Psychometric Analyses of the 2006 MCAS High School Biology Test[1,2]

Zachary R. Smith and Ronald K. Hambleton
University of Massachusetts Amherst

February 19, 2008

[1]Financial support for completing this research came from Measured Progress and the Massachusetts Department of Education. The authors are responsible for the contents of the report and for any errors or misinterpretations of the data that may remain. .
[2]**Center for Educational Assessment Research Report No. 645**. Amherst, MA: University of Massachusetts, Center for Educational Assessment.

1. Goal of the Psychometric Analyses

The primary goal of our work has been to provide readers with a number of worthwhile psychometric analyses of the 2006 MCAS High School Biology Test. These analyses provide more detail on the Biology Test than it was possible to provide in the summary report prepared by Hambleton, Zhao, Smith, Lam, and Deng (2008). These analyses include (1) an item analysis, (2) descriptive statistics on the test scores including break-outs for several subgroups of students, (3) classical reliability analyses for the test scores organized by item format, and for the total test, (4) two investigations of test dimensionality, (5) item response theory (IRT) item calibrations obtained from fitting the three-parameter logistic model to binary-scored items and the graded response model to polytomously-scored items, (6) various item and test level model fit findings, (7) test information and conditional standard errors, and (8) the identification of differentially functioning test items.

2. Brief Test Description

The 2006 MCAS Biology Assessment included 45 items, 40 of which were multiple-choice items (dichotomously scored) and five that were open-response items (polytomously scored). The maximum score for the multiple-choice items was 1 point and the maximum for the open-response items was 4 points, for a maximum raw score of 60 points.

Table 2.1 Test Information

Item Type	Number of Items	Number of Points
Multiple-Choice	40	40
Open-Response	5	20
Total	45	60

The open-response items included numbers 11, 25, 26, 32, and 39. The test was broken up into two sessions. Session 1 included items 1 – 26 and session 2 included items 27 – 45.

Table 2.2 shows the six learning standards for Biology: The Chemistry of Life, Structure and Function of Cells, Genetics, Human Anatomy and Physiology, Evolution and Biodiversity, and Ecology. The topic of Human Anatomy and Physiology was not covered on this assessment, but the split of items was fairly even among the remaining five learning standards.

Table 2.2 Descriptive Statistics of the MCAS Biology Learning Standards

Standards	# of Items	# of Points
1. The Chemistry of Life	6	9
2. Structure and Function of Cells	12	15

3. Genetics	9	12
4. Human Anatomy and Physiology	0	0
5. Evolution and Biodiversity	9	12
6. Ecology	9	12

3. Item Analyses

For all analyses, examinees were excluded if their raw score was equal to zero or left blank. This reduced the original sample of N=58,441 examinees down to the adjusted sample of N=55,673. For analyses following the item analysis, a random sample of N=5,517 was drawn from the sample of 55,673 examinees. The full sample was not needed.

We carried out an item analysis on the adjusted Biology sample (N=55,673). The results are shown in Tables 3.1 to 3.6. The *p*- and *r* values were calculated and can be seen in Table 3.1. The *p*-values (or item means) ranged from 0.24 to 0.93 for the multiple-choice items and from 0.81 to 2.00 for the open-response items. The *p*-values for the polytomous items are on a 0 to 4 point scale, hence the reason for means above 1 in many cases. The large span of *p*-values for the multiple choice items highlights the fact that items ranging from very easy to very hard were included in the assessment. This is important so that information will be available across the scale for confidently assigning students to performance categories. Figure 3.1 shows the distribution of the *p*-values for the multiple-choice items.

The item *r* values were good to excellent. The *r* values for the multiple-choice items ranged from 0.23 to 0.53. The polytomous items consistently had the highest *r* values overall, ranging from 0.67 to 0.73. Figure 3.2 shows the distribution of the *r* values.

A distractor analysis was also carried out using the computer software known as the Test Analysis Program (TAP; Brooks & Johanson, 2003). The purpose of this software is to determine the merits of each distractor for the multiple choice items on a particular assessment, as well as recording the percentage of students receiving particular scores on the polytomously scored items. For the Biology Test, we chose to look at the top and bottom 25 percent of examinees to determine how they were answering the items. The full distractor analysis can be found in Appendix A. The first number presented in each column represents the number of students making that answer choice for dichotomously scored items or the number of students receiving that score on the polytomously scored items. The second number in each column, the one in parentheses, represents the percentage of students out of the sample for that answer choice. Consider item 35, for example. The majority of students did not answer this item correctly. In fact, examinees chose two of the other options more often than the correct answer. Still, we were able to see that the highest scoring students chose the correct answer 43.3% of the time while the lowest scoring students were more drawn to the other options, only answering correctly 16.9% of the time. Clearly the discriminating power of the test item is evident.

Table 3.1 Summary of Descriptive Item Statistics

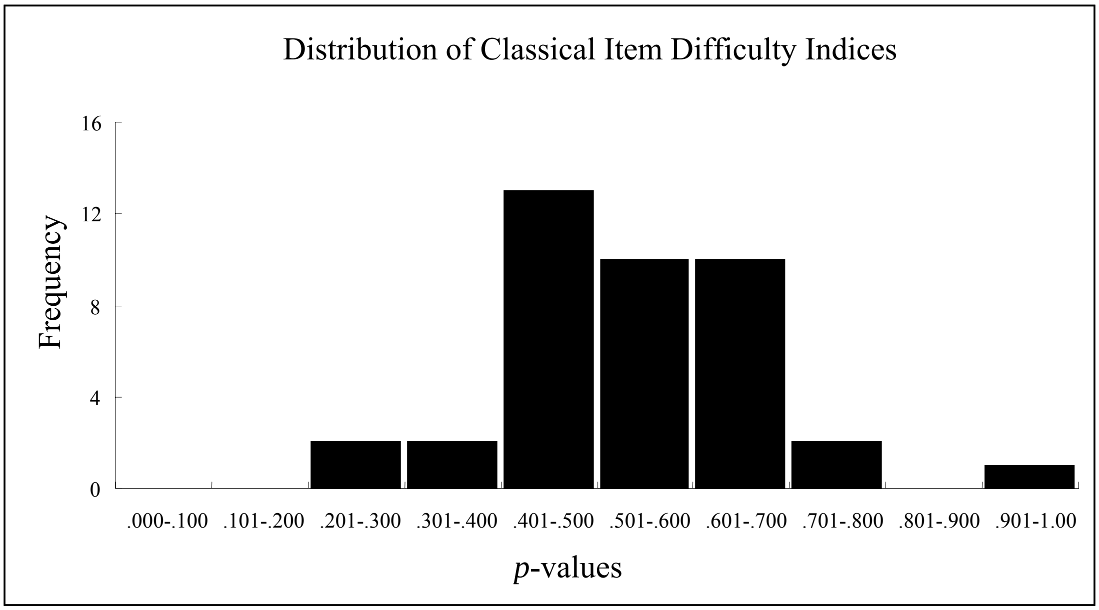
		Item Difficulty	Item Discrimination
Multiple-Choice	$\overline{p}$	0.53	0.41
	SD	0.14	0.07
Open Response	$\overline{p}$	1.61	0.69
	SD	0.48	0.03

Table 3.2 Descriptive Item Statistics

Item	*p*	*R*	Min	Max
1	0.93	0.29	0	1
2	0.65	0.35	0	1
3	0.46	0.30	0	1
4	0.24	0.40	0	1
5	0.64	0.40	0	1
6	0.66	0.42	0	1
7	0.55	0.44	0	1
8	0.54	0.38	0	1
9	0.65	0.39	0	1
10	0.52	0.28	0	1
11	2.00	0.72	0	4
12	0.49	0.39	0	1
13	0.49	0.44	0	1
14	0.42	0.41	0	1
15	0.57	0.46	0	1
16	0.40	0.32	0	1
17	0.56	0.51	0	1
18	0.61	0.51	0	1
19	0.44	0.37	0	1
20	0.44	0.31	0	1
21	0.43	0.43	0	1
22	0.50	0.47	0	1
23	0.50	0.23	0	1
24	0.51	0.42	0	1
25	1.92	0.67	0	4
26	1.75	0.73	0	4
27	0.64	0.44	0	1
28	0.70	0.43	0	1
29	0.36	0.41	0	1
30	0.79	0.48	0	1
31	0.60	0.42	0	1
32	0.81	0.67	0	4
33	0.42	0.35	0	1
34	0.42	0.43	0	1
35	0.25	0.29	0	1
36	0.71	0.53	0	1
37	0.63	0.47	0	1

38	0.44	0.39	0	1
39	1.56	0.69	0	4
40	0.68	0.52	0	1
41	0.64	0.52	0	1
42	0.47	0.45	0	1
43	0.43	0.49	0	1
44	0.38	0.38	0	1
45	0.57	0.47	0	1

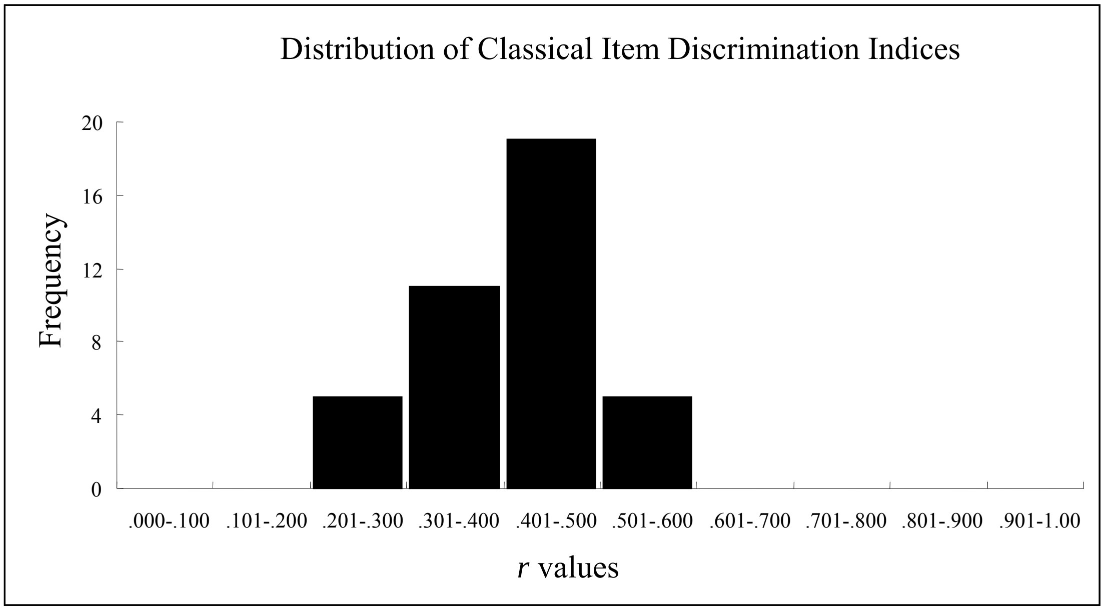
Figure 3.1 Histogram Showing the Distribution of Classical Item Difficulty Indices



Note: Only multiple-choice items were included in this histogram.

Students were excluded from the analysis if their raw score was equal to zero or left blank. The new sample included 55,673 students.

Figure 3.2 Histogram Showing the Distribution of Classical Item Discrimination Indices



Note: Only multiple-choice items were included in this histogram.

4. Basic Test Statistics and Reliability

The raw score distribution for the adjusted sample of students had a mean score of 29.3 with a standard deviation of 12.5. Figure 4.1 shows that the distribution of the raw scores is skewed positively (skewness=0.183)—not a surprising result with new tests.

Tables 4.1 and 4.2 display the descriptive statistics for gender and ethnicity, respectively. Females performed on average one test score point higher than males on the Biology Test. For ethnicity, Asians performed higher than any other demographic group with a mean score over 34. The mean score for whites was over 31, and then came Native Americans, Blacks, and Hispanics, in that order.

Figure 4.1 Histogram Showing the Distribution of the Raw Scores

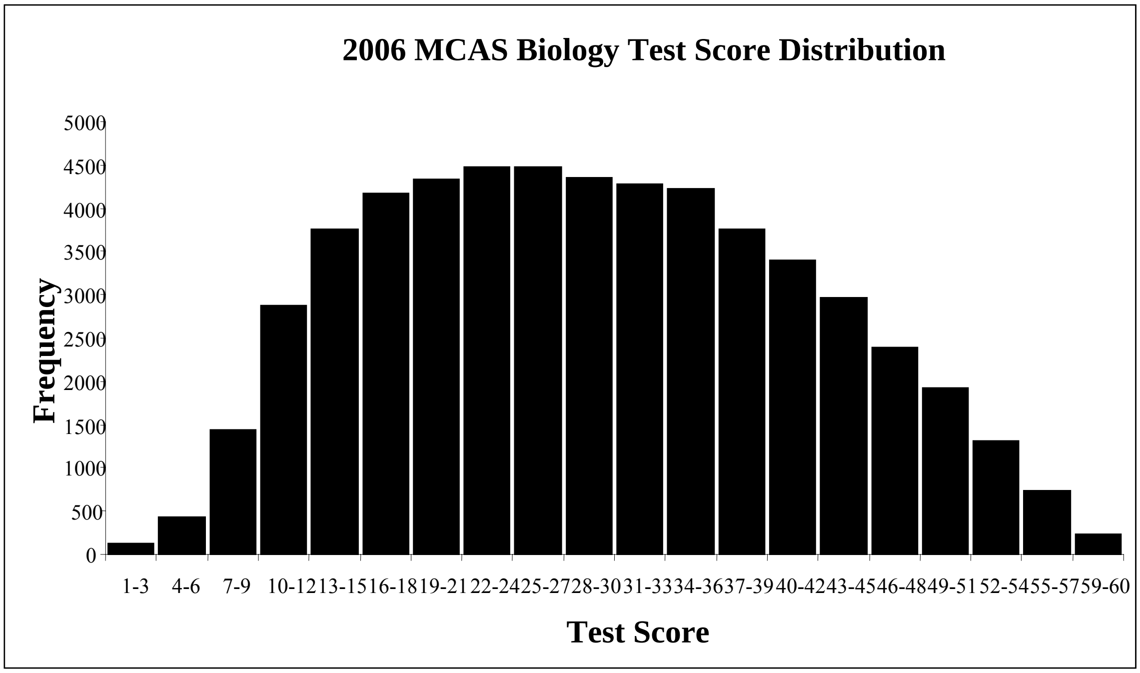


Table 4.1 Test Statistics for the Total Sample of Students

	N	Mean	SD	Min	Max	Skewness
Overall	55,673	29.33	12.46	1	60	.18

Table 4.2 Descriptive Test Statistics by Gender

Group	N	Mean	SD	Min	Max	Skewness	Percent
No Response	879	20.59	11.10	1	58	0.84	1.6
Females	27,233	29.95	11.92	1	60	0.17	48.9
Males	27,561	28.99	12.89	1	60	0.20	49.5
Total	55,673						100.0

Table 4.3 Descriptive Test Statistics by Ethnicity

Ethnicity	N	Mean	SD	Min	Max	Skewness	Percent
No Response	901	20.74	11.14	1	58	0.82	1.6
Asian	2,155	34.16	13.05	1	60	-0.03	3.9
Black	4,359	21.61	10.14	1	60	0.65	7.8
Hispanic	4,885	20.75	10.01	1	58	0.82	8.8
Native American	156	26.72	11.55	4	58	0.38	.3
White	43,217	31.02	12.11	1	60	0.07	77.6
Total	55,673						100.0

Score reliability was calculated for the assessment as a whole using Cronbach's coefficient alpha, as well as for the multiple-choice and the open-response items separately. The results are displayed in Table 4.4. There was a high overall reliability (α =.91), and the reliabilities for the different item types was also high too (MCQ: α=.88; Open-Response: α=.81), though these subtest scores are not reported to students.

Table 4.4 Reliability Analysis

Variable	α
Total	.91
Multiple-Choice	.88
Open-Response	.81

5. Test Dimensionality

An initial check of Biology Test dimensionality was run on the sample of N=5,517 by examining the eigenvalues of the matrix of inter-item correlations using the program R version 2.4.1. (We are grateful to Yue Zhao for carrying out these analyses.) Table 5.1 displays the top ten eigenvalues found from this analysis along with the variance accounted for by each of the factors. Figure 5.1 also displays this result, showing a large first factor of 13.80 and a second factor of 1.65 and a third factor of 1.44.

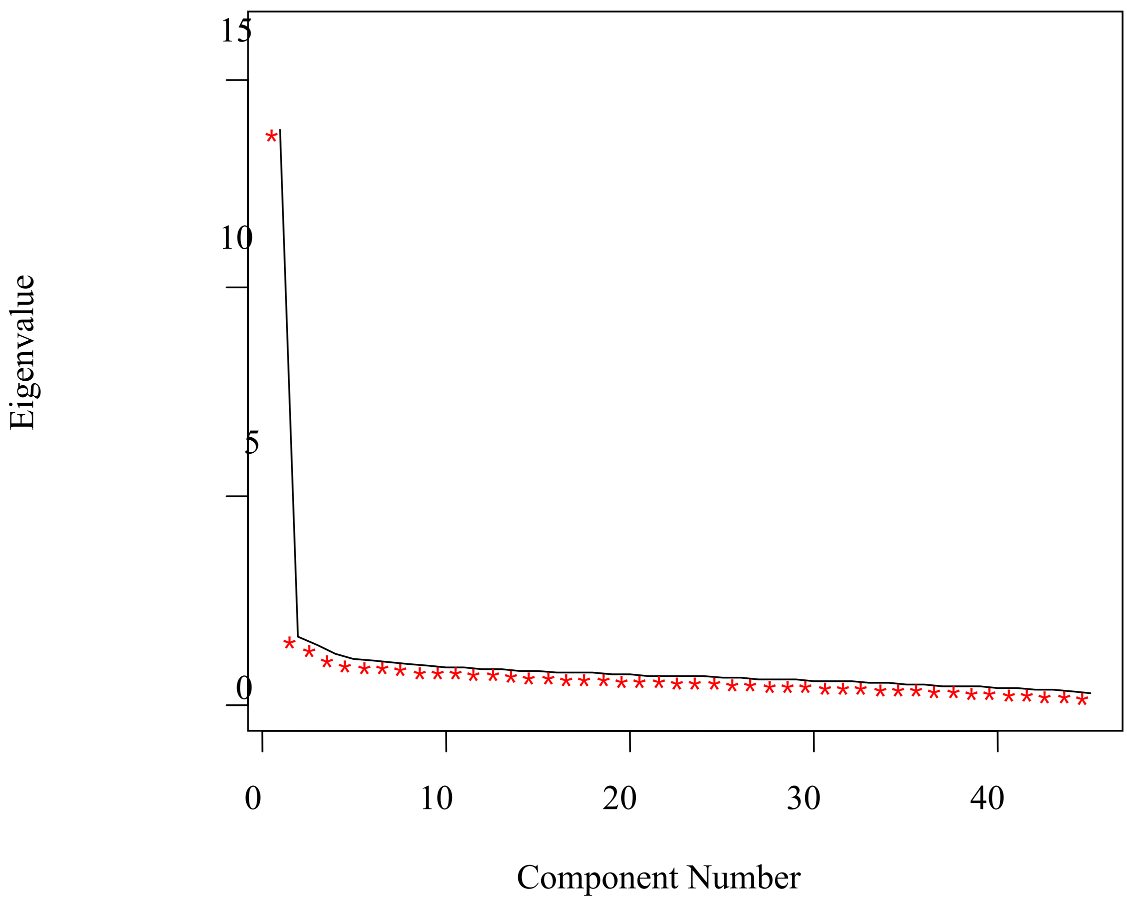
Table 5.1 Eigenvalues and Corresponding Variance

Item	Eigenvalue	Variance Accounted For
1	13.80	0.31
2	1.65	0.04
3	1.44	0.03
4	1.20	0.03
5	1.08	0.02
6	1.04	0.02

7	1.02	0.02
8	0.98	0.02
9	0.93	0.02
10	0.91	0.02

Note: Eigenvalues were calculated using a random sample of N=5,517 from the original sample of N=55,673.

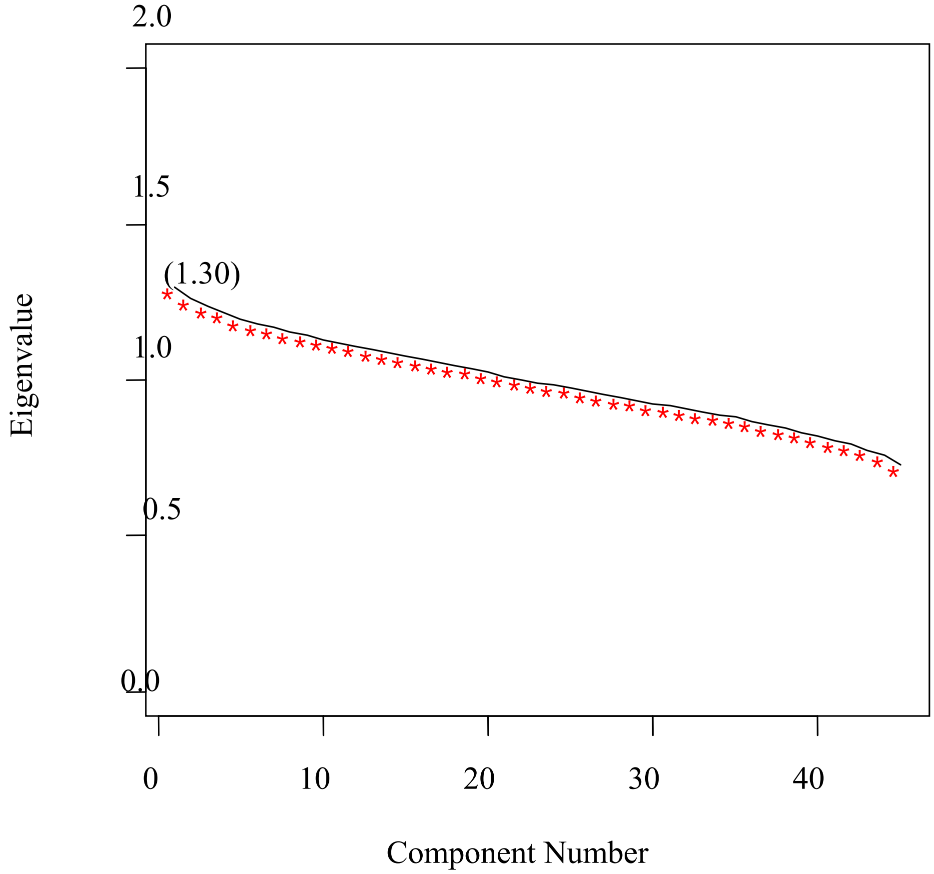
Figure 5.1 A Plot of the Eigenvalues (in order from highest to lowest) for the 2006 MCAS Biology Assessment



Note: Eigenvalues were calculated using a random sample of N=5,517 from the original sample of N=55,673.

To determine a cut point for our evaluation of the number of dimensions, we ran a parallel analysis. We created a sample of random normal deviates with R that was equal to the sample size used in the eigenvalue analysis (N=5,517). Using the probabilities from our earlier calculations, we were able to preserve the *p*-values for each item and create a more precise sample of deviates for our purposes. The percentages of students at each score level for the polytomous items were also preserved. Ten replications were run and then the eigenvalues were averaged for each item. In Figure 5.2, all of the values were arranged from highest to lowest, with the top average score being 1.30 and represented in parentheses.

Figure 5.2 Parallel Analysis of the 45 Item 2006 MCAS Biology Assessment Using Random Normal Deviates (5,517 Generated Cases). (Ten sets of random normal deviates were used to generate 10 sets of eigenvalues. The average of the largest eigenvalue was 1.30)



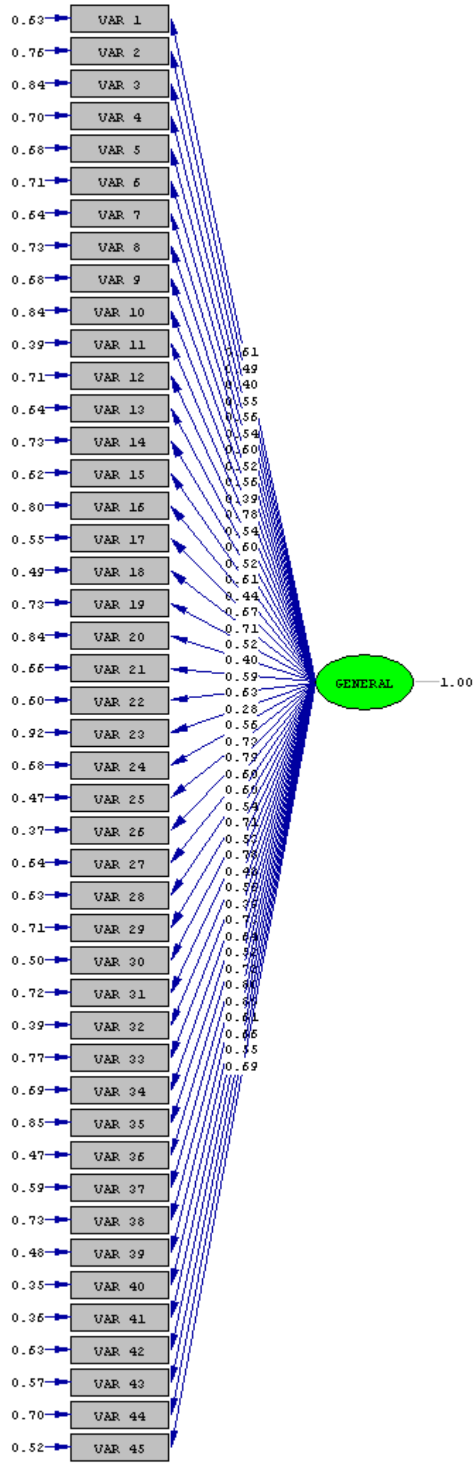
The highest value became the cutoff point to which we compared the eigenvalues. The analysis showed that the first eigenvalue of 13.80 is still prominent at more than an 8:1 ratio with the second eigenvalue. This is enough to claim unidimensionality with the first factor with an eigenvalue of 13.80. With the cutoff at 1.30, it is possible that the second and third eigenvalues are still representing very small dimensions. We determined that the third eigenvalue was too close to the cutoff and this difference can be attributed to error. This leaves a minor second factor from the eigenvalue of 1.65. But the main finding was clear: A very big first factor, and high enough to carry on with the unidimensional IRT analyses.

We also carried out a confirmatory factor analysis assuming a one-factor solution. The results are presented below, but this analysis too, reported in Table 5.2 and Figure 5.3, shows a very strong first factor with all of the loadings very high.

Table 5.2 Factor Loadings for Each Item

Item	Factor Loadings
1	0.61
2	0.49
3	0.40
4	0.55
5	0.56
6	0.54
7	0.60
8	0.52
9	0.56
10	0.39
11	0.78
12	0.54
13	0.60
14	0.52
15	0.61
16	0.44
17	0.67
18	0.71
19	0.52
20	0.40
21	0.59
22	0.63
23	0.28
24	0.56
25	0.73
26	0.79
27	0.60
28	0.60
29	0.54
30	0.71
31	0.53
32	0.78
33	0.48
34	0.56
35	0.39
36	0.73
37	0.64
38	0.52
39	0.72
40	0.80
41	0.80
42	0.61
43	0.66
44	0.55
45	0.69

Figure 5.3. Graphical Display of the One Factor Model Fit to the Data



Chi-Square=3733.57, df=945, P-value=0.00000, RMSEA=0.023

6. Item Calibration and Model Fit

Table 6.1 gives the mean and standard deviation for all of the standardized residual points combined (1950 total). The percentage of standardized residuals that do not fit is also provided in the table. This means that out of the 1950 standardized residual points, only 6% fell outside two standard deviations which is just about what would be expected under the null hypothesis that the model fit the test data.

Table 6.2 displays the parameter estimates from PARSCALE along with the corresponding standard errors (SEs). In Table 6.3, the chi-square statistics and the corresponding probabilities are provided. These probabilities are a good starting point for examining model fit, but the statistic is too stringent to be used to make a final judgment. Therefore, we must look at the fit plots.

The residuals and item characteristic curves (ICCs) can be found in Appendix B-D. Appendix B and C display the residuals and ICCs for the dichotomously scored items, respectively. After examining these plots, the majority of the multiple-choice items fit the model very well. Item 35 seemed to be the only one that was problematic at the low end of the proficiency scale. The first nine points are above zero for the residual and above the ICC line while the next eight are below. This is indicative of a problem at the low end of the scale. The lower achievers were generally scoring higher on this item than those with proficiency scores closer to zero.

The fit plots provide data for the polytomous items at each of the score categories, which can be seen in Appendix D. These items tended to fit the model more poorly than the dichotomous items. Score category 2 is problematic at the low end of the proficiency scale for item 11. For item 25, score category 2 is problematic at the low end, while score categories 3, 4, and 5 are problematic at the high end. There could also be some issues with item 26 at the high end for score categories 4 and 5. Item 32 seems to fit well for the most part, but item 39 exhibited problems at all score categories.

Figure 6.1 shows the test score distribution, which is a comparison of the predicted scores and the observed scores. The fit is excellent. The match in Figure 6.2 of the cumulative observed and predicted score distributions assuming the model to be true is just about perfect. Often Figure 6.2 is more useful to review than Figure 6.1 because of the smoothing that is reflected in the cumulative distributions.

Table 6.1 Descriptive Information for the Standardized Residuals from PARSCALE

Mean	SD	Percentage Not Fit
-0.06	1.05	0.06

Table 6.2 Item Parameter Estimates

Item	Slope (a)	SE	Location (b)	SE	Guessing (c)	SE
1	0.993	0.061	-1.923	0.140	0.290	0.078
2	0.749	0.057	0.027	0.100	0.327	0.032
3	0.625	0.059	0.946	0.084	0.237	0.026
4	1.261	0.079	1.325	0.033	0.104	0.008
5	0.941	0.062	0.161	0.063	0.312	0.023
6	0.649	0.037	-0.502	0.099	0.129	0.038
7	1.139	0.068	0.417	0.041	0.262	0.016
8	0.731	0.054	0.405	0.076	0.231	0.026
9	0.871	0.059	0.025	0.076	0.318	0.027
10	0.818	0.077	1.025	0.066	0.344	0.019
11	1.080	0.016	0.021	0.014	0.000	0.000
12	1.403	0.091	0.776	0.033	0.296	0.012
13	0.815	0.047	0.329	0.052	0.134	0.021
14	1.014	0.066	0.842	0.040	0.200	0.014
15	0.848	0.047	-0.007	0.059	0.153	0.024
16	1.345	0.101	1.164	0.036	0.267	0.011
17	1.264	0.067	0.268	0.035	0.234	0.016
18	1.131	0.056	-0.104	0.043	0.194	0.020
19	0.924	0.065	0.842	0.047	0.229	0.016
20	1.092	0.095	1.238	0.048	0.315	0.013
21	1.342	0.080	0.784	0.031	0.224	0.012
22	0.910	0.051	0.335	0.046	0.152	0.019
23	0.677	0.085	1.391	0.091	0.381	0.021
24	0.952	0.064	0.577	0.050	0.265	0.018
25	0.920	0.013	0.130	0.016	0.000	0.000
26	1.097	0.016	0.312	0.014	0.000	0.000
27	0.767	0.040	-0.371	0.073	0.126	0.031
28	0.766	0.042	-0.613	0.090	0.162	0.038
29	1.236	0.077	1.002	0.032	0.180	0.011
30	1.202	0.056	-0.872	0.055	0.175	0.031
31	0.788	0.053	0.130	0.076	0.248	0.028
32	1.126	0.020	1.382	0.016	0.000	0.000
33	0.549	0.034	0.633	0.064	0.060	0.020
34	0.821	0.051	0.718	0.046	0.132	0.017
35	1.338	0.108	1.600	0.042	0.160	0.008
36	1.136	0.050	-0.600	0.046	0.131	0.025
37	0.816	0.040	-0.339	0.061	0.105	0.026
38	0.968	0.065	0.833	0.043	0.213	0.015
39	0.907	0.015	0.412	0.017	0.000	0.000
40	1.113	0.049	-0.488	0.045	0.130	0.023
41	1.161	0.054	-0.248	0.041	0.160	0.021
42	0.863	0.050	0.466	0.046	0.134	0.018
43	1.145	0.059	0.550	0.031	0.135	0.013
44	0.910	0.061	0.926	0.042	0.164	0.015
45	0.841	0.043	-0.029	0.054	0.119	0.023

Table 6.3 Item Chi-Square Model Fit Statistics and the Corresponding Probabilities

Item	Chi-Square	*df*	Probability
1	23.621	17	0.130
2	22.293	28	0.768
3	33.989	30	0.281
4	54.060	30	0.005
5	24.378	27	0.610
6	36.668	28	0.126
7	36.546	27	0.104
8	25.959	30	0.678
9	17.562	27	0.916
10	32.468	30	0.346
11	181.999	99	0.000
12	38.786	27	0.066
13	28.392	29	0.497
14	25.847	30	0.683
15	21.344	27	0.770
16	34.709	30	0.253
17	20.164	26	0.784
18	27.021	25	0.355
19	43.686	30	0.051
20	66.127	30	0.000
21	20.349	28	0.852
22	27.784	29	0.530
23	45.079	30	0.038
24	24.796	29	0.689
25	236.530	103	0.000
26	166.049	99	0.000
27	20.286	27	0.819
28	22.928	26	0.637
29	48.751	30	0.017
30	33.961	21	0.037
31	29.129	29	0.458
32	136.973	92	0.002
33	57.207	30	0.002
34	33.380	30	0.306
35	38.153	30	0.146
36	24.852	23	0.358
37	27.561	27	0.434
38	36.321	30	0.197
39	397.447	106	0.000
40	29.667	24	0.196
41	31.589	25	0.170
42	36.525	30	0.191
43	31.171	28	0.309
44	23.421	30	0.798
45	17.291	27	0.924
Total	2392.817	610	0.000

Figure 6.1 Observed and Predicted Test Score Distributions

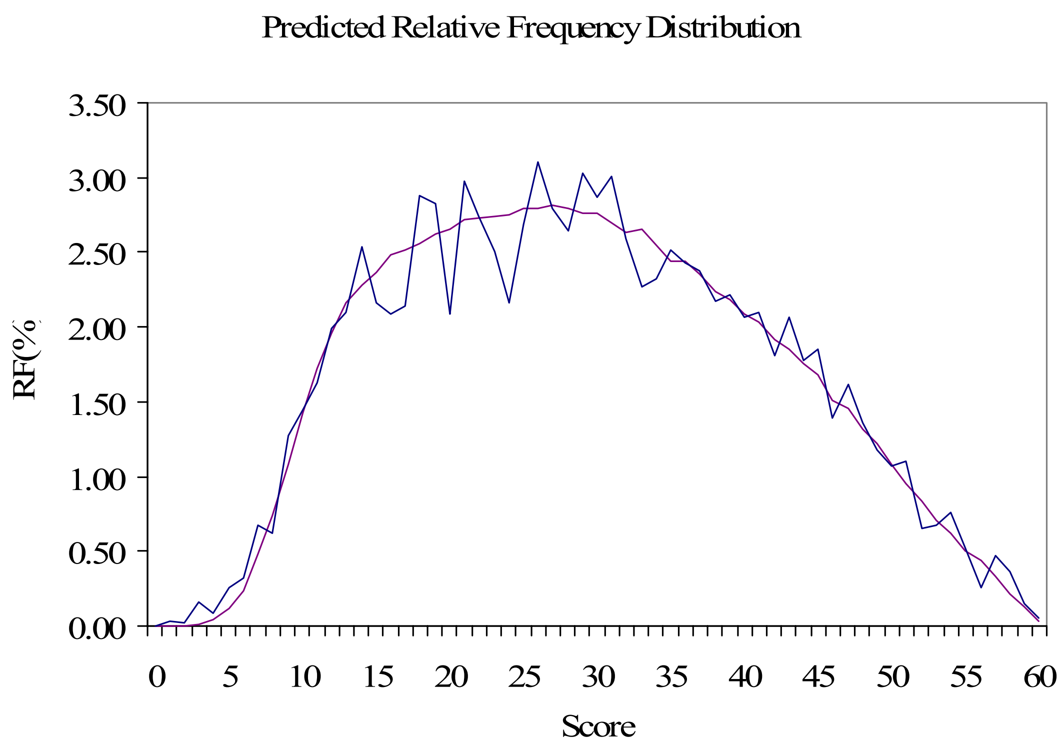
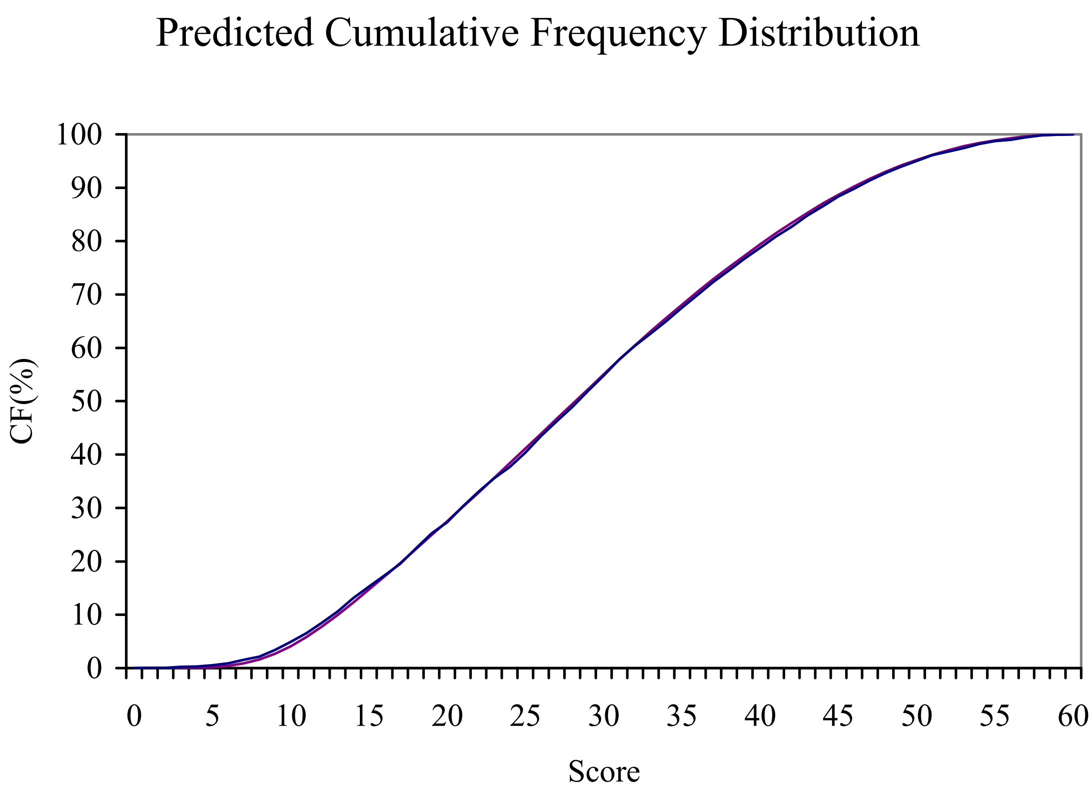


Figure 6.2 Observed and Predicted Cumulative Frequency Distributions



7. Test Information and Conditional Standard Errors

The test characteristic curve, test information function, and standard error of measurement are displayed in the following three figures, 7.1, 7.2, and 7.3. They were produced from the parameter file created in PARSCALE and plotted in a software package called IRT Painter.

Figure 7.1 Test Characteristic Curve for the Biology Test

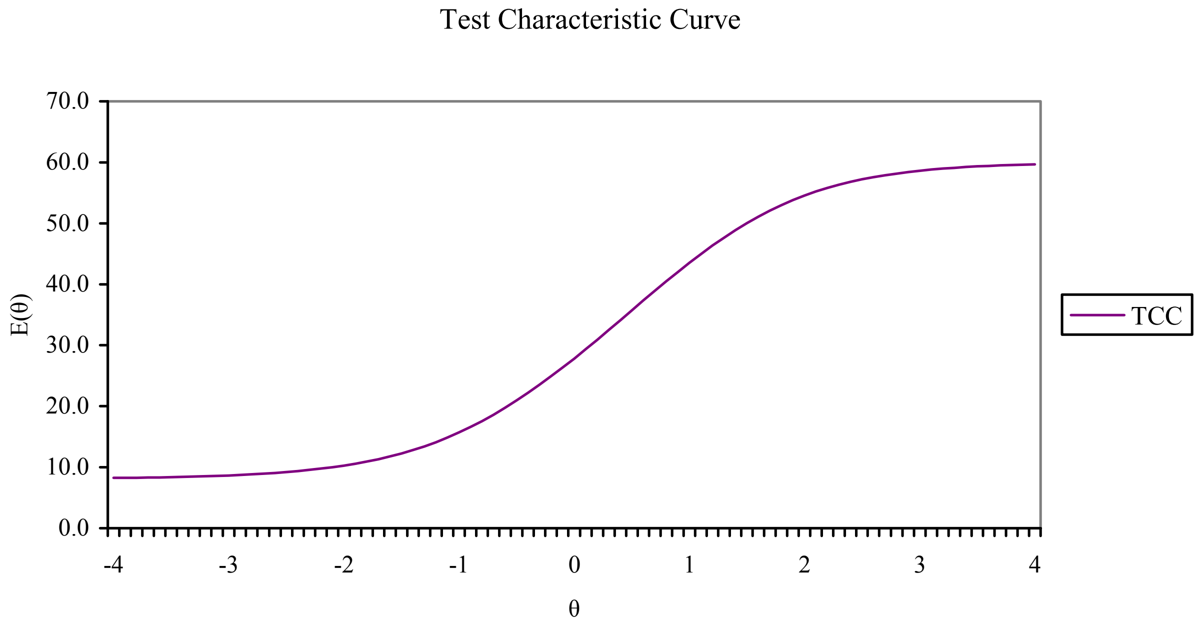


Figure 7.2 Test Information Function for the Biology Test

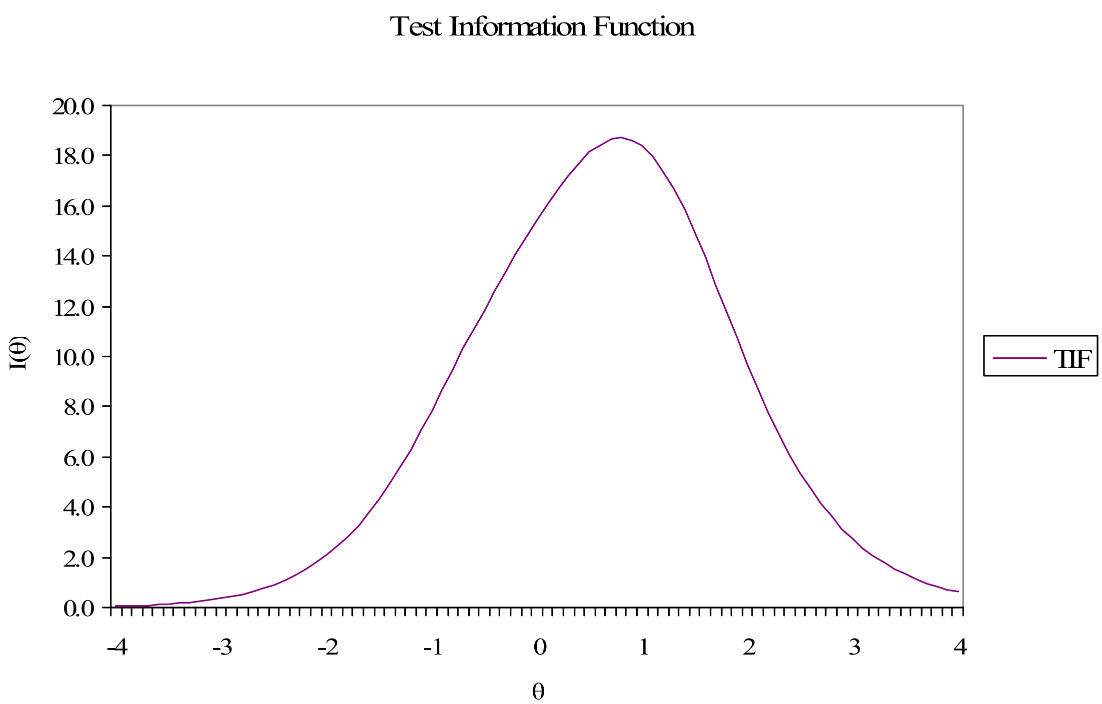
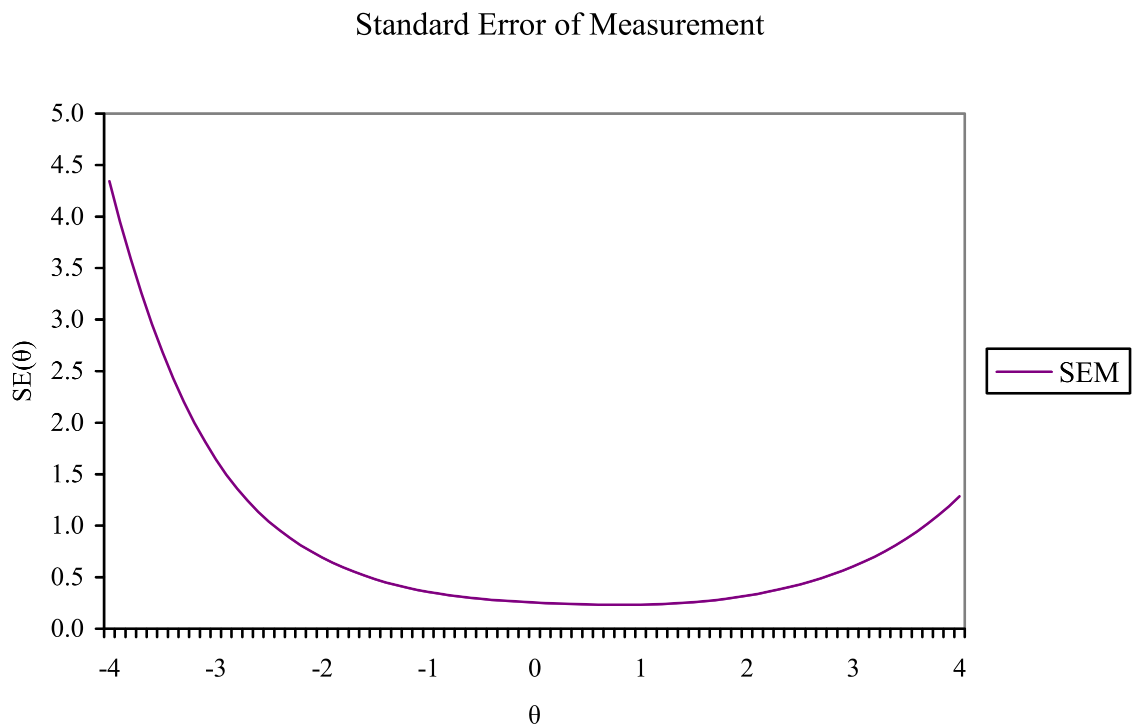


Figure 7.3 Conditional Standard Errors for the Biology Test



8. Identification of Differentially Functioning Test Items

The test was examined for differentially functioning items using the computer program STDIF (Zenisky & Hambleton, 2007). The program calculates both the SDIF and UDIF statistics. For the purposes of this study, the UDIF statistic was used. This analysis is done in two stages. First, the program is run including all of the items when calculating the statistics. Then, for the second stage, the items that showed DIF from the first stage are excluded from the matching (conditioning) variable, and then the analysis is repeated for all of the test items.

