

Graphical Rasch models

Rasch models embedded in a multivariate graphical model

Definition of IRT and Rasch graphs

Tests of conditional independence

Parametrizing Rasch models

Testing fit of items to Rasch models

Estimating person parameters and assessment of measurement quality

Demo

What to do when fit is rejected

The CONF08 example

The data for this project originated in the Danish component of the European values studies in 2008.

We use five polytomous items measuring confidence in the following public institutions.

D) The police

E) The parliament

G) The social security system

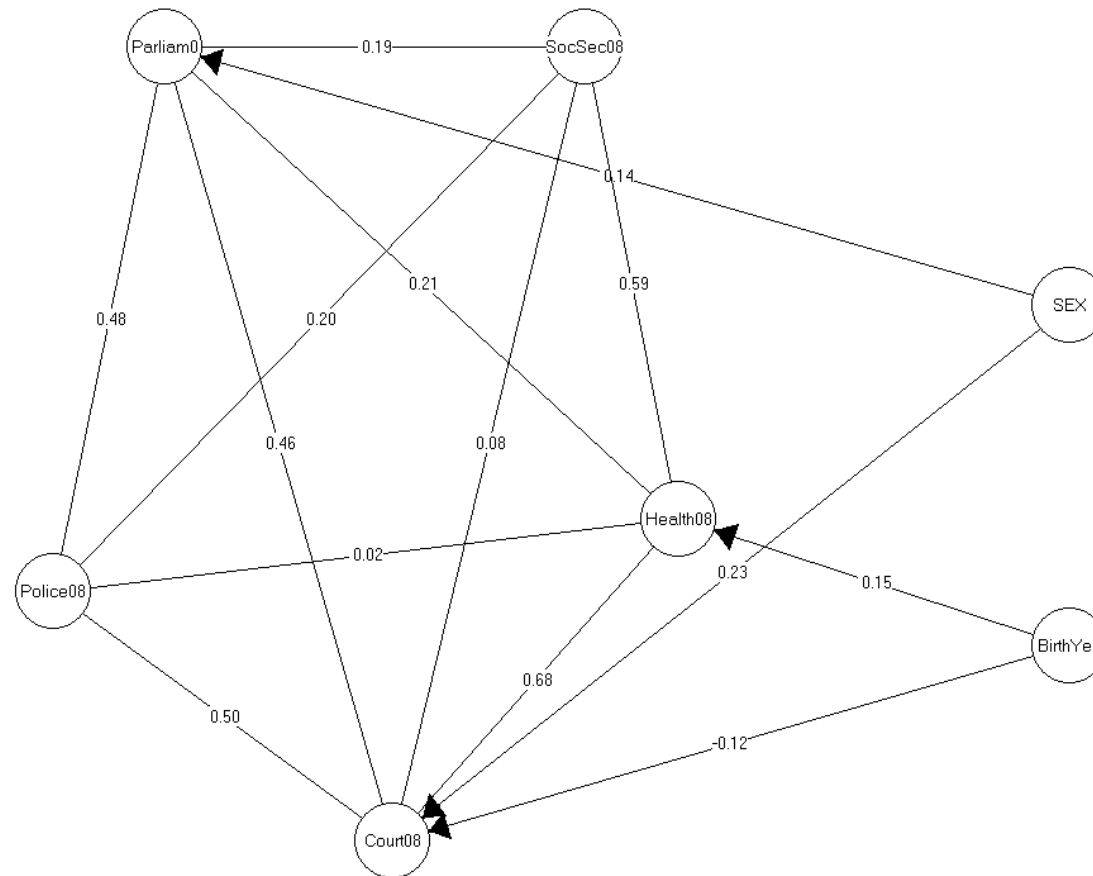
K) The health system

L) The courts.

Response categories were “High degree of confidence” (0), “Some degree of confidence” (1), “Little degree of confidence” (2) and “No confidence” (3). Responses were coded in such a way that a high item score indicate **a high degree of distrust**. We refer to the instrument as the CONF scale even though a high CONF score is an indication of lack of confidence of public institutions.

In addition to these variables, we have included information on sex (Y) and age (Z) to illustrate graphical Rasch models with exogenous items.

The *graphical* model defined by screening



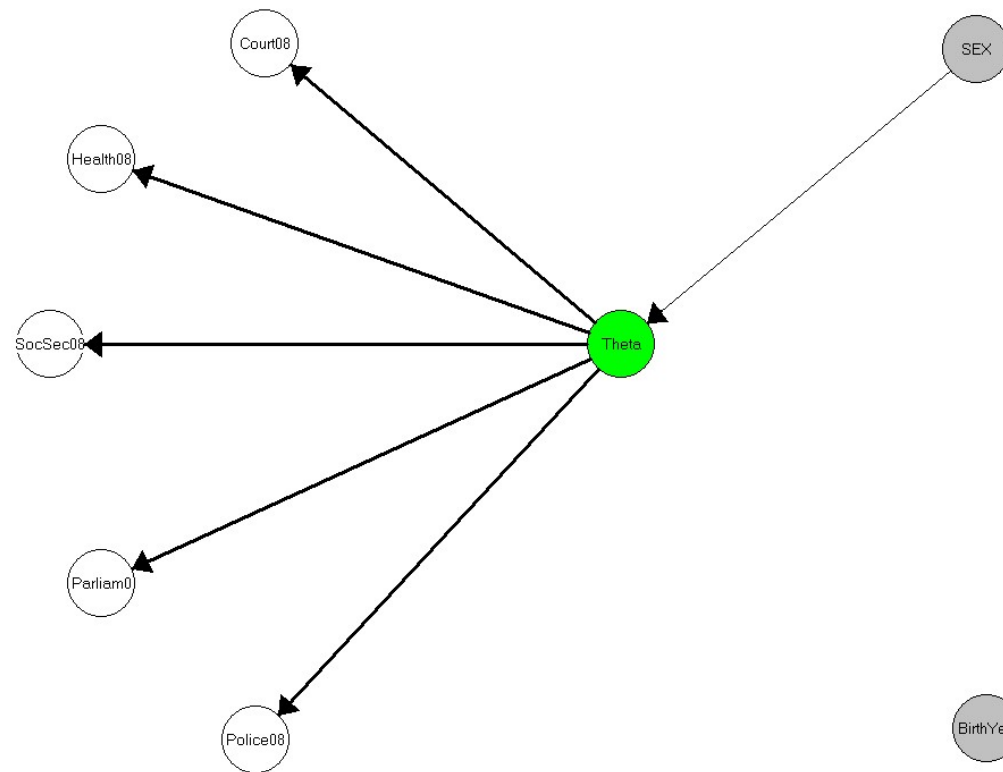
Can we simplify things by replacing the five questions concerning public institutions with a summery confidence score?

Yes, if items fit a Rasch model

The IRT graph

IRT and Rasch models assume that items depend on an unobserved latent variable T that explains the correlation among items and among items and exogenous variables.

