

UNIVERSIDADE
SÃO FRANCISCO

WINSTEPS: Conceitos básicos e Interpretações das principais tabelas

Dr. Ricardo Primi

Programa de Mestrado e Doutorado em Avaliação Psicológica

Universidade São Francisco

Análises básicas (sugestão de sequência de análise)

- Análise dos itens

- Dificuldade, correlação item total, infit outfit

- Critérios para Infit, Outfit:

- >2,0: Degrada o sistema de mensuração,
 - 1,5 a 2,0: Não produtivo para medida
 - 0,5 a 1,5: Produtivo
 - <0,5 menor produtivo mas menos preocupante. Pode produzir artificialmente altas precisões

$$Outfit_i = \frac{\sum_{n=1}^N \frac{r_{ni}^2}{V_{ni}}}{N}$$

$$Infit_i = \frac{\sum_{n=1}^N r_{ni}^2}{\sum_{n=1}^N V_{ni}}$$

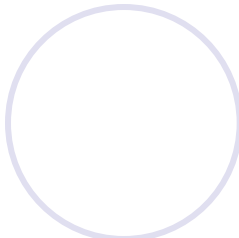
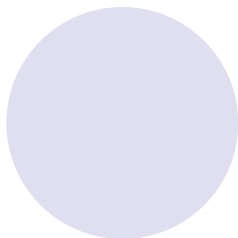
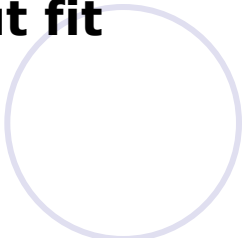
- Análise da escala

- Precisão
 - Informação / Precisão Local
 - Mapa de itens

- Salvar escores para SPSS

	i02	i05	i01	i09	i07	i03	i06	i10	i04	i08	i14	i11	i15	i13	i12	i16						Esc	Desv	Desv ²
<i>a</i>	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1						1	0.5	0.25
<i>b</i>	-1.03	-0.88	-0.73	-0.52	-0.44	-0.1	-0.09	0.07	0.15	0.26	0.46	0.6	1.02	1.16	1.21	1.22						1	0.5	0.25
<i>c</i>	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0						1	0.5	0.25
Theta estimado / Valores esperados																						1	0.5	0.25
-0.3	0.78	0.73	0.68	0.59	0.56	0.42	0.41	0.35	0.32	0.28	0.22	0.18	0.10	0.08	0.07	0.07	5.810					0	-0.5	0.25
Vetor de resposta																						0	-0.5	0.25
	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	6.000					0	-0.5	0.25
Resíduos																						0	-0.5	0.25
	0.22	0.27	0.32	0.41	0.44	0.58	-0.41	-0.35	-0.32	-0.28	-0.22	-0.18	-0.10	-0.08	-0.07	-0.07	0.190					Média	0.5	
																						Soma		2.5
Variância esperada																								
	0.17	0.20	0.22	0.24	0.25	0.24	0.24	0.23	0.22	0.20	0.17	0.15	0.09	0.07	0.07	0.07	2.813					Variância		0.28
																						Desv. Padr		0.53
Desvio padrão																								
	0.417	0.445	0.468	0.491	0.496	0.493	0.492	0.476	0.466	0.448	0.411	0.383	0.294	0.267	0.257	0.255	6.561							
Resíduo ao quadrado																								
	0.05	0.074	0.106	0.166	0.194	0.341	0.169	0.121	0.101	0.078	0.046	0.032	0.009	0.006	0.005	0.005								
Resíduo ao quadrado / Variância (divide o resíduo ao quadrado pela variância, isto é pela expectativa modelada do resíduo)																	Qui Quadrad	Outfit	Infit					
	0.29	0.37	0.48	0.69	0.79	1.40	0.70	0.53	0.47	0.39	0.27	0.22	0.11	0.08	0.08	0.08	6.942	0.43	0.53			0.53		
Resíduo ao quadrado/ Variância multiplicado por um peso da "informação" para o cálculo do Infit																								
	0.05	0.074	0.106	0.166	0.194	0.341	0.169	0.121	0.101	0.078	0.046	0.032	0.009	0.006	0.005	0.005								
Resíduo/Desvio padrão																		Zoutfit	Zinfit					
	0.538	0.611	0.694	0.829	0.888	1.185	-0.84	-0.73	-0.68	-0.62	-0.52	-0.47	-0.33	-0.29	-0.28	-0.27		-0.02	0.07					
Resíduo/Desvio padrão multiplicado pelo peso da Informação																								
	0.094	0.121	0.152	0.2	0.219	0.288	-0.2	-0.17	-0.15	-0.12	-0.09	-0.07	-0.03	-0.02	-0.02	-0.02								

