Analysis of person fit in graphical Rasch models

Person fit issues

Conditional distributions in Rasch models

Person fit in DIGRAM

Examples

Person fit issues

"The objective of person-fit measurement is to detect item-score patterns that are improbable given an IRT model".

(Meier and Sijtsma, 2001, page 130)

Two types of applications

Clinical applications of validated scales

Statistical applications and model fit

Statistical issues

Selecting a fit statistic

Assessment of significance

Clinical applications of validated scales

Examples

Application of educational test

Applications of health related scales during examination or assessment of treatment of patients

The problem

Measurement is invalid and should be disregarded if the respondent has responded to items in a way that does not agree with the Rasch model

Statistical applications and model fit

Misfitting persons influence estimation of item parameters and tests of model fit in ways that may provide spurious evidence of local dependence and DIF

Misfitting persons will influence statistical analyses of the association between the latent variable and exogenous variables

Eliminating the problems caused by misfitting persons is not easy

Distinguishing between genuine and spurious person misfit is difficult

The PF subscale of SF-36

Data from a Danish health survey

The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

A) PF1: Vigorous activities, such as running, lifting heavy objects, participating in sports
B) PF2: Moderate activities, moving a table, pushing a vacuum cleaner, or playing golf
C) PF3: Lifting or carrying groceries
D) PF4: Climbing several flights of stairs
E) PF5: Climbing one flight of stairs
F) PF6: Bending, kneeling, or stooping
G) PF7: Walking more than a mile
H) PF8: Walking several blocks
I) PF9: Walking one block
J) PF10: Bathing or dressing yourself

Responses were coded so that in such a way that a low score indicates physical impairment because of health problems.

0 : Yes, limited a lot   1: yes, limited a little   2: No, not limited at all

Model fit issues

What does the PF subscale measure? Physical functioning or limitations do to health problems.

Is PF unidimensional?

Do all items fit a Rasch model without local dependence and DIF?

The study population consist of Danes from age 18 with or without health problems. Does PF measure the same construct for all these persons?

Did all respondents read and understand the question?

The items are about activities you might do during a typical day.
Does *your health now* limit you in these activities? If so, how much?

We will assume that PF fits the Rasch model and show what tests of person fit may tell us.

The person-fit statistic

Let $X_1,\ldots,X_K$ be the items. The natural measure of improbability is

$$\mathbf{P = Pr(X_{i1} = x_1, \ldots, X_{ik} = x_k \mid \theta)}$$

Since θ is unknown, IRT theory suggests that we should use an estimate of θ instead and calculate.

$$\hat{\mathbf{P}}\left(\mathbf{x_1,\ldots,x_k}\right) = \mathbf{Pr}\left(\mathbf{X_{i1} = x_1, \ldots, X_{ik} = x_k \mid \hat{\theta}\left(x_1,\ldots,x_k\right)}\right)$$

In Rasch models, this leads to a test statistics where we know and can estimate the exact distribution of the fit statistic.

Assessment of significance

To assess the significance of $\hat{P}(x_1,...,x_k)$ we need to know its distribution

Let $R=\sum_{1=1}^{k} X_i$ the sufficient person score

The principle of conditional inference in Rasch models insists that we should assess the significance of $\hat{P}$ conditional given the total score over all items.

$$\Pr\left(\hat{P}(x_1,...,x_k) = p \mid R = r\right)$$

***All* estimates of the person parameter in the Rasch model are monotone functions of R**

$$\hat{\theta} = f(r) \;\; \text{or} \;\; \hat{\theta} = \hat{\theta}_r = f(r) \;\; \Leftrightarrow \;\; R = r$$

In other words. Our measure of improbability is defined by the conditional distribution of the response pattern given the score

$$\hat{P}(x_1,\ldots,x_k) = \Pr(X_{i1} = x_1,\ldots,X_{ik} = x_k \mid \hat{\theta} = f(r))$$

$$=$$

$$\Pr(X_{i1} = x_1,\ldots,X_{ik} = x_k \mid R{=}r)$$

The conditional probability of response patterns provide both the test statistic and the exact distribution of the test statistic

Conditional distributions in Rasch models

The joint distribution of items is

$$\Pr(x_1,\ldots,x_k \mid \exp(\theta) = \xi) = \prod_{i=1}^{k} \frac{\xi^{x_i}\delta_{ix_i}}{\sum_{z=0}^{m_i} \xi^z \delta_{iz}} = \xi^r \frac{\prod_{i=1}^{k} \delta_{ix_i}}{\prod_{i=1}^{k} G(\xi,\delta_i)}$$