Four group comparisons were completed using the sample of N=55,673, which included Male/Female, White/Asian, White/Black, and White/Hispanic. The comparison results are reported in Table 8.1.

For the Male/Female comparison, males were the reference group. This analysis showed three DIF items (4, 29, 38) at Stage 1, which dropped to zero items at Stage 2 using the .10 criterion for DIF detection. Four items (4, 13, 29, 38) had statistics between .075 and .10 at Stage 2. Figures 8.1 – 8.4 displays the comparison for the four items under the gender category. Males outperformed females consistently for all of the flagged items.

For the ethnicity comparisons, the White sample formed the reference group. When the White sample was compared to the Asian sample, there were two DIF items at

Stage 1 and none using the .10 criterion at Stage 2. The White sample had much more stable curves, while the Asian sample jumped around between score points. This is attributed to the considerably smaller sample size of the Asian group. There were many fewer examinees at each test score point, causing the erratic line. In Figure 8.5, we have shown the most problematic of the items, item 8, in the White-Asian comparisons--the White sample outperformed the Asian sample for item 8. But this item did not reach the .10 criterion for being labeled as a DIF item. The White/Black and White/Hispanic comparison showed no DIF at Stage 2 either.

Figure 8.6 displays the gender DIF indices for all items on the assessment. The top of the figure shows which items tend to favor males and the bottom of the graph shows the items that tend to favor females. Most indices hover around zero, but it is noticeable that there are a couple items that are favoring males and these were the four items we displayed in the figures.

Table 8.1 Flagged Items Using the UDIF Statistic

	Stage 1		Stage 2	
Comparison	# Flagged	Item #'s	# Flagged	Item #'s
M/F	3	4, 29, 38	0	N/A
W/A	2	8, 45	0	N/A
W/B	0	N/A	0	N/A
W/H	1	10	0	N/A

Gender

Figure 8.1 Conditional *p*-Value Plot for Item 4

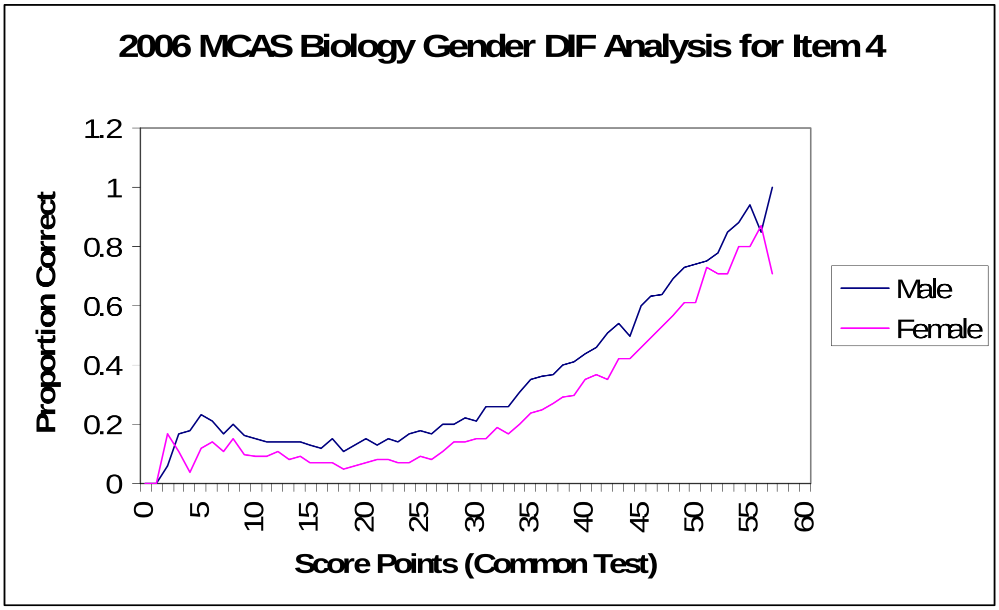


Figure 8.2 Conditional *p*-Value Plot for Item 13

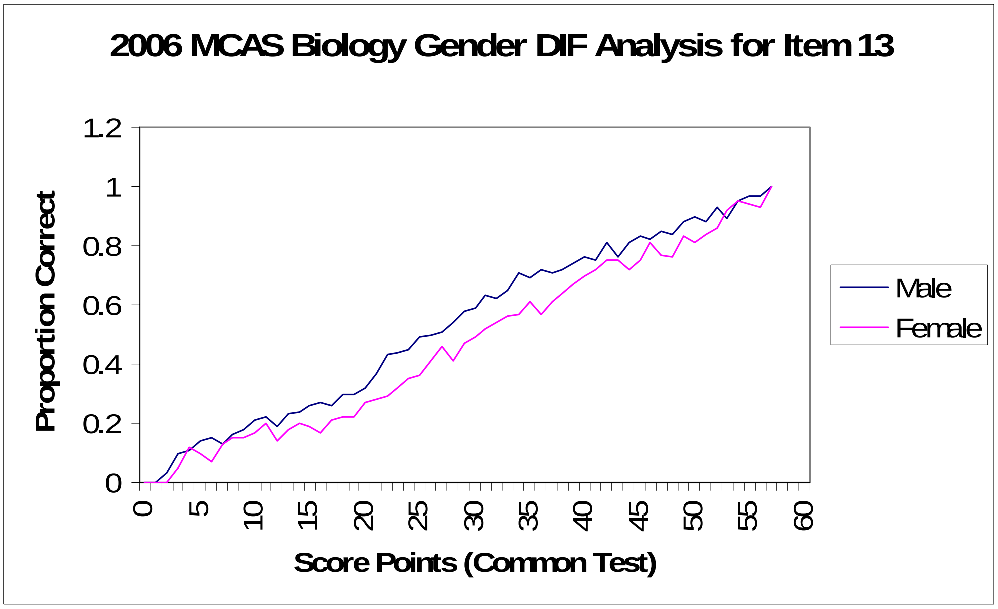


Figure 8.3 Conditional *p*-Value Plot for Item 29

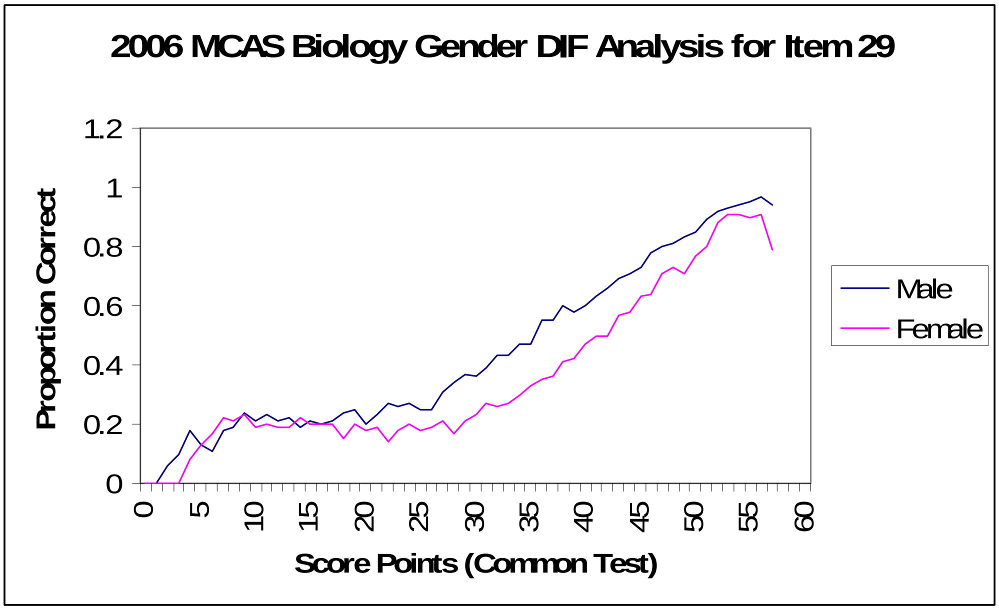
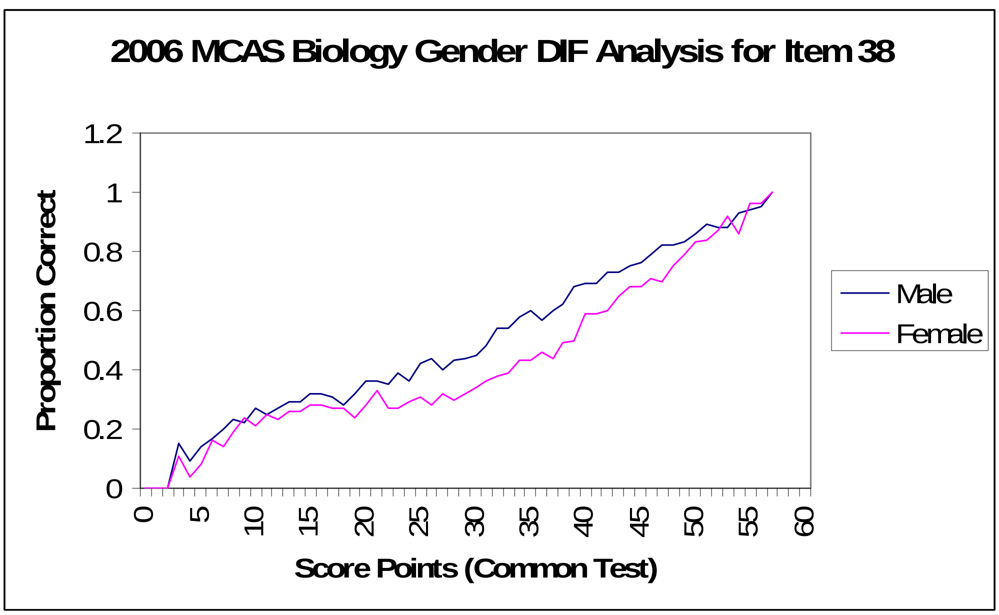


Figure 8.4 Conditional *p*-Value Plot for Item 38



Ethnicity

Figure 8.5 Conditional *p*-Value Plot for Item 8

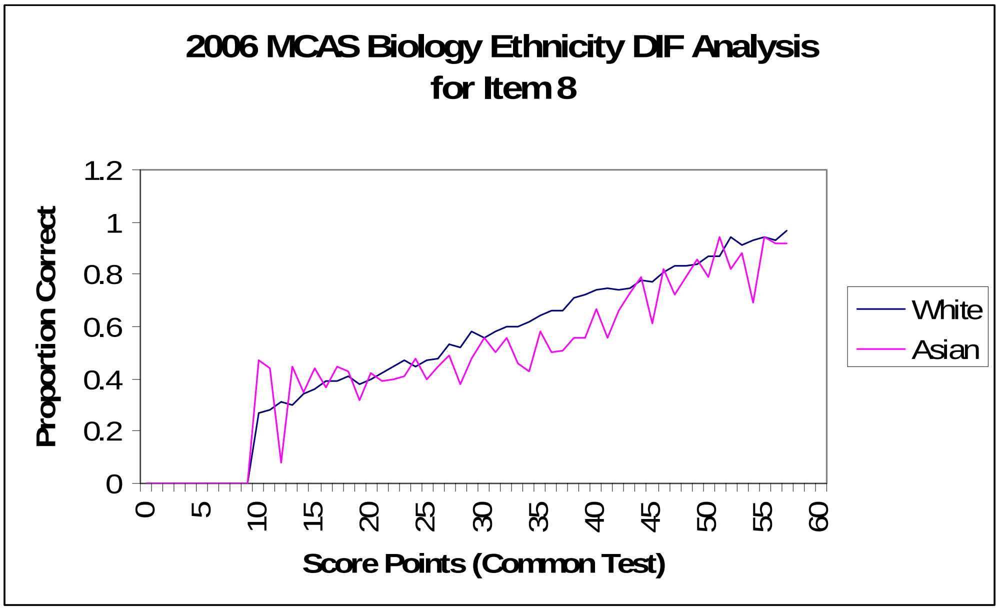
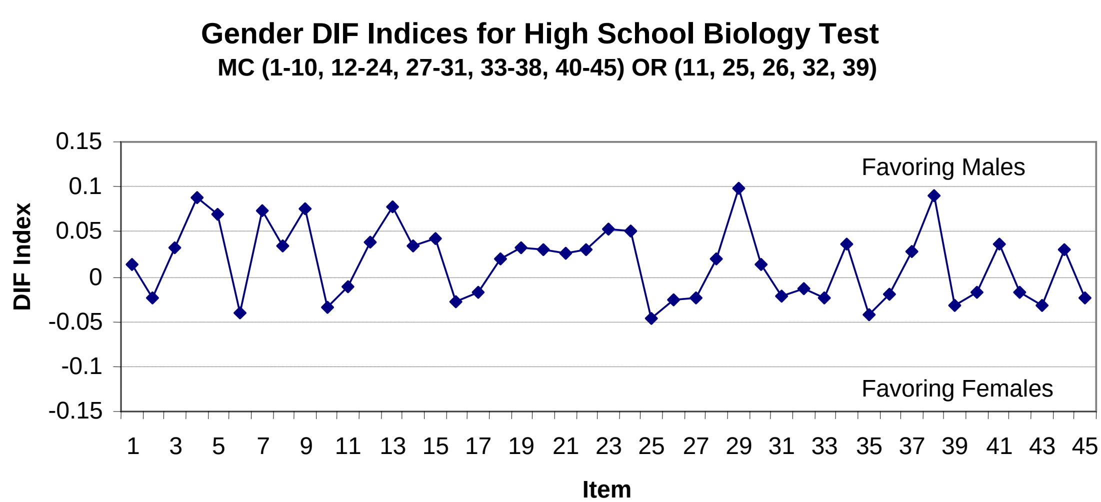


Figure 8.6 Summary of the Gender DIF Indices



9. Conclusions

The goal of this report was to give readers several meaningful psychometric analyses of the 2006 MCAS High School Biology Test. Statistics were provided throughout the eight sections of the report and these statistics provided the basis for the summary provided in Hambleton, Zhao, Smith, Lam, and Deng (2008).

The item analysis indicated good to excellent *p*- and *r* values for the purpose of the test, and showed a wide range in item difficulty, excellent for enhancing the score precision along the proficiency score scale. Reliability was obtained using Cronbach's coefficient alpha and indicated that score reliability was high, as well as for each of the item types (dichotomously scored and polytomously scored). The Biology Test was strongly unidimensional, with a large first factor shown from the eigenvalue calculations, with only a slight possibility of a second or third factor. The confirmatory factor analysis work also showed a strong first factor. These findings strongly support the unidimensionality assumption underlying the IRT model analyses. The model fit was excellent, as shown by the residuals and ICCs in Appendices B, C, and D. The polytomous items fit more poorly than the dichotomous items. The reason was not clear and should be further investigated. Differential item functioning (DIF) analyses revealed no DIF items across four analyses of the 45 item test using the .10 criterion. This analysis showed too that a small number of test items were approaching DIF using a less stringent criterion of .075 using the UDIF statistic (Zenisky & Hambleton, 2007). Of minor concern was the presence of four items that showed a tendency for males to outperform females, when comparisons were made for groups matched on Biology proficiency (using adjusted test scores). These four items might be followed up to identify possible causes, but clearly the level of DIF was very small.

The psychometric analyses that we carried out indicated that the technical aspects of the Biology Test are high. Fit of the polytomous items and the small level of DIF in the male-female comparisons could be further studied. On this latter point, something might be learned that could be passed on to the Biology Test development committee and might be helpful when constructing future Biology Tests.

10. References

Brooks, G. P., & Johanson, G. A. (2003). Test analysis program. *Applied Psychological Measurement, 27*, 305-306.

Hambleton, R. K., Zhao, Y., Smith, Z. R., Lam, W., & Deng, N. (2008). *Psychometric analyses of the 2006 MCAS high school science tests* (Center for Educational Assessment Research Report No. 649). Amherst, MA: University of Massachusetts, Center for Educational Assessment.

Zenisky, A. L., & Hambleton, R. K. (2007). *Differential item functioning analyses with STDIF: User's guide* (Unpublished report). Amherst, MA: University of Massachusetts, Center for Educational Assessment.

Appendix A

Table A.1 Distractor Analysis: Number of examinees (Percentage of examinees)

Item	MC Key	Group	A/1	B/2	C/3	D/4	/0
			Answer Choice/Open-Response Score				
1	B	TOTAL	143 (0.026)	5149*(0.933)	150 (0.027)	58 (0.011)	
		High	3 (0.002)	1489 (0.996)	3 (0.002)	0 (0.000)	
		Low	114 (0.076)	1199 (0.801)	116 (0.077)	51 (0.034)	
		Diff	-111(-0.074)	290 (0.195)	-113(-0.075)	-51(-0.034)	
2	C	TOTAL	422 (0.076)	1324 (0.240)	3638*(0.659)	114 (0.021)	
		High	8 (0.005)	171 (0.114)	1311 (0.877)	5 (0.003)	
		Low	276 (0.184)	487 (0.325)	643 (0.430)	74 (0.049)	
		Diff	-268(-0.179)	-316(-0.211)	668 (0.447)	-69(-0.046)	
3	C	TOTAL	833 (0.151)	348 (0.063)	2587*(0.469)	1724 (0.312)	
		High	39 (0.026)	15 (0.010)	1006 (0.673)	435 (0.291)	
		Low	411 (0.275)	211 (0.141)	427 (0.285)	424 (0.283)	
		Diff	-372(-0.248)	-196(-0.131)	579 (0.388)	11 (0.008)	
4	A	TOTAL	1302*(0.236)	1486 (0.269)	987 (0.179)	1715 (0.311)	
		High	758 (0.507)	337 (0.225)	152 (0.102)	247 (0.165)	
		Low	163 (0.109)	399 (0.267)	379 (0.253)	538 (0.359)	
		Diff	595 (0.398)	-62(-0.041)	-227(-0.152)	-291(-0.194)	
5	B	TOTAL	551 (0.100)	3457*(0.627)	202 (0.037)	1286 (0.233)	
		High	24 (0.016)	1325 (0.886)	7 (0.005)	139 (0.093)	
		Low	303 (0.202)	547 (0.365)	150 (0.100)	479 (0.320)	
		Diff	-279(-0.186)	778 (0.521)	-143(-0.096)	-340(-0.227)	
6	C	TOTAL	820 (0.149)	408 (0.074)	3636*(0.659)	632 (0.115)	
		High	98 (0.066)	23 (0.015)	1308 (0.875)	65 (0.043)	
		Low	334 (0.223)	223 (0.149)	592 (0.395)	332 (0.222)	
		Diff	-236(-0.158)	-200(-0.134)	716 (0.479)	-267(-0.178)	
7	A	TOTAL	2986*(0.541)	1007 (0.183)	506 (0.092)	992 (0.180)	
		High	1264 (0.845)	86 (0.058)	60 (0.040)	85 (0.057)	
		Low	410 (0.274)	437 (0.292)	237 (0.158)	390 (0.261)	
		Diff	854 (0.572)	-351(-0.234)	-177(-0.118)	-305(-0.204)	
8	D	TOTAL	500 (0.091)	558 (0.101)	1452 (0.263)	2990*(0.542)	
		High	46 (0.031)	38 (0.025)	229 (0.153)	1182 (0.791)	
		Low	258 (0.172)	296 (0.198)	484 (0.323)	442 (0.295)	
		Diff	-212(-0.142)	-258(-0.172)	-255(-0.170)	740 (0.495)	

9	C	TOTAL	328 (0.059)	307 (0.056)	3617*(0.656)	1241 (0.225)	
		High	34 (0.023)	6 (0.004)	1337 (0.894)	118 (0.079)	
		Low	188 (0.126)	229 (0.153)	598 (0.399)	462 (0.309)	
		Diff	-154(-0.103)	-223(-0.149)	739 (0.495)	-344(-0.230)	
10	C	TOTAL	392 (0.071)	1760 (0.319)	2826*(0.512)	506 (0.092)	
		High	28 (0.019)	351 (0.235)	1083 (0.724)	31 (0.021)	
		Low	214 (0.143)	482 (0.322)	510 (0.341)	264 (0.176)	
		Diff	-186(-0.124)	-131(-0.087)	573 (0.384)	-233(-0.156)	
11		TOTAL	702 (0.127)	1455 (0.264)	1164 (0.211)	986*(0.179)	1210 (0.219)
		High	58 (0.039)	272 (0.182)	440 (0.294)	696 (0.466)	29 (0.019)
		Low	239 (0.160)	295 (0.197)	126 (0.084)	20 (0.013)	817 (0.546)
		Diff	-181(-0.121)	-23(-0.015)	314 (0.210)	676 (0.452)	-788(-0.526)
12	A	TOTAL	2661*(0.482)	629 (0.114)	1167 (0.212)	1034 (0.187)	
		High	1162 (0.777)	74 (0.049)	123 (0.082)	136 (0.091)	
		Low	410 (0.274)	261 (0.174)	484 (0.323)	318 (0.212)	
		Diff	752 (0.503)	-187(-0.125)	-361(-0.241)	-182(-0.121)	
13	A	TOTAL	2731*(0.495)	543 (0.098)	696 (0.126)	1515 (0.275)	
		High	1196 (0.800)	22 (0.015)	67 (0.045)	210 (0.140)	
		Low	314 (0.210)	321 (0.214)	317 (0.212)	515 (0.344)	
		Diff	882 (0.590)	-299(-0.200)	-250(-0.167)	-305(-0.204)	
14	B	TOTAL	1025 (0.186)	2311*(0.419)	1651 (0.299)	501 (0.091)	
		High	102 (0.068)	1062 (0.710)	266 (0.178)	63 (0.042)	
		Low	409 (0.273)	315 (0.210)	519 (0.347)	228 (0.152)	
		Diff	-307(-0.205)	747 (0.500)	-253(-0.169)	-165(-0.110)	
15	C	TOTAL	635 (0.115)	792 (0.144)	3192*(0.579)	868 (0.157)	
		High	82 (0.055)	39 (0.026)	1291 (0.864)	83 (0.056)	
		Low	282 (0.188)	380 (0.254)	410 (0.274)	396 (0.265)	
		Diff	-200(-0.134)	-341(-0.228)	881 (0.590)	-313(-0.209)	
16	C	TOTAL	1123 (0.204)	1610 (0.292)	2173*(0.394)	569 (0.103)	
		High	119 (0.080)	307 (0.205)	982 (0.657)	85 (0.057)	
		Low	394 (0.263)	477 (0.319)	376 (0.251)	216 (0.144)	
		Diff	-275(-0.184)	-170(-0.113)	606 (0.406)	-131(-0.087)	
17	A	TOTAL	3063*(0.555)	889 (0.161)	915 (0.166)	605 (0.110)	
		High	1318 (0.882)	43 (0.029)	93 (0.062)	40 (0.027)	
		Low	384 (0.257)	441 (0.295)	366 (0.244)	270 (0.180)	
		Diff	934 (0.625)	-398(-0.266)	-273(-0.182)	-230(-0.154)	
18	D	TOTAL	548 (0.099)	845 (0.153)	643 (0.117)	3443*(0.624)	
		High	27 (0.018)	66 (0.044)	21 (0.014)	1380 (0.923)	
		Low	314 (0.210)	382 (0.255)	347 (0.232)	419 (0.280)	
		Diff	-287(-0.192)	-316(-0.211)	-326(-0.218)	961 (0.643)	

19	D	TOTAL	1846 (0.335)	782 (0.142)	392 (0.071)	2464*(0.447)	
		High	319 (0.213)	70 (0.047)	32 (0.021)	1073 (0.718)	
		Low	514 (0.343)	354 (0.236)	236 (0.158)	363 (0.242)	
		Diff	-195(-0.130)	-284(-0.190)	-204(-0.136)	710 (0.475)	
20	A	TOTAL	2418*(0.438)	1050 (0.190)	631 (0.114)	1369 (0.248)	
		High	960 (0.642)	243 (0.163)	67 (0.045)	224 (0.150)	
		Low	409 (0.273)	365 (0.244)	297 (0.198)	385 (0.257)	
		Diff	551 (0.369)	-122(-0.081)	-230(-0.154)	-161(-0.107)	
21	B	TOTAL	1567 (0.284)	2369*(0.429)	796 (0.144)	723 (0.131)	
		High	214 (0.143)	1122 (0.751)	69 (0.046)	85 (0.057)	
		Low	482 (0.322)	327 (0.218)	351 (0.234)	293 (0.196)	
		Diff	-268(-0.179)	795 (0.532)	-282(-0.188)	-208(-0.139)	
22	D	TOTAL	706 (0.128)	977 (0.177)	1002 (0.182)	2758*(0.500)	
		High	50 (0.033)	51 (0.034)	189 (0.126)	1203 (0.805)	
		Low	326 (0.218)	454 (0.303)	348 (0.232)	312 (0.208)	
		Diff	-276(-0.184)	-403(-0.269)	-159(-0.106)	891 (0.596)	
23	C	TOTAL	718 (0.130)	1262 (0.229)	2836*(0.514)	634 (0.115)	
		High	141 (0.094)	301 (0.201)	988 (0.661)	64 (0.043)	
		Low	258 (0.172)	351 (0.234)	551 (0.368)	283 (0.189)	
		Diff	-117(-0.078)	-50(-0.033)	437 (0.293)	-219(-0.146)	
24	A	TOTAL	2861*(0.519)	828 (0.150)	1240 (0.225)	504 (0.091)	
		High	1187 (0.794)	88 (0.059)	169 (0.113)	47 (0.031)	
		Low	417 (0.279)	364 (0.243)	418 (0.279)	241 (0.161)	
		Diff	770 (0.515)	-276(-0.184)	-249(-0.166)	-194(-0.130)	
25		TOTAL	1156 (0.210)	1562 (0.283)	1348 (0.244)	564*(0.102)	887 (0.161)
		High	128 (0.086)	391 (0.262)	575 (0.385)	375 (0.251)	26 (0.017)
		Low	457 (0.305)	293 (0.196)	109 (0.073)	10 (0.007)	628 (0.420)
		Diff	-329(-0.220)	98 (0.066)	466#(0.312)	365 (0.244)	-602(-0.402)
26		TOTAL	947 (0.172)	1045 (0.189)	1661 (0.301)	426*(0.077)	1438 (0.261)
		High	81 (0.054)	240 (0.161)	792 (0.530)	323 (0.216)	59 (0.039)
		Low	333 (0.222)	190 (0.127)	99 (0.066)	8 (0.005)	867 (0.579)
		Diff	-252(-0.168)	50 (0.034)	693#(0.464)	315 (0.211)	-808(-0.540)
27	B	TOTAL	500 (0.091)	3541*(0.642)	401 (0.073)	970 (0.176)	
		High	34 (0.023)	1340 (0.896)	20 (0.013)	101 (0.068)	
		Low	267 (0.178)	513 (0.343)	236 (0.158)	383 (0.256)	
		Diff	-233(-0.156)	827 (0.554)	-216(-0.144)	-282(-0.188)	
28	C	TOTAL	590 (0.107)	570 (0.103)	3882*(0.704)	358 (0.065)	
		High	60 (0.040)	37 (0.025)	1373 (0.918)	23 (0.015)	
		Low	230 (0.154)	338 (0.226)	621 (0.415)	204 (0.136)	
		Diff	-170(-0.114)	-301(-0.201)	752 (0.504)	-181(-0.121)	

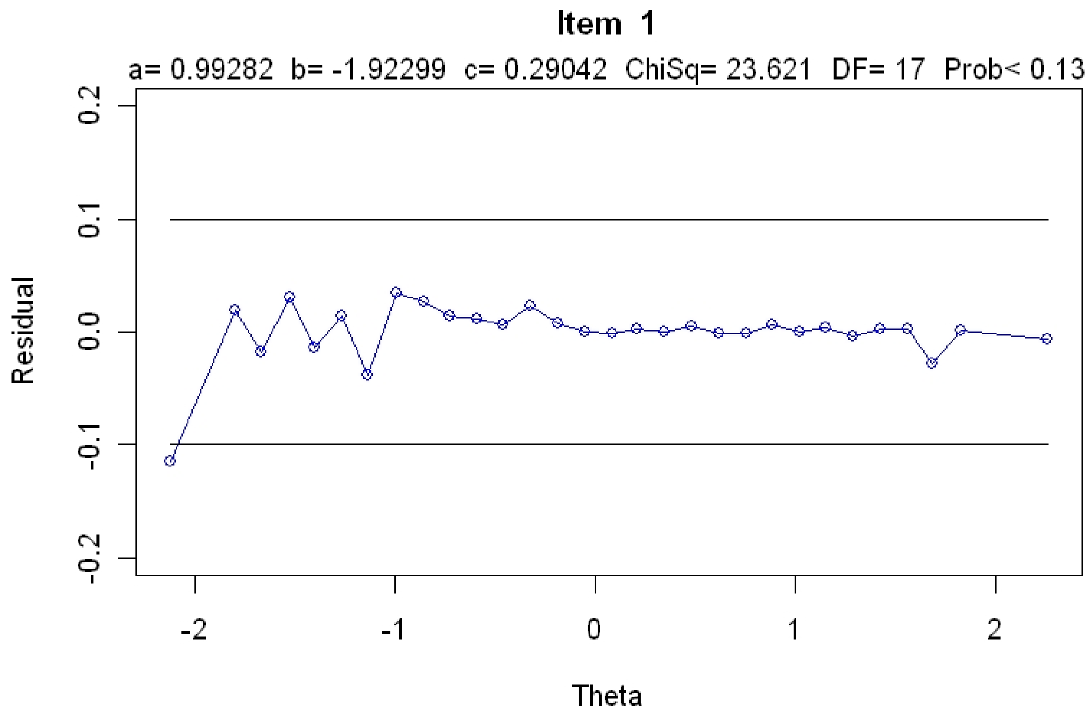
29	B	TOTAL	748 (0.136)	1974*(0.358)	1092 (0.198)	1588 (0.288)	
		High	79 (0.053)	993 (0.664)	157 (0.105)	262 (0.175)	
		Low	327 (0.218)	262 (0.175)	394 (0.263)	414 (0.277)	
		Diff	-248(-0.166)	731 (0.489)	-237(-0.158)	-152(-0.101)	
30	B	TOTAL	342 (0.062)	4366*(0.791)	368 (0.067)	330 (0.060)	
		High	13 (0.009)	1460 (0.977)	13 (0.009)	9 (0.006)	
		Low	227 (0.152)	696 (0.465)	260 (0.174)	211 (0.141)	
		Diff	-214(-0.143)	764 (0.512)	-247(-0.165)	-202(-0.135)	
31	C	TOTAL	1026 (0.186)	470 (0.085)	3309*(0.600)	597 (0.108)	
		High	135 (0.090)	42 (0.028)	1275 (0.853)	42 (0.028)	
		Low	337 (0.225)	279 (0.186)	525 (0.351)	251 (0.168)	
		Diff	-202(-0.135)	-237(-0.158)	750 (0.502)	-209(-0.140)	
32		TOTAL	869 (0.158)	946 (0.171)	421 (0.076)	115*(0.021)	3166 (0.574)
		High	252 (0.169)	467 (0.312)	334 (0.223)	112 (0.075)	330 (0.221)
		Low	103 (0.069)	54 (0.036)	3 (0.002)	0 (0.000)	1337 (0.893)
		Diff	149#(0.100)	413#(0.276)	331#(0.221)	112 (0.075)	-1007(-0.672)
33	D	TOTAL	814 (0.148)	722 (0.131)	1547 (0.280)	2282*(0.414)	
		High	118 (0.079)	70 (0.047)	344 (0.230)	960 (0.642)	
		Low	325 (0.217)	364 (0.243)	414 (0.277)	262 (0.175)	
		Diff	-207(-0.138)	-294(-0.196)	-70(-0.046)	698 (0.467)	
34	A	TOTAL	2270*(0.411)	1646 (0.298)	617 (0.112)	827 (0.150)	
		High	1052 (0.704)	329 (0.220)	53 (0.035)	59 (0.039)	
		Low	249 (0.166)	440 (0.294)	315 (0.210)	358 (0.239)	
		Diff	803 (0.537)	-111(-0.074)	-262(-0.175)	-299(-0.200)	
35	A	TOTAL	1334*(0.242)	981 (0.178)	1442 (0.261)	1582 (0.287)	
		High	647 (0.433)	155 (0.104)	294 (0.197)	389 (0.260)	
		Low	253 (0.169)	360 (0.240)	412 (0.275)	329 (0.220)	
		Diff	394 (0.264)	-205(-0.137)	-118(-0.079)	60 (0.040)	
36	B	TOTAL	626 (0.113)	3955*(0.717)	423 (0.077)	345 (0.063)	
		High	26 (0.017)	1435 (0.960)	22 (0.015)	10 (0.007)	
		Low	348 (0.232)	546 (0.365)	235 (0.157)	223 (0.149)	
		Diff	-322(-0.215)	889 (0.595)	-213(-0.142)	-213(-0.142)	
37	B	TOTAL	386 (0.070)	3471*(0.629)	579 (0.105)	916 (0.166)	
		High	21 (0.014)	1351 (0.904)	21 (0.014)	102 (0.068)	
		Low	221 (0.148)	462 (0.309)	315 (0.210)	353 (0.236)	
		Diff	-200(-0.134)	889 (0.595)	-294(-0.196)	-251(-0.168)	
38	C	TOTAL	1468 (0.266)	474 (0.086)	2390*(0.433)	1010 (0.183)	
		High	227 (0.152)	57 (0.038)	1074 (0.718)	135 (0.090)	
		Low	386 (0.258)	243 (0.162)	329 (0.220)	393 (0.263)	
		Diff	-159(-0.106)	-186(-0.124)	745 (0.499)	-258(-0.172)	

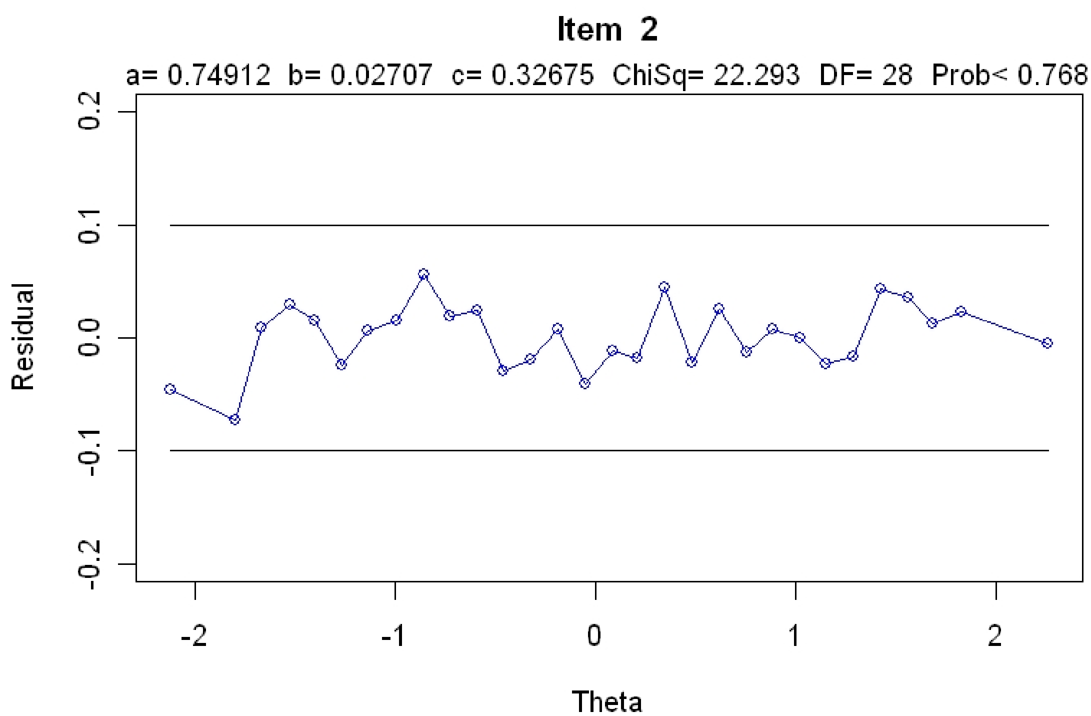
39		TOTAL	815 (0.148)	676 (0.123)	869 (0.158)	959*(0.174)	2198 (0.398)
		High	97 (0.065)	133 (0.089)	354 (0.237)	705 (0.472)	206 (0.138)
		Low	306 (0.204)	138 (0.092)	55 (0.037)	17 (0.011)	981 (0.655)
		Diff	-209(-0.140)	-5(-0.003)	299 (0.200)	688 (0.460)	-775(-0.518)
40	D	TOTAL	299 (0.054)	515 (0.093)	396 (0.072)	3801*(0.689)	
		High	9 (0.006)	19 (0.013)	6 (0.004)	1428 (0.955)	
		Low	186 (0.124)	279 (0.186)	254 (0.170)	474 (0.317)	
		Diff	-177(-0.118)	-260(-0.174)	-248(-0.166)	954 (0.639)	
41	A	TOTAL	3555*(0.644)	566 (0.103)	584 (0.106)	320 (0.058)	
		High	1408 (0.942)	26 (0.017)	18 (0.012)	10 (0.007)	
		Low	408 (0.273)	285 (0.190)	314 (0.210)	197 (0.132)	
		Diff	1000 (0.669)	-259(-0.173)	-296(-0.198)	-187(-0.125)	
42	D	TOTAL	734 (0.133)	1070 (0.194)	652 (0.118)	2552*(0.463)	
		High	152 (0.102)	121 (0.081)	44 (0.029)	1144 (0.765)	
		Low	247 (0.165)	359 (0.240)	298 (0.199)	286 (0.191)	
		Diff	-95(-0.063)	-238(-0.159)	-254(-0.170)	858 (0.574)	
43	D	TOTAL	1158 (0.210)	868 (0.157)	631 (0.114)	2366*(0.429)	
		High	146 (0.098)	127 (0.085)	30 (0.020)	1158 (0.775)	
		Low	371 (0.248)	282 (0.188)	322 (0.215)	233 (0.156)	
		Diff	-225(-0.150)	-155(-0.103)	-292(-0.195)	925 (0.619)	
44	B	TOTAL	1037 (0.188)	2127*(0.386)	921 (0.167)	906 (0.164)	
		High	157 (0.105)	999 (0.668)	136 (0.091)	164 (0.110)	
		Low	334 (0.223)	270 (0.180)	311 (0.208)	276 (0.184)	
		Diff	-177(-0.118)	729 (0.488)	-175(-0.117)	-112(-0.075)	
45	A	TOTAL	3125*(0.566)	466 (0.084)	1000 (0.181)	421 (0.076)	
		High	1278 (0.855)	21 (0.014)	135 (0.090)	28 (0.019)	
		Low	374 (0.250)	272 (0.182)	348 (0.232)	202 (0.135)	
		Diff	904 (0.605)	-251(-0.168)	-213(-0.142)	-174(-0.116)	

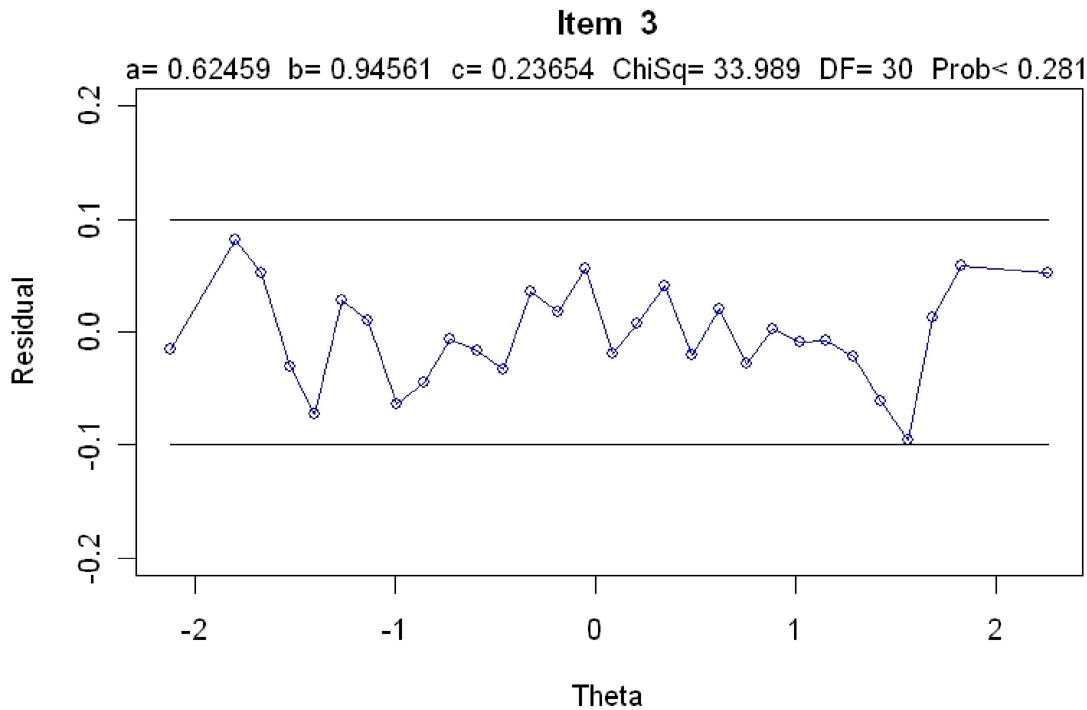
Note: Distractor analysis carried out with a sample of N=5,517. High includes the top 25% of examinees and low includes the bottom 25% of examinees.

