Contents

Preface vii
Acknowledgements ix

1 An Introduction to Measurement Models 1
1.1 Introduction 1
1.2 Statistical models for physical measurements 3
1.3 Psychometric tests: the common-factor model 7
1.4 Poisson process models 9
1.5 Binary response models 11
1.6 Identification and diagnosis 13
1.7 Comparison of measurements with a standard 13
1.8 Implications of measurement error 16
1.9 Exercises 18

2 Measures of Proximity, Agreement and Disagreement 19
2.1 Introduction 19
2.2 Similarity of interval- and ratio-scaled measurements 20
2.3 Similarity of binary responses 21
2.4 Similarity of categorical responses 23
2.5 Similarity of response profiles 25
2.6 Rater uncertainty 27
2.7 Chance agreement 28
2.8 Chance-corrected agreement 30
2.9 Intra-class correlation 34
2.10 Monotonic measures of agreement for ranked data 35
2.11 What is good agreement? 37
2.12 Exercises 38

3 Reliability, Consistency and Stability 40
3.1 Introduction to reliability 40
3.2 Reliability of composite scores 44
3.3 Split halves and the internal consistency of tests 45
3.4 Repeated testing and stability 50
3.5 Reliability of counts and proportions 54
3.6 Generalisability theory 56
3.7 Exercises 57

4 Designs for Reliability Studies 59
4.1 Introduction 59
4.2 Choice of subjects 60

4.3	Selection of measuring instruments and measurement conditions	61
4.4	Simple replications	63
4.5	Multiple indicators	64
4.6	Randomised blocks designs	65
4.7	Crossing and nesting	68
4.8	'Case-control' sampling	69
4.9	Sample size requirements	71
5	**Covariance-Components Models**	**73**
5.1	Introduction: a model for string lengths	73
5.2	Precision of reliability estimation	77
5.3	Pairwise comparisons	80
5.4	Maximum likelihood estimation	84
5.5	Confirmatory factor analysis	87
5.6	Pitfalls for the unwary	94
5.7	Simultaneous factor analysis in several populations	96
5.8	Analysis of data with missing values	101
5.9	Longitudinal factor analysis and structural equation models	105
5.10	Further consideration of sample size requirements	109
5.11	Exercises	112
6	**Components of Variance**	**114**
6.1	A simple replication study	114
6.2	Intra-class correlation	117
6.3	Expected mean squares	118
6.4	Maximum likelihood estimators	120
6.5	Lack of balance (missing data)	122
6.6	Two-way mixed effects models	125
6.7	The general mixed model	127
6.8	Analysis of nested data	129
6.9	Generalisability studies	133
6.10	Exercises	134
7	**Statistical Methods for Categorical Measurements**	**136**
7.1	Assessment of diagnostic tests	136
7.2	Tests of observer bias	141
7.3	Variance components models for binary data	145
7.4	Measures of agreement for multinomial data	148
7.5	Reliability of unanimous and majority opinions	158
7.6	Exercises	160
Appendix 1	Expected Values	162
Appendix 2	Linear Combinations of Measurements	165
Appendix 3	The δ Technique	170
Appendix 4	Computer Software	177
Selected Bibliography		180
Index		191