The distribution of the score is

$$\Pr(R=r \mid \exp(T)=\xi) = \frac{\xi^x \gamma_r}{\sum_{z=0}^{m} \xi^z \gamma_r} \quad \text{where} \quad \gamma_r = \sum_{(x_1,\ldots,x_k):\sum_i x_i = r} \prod_{i=1}^{k} \delta_{ix_i}$$

$$\Pr(X_{i1}=x_1,\ldots,X_{ik}=x_k \mid R=r) = \frac{\prod_{i=1}^{k} \delta_{ix_i}}{\gamma_r}$$

Assessment of significance

The conditional probability of response patterns provide both the test statistic and the exact distribution of the test statistic

Let the observed conditional probability be

$$\hat{p} = \hat{P}\left(x_1,\ldots,x_k\right) = \Pr\left(X_{i1}=x_1,\ldots,X_{ik}=x_k \mid R=r\right)$$

Calculate the conditional probability of all response patterns given the observed score R = r and rank them according to increasing probabilities.

The p-value of the observed response pattern is

$$p = \sum_{(x_1,\ldots,x_k):\hat{P}(x_1,\ldots,x_k)\le\hat{p}} \hat{P}\left(x_1,\ldots,x_k\right)$$

Analysis of person fit in DIGRAM

Estimate the item parameters

Generate all response patterns and calculate

$$\hat{P}\left(x_1,\ldots,x_k\right) = Pr\left(X_{i1}=x_1,\ldots,X_{ik}=x_k \mid R=r\right)$$

Calculate $\hat{p} = Pr\left(X_{i1}=x_1,\ldots,X_{ik}=x_k \mid R=r\right)$ for each observed response pattern and calculate and report the p-value as the sum of the probabilities of the patterns that are as improbable as the observed pattern.

Example: The PF subscale of SF-36

The following items are about activities you might do during a typical day. Does your health now limit you in these activities? If so, how much?

A) PF1: Vigorous activities, such as running, lifting heavy objects, participating in sports
B) PF2: Moderate activities, moving a table, pushing a vacuum cleaner, or playing golf
C) PF3: Lifting or carrying groceries
D) PF4: Climbing several flights of stairs
E) PF5: Climbing one flight of stairs
F) PF6: Bending, kneeling, or stooping
G) PF7: Walking more than a mile
H) PF8: Walking several blocks
I) PF9: Walking one block
J) PF10: Bathing or dressing yourself

Responses were coded so that in such a way that a low score indicates physical impairment because of health problems.

0 : Yes, limited a lot 1: yes, limited a little 2: No, not limited at all

Model fit issues

What does the PF subscale measure? Physical functioning or limitations do to health problems.

Is PF unidimensional?

Do all items fit a Rasch model without local dependence and DIF?

The study population consist of Danes from age 18 with or without health problems. Does PF measure the same construct for all these persons?

Did all respondents read and understand the question?

The items are about activities you might do during a typical day.
Does *your health now* limit you in these activities? If so, how much?

We will assume that PF fits the Rasch model and show what tests of person fit may tell us.

Distributions of items

	Item	0	1	2	Mean	Average
J	Bathing	1.4	4.7	94.0	1.93	0.96
I	Walk 1bl	1.9	4.6	93.4	1.92	0.96
H	Walk 2+b	3.0	5.5	91.4	1.88	0.94
E	Stair1	2.4	7.3	90.3	1.88	0.94
G	Walk 1m	6.5	8.6	84.9	1.78	0.89
C	Liftgroc	4.9	12.9	82.2	1.77	0.89
B	Mod.act	5.7	13.0	81.3	1.76	0.88
F	Bending	6.2	14.5	79.3	1.73	0.87
D	Stair2+	6.3	14.7	78.9	1.73	0.86
A	Vig.act.	18.9	29.9	51.3	1.32	0.66