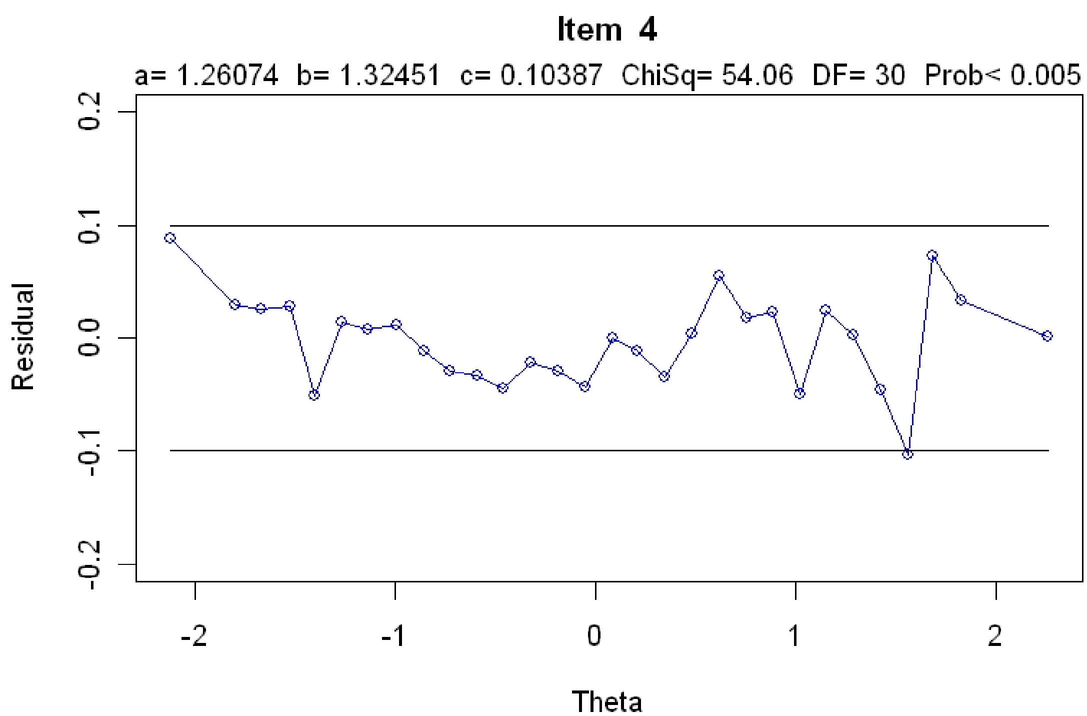
Appendix B

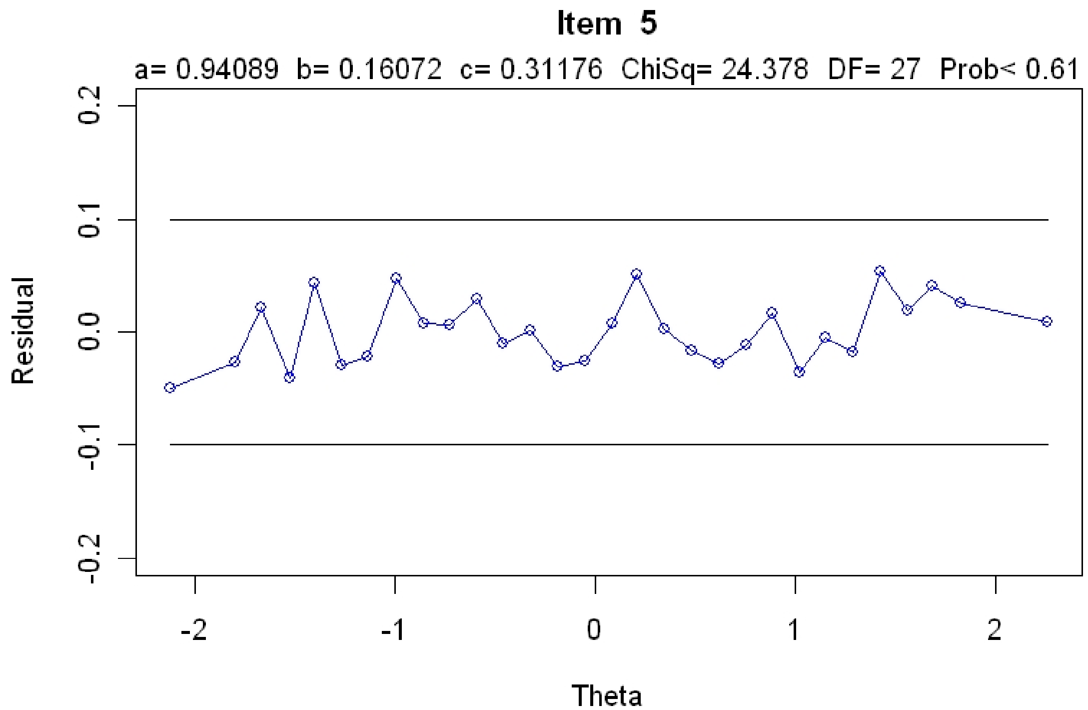
Residuals for Dichotomously Scored Items



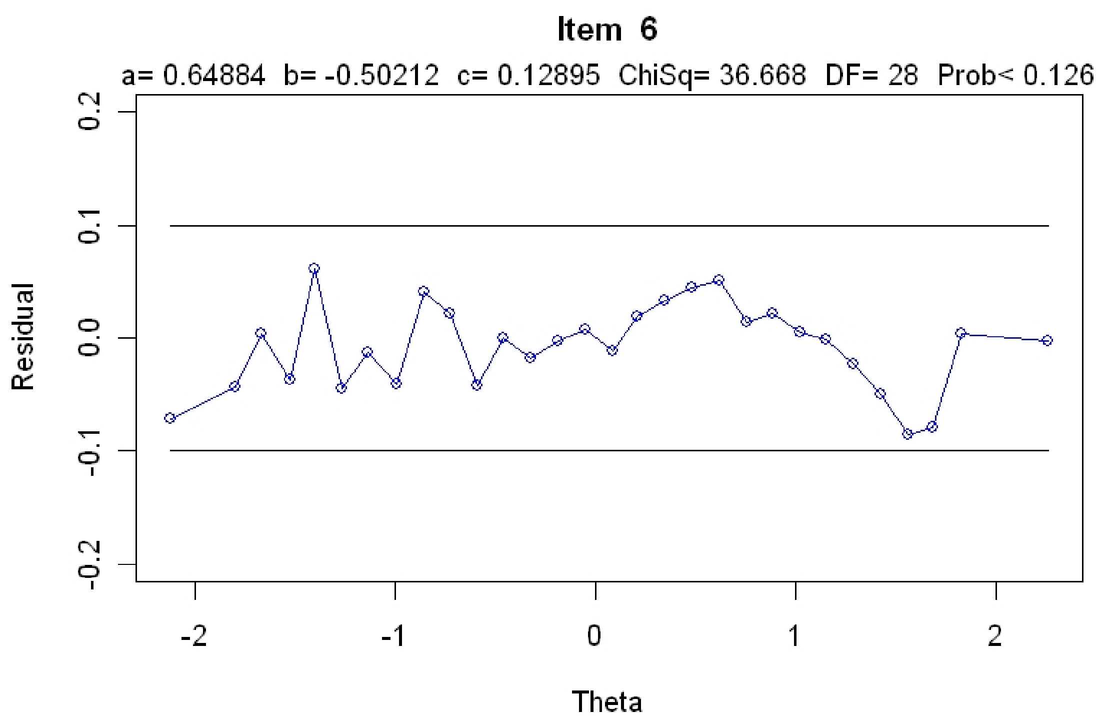



Item 3
a= 0.62459 b= 0.94561 c= 0.23654 ChiSq= 33.989 DF= 30 Prob< 0.281
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta

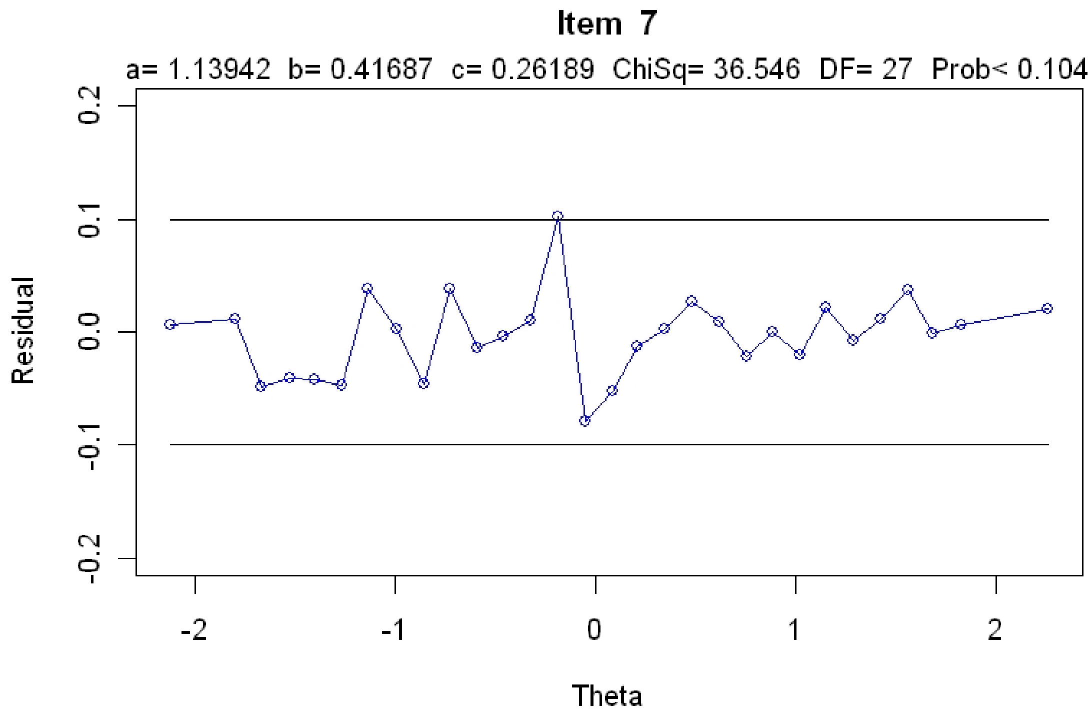

Item 4
a= 1.26074 b= 1.32451 c= 0.10387 ChiSq= 54.06 DF= 30 Prob< 0.005
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta

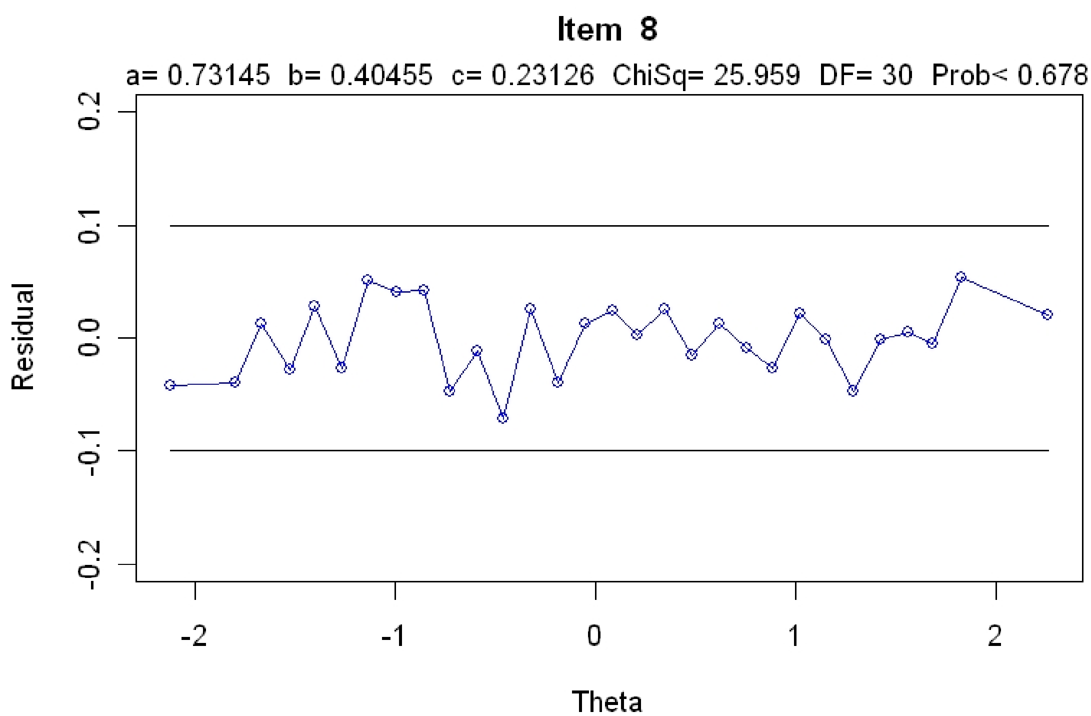


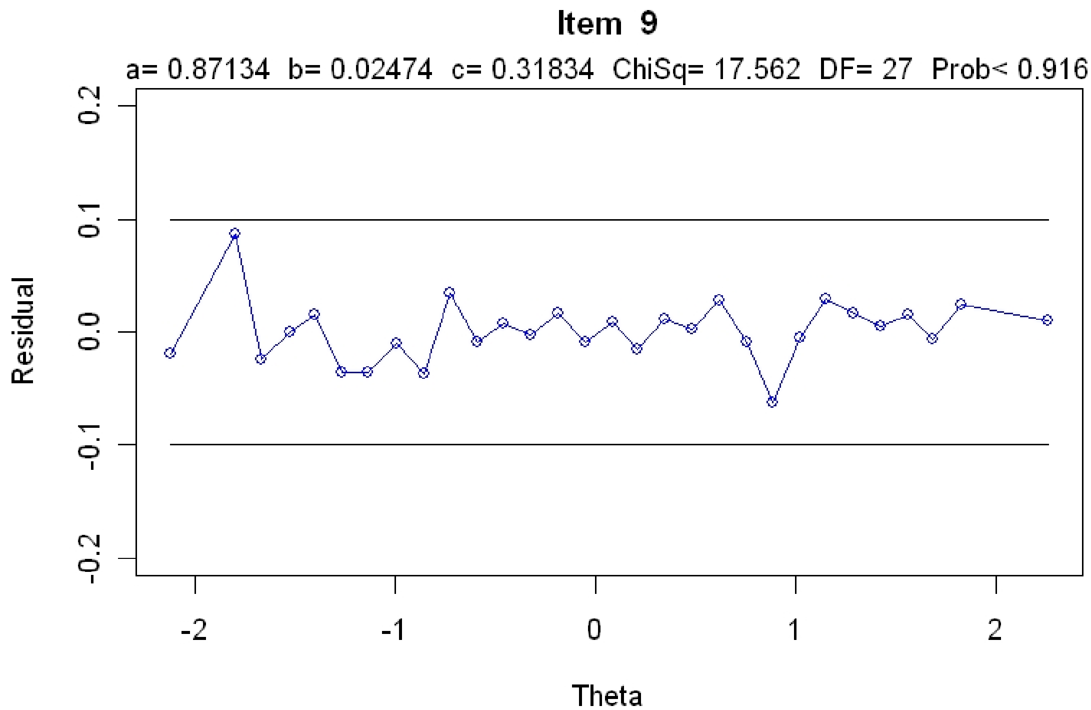
Item 5
a= 0.94089 b= 0.16072 c= 0.31176 ChiSq= 24.378 DF= 27 Prob< 0.61
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta

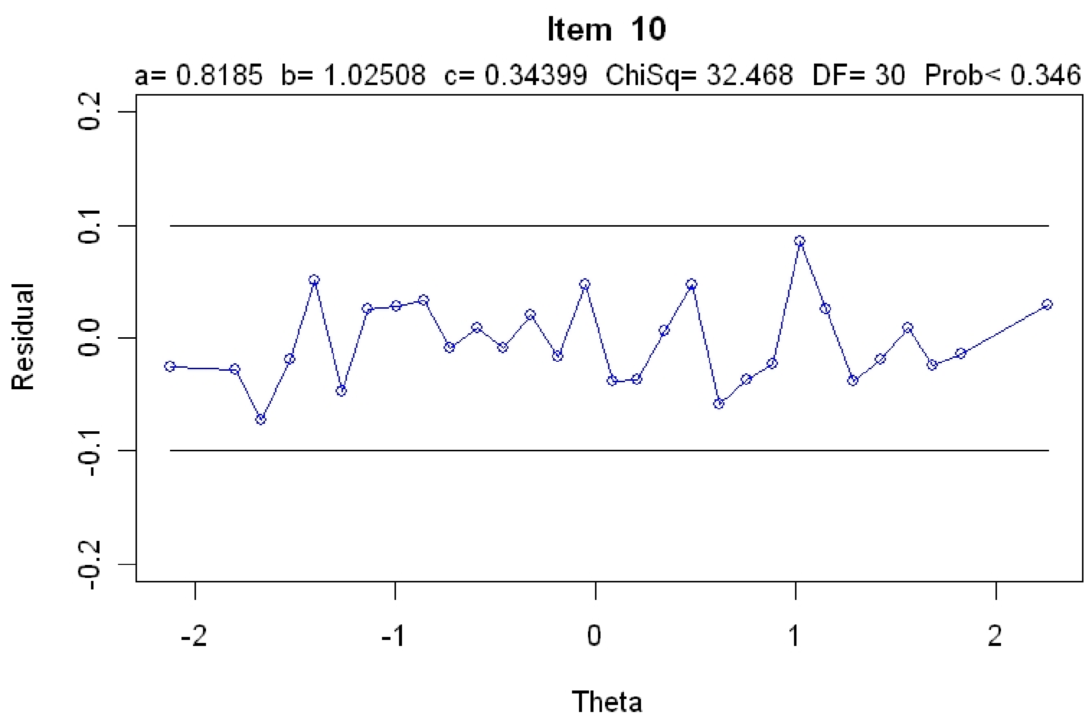


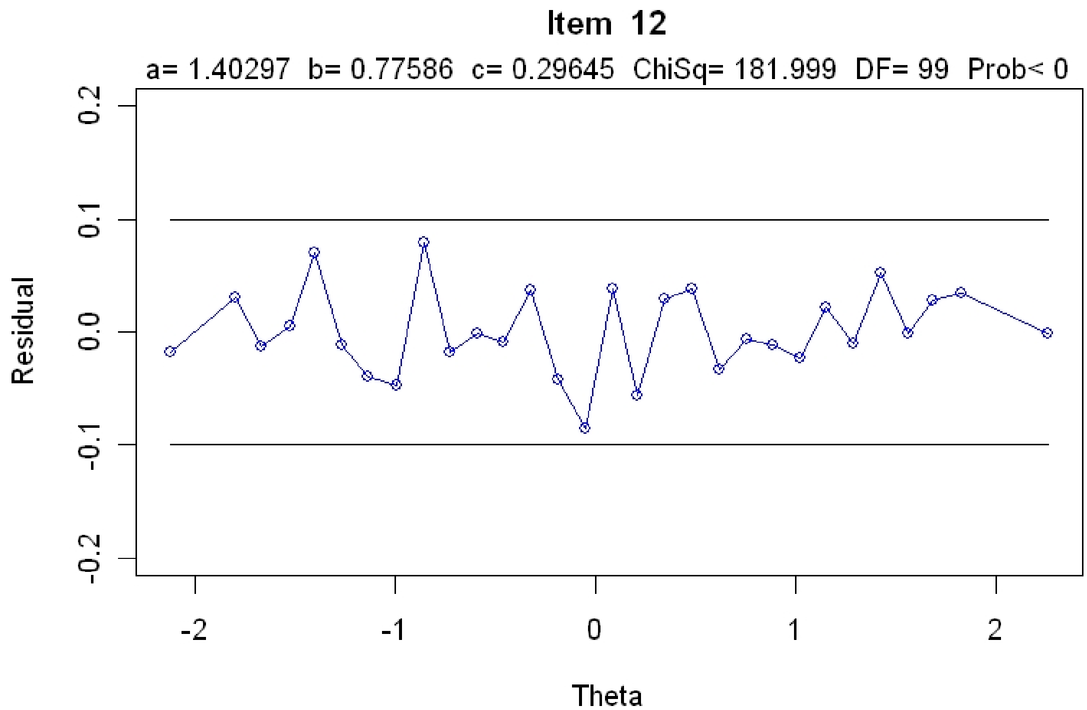
Item 6
a= 0.64884 b= -0.50212 c= 0.12895 ChiSq= 36.668 DF= 28 Prob< 0.126
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta

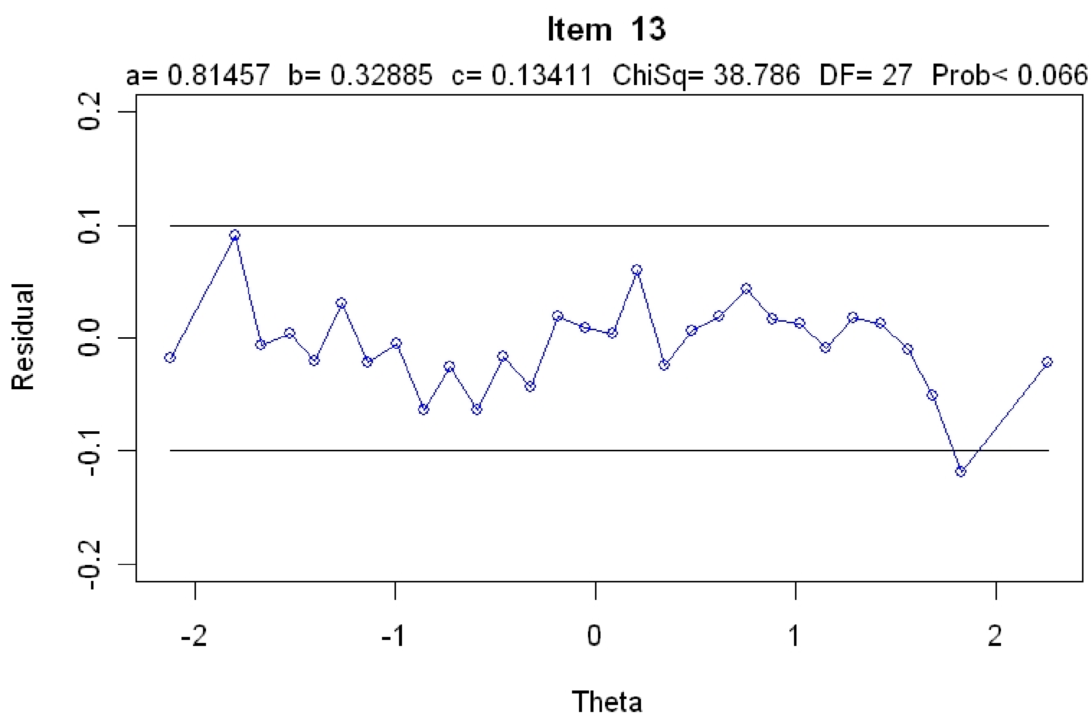

Item 7
a= 1.13942 b= 0.41687 c= 0.26189 ChiSq= 36.546 DF= 27 Prob< 0.104
Residual
Theta
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2

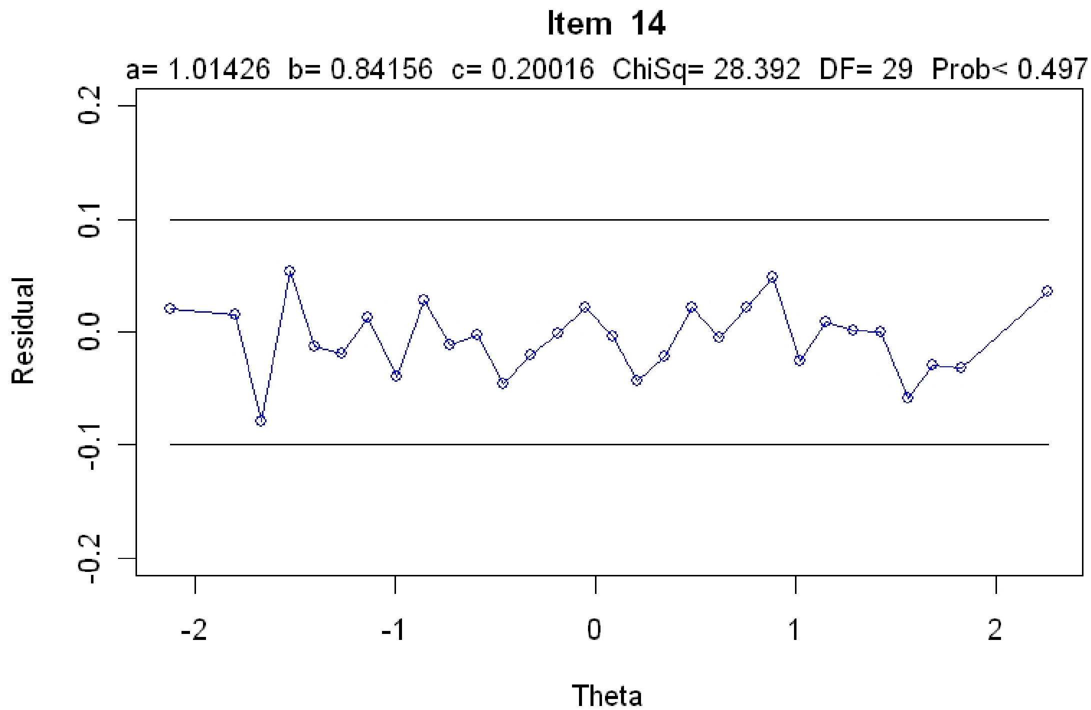

Item 8
a= 0.73145 b= 0.40455 c= 0.23126 ChiSq= 25.959 DF= 30 Prob< 0.678
Residual
Theta
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2

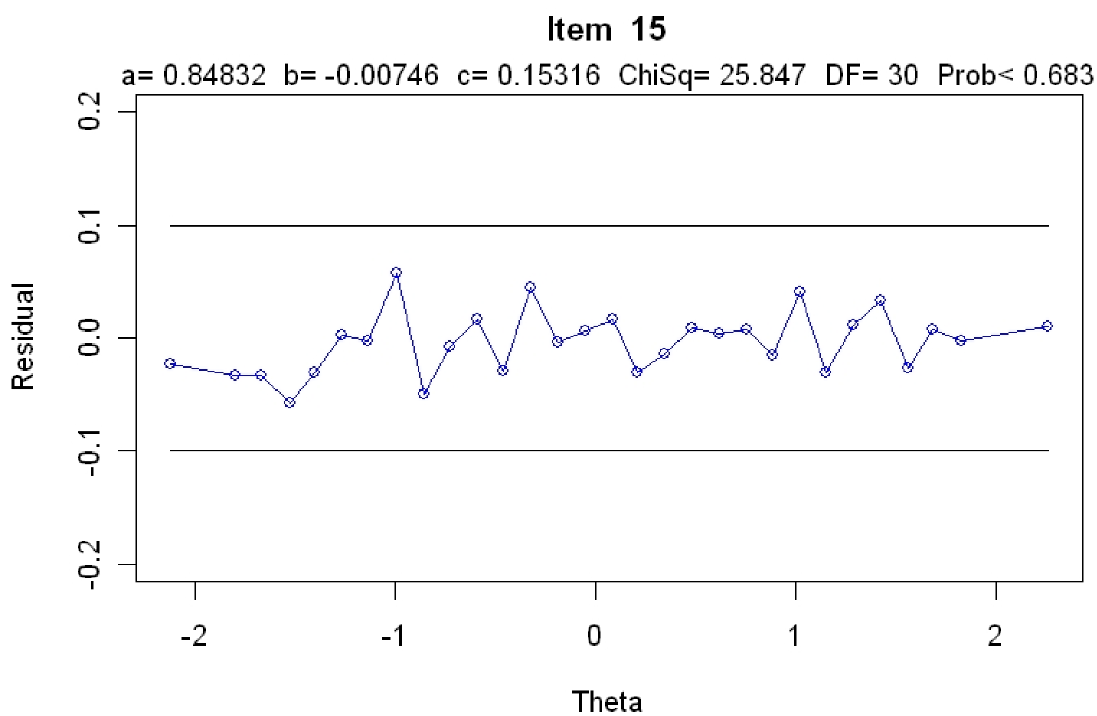

Item 9
a= 0.87134 b= 0.02474 c= 0.31834 ChiSq= 17.562 DF= 27 Prob< 0.916
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta

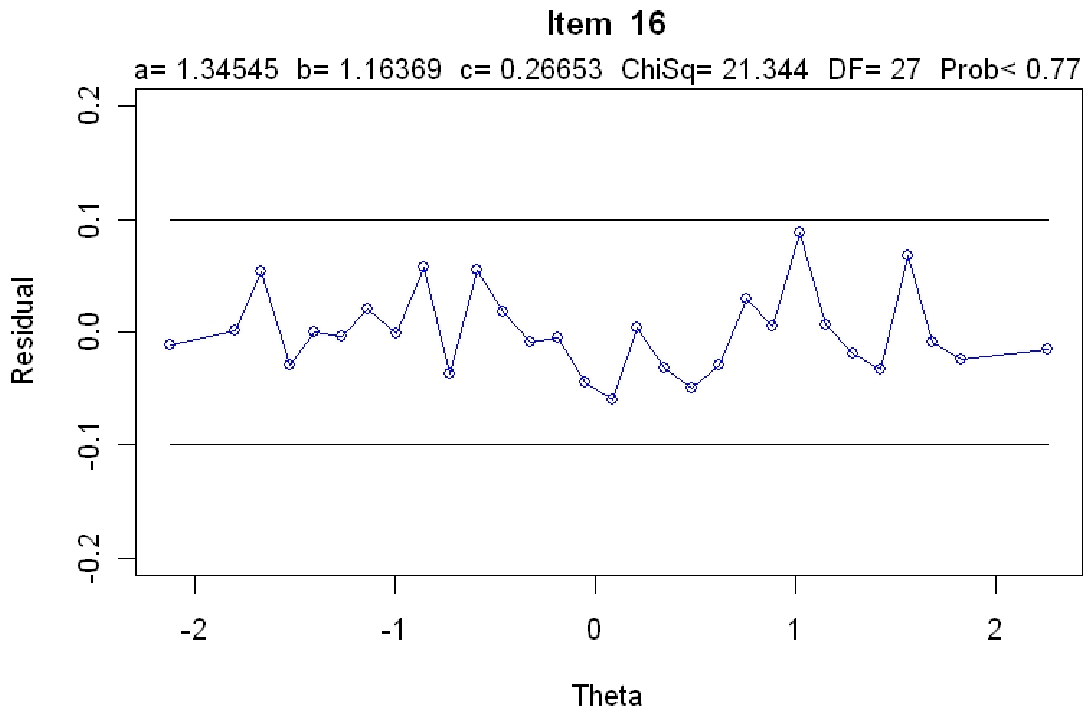

Item 10
a= 0.8185 b= 1.02508 c= 0.34399 ChiSq= 32.468 DF= 30 Prob< 0.346
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta

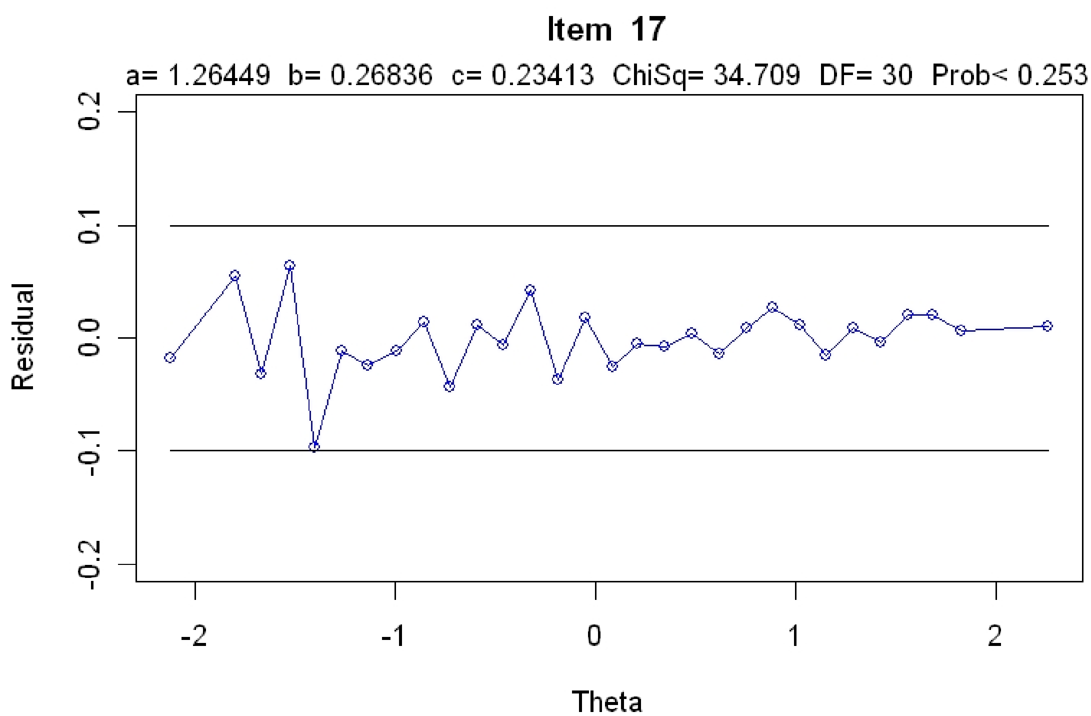

Item 12
a= 1.40297 b= 0.77586 c= 0.29645 ChiSq= 181.999 DF= 99 Prob< 0
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta

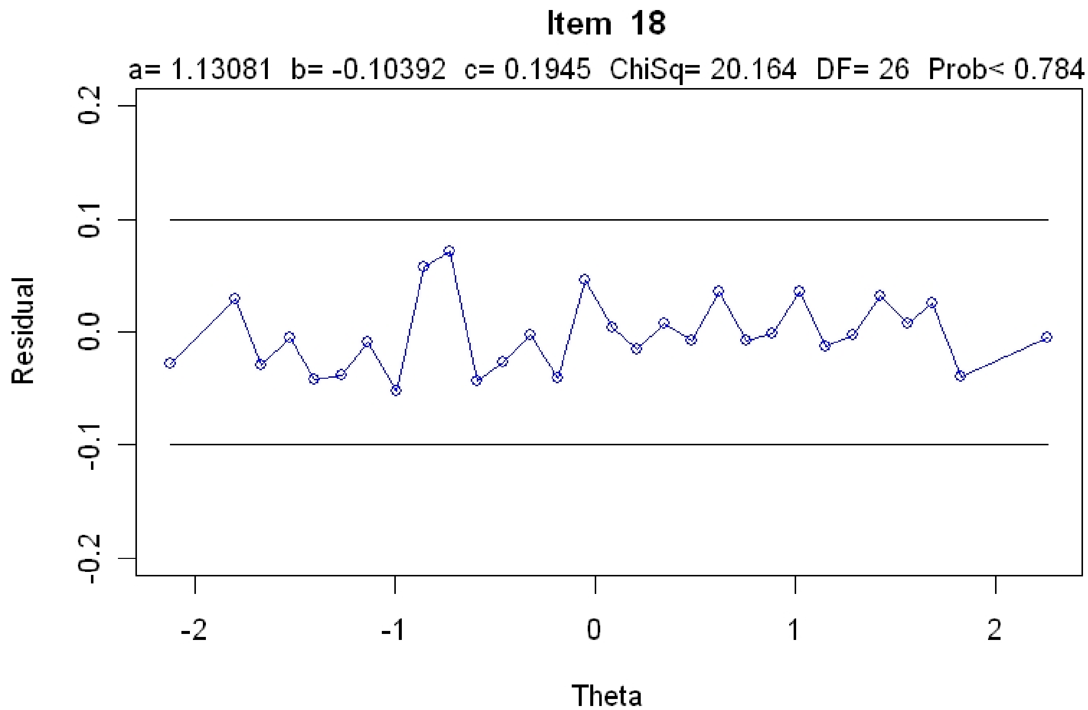

Item 13
a= 0.81457 b= 0.32885 c= 0.13411 ChiSq= 38.786 DF= 27 Prob< 0.066
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta

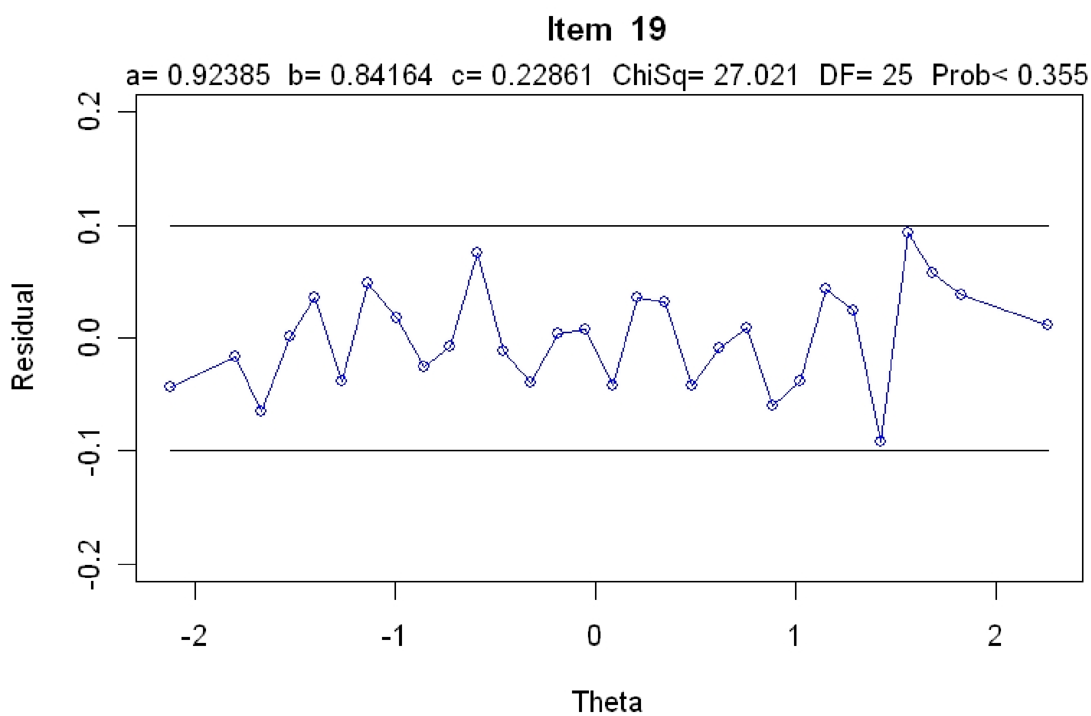

Item 14
a= 1.01426 b= 0.84156 c= 0.20016 ChiSq= 28.392 DF= 29 Prob< 0.497
Residual
-0.2
-0.1
0.0
0.1
0.2
-2
-1
0
1
2
Theta

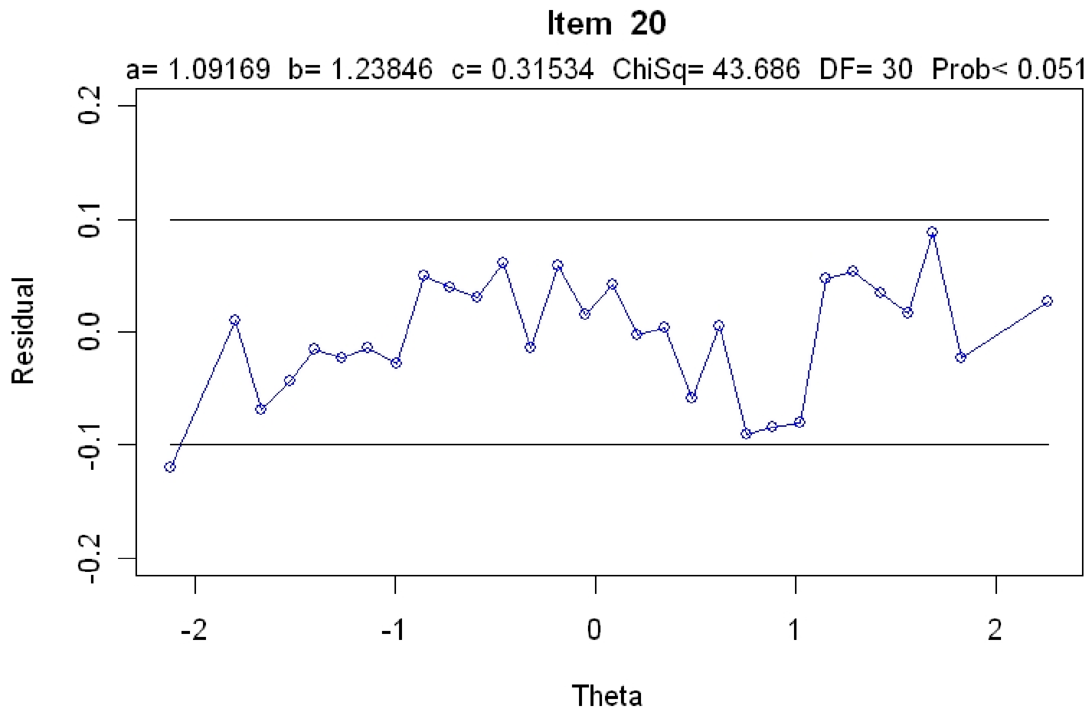

Item 15
a= 0.84832 b= -0.00746 c= 0.15316 ChiSq= 25.847 DF= 30 Prob< 0.683
Residual
-0.2
-0.1
0.0
0.1
0.2
-2
-1
0
1
2
Theta

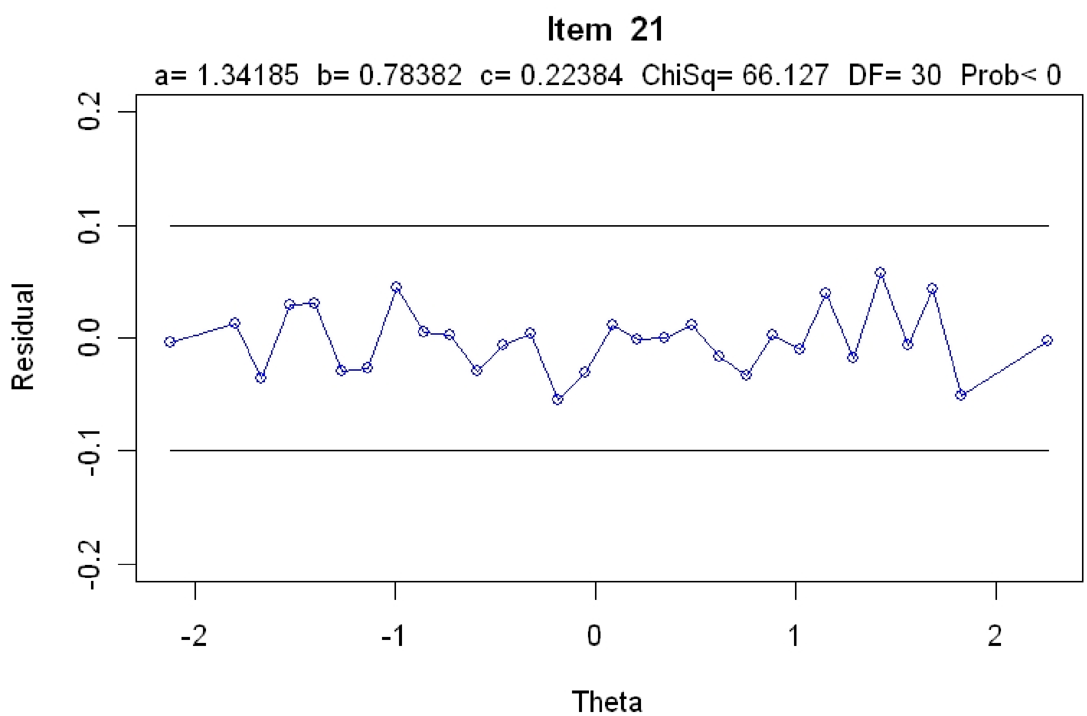

Item 16
a= 1.34545 b= 1.16369 c= 0.26653 ChiSq= 21.344 DF= 27 Prob< 0.77
Residual
Theta

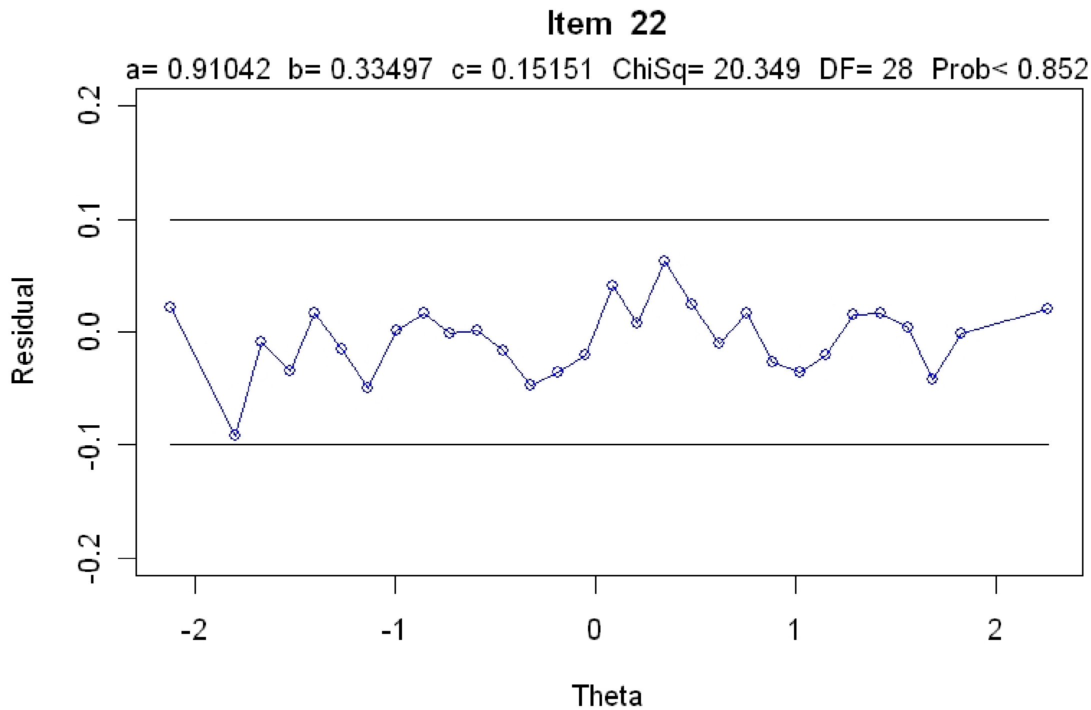

Item 17
a= 1.26449 b= 0.26836 c= 0.23413 ChiSq= 34.709 DF= 30 Prob< 0.253
Residual
Theta


Item 18
a= 1.13081 b= -0.10392 c= 0.1945 ChiSq= 20.164 DF= 26 Prob< 0.784
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta

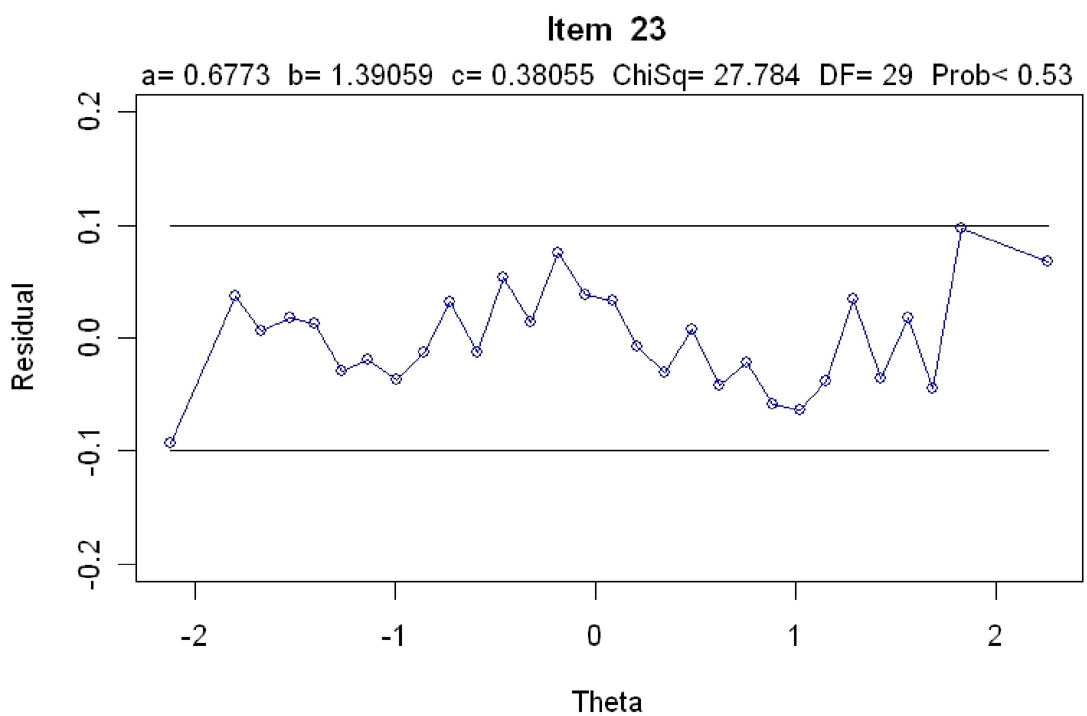

Item 19
a= 0.92385 b= 0.84164 c= 0.22861 ChiSq= 27.021 DF= 25 Prob< 0.355
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta


Item 20
a= 1.09169 b= 1.23846 c= 0.31534 ChiSq= 43.686 DF= 30 Prob< 0.051
Residual
Theta
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2

