

Contents

<u>chap:acknow</u> Acknowledgment	xvii
Part I: Preliminaries	1
1 Introduction	2
1.1 <i>What is Inter-Rater Reliability?</i>	4
1.2 <i>Scope and Design of Inter-Rater Reliability Experiments</i>	10
1.2.1 <i>Scope of the Investigation</i>	11
1.2.2 <i>Experimental Design</i>	14
1.3 <i>Scoring of Subjects/Items</i>	17
1.4 <i>Formulation of Agreement Coefficients</i>	23
1.4.1 <i>Analysis of Variance Models</i>	24
1.4.2 <i>Intraclass Correlation Coefficients</i>	25
1.5 <i>Different Reliability Types</i>	26
1.5.1 <i>Measures of Association</i>	26
1.5.2 <i>Univariate Inter-Rater Reliability</i>	27
1.5.3 <i>Multivariate Inter-Rater Reliability</i>	27
1.5.4 <i>Special Agreement Coefficients for Quantitative Ratings</i>	28
1.6 <i>Statistical Inference</i>	29
1.7 <i>Book's Structure</i>	30
1.8 <i>Choosing the Right Method</i>	32
2 Setting Up Databases of Ratings for Analysis	35
2.1 <i>Introduction</i>	36
2.2 <i>The Notions of Subject and Characteristic</i>	40
2.3 <i>The Notion of Agreement in a Multiple-Level Rating Process</i>	44
2.4 <i>Multiple Ratings per Rater and per Subject</i>	46
2.5 <i>Dealing with the Notion of Rater</i>	47
2.6 <i>Concluding Remarks</i>	49
Part II: Intraclass Correlation Coefficients	51

3	Intraclass Correlation: A Measure of Raters' Agreement	52
3.1	<i>Introduction</i>	53
3.2	<i>Statistical Models</i>	54
3.3	<i>The Bland-Altman Plot</i>	57
3.4	<i>Sample Size Calculations</i>	60
3.5	<i>Multivariate Analysis</i>	61
3.5.1	<i>The Principal Component</i>	63
3.5.2	<i>The Multivariate ICC</i>	67
3.6	<i>Concluding Remarks</i>	68
4	Intraclass Correlations in One-Factor Studies	72
4.1	<i>Intraclass Correlation under Model 1A</i>	74
4.1.1	<i>Defining Inter-Rater Reliability</i>	74
4.1.2	<i>Calculating Inter-Rater Reliability</i>	75
4.1.3	<i>Defining Intra-Rater Reliability</i>	79
4.1.4	<i>Recommendations</i>	79
4.2	<i>Intraclass Correlation under Model 1B</i>	79
4.2.1	<i>Defining Intra-Rater Reliability</i>	80
4.2.2	<i>Calculating Intra-Rater Reliability</i>	82
4.3	<i>Statistical Inference about ICC under Models 1A and 1B</i>	85
4.3.1	<i>Confidence Interval for ρ under Model 1A</i>	86
4.3.2	<i>p-Value for ρ under Model 1A</i>	89
4.3.3	<i>Confidence Interval for γ under Model 1B</i>	91
4.3.4	<i>p-Value for γ under Model 1B</i>	94
4.4	<i>Sample Size Calculations</i>	96
4.4.1	<i>Sample Size Calculations under Model 1A: The Statistical Power Approach</i>	96
4.4.2	<i>Sample Size Calculations under Model 1A: The Confidence Interval Approach</i>	102
4.4.3	<i>Sample Size Calculations under Model 1B</i>	110
4.5	<i>Concluding Remarks</i>	111
5	Intraclass Correlations under the Random Factorial Design	113
5.1	<i>The Issues</i>	115
5.2	<i>The Intraclass Correlation Coefficients</i>	117
5.2.1	<i>Inter-Rater Reliability Coefficient</i>	120
5.2.2	<i>Intra-Rater Reliability Coefficient</i>	126
5.3	<i>Statistical Inference About the ICC</i>	128
5.3.1	<i>Statistical Inference about Inter-Rater Reliability Coefficient ρ</i>	129
5.3.2	<i>Statistical Inference about Intra-Rater Reliability Coefficient γ</i>	135
5.4	<i>Sample Size Calculations</i>	138

