

Contents

Preface	xiii
Notation	xvii
Acronyms	xix
Author bios	xxiii
1 Introduction	1
1.1 Brief history of psychometrics	1
1.2 Measurement in social sciences	3
1.2.1 Educational measurement	4
1.2.2 Psychological assessment	6
1.2.3 Health-related outcome measures	6
1.2.4 Other areas of measurement	7
1.3 Measurement data in this book	7
1.4 Psychometrics with R	10
1.5 Exploring measurement data	12
1.5.1 Item scores	12
1.5.2 Test scores	16
1.5.3 Covariates and more complex data structures	18
1.6 Modeling measurement data	20
1.7 ShinyItemAnalysis interactive application	22
1.8 Summary	23
2 Validity	25
2.1 Introduction	25
2.2 Sources of validity-supporting evidence	25
2.2.1 Evidence based on test content	26
2.2.2 Evidence based on relations to other variables	26
2.2.3 Evidence based on internal structure	27
2.3 Statistical methods in test validation	27
2.3.1 Inferences based on ratios	27
2.3.2 One sample and a paired <i>t</i> test	29
2.3.3 Two sample <i>t</i> test	32
2.3.4 More samples – ANOVA	34
2.3.5 Correlation coefficients	37
2.3.6 Regression models	41
2.3.6.1 Simple linear regression	41
2.3.6.2 Multiple regression	45
2.3.6.3 More complex designs	47
2.4 Further issues	48

2.4.1	Estimation of model parameters	48
2.4.1.1	Ordinary least squares method	48
2.4.1.2	Maximum likelihood method	49
2.4.2	Model selection and model fit	50
2.4.3	Correction for range restriction	51
2.5	Validity in interactive application	53
2.6	Summary	54
3	Internal structure of the test and factor analysis	55
3.1	Introduction	55
3.2	Correlation structure	55
3.3	Cluster analysis	59
3.4	Factor analysis	61
3.4.1	Exploratory factor analysis	62
3.4.1.1	The single factor model	62
3.4.1.2	More factors	64
3.4.2	Factor rotation	65
3.4.3	Factor scores	67
3.4.4	The number of factors	68
3.4.5	Model selection and model fit	69
3.4.6	Confirmatory factor analysis	70
3.4.7	Hierarchical and more complex structures	74
3.5	Internal structure and FA in interactive application	75
3.6	Summary	76
4	Reliability	77
4.1	Introduction	77
4.2	Formal definition and properties in CTT	78
4.2.1	Definition of reliability	78
4.2.2	Reliability as correlation between measurements	79
4.2.3	Implications of low reliability	80
4.2.4	Rules of thumb	80
4.2.5	Reliability of composites	81
4.2.5.1	Spearman-Brown prophecy formula	81
4.2.6	Increasing reliability	83
4.3	Estimation of reliability	84
4.3.1	Reliability estimation with correlation coefficients	84
4.3.1.1	Test-retest reliability	84
4.3.1.2	Parallel forms	85
4.3.1.3	Split-half coefficient	85
4.3.2	Cronbach's alpha	87
4.3.2.1	Cronbach's alpha and inter-item correlations	88
4.4	Estimation of reliability with variance components	89
4.4.1	ANOVA method of estimation	89
4.4.1.1	One-way ANOVA	89
4.4.1.2	Two-way ANOVA and Cronbach's alpha	92
4.4.2	Maximum likelihood	96
4.4.3	Restricted maximum likelihood	97
4.4.4	Bootstrap confidence intervals	99
4.4.5	Bayesian estimation	100
4.5	More sources of error and G-theory	101

4.5.1	A one-facet study	103
4.5.2	A two-facet study	107
4.6	Other estimates of reliability	107
4.7	Reliability in interactive application	108
4.8	Summary	109
5	Traditional item analysis	111
5.1	Introduction	111
5.2	Item difficulty	111
5.2.1	Difficulty in binary items	112
5.2.2	Difficulty in ordinal items	113
5.3	Item discrimination	114
5.3.1	Correlation between item and total score	115
5.3.2	Difference between upper and lower group	116
5.4	Item characteristic curve	117
5.5	Distractor analysis	119
5.6	Reliability if an item is dropped	121
5.7	Item validity	121
5.8	Missed items	123
5.9	Item analysis in interactive application	124
5.10	Summary	125
6	Item analysis with regression models	127
6.1	Introduction	127
6.2	Model specification	127
6.3	Models for continuous items	129
6.3.1	Linear regression model	129
6.4	Models for binary items	131
6.4.1	Logistic regression model	132
6.4.2	Other link functions, probit regression model	133
6.4.3	IRT parametrization	135
6.4.4	Nonlinear regression models	136
6.5	Estimation of item parameters	139
6.5.1	Nonlinear least squares	139
6.5.2	Maximum likelihood method	140
6.6	Model selection	140
6.6.1	Likelihood-ratio test	141
6.6.2	Akaike information criterion	141
6.6.3	Bayesian information criterion	141
6.7	Models for polytomous items	142
6.7.1	Ordinal regression models	142
6.7.1.1	Cumulative logit model	142
6.7.1.2	Adjacent-categories logit model	145
6.7.2	Multinomial regression models	146
6.8	Joint model	150
6.8.1	Person-item map	152
6.9	Regression models in interactive application	153
6.10	Summary	154