Padrões de In-Out fit



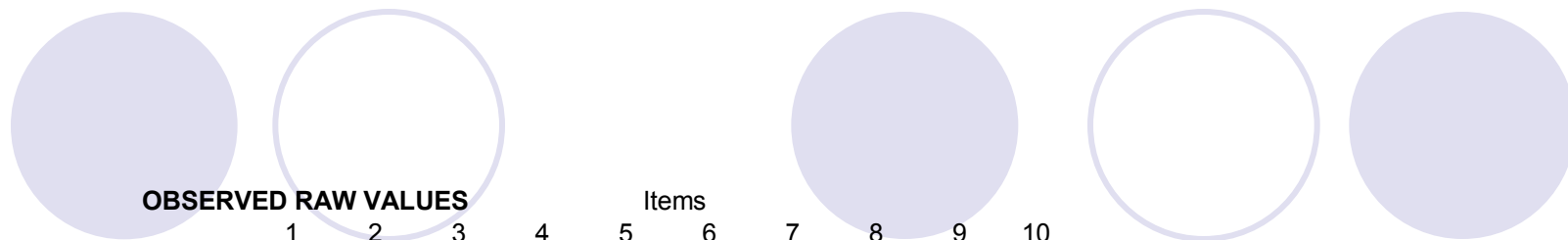
Responses:	Diagnosis	INFIT	OUTFIT
Easy--Items--Hard	Pattern	MnSq	MnSq

111 0110110100 000	Modelled/Ideal	1.1	1.0
000 0000011111 111	Miscode	4.3	12.6
011 1111110000 000	Carelessness/Sleeping	1.0	3.8
111 1111000000 001	Lucky Guessing	1.0	3.8
101 0101010101 010	Response set/Miskey	2.3	4.0
111 1000011110 000	Special knowledge	1.3	0.9
111 1111100000 000	Guttman/Deterministic	0.5	0.3
111 1010110010 000	Imputed outliers *	1.0	0.6

Right Transition Wrong	Expectation: 1.0		

Diagnosing Misfit				
Classification	INFIT	OUTFIT	Explanation	Investigation
	Noisy	Noisy	Lack of convergence Loss of precision Anchoring	Final values in Table 0 large? Many categories? Large logit range? Displacements reported?
Hard Item	Noisy	Noisy	Bad item	Ambiguous or negative wording? Debatable or misleading options?
	Muted	Muted	Only answered by top people	At end of test?
Item	Noisy	Noisy	Qualitatively different item Incompatible anchor value	Different process or content? Anchor value incorrectly applied?
		?	Biased (DIF) item	Stratify residuals by person group?
		Muted	Curriculum interaction	Are there alternative curricula?
	Muted	?	Redundant item	Similar items? One item answers another? Item correlated with other variable?
Rating scale	Noisy	Noisy	Extreme category overuse	Poor category wording? Combine or omit categories? Wrong model for scale?
	Muted	Muted	Middle category overuse	
Person	Noisy	?	Processing error Clerical error Idiosyncratic person	Scanner failure? Form markings misaligned? Qualitatively different person?