The distribution of the PF score

Score distribution. Mean = 17.7, sd = 3.95, skewness -2.38

Score	Count	Percent	Cumulated
0	16	0.6	0.6
1	11	0.4	1.1
2	7	0.3	1.3
3	11	0.4	1.8
4	11	0.4	2.2
5	11	0.4	2.6
6	15	0.6	3.2
7	28	1.1	4.3
8	21	0.8	5.1
9	29	1.1	6.3
10	36	1.4	7.7
11	33	1.3	9.0
12	31	1.2	10.2
13	35	1.4	11.6
14	57	2.2	13.8
15	75	2.9	16.8
16	104	4.1	20.9
17	115	4.5	25.4
18	218	8.6	33.9
19	465	18.3	52.2
20	1217	47.8	100.0
Total	2546	100.0	

CML estimates of PCM thresholds

item	1	2	Location
A: Vig.act.	2.038	5.248	3.643
B: Mod.act	-0.590	1.656	0.533
C: Liftgroc	-1.020	1.539	0.260
D: Stair2+	-0.593	2.022	0.715
E: Stair1	-2.480	-0.030	-1.255
F: Bending	-0.634	1.966	0.666
G: Walk 1m	-0.144	0.911	0.384
H: Walk 2+b	-1.543	-0.419	-0.981
I: Walk 1bl	-2.463	-0.926	-1.694
J: Bathing	-3.496	-1.042	-2.269

CML estimates of multiplicative parameters

item	0	1	2
A: Vig.act.	1.000	0.130	0.001
B: Mod.act	1.000	1.804	0.344
C: Liftgroc	1.000	2.774	0.595
D: Stair2+	1.000	1.809	0.240
E: Stair1	1.000	11.937	12.299
F: Bending	1.000	1.884	0.264
G: Walk 1m	1.000	1.155	0.464
H: Walk 2+b	1.000	4.681	7.119
I: Walk 1bl	1.000	11.740	29.634
J: Bathing	1.000	32.976	93.510

The score parameters of the PF score

score	gamma
0	1.0000
1	70.8919
2	1952.8930
3	29177.1080
4	272225.3872
5	1714738.7469
6	7644137.6048
7	24823957.3008
8	59719300.7794
9	107292048.9956
10	144156268.4685
11	144252254.5792
12	106529172.9351
13	57229614.7680
14	21905625.1053
15	5806203.0199
16	1027635.1758
17	115181.1417
18	7426.9097
19	219.6006
20	1.0000

The WML estimate of the person parameter

Score	Theta estimate	True score	Bias	RMSE	Score SEM
0	-5.031	0.42	0.521	0.889	0.62
1	-3.700	1.32	0.063	0.873	1.02
2	-2.979	2.27	0.008	0.799	1.26
3	-2.454	3.23	-0.000	0.713	1.44
4	-2.029	4.19	-0.001	0.649	1.57
5	-1.663	5.15	-0.001	0.606	1.67
6	-1.333	6.11	-0.000	0.578	1.74
7	-1.026	7.07	0.000	0.561	1.79
8	-0.730	8.03	0.001	0.551	1.82
9	-0.439	9.00	0.001	0.548	1.83
10	-0.148	9.97	0.001	0.551	1.82
11	0.150	10.94	0.001	0.558	1.79
12	0.460	11.92	0.001	0.570	1.76
13	0.788	12.90	0.000	0.588	1.70
14	1.138	13.88	-0.000	0.614	1.64
15	1.520	14.85	-0.000	0.652	1.56
16	1.945	15.82	0.002	0.712	1.45
17	2.443	16.77	0.007	0.802	1.31
18	3.092	17.72	0.007	0.931	1.11
19	4.222	18.76	-0.077	1.122	0.82
20	6.316	19.69	-0.641	1.187	0.50

Illustration: response pattern (2,2,1,2,2,2,1,2 2 2 2) with PF = 18

Is this significant evidence against the Rasch model?

PR(2,2,1,2,2,2,1,2,2,2,2 | PF = 18) = 0.0044831

There are 55 response patterns with PF = 18. 40 with probabilities below 0.0044831 and 14 with larger probabilities

The probability of a response pattern with probability less than or equal to 0.0044831 is 0.027 providing weakly significant against person fit.

Why? Is it really misfit or just random error?

To understand what has happened we have to look at the most probable response pattern and the expected response pattern when the PF score is equal to 18.

	PF1	PF2	PF3	PF4	PF5	PF6	PF7	PF8	PF9	PF10	$\hat{P}$
Obs	2	2	**1**	2	2	**1**	2	2	2	2	0.00448
Mode	0	2	**2**	2	2	**2**	2	2	2	2	0.19641
Exp	0.85	1.85	**1.86**	1.78	1.97	**1.79**	1.93	1.98	1.99	1.99	-

Obviously, the response to PF3 (carrying groceries) is inconsistent compared to the responses to PF1 and PF2.