7	Item response theory models	155
7.1	Introduction	155
7.2	General concepts and assumptions	156
7.2.1	IRT model assumptions	156
7.2.2	IRT models for binary data	158
7.2.2.1	Rasch or 1PL IRT model	158
7.2.2.2	2PL IRT model	159
7.2.2.3	Normal-ogive model	159
7.2.2.4	3PL IRT model	160
7.2.2.5	4PL IRT model	161
7.3	Estimation methods for IRT models	161
7.3.1	Heuristic methods and starting values	162
7.3.2	Joint maximum likelihood	162
7.3.3	Conditional maximum likelihood	165
7.3.4	Marginal maximum likelihood	166
7.3.4.1	Estimation of person abilities in MML	167
7.3.5	Bayesian IRT models	168
7.3.6	Item and test information	168
7.3.7	Model selection and model fit	169
7.4	Binary IRT models in R	171
7.4.1	The <code>mirt</code> package	171
7.4.2	The <code>ltm</code> package	176
7.4.3	The <code>eRm</code> package	178
7.4.4	Other IRT packages	179
7.4.4.1	The TAM package	179
7.4.4.2	The <code>ShinyItemAnalysis</code> package	179
7.4.5	The <code>lme4</code> and <code>nlme</code> packages	180
7.4.6	Bayesian IRT with the <code>brms</code> package	181
7.5	Relationship between IRT and factor analysis	181
7.6	IRT models in interactive application	183
7.7	Summary	186
8	More complex IRT models	187
8.1	Introduction	187
8.2	IRT models for polytomous items	187
8.2.1	Cumulative logit IRT models	187
8.2.1.1	Graded response model	187
8.2.1.2	Graded rating scale model	192
8.2.2	Adjacent-categories logit IRT models	193
8.2.2.1	Generalized partial credit model	193
8.2.2.2	Partial credit model	195
8.2.2.3	Rating scale model	196
8.2.3	Baseline-category logit IRT models	198
8.2.3.1	Nominal response model	198
8.2.4	Other IRT models for polytomous data	201
8.2.5	Item-specific IRT models	201
8.3	Multidimensional IRT models	202
8.3.1	Multidimensional 2PL model	202
8.3.2	Multidimensional graded response model	204
8.3.3	Confirmatory multidimensional IRT models	206
8.4	Estimation in more complex IRT models	208

8.4.1	Maximum likelihood methods	208
8.4.2	Regularization methods	208
8.4.3	Bayesian methods with MCMC and MH-RM	208
8.4.4	Model selection and model fit	209
8.5	More complex IRT models in interactive application	209
8.6	Summary	210
9	Differential item functioning	211
9.1	Introduction	211
9.2	Definition	211
9.2.1	DIF examples and interpretations	212
9.2.2	Matching criterion	214
9.3	Traditional DIF detection methods	214
9.3.1	Delta plot method	214
9.3.2	Mantel-Haenszel test	219
9.3.3	SIBTEST	223
9.4	DIF detection based on regression models	225
9.4.1	Logistic regression	226
9.4.1.1	Testing for DIF	227
9.4.2	Generalized logistic regression models	234
9.4.3	Group-specific cumulative logit model	235
9.4.4	Group-specific adjacent category logit model	238
9.4.5	Group-specific multinomial regression model	240
9.5	IRT-based DIF detection methods	243
9.5.1	Group-specific IRT models	243
9.5.2	Lord's test	246
9.5.3	Likelihood-ratio test	248
9.5.4	Raju's test	248
9.6	Other methods	251
9.6.1	Iterative hybrid ordinal logistic regression with IRT	251
9.6.2	Regularization approach for DIF detection	252
9.6.3	Measurement invariance: Factor analytic approach	253
9.7	DIF detection in interactive application	253
9.8	Summary	254
10	Outlook on applications and more advanced psychometric topics	255
10.1	Introduction	255
10.2	Computerized adaptive testing	255
10.2.1	Item bank	257
10.2.2	Ability estimation	257
10.2.3	Item selection algorithms	257
10.2.4	Stopping rules	258
10.2.5	CAT implementation in interactive application	259
10.2.6	Post-hoc analysis	261
10.2.7	CAT simulation study with MCMC	264
10.3	Test equating and linking	267
10.4	Generalizing latent variable models	267
10.5	Big data and computational psychometrics	268
10.6	Interactive psychometric modules	269
10.7	Summary	269

A Introduction to R	271
A.1 Obtaining and running R and RStudio	271
A.2 Starting with R	272
A.3 Installation of R packages	273
A.4 Data handling in R	274
A.4.1 Data types in R	276
A.4.2 Wide and long data format	277
A.4.3 Data handling with tidyverse	278
A.5 Graphics in R	279
A.5.1 Graphics in base R	279
A.5.2 Graphics with the ggplot2 package	280
A.5.3 Trellis graphics with the lattice package	281
A.6 Interactive Shiny applications	281
B Descriptive statistics	283
C Distributions of random variables	285
C.1 Discrete random variables	285
C.2 Continuous random variables	289
D Measurement data in ShinyItemAnalysis	293
E Exercises	295
E.1 Exercises for Chapter 1	295
E.2 Exercises for Chapter 2	296
E.3 Exercises for Chapter 3	297
E.4 Exercises for Chapter 4	297
E.5 Exercises for Chapter 5	299
E.6 Exercises for Chapter 6	299
E.7 Exercises for Chapter 7	301
E.8 Exercises for Chapter 8	302
E.9 Exercises for Chapter 9	303
E.10 Exercises for Chapter 10	304
References	305
Index	319