Diagnosing Misfit				
Classification	INFIT	OUTFIT	Explanation	Investigation
	Muted	Muted	Middle category overuse	
Person	Noisy	?	Processing error Clerical error Idiosyncratic person	Scanner failure? Form markings misaligned? Qualitatively different person?
High Person	?	Noisy	Careless Sleeping Rushing	Unexpected wrong answers? Unexpected errors at start? Unexpected errors at end?
Low Person	?	Noisy	Guessing Response set "Special" knowledge	Unexpected right answers? Systematic response pattern? Content of unexpected answers?
	Muted	?	Plodding Caution	Did not reach end of test? Only answered easy items?
Person/Judge Rating	Noisy	Noisy	Extreme category overuse	Extremism? Defiance?
	Muted	Muted	Middle category overuse	Conservatism? Resistance?
Judge Rating			Apparent unanimity	Collusion?
INFIT: information-weighted mean-square, sensitive to irregular inlying patterns OUTFIT: usual unweighted mean-square, sensitive to unexpected rare extremes Muted: un-modeled dependence, redundancy, error trends Noisy: unexpected unrelated irregularities				



		OBSERVED RAW VALUES									
		Items									
		1	2	3	4	5	6	7	8	9	10
Persons	A	.	1	1	1	1	1	1	1	0	1
	B	1	1	1	1	1	1	1	0	1	0
	C	1	1	1	1	1	1	0	1	0	0
	D	1	1	1	1	1	1	0	1	0	0
	E	1	1	1	1	1	1	0	1	0	0
	F	1	1	1	1	1	0	1	0	0	0
	G	1	1	1	1	0	1	0	0	0	0
	H	1	0	1	0	1	0	0	0	0	0
	I	0	1	0	1	0	0	0	0	0	0

		EXPECTED VALUES FROM LAST ITERATION										Final Person / Person	
		1	2	3	4	5	6	7	8	9	10		
A		1,00	1,00	1,00	1,00	1,00	0,99	0,91	0,96	0,57	0,57	4,63	1,27
B		1,00	1,00	1,00	1,00	0,99	0,97	0,72	0,85	0,24	0,24	3,22	1,15
C		0,99	0,99	0,99	0,99	0,96	0,90	0,41	0,60	0,08	0,08	1,93	1,12
D		0,99	0,99	0,99	0,99	0,96	0,90	0,41	0,60	0,08	0,08	1,93	1,12
E		0,99	0,99	0,99	0,99	0,96	0,90	0,41	0,60	0,08	0,08	1,93	1,12
F		0,97	0,97	0,97	0,97	0,88	0,71	0,17	0,31	0,03	0,03	0,69	1,10
G		0,91	0,91	0,91	0,91	0,71	0,45	0,06	0,13	0,01	0,01	-0,43	1,02
H		0,63	0,63	0,63	0,63	0,30	0,12	0,01	0,02	0,00	0,00	-2,18	0,88
I		0,44	0,44	0,44	0,44	0,16	0,06	0,01	0,01	0,00	0,00	-2,97	0,91
Final Item Difficulty Logi		-2,91	-2,91	-2,91	-2,91	-1,44	-0,21	2,46	1,64	4,59	4,59		
Item Standard Error		1,27	1,27	1,27	1,27	1,14	1,03	0,90	0,90	1,20	1,20		

MarkM:

Each cell is the squared residual divided by the variance for that cell. The red highlighted cells are misfitting (>1.3), though these misfits are somewhat artifactual.

MarkM:

Outfit is simply the average of the cell fits for that row, sensitive to outliers.

CELL FIT

	Items											
	1	2	3	4	5	6	7	8	9	10	OutFit	InFit
A		0,00	0,00	0,00	0,00	0,01	0,10	0,04	1,33	0,75	0,25	0,83
B	0,00	0,00	0,00	0,00	0,01	0,03	0,39	5,53	3,10	0,32	0,94	1,90
C	0,01	0,01	0,01	0,01	0,04	0,12	0,70	0,66	0,09	0,09	0,17	0,44
D	0,01	0,01	0,01	0,01	0,04	0,12	0,70	0,66	0,09	0,09	0,17	0,44
E	0,01	0,01	0,01	0,01	0,04	0,12	0,70	0,66	0,09	0,09	0,17	0,44
F	0,03	0,03	0,03	0,03	0,13	2,49	4,91	0,44	0,03	0,03	0,82	1,58
G	0,10	0,10	0,10	0,10	2,49	1,24	0,07	0,14	0,01	0,01	0,44	0,90
H	0,58	1,73	0,58	1,73	2,31	0,14	0,01	0,03	0,00	0,00	0,71	1,23
I	0,78	1,27	0,79	1,27	0,20	0,06	0,01	0,01	0,00	0,00	0,44	0,87
OutFit	0,19	0,35	0,17	0,35	0,58	0,48	0,84	0,91	0,53	0,15		
InFit	0,54	1,16	0,54	1,16	1,35	0,93	1,05	1,04	1,33	0,38		