To assess the response to PF6 is difficult. It may help to look at the complete range of response patterns with PF = 18

A	B	C	D	E	F	G	H	I	J	Cprobs	Cumulated	L0	Lz	
2	2	2	2	2	2	2	2	2	0	0.0000014	0.0000014	-13.479	-10.410	
2	2	2	2	2	2	2	2	0	2	0.0000045	0.0000059	-12.311	-9.333	
2	2	2	2	0	2	2	2	2	2	0.0000109	0.0000168	-11.427	-8.517	
2	2	2	2	2	2	2	2	1	1	0.0000188	0.0000356	-10.882	-8.015	
2	2	2	2	2	2	2	0	2	2	0.0000189	0.0000545	-10.876	-8.010	
2	2	2	2	2	2	2	1	2	1	0.0000312	0.0000857	-10.375	-7.548	
2	2	2	2	2	2	2	1	1	2	0.0000351	0.0001208	-10.257	-7.439	
2	2	2	2	1	2	2	2	2	1	0.0000461	0.0001669	-9.985	-7.188	
2	2	2	2	1	2	2	2	1	2	0.0000518	0.0002187	-9.868	-7.080	
2	2	2	2	1	2	2	1	2	2	0.0000859	0.0003046	-9.362	-6.614	
2	2	2	2	2	2	1	2	2	1	0.0001181	0.0004227	-9.044	-6.320	
2	2	2	2	2	2	1	2	1	2	0.0001327	0.0005554	-8.927	-6.213	
2	2	2	2	2	2	1	1	2	2	0.0002202	0.0007756	-8.421	-5.746	
2	2	1	2	2	2	2	2	2	1	0.0002213	0.0009969	-8.416	-5.741	
2	2	0	2	2	2	2	2	2	2	0.0002263	0.0012232	-8.394	-5.721	
2	2	1	2	2	2	2	2	1	2	0.0002487	0.0014719	-8.299	-5.633	
2	1	2	2	2	2	2	2	2	1	0.0002488	0.0017207	-8.299	-5.633	
2	1	2	2	2	2	2	2	1	2	0.0002795	0.0020002	-8.183	-5.526	
2	2	2	2	2	2	0	2	2	2	0.0002900	0.0022902	-8.146	-5.492	
2	2	2	2	1	2	1	2	2	2	0.0003250	0.0026152	-8.032	-5.387	
2	2	2	2	2	1	2	2	2	1	0.0003391	0.0029543	-7.989	-5.348	
2	2	2	1	2	2	2	2	2	1	0.0003586	0.0033129	-7.933	-5.296	
2	2	2	2	2	1	2	2	1	2	0.0003810	0.0036939	-7.873	-5.240	
2	0	2	2	2	2	2	2	2	2	0.0003911	0.0040850	-7.847	-5.216	
2	2	2	1	2	2	2	2	1	2	0.0004029	0.0044879	-7.817	-5.189	
2	2	1	2	2	2	2	1	2	2	0.0004127	0.0049006	-7.793	-5.166	
2	1	2	2	2	2	2	1	2	2	0.0004639	0.0053645	-7.676	-5.059	
2	2	2	2	2	0	2	2	2	2	0.0005103	0.0058748	-7.581	-4.971	
2	2	2	0	2	2	2	2	2	2	0.0005622	0.0064370	-7.484	-4.881	
2	2	1	2	1	2	2	2	2	2	0.0006092	0.0070462	-7.403	-4.807	
2	2	2	2	2	1	2	1	2	2	0.0006323	0.0076785	-7.366	-4.773	
2	2	2	1	2	2	2	1	2	2	0.0006687	0.0083472	-7.310	-4.721	
2	1	2	2	1	2	2	2	2	2	0.0006849	0.0090321	-7.286	-4.699	
2	2	2	2	1	1	2	2	2	2	0.0009335	0.0099656	-6.977	-4.414	
2	2	2	1	1	2	2	2	2	2	0.0009871	0.0109527	-6.921	-4.362	
2	2	1	2	2	2	1	2	2	2	0.0015609	0.0125136	-6.462	-3.940	
2	1	2	2	2	2	1	2	2	2	0.0017547	0.0142683	-6.345	-3.832	
2	2	2	2	2	1	1	2	2	2	0.0023916	0.0166599	-6.036	-3.546	
2	2	2	1	2	2	1	2	2	2	0.0025290	0.0191889	-5.980	-3.495	
2	1	1	2	2	2	2	2	2	2	0.0032892	0.0224781	-5.717	-3.252	
2	**2**	**1**	**2**	**2**	**1**	**2**	**2**	**2**	**2**	**0.0044831**	**0.0269612**	**-5.407**	**-2.967**	**Observed**
2	2	1	1	2	2	2	2	2	2	0.0047407	0.0317019	-5.352	-2.915	
2	1	2	2	2	1	2	2	2	2	0.0050398	0.0367417	-5.290	-2.859	
2	1	2	1	2	2	2	2	2	2	0.0053294	0.0420711	-5.235	-2.807	
2	**2**	**2**	**1**	**2**	**1**	**2**	**2**	**2**	**2**	**0.0072638**	**0.0493349**	**-4.925**	**-2.522**	**Critical**
1	2	2	2	2	2	2	2	2	1	0.0090283	0.0583632	-4.707	-2.321	
1	2	2	2	2	2	2	2	1	2	0.0101425	0.0685057	-4.591	-2.214	
1	2	2	2	2	2	2	1	2	2	0.0168332	0.0853389	-4.084	-1.747	
1	2	2	2	1	2	2	2	2	2	0.0248497	0.1101886	-3.695	-1.388	
1	2	2	2	2	2	1	2	2	2	0.0636662	0.1738548	-2.754	-0.520	
1	2	1	2	2	2	2	2	2	2	0.1193433	0.2931981	-2.126	0.059	
1	1	2	2	2	2	2	2	2	2	0.1341626	0.4273607	-2.009	0.167	
1	**2**	**2**	**2**	**2**	**1**	**2**	**2**	**2**	**2**	**0.1828586**	**0.6102193**	**-1.699**	**0.453**	**Consistent**
1	2	2	1	2	2	2	2	2	2	0.1933671	0.8035864	-1.643	0.504	
0	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**2**	**0.1964133**	**0.9999997**	**-1.628**	**0.519**	**Mode**