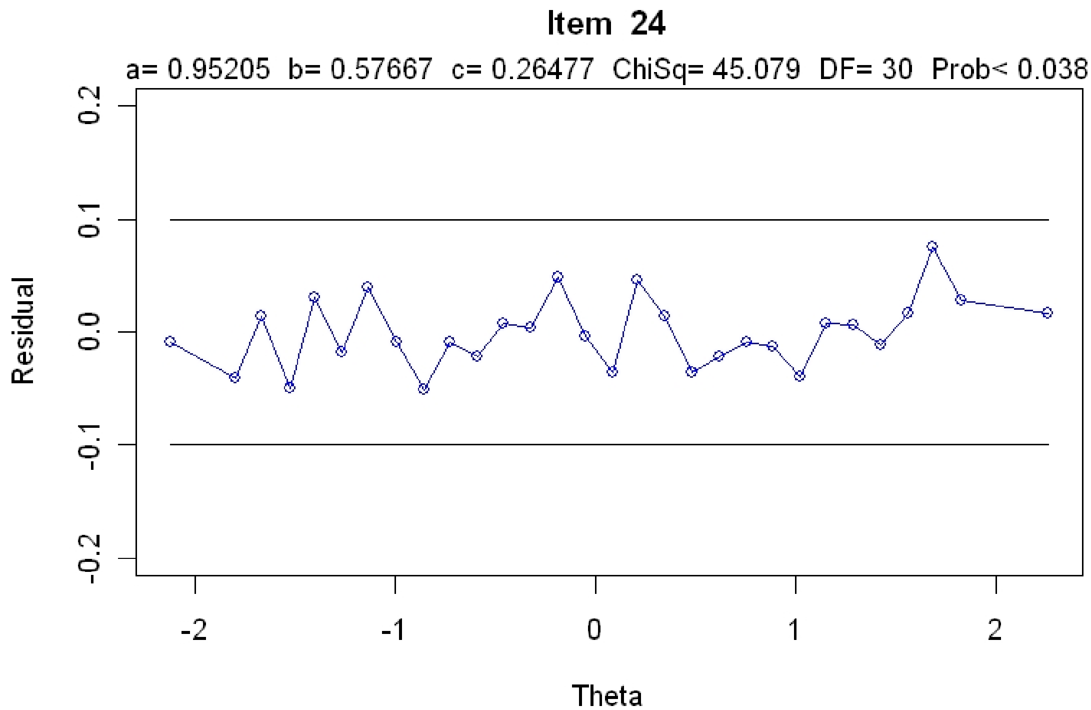

Item 21
a= 1.34185 b= 0.78382 c= 0.22384 ChiSq= 66.127 DF= 30 Prob< 0
Residual
Theta
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2

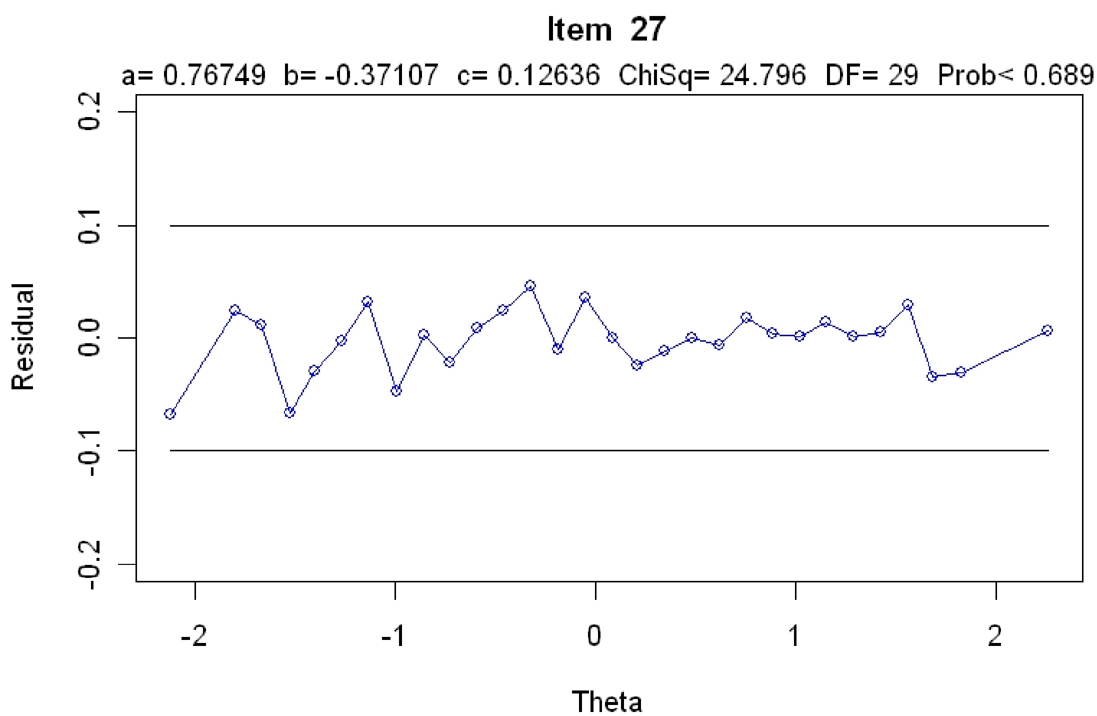


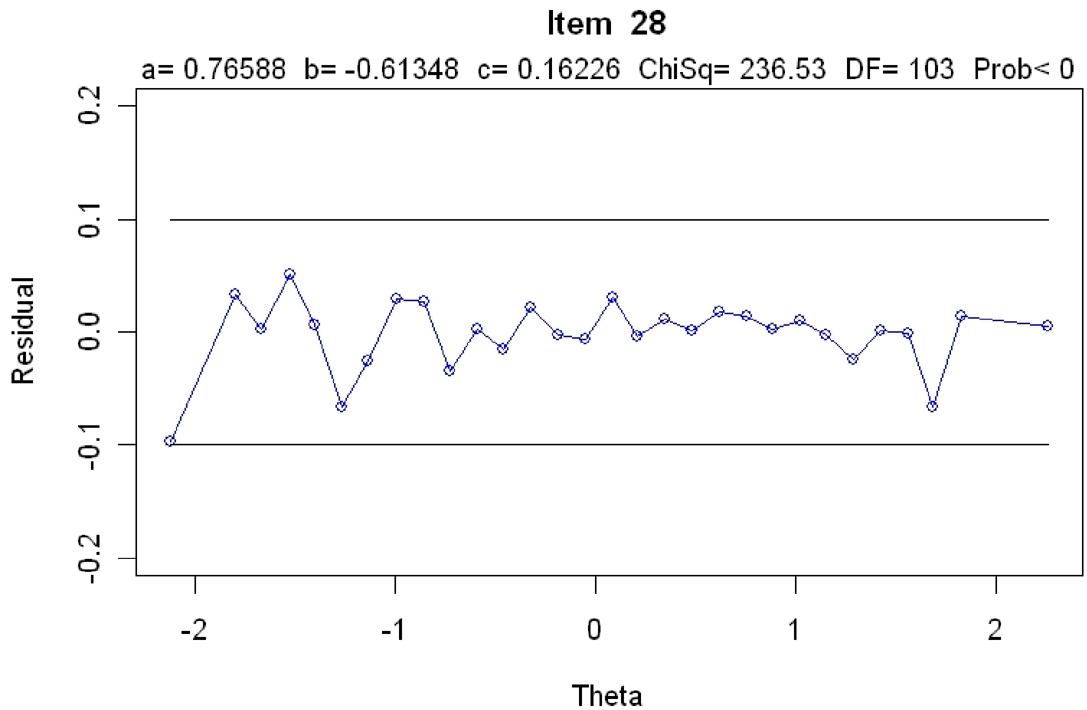
Item 22
a= 0.91042 b= 0.33497 c= 0.15151 ChiSq= 20.349 DF= 28 Prob< 0.852
Residual
Theta

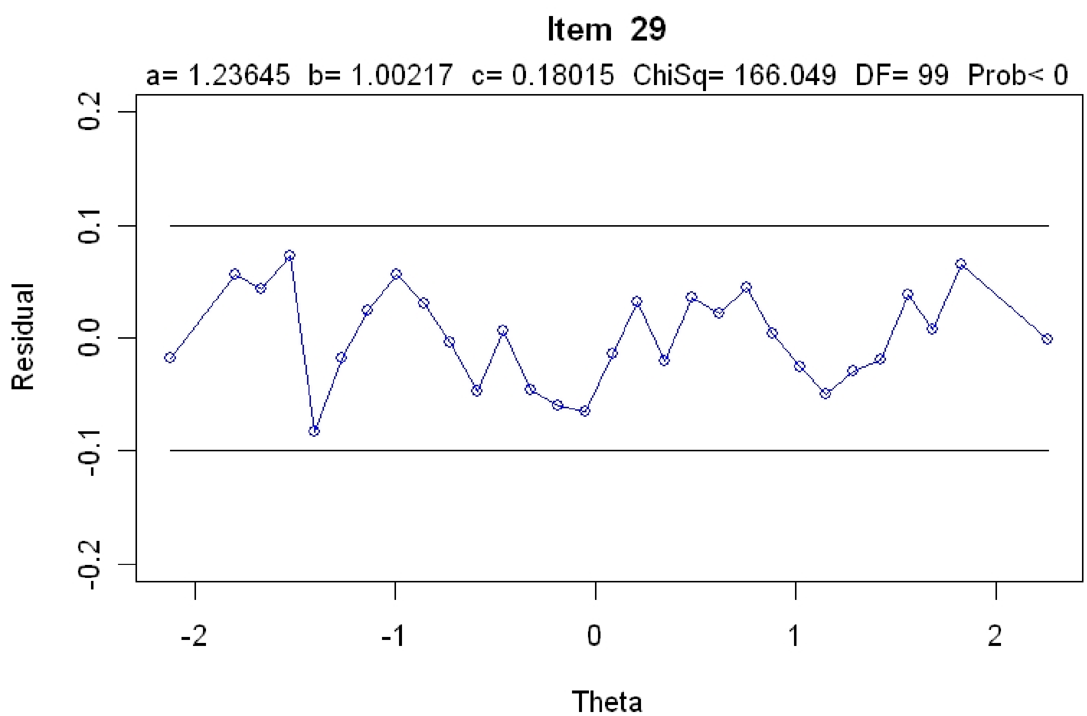


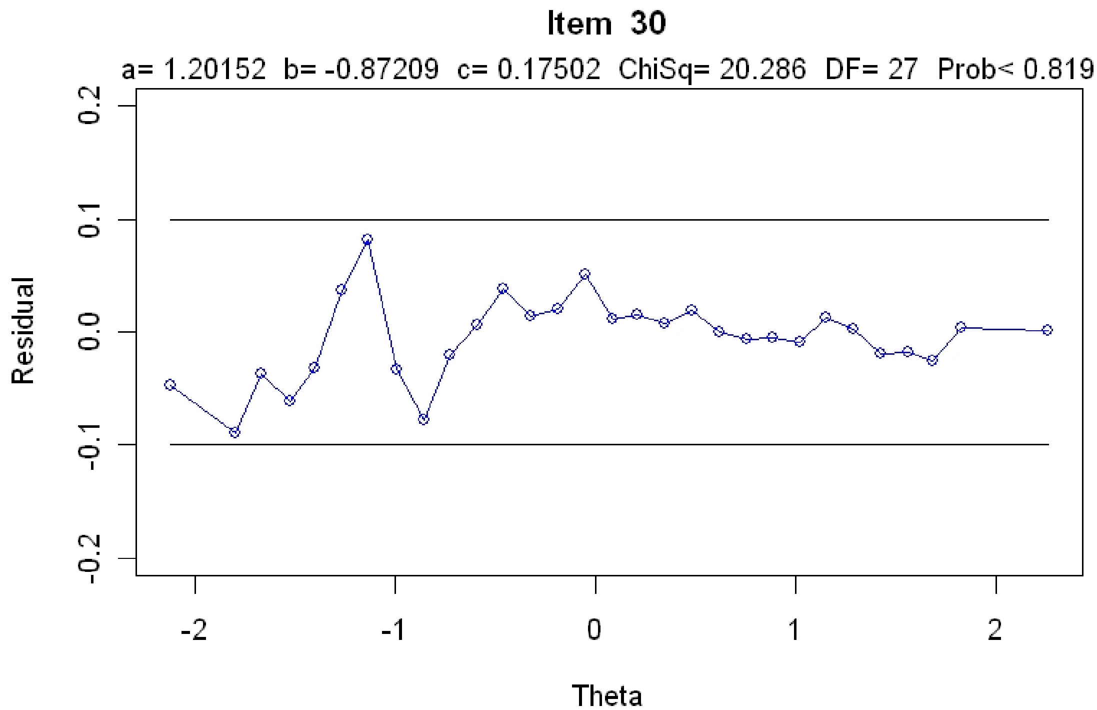
Item 23
a= 0.6773 b= 1.39059 c= 0.38055 ChiSq= 27.784 DF= 29 Prob< 0.53
Residual
Theta

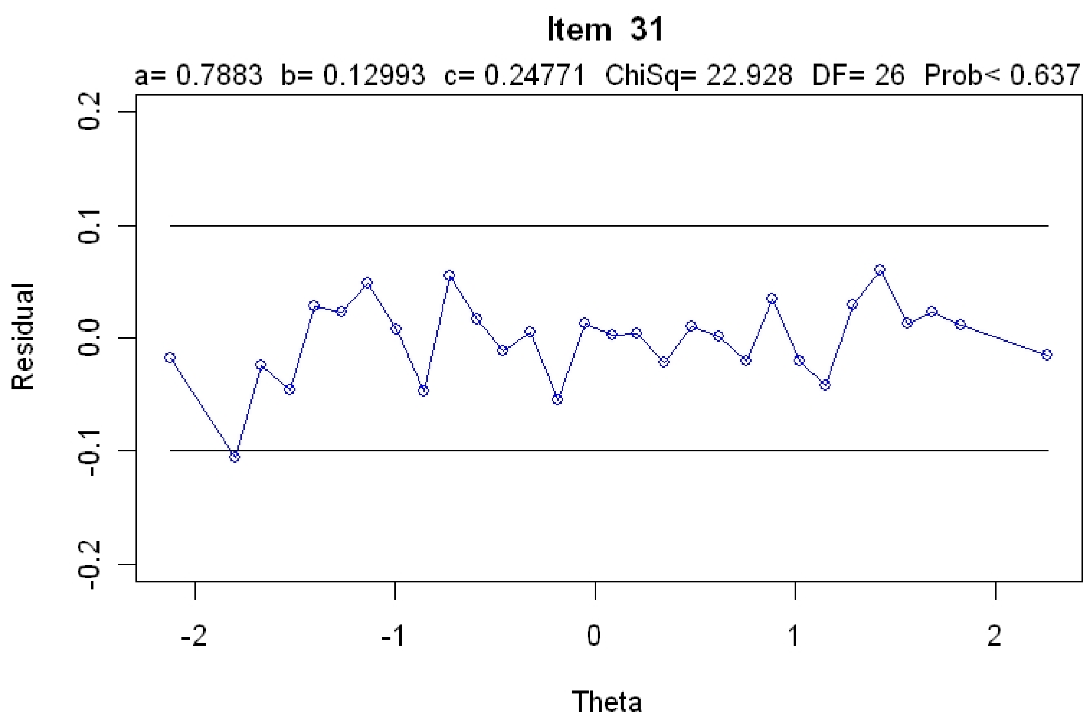

Item 24
a= 0.95205 b= 0.57667 c= 0.26477 ChiSq= 45.079 DF= 30 Prob< 0.038
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta

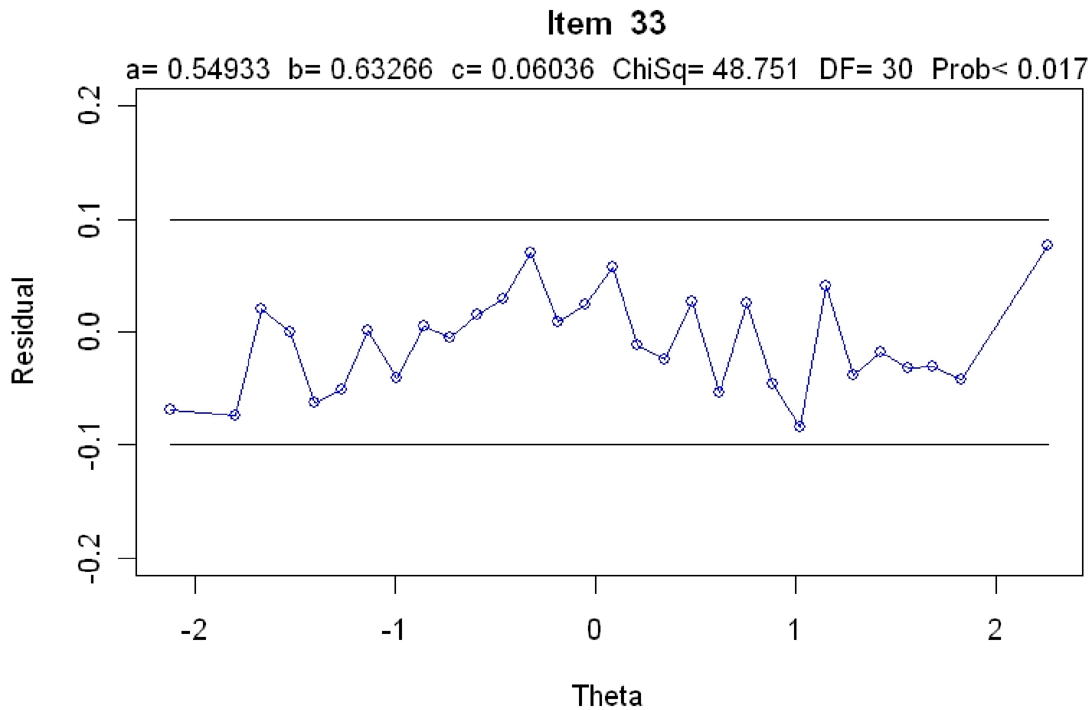

Item 27
a= 0.76749 b= -0.37107 c= 0.12636 ChiSq= 24.796 DF= 29 Prob< 0.689
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta

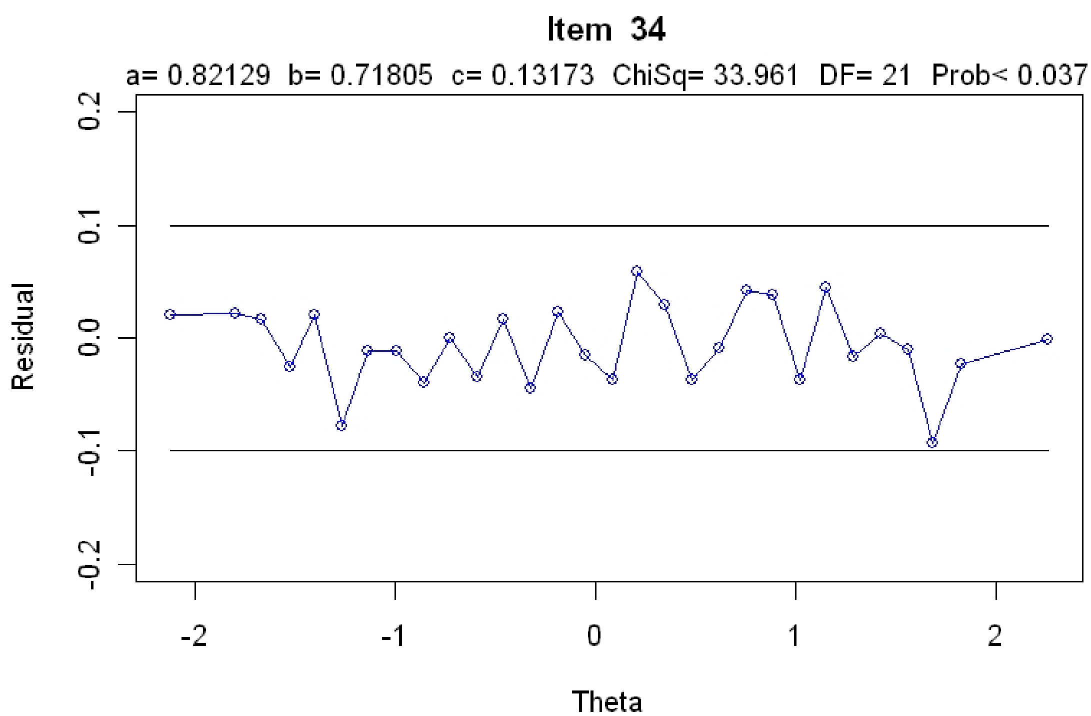

Item 28
a= 0.76588 b= -0.61348 c= 0.16226 ChiSq= 236.53 DF= 103 Prob< 0
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta

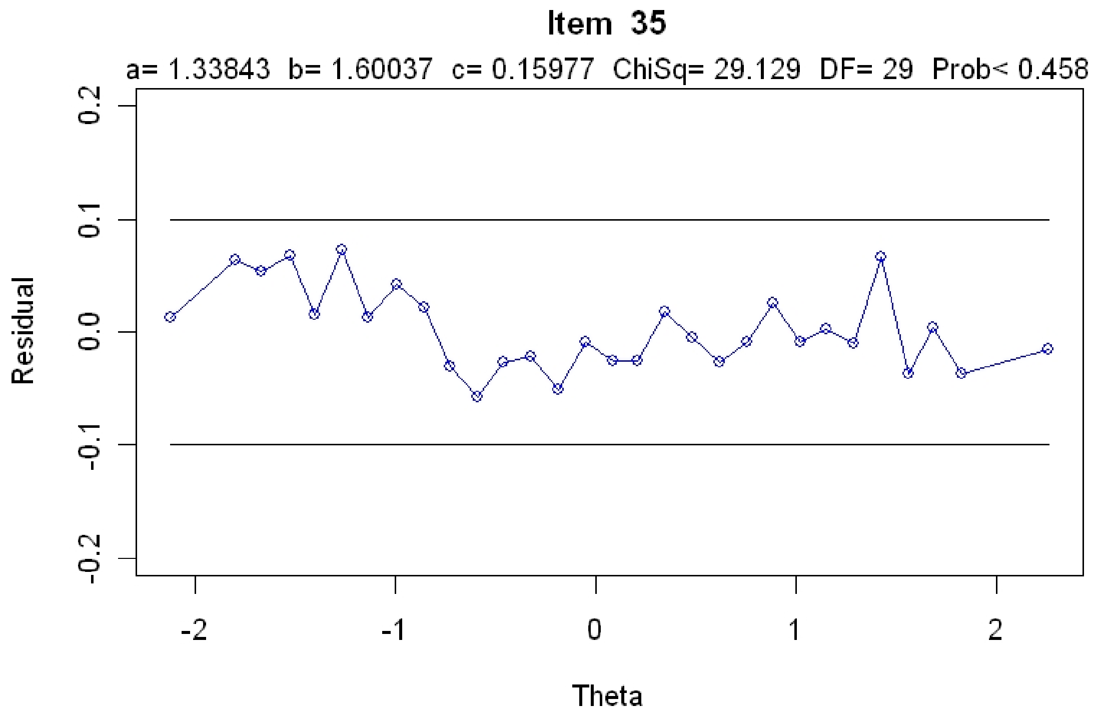

Item 29
a= 1.23645 b= 1.00217 c= 0.18015 ChiSq= 166.049 DF= 99 Prob< 0
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta

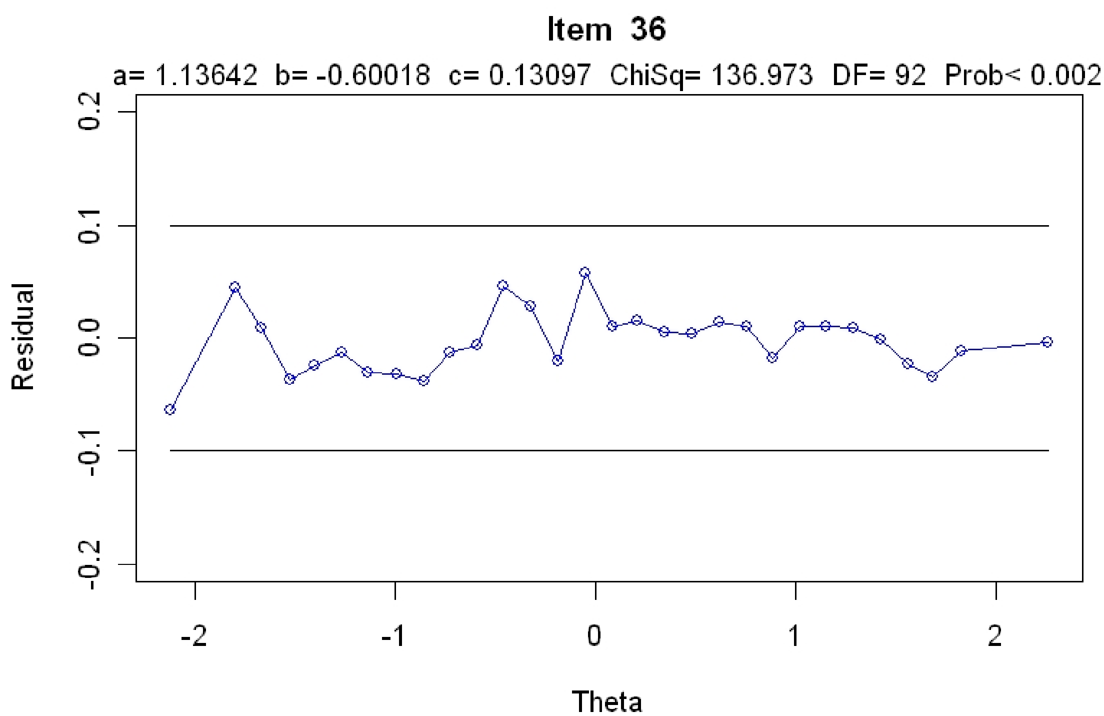

Item 30
a= 1.20152 b= -0.87209 c= 0.17502 ChiSq= 20.286 DF= 27 Prob< 0.819
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta

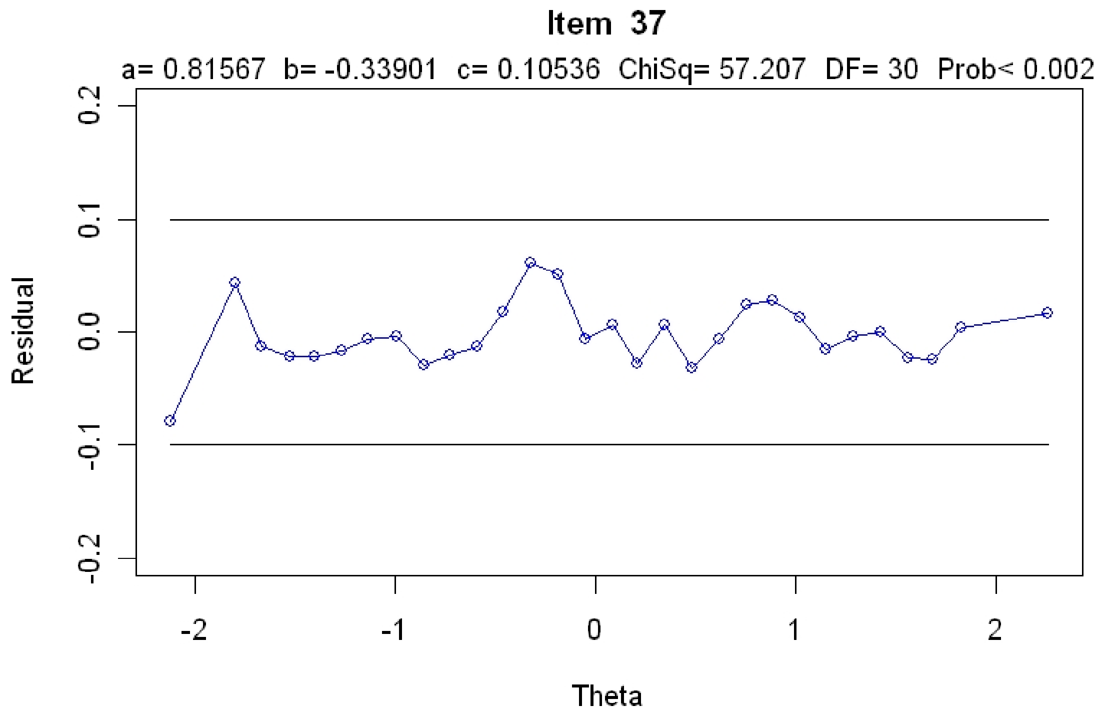

Item 31
a= 0.7883 b= 0.12993 c= 0.24771 ChiSq= 22.928 DF= 26 Prob< 0.637
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta

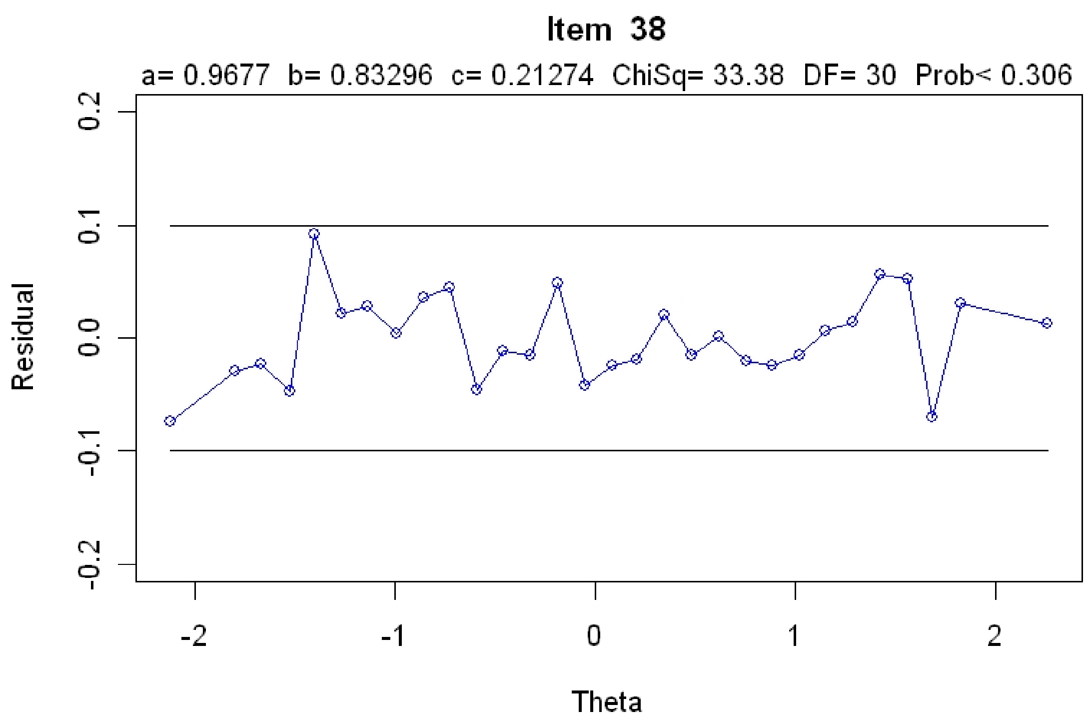

Item 33
a= 0.54933 b= 0.63266 c= 0.06036 ChiSq= 48.751 DF= 30 Prob< 0.017
Residual
Theta
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2

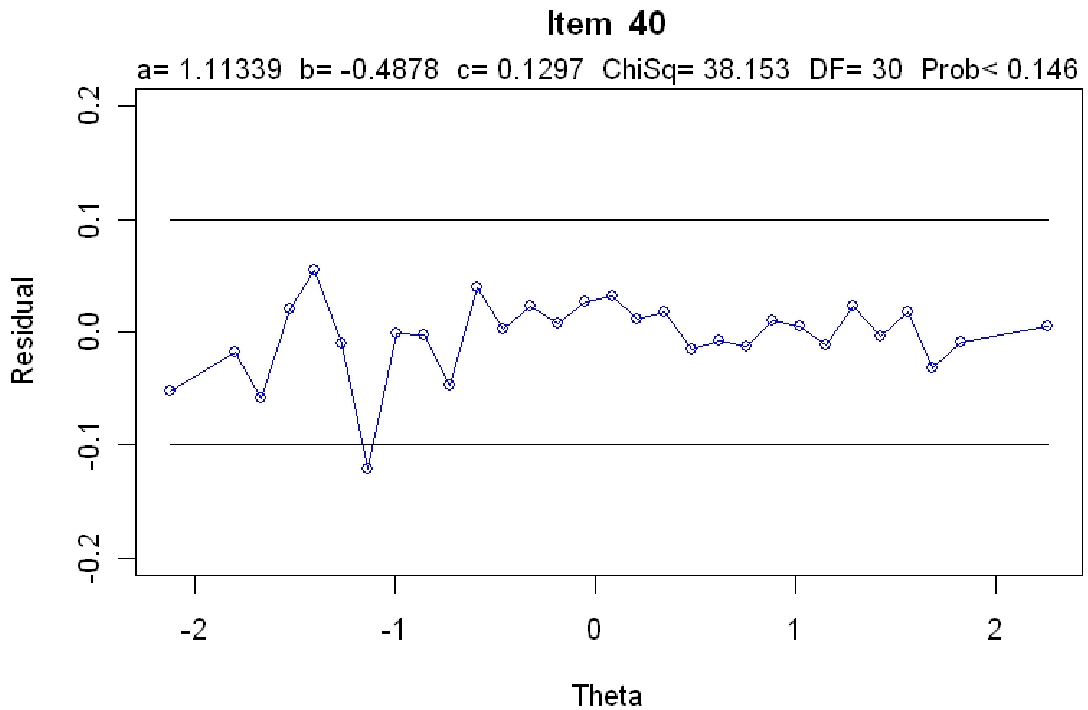

Item 34
a= 0.82129 b= 0.71805 c= 0.13173 ChiSq= 33.961 DF= 21 Prob< 0.037
Residual
Theta
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2

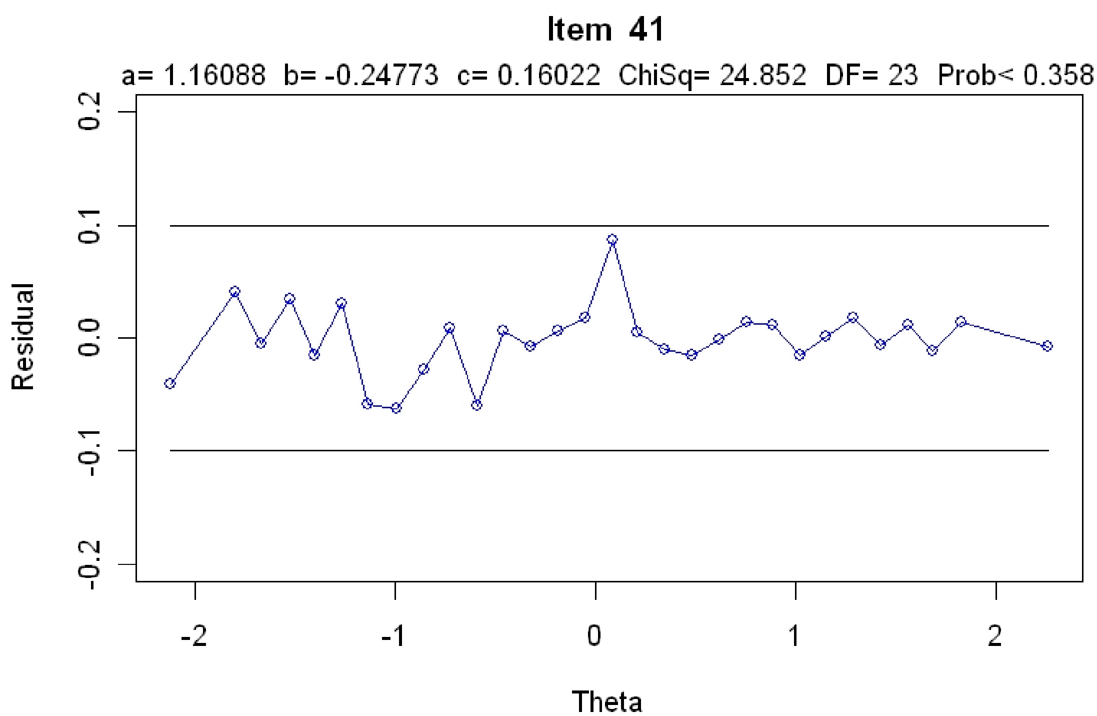

Item 35
a= 1.33843 b= 1.60037 c= 0.15977 ChiSq= 29.129 DF= 29 Prob< 0.458
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta

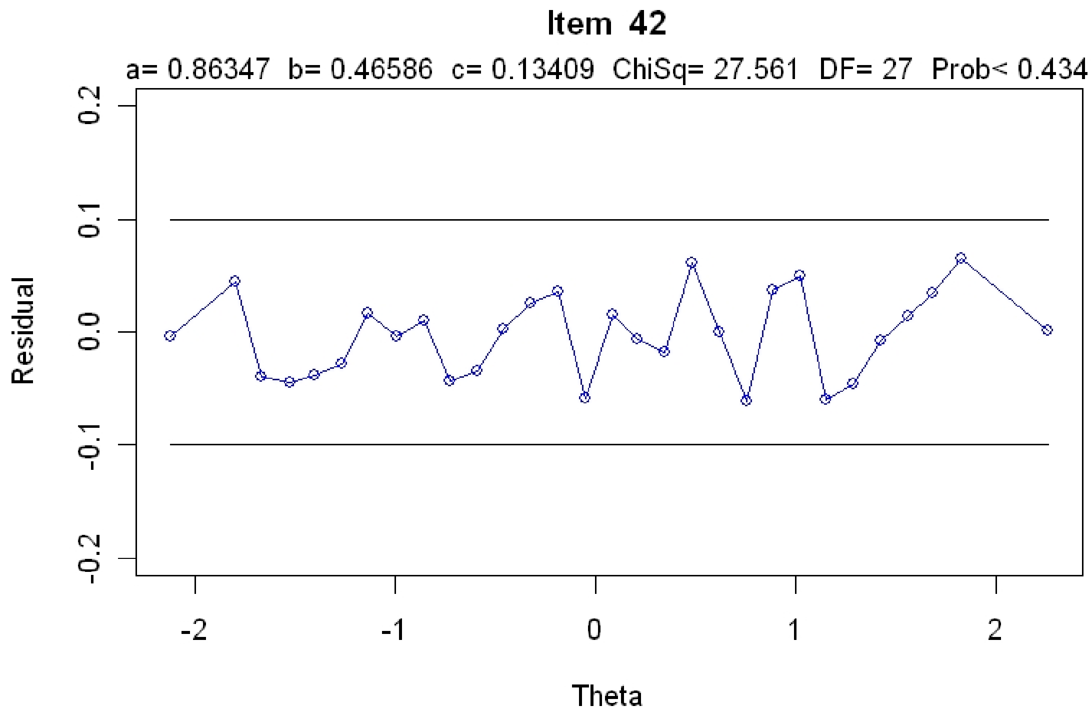

Item 36
a= 1.13642 b= -0.60018 c= 0.13097 ChiSq= 136.973 DF= 92 Prob< 0.002
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta

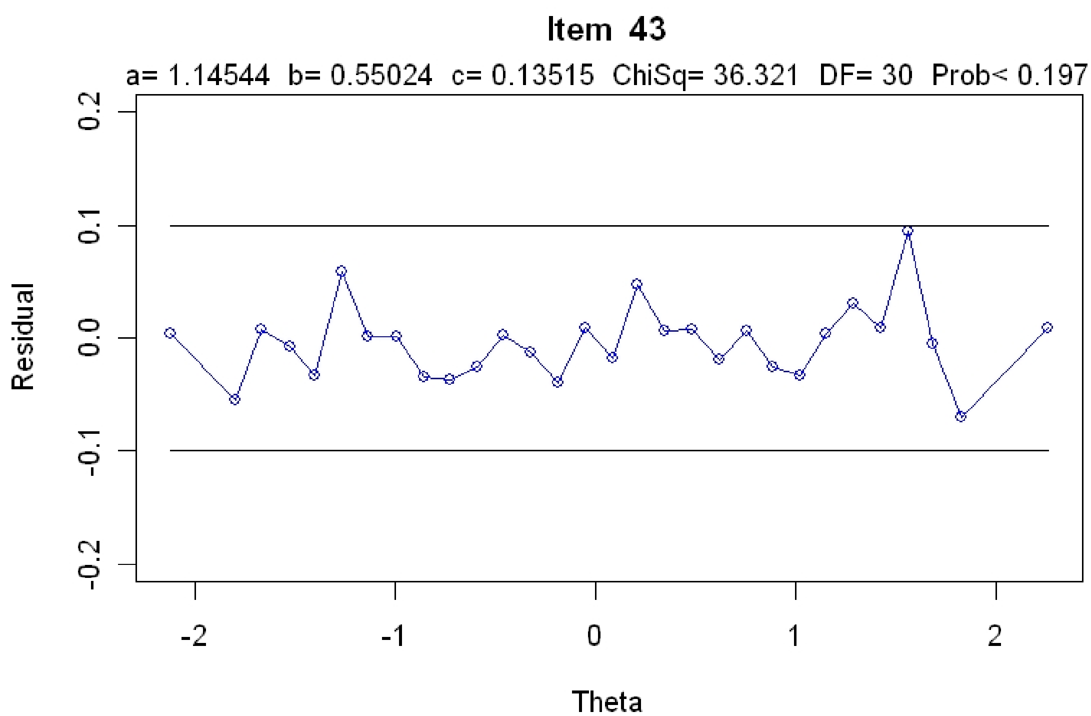

Item 37
a= 0.81567 b= -0.33901 c= 0.10536 ChiSq= 57.207 DF= 30 Prob< 0.002
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta

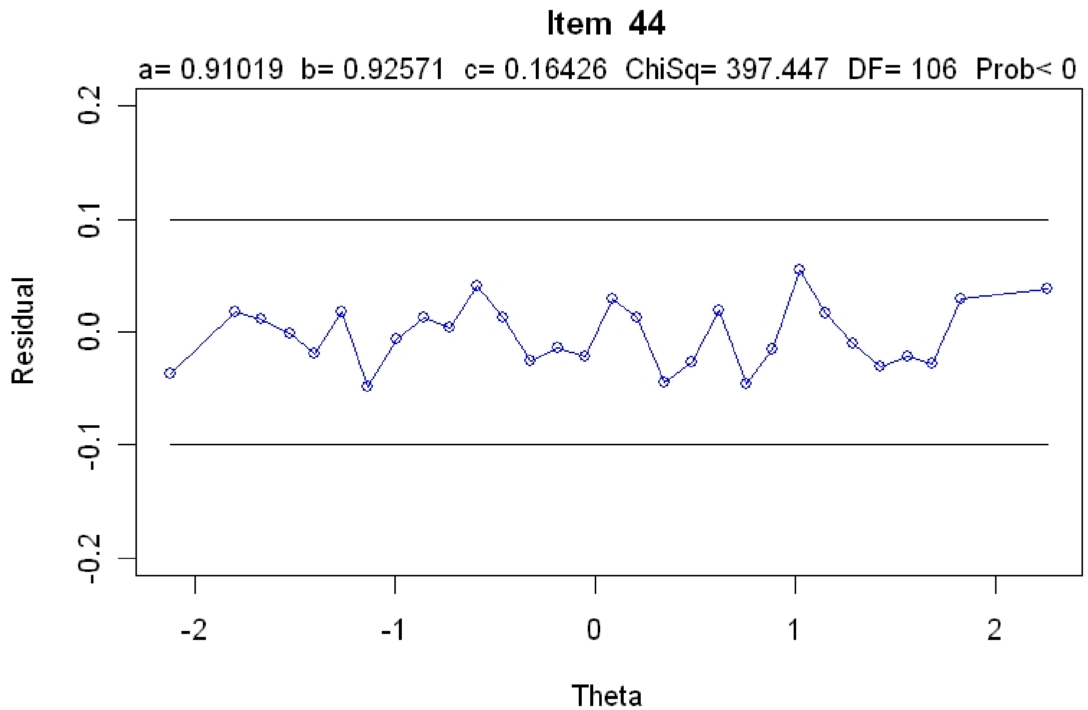

Item 38
a= 0.9677 b= 0.83296 c= 0.21274 ChiSq= 33.38 DF= 30 Prob< 0.306
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta

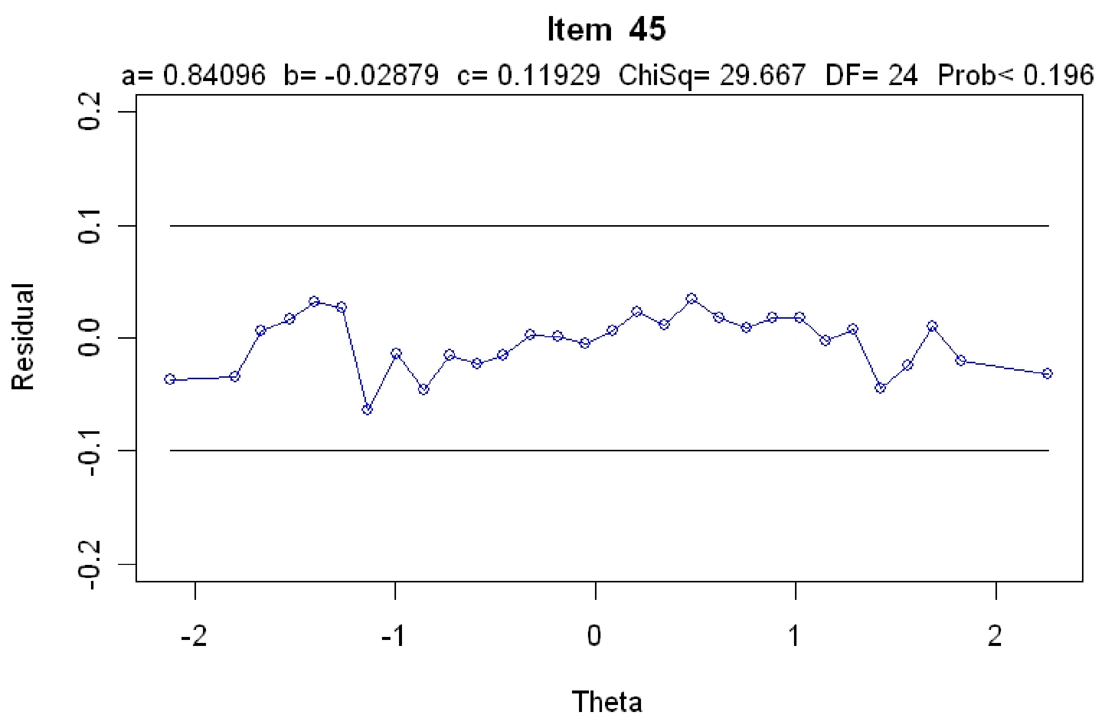

Item 40
a= 1.11339 b= -0.4878 c= 0.1297 ChiSq= 38.153 DF= 30 Prob< 0.146
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta


Item 41
a= 1.16088 b= -0.24773 c= 0.16022 ChiSq= 24.852 DF= 23 Prob< 0.358
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta


Item 42
a= 0.86347 b= 0.46586 c= 0.13409 ChiSq= 27.561 DF= 27 Prob< 0.434
Residual
Theta
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2

