

Item Response Theory Models in Stata

Meghan K. Cain, Ph.D. | September 7th, 2022

You can download the slides and other materials here:

<https://tinyurl.com/StataIRT22>



Overview

- What is item response theory?
- Math example
 - 1PL/Rasch model
 - 2PL
 - 3PL
- Differential item functioning
- Graded response models



What is item response theory?

- Item response theory (IRT) is a measurement model framework.
- Frequently, we assume that our observed variables are perfect representations of their underlying constructs we wish to study:
 - Ability
 - Beliefs
 - Personality
- This is rarely true in practice.



What is item response theory?

- IRT uses responses to a series of items to measure
 - item properties
 - respondents' latent trait levels
- In research, these are often estimated jointly.
- In standardized testing, item calibration occurs during test standardization. These item parameters are used subsequently to measure respondent traits.

Junior Certificate Mathematics

Adedoyin & Mokobi (2013)

Item 1

Which class of numbers is listed below?

2, 3, 5, 7,...

- a) Rectangle numbers
- b) Even numbers
- c) Prime numbers
- d) Odd numbers

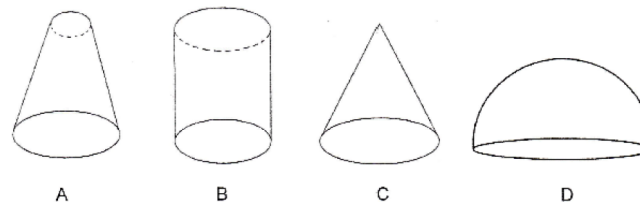
Item 25

The number of students in a school decreased by 5%.
After the decrease there are 475 students in the school.
Calculate the number of students before the decrease.

- a) 570
- b) 500
- c) 480
- d) 740

Item 15

Which of the following is a prism?



Item 21

Expand $(a+x)(a+2)$

- a) $4ax+3ax$
- b) $2a+ax+2x$
- c) $a^2+3ax+2x$
- d) $a^2+2a+ax+2x$

JC Math example

```
. use math
```

```
. codebook, compact
```

Variable	Obs	Unique	Mean	Min	Max	Label
item1	1000	2	.885	0	1	Item 1
item2	1000	2	.493	0	1	Item 2
item3	1000	2	.656	0	1	Item 3
item4	1000	2	.615	0	1	Item 4
item5	1000	2	.791	0	1	Item 5
item6	1000	2	.313	0	1	Item 6
esl	1000	2	.5	0	1	ESL student (1=yes)

Math proficiency

```
. generate score = item1 + item2 + item3 + item4 + item5 + item6  
  
. list item* score in 1/10
```

	item1	item2	item3	item4	item5	item6	score
1.	0	0	0	0	1	0	1
2.	1	1	0	1	1	0	4
3.	0	0	1	1	0	0	2
4.	1	1	1	0	1	1	5
5.	1	1	1	1	1	0	5
6.	1	0	0	1	1	1	4
7.	1	1	0	0	1	0	3
8.	1	0	1	1	1	0	4
9.	1	0	0	0	1	0	2
10.	1	1	0	0	1	1	4

Item difficulty

```
. summarize item*
```

Variable	Obs	Mean	Std. dev.	Min	Max
item1	1,000	.885	.3191816	0	1
item2	1,000	.493	.5002012	0	1
item3	1,000	.656	.4752787	0	1
item4	1,000	.615	.4868388	0	1
item5	1,000	.791	.4067978	0	1
item6	1,000	.313	.4639464	0	1



Advantages of IRT

- Put item difficulty and ability on the same scale.
- Estimate standard errors for each response pattern.
- Compare model fit.
- Link tests with different items.
- Choose an optimum set of items for each respondent from a pool, i.e. Computerized adaptive testing.

1PL model

- One-parameter logistic model
- We model the probability of respondent j answering item i correctly, $\pi_{ij} = \Pr(Y_{ij} = 1 | b_i, \theta_j)$

$$\pi_{ij} = \frac{\exp\{\theta_j - b_i\}}{1 + \exp\{\theta_j - b_i\}}$$

θ_j respondent j 's latent trait

b_i item i 's difficulty (location)

1PL model

- One-parameter logistic model
- We model the probability of respondent j answering item i correctly, $\pi_{ij} = \Pr(Y_{ij} = 1 | b_i, \theta_j)$

$$\pi_{ij} = \frac{\exp\{\theta_j - b_i\}}{\underbrace{1 + \exp\{\theta_j - b_i\}}_{(0,1)}}$$

θ_j respondent j 's latent trait

b_i item i 's difficulty

1 PL model – item difficulty

```
. irt 1pl item1-item6
```

One-parameter logistic model
Log likelihood = -3439.2673

Number of obs = 1,000

		Coefficient	Std. err.	z	P> z	[95% conf. interval]	
	Discrim	.7684024	.0531589	14.45	0.000	.6642129	.8725919
item1	Diff	-2.934959	.218892	-13.41	0.000	-3.36398	-2.505939
item2	Diff	.0413469	.0931648	0.44	0.657	-.1412529	.2239466
item3	Diff	-.9475985	.1108453	-8.55	0.000	-1.164851	-.7303456
item4	Diff	-.6884687	.1027827	-6.70	0.000	-.889919	-.4870184
item5	Diff	-1.937404	.1565575	-12.38	0.000	-2.244251	-1.630556
item6	Diff	1.152635	.1106601	10.37	0.000	.932046	1.385223

1 PL model – item difficulty

```
. estat report, sort(b) byparm
```

One-parameter logistic model
Log likelihood = -3439.2673

Number of obs = 1,000

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
Discrim	.7684024	.0531589	14.45	0.000	.6642129	.8725919
Diff						
item1	-2.934959	.218892	-13.41	0.000	-3.36398	-2.505939
item5	-1.937404	.1565575	-12.38	0.000	-2.244251	-1.630556
item3	-.9475985	.1108453	-8.55	0.000	-1.164851	-.7303456
item4	-.6884687	.1027827	-6.70	0.000	-.889919	-.4870184
item2	.0413469	.0931648	0.44	0.657	-.1412529	.2239466
item6	1.152635	.1186691	9.71	0.000	.920048	1.385222

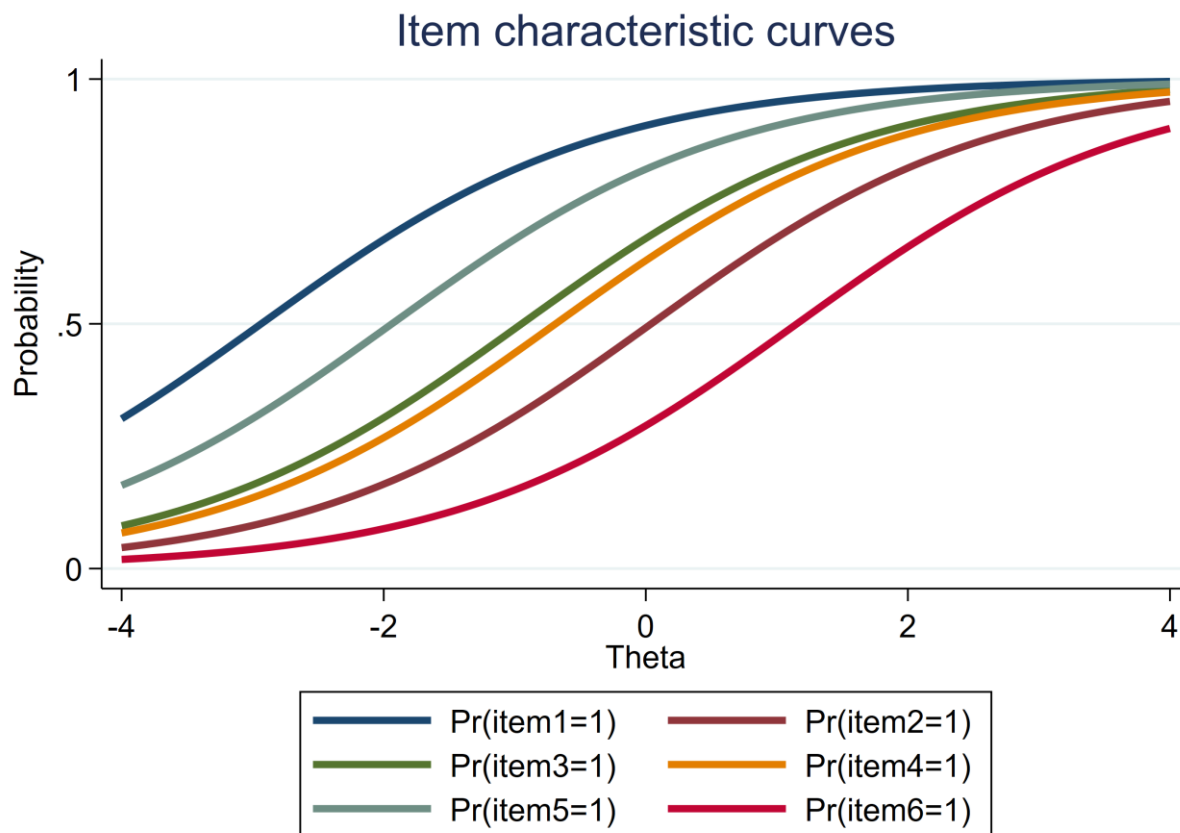
1PL model – math proficiency

```
. predict theta1, latent se(th1_se)  
. list score theta1 th1_se in 1/10
```

	score	theta1	th1_se
1.	1	-1.271147	.7720639
2.	4	.1061837	.7814902
3.	2	-.814758	.7701293
4.	5	.5830028	.7948129
5.	5	.5830028	.7948129
6.	4	.1061837	.7814902
7.	3	-.3576659	.7732583
8.	4	.1061837	.7814902
9.	2	-.814758	.7701293
10.	4	.1061837	.7814902