Adding the latent variable to the model defines an extended graphical model.



The graphical Rasch model

$$\Pr(D,E,G,K,L \mid T=\theta,Y,Z) = \Pr(D \mid T=\theta) \Pr(E \mid T=\theta) \Pr(G \mid T=\theta) \Pr(K \mid T=\theta) \Pr(L \mid T=\theta)$$

The outcome of the latent variable (θ) appear as an unknown parameter in the conditional distribution of the items given θ

The IRT model is a Rasch model

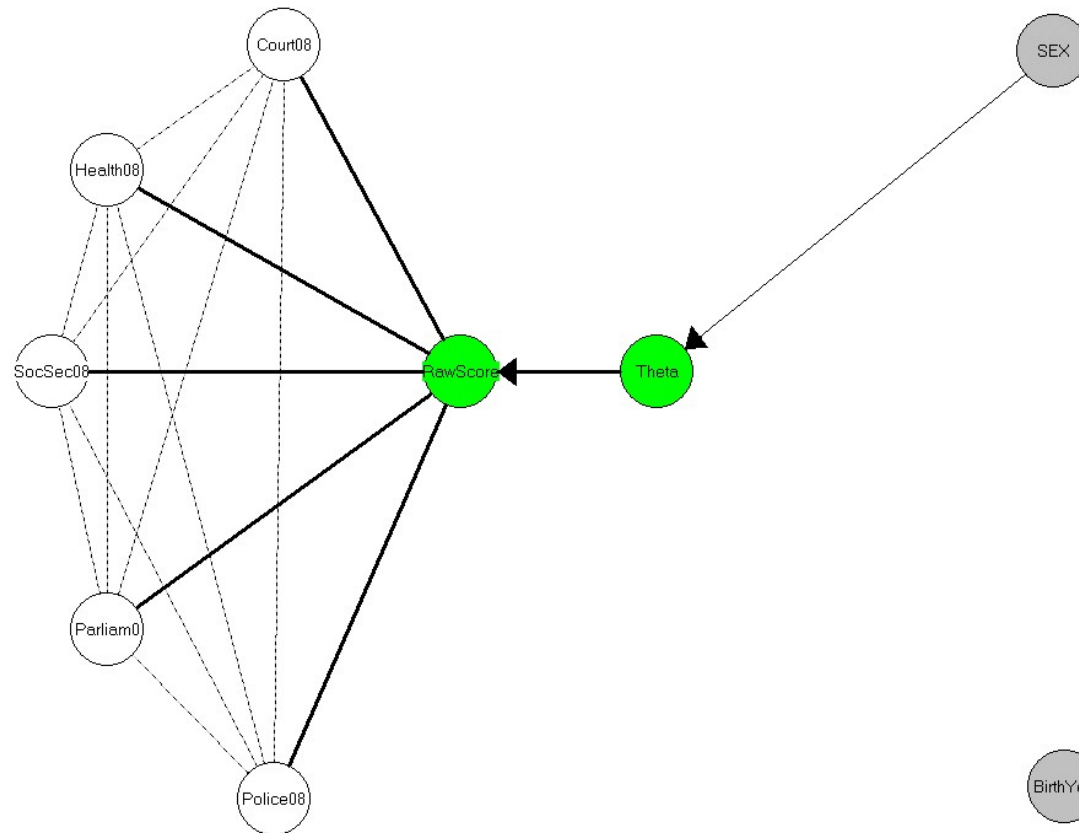
if and only if

the score over all items is sufficient for θ

This implies that items are conditionally independent of T given the score

Adding the score to the model generates a new graphical model.

The Rasch graph



This model implies that every item is conditionally independent of sex and age given the raw score (no DIF)

Item $\perp$ Sex | Score and Item $\perp$ Age | Score

Tests of no differential item functioning in DIGRAM

Analysis of DIF relative to Y: SEX

Scale : # - RawScore

Item	X ²	df	asyp	exact	gamma	asyp	exact	nsim	
D:Police08	28.2	19	0.080	0.071	-0.27	0.003	0.004	1000	--
E:Parliam0	30.4	21	0.084	0.075	0.20	0.023	0.022	1000	+
G:SocSec08	15.3	20	0.760	0.667	-0.06	0.489	0.571	21	
K:Health08	19.1	22	0.641	0.524	0.07	0.408	0.333	21	
L: Court08	29.8	19	0.055	0.036	0.03	0.723	0.725	1000	*

Analysis of DIF relative to Z: BirthYea

Scale : # - RawScore

Item	X ²	df	asyp	exact	gamma	asyp	exact	nsim	
D:Police08	76.4	73	0.370	0.524	-0.04	0.542	0.619	21	
E:Parliam0	109.5	81	0.019	0.018	-0.11	0.104	0.081	1000	*
G:SocSec08	87.5	78	0.216	0.216	0.11	0.101	0.068	1000	
K:Health08	109.5	85	0.038	0.032	0.12	0.058	0.059	1000	*
L: Court08	78.2	74	0.347	0.429	-0.07	0.347	0.286	21	

Parameterizing Rasch models

Let X be an item. $\Pr(X=x|\theta)$ can be parametrized in five different ways.

Exponential family **Partial credit version** Rasch's version

$$\Pr(X_i = x | \theta) = \frac{\exp(x\theta + \sigma_{ix})}{H_i(\theta)} = \frac{\exp(x\theta - \sum_{j=1}^x \beta_{ij})}{H_i(\theta)} = \frac{\exp(x(\theta + \sigma_i) + \alpha_{ix})}{H_i(\theta)}$$

Power series versions

$$\Pr(X_i = x | \xi = e^\theta) = \frac{\xi^x \delta_{ix}}{H_i(\theta)} = \frac{\xi^x \delta_i^x \omega_{ix}}{H_i(\theta)}$$

DIGRAM estimates ξ and δ_{ix} and calculates the parameters of the other versions

DIGRAM commands for Rasch analysis

Items DEGKL defines the items and calculate to total score

EXO YZ defines the exogenous variables

DIF calculates tests of differential item functioning

GRM invokes the Rasch dialog and calculates the estimates of the item parameters

After ITEMS DEGKL

The IRT and “Graphical Rasch models” buttons have been enabled and the GLLRM is a Rasch model.

The screenshot displays the DIGRAM software interface, version 5.01.01, dated January 22nd, 2023. The main window is titled "Project = CONF08" and shows the following details:

- Project: CONF08
- Path: C:\DIGRAM 5 - user
- Data file: CONF2.DAT
- New project generated by Conf9
- 560 variables in the dataset
- All cases will be used
- 7 project variables:
 - D Police08.. 4 ordinal categor
 - E Parliam0.. 4 ordinal categor
 - G SocSec08.. 4 ordinal categor
 - K Health08.. 4 ordinal categor
 - L Court08.. 4 ordinal categor
 - Y SEX..... binary
 - Z BirthYea.. 5 ordinal categor
- DEGKL <- YZ
- 980 cases in data set
- 496 cells in TAB data

Below the project details, the interface shows "1-sided tests" and "Rep. MC tests NSIM = 1000 SEED = 9".