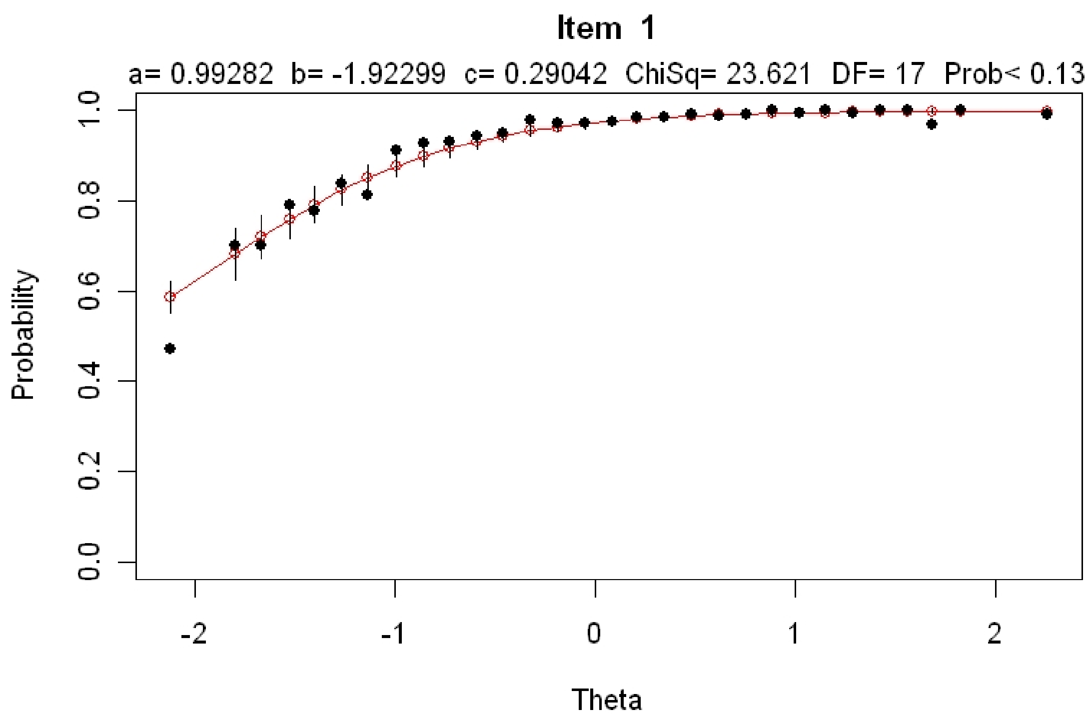

Item 43
a= 1.14544 b= 0.55024 c= 0.13515 ChiSq= 36.321 DF= 30 Prob< 0.197
Residual
Theta
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2

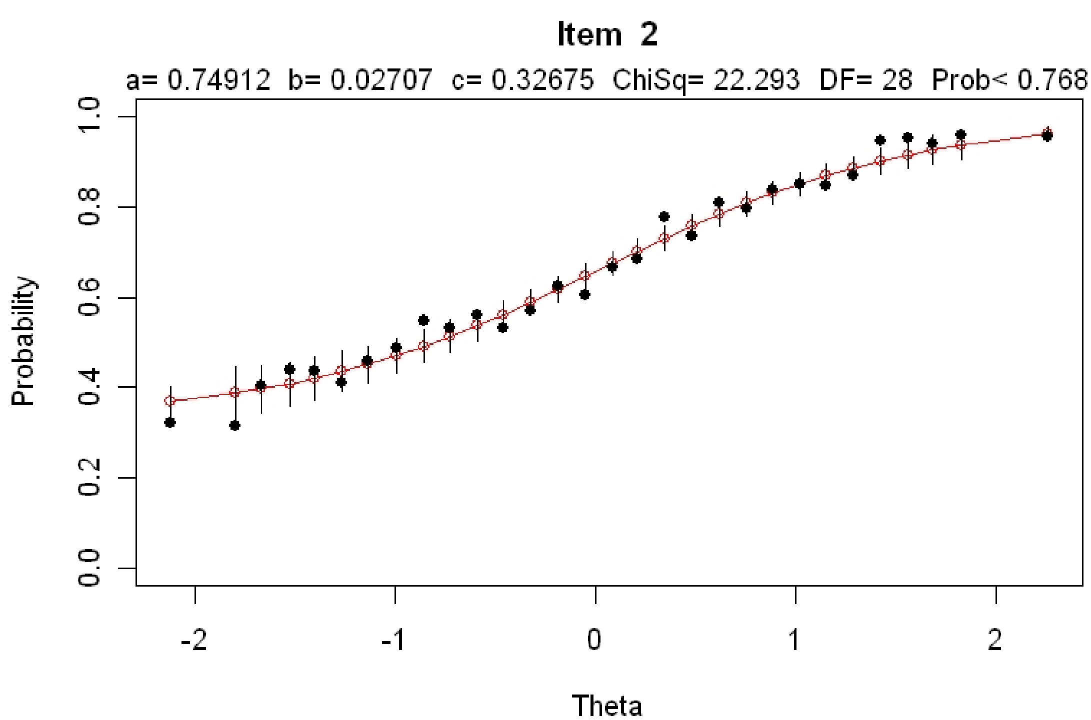

Item 44
a= 0.91019 b= 0.92571 c= 0.16426 ChiSq= 397.447 DF= 106 Prob< 0
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta

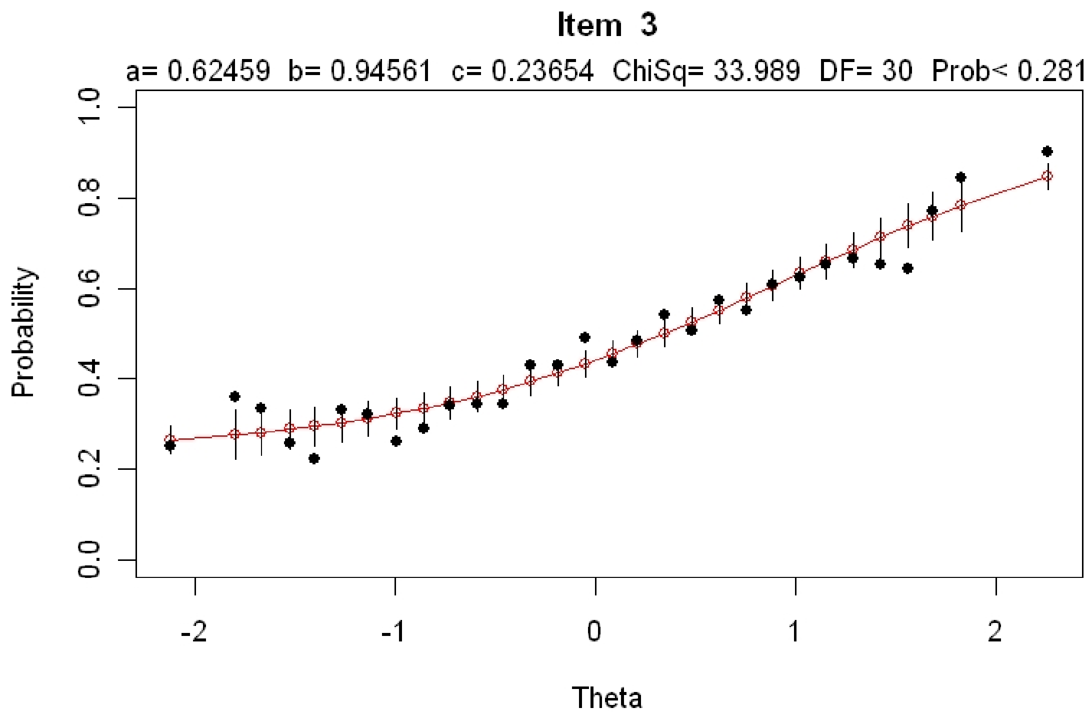

Item 45
a= 0.84096 b= -0.02879 c= 0.11929 ChiSq= 29.667 DF= 24 Prob< 0.196
Residual
0.2
0.1
0.0
-0.1
-0.2
-2
-1
0
1
2
Theta

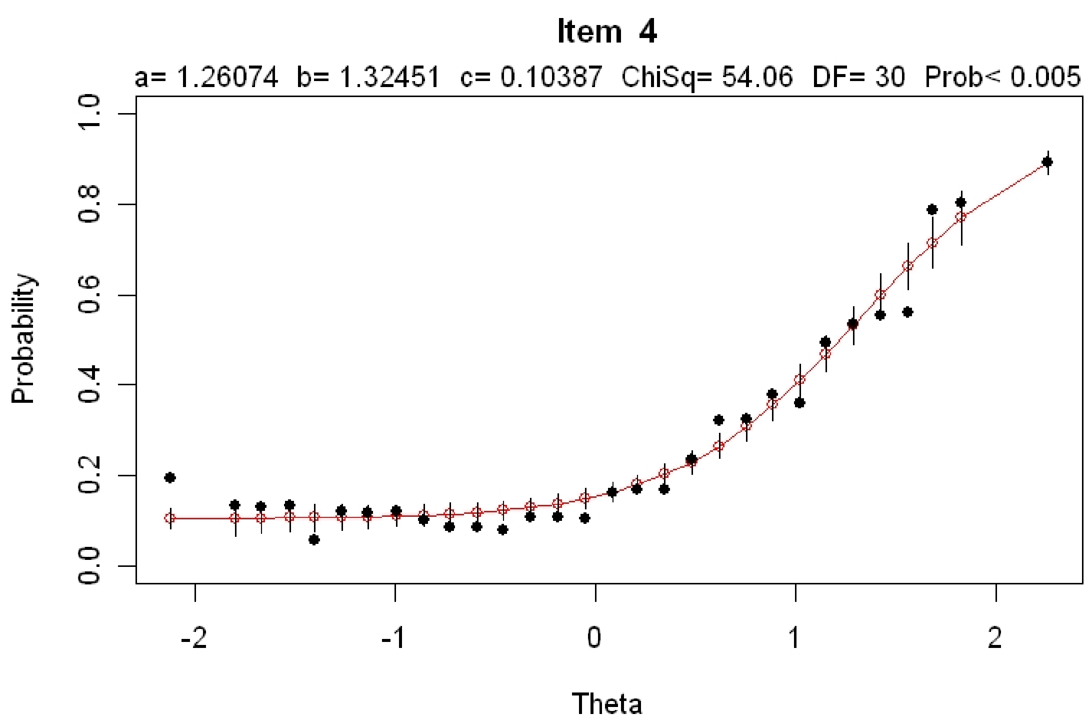
Appendix C

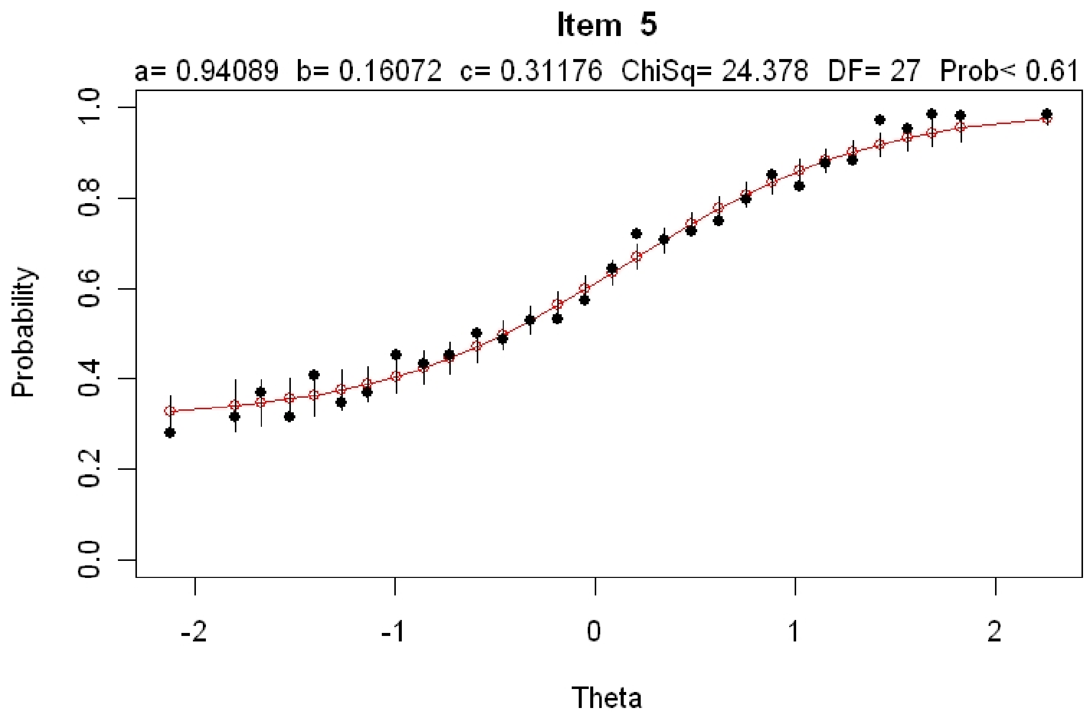
ICCs for Dichotomously Scored Items

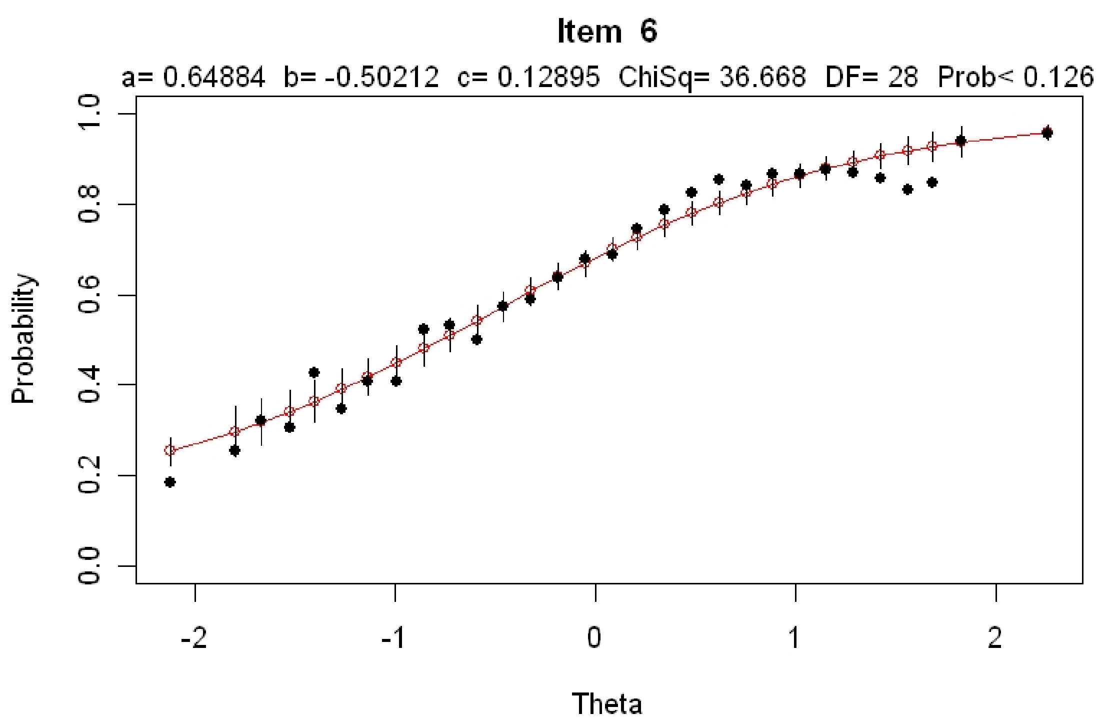


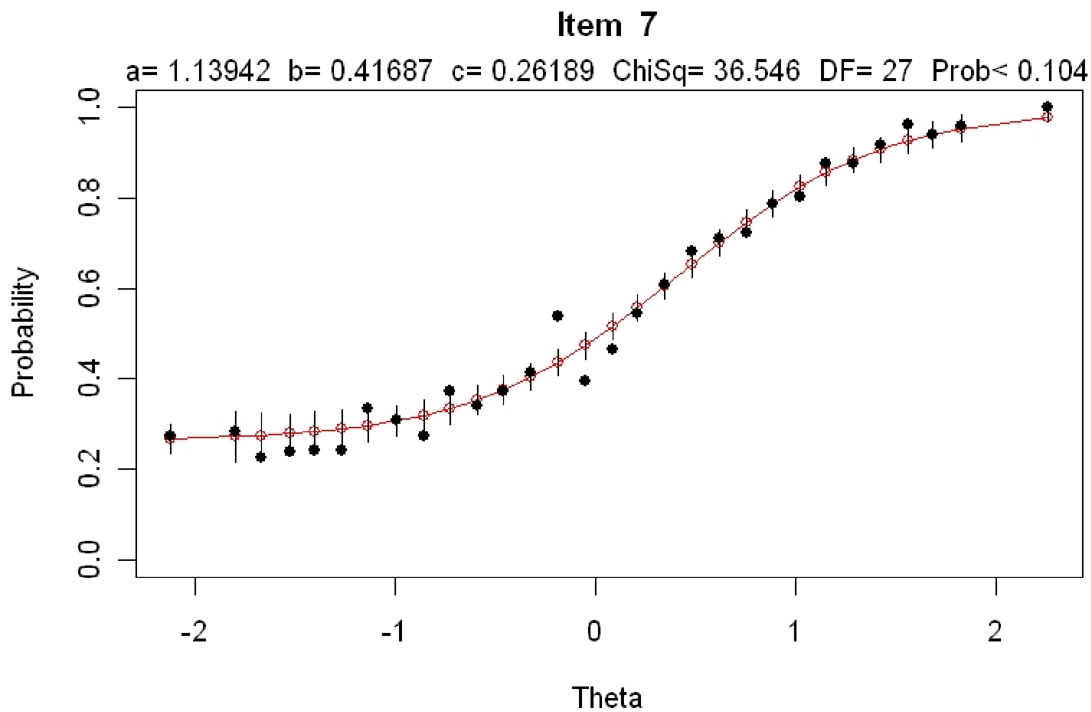


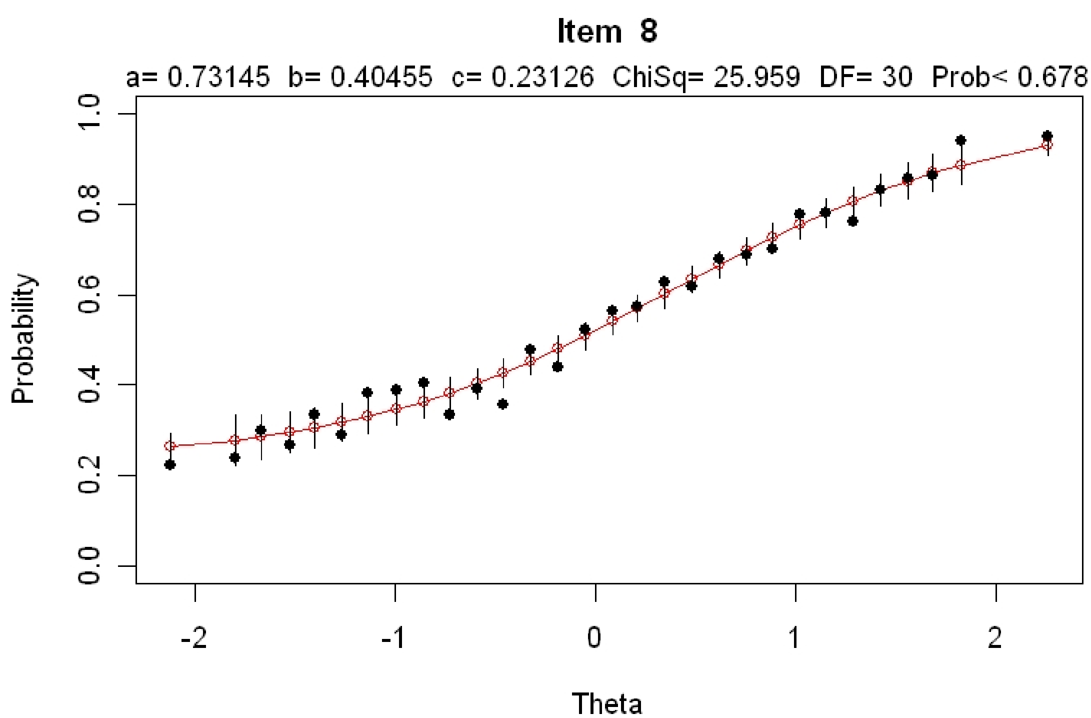

Item 3
a= 0.62459 b= 0.94561 c= 0.23654 ChiSq= 33.989 DF= 30 Prob< 0.281
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

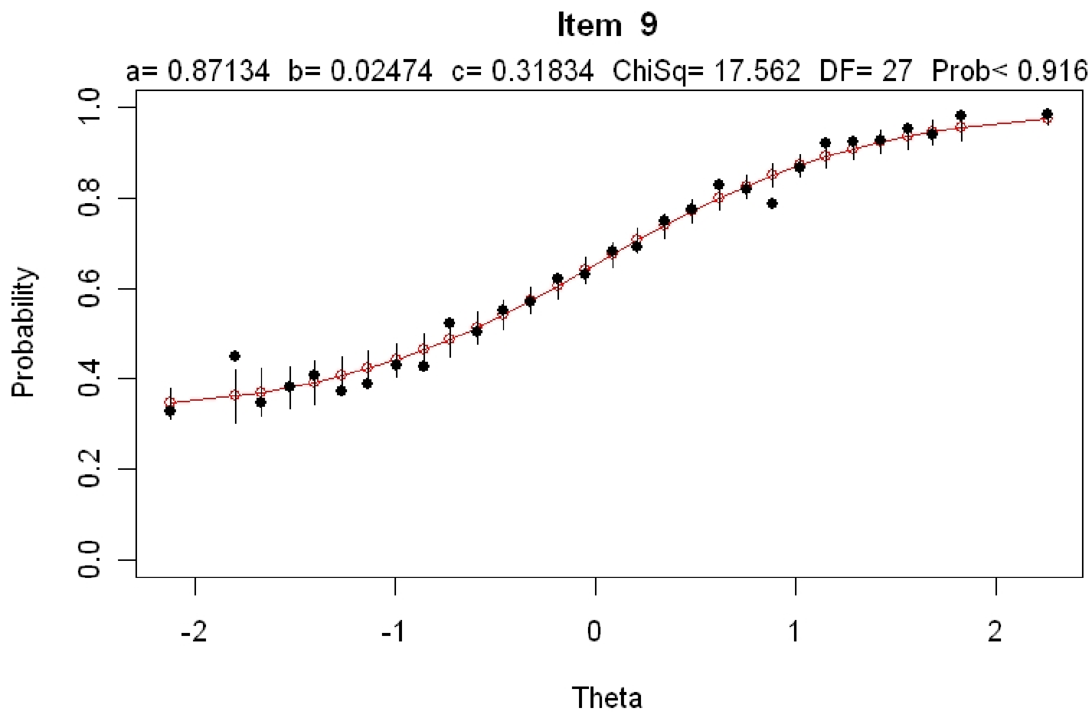

Item 4
a= 1.26074 b= 1.32451 c= 0.10387 ChiSq= 54.06 DF= 30 Prob< 0.005
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

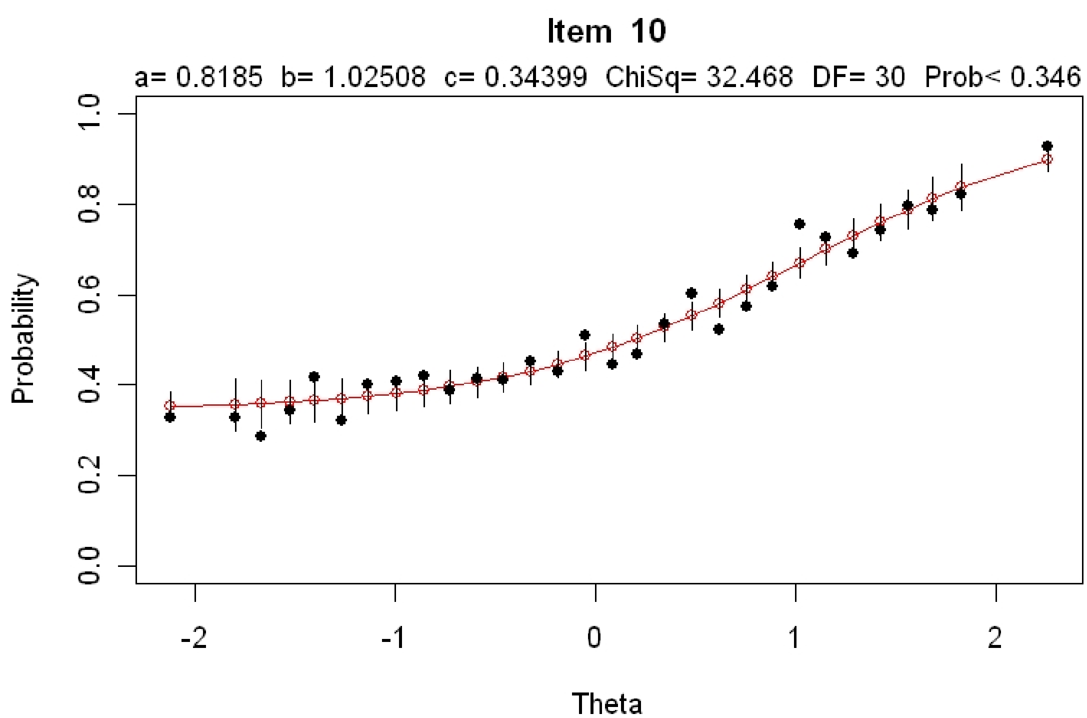

Item 5
a= 0.94089 b= 0.16072 c= 0.31176 ChiSq= 24.378 DF= 27 Prob< 0.61
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

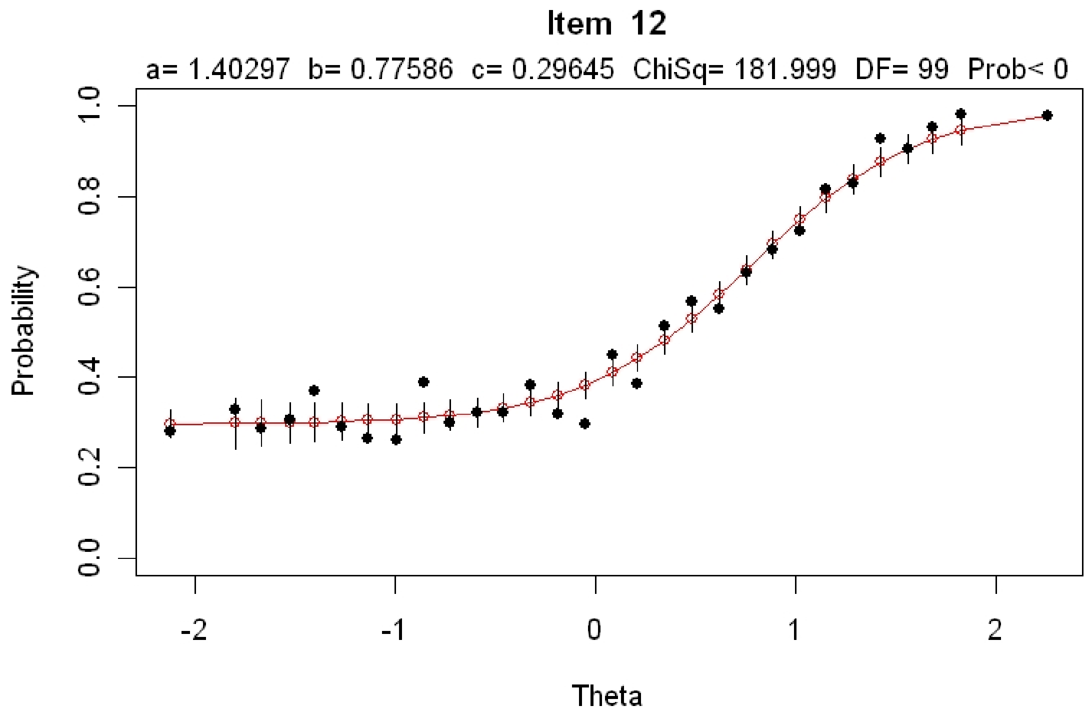

Item 6
a= 0.64884 b= -0.50212 c= 0.12895 ChiSq= 36.668 DF= 28 Prob< 0.126
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

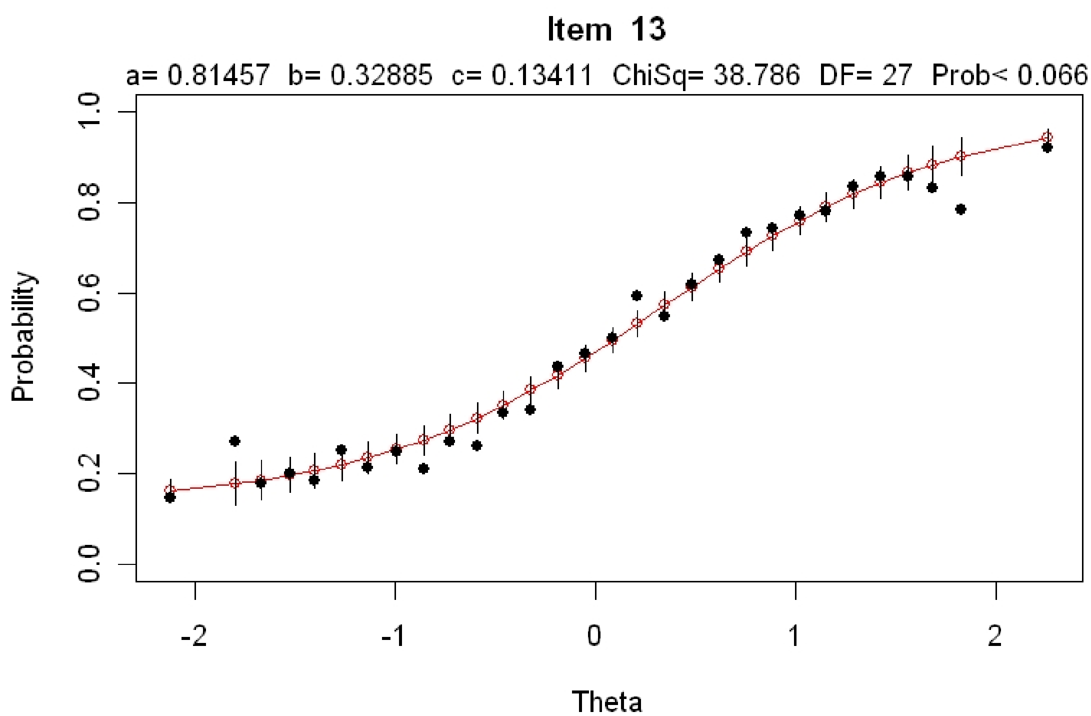

Item 7
a= 1.13942 b= 0.41687 c= 0.26189 ChiSq= 36.546 DF= 27 Prob< 0.104
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

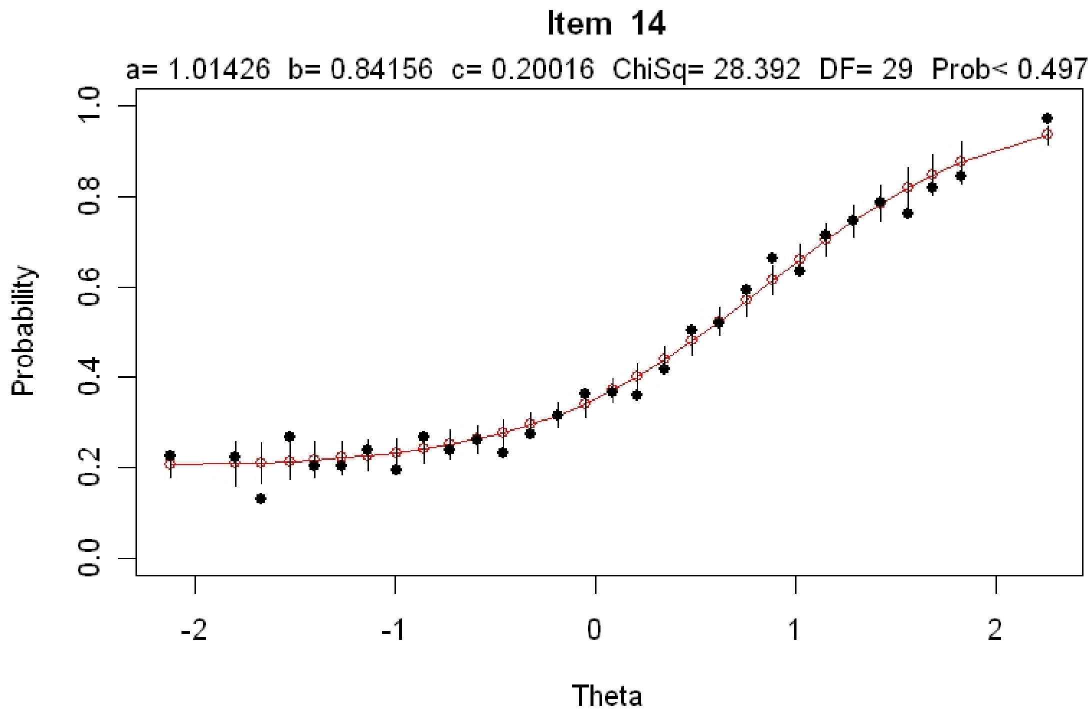

Item 8
a= 0.73145 b= 0.40455 c= 0.23126 ChiSq= 25.959 DF= 30 Prob< 0.678
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

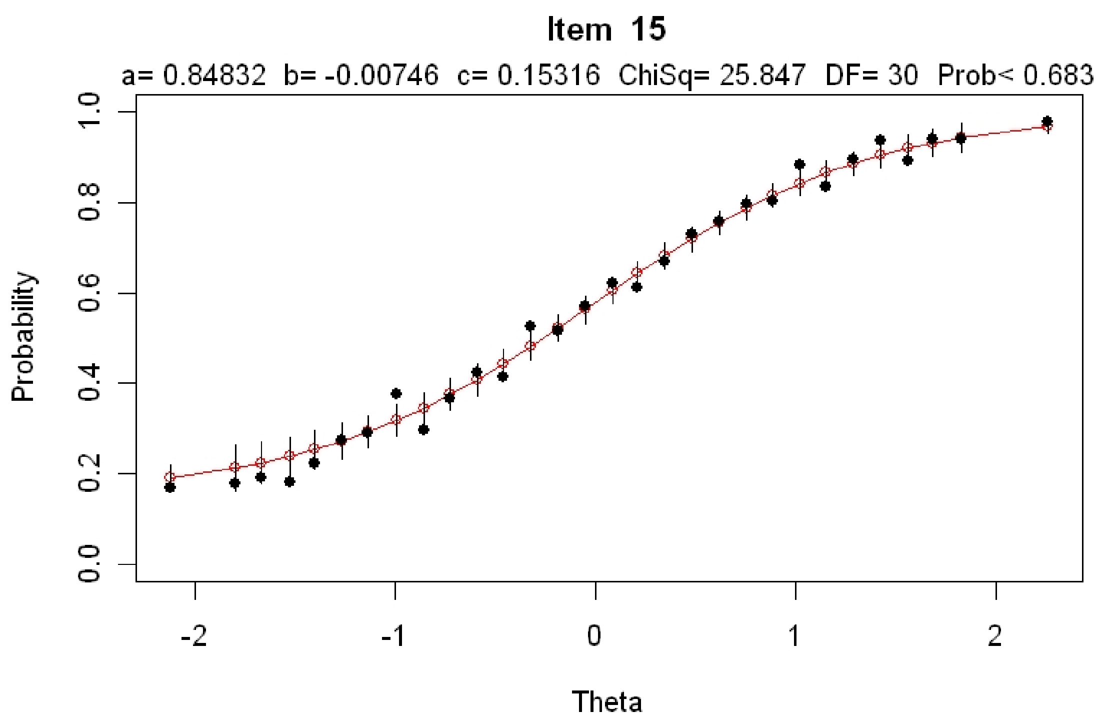

Item 9
a= 0.87134 b= 0.02474 c= 0.31834 ChiSq= 17.562 DF= 27 Prob< 0.916
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

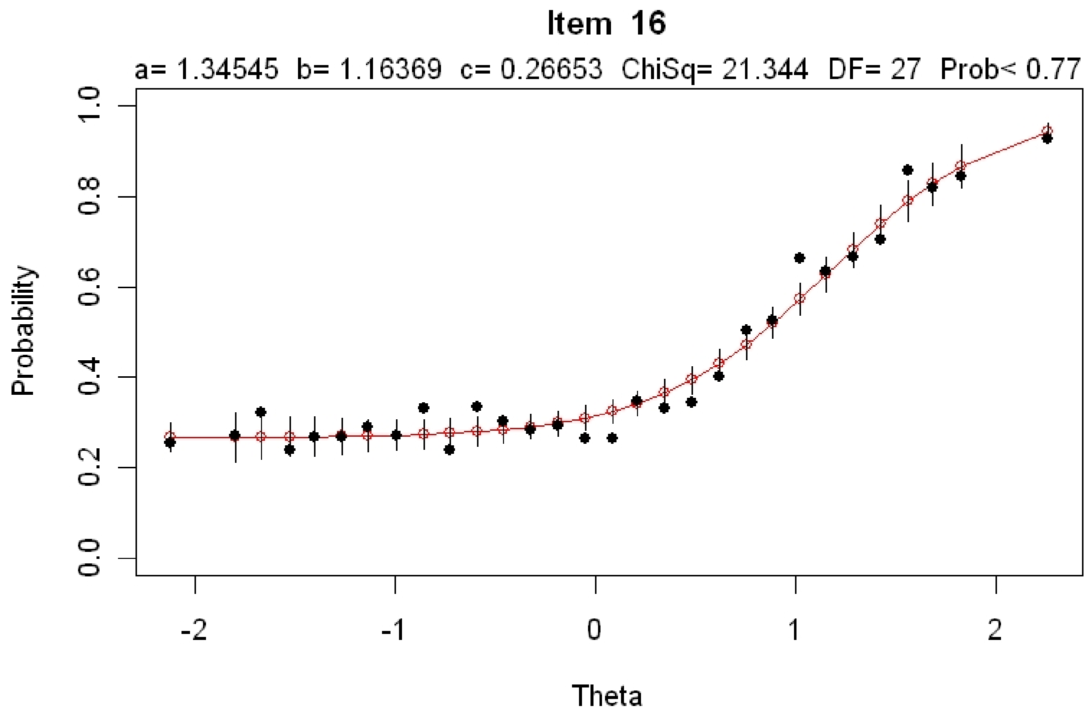

Item 10
a= 0.8185 b= 1.02508 c= 0.34399 ChiSq= 32.468 DF= 30 Prob< 0.346
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

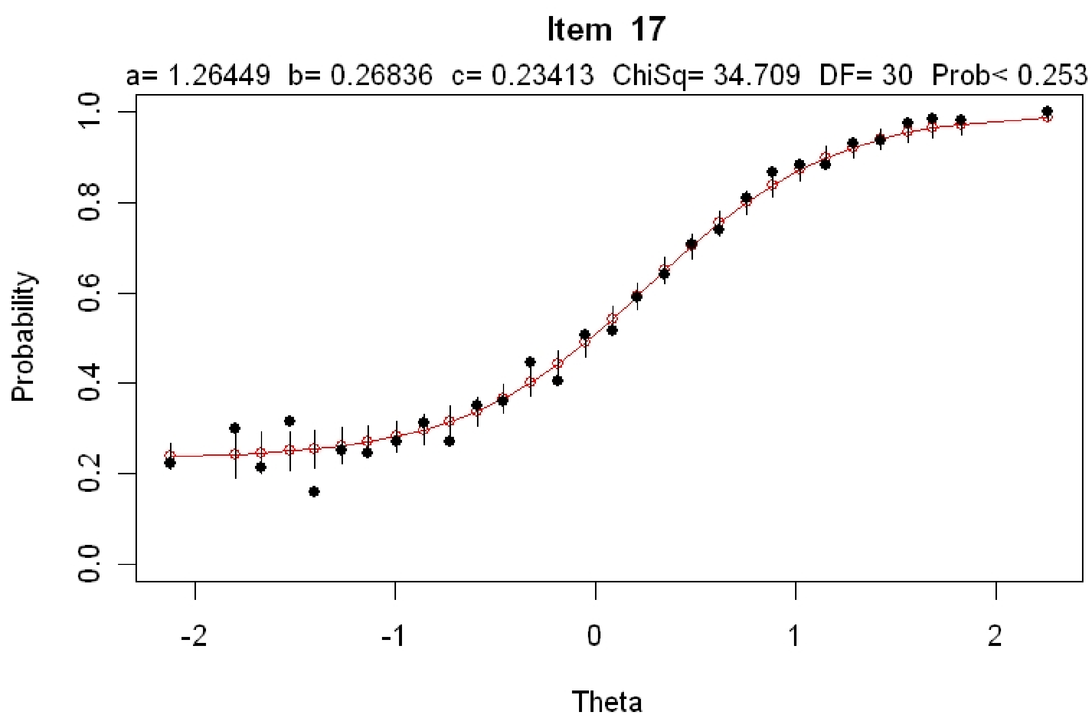

Item 12
a= 1.40297 b= 0.77586 c= 0.29645 ChiSq= 181.999 DF= 99 Prob< 0
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

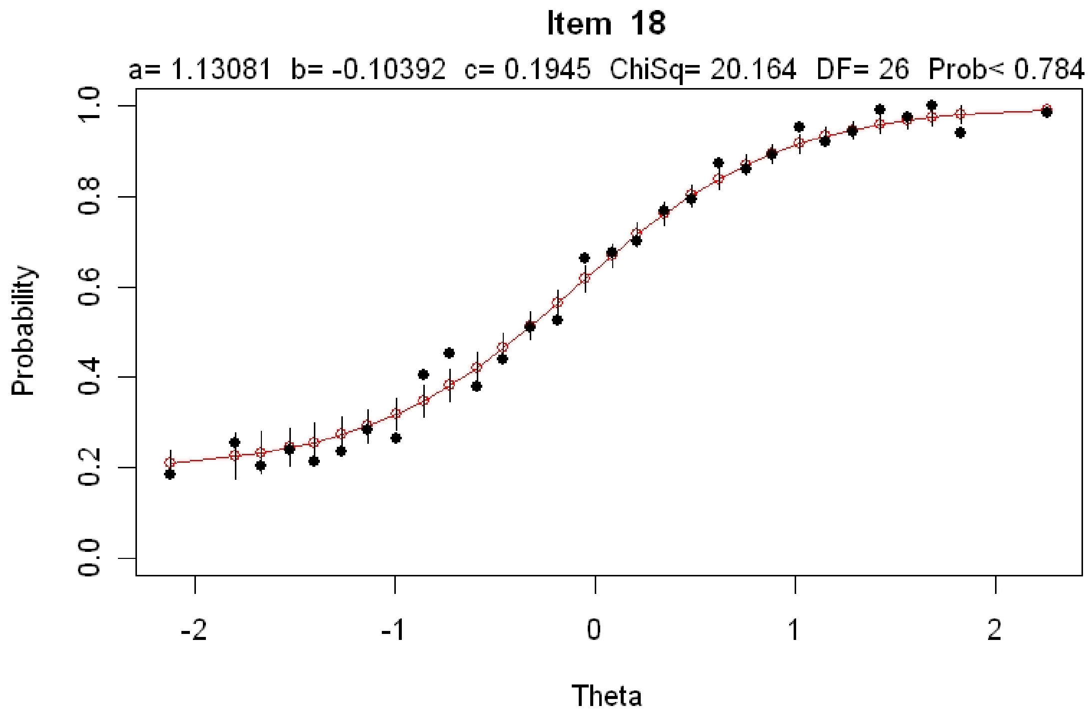

Item 13
a= 0.81457 b= 0.32885 c= 0.13411 ChiSq= 38.786 DF= 27 Prob< 0.066
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

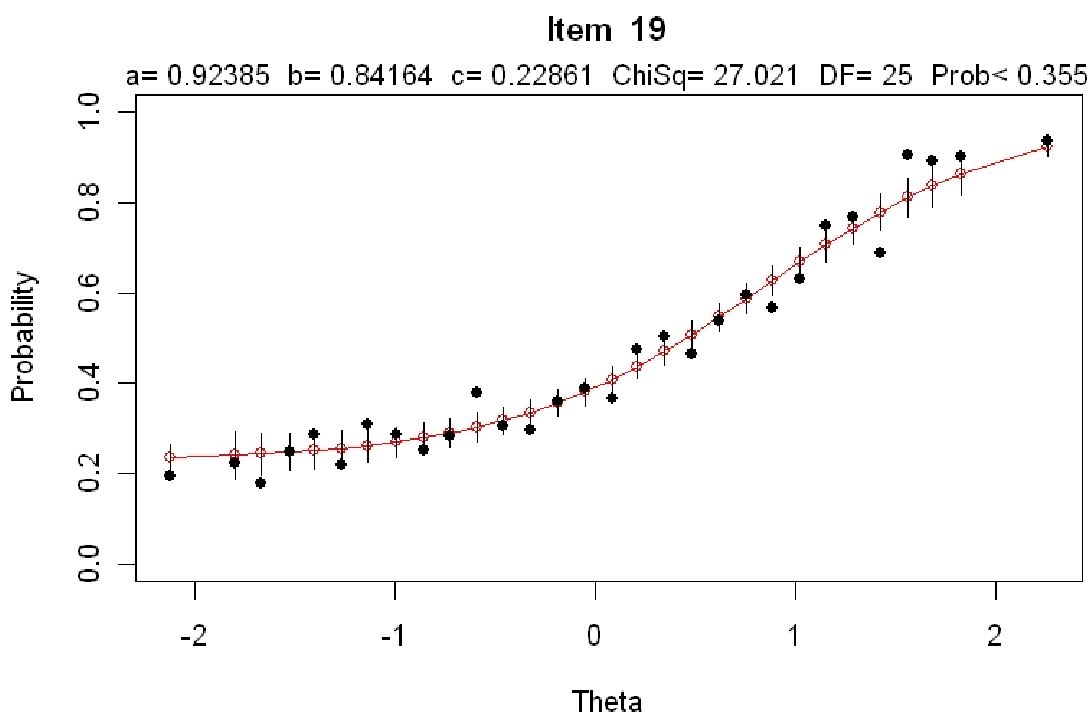

Item 14
a= 1.01426 b= 0.84156 c= 0.20016 ChiSq= 28.392 DF= 29 Prob< 0.497
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