5.4.1	<i>Sample Sizes for Inter-Rater Reliability Studies: the Statistical Test Power Approach</i>	139
5.4.2	<i>Sample Sizes for Intra-Rater Reliability Studies: the Statistical Test Power Approach</i>	144
5.4.3	<i>Sample Size for Inter-Rater Reliability Studies: The Confidence Interval Approach</i>	147
5.4.4	<i>Sample Size for Intra-Rater Reliability Studies: The Confidence Interval Approach</i>	154
5.5	<i>Special Topics</i>	158
5.5.1	<i>Rater Reliability for a Random Factorial Model Without Interaction</i>	158
5.5.2	<i>How are the Power Curves Obtained?</i>	165
5.6	<i>Concluding Remarks</i>	169
6	Intraclass Correlations under the Mixed Factorial Design	171
6.1	<i>The Problem</i>	173
6.2	<i>Intraclass Correlation Coefficient</i>	174
6.2.1	<i>Defining Inter-Rater Reliability</i>	175
6.2.2	<i>Defining Intra-Rater Reliability</i>	178
6.2.3	<i>Calculating Inter-Rater and Intra-Rater Reliability Coefficients</i>	179
6.3	<i>Statistical Inference About the ICC</i>	185
6.3.1	<i>Confidence Interval for the Inter-Rater Reliability Coefficient</i>	186
6.3.2	<i>Confidence Interval for the Intra-Rater Reliability Coefficient</i>	188
6.3.3	<i>p-Value of Inter-Rater Reliability Coefficient</i>	191
6.3.4	<i>p-Value of Intra-Rater Reliability Coefficient</i>	193
6.4	<i>Sample Size Calculations</i>	195
6.4.1	<i>Inter-Rater Reliability Sample Size Calculation: the Statistical Test Power Approach</i>	195
6.4.2	<i>Intra-Rater Reliability Sample Size Calculation: the Statistical Test Power Approach</i>	198
6.4.3	<i>Sample Size Calculations: the Confidence Interval Approach</i>	201
6.4.4	<i>Inter-Rater Reliability Sample Size Calculation: the Confidence Interval Approach</i>	202
6.4.5	<i>Intra-Rater Reliability Sample Size Calculation: the Confidence Interval Approach</i>	208
6.5	<i>Special Topics</i>	214
6.5.1	<i>Calculations of RSS and h_6 Related to Equation 6.2.12</i>	214
6.5.2	<i>How are the Power Curves Obtained ?</i>	216
6.6	<i>Concluding Remarks</i>	220

7	Finn's Coefficient of Reliability	222
7.1	<i>The Problem</i>	223
7.2	<i>Finn's Coefficient</i>	225
7.2.1	<i>Finn's Coefficient: Computations</i>	225
7.2.2	<i>Finn's Coefficient: General Definition</i>	227
7.3	<i>Statistical Inference</i>	228
7.3.1	<i>Inference with a homogeneous subject sample</i>	229
7.3.2	<i>Inference with a representative subject sample</i>	233
7.3.3	<i>Sample size for a given error margin E</i>	234
7.3.4	<i>Sample size for a given MDD</i>	235
7.4	<i>Generalized Finn's Coefficient</i>	239
7.4.1	<i>Generalized Finn's Inter-Rater Reliability Coefficient</i>	240
7.4.2	<i>Influence of Individual Raters on Inter-Rater Agreement</i>	242
7.4.3	<i>Rater Agreement with the Gold Standard</i>	244
7.5	<i>Advantages, Disadvantages & Concluding Remarks</i>	245
8	Measures of Association and Concordance	247
8.1	<i>Overview</i>	248
8