The left panel displays the following statistics:

Item	Count	Mean	SD	Skewness	Missing
8	68	7.3	93.3		
9	39	4.2	97.4		
10	15	1.6	99.0		
11	6	0.6	99.7		
12	1	0.1	99.8		
13	1	0.1	99.9		
14	1	0.1	100.0		
Total	936	100.0			

Mean = 5.37
Variance = 4.33
s.d. = 2.08
skewness = 0.09
Missing = 44

Chronbach's Alpha = 0.698

Score groups for tests of Rasch models

New score groups by cutpoints.

Observed:
Minimum score = 0
maximum score = 14

ScoreGrp: 936 Cases

Score	Count	Percent	Cumulative
0-5	530	56.6	56.6
6-14	406	43.4	100.0
Total	936	100.0	
Missing	44		

The bottom panel contains several buttons for file operations and analysis:

- Output: Print, Erase, Read, Log on/off, Append, Save
- Project: Save model/graph, Open, Open project and run cmd file
- Command File: Edit file, Run
- Graph: Graph, IRT graph, Graphical Rasch models
- Tests: EXA, REP, Asymp

The status bar at the bottom shows: CONF08, No table, Items DEGKL, No exogenous.

Recall that the test of DIF provided som evidence of DIF

Analysis of DIF relative to Y: SEX

Scale : # - RawScore

Item	X ²	df	asyp	exact	gamma	asyp	exact	nsim	
D:Police08	28.2	19	0.080	0.071	-0.27	0.003	0.004	1000	--
E:Parliam0	30.4	21	0.084	0.075	0.20	0.023	0.022	1000	+
G:SocSec08	15.3	20	0.760	0.667	-0.06	0.489	0.571	21	
K:Health08	19.1	22	0.641	0.524	0.07	0.408	0.333	21	
L: Court08	29.8	19	0.055	0.036	0.03	0.723	0.725	1000	*

Analysis of DIF relative to Z: BirthYea

Scale : # - RawScore

Item	X ²	df	asyp	exact	gamma	asyp	exact	nsim	
D:Police08	76.4	73	0.370	0.524	-0.04	0.542	0.619	21	
E:Parliam0	109.5	81	0.019	0.018	-0.11	0.104	0.081	1000	*
G:SocSec08	87.5	78	0.216	0.216	0.11	0.101	0.068	1000	
K:Health08	109.5	85	0.038	0.032	0.12	0.058	0.059	1000	*
L: Court08	78.2	74	0.347	0.429	-0.07	0.347	0.286	21	

After the GRM command or a click on “Graphical Rasch models”

Graphical and log-linear Rasch models

Latent: **Theta**

Items

- D: Police08 (4)
- E: Parliam0 (4)
- G: SocSec08 (4)
- K: Health08 (4)
- L: Court08 (4)

Exogenous variables

Y: SEX(2)
Z: BirthYea(5)

Score groups

2 score groups
1: 0 - 5

1 14

Filter

The current model

Use current model The new model Change model

Options

- ☐ Reduce model
- ☐ Check local independence
- ☐ Check missing DIF
- ☐ Test of Global homogeneity
- ☒ Test of Local homogeneity
- ☐ Test of Global Invariance
- ☐ Test of Local Invariance
- ☐ Test Unidimensionality
- ☐ Item correlations
- ☐ Item fits
- ☐ Estimate person parameters
- ☐ Show margins of main model
- ☐ Show margins of all models
- ☐ Show estimates for all models
- ☐ Export data for latent regression
- ☐ Export data for ICC curves
- ☐ Include incomplete responses
- ☐ Test information and targeting
- ☐ Item-by-scores tables
- ☐ Not ready yet 1
- ☐ Extended info on estimates (not ready)

Start

Save current model

Test model terms:

☐ Extended output

☐ Save person estimates

☐ Output on files

Minimum number of iterative steps: 50 Erase Save and erase Save Exit

Convergence: 0.0001 Clear DIGRAM memo, save and erase

Item target = the person parameter where item information is maximized.
Midpoint = the person parameter where the expected score = max score/2

Locally independent items

Item	Midpoint	Local targets	Target	Info	(0.12)
D - Police08	1.340	Target = -2.357	Target = 2.535	Info = 0.372	(0.12)
E - Parliam0	-0.242	Target = -3.746	Target = 0.356	Info = 0.333	(0.11)
G - SocSec08	-0.405	Target = -4.059	Target = 0.104	Info = 0.324	(0.11)
K - Health08	-0.213	Target = -3.551	Target = 0.613	Info = 0.351	(0.12)
L - Court08	0.928	Target = -2.539	Target = 1.986	Info = 0.362	(0.12)

Nstep = 29 Delta = 0.0000754 14 estimated parameters

-ln(likelihood) = 2020.571
AIC = 4069.143
BIC = 4136.760
n = 925

Incomplete response patterns were not included during estimation of item parameters

Copy of GLLRM ready

Item parameters were estimated. Select options and click START to continue

```

+-----+
|
| Multiplicative score parameters |
|
+-----+

```

item	0	1	2	3
D: Police08	1.000	12.680	2.889	0.056
E: Parliam0	1.000	53.375	62.910	4.221
G: SocSec08	1.000	70.883	98.992	7.035
K: Health08	1.000	45.504	51.286	3.969
L: Court08	1.000	15.986	5.635	0.153

=====

```

+-----+
|
| PCM thresholds and locations |
|
+-----+

```

item	1	2	3	Location
D: Police08	-2.540	1.479	3.951	0.963
E: Parliam0	-3.977	-0.164	2.702	-0.480
G: SocSec08	-4.261	-0.334	2.644	-0.650
K: Health08	-3.818	-0.120	2.559	-0.459
L: Court08	-2.772	1.043	3.609	0.627

=====

```

+-----+
|
| Item and category effects |
|
+-----+

```

item	0	1	2	3	Item effect
D: Police08	0.000	3.503	2.987	-0.000	-0.963
E: Parliam0	0.000	3.497	3.182	-0.000	0.480
G: SocSec08	0.000	3.611	3.294	0.000	0.650
K: Health08	0.000	3.358	3.018	0.000	0.459
L: Court08	0.000	3.398	2.982	0.000	-0.627

---- MICE effects ----

D: Police08	1.000	33.222	19.835	1.000	0.382
E: Parliam0	1.000	33.027	24.087	1.000	1.616
G: SocSec08	1.000	36.993	26.962	1.000	1.916
K: Health08	1.000	28.741	20.459	1.000	1.583
L: Court08	1.000	29.914	19.730	1.000	0.534

```
=====
```

**What do these parameters tell us about the ways
that the CONF items function?**

Relevant item information

```

+-----+
|
| Midpoints and targets |
|
+-----+

```

Item target = the person parameter where item information is maximized.
Midpoint = the person parameter where the expected score = max score/2