The conditional distributions of response patterns disclose three interesting things.

The most probable response pattern has zero points on PF1 and two points on all items.

All response patterns with two points on PF1 are significant

The conditional probability with consistent responses on PF1, PF2 and PF3 but PF6 = 1 is far from significant (p = 0.61)

Person fit in DIGRAM

Invoke the PERsonfit command, to obtain

A table with the most probable response patterns in each score group

A table with the expected item scores in each score group

A table with the critical sizes of test that reject fit if $p \leqq 0.05$

A table with all observed response patterns with $p \leqq 0.05$

A tables with the number of observed and expected patterns with significant tests of person fitresponse patterns

A test comparing the observed and expected number of significant tests

The most probable response patterns

			A	B	C	D	E	F	G	H	I	J
1	0.4651646	pattern =	0	0	0	0	0	0	0	0	0	1
2	0.2015741	pattern =	0	0	0	0	1	0	0	0	0	1
3	0.1583958	pattern =	0	0	0	0	1	0	0	0	1	1
4	0.0794657	pattern =	0	0	0	0	1	0	0	1	1	1
5	0.0357740	pattern =	0	0	0	0	1	0	0	1	1	2
6	0.0222620	pattern =	0	0	1	0	1	0	0	1	1	2
7	0.0173037	pattern =	0	0	1	0	1	0	0	1	2	2
8	0.0135547	pattern =	0	0	1	0	1	1	0	1	2	2
9	0.0136486	pattern =	0	0	1	1	1	1	0	1	2	2
10	0.0183294	pattern =	0	1	1	1	1	1	0	1	2	2
11	0.0278583	pattern =	0	1	1	1	1	1	0	2	2	2
12	0.0435598	pattern =	0	1	1	1	1	1	1	2	2	2
13	0.0835367	pattern =	0	1	1	1	2	1	1	2	2	2
14	0.0877600	pattern =	0	1	1	1	2	1	2	2	2	2
15	0.0710273	pattern =	0	1	2	1	2	1	2	2	2	2
16	0.0765792	pattern =	0	2	2	1	2	1	2	2	2	2
17	0.0956570	pattern =	0	2	2	1	2	2	2	2	2	2
18	0.1964133	pattern =	0	2	2	2	2	2	2	2	2	2
19	0.8658412	pattern =	1	2	2	2	2	2	2	2	2	2