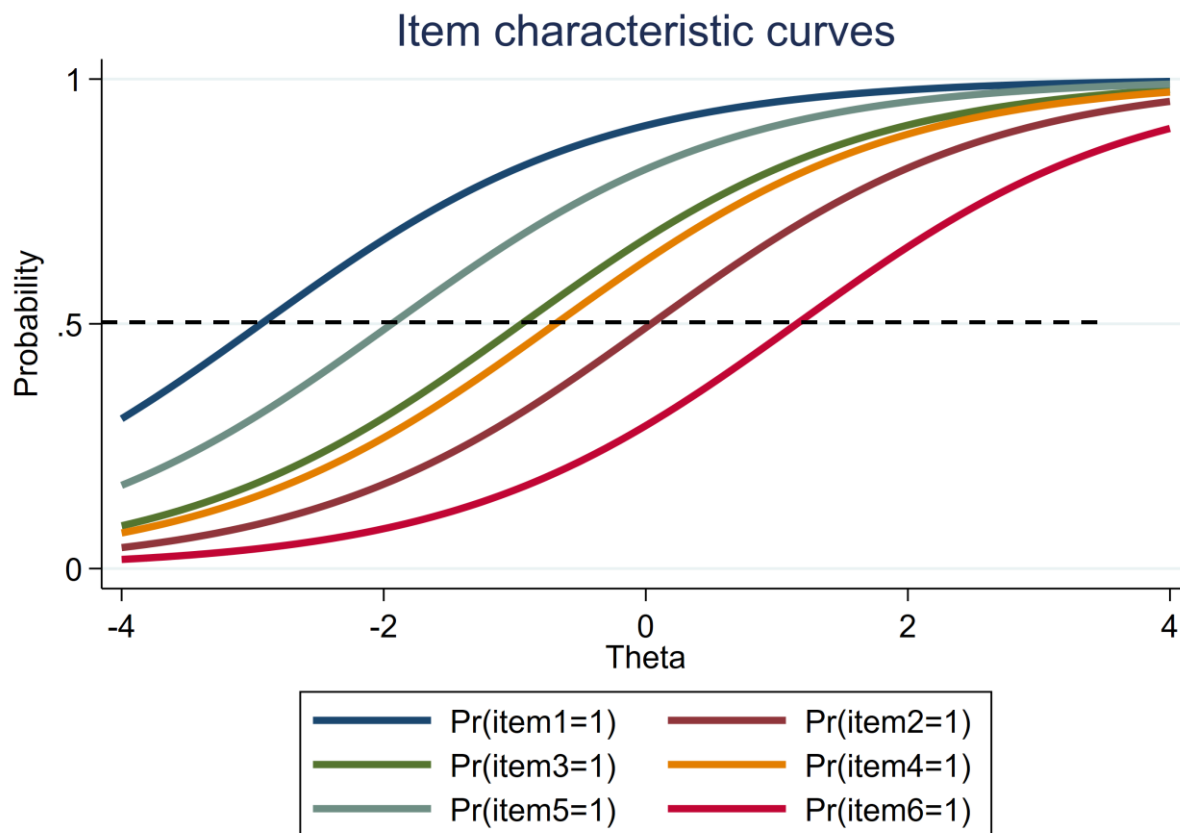
1PL model

```
. irtgraph icc item1-item6
```



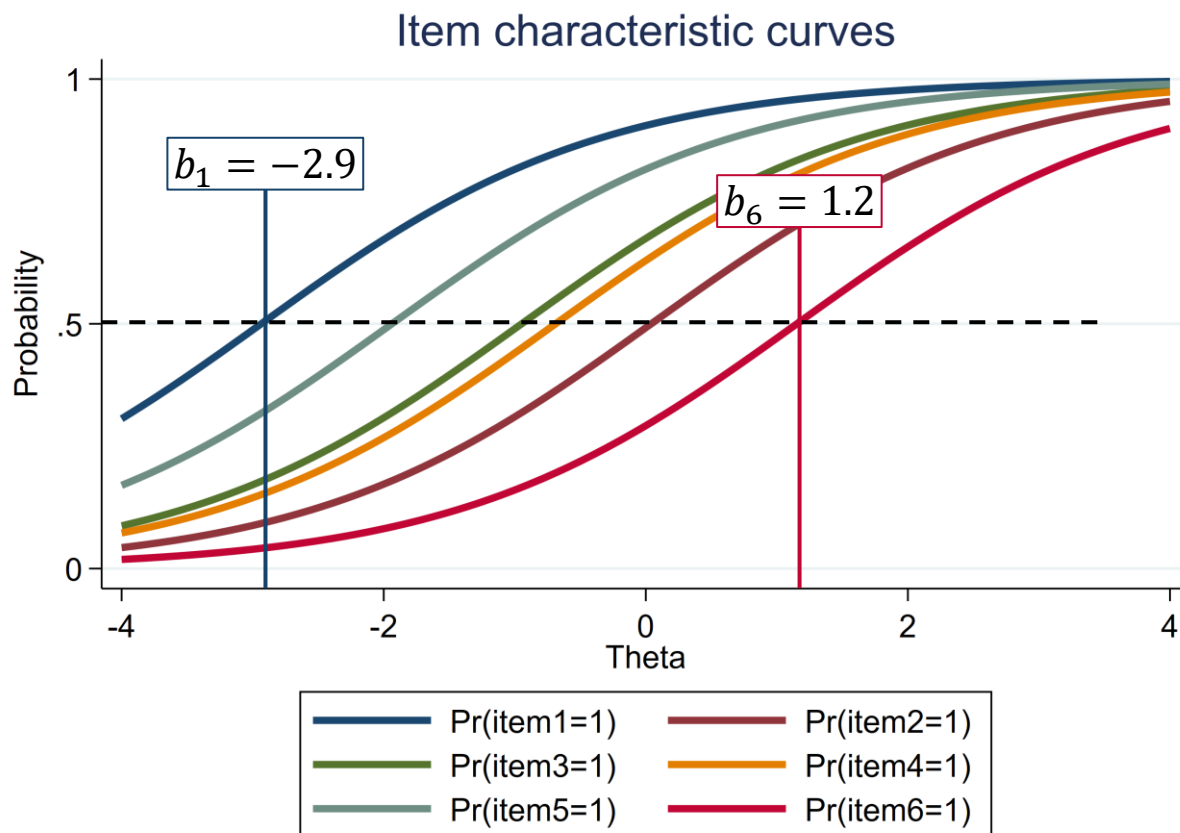
1PL model

```
. irtgraph icc item1-item6
```



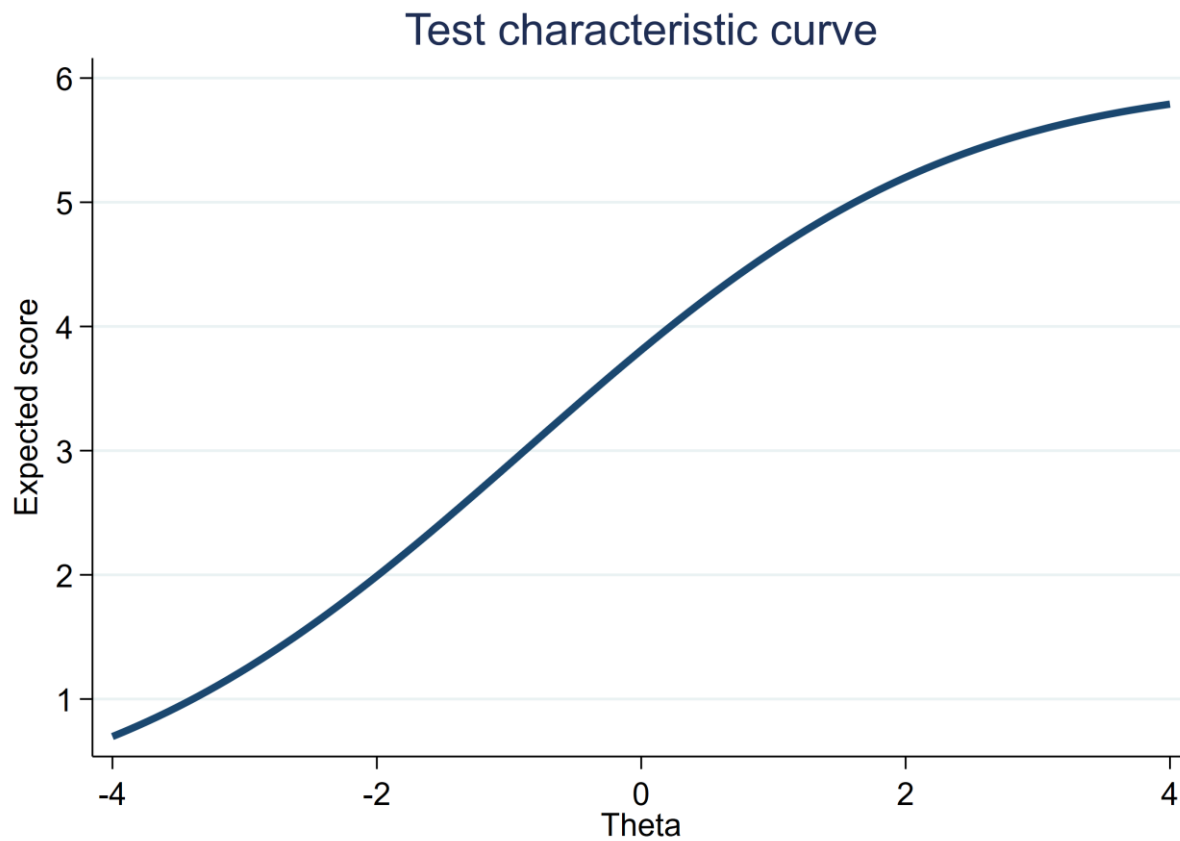
1PL model

```
. irtgraph icc item1-item6
```