D - Police08	Midpoint = 1.340	Target = 2.535	Info = 0.372	(0.12)
	Local targets:	Target = -2.357	Info = 0.269	
E - Parliam0	Midpoint = -0.242	Target = 0.356	Info = 0.333	(0.11)
	Local targets:	Target = -3.746	Info = 0.274	
G - SocSec08	Midpoint = -0.405	Target = 0.104	Info = 0.324	(0.11)
	Local targets:	Target = -4.059	Info = 0.271	
		Target = 1.890	Info = 0.315	
K - Health08	Midpoint = -0.213	Target = 0.613	Info = 0.351	(0.12)
	Local targets:	Target = -3.551	Info = 0.278	
L - Court08	Midpoint = 0.928	Target = 1.986	Info = 0.362	(0.12)
	Local targets:	Target = -2.539	Info = 0.274	

=====

Available options for analysis by Rasch models

Test global homogeneity and invariance

Item fits

Check missing DIF

Estimate Person parameters

(Export data for ICC curves)

Test info and targeting

Test global homogeneity and invariance

Erling B. Andersen's conditional likelihood ratio test comparing estimates of item parameters in different score groups and groups defined by outcomes on exogenous variables.

	CLR	df	p

scoregroups	17.6	14	0.227
Y: SEX	36.1	14	0.001
Z: BirthYea	99.9	56	0.000

Items appear to function in the same way for persons with low and high scores, but invariance relative to Sex and age is rejected.

Item fits

```
+-----+
|
| ** Conditional outfits and infits |
|
+-----+
```

Item	Outfit			Infit		
	observed	sd	p	observed	sd	p
D - Police08	0.988	0.058	0.84118	0.991	0.051	0.86451
E - Parliam0	1.012	0.054	0.82591	1.033	0.048	0.49962
G - SocSec08	0.997	0.053	0.95067	1.004	0.048	0.93122
K - Health08	0.995	0.054	0.92024	1.020	0.048	0.67338
L - Court08	0.889	0.057	0.05030	0.906	0.051	0.06568

```

+-----+
|
| Item restscore association |
|
+-----+

```

Item	Item-restscore		gamma	p
	observed	expected	sd	
D - Police08	0.543	0.531	0.035	0.74595
E - Parliam0	0.531	0.539	0.033	0.81189
G - SocSec08	0.528	0.531	0.034	0.91228
K - Health08	0.556	0.541	0.033	0.65789
L - Court08	0.621	0.542	0.034	0.02031

Critical levels adjusted by the Benjamini-Hochberg:

* < 5 % FDR, ** < 1 % FDR, *** = FDR < 0.1 % FDR

Item fit is accepted

Check missing DIF

Conditional likelihood ratio tests that the comparing item parameters of on item in different groups under the assumption that item parameters are the same for the other items

Check assumptions of no DIF

Test results will only be shown if $p \leq 0.05$.

Select extended output if you want to see all testresults.

Significant test results:

D & Y:	lr =	17.69	df =	3	p =	0.0005
E & Y:	lr =	10.67	df =	3	p =	0.0137
E & Z:	lr =	27.49	df =	12	p =	0.0066
K & Z:	lr =	21.43	df =	12	p =	0.0444

Benjamini & Hochberg rejects at 0.01500

DIF: **DY EY EZ**

Estimate Person parameters

ML estimates and WML estimates

Includes estimates of bias and standard errors (of measurement)

```
+-----+
|       |
| ML estimates |
|       |
+-----+
```

Pseudo ML estimates for extreme scores:

```
Score = 0      theta = -6.613
Score = 15     theta = 6.202
```

Properties of ML estimates

Score	Theta estimate	True score	Bias	RMSE	Score SEM
0	-6.613	0.25	0.386	0.824	0.48
1	-5.014	1.00	-0.194	1.112	0.88
2	-3.986	2.00	-0.127	1.077	1.07
3	-3.168	3.00	-0.035	0.967	1.13
4	-2.376	4.00	-0.001	0.900	1.12
5	-1.566	5.00	-0.019	0.867	1.11
6	-0.786	6.00	-0.041	0.853	1.16
7	-0.079	7.00	-0.040	0.835	1.22
8	0.570	8.00	-0.029	0.810	1.26
9	1.184	9.00	-0.016	0.793	1.29
10	1.777	10.00	0.000	0.795	1.30
11	2.365	11.00	0.030	0.827	1.30
12	2.978	12.00	0.084	0.906	1.25
13	3.675	13.00	0.170	1.032	1.13
14	4.633	14.00	0.217	1.079	0.89
15	6.202	14.75	-0.378	0.806	0.49

```

+-----+
|       |
| WML estimates |
|       |
+-----+

```

Weighted ML estimates

Score	Theta estimate	True score	Bias	RMSE	Score SEM
0	-6.090	0.40	0.519	0.902	0.60
1	-4.732	1.23	0.063	0.942	0.94
2	-3.897	2.10	0.015	0.922	1.08
3	-3.158	3.01	0.015	0.894	1.13
4	-2.393	3.98	0.014	0.890	1.12
5	-1.545	5.03	0.002	0.883	1.11
6	-0.732	6.07	-0.007	0.860	1.16
7	-0.038	7.06	-0.004	0.826	1.22
8	0.597	8.04	-0.001	0.794	1.26
9	1.201	9.03	-0.002	0.773	1.29
10	1.782	10.01	-0.004	0.765	1.30
11	2.348	10.97	-0.005	0.772	1.30
12	2.922	11.91	-0.005	0.802	1.26
13	3.547	12.83	-0.010	0.860	1.16
14	4.322	13.73	-0.063	0.906	0.98
15	5.659	14.59	-0.518	0.892	0.61

```

-----

Test midpoint    =    0.251
Test target      =    1.937    Test information =    1.705

-----

```

The distribution of the person parameter and reliability

```
+-----+
|
| Monte Carlo estimates of Reliability |
|
+-----+
```