MarkM:

Infit is the sum of the squared residuals divided by the sum of the variances. It is less sensitive to outliers and more sensitive to structural problems.

Ability Logit**Standard Error****MarkM:**

Infit is the sum of the squared residuals divided by the sum of the variances. It is less sensitive to outliers and more sensitive to structural problems.

MarkM:

Outfit is simply the average of the cell fits for this column, sensitive to outliers.

ITERATION 9: VARIANCE OF EXPECTED VALUES

	1	2	3	4	5	6	7	8	9	10	(-1)*(S
A	0,00	0,00	0,00	0,00	0,00	0,01	0,08	0,04	0,25	0,25	-0,62
B	0,00	0,00	0,00	0,00	0,01	0,03	0,20	0,13	0,18	0,18	-0,75
C	0,01	0,01	0,01	0,01	0,04	0,09	0,24	0,24	0,08	0,08	-0,80
D	0,01	0,01	0,01	0,01	0,04	0,09	0,24	0,24	0,08	0,08	-0,80
E	0,01	0,01	0,01	0,01	0,04	0,09	0,24	0,24	0,08	0,08	-0,80
F	0,03	0,03	0,03	0,03	0,10	0,20	0,14	0,21	0,02	0,02	-0,83
G	0,08	0,08	0,08	0,08	0,20	0,25	0,06	0,11	0,01	0,01	-0,97
H	0,23	0,23	0,23	0,23	0,21	0,11	0,01	0,02	0,00	0,00	-1,29
I	0,25	0,25	0,25	0,25	0,14	0,06	0,01	0,01	0,00	0,00	-1,20
(-1)*(S	-0,62	-0,62	-0,62	-0,62	-0,77	-0,94	-1,22	-1,25	-0,69	-0,69	

Tabela 3.1 Summary of persons and Items

TABLE 3.1 GfRLD Forma A

ZOU504WS.TXT Oct 4 23:03 2010

INPUT: 528 P 46 I MEASURED: 528 P 45 I 2 CATS

WINSTEPS 3.69.1

SUMMARY OF 528 MEASURED P								
	RAW			MODEL	INFIT		OUTFIT	
	SCORE	COUNT	MEASURE	ERROR	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	22.1	43.6	.12	.37	1.00	-.1	1.04	.0
S.D.	5.7	3.8	.68	.03	.25	1.4	.55	1.2
MAX.	42.0	45.0	3.28	.67	2.66	7.2	5.27	7.2
MIN.	8.0	15.0	-1.40	.35	.53	-3.3	.42	-2.5
REAL RMSE	.39	TRUE SD	.56	SEPARATION	1.45	P	RELIABILITY	.68
MODEL RMSE	.37	TRUE SD	.57	SEPARATION	1.54	P	RELIABILITY	.70
S.E. OF P MEAN = .03								

VALID RESPONSES: 96.9%

P RAW SCORE-TO-MEASURE CORRELATION = .95 (approximate due to missing data)

CRONBACH ALPHA (KR-20) P RAW SCORE RELIABILITY = .75 (approximate due to missing data)

SUMMARY OF 45 MEASURED I

	RAW			MODEL	INFIT		OUTFIT	
	SCORE	COUNT	MEASURE	ERROR	MNSQ	ZSTD	MNSQ	ZSTD
MEAN	259.6	511.8	.00	.11	1.00	-.2	1.03	.1
S.D.	128.3	9.9	1.42	.04	.06	1.7	.14	1.9
MAX.	512.0	524.0	2.45	.34	1.13	4.1	1.39	4.4
MIN.	55.0	471.0	-4.11	.09	.89	-3.7	.82	-3.6
REAL RMSE	.12	TRUE SD	1.41	SEPARATION	11.54	I	RELIABILITY	.99
MODEL RMSE	.12	TRUE SD	1.41	SEPARATION	11.67	I	RELIABILITY	.99
S.E. OF I MEAN = .21								