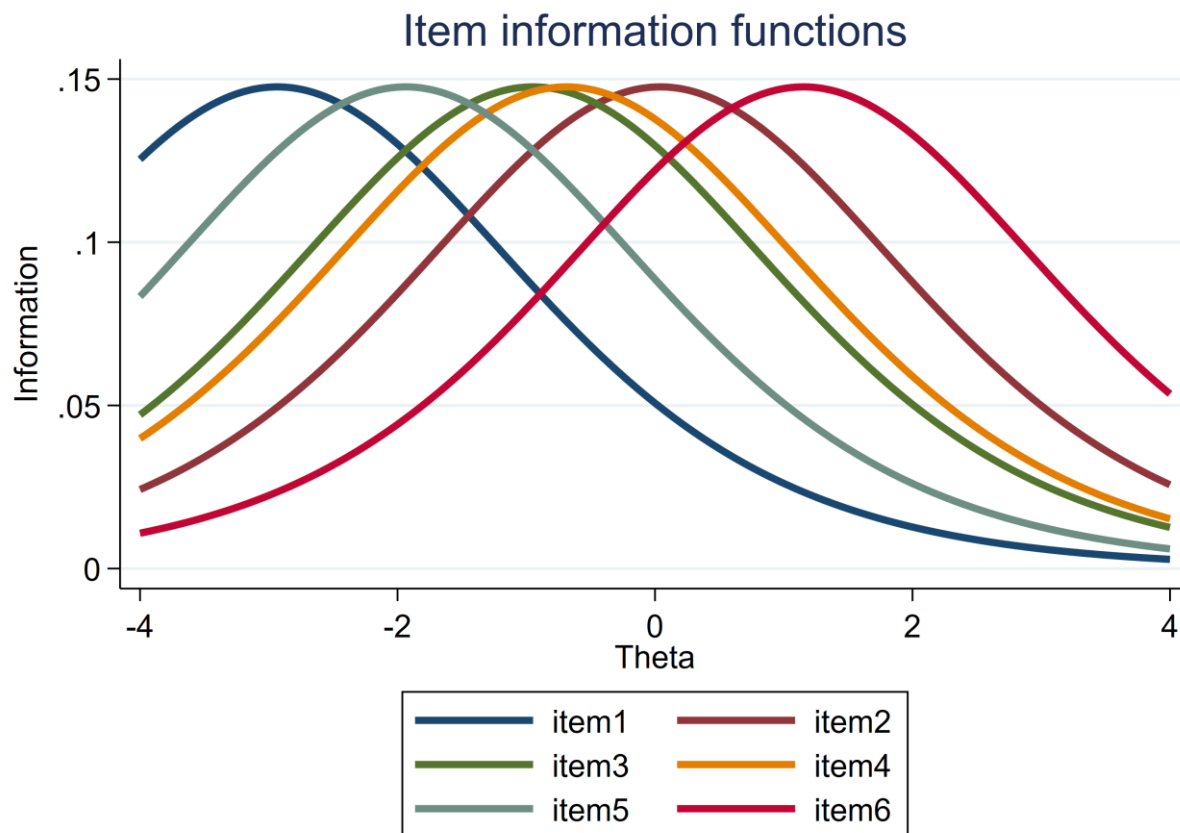
1PL model

```
. irtgraph tcc
```



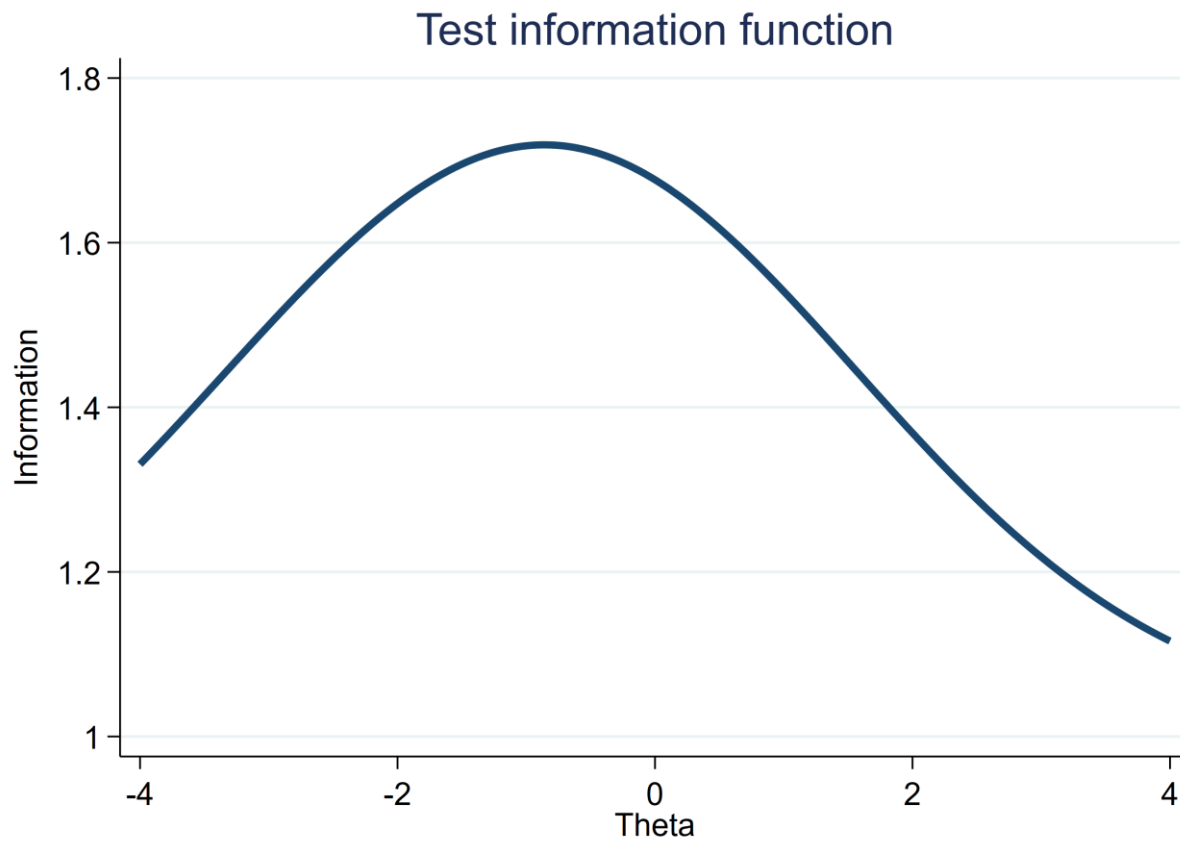
1PL model

```
. irtgraph iif item1-item6
```



1PL model

```
. irtgraph tif
```



2PL model

- We model the probability of respondent j answering item i correctly, $\pi_{ij} = \Pr(Y_{ij} = 1 | a_i, b_i, \theta_j)$

$$\pi_{ij} = \frac{\exp\{a_i(\theta_j - b_i)\}}{1 + \exp\{a_i(\theta_j - b_i)\}}$$

θ_j respondent j 's latent trait

a_i item i 's discrimination (scale)

b_i item i 's difficulty (location)

2PL model

```
. irt 2pl item1-item6  
. estat report, sort(a, desc)
```

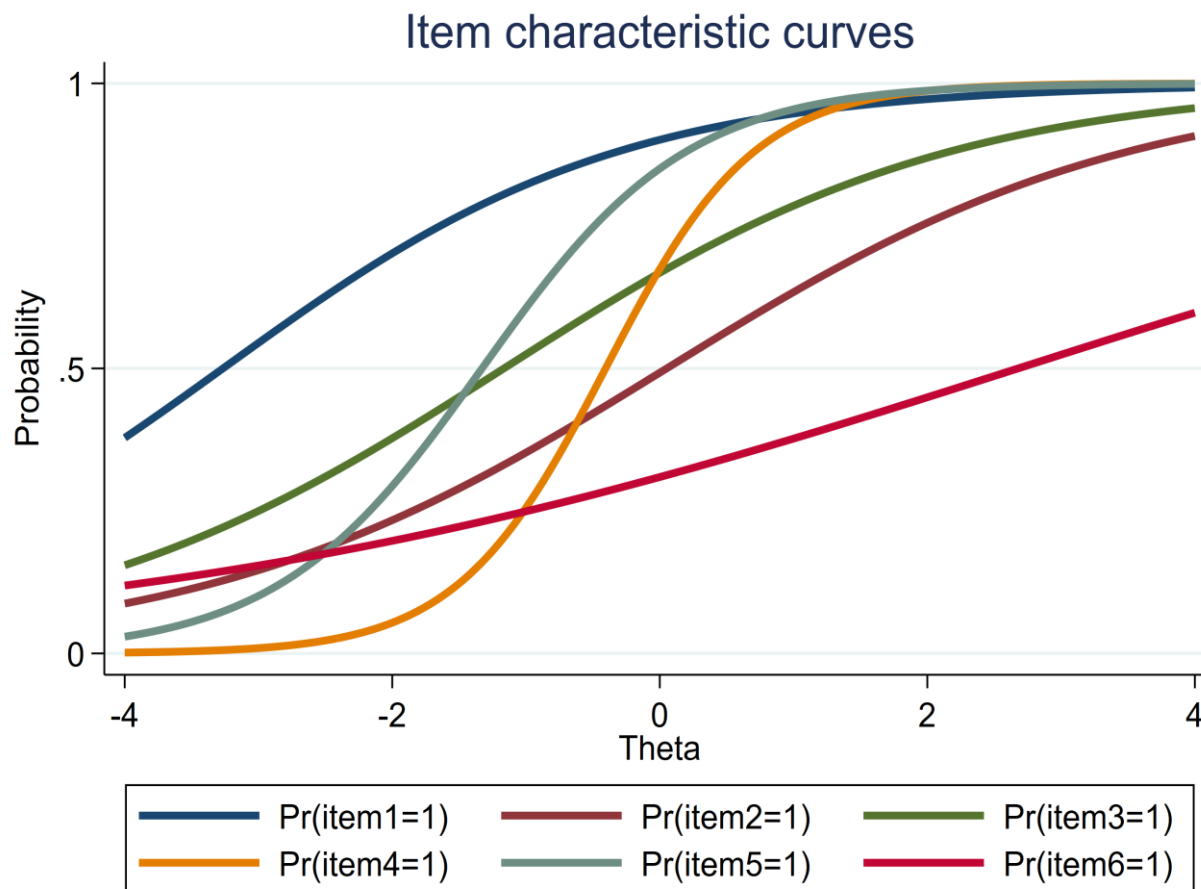
Two-parameter logistic model
Log likelihood = -3417.0828

Number of obs = 1,000

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
item4						
Discrim	1.79582	.3923366	4.58	0.000	1.026855	2.564786
Diff	-.4052651	.0679796	-5.96	0.000	-.5385028	-.2720275
item5						
Discrim	1.307999	.2342803	5.58	0.000	.8488179	1.76718
Diff	-1.329003	.1629475	-8.16	0.000	-1.648375	-1.009632
item1						
Discrim	.6764578	.1640356	4.12	0.000	.354954	.9979616
Diff	-3.265535	.6949543	-4.70	0.000	-4.62762	-1.90345
item3						
Discrim	.5989938	.1164829	5.14	0.000	.3706916	.8272961
Diff	-1.165294	.2287745	-5.09	0.000	-1.613684	-.7169041
item2						

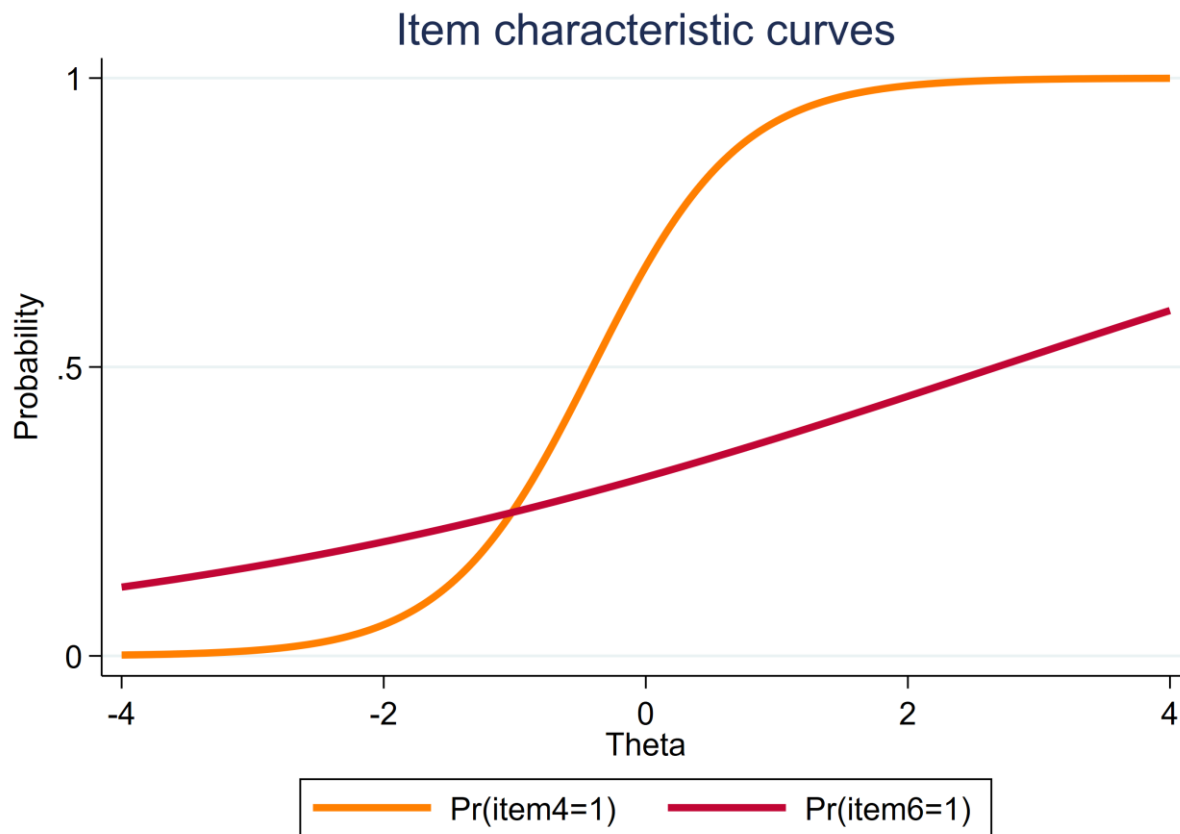
2PL model

```
. irtgraph icc item1-item6
```