Normal trait distribution mean = -1.33 s.d.= 1.30

Expected True Score = 5.37 Variance = 3.07

Observed mean Score = 5.37 Variance = 4.39

Ratio = 0.70

Test-retest correlation = 0.69

Test-true score correlation = 0.84

Test information and targeting

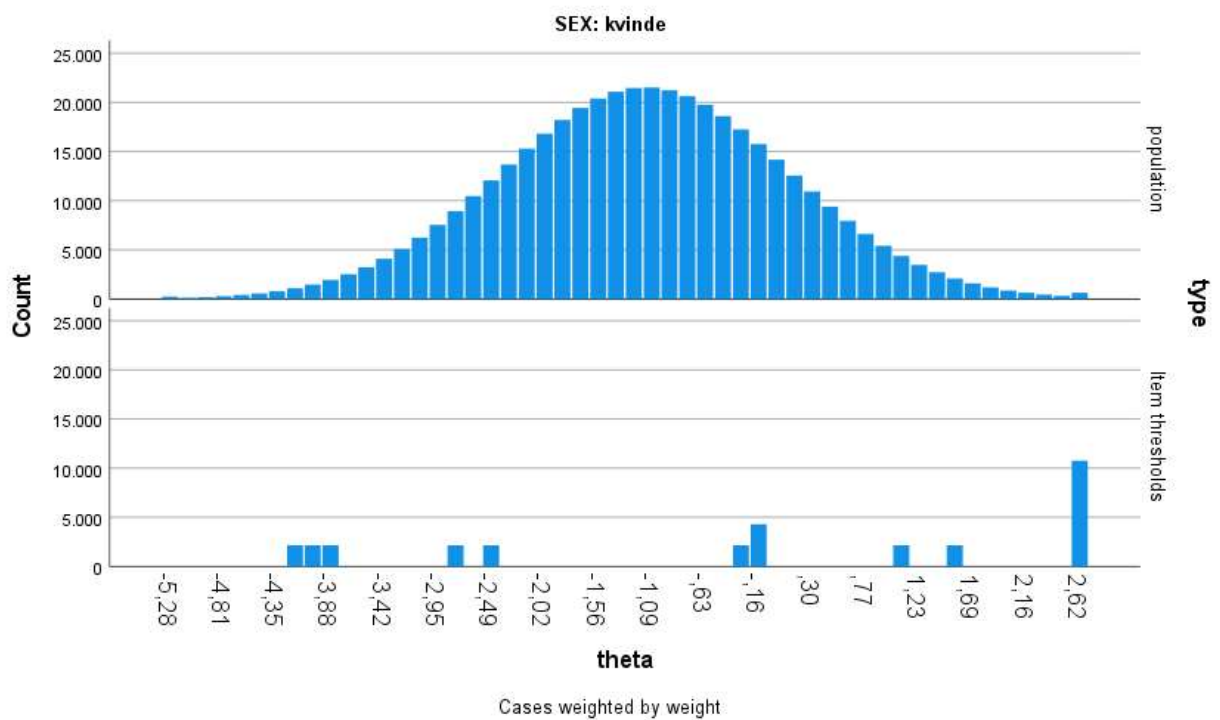
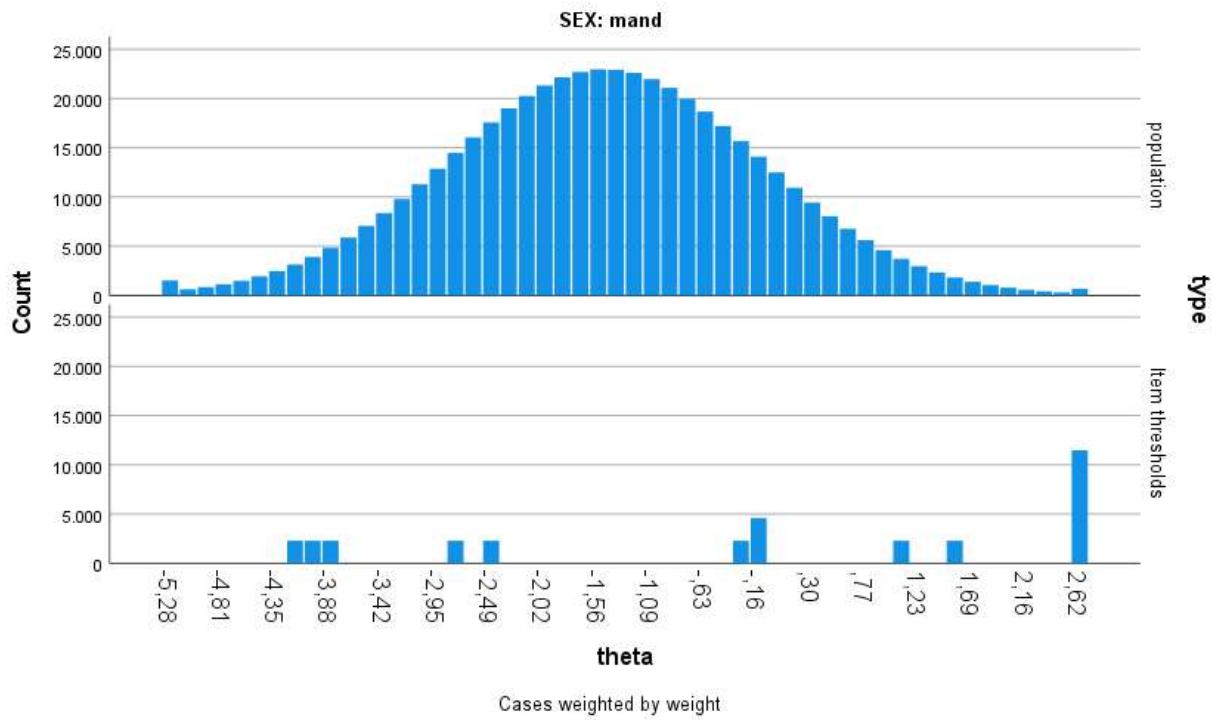
```
+-----+
|
| Targeting and test information |
|
+-----+
```

SEX	Target	n	theta		test info		target index	RMSE (WML)		target index	PSI
			Mean	sd	Mean	max		Mean	min		
mand	1.94	502	-1.50	1.35	1.317	1.705	0.772	0.869	0.766	0.881	0.728
kvinde	1.94	434	-1.15	1.24	1.348	1.705	0.790	0.861	0.766	0.890	0.697