The expected response patterns

score	A	B	C	D	E	F	G	H	I	J
1	0.00	0.03	0.04	0.03	0.17	0.03	0.02	0.07	0.17	0.47
2	0.00	0.06	0.10	0.06	0.37	0.07	0.04	0.17	0.39	0.74
3	0.01	0.11	0.17	0.11	0.56	0.12	0.08	0.29	0.61	0.93
4	0.01	0.18	0.25	0.18	0.71	0.18	0.12	0.45	0.83	1.09
5	0.02	0.25	0.35	0.24	0.85	0.25	0.18	0.62	1.02	1.23
6	0.03	0.33	0.44	0.32	0.97	0.33	0.24	0.80	1.19	1.35
7	0.04	0.42	0.54	0.41	1.09	0.42	0.33	0.97	1.33	1.46
8	0.05	0.51	0.65	0.50	1.20	0.51	0.43	1.15	1.46	1.55
9	0.07	0.62	0.75	0.59	1.30	0.61	0.55	1.31	1.57	1.64
10	0.09	0.73	0.85	0.70	1.41	0.71	0.69	1.45	1.66	1.71
11	0.12	0.85	0.96	0.80	1.51	0.82	0.86	1.57	1.74	1.78
12	0.16	0.97	1.07	0.91	1.61	0.93	1.04	1.68	1.80	1.83
13	0.21	1.10	1.19	1.03	1.70	1.04	1.24	1.76	1.86	1.87
14	0.28	1.24	1.32	1.15	1.78	1.17	1.43	1.83	1.90	1.91
15	0.37	1.39	1.45	1.28	1.84	1.30	1.60	1.89	1.93	1.94
16	0.49	1.55	1.59	1.44	1.90	1.45	1.73	1.93	1.96	1.96
17	0.65	1.70	1.73	1.60	1.94	1.62	1.84	1.96	1.97	1.98
18	0.85	1.85	1.86	1.78	1.97	1.79	1.93	1.98	1.99	1.99
19	1.13	1.98	1.98	1.97	2.00	1.97	1.99	2.00	2.00	2.00
Average	0.31	0.85	0.92	0.81	1.28	0.82	0.87	1.23	1.40	1.50
Weighted	1.33	1.76	1.78	1.74	1.89	1.74	1.80	1.89	1.92	1.93

Critical sizes of test of person fit

Score	number of patterns	critical size
1	10	0.0436
2	55	0.0494
3	210	0.0499
4	615	0.0496
5	1452	0.0500
6	2850	0.0500
7	4740	0.0499
8	6765	0.0500
9	8350	0.0499
10	8953	0.0500
11	8350	0.0499
12	6765	0.0500
13	4740	0.0500
14	2850	0.0500
15	1452	0.0496
16	615	0.0497
17	210	0.0465
18	55	0.0493
19	10	0.0434

