. (2001). *Reducing Missing Data in Surveys: An Overview of Methods,* Quality and Quantity 35 (2):147-160.

[5] Lord, F. M. (1980). *Applications of item response theory to practical testing problems,* Hillsdale, NJ: Lawrence Erlbaum Associates.

[6] 'R' Development Core Team (2010). *R: A Language and Environment for Statistical Computing,* Vienna, Austria: R Foundation for Statistical Computing. http://www.R-project.org

[7] Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests,* Chicago: MESA.

[8] Ren, L. (2009). *Surveying Public Opinion in Transitional China: An Examination of Survey Response,* Pittsburgh: University of Pittsburgh. PhD Thesis

[9] Rizopoulos, D. (2006). *ltm: An R Package for Latent Variable Modeling and Item Response Theory Analyses,* Catholic University of Leuven: Journal of Statistical Software, Vol. 17, Issue 5. http://www.jstatsoft.org

[10] Rubin, D. B. (1976). *Inference and missing data,* Biometrika, *63*(3), 581–592.

[11] Scott, D. (1979). *On optimal and data-based histograms,* Biometrika 66: 605–610.

[12] World Values Survey Data (2010, October 25th) [online] –URL: http://www.worldvaluessurvey.org