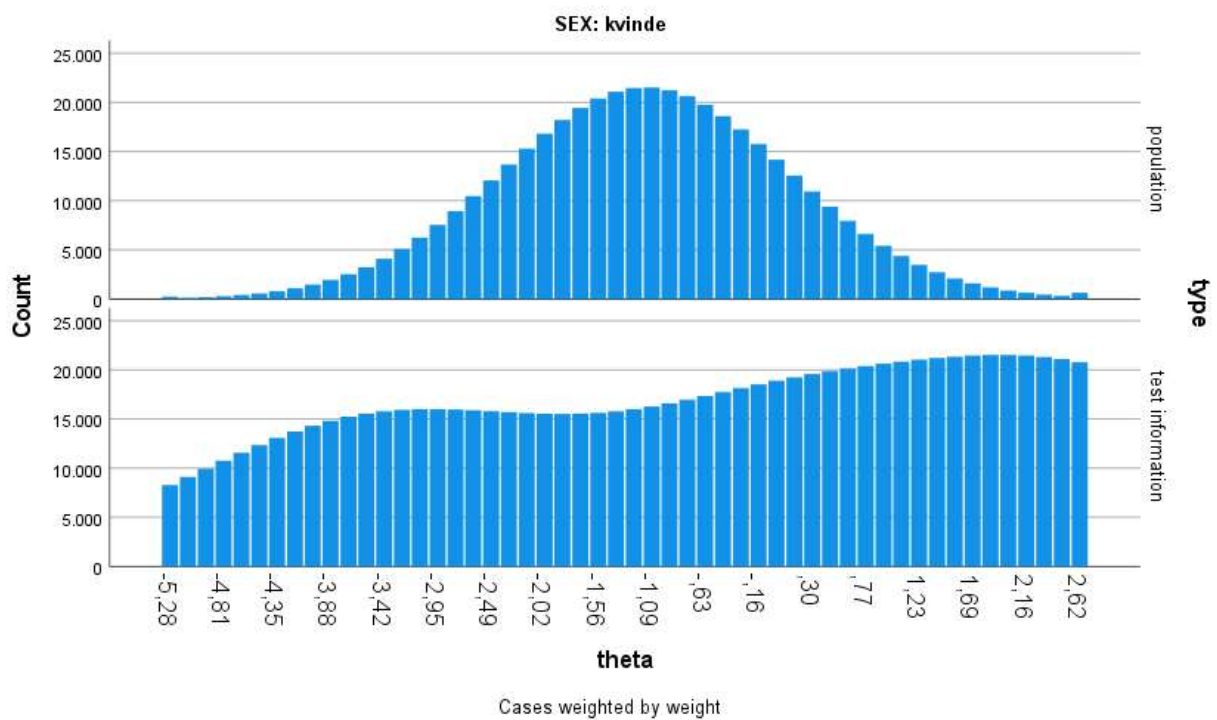
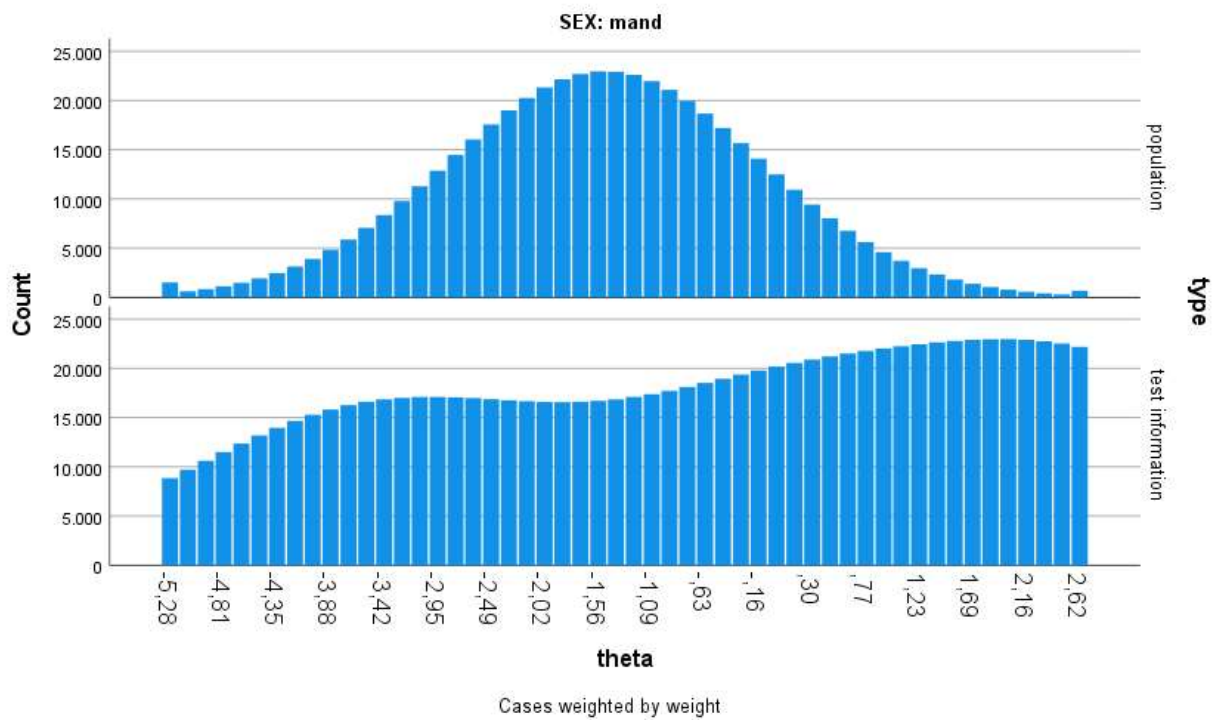
```
+-----+
|
| True score estimation and reliability |
|
+-----+
```

SEX	Target	n	Score		reliability	separation prob	no sep. prob	True score SEM	
			Mean	sd				Target	Mean
mand	10.27	502	5.16	2.11	0.693	0.743	0.136	1.31	1.15
kvinde	10.27	434	5.61	2.03	0.664	0.726	0.143	1.31	1.16