DELETED: 1 I

UMEAN=.0000 USCALE=1.0000

I RAW SCORE-TO-MEASURE CORRELATION = -.98 (approximate due to missing data)

23029 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 23797.51 with 22457 d.f. p=.0000

Global Root-Mean-Square Residual (excluding extreme scores): .4139

Precisão no Modelo de Rasch calculada pelo WINSTEPS

- Precisão = Variância Verdadeira / Variância Total
 - Variância Total = Variância Verdadeira + Variância de Erro
 - Ou melhor: parcela (em proporção) da variância observada que é verdadeira
- Precisão do teste = precisão das medidas das pessoas. Ainda a precisão não é uma medida de qualidade e sim de reprodutibilidade
- Reliability (separation index) means "reproducibility of relative measure location".
- Precisão do teste depende (das medidas das pessoas) :
 - Variância da amostra
 - Número de itens
 - Mais categorias por item
 - Correspondência itens-pessoas
- Precisão dos itens
 - Variância nos índices de dificuldade
 - Tamanho da amostra
- Diferentes métodos usam cálculos diferentes da variância verdadeira:
 - KR-20: cálculo aproximado pelas correlações bisseriais
 - Alfa: de cálculo aproximado empregando análise da variância (var. do escore total e var da soma dos itens)
 - Rasch: cálculo aproximado pelo erro padrão das estimativas

Precisão no Modelo de Rasch calculada pelo WINSTEPS

- WINSTEPS calcula dois coeficientes baseado em dois tipos de erro:
 - Upper bound (valor mais alto possível): Model reliability baseado em erros modelados
 - baseada na coesão/correspondência entre pessoas/itens
 - Supõe que os dados se adequam perfeitamente ao modelo
 - Linacre, 1997 (variância modelada):
 - Lower bound (valor real mais baixo que o anterior): Real reliability baseada em erros modelados + desajuste
 - Além do erro anterior acrescenta o desajuste (misfit)
$$X_{ni} = P_{ni} \pm \sqrt{(P_{ni}(1 - P_{ni}))}$$
 - Modelada X $\max(1.0, \text{INFIT Mean square})$
- Precisão no Rasch = Variância Observada das Medidas – Variância de Erro Real / Variância Observada das Medidas
- Cronbach sempre superestima a precisão. Os dois valores serão diferentes tanto mais há escores extremos e dados incompletos (missing data)
- Local reliability (Daniel, 1999 p. 54)
 - COMPUTE Local Rel. = $1 - ((\text{erro rasch} ** 2) / (\text{Desv. Padr.} ** 2))$.

Precisão no modelo de Rasch

- Separação é um coeficiente que compara a dispersão verdadeira das medidas com o erro de medida. Indica a dispersão da amostra de pessoas em unidades de erro.
- Indica a taxa sinal / ruído
 - Separação (G)= Desvio padrão Verdadeiro / Erro (RMSE).
- Média dos Erros ao Quadrado = variância dos erros , a raiz dessa quantidade corresponde ao erro padrão médio (RMSE).
- $(\text{Desvio Padrão Verdadeiro})^2 = (\text{Desvio Padrão das Medidas})^2 - (\text{RMSE})^2$
- *Distinct Strata* = $(4G+1)/3$

Separation	KR-20, Alpha:	% Variance:	Distinct Strata:
Ratio: G	G^2/(1+G^2)	Not Due Error/Due Error	(4G+1)/3

0	.00	0/100	1
1	.50	50/50	1
1.5	.70	70/30	2
2	.80	80/20	3
3	.90	90/10	4
4	.94	94/6	5
5	.96	96/4	7
6	.97	97/3	8
7	.98	98/2	9

Item / Person Strata