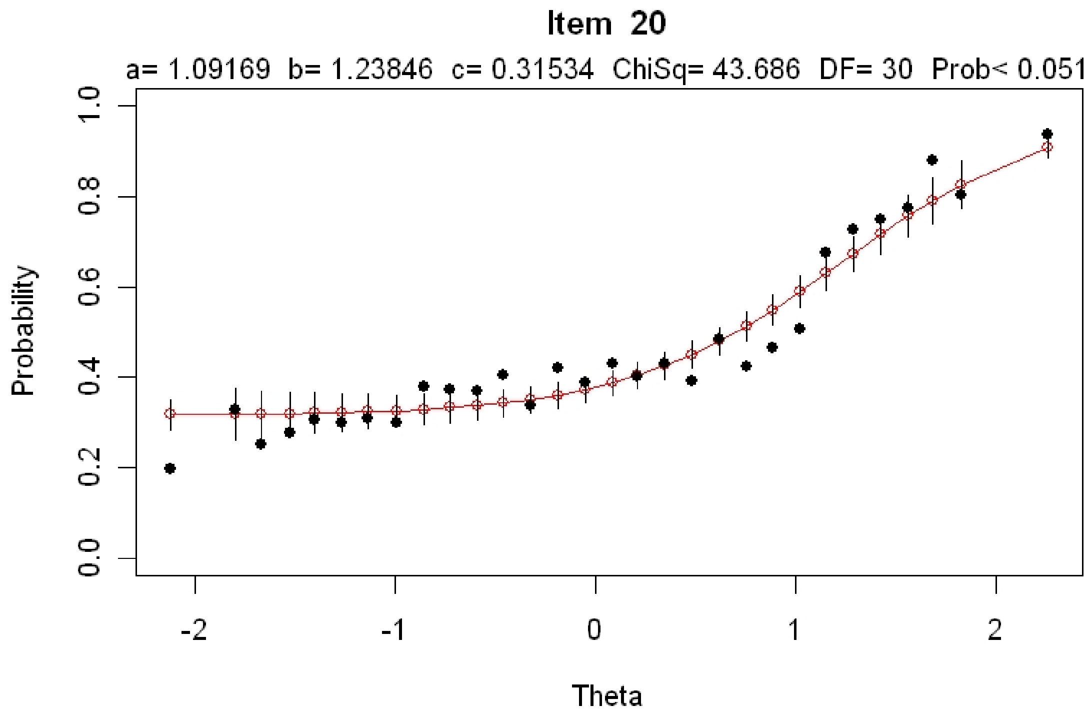

Item 15
a= 0.84832 b= -0.00746 c= 0.15316 ChiSq= 25.847 DF= 30 Prob< 0.683
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta


Item 16
a= 1.34545 b= 1.16369 c= 0.26653 ChiSq= 21.344 DF= 27 Prob< 0.77
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

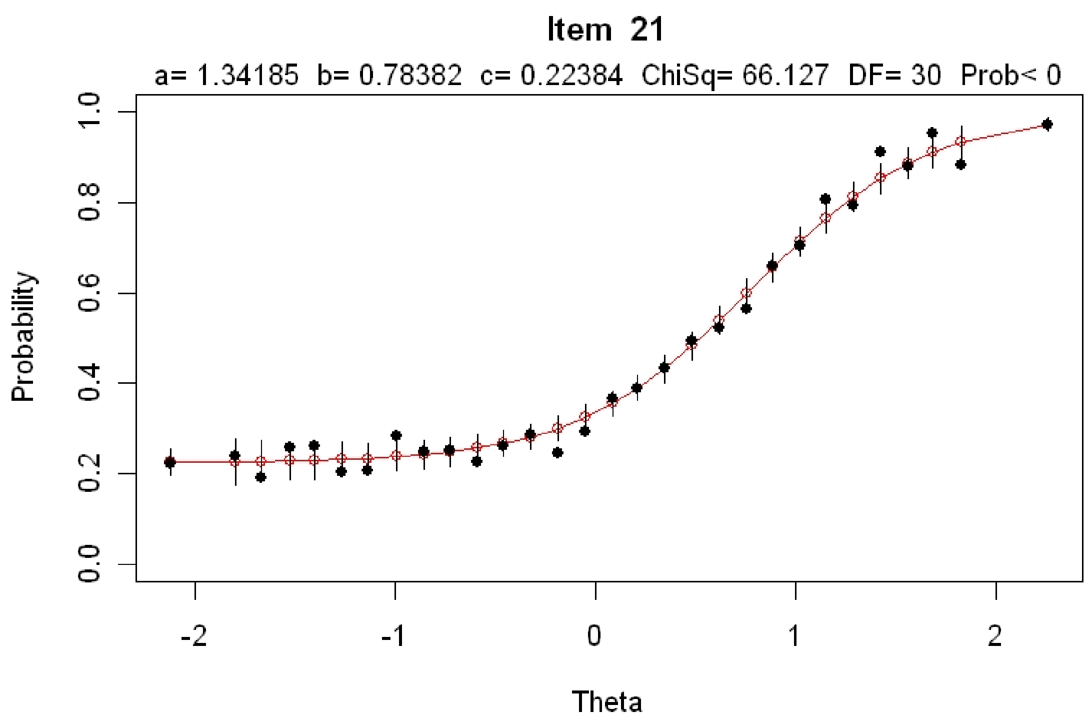

Item 17
a= 1.26449 b= 0.26836 c= 0.23413 ChiSq= 34.709 DF= 30 Prob< 0.253
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta


Item 18
a= 1.13081 b= -0.10392 c= 0.1945 ChiSq= 20.164 DF= 26 Prob< 0.784
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

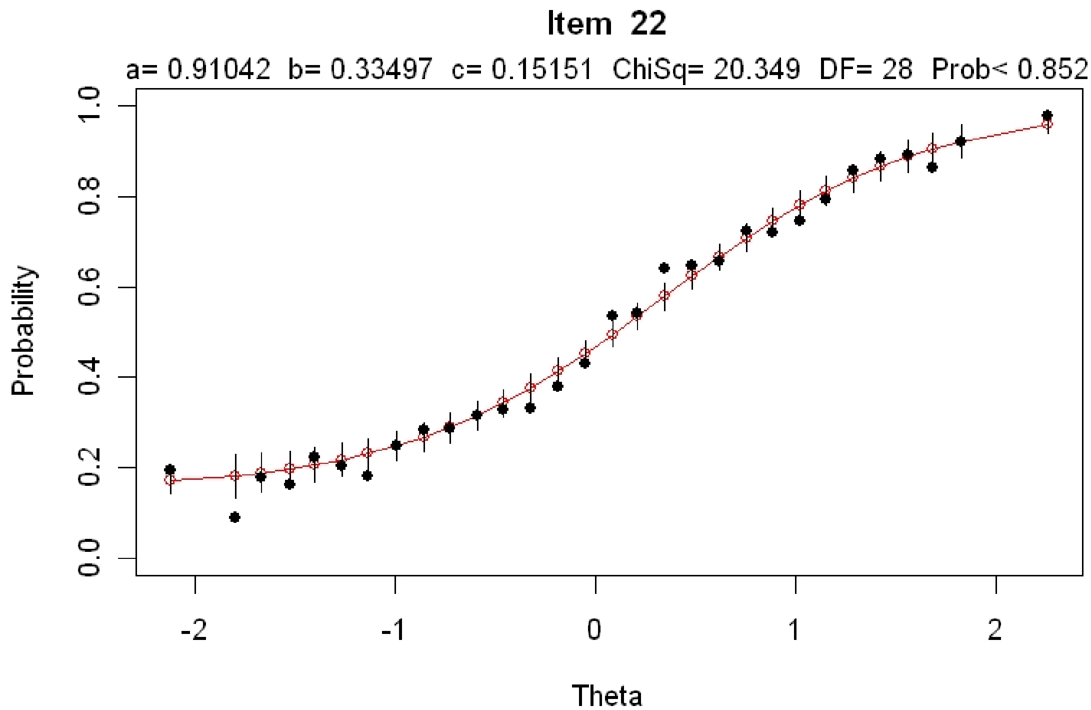

Item 19
a= 0.92385 b= 0.84164 c= 0.22861 ChiSq= 27.021 DF= 25 Prob< 0.355
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

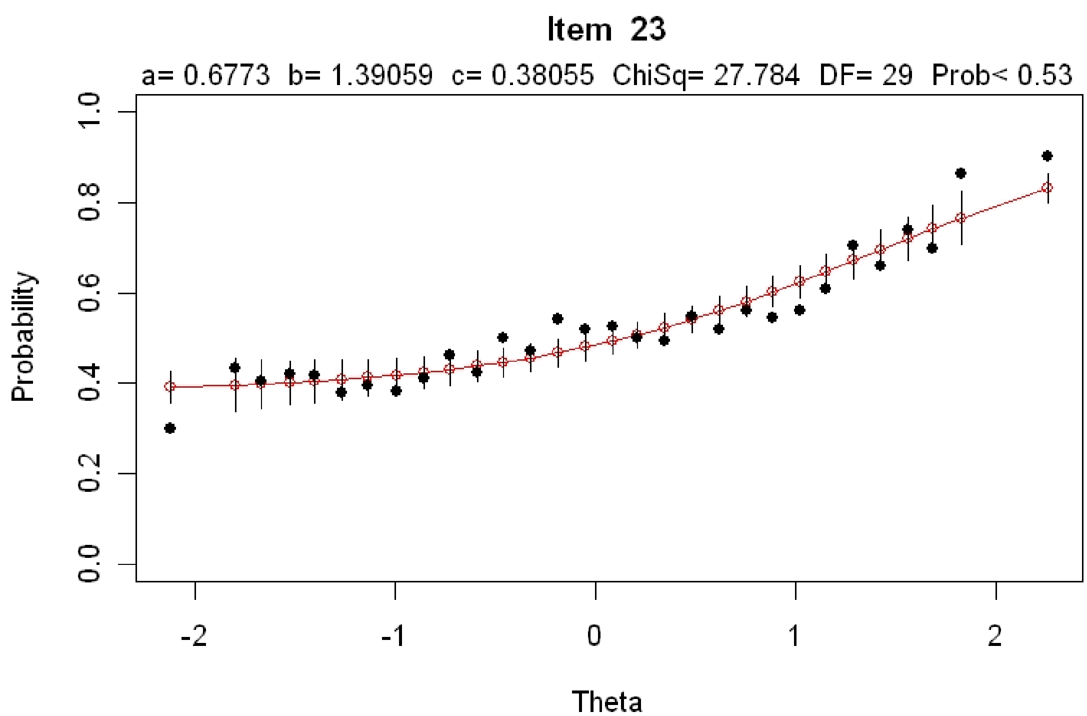


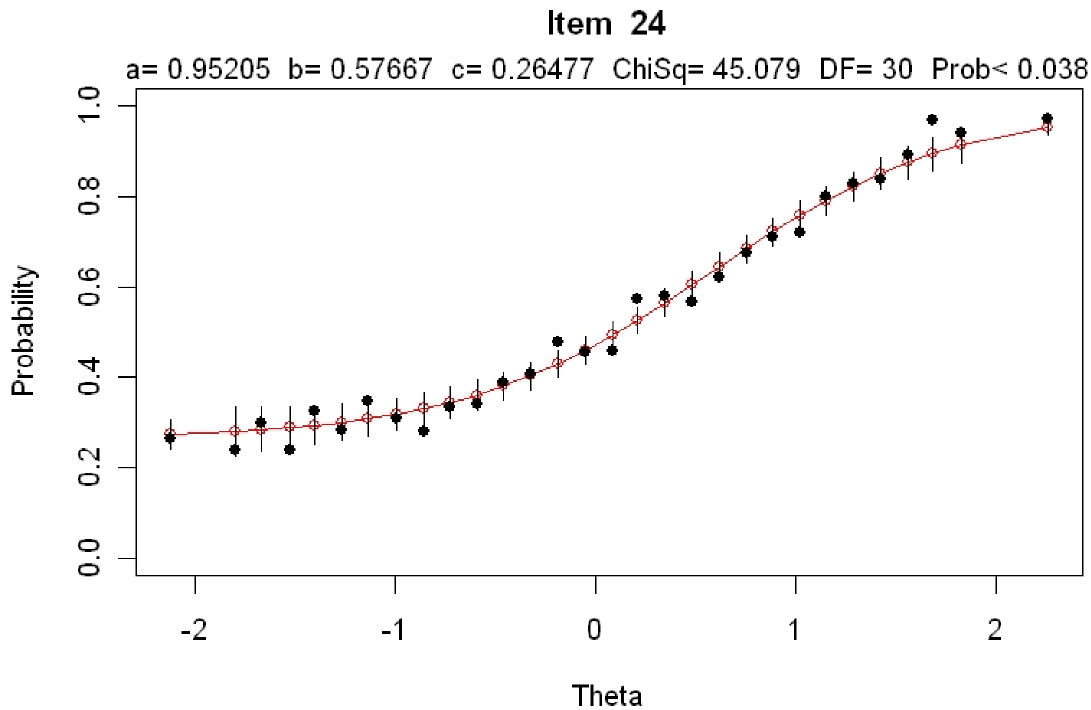
Item 20
a= 1.09169 b= 1.23846 c= 0.31534 ChiSq= 43.686 DF= 30 Prob< 0.051
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

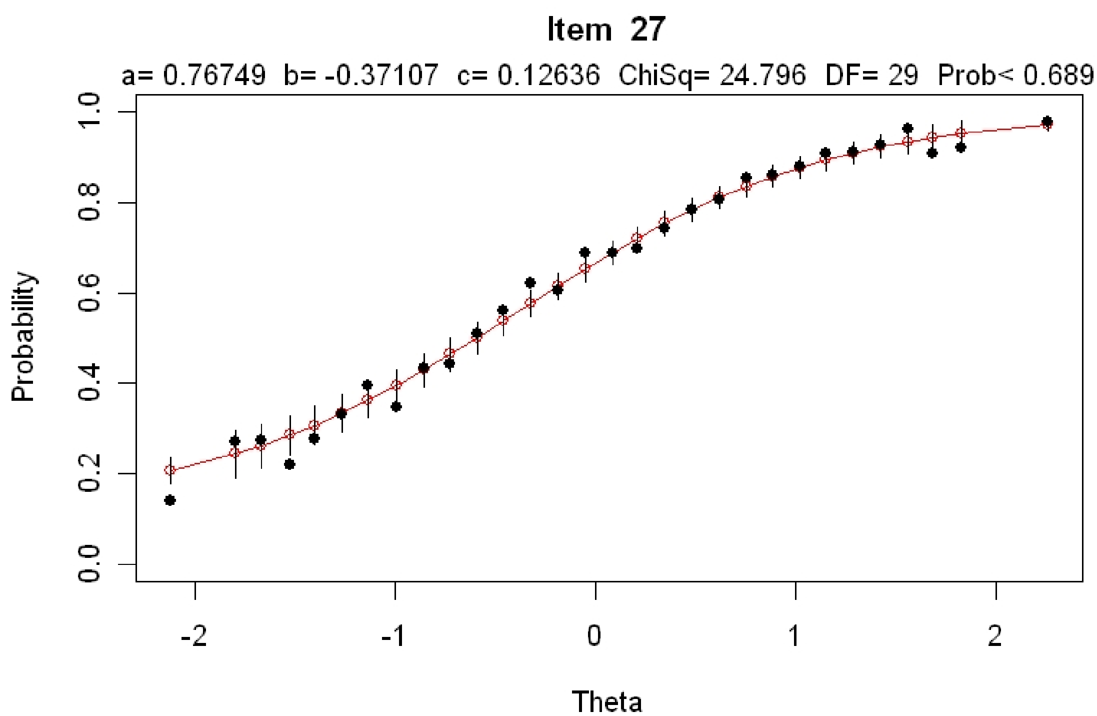


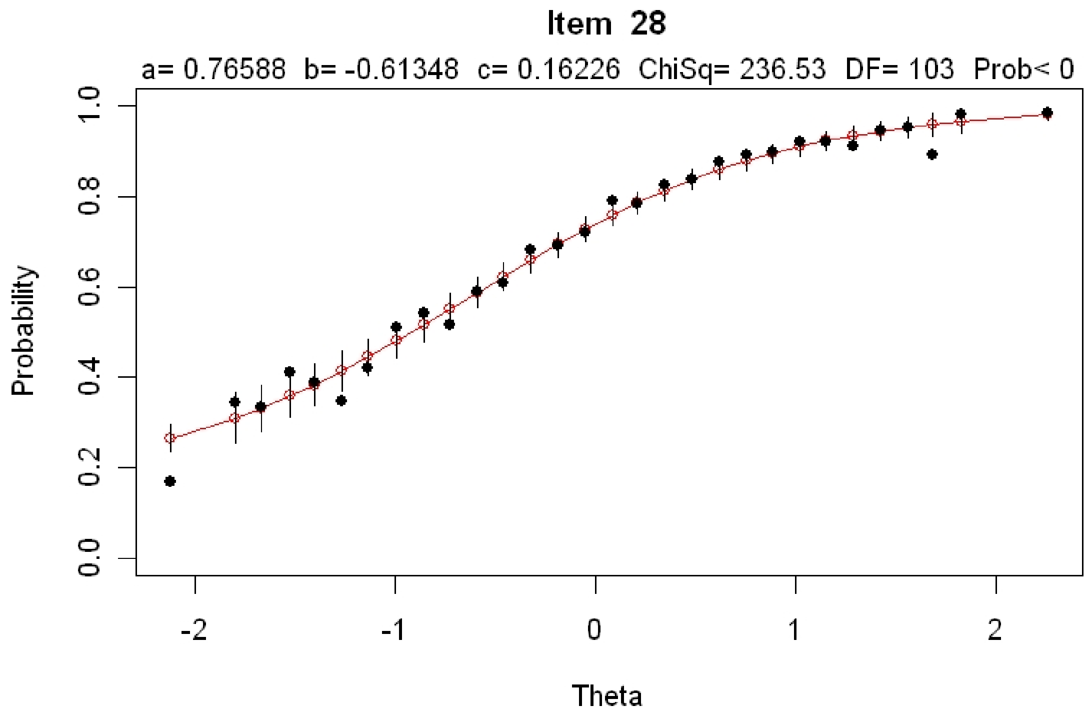
Item 21
a= 1.34185 b= 0.78382 c= 0.22384 ChiSq= 66.127 DF= 30 Prob< 0
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

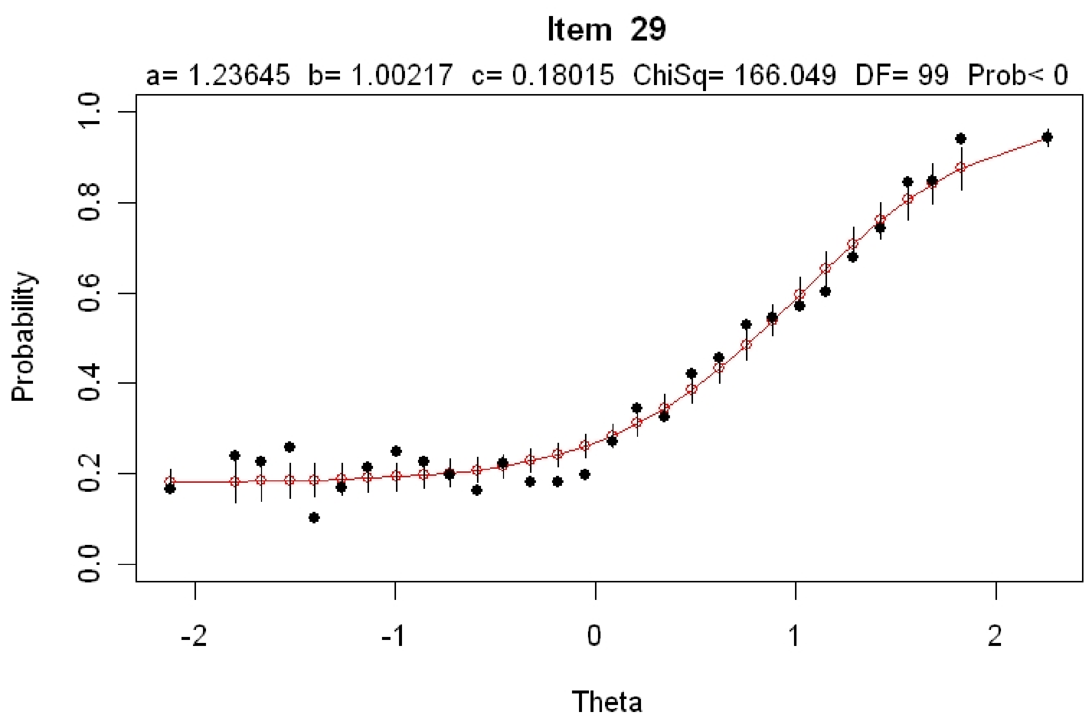

Item 22
a= 0.91042 b= 0.33497 c= 0.15151 ChiSq= 20.349 DF= 28 Prob< 0.852
Probability
Theta
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2

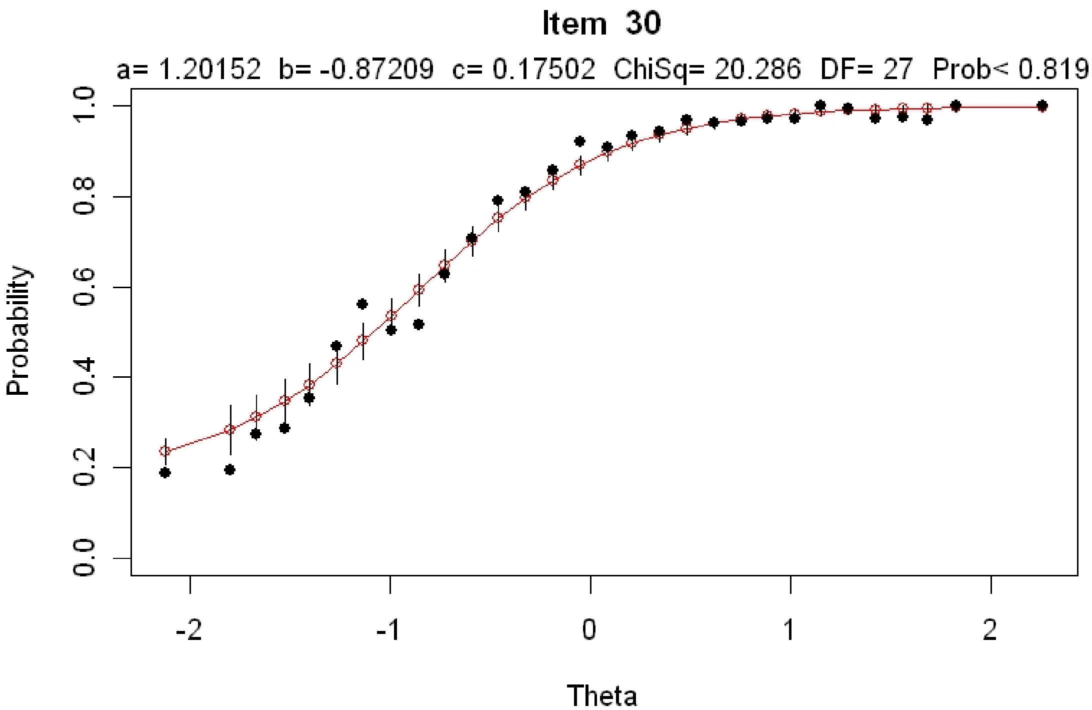

Item 23
a= 0.6773 b= 1.39059 c= 0.38055 ChiSq= 27.784 DF= 29 Prob< 0.53
Probability
Theta
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2

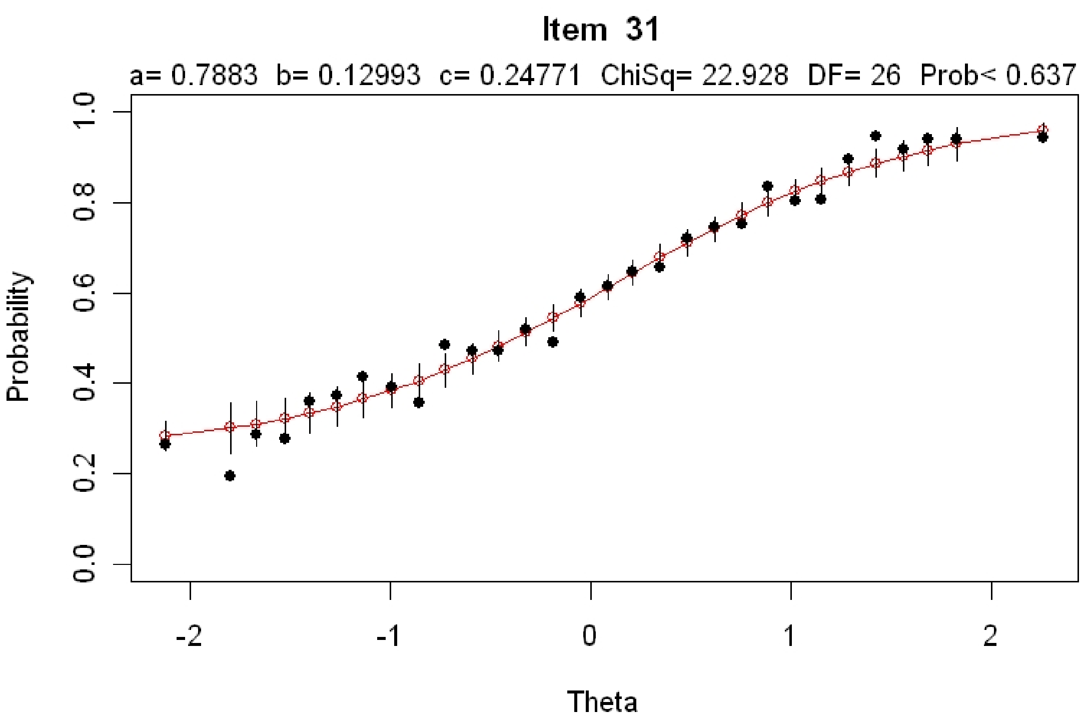

Item 24
a= 0.95205 b= 0.57667 c= 0.26477 ChiSq= 45.079 DF= 30 Prob< 0.038
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

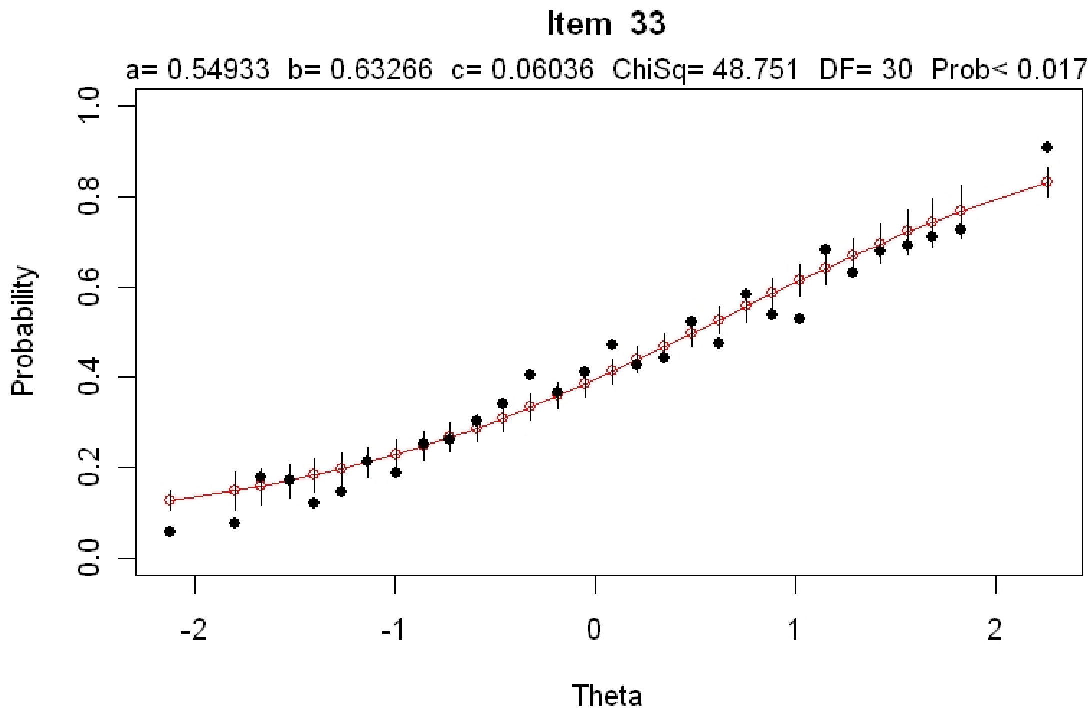

Item 27
a= 0.76749 b= -0.37107 c= 0.12636 ChiSq= 24.796 DF= 29 Prob< 0.689
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

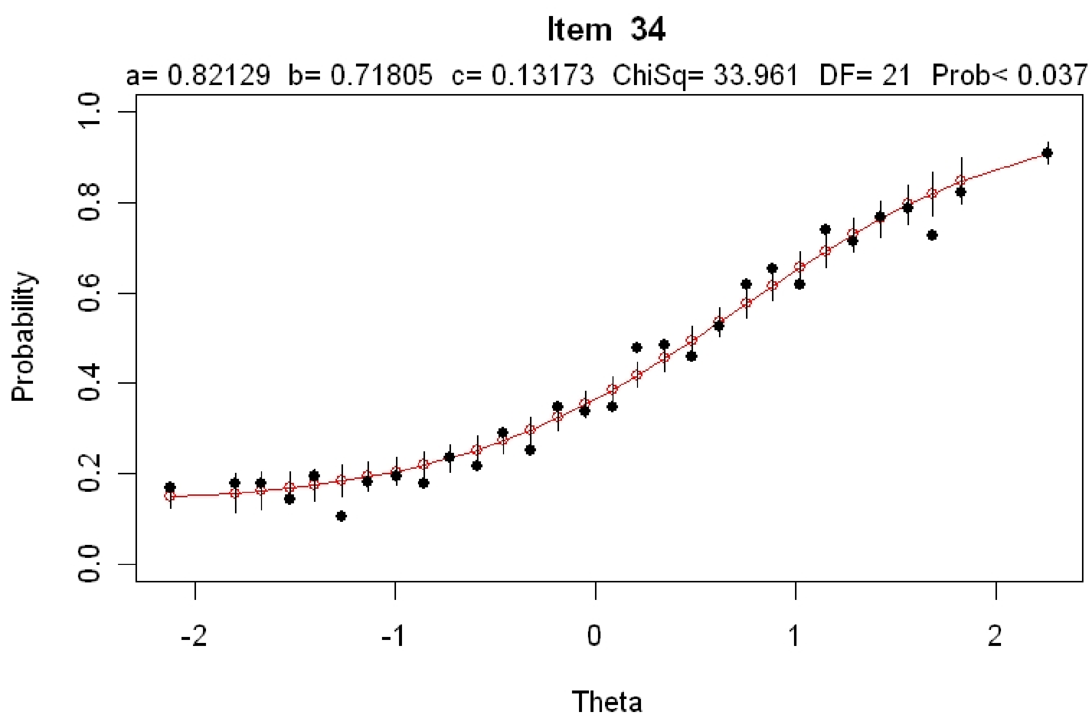

Item 28
a= 0.76588 b= -0.61348 c= 0.16226 ChiSq= 236.53 DF= 103 Prob< 0
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

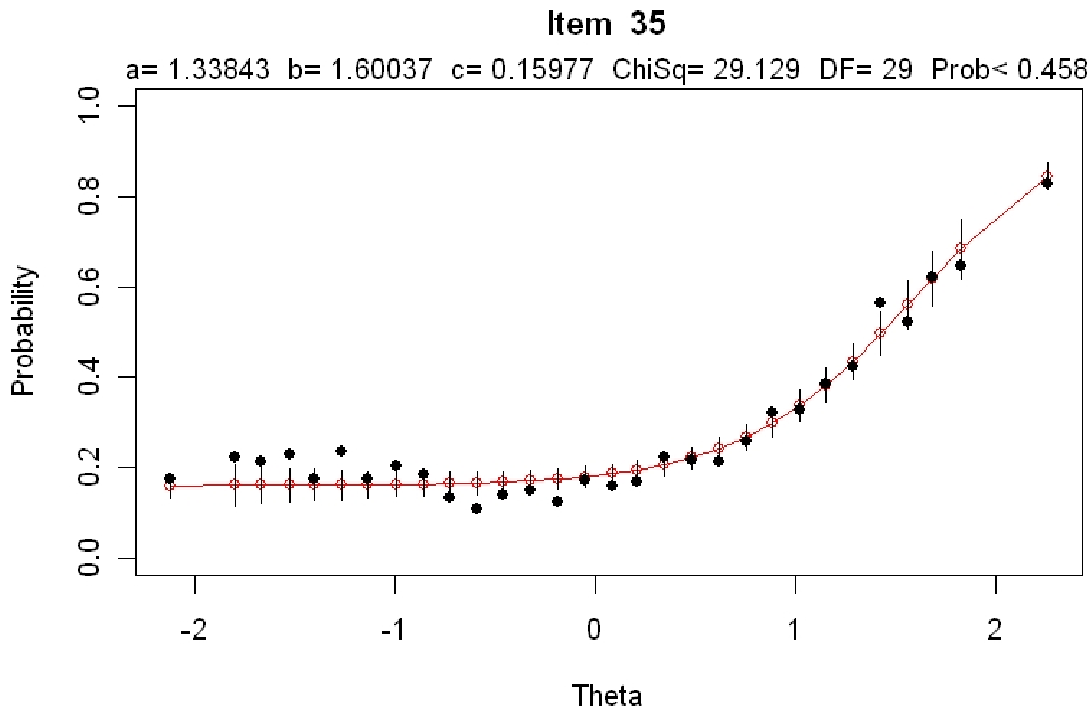

Item 29
a= 1.23645 b= 1.00217 c= 0.18015 ChiSq= 166.049 DF= 99 Prob< 0
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

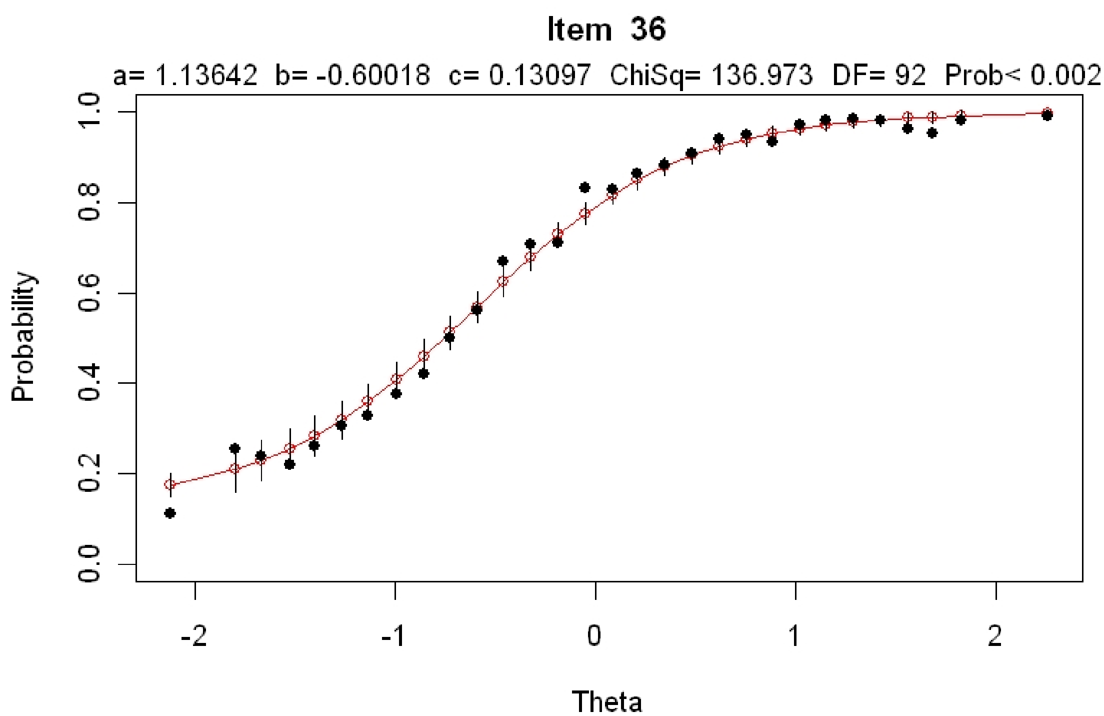

Item 30
a= 1.20152 b= -0.87209 c= 0.17502 ChiSq= 20.286 DF= 27 Prob< 0.819
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

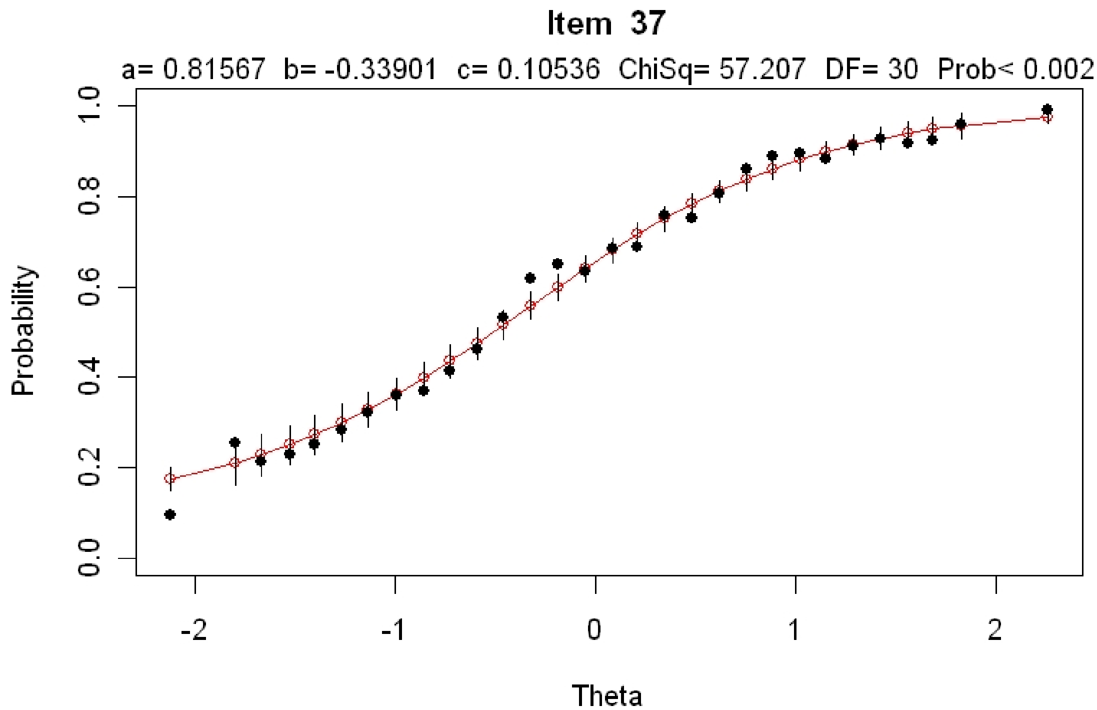

Item 31
a= 0.7883 b= 0.12993 c= 0.24771 ChiSq= 22.928 DF= 26 Prob< 0.637
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

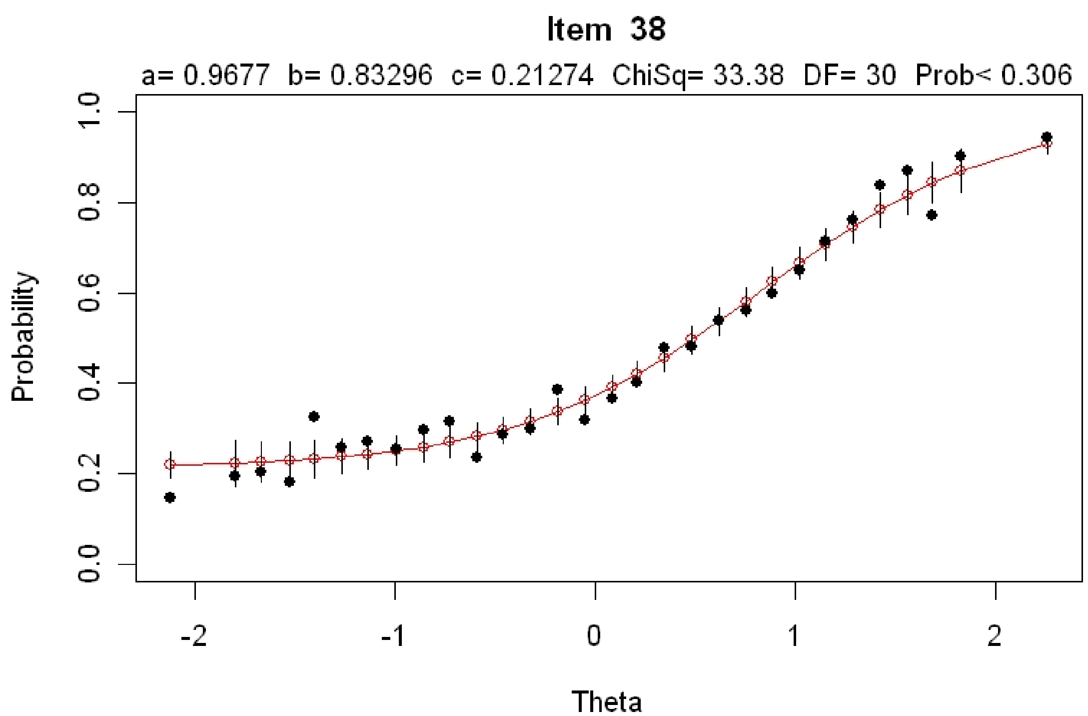

Item 33
a= 0.54933 b= 0.63266 c= 0.06036 ChiSq= 48.751 DF= 30 Prob< 0.017
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

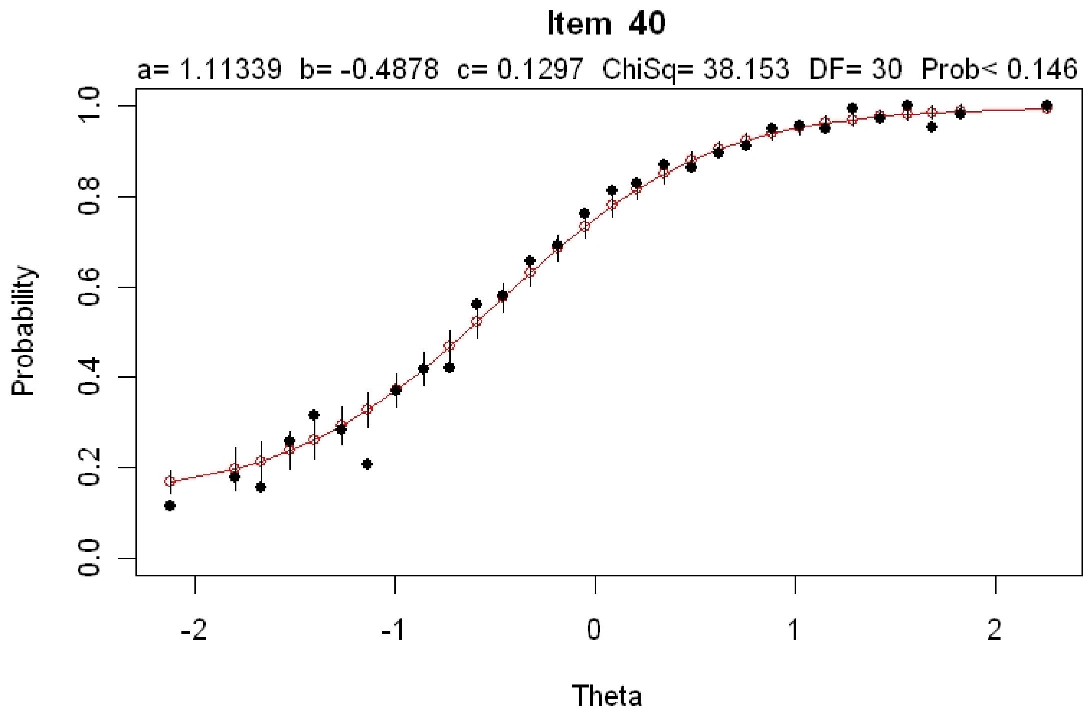

Item 34
a= 0.82129 b= 0.71805 c= 0.13173 ChiSq= 33.961 DF= 21 Prob< 0.037
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

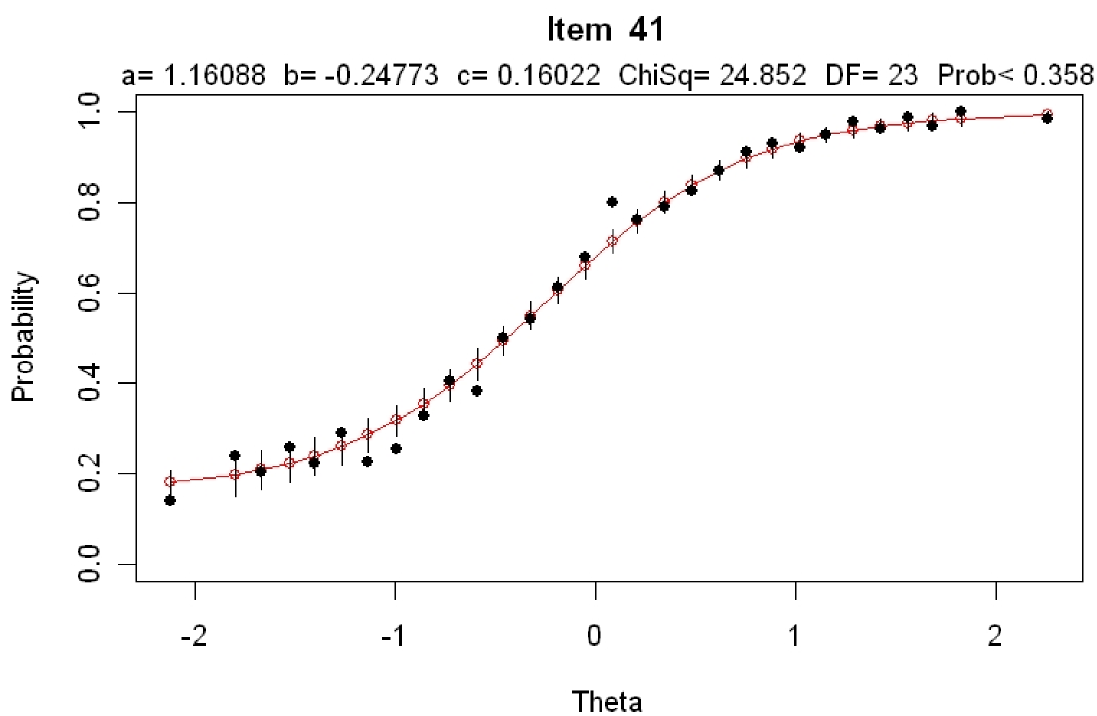

Item 35
a= 1.33843 b= 1.60037 c= 0.15977 ChiSq= 29.129 DF= 29 Prob< 0.458
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