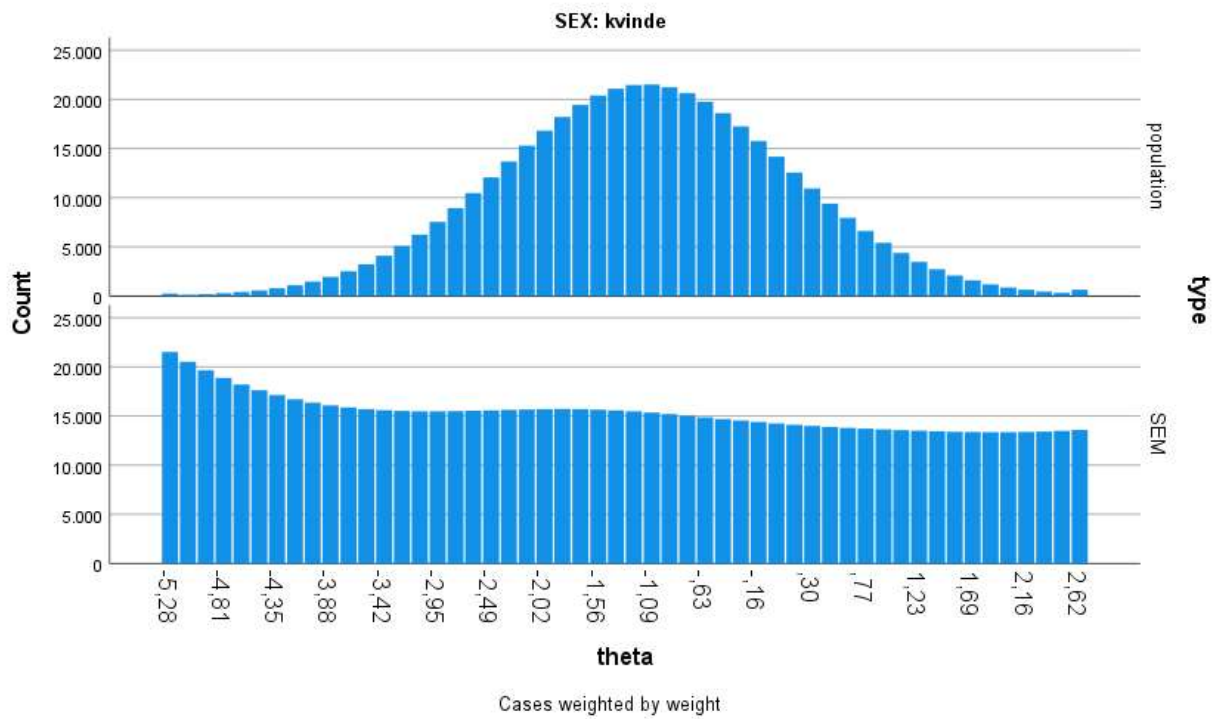
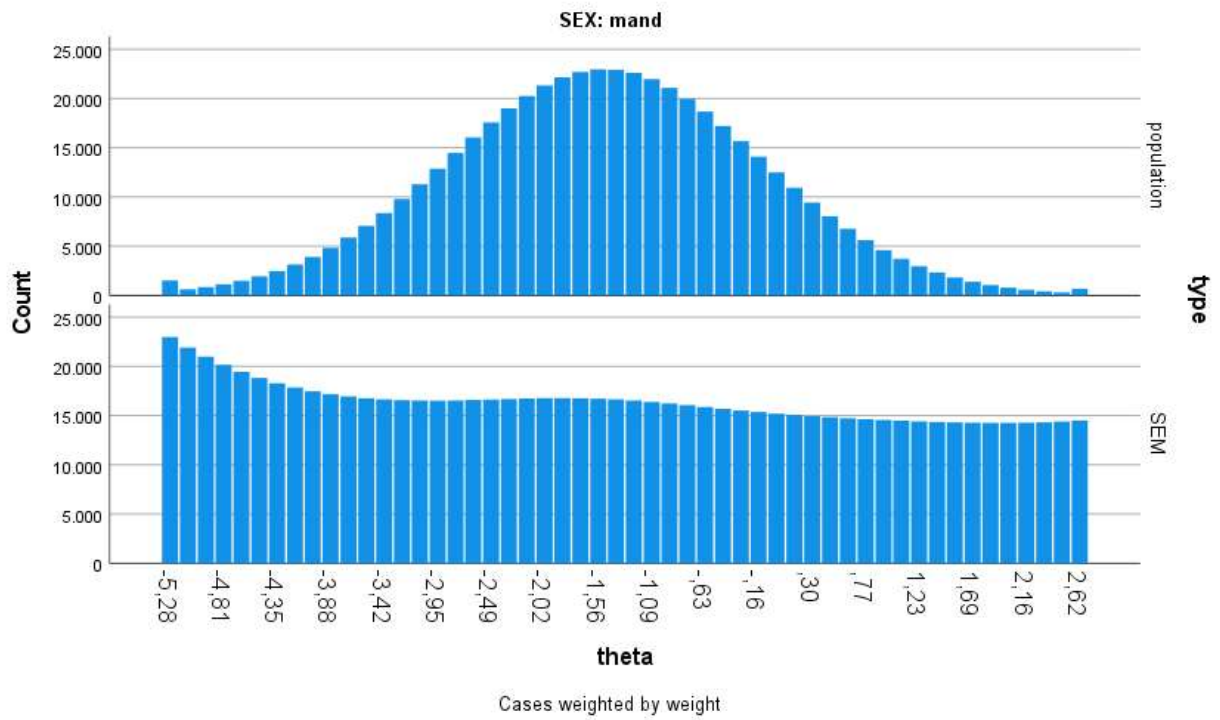
Item Maps



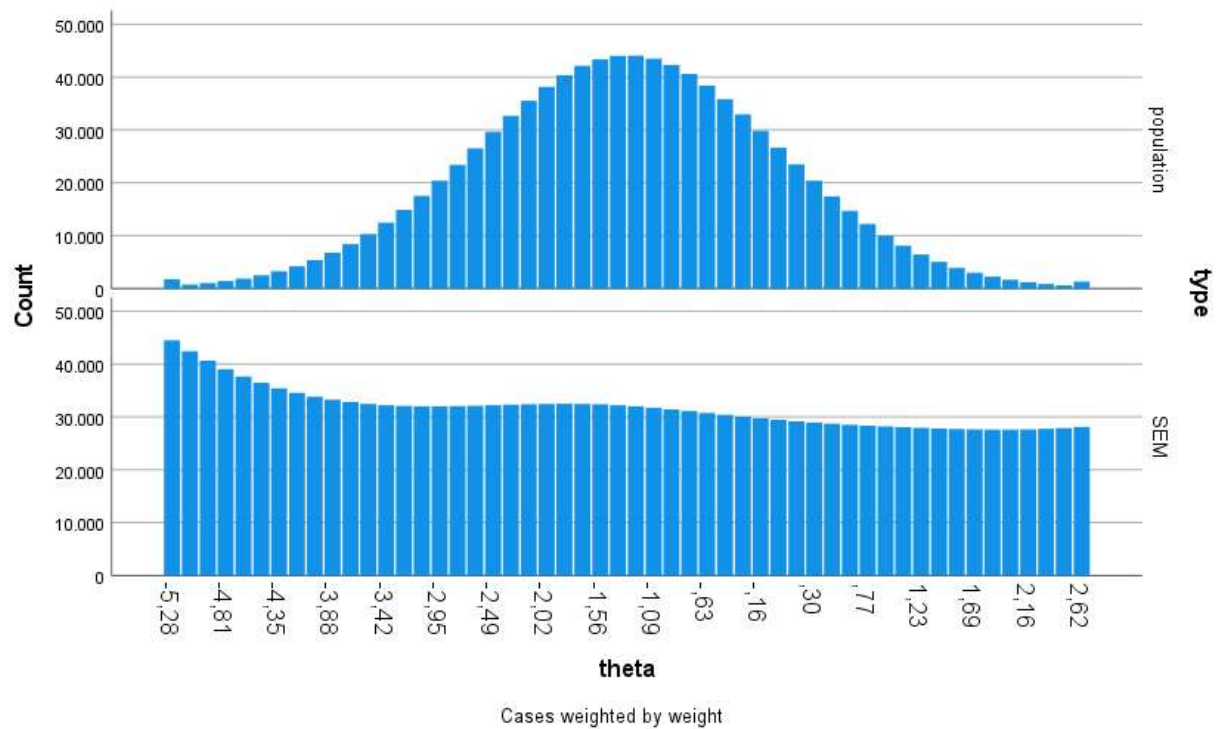
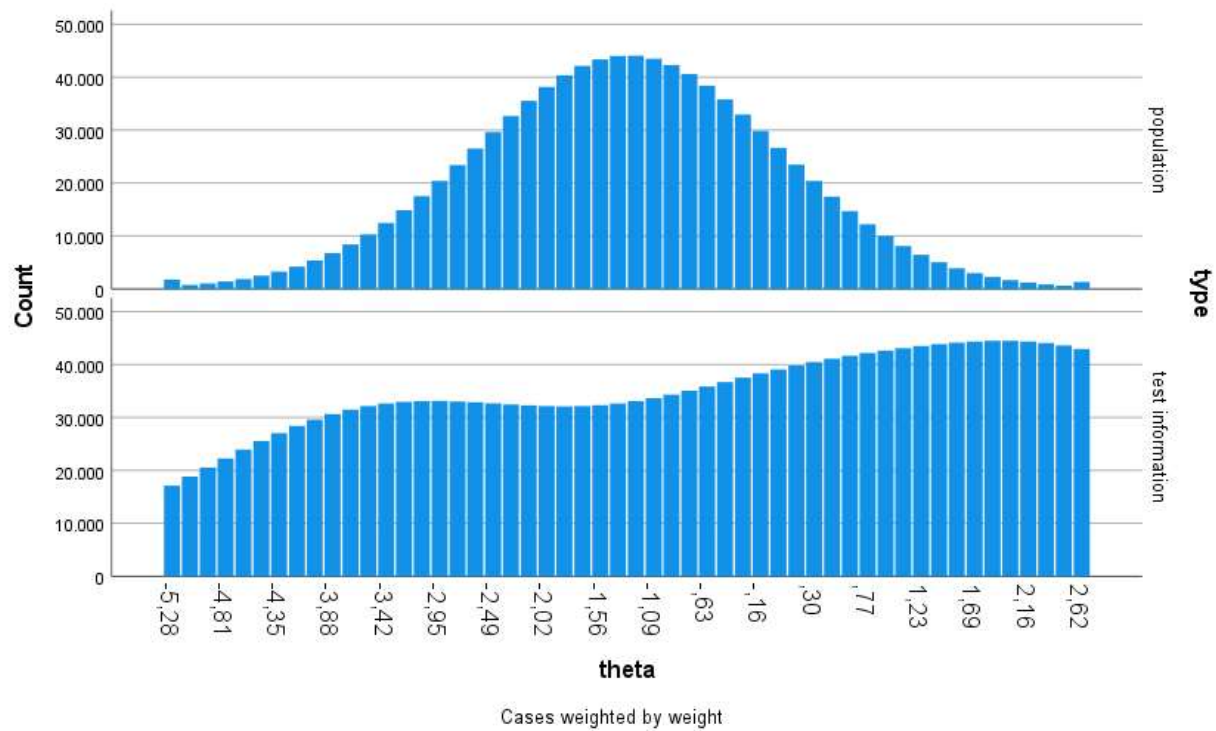
Test info maps