Highly significant tests of person fit

K, L, M, N, O refer to exogenous variables

J	I	E	H	C	G	B	F	D	A	score	Prob	count	p	size	K	L	M	N	O
2	0	2	0	2	0	0	2	2	0	10	0.00000	1	0.000	0.050	3	2	1	2	2
1	0	2	0	2	2	0	1	2	1	11	0.00000	1	0.000	0.050	2	3	3	1	2
0	0	0	0	2	2	1	1	0	1	7	0.00000	1	0.000	0.050	2	2	3	2	2
0	0	0	0	0	0	0	0	0	2	2	0.00000	1	0.000	0.049	4	3	3	2	2
1	2	0	1	2	1	0	1	1	2	11	0.00000	1	0.000	0.050	1	*	3	1	2
2	1	0	1	0	0	1	0	0	2	7	0.00000	1	0.000	0.050	1	3	1	2	*
2	1	0	1	0	1	1	0	0	2	8	0.00000	1	0.000	0.050	1	2	*	1	2
2	1	2	0	0	0	1	1	2	2	11	0.00000	1	0.000	0.050	2	1	3	1	2
2	1	0	0	1	0	1	0	0	2	7	0.00000	1	0.000	0.050	3	2	3	1	5
0	0	0	0	1	0	1	0	2	2	6	0.00000	1	0.000	0.050	3	3	*	1	3
0	0	1	0	1	1	1	1	2	2	9	0.00000	1	0.000	0.050	4	3	3	1	2
1	0	2	0	2	1	1	0	0	2	9	0.00000	1	0.000	0.050	3	*	3	2	2
0	0	0	0	1	0	2	1	0	2	6	0.00000	1	0.000	0.050	2	*	1	1	1
2	2	1	0	1	0	2	2	0	2	12	0.00000	1	0.000	0.050	3	*	3	1	5
1	1	2	1	1	2	2	2	2	2	16	0.00000	1	0.000	0.050	3	2	1	1	3
2	2	2	1	2	0	2	2	0	2	15	0.00000	1	0.000	0.050	3	2	3	1	3
0	1	1	1	2	1	2	1	1	2	12	0.00000	1	0.000	0.050	4	3	3	1	4
2	1	1	1	2	1	2	1	1	2	14	0.00000	1	0.000	0.050	4	*	3	2	*
0	2	2	1	2	0	2	1	1	2	13	0.00000	1	0.000	0.050	1	*	3	2	1
2	1	1	1	2	2	2	1	2	2	16	0.00000	1	0.000	0.050	2	2	3	2	2
2	1	2	1	2	1	2	1	2	2	16	0.00000	1	0.000	0.050	3	*	3	2	1
2	0	2	0	2	0	2	2	2	2	14	0.00000	1	0.000	0.050	3	2	3	2	2

Observed and expected significant tests in each score group

Score	n	Pcrit	Significant Expected	Significant Observed
1	11	0.0436	0.48	0
2	7	0.0494	0.35	1
3	11	0.0499	0.55	0
4	11	0.0496	0.55	0
5	11	0.0500	0.55	1
6	15	0.0500	0.75	2
7	28	0.0499	1.40	6
8	21	0.0500	1.05	1
9	29	0.0499	1.45	3
10	36	0.0500	1.80	7
11	33	0.0499	1.65	4
12	31	0.0500	1.55	5
13	35	0.0500	1.75	3
14	57	0.0500	2.85	6
15	75	0.0496	3.72	6
16	104	0.0497	5.17	11
17	115	0.0465	5.35	7
18	218	0.0493	10.76	14
19	465	0.0434	20.17	10

Observed and expected significant tests in groups defined by exogenous variables

```
+---------+
|         |
| K - srh |
|         |
+---------+
```

srh	n	Obs	exp	
VeryGood	199	14	9.2	p = 0.1078
Good	616	30	28.6	p = 0.7907
Fair	394	33	19.0	p = 0.0010
Bad	79	8	3.9	p = 0.0309

```
+---------+
|         |
| L - BMI |
|         |
+---------+
```

BMI	n	Obs	exp	
10 - 22	366	21	17.1	p = 0.3320
23 - 25	414	27	19.4	p = 0.0787
26 - 30	348	19	16.5	p = 0.5280
31+	112	7	5.4	p = 0.4750

```
+-------------+
|             |
| M - Smoking |
|             |
+-------------+
```

Smoking	n	Obs	exp		
Yes	498	28	23.4	p =	0.3271
NotDaily	91	2	4.2	p =	0.2657
No	684	53	32.4	p =	0.0002

```
+---------+
|         |
| N - Sex |
|         |
+---------+
```

Sex	n	Obs	exp		
Male	524	29	24.5	p =	0.3523
Female	772	57	36.6	p =	0.0005

```
+----------+
|          |
| O - Age  |
|          |
+----------+
```

Age	n	Obs	exp		
18 - 29	307	17	14.1	p =	0.4367
30 - 49	422	34	19.7	p =	0.0009
50 - 59	183	10	8.7	p =	0.6607
60 - 69	147	7	7.0	p =	0.9868
70+	233	16	11.3	p =	0.1543

A total of 87 response patterns that disagreed with the Rasch model

Only 61.9 was expected.

Since this is significant (p = 0.0011) we conclude that person misfit is an issue for the PF scale and/or the Rasch model.

Person misfit issues

Do we have measurement issues or is it a problem with a misfitting statistical model?

If it is a measurement problem, how do we distinguish between random errors generated by the model and response patterns with irregular response behavior?

If it is a problem with the model, how do we deal with it if we had done our best to test the model? Do we eliminate persons?