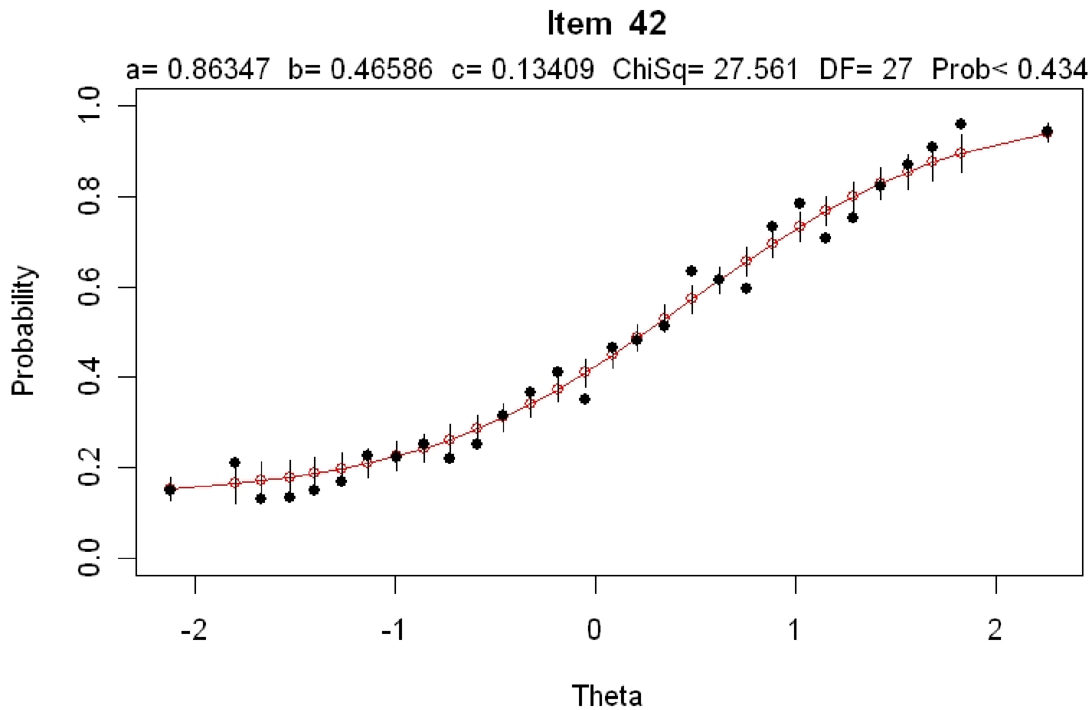

Item 36
a= 1.13642 b= -0.60018 c= 0.13097 ChiSq= 136.973 DF= 92 Prob< 0.002
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

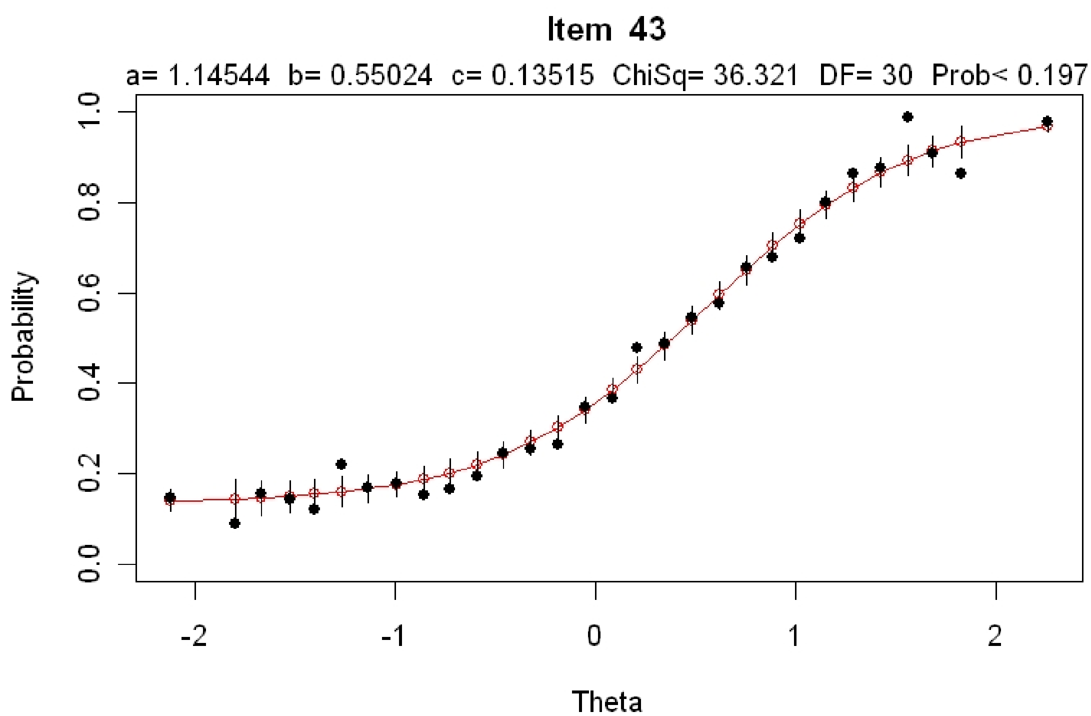

Item 37
a= 0.81567 b= -0.33901 c= 0.10536 ChiSq= 57.207 DF= 30 Prob< 0.002
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

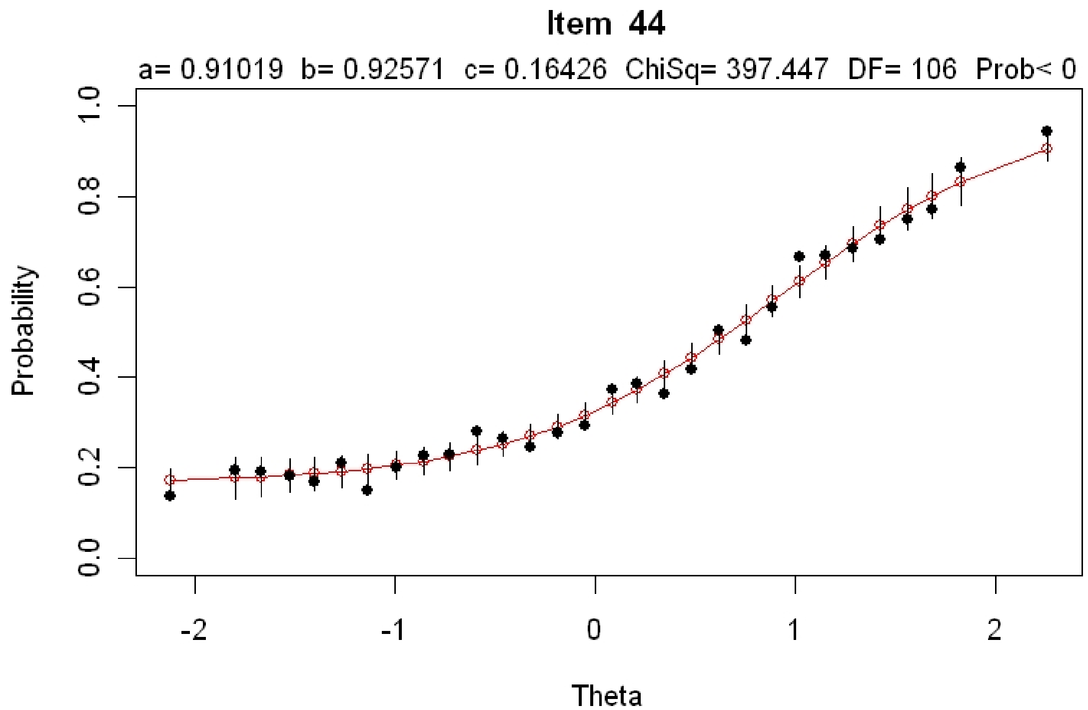

Item 38
a= 0.9677 b= 0.83296 c= 0.21274 ChiSq= 33.38 DF= 30 Prob< 0.306
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

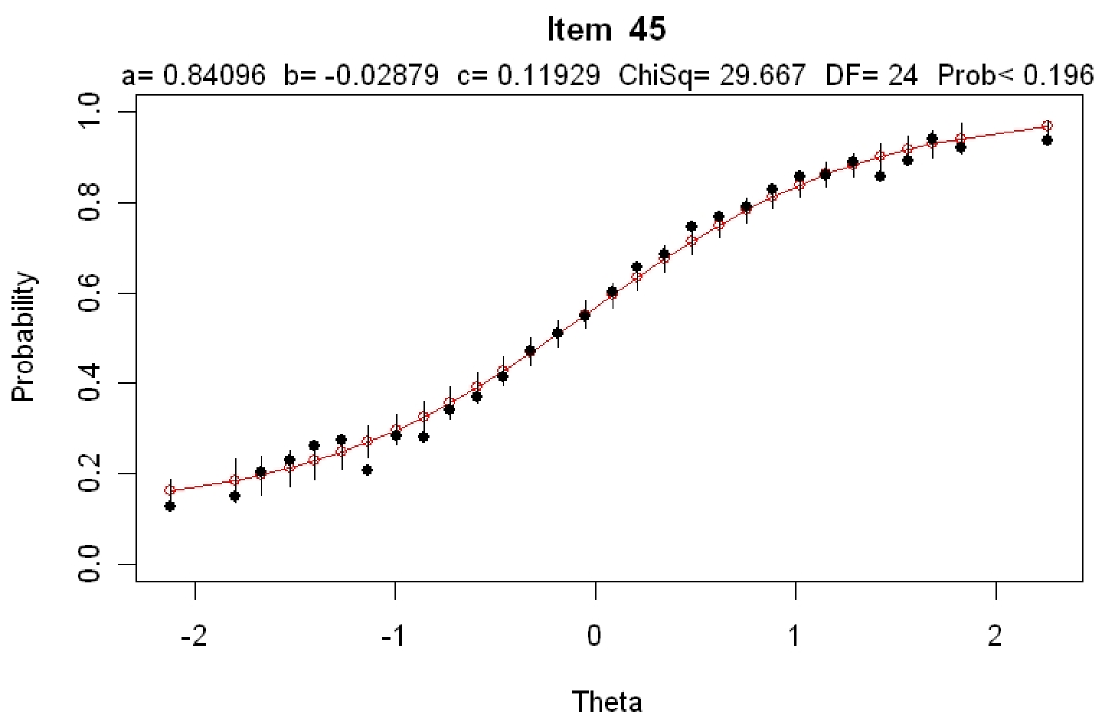

Item 40
a= 1.11339 b= -0.4878 c= 0.1297 ChiSq= 38.153 DF= 30 Prob< 0.146
Probability
Theta
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2


Item 41
a= 1.16088 b= -0.24773 c= 0.16022 ChiSq= 24.852 DF= 23 Prob< 0.358
Probability
Theta
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2

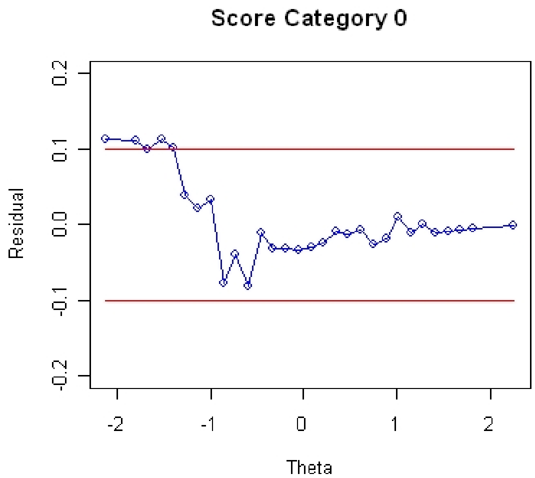

Item 42
a= 0.86347 b= 0.46586 c= 0.13409 ChiSq= 27.561 DF= 27 Prob< 0.434
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

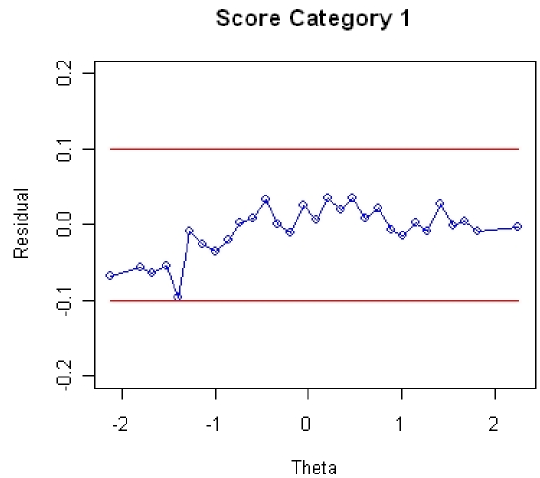

Item 43
a= 1.14544 b= 0.55024 c= 0.13515 ChiSq= 36.321 DF= 30 Prob< 0.197
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

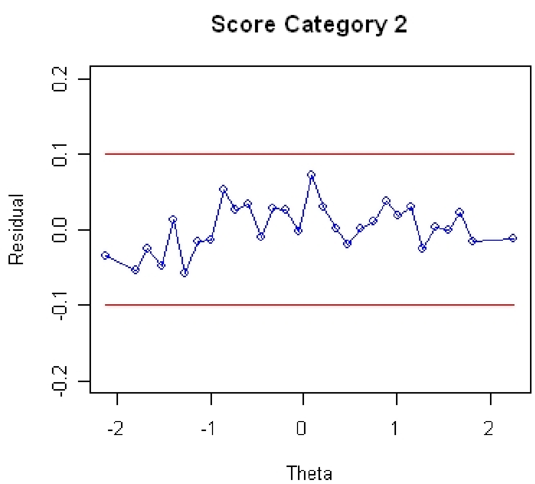

Item 44
a= 0.91019 b= 0.92571 c= 0.16426 ChiSq= 397.447 DF= 106 Prob< 0
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

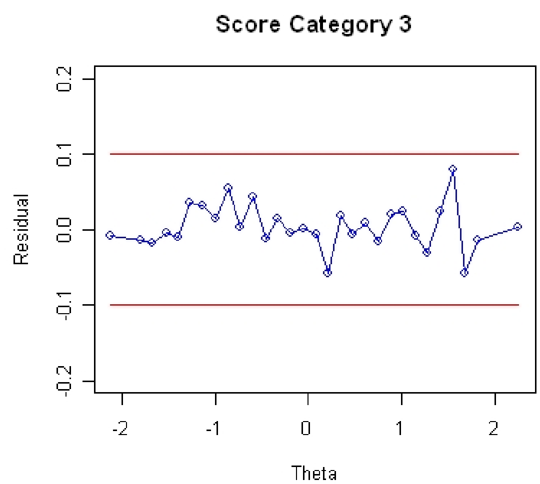

Item 45
a= 0.84096 b= -0.02879 c= 0.11929 ChiSq= 29.667 DF= 24 Prob< 0.196
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

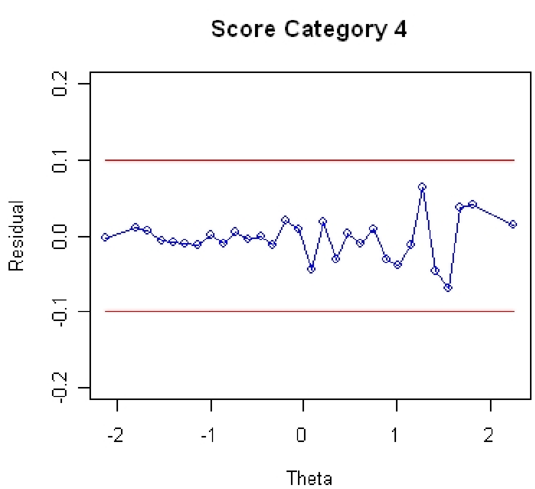
Appendix D
Residuals and ICCs for Polytomously Scored Items
Item 11 Residuals



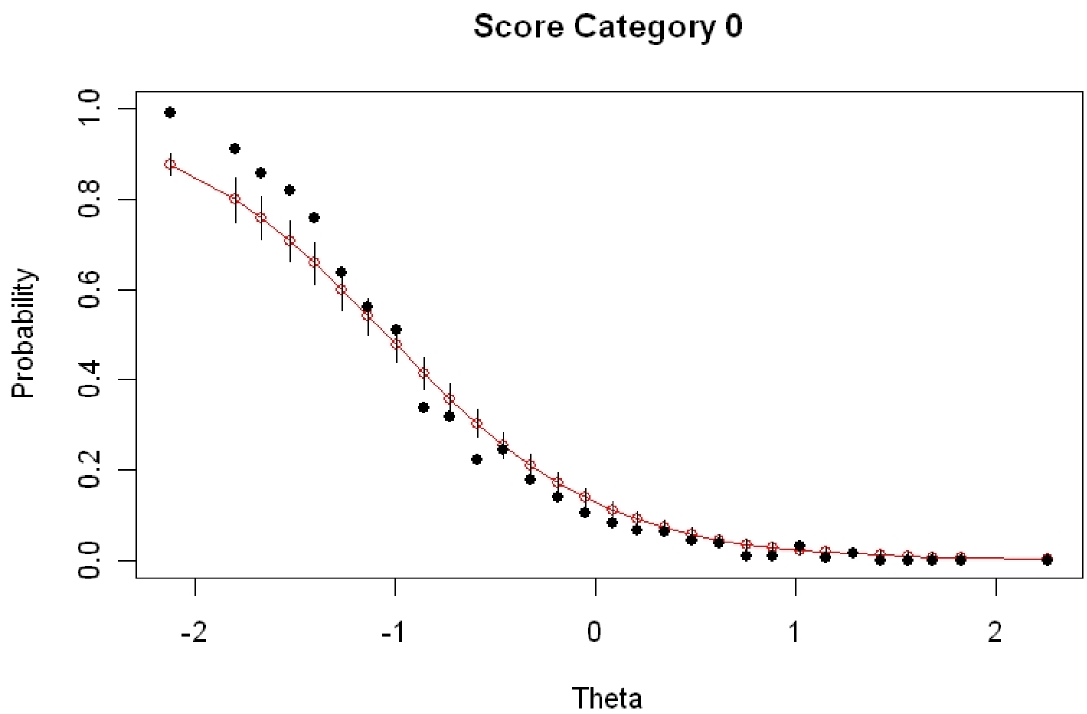


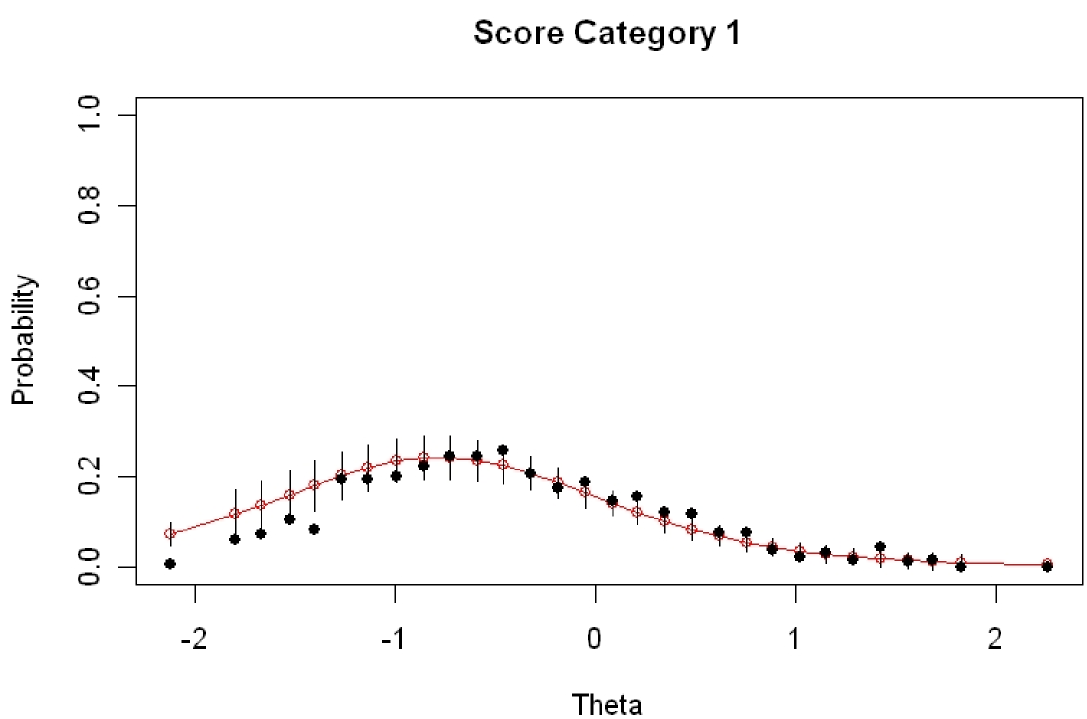


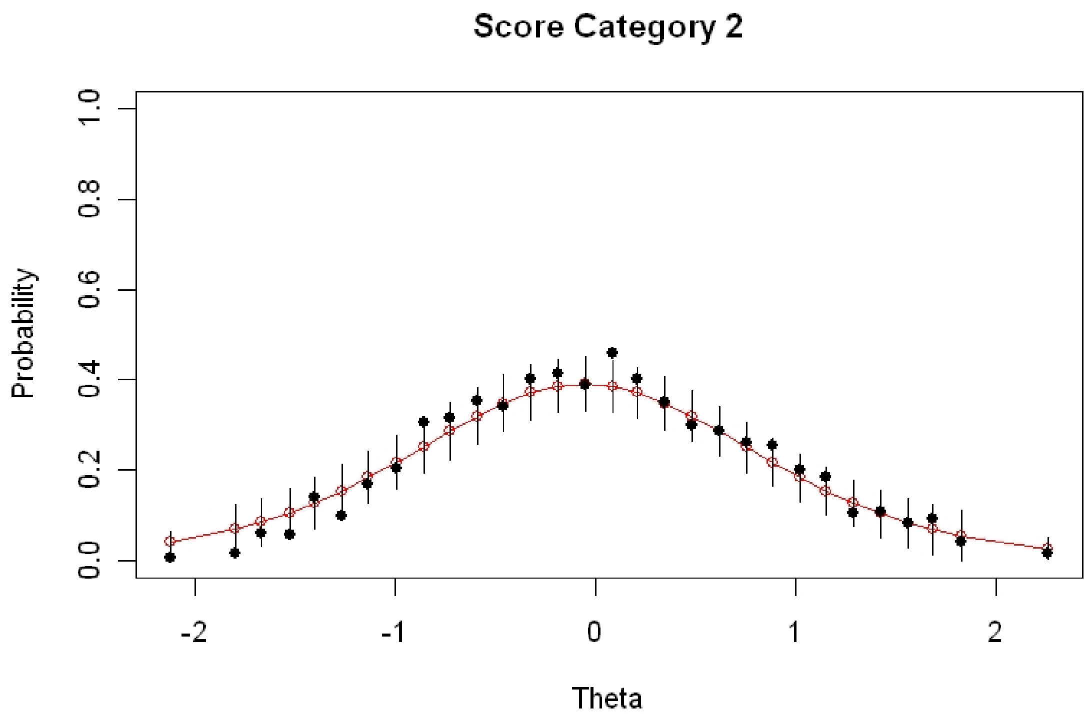


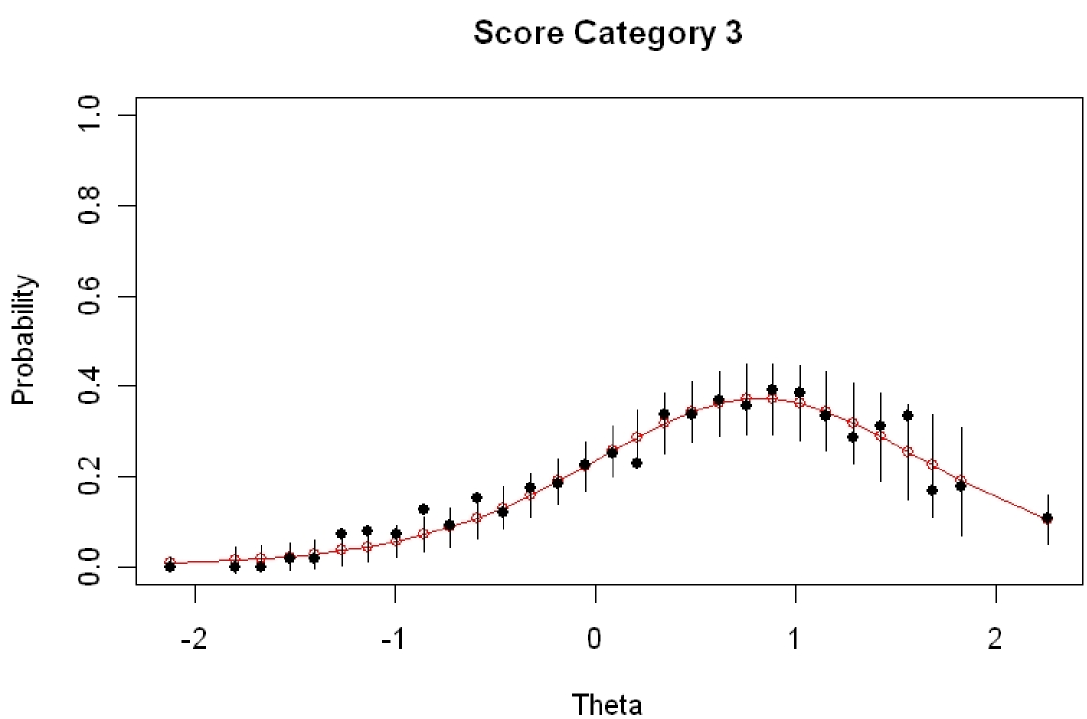


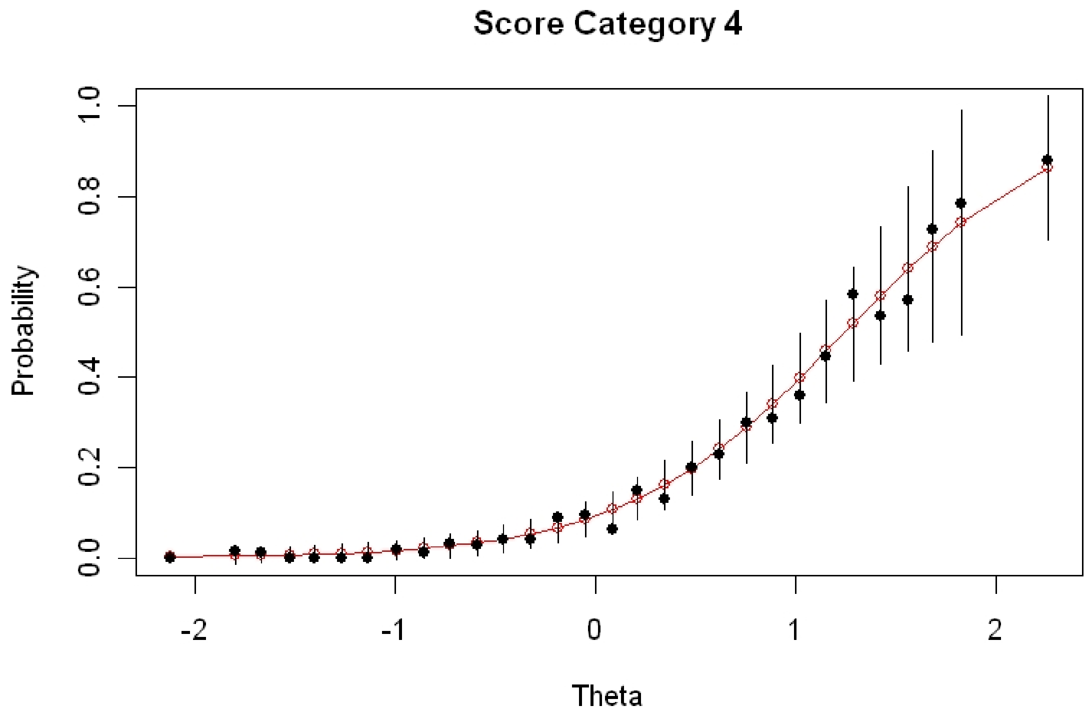
Item 11 ICCs


Score Category 0
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

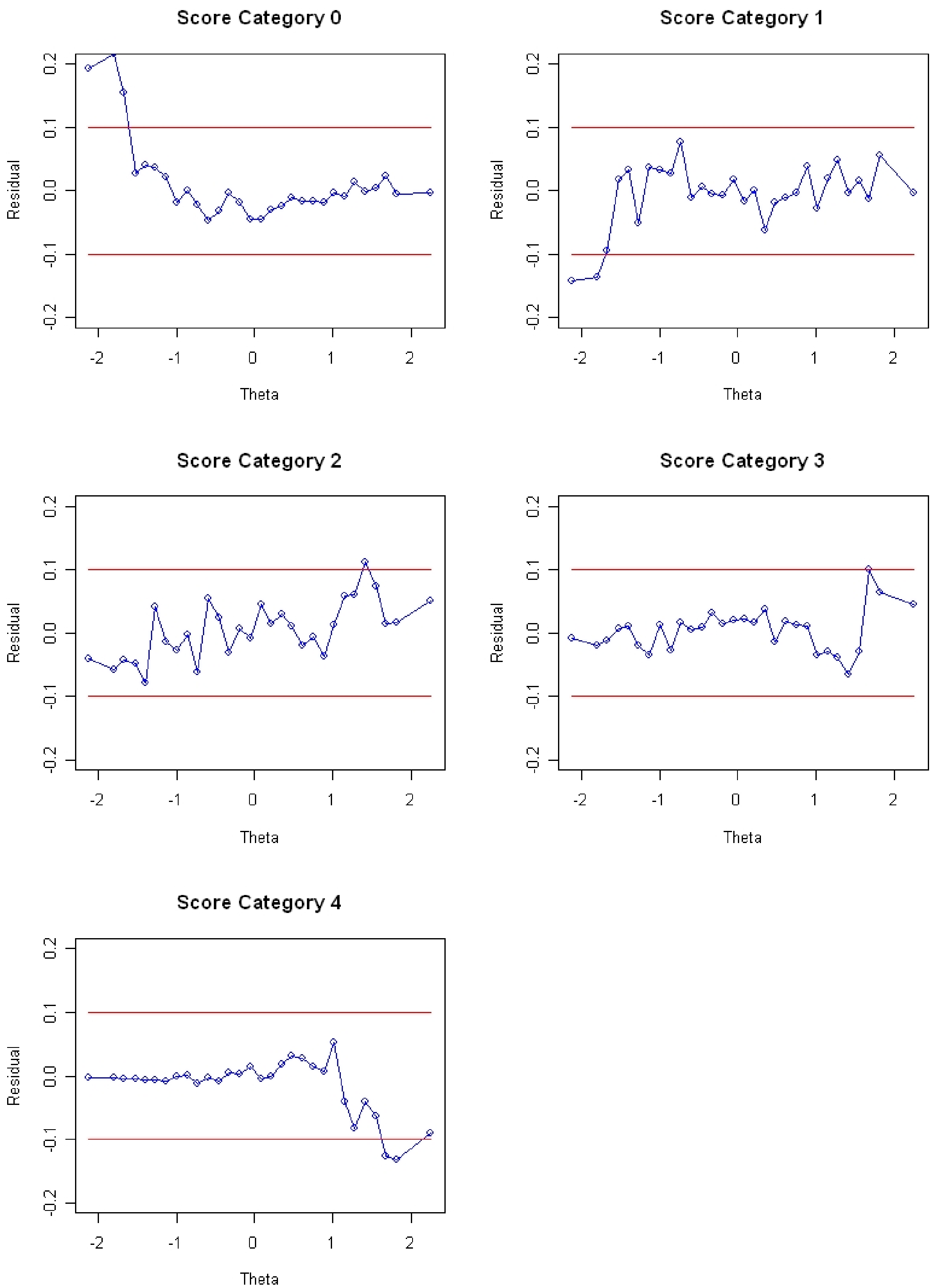

Score Category 1
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta


Score Category 2
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

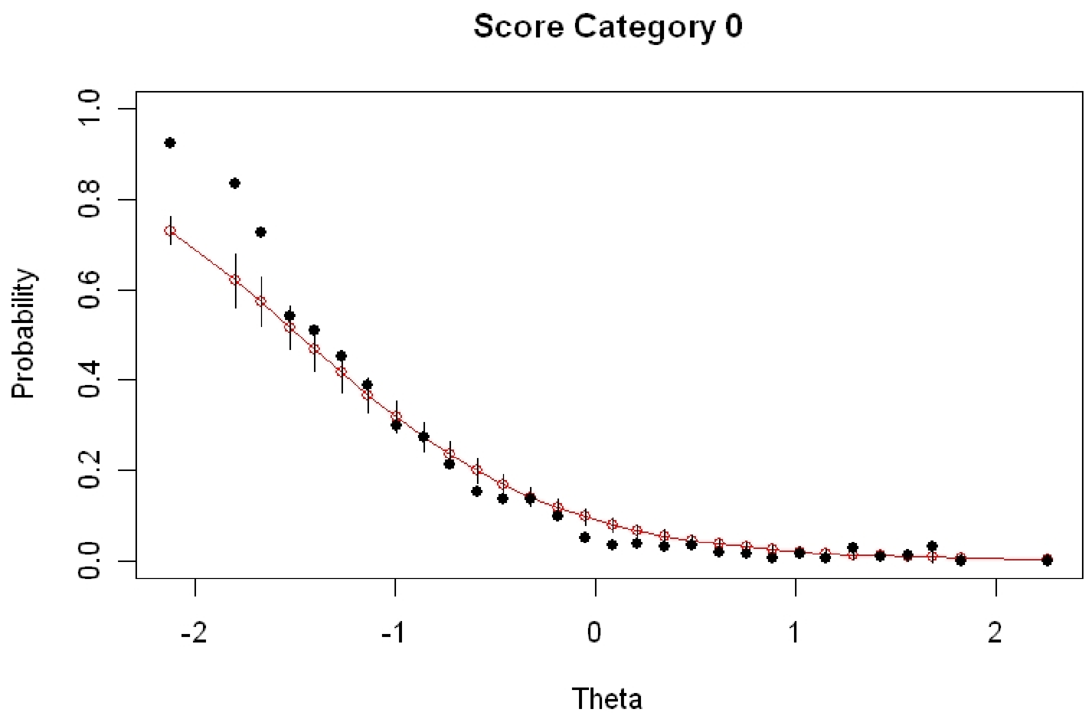

Score Category 3
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

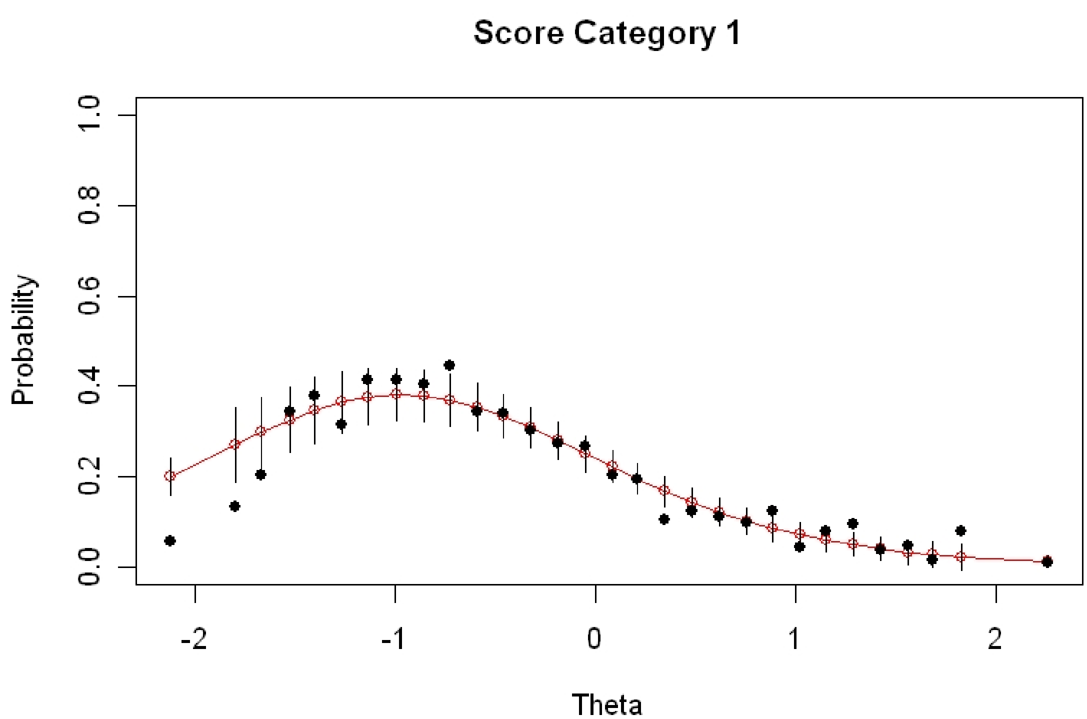

Score Category 4
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

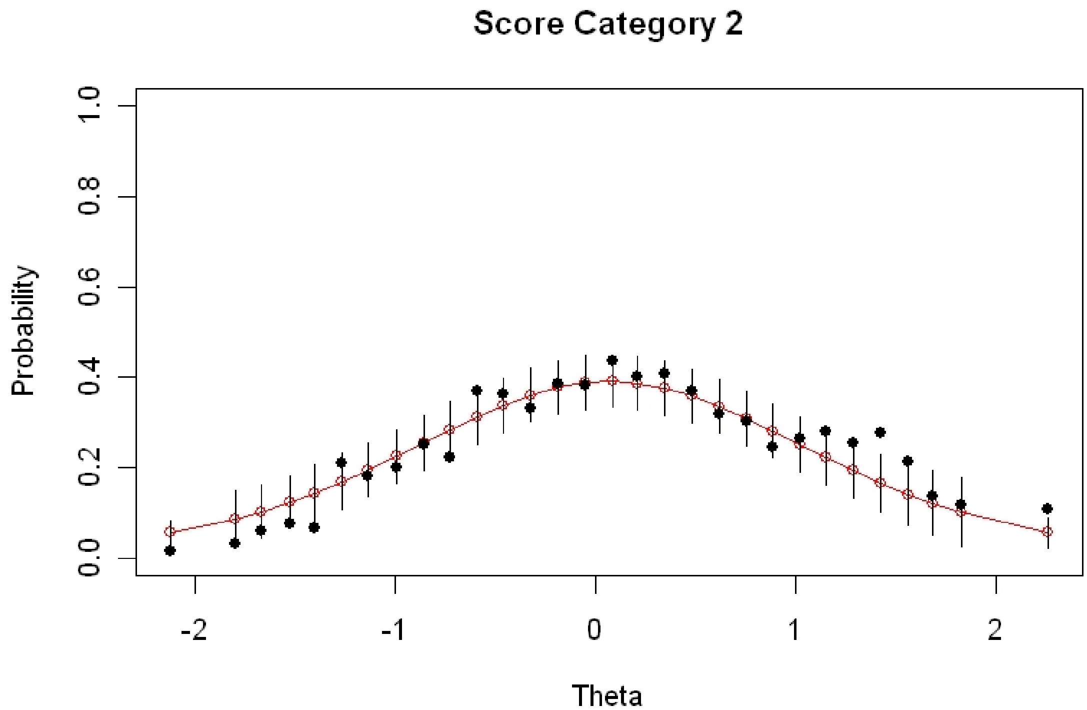
Item 25 Residuals

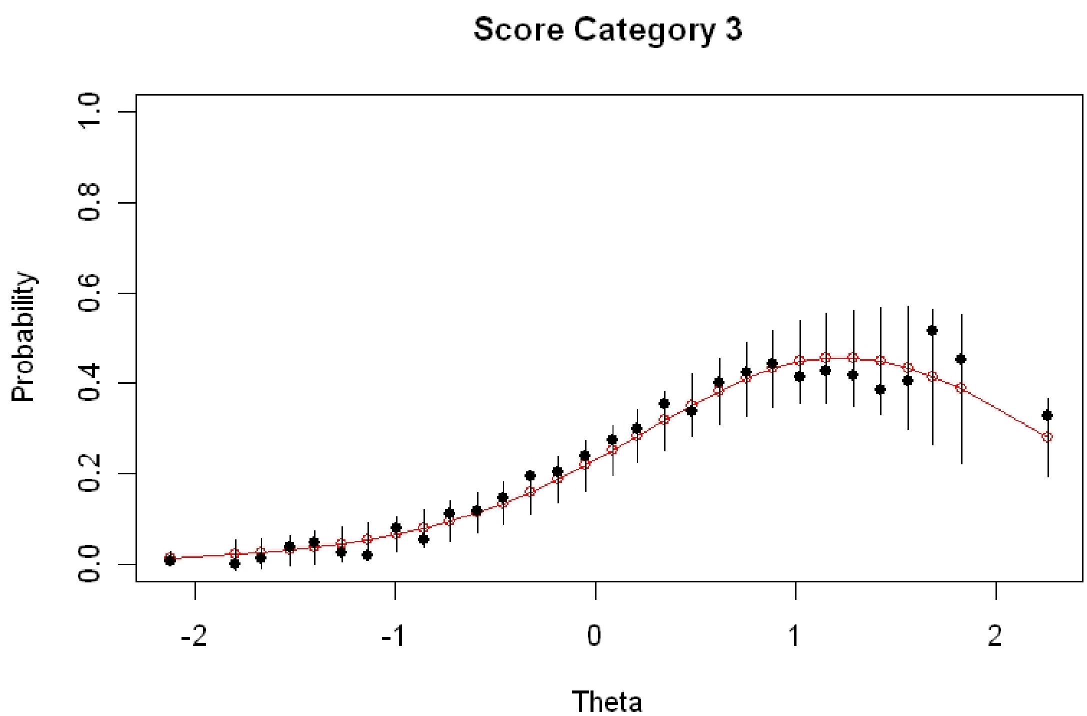


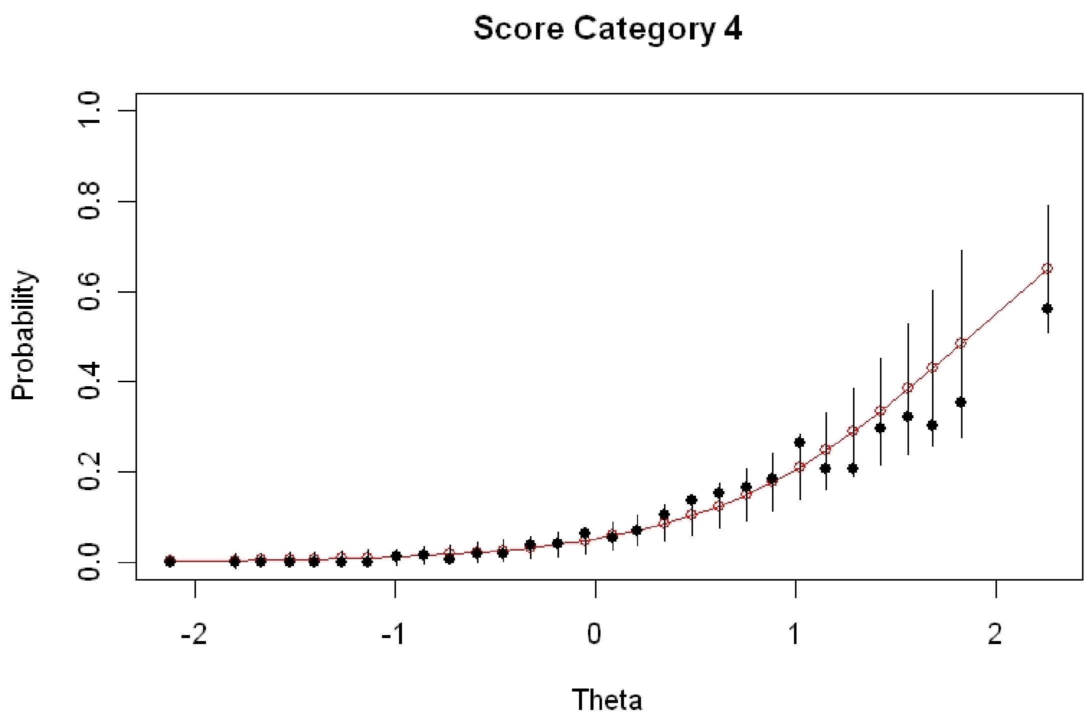
Item 25 ICCs


Score Category 0
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

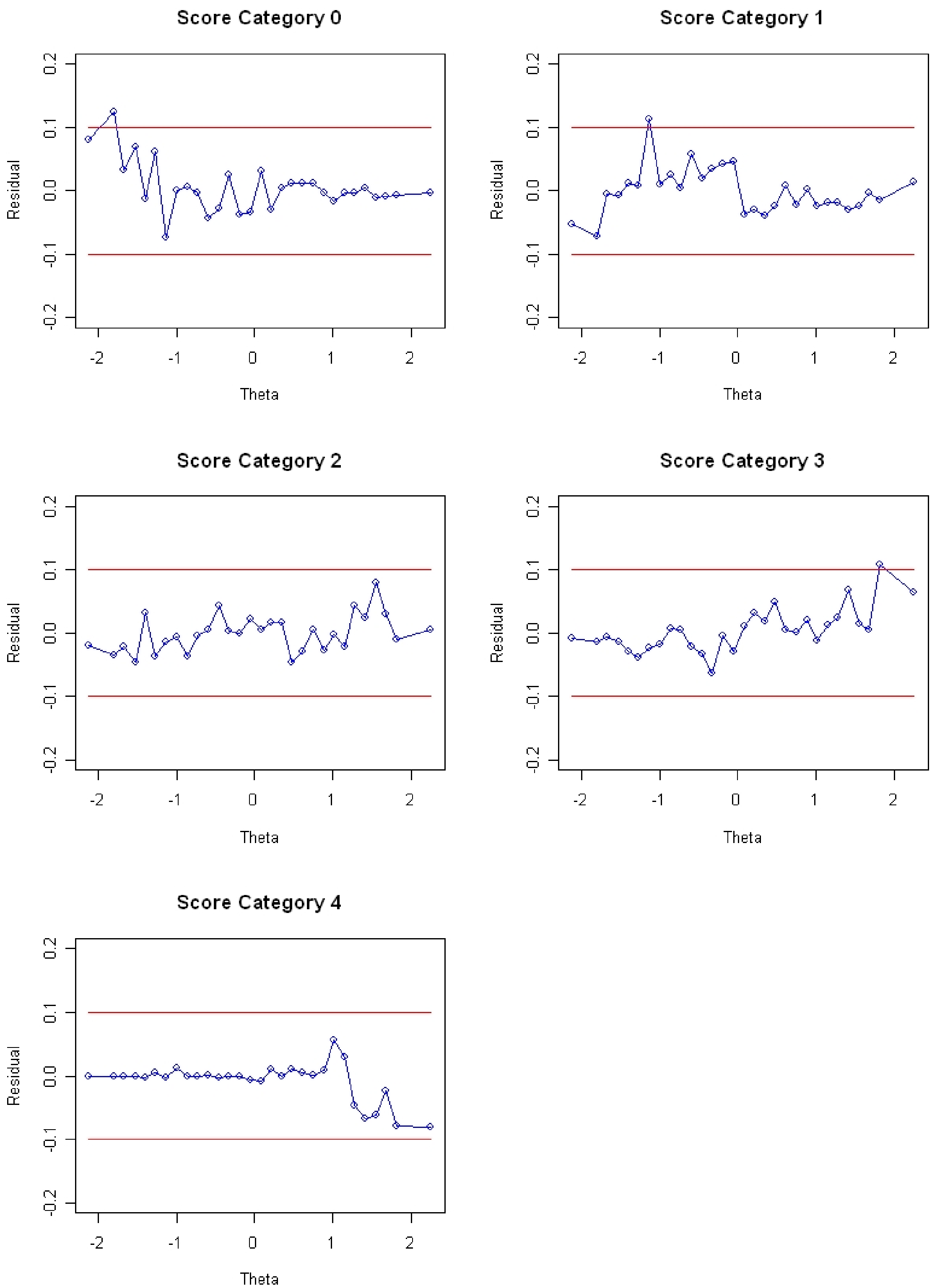

Score Category 1
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta


Score Category 2
Probability
Theta
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2