SEM maps

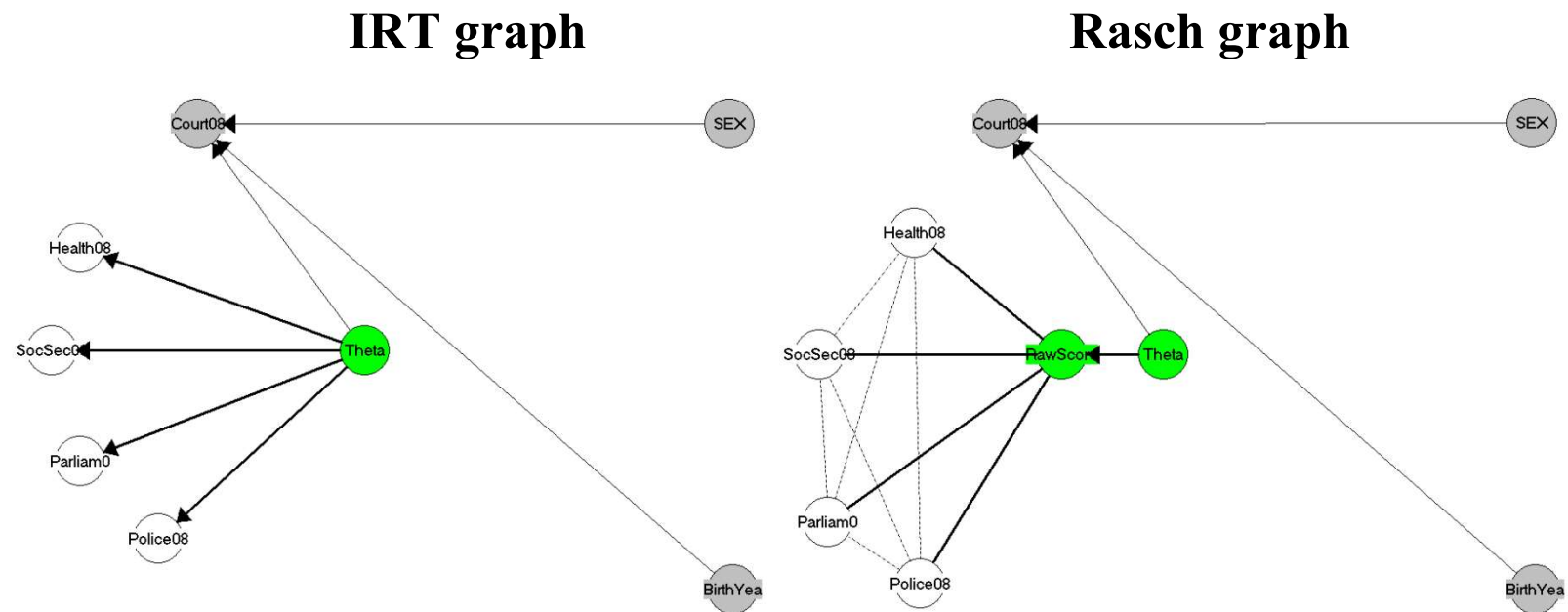


Marginal Info & SEM maps



What about the assumption of local independence?

Subtract one item from the model and treat it as an exogenous variable



If items fit the Rasch models, an item will be conditionally independent of all other items given the rest-score without the item.

Invoking DIGRAM's "SCREEN I" command provide tests of local dependence in Rasch models.

Tests of local independence
 - the row item has been subtracted from the score

			D	E	G	K	L

D	Police08	Gamma		0.234	-0.124	-0.314	0.163
		p		0.008	0.167	0.001	0.106
E	Parliam0	Gamma	0.219		-0.146	-0.313	0.239
		p	0.013		0.088	0.000	0.011
G	SocSec08	Gamma	-0.056	-0.203		0.398	-0.222
		p	0.548	0.020		0.000	0.026
K	Health08	Gamma	-0.341	-0.278	0.286		0.265
		p	0.000	0.000	0.000		0.002
L	Court08	Gamma	0.097	0.101	-0.345	0.147	
		p	0.265	0.258	0.000	0.086	

Even though the CLR test of homogeneity and the tests of item fit accepted the Rasch model, the evidence of DIF and the overwhelming evidence of local dependence show reject the Rasch model.

A graphical and log-linear Rasch model is one option