

Contents

1

Introduction 1

- 1.1 The Use of Tests 1
- 1.2 Definition of Measurement 2
- 1.3 History of Testing and Measurement 2
- 1.4 Organization of This Book 4
- 1.5 Standards for Test Users 4
- 1.6 Vocabulary 5
- 1.7 Study Questions 5

2

A Review of Basic Statistical Concepts 6

- 2.1 Introduction 6
- 2.2 Levels of Measurement 6
- 2.3 Common Statistical Notation and Definitions 9
- 2.4 Distributions and Probabilities 12

2.5	Descriptive Statistics	16
2.6	Inferential Statistics	18
2.7	The Normal Distribution	20
2.8	The Pearson Correlation Coefficient	23
2.9	Linear Regression	31
2.10	Influences on Correlation Coefficients	34
2.11	Alternative Measures of Association	36
2.12	Introduction to Advanced Topics	42
2.13	(Optional) Joint and Conditional Probabilities	42
2.14	(Optional) The Algebra of Expectations	44
2.15	Summary	48
2.16	Vocabulary	49
2.17	Study Questions	50
2.18	Computational Problems	52

3 Classical True-Score Theory 56

3.1	The Assumptions of Classical True-Score Theory	56
3.2	Summary of Classical True-Score Theory	60
3.3	Conclusions Derived from Classical True-Score Theory	60
3.4	(Optional) Proofs of Conclusions Derived from Classical True-Score Theory	65
3.5	Vocabulary	70
3.6	Study Questions	71
3.7	Computational Problems	71

4 Reliability 72

4.1	Introduction	72
4.2	Test/Retest Reliability Estimates	76
4.3	Parallel-Forms and Alternate-Forms Reliability Estimates	77
4.4	Internal-Consistency Estimates of Reliability: Split Halves	78
4.5	Internal-Consistency Reliability: The General Case	83
4.6	The Spearman-Brown Formula: The General Case	85

4.7	Comparison of Methods of Estimating Reliabilities	88
4.8	Standard Errors of Measurement and Confidence Intervals for True Scores	88
4.9	Summary	91
4.10	Vocabulary	92
4.11	Study Questions	92
4.12	Computational Problems	93

5 **Validity** **95**

5.1	Introduction	95
5.2	Content Validity	95
5.3	Criterion-Related Validity	97
5.4	Correction for Attenuation	98
5.5	Estimating a Criterion from a Predictor	99
5.6	Screening into Dichotomous Categories	101
5.7	Taylor-Russell Tables	104
5.8	Multiple Predictors or Criteria	108
5.9	Construct Validity	108
5.10	Multitrait-Multimethod Validity	109
5.11	Factorial Validity	111
5.12	Summary	113
5.13	Vocabulary	114
5.14	Study Questions	115
5.15	Computational Problems	116

6 **Principles of Test Construction** **118**

6.1	Introduction	118
6.2	Item Formats	119
6.3	Item Analysis Using Indices of Item Difficulty and Discrimination	120
6.4	Item Analysis Using Item Reliability and Validity Indices	124
6.5	Item Analysis Using Item-Characteristic Curves	127
6.6	Using Factor Analysis for Item Analysis and Test Development	130
6.7	Distributions of Test Scores	133
6.8	Some General Considerations in Item Analysis	138

6.9	Comparison of Item-Analysis Methods	140
6.10	Selecting a Published Test	140
6.11	Summary	142
6.12	Vocabulary	144
6.13	Study Questions	144
6.14	Computational Problems	145

7 Transforming and Equating Test Scores 148

7.1	Introduction	148
7.2	Percentiles	150
7.3	Age and Grade Scores	157
7.4	Expectancy Tables	160
7.5	Standard and Standardized Scores	161
7.6	Normalized Scores	163
7.7	Corrections for Guessing and Omissions	166
7.8	Equal-Interval Scales	168
7.9	Equating Test Scores	170
7.10	Summary	173
7.11	Vocabulary	174
7.12	Study Questions	175
7.13	Computational Problems	175

8 Scaling 178

8.1	Introduction	178
8.2	Determining the Level of Measurement	179
8.3	Levels of Measurement Produced by Selected Scales	182
8.4	Additional Methods for Obtaining Scales	184
8.5	Summary	190
8.6	Vocabulary	191
8.7	Study Questions	192
8.8	Computational Problems	192

9 Special Considerations in the Use of Measurements 193

9.1	Introduction	193
9.2	Group Heterogeneity	193
9.3	Combining Scores; Multiple Regression	200

- 9.4 Suppressor Variables, Moderator Variables, and Bias in Selection 204
- 9.5 Measuring Change 208
- 9.6 Profile Analysis 211
- 9.7 Implications of Measurement Theory for the Design and Interpretation of Research Studies 213
- 9.8 Summary 215
- 9.9 Vocabulary 216
- 9.10 Study Questions 217
- 9.11 Computational Problems 218

10 Controversies and Current Developments in Measurement 222

- 10.1 Introduction 222
- 10.2 Clinical and Statistical Prediction 222
- 10.3 Homogeneous and Nonhomogeneous Tests 224
- 10.4 Criticisms of Standardized Testing 225
- 10.5 Criterion-Referenced Tests 227
- 10.6 Strong True-Score Theories and Latent-Trait Theories 229
- 10.7 Generalizability Theory 230
- 10.8 Bayesian Methods 232
- 10.9 Tailored Testing 232
- 10.10 Computer-Interpreted Testing 234
- 10.11 Summary 235
- 10.12 Vocabulary 236
- 10.13 Study Questions 237

11 Strong True-Score Theories and Latent-Trait Models 239

- 11.1 Introduction 239
- 11.2 Local Independence 241
- 11.3 Binomial-Error Models 242
- 11.4 Poisson Models 249
- 11.5 Normal-Ogive and Logistic Models 253
- 11.6 Rasch's Logistic Model 260
- 11.7 Test Information 262
- 11.8 Information Functions and the Logistic Models 267
- 11.9 Summary 269

11.10	Vocabulary	270
11.11	Study Questions	270
11.12	Computational Problems	272
Appendix: Areas under the Standard Normal Curve		275
Glossary of Symbols		278
Answers to Computational Problems		284
References		291
Name Index		296
Subject Index		298