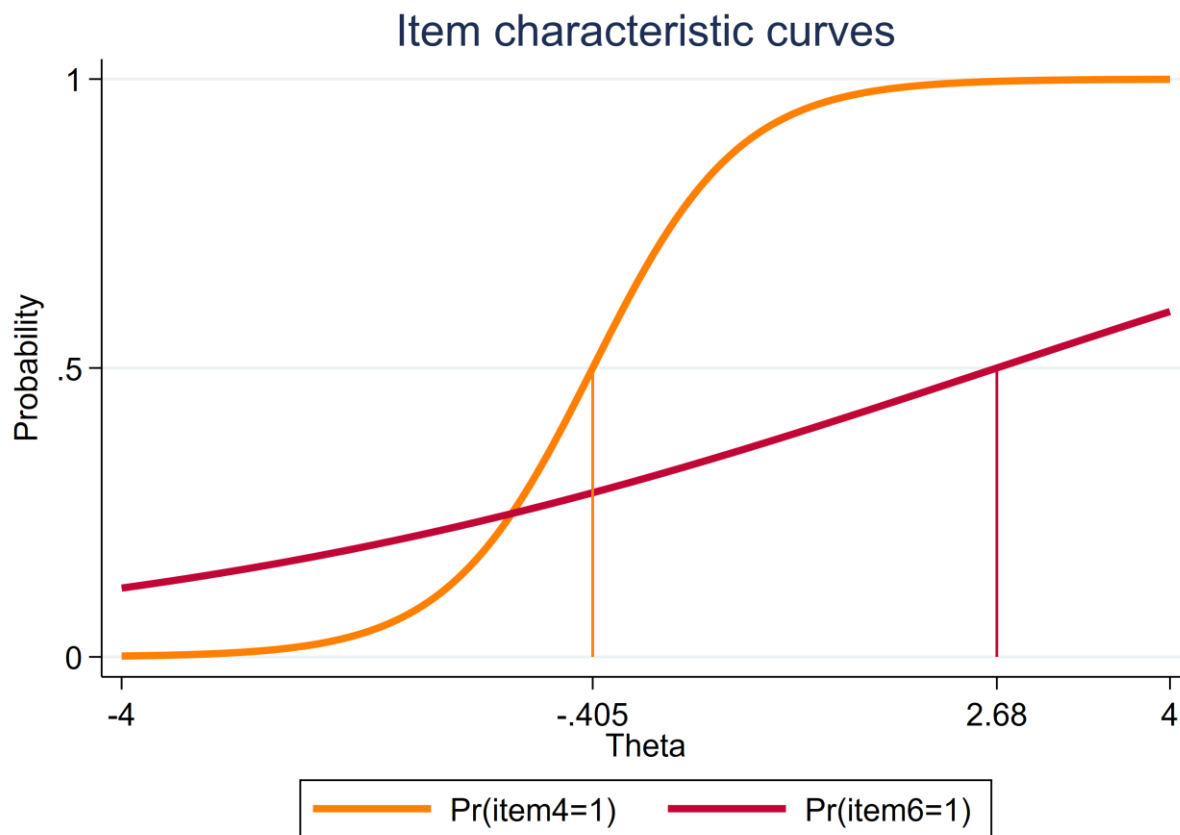
2PL model

```
. irtgraph icc item4 item6
```



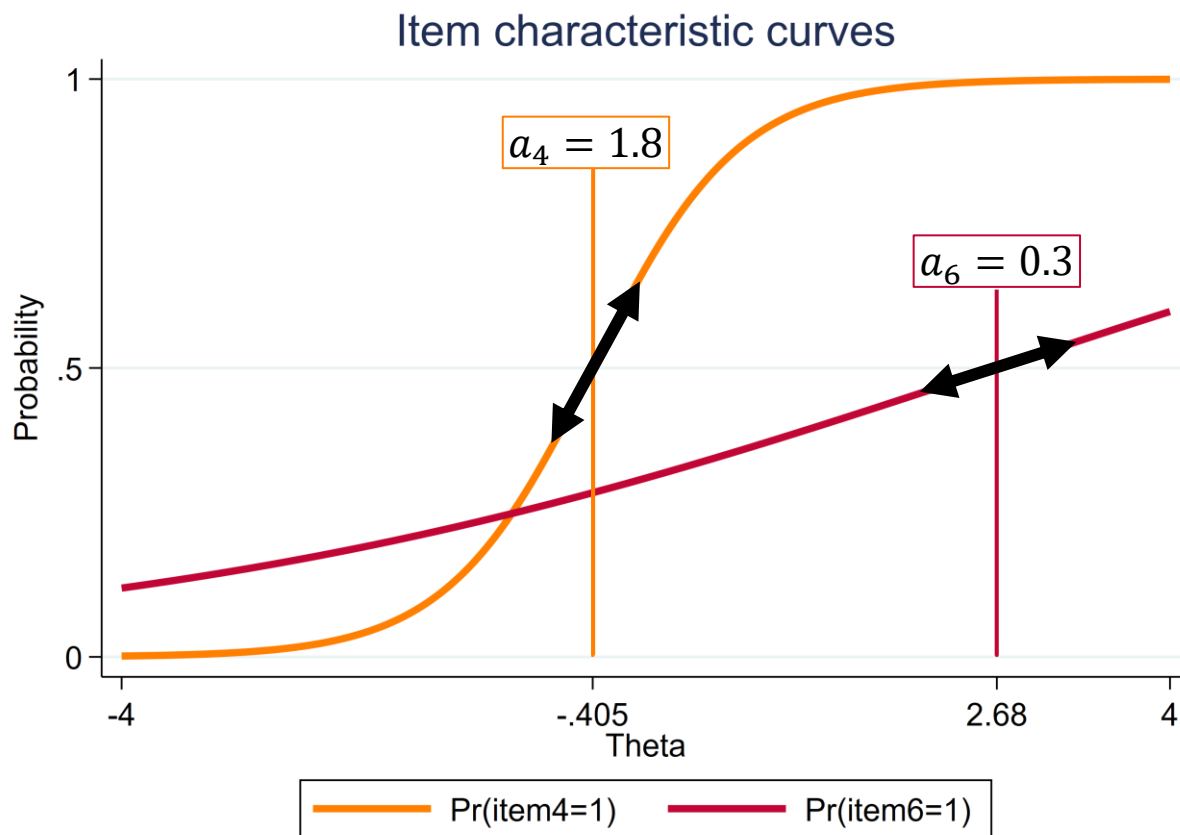
2PL model

```
. irtgraph icc item4 item6, bloc
```



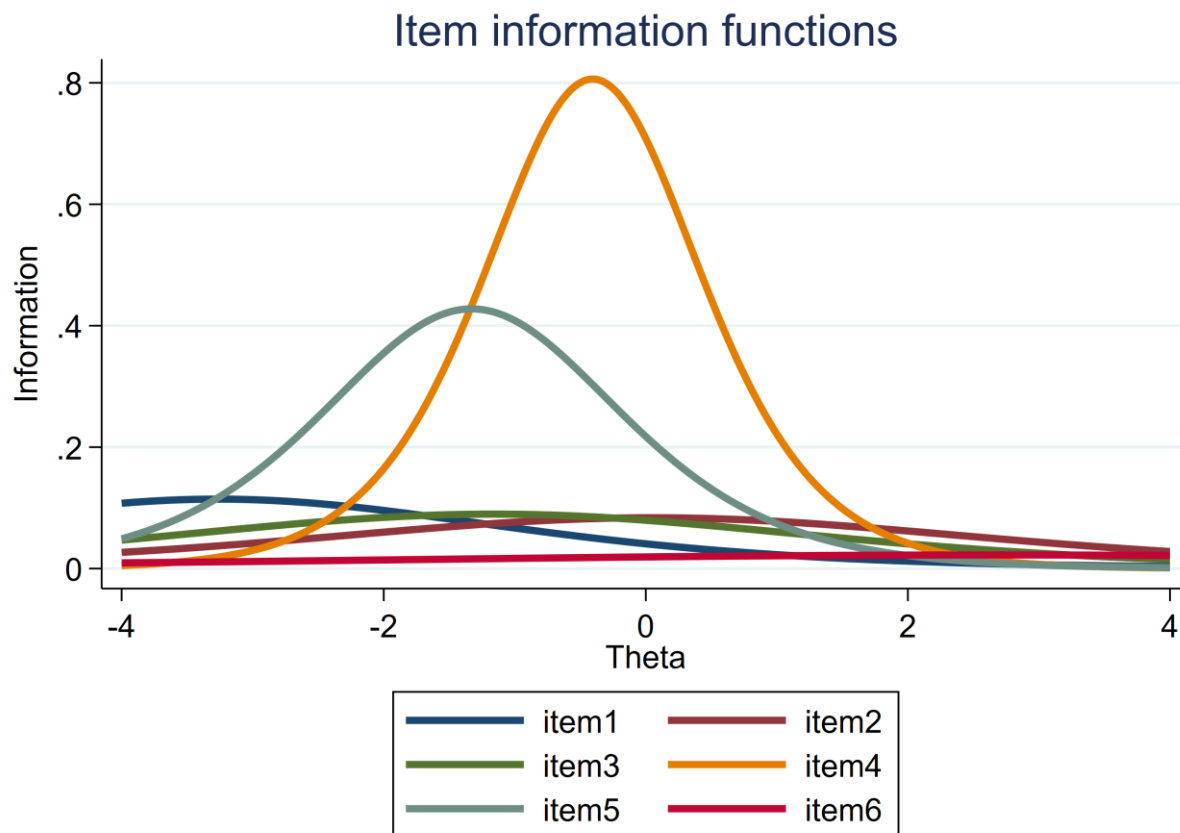
2PL model

```
. irtgraph icc item4 item6, bloc
```



2PL model

```
. irtgraph iif item1-item6
```



2PL model – math proficiency

```
. predict theta2, latent se(th2_se)  
. list score theta* th1_se th2_se in 1/10
```

	score	theta1	theta2	th1_se	th2_se
1.	1	-1.271147	-1.074278	.7720639	.6881211
2.	4	.1061837	.4044152	.7814902	.7424229
3.	2	-.814758	-.5683309	.7701293	.6808875
4.	5	.5830028	-.0614431	.7948129	.7022093
5.	5	.5830028	.7498873	.7948129	.778083
6.	4	.1061837	.2537876	.7814902	.7279438
7.	3	-.3576659	-.4900851	.7732583	.6822844
8.	4	.1061837	.4155966	.7814902	.7435357
9.	2	-.814758	-.758338	.7701293	.6804264
10.	4	.1061837	-.3497283	.7814902	.6865731