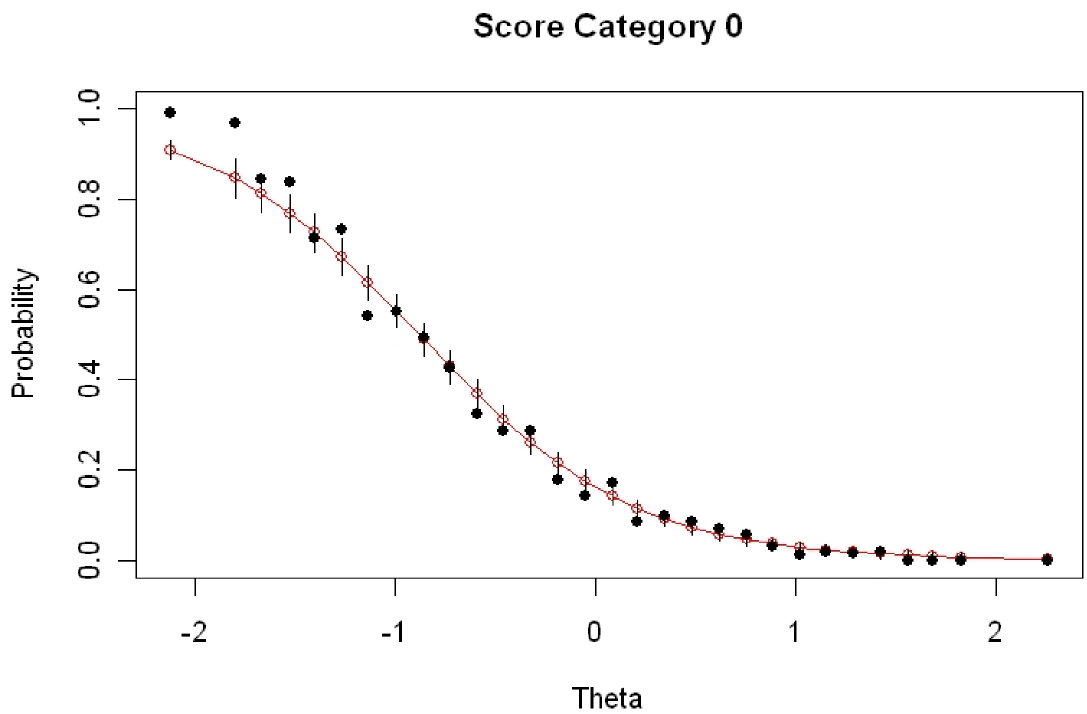

Score Category 3
Probability
Theta
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2


Score Category 4
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

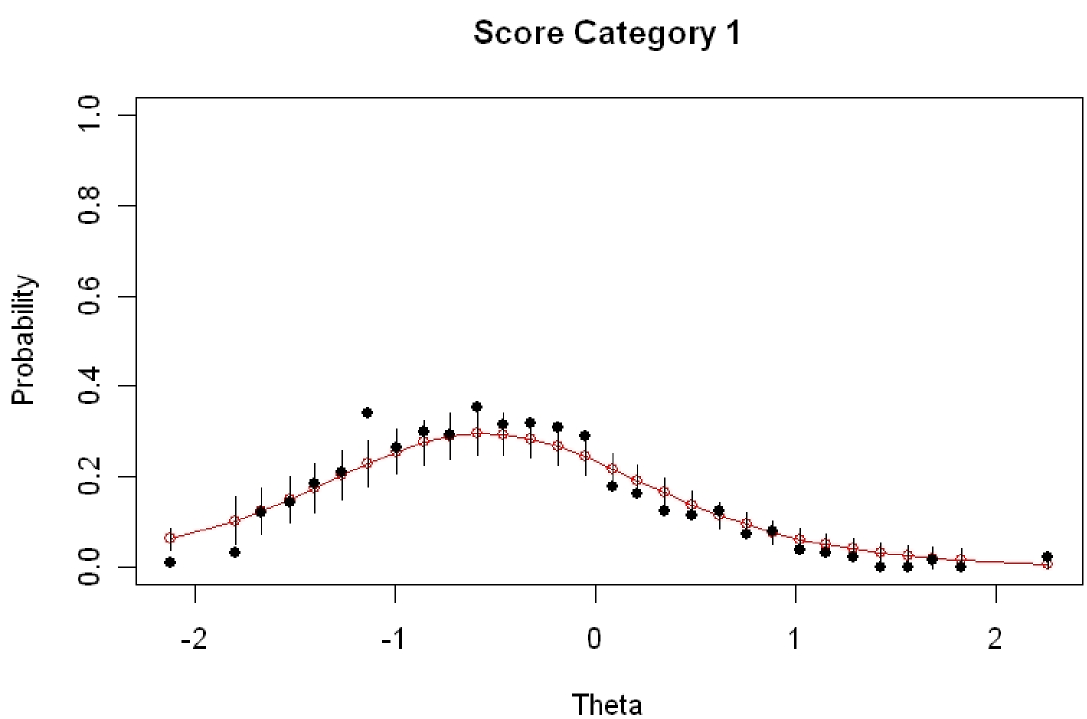
Item 26 Residuals



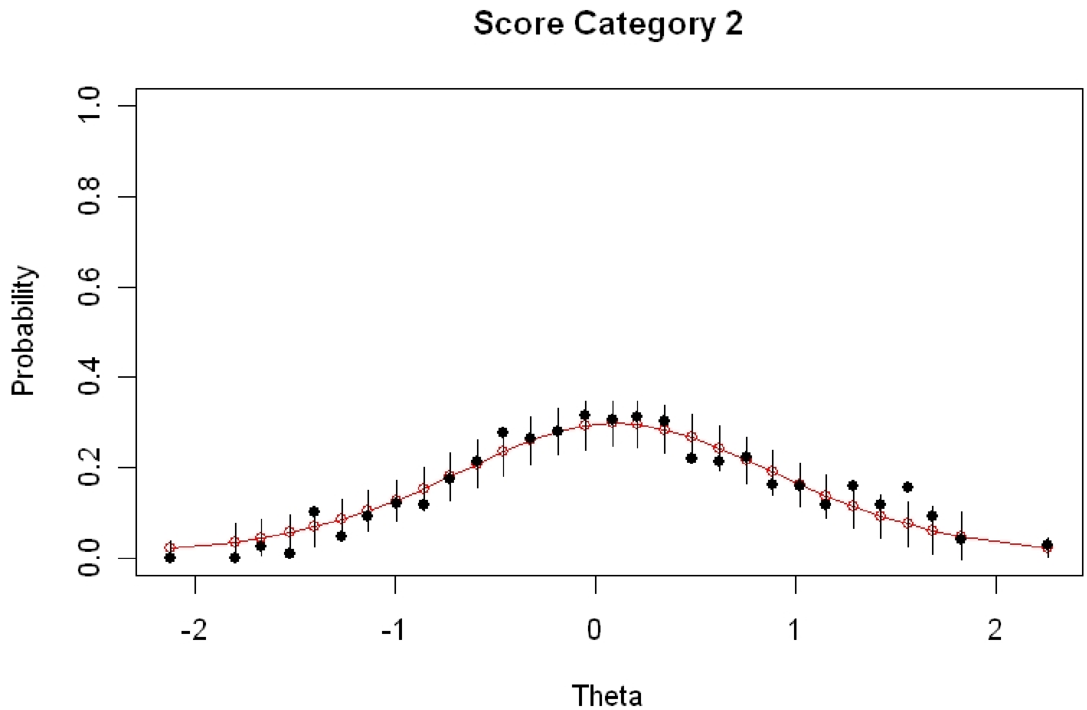
Item 26 ICCs



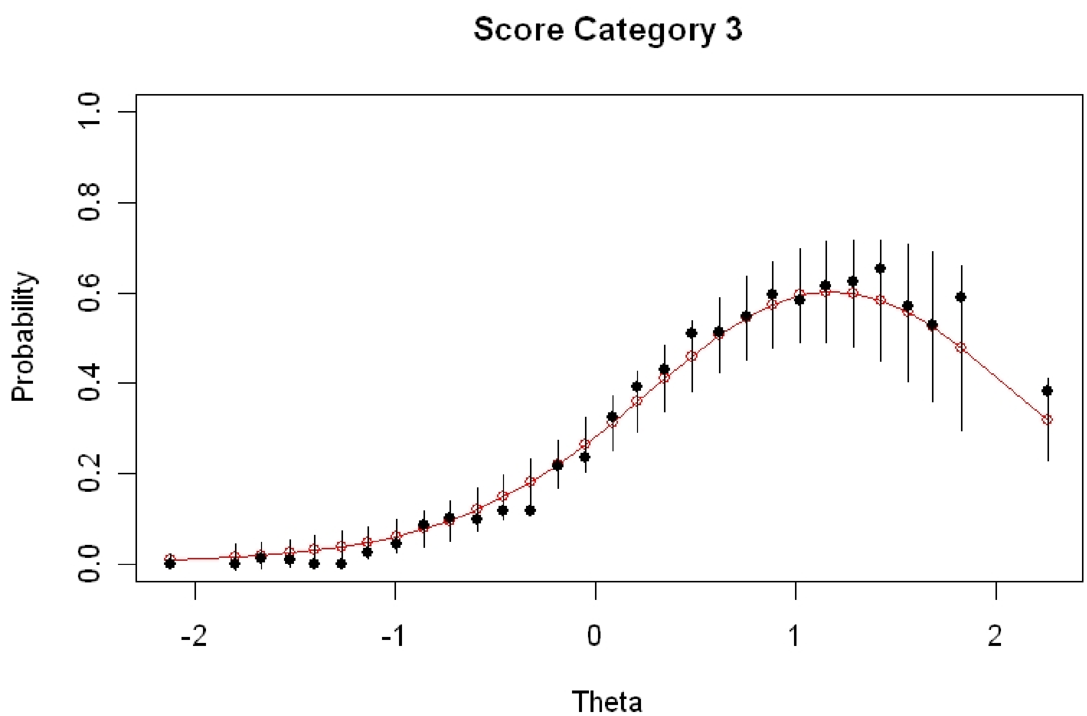
Score Category 0
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta



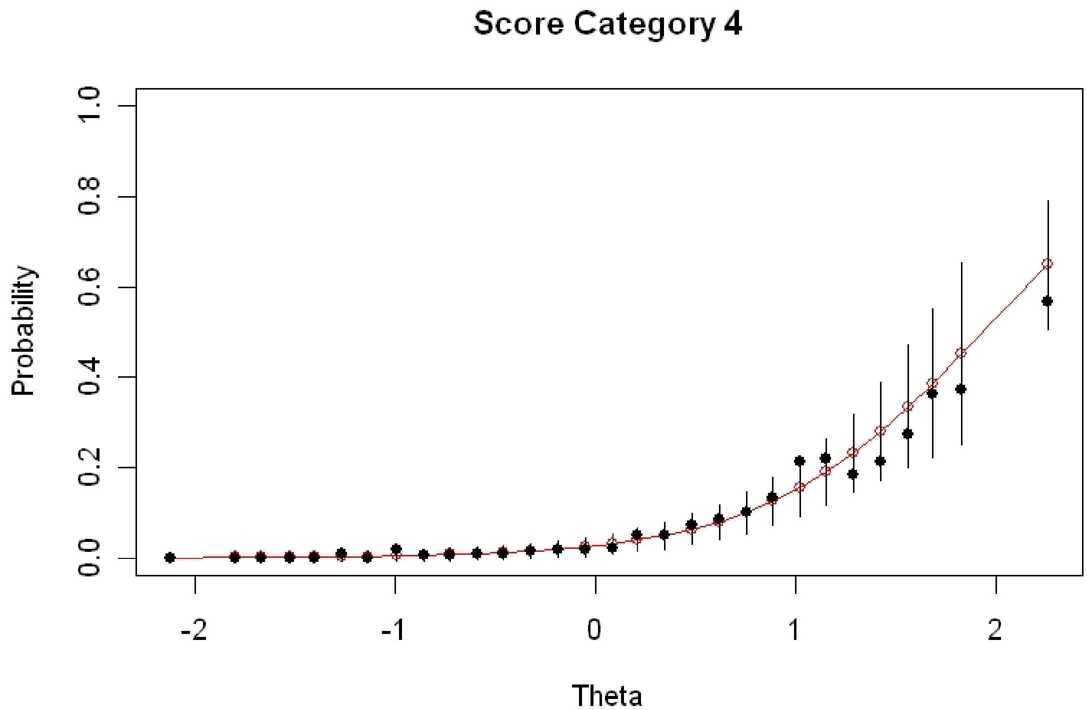
Score Category 1
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta



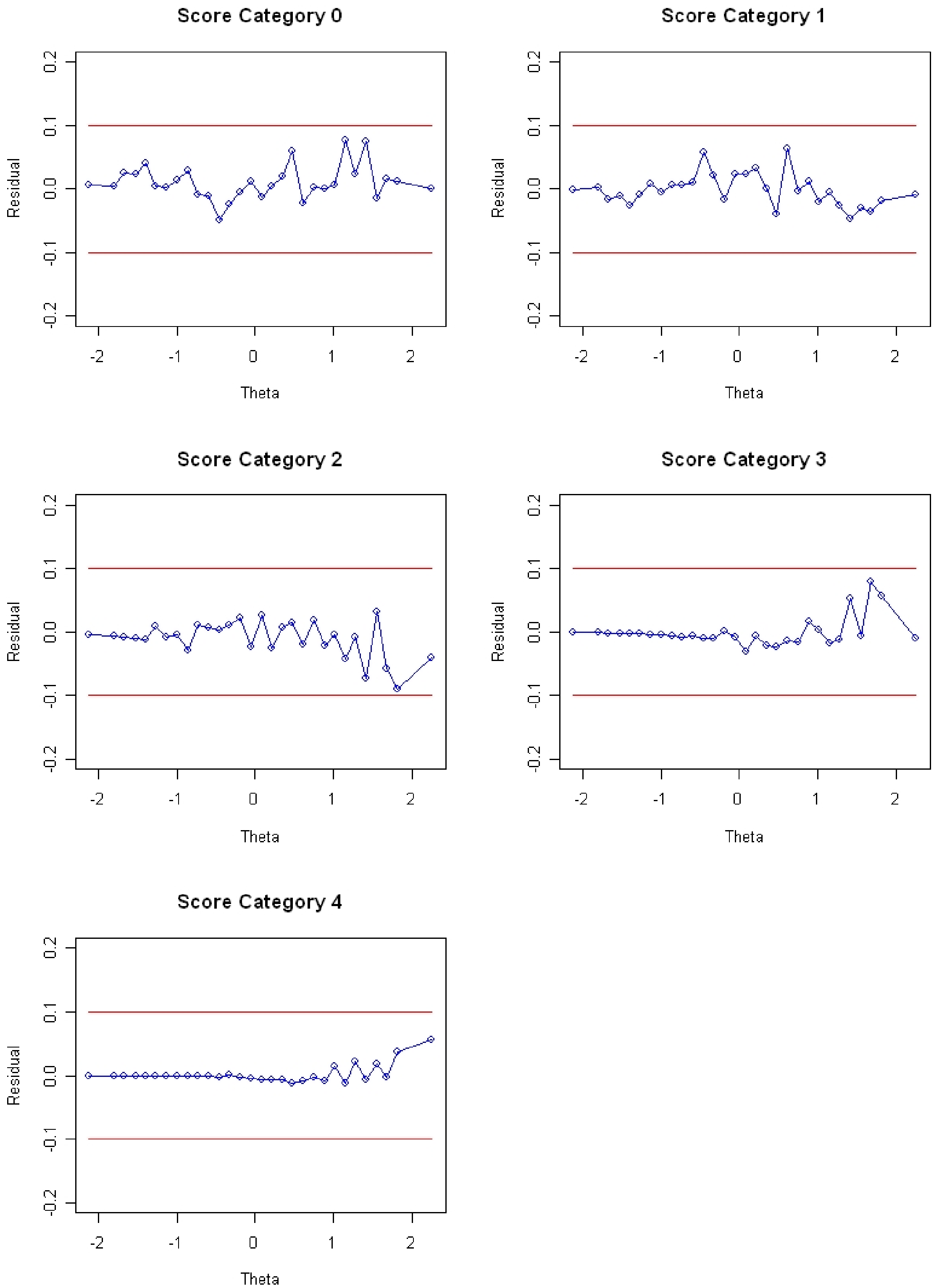
Score Category 2
Probability
1.0
0.8
0.6
0.4
0.2
0.0
-2
-1
0
1
2
Theta



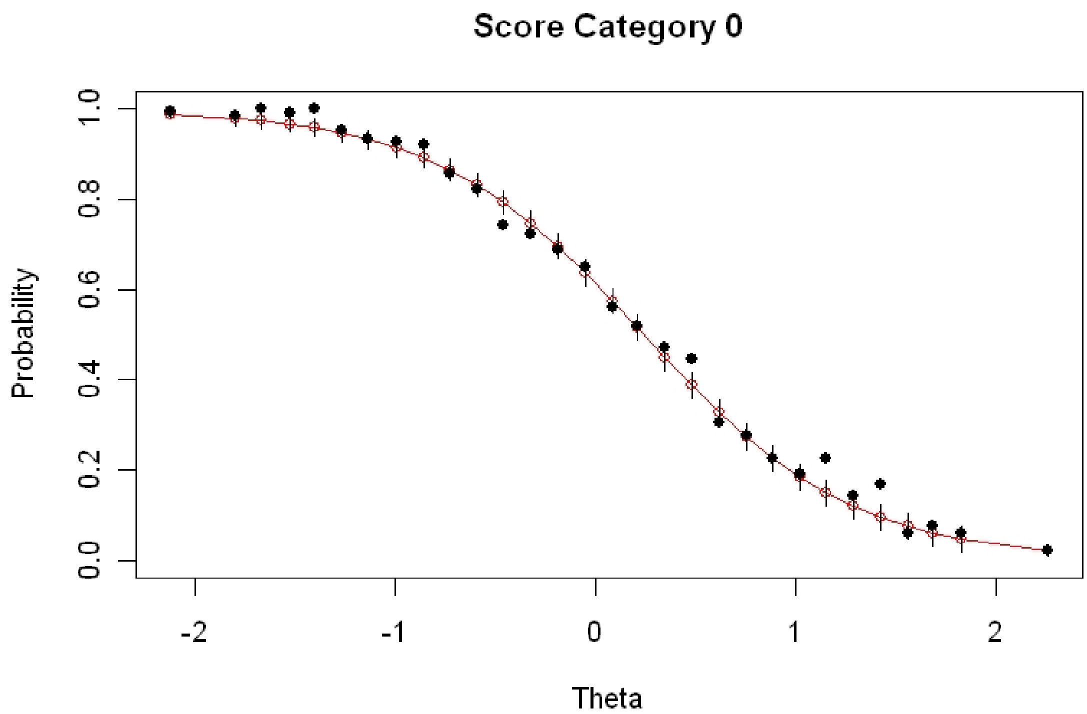
Score Category 3
Probability
1.0
0.8
0.6
0.4
0.2
0.0
-2
-1
0
1
2
Theta


Score Category 4
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

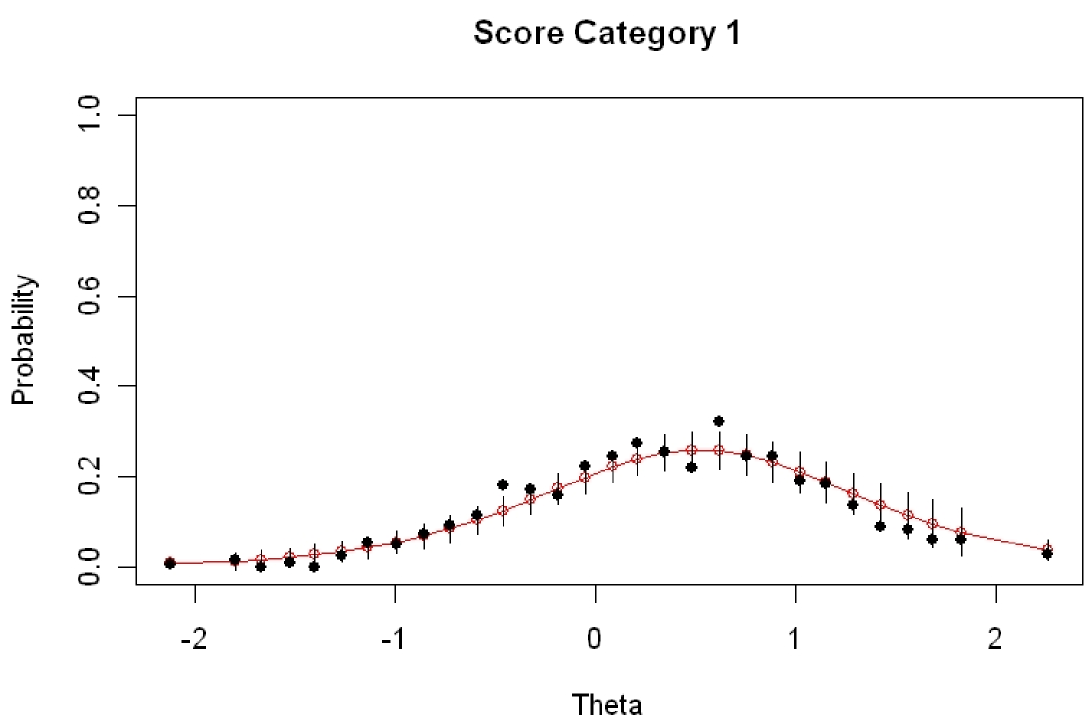
Item 32 Residuals



Item 32 ICCs

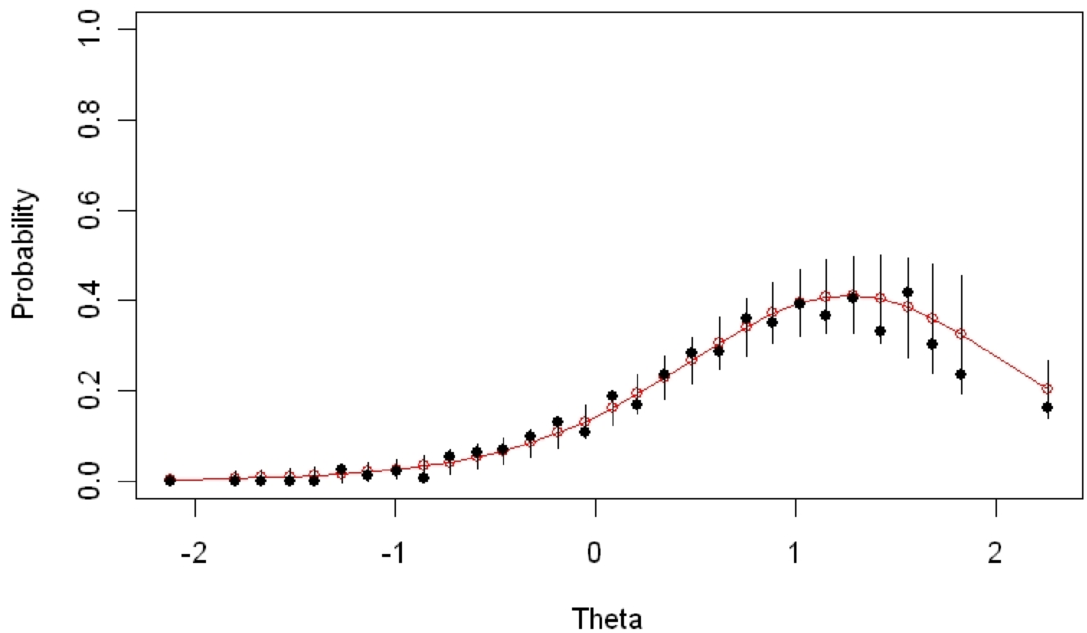


Score Category 0
Probability
1.0
0.8
0.6
0.4
0.2
0.0
-2
-1
0
1
2
Theta

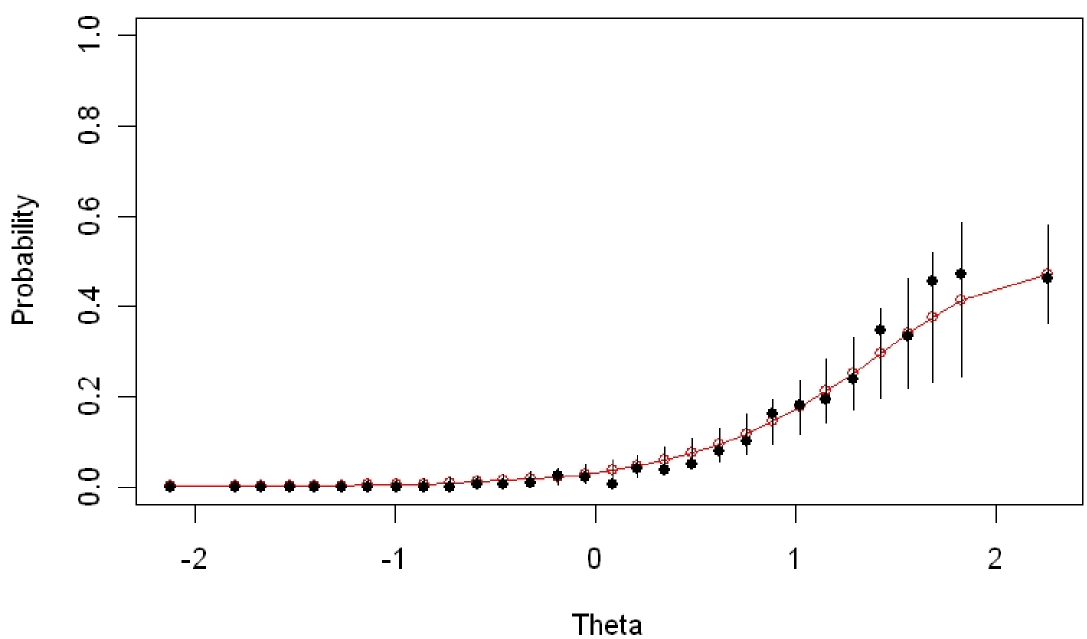


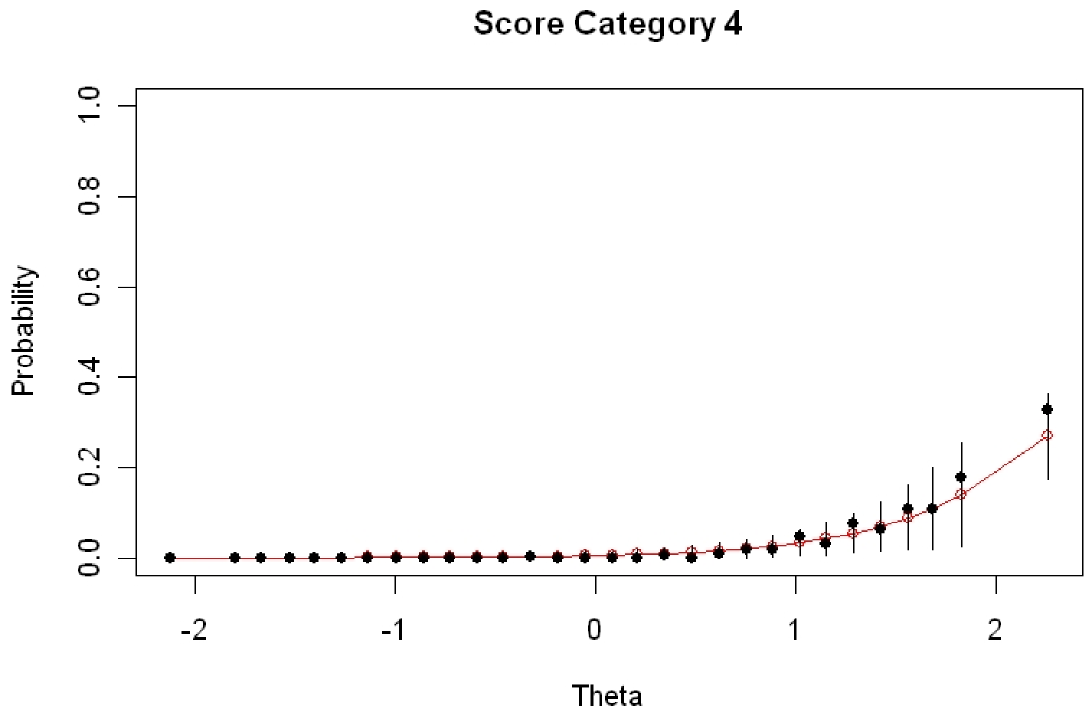
Score Category 1
Probability
1.0
0.8
0.6
0.4
0.2
0.0
-2
-1
0
1
2
Theta

Score Category 2

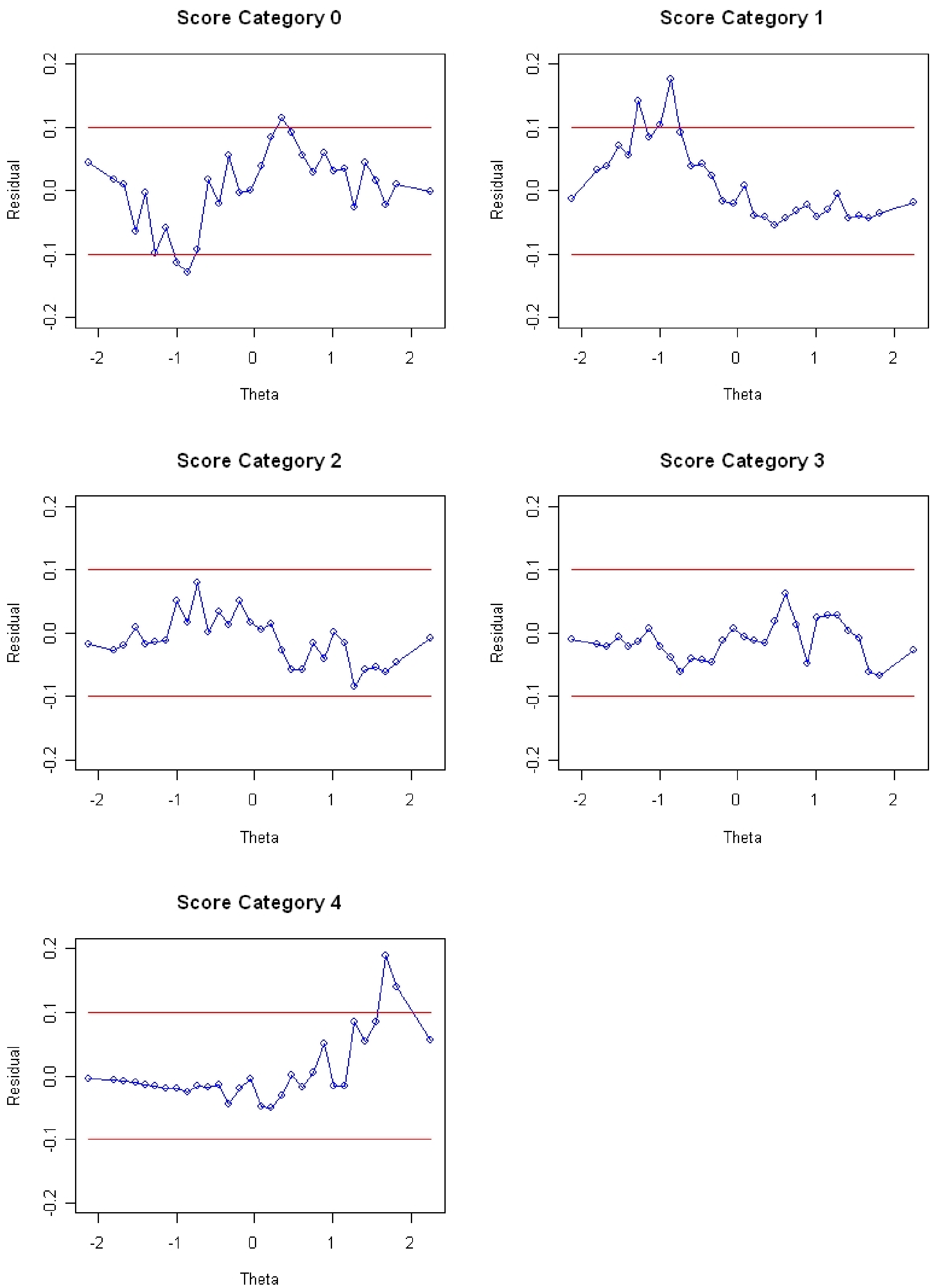


Score Category 3

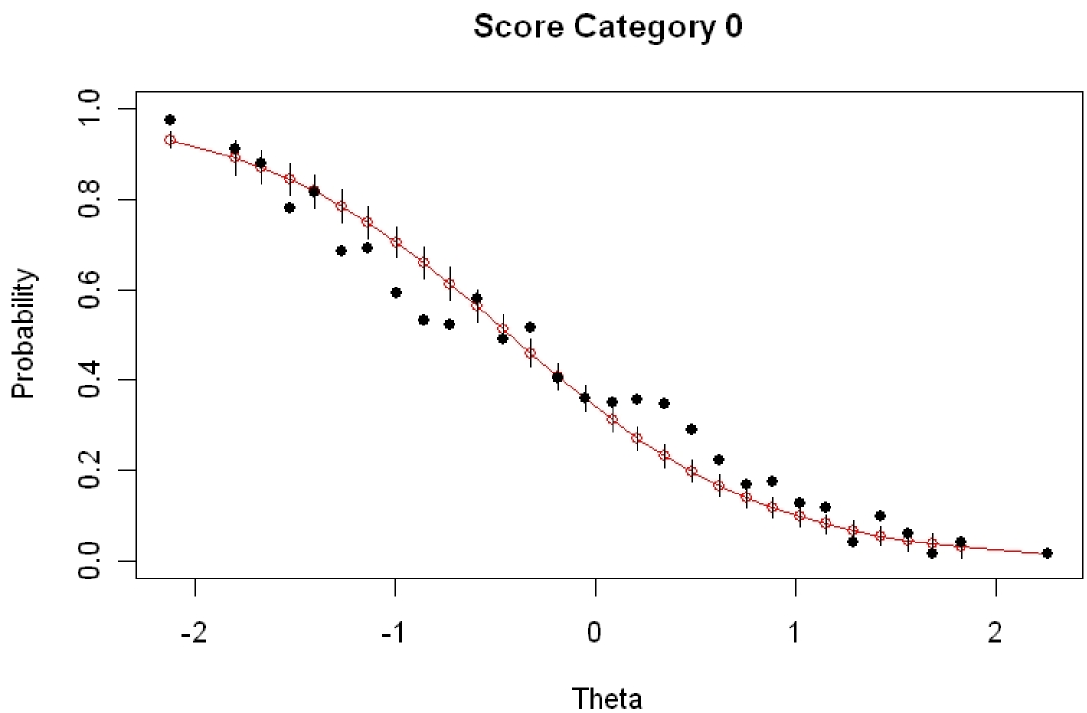


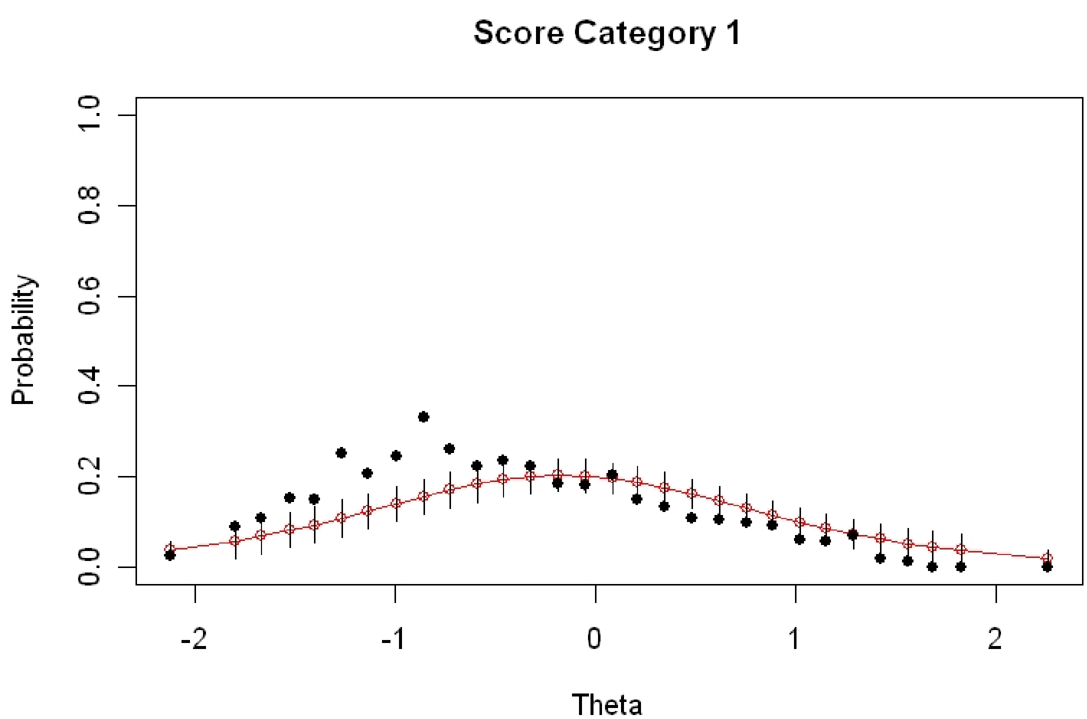

Score Category 4
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta

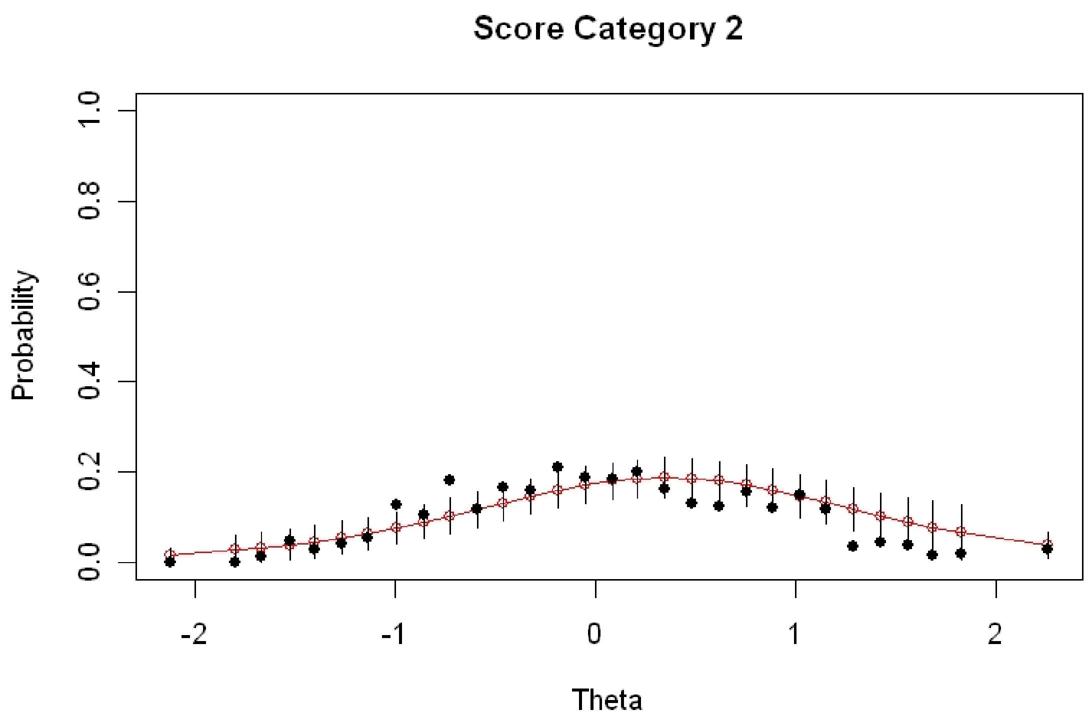
Item 39 Residuals

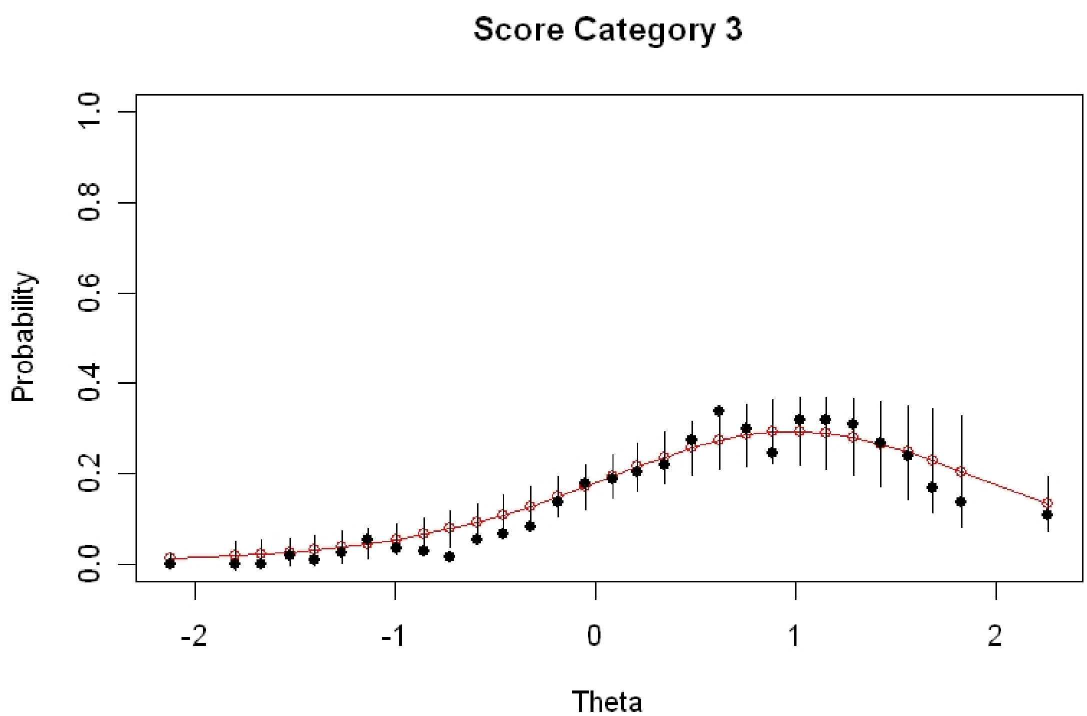


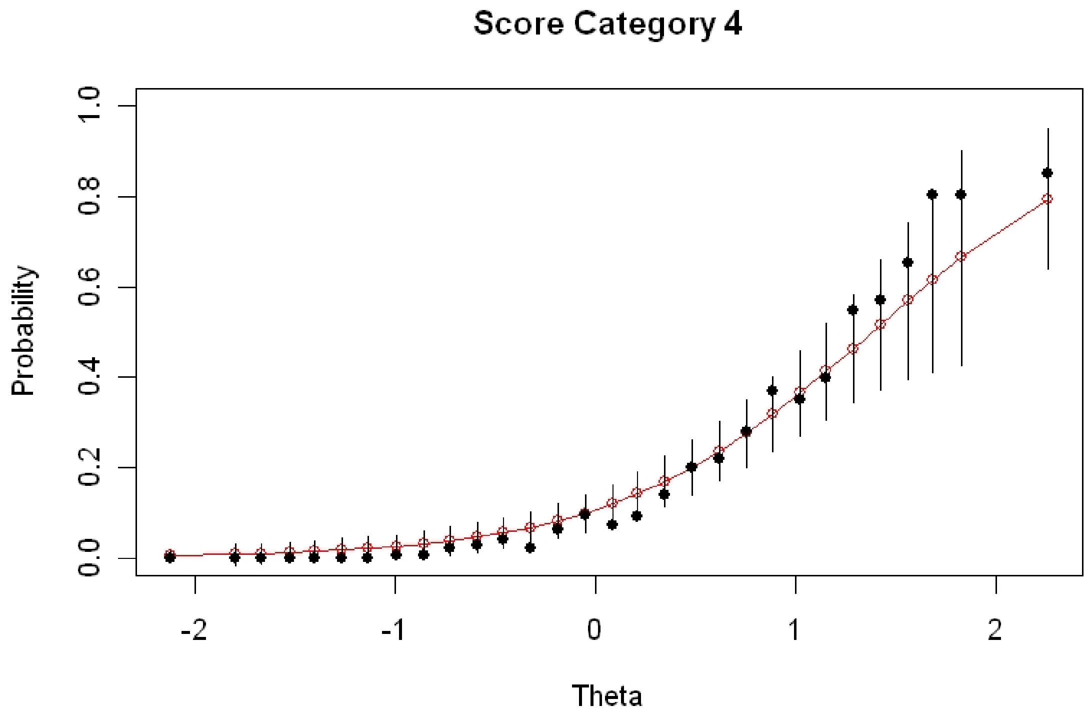
Item 39 ICCs


Score Category 0
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta


Score Category 1
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta


Score Category 2
Probability
1.0
0.8
0.6
0.4
0.2
0.0
-2
-1
0
1
2
Theta


Score Category 3
Probability
1.0
0.8
0.6
0.4
0.2
0.0
-2
-1
0
1
2
Theta


Score Category 4
Probability
0.0
0.2
0.4
0.6
0.8
1.0
-2
-1
0
1
2
Theta