3PL model

- We model the probability of respondent j answering item i correctly, $\pi_{ij} = \Pr(Y_{ij} = 1 | a_i, b_i, c_i, \theta_j)$

$$\pi_{ij} = c_i + (1 - c_i) \frac{\exp\{a_i(\theta_j - b_i)\}}{1 + \exp\{a_i(\theta_j - b_i)\}}$$

θ_j respondent j 's latent trait

a_i item i 's discrimination

b_i item i 's difficulty

c_i item i 's pseudoguessing probability

3PL model

- We model the probability of respondent j answering item i correctly, $\pi_{ij} = \Pr(Y_{ij} = 1 | a_i, b_i, c_i, \theta_j)$

$$\pi_{ij} = c_i + (1 - c_i) \underbrace{\frac{\exp\{a_i(\theta_j - b_i)\}}{1 + \exp\{a_i(\theta_j - b_i)\}}}_{(0,1)}$$

θ_j respondent j 's latent trait

a_i item i 's discrimination

b_i item i 's difficulty

c_i item i 's pseudoguessing probability



3PL model

- Estimate c_i for each item

```
. irt 3pl item1-item6, sepguess
```

- Estimate constant c across all items

```
. irt 3pl item1-item6
```

- Constrain c to a particular probability

```
. irt 3pl item1-item6, cns(c@.25)
```

3PL model

```
. irt 3pl item1-item6, cns(c@.25)
. estat report, byparm
```

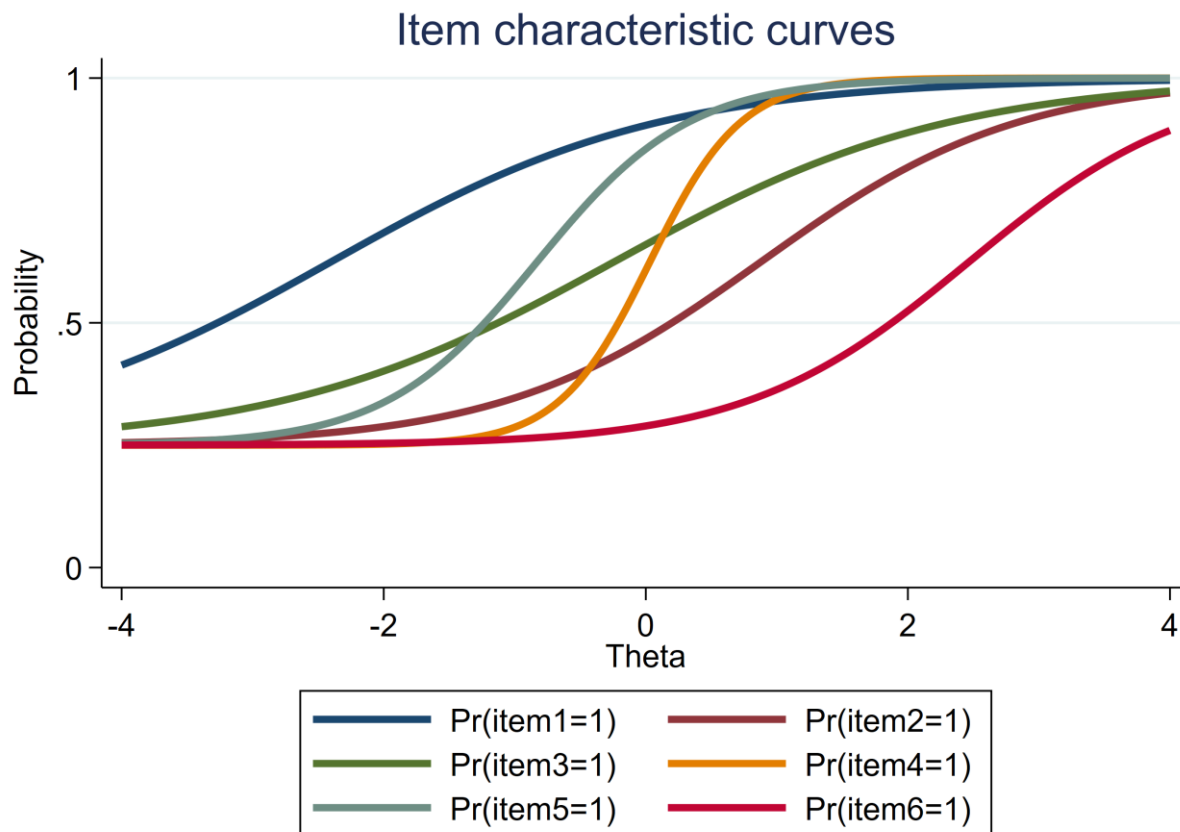
Three-parameter logistic model
Log likelihood = -3417.0697

Number of obs = 1,000

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
Discrim						
item1	.7974485	.1969278	4.05	0.000	.411477	1.18342
item2	1.015884	.2386842	4.26	0.000	.5480711	1.483696
item3	.7801923	.1651334	4.72	0.000	.4565368	1.103848
item4	2.872629	.8360044	3.44	0.001	1.23409	4.511167
item5	1.726012	.3938234	4.38	0.000	.9541319	2.497891
item6	1.172213	.6001308	1.95	0.051	-.004022	2.348448
Diff						
item1	-2.398791	.4992203	-4.81	0.000	-3.377245	-1.420338
item2	.8811249	.165145	5.34	0.000	.5574466	1.204803
item3	-.2388288	.1273723	-1.88	0.061	-.488474	.0108164
item4	.0319238	.0602117	0.53	0.596	-.086089	.1499366
item5	-.8279893	.1183433	-7.00	0.000	-1.059938	-.5960407
item6	2.468735	.7914398	3.12	0.002	.9175413	4.019928
Guess	.25 (constrained)					

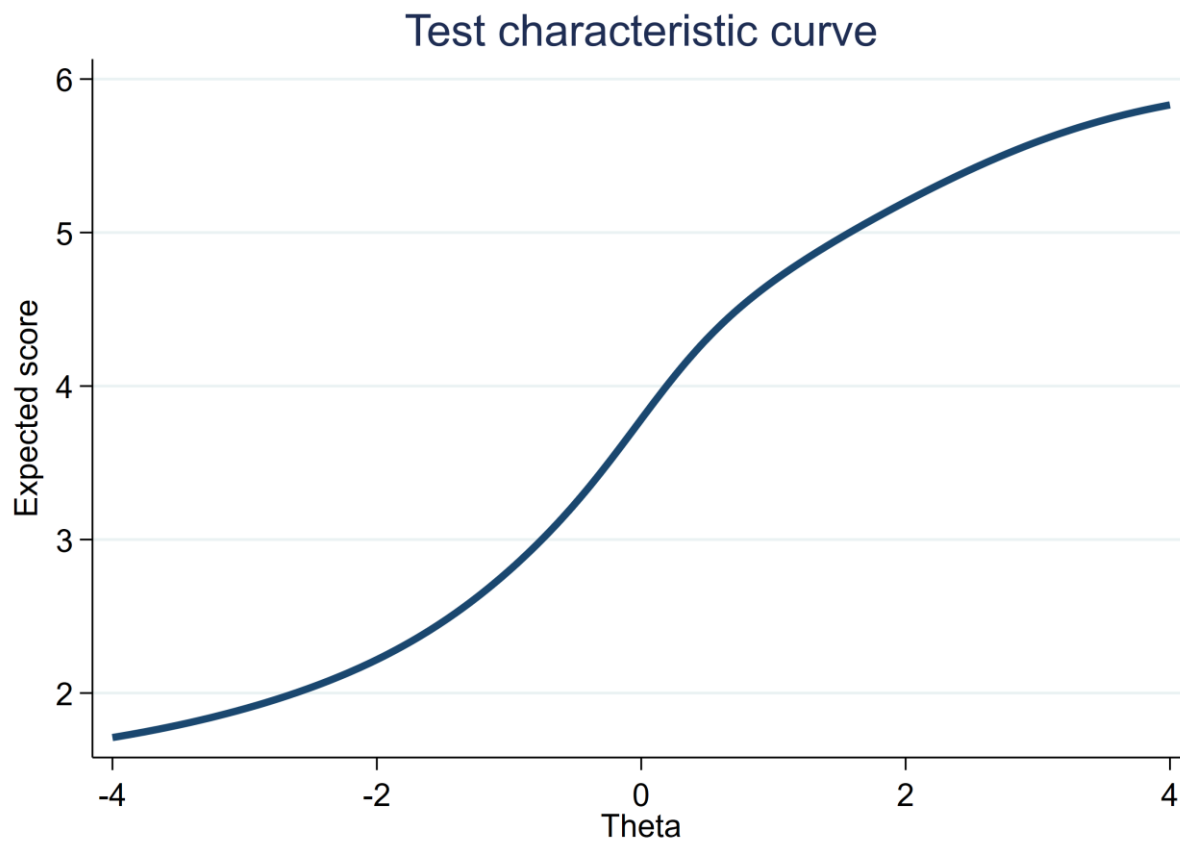
3PL model

```
. irtgraph icc item1-item6
```



3PL model

```
. irtgraph tcc
```



Model comparison

```
. irt 1pl item1-item6  
. estimates store _1pl
```

```
. irt 2pl item1-item6  
. estimates store _2pl
```

```
. irt 3pl item1-item6  
. estimates store _3pl
```

```
. irt 3pl item1-item6, cns(c@.25)  
. estimates store _c3pl
```

Model comparison

```
. lrtest _1pl _2pl
```

Likelihood-ratio test

Assumption: _1pl nested within _2pl

LR chi2(5) = 44.37

Prob > chi2 = 0.0000

```
. lrtest _2pl _3pl
```

Likelihood-ratio test

Assumption: _2pl nested within _3pl

LR chi2(1) = 0.77

Prob > chi2 = 0.3792

Model comparison

```
. estimates stats _2pl _c3pl
```

Akaike's information criterion and Bayesian information criterion

Model	N	ll(null)	ll(model)	df	AIC	BIC
<u>_2pl</u>	1,000	.	-3417.083	12	6858.166	6917.059
<u>_c3pl</u>	1,000	.	-3417.07	12	6858.139	6917.033

Differential item functioning

Journal of Educational Measurement
Spring 1990, Vol. 27, No. 1, pp. 67–81

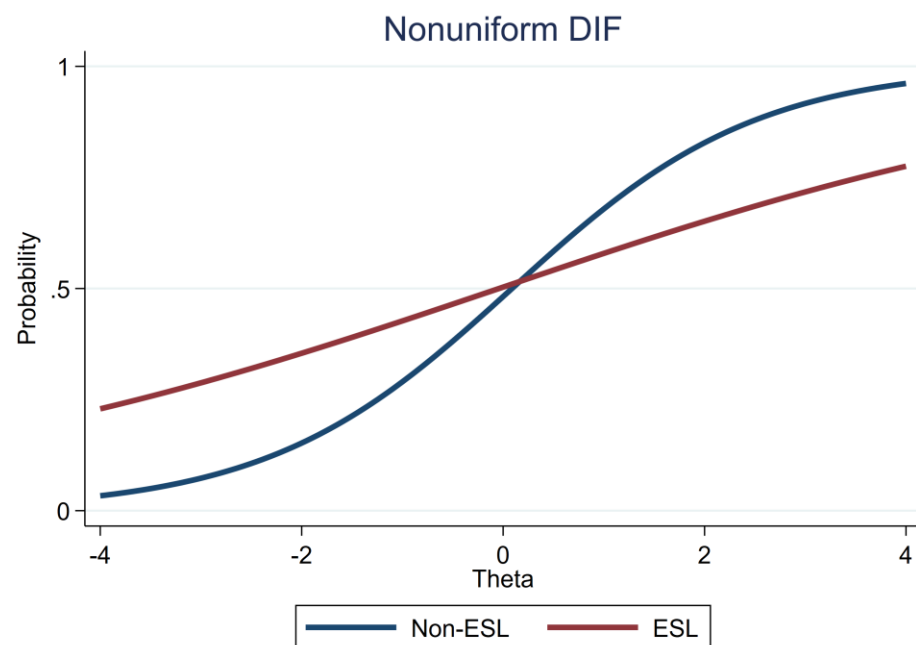
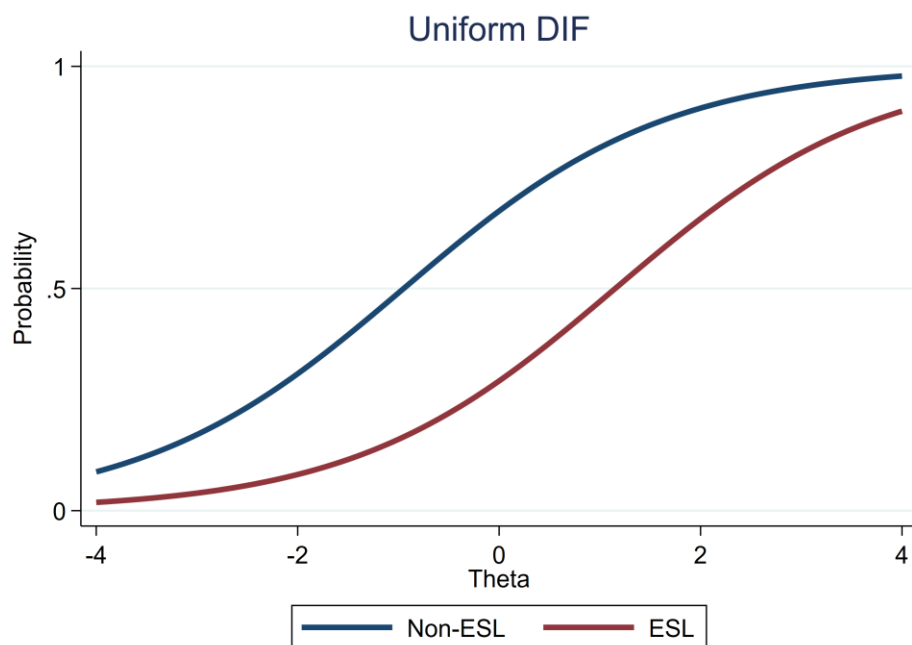
Differential Item Functioning for Minority Examinees on the SAT

Alicia P. Schmitt and Neil J. Dorans
Educational Testing Service

The standardization approach to assessing differential item functioning (DIF), including standardized distractor analysis, is described. The results of studies conducted on Asian Americans, Hispanics (Mexican Americans and Puerto Ricans), and Blacks on the Scholastic Aptitude Test (SAT) are described and then synthesized across studies. Where the groups were limited to include only examinees who spoke English as their best language, very few items across forms and ethnic groups exhibited large DIF. Major findings include evidence of differential speededness (where minority examinees did not complete SAT-Verbal sections at the same rate as White students with comparable SAT-Verbal scores) for Blacks and Hispanics and, when the item content is of special interest, advantages for the relevant ethnic group. In addition, homographs tend to disadvantage all three ethnic groups, but the effect of vertical relation-

Differential item functioning

- Probability profile differs by group



Differential item functioning

```
. difmh item1-item6, group(esl)
```

Mantel-Haenszel DIF analysis

Item	chi2	Prob.	Odds ratio	[95% conf. interval]	
item1	0.54	0.4626	1.2042	0.7823	1.8536
item2	1.08	0.2996	1.1815	0.8808	1.5849
item3	0.00	0.9500	0.9978	0.7376	1.3497
item4	0.09	0.7652	0.9377	0.6746	1.3035
item5	10.21	0.0014	0.5307	0.3632	0.7755
item6	1.61	0.2041	1.2381	0.9072	1.6896

Differential item functioning

```
. diflogistic item1-item6, group(es1)
```

Logistic regression DIF analysis

Item	Nonuniform		Uniform	
	chi2	prob.	chi2	prob.
item1	0.88	0.3493	0.73	0.3929
item2	4.14	0.0420	1.39	0.2387
item3	1.96	0.1610	0.00	0.9982
item4	0.16	0.6870	0.16	0.6920
item5	0.00	0.9904	11.28	0.0008
item6	1.24	0.2662	0.64	0.4248

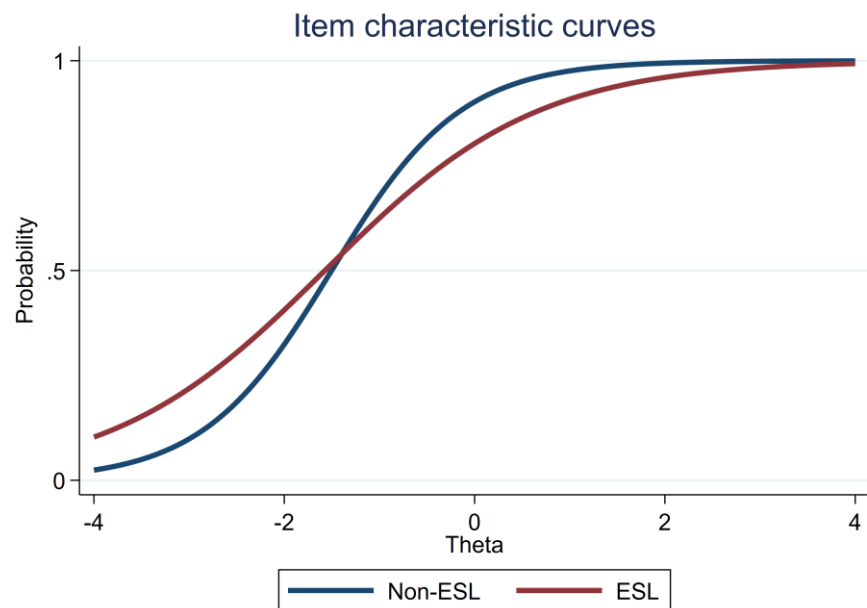
Group IRT – nonuniform DIF

```
. irt (0: 2pl item2 item5) (1: 2pl item2 item5) (2pl item1 item3 item4 item6), group(esl)  
. estat greport, byparm
```

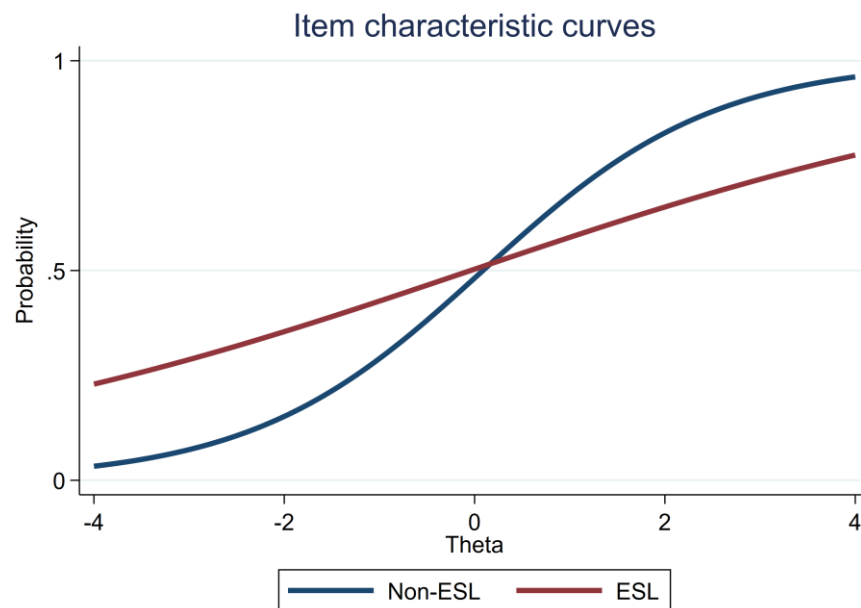
Parameter	0.esl	1.esl
Discrim		
item2	.82235935	.30631077
item5	1.4774125	.89161637
item1	.51966513	.51966513
item3	.50706535	.50706535
item4	1.4898238	1.4898238
item6	.23127462	.23127462
Diff		
item2	.08817955	-.04117464
item5	-1.5019333	-1.5726109
item1	-4.2175398	-4.2175398
item3	-1.3875508	-1.3875508
item4	-.49074734	-.49074734
item6	3.4549245	3.4549245
mean(Theta)	0	-.01294988
var(Theta)	1	1.961058

Nonuniform DIF

```
. irtgraph icc item5
```



```
. irtgraph icc item2
```



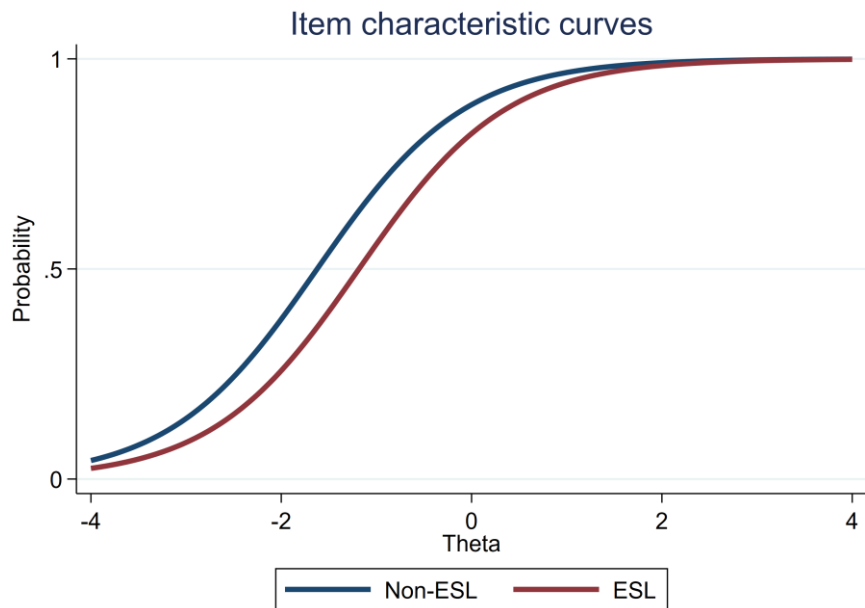
Group IRT – uniform DIF

```
. irt (0: 2pl item2, cns(a@a2)) (1: 2pl item2, cns(a@a2)) ///
>    (0: 2pl item5, cns(a@a5)) (1: 2pl item5, cns(a@a5)) ///
>    (2pl item1 item3 item4 item6), group(es1)
. estat greport, byparm
```

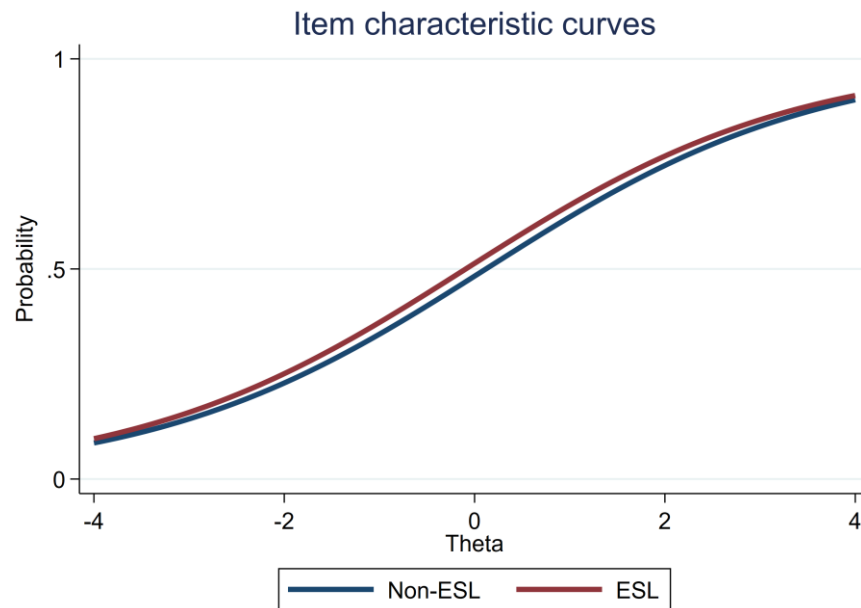
Parameter	0.es1	1.es1
Discrim		
item2	.57378099	.57378099
item5	1.29397	1.29397
item1	.66325384	.66325384
item3	.58816608	.58816608
item4	1.7458628	1.7458628
item6	.29324767	.29324767
Diff		
item2	.11924897	-.09271122
item5	-1.6228392	-1.1834184
item1	-3.3693264	-3.3693264
item3	-1.2255688	-1.2255688
item4	-.45311166	-.45311166
item6	2.6978003	2.6978003
mean(Theta)	0	-.07737261
var(Theta)	1	1.0776284

Uniform DIF

```
. irtgraph icc item5
```



```
. irtgraph icc item2
```





Charity example

- A survey measuring the public's faith and trust in charity organizations was administered to 949 Michigan citizens (Zheng & Rabe-Hesketh, 2007).
 1. Charitable organizations are more effective now in providing services than they were 5 years ago.
 2. I place a low degree of trust in charitable organizations.
 3. Most charitable organizations are honest and ethical in their use of donated funds.
 4. Generally, charitable organizations play a major role in making our communities better places to live.
 5. On the whole, charitable organizations do not do a very good job in helping those who need help.

Charity example

```
. use charity, clear  
. tabulate ta1
```

Charitable organizations more effective	Freq.	Percent	Cum.
Strongly agree	203	22.94	22.94
Agree	447	50.51	73.45
Disagree	177	20.00	93.45
Strongly disagree	58	6.55	100.00
Total	885	100.00	

Graded response model

- We model the probability of respondent j choosing outcome category k or higher for item i , $\pi_{ijk} = \Pr(Y_{ij} \geq k | \theta_j)$

$$\pi_{ijk} = \frac{\exp\{a_i(\theta_j - b_{ik})\}}{1 + \exp\{a_i(\theta_j - b_{ik})\}}$$

θ_j respondent j 's latent trait

a_i item i 's discrimination

b_{ik} the k th cutpoint for item i

Graded response model

```
. irt grm ta1-ta5
```

Graded response model

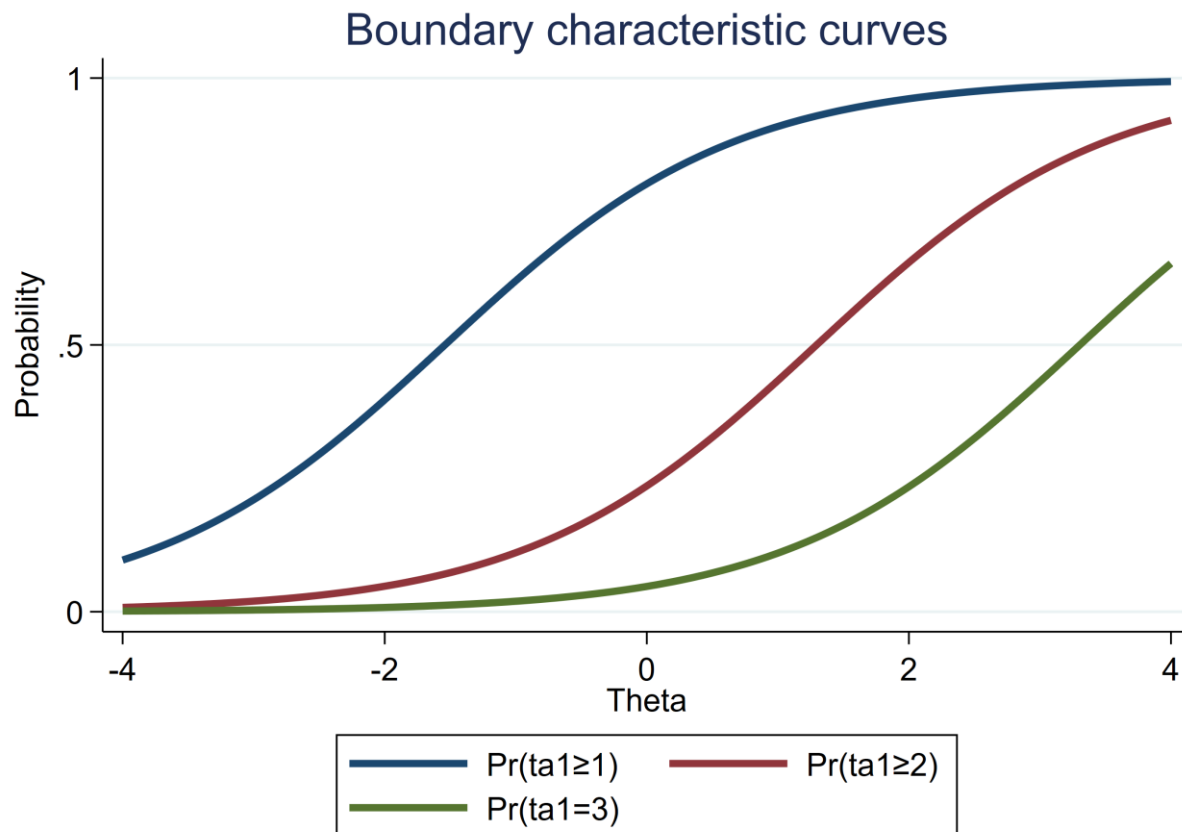
Number of obs = 945

Log likelihood = -5159.2791

	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
ta1						
Discrim	.907542	.0955772	9.50	0.000	.7202142	1.09487
Diff						
>=1	-1.540098	.1639425			-1.861419	-1.218776
>=2	1.296135	.1427535			1.016343	1.575927
=3	3.305059	.3248468			2.668371	3.941747
ta2						
Discrim	.9434675	.0967483	9.75	0.000	.7538444	1.133091
Diff						
>=1	-1.661331	.167878			-1.990366	-1.332296
>=2	.0068314	.082222			-.1543208	.1679836
=3	2.531091	.2412513			2.058247	3.003935
ta3						
Discrim	1.734201	.1554383	11.16	0.000	1.429548	2.038855
Diff						
>=1	-1.080079	.0835119			-1.243759	-.9163983
>=2	1.016567	.0796635			.8604297	1.172705

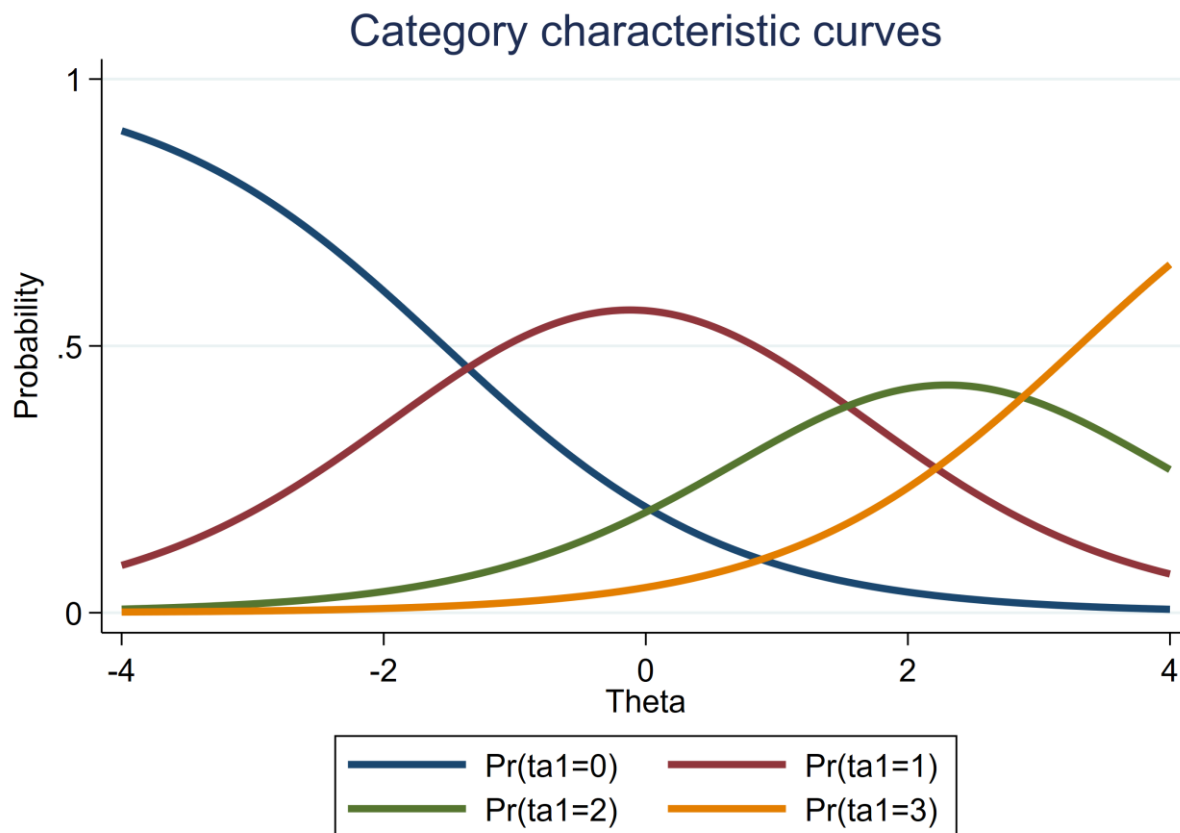
Graded response model

```
. irtgraph icc ta1, bcc
```



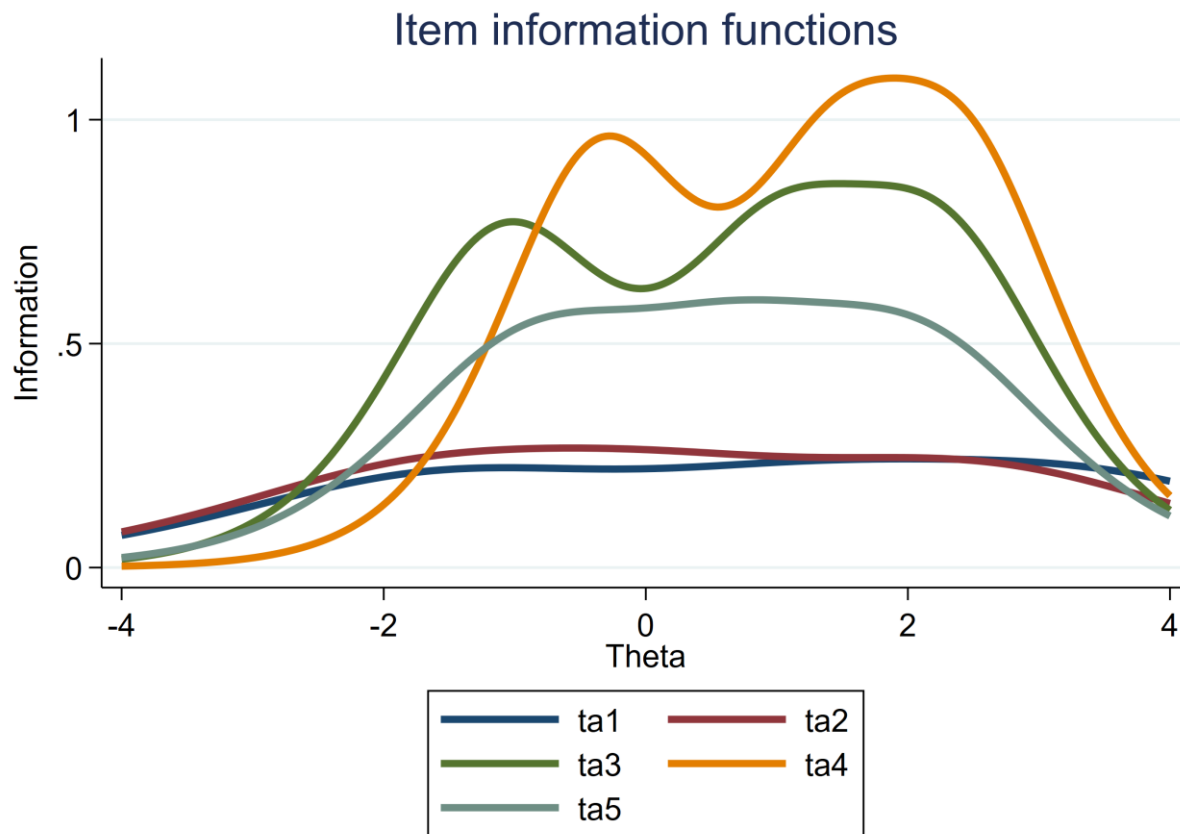
Graded response model

```
. irtgraph icc ta1
```



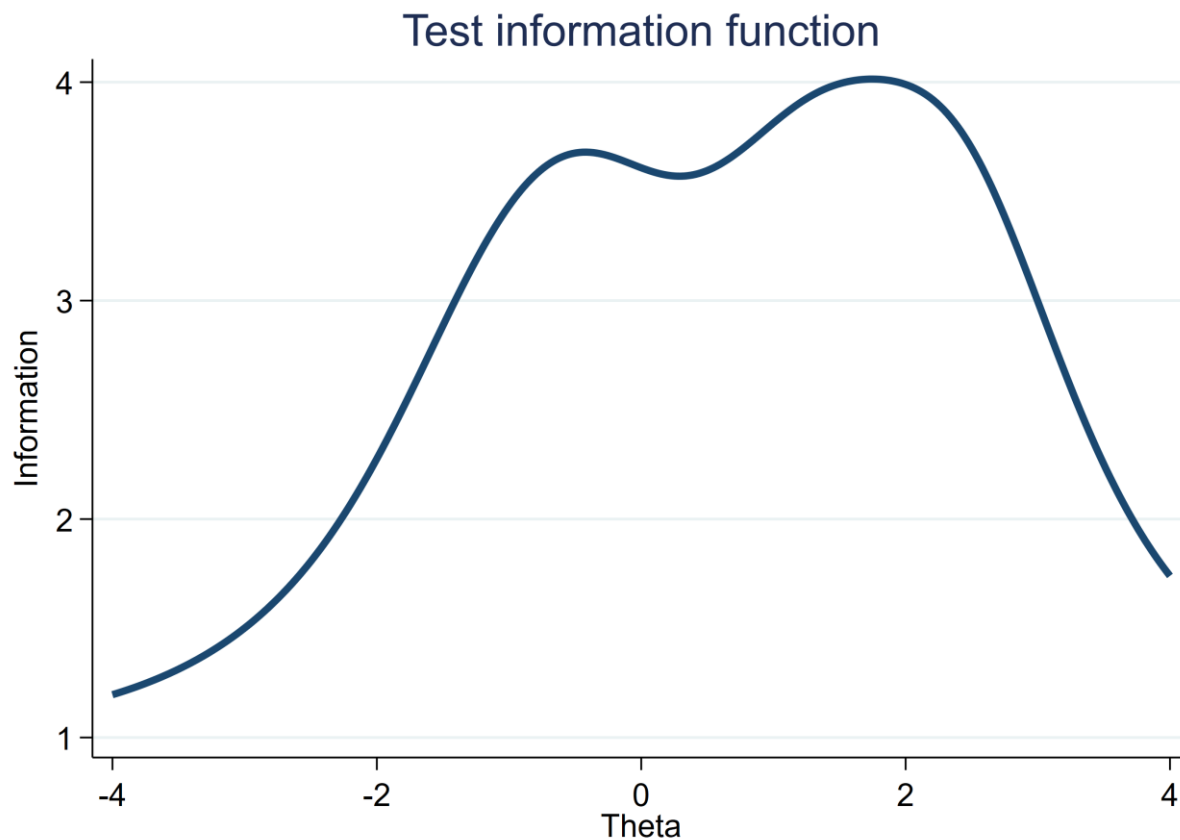
Graded response model

```
. irtgraph iif ta1-ta5
```



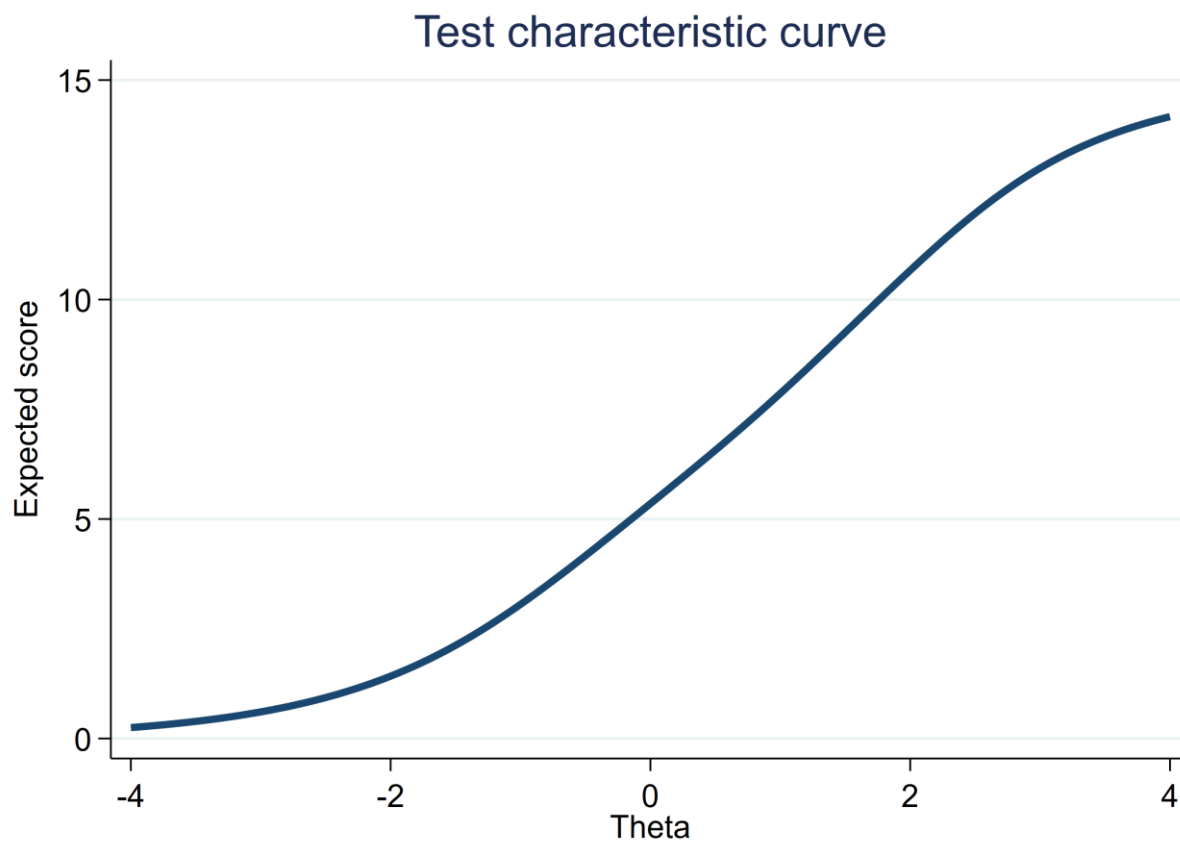
Graded response model

```
. irtgraph tif
```



Graded response model

```
. irtgraph tcc
```



More IRT models in Stata

irt - Item response theory (IRT) models

Start command log

Model

Report

Graph

DIF

Finish

Models

Binary item models

- ☐ One-parameter logistic model (1PL)
- ☐ Two-parameter logistic model (2PL)
- ☐ Three-parameter logistic model (3PL)

Ordered item models

- ☒ Graded response model (GRM)
- ☐ Partial credit model (PCM)
- ☐ Generalized partial credit model (GPCM)
- ☐ Rating scale model (RSM)

Unordered categorical item model

- ☐ Nominal response model (NRM)

Hybrid models

- ☐ Hybrid models

Items:

Group variable:

Fit model

Advanced options

Model: Graded response model

Close

Bayesian IRT

<https://blog.stata.com/2016/01/18/bayesian-binary-item-response-theory-models-using-bayesmh/>



[Home](#) > [Statistics](#) > Bayesian binary item response theory models using bayesmh

Bayesian binary item response theory models using bayesmh

18 January 2016 Nikolay Balov, Associate Director, Bayesian Statistics 12 Comments

Like 3

Tweet

This post was written jointly with Yulia Marchenko, Executive Director of Statistics, StataCorp.

Table of Contents

- Overview
- 1PL model
- 2PL model
- 3PL model
- 4PL model
- 5PL model
- Conclusion

Type text to search here...

RSS Twitter Facebook

Following 9,107

Subscribe to the Stata Blog

Receive email notifications of new blog posts

Name

Email Address*

Recent articles

- Just released from Stata Press:

Thank you!

You can download the dataset, do-file, and slides here:
<https://tinyurl.com/StataIRT22>

You can contact tech support at tech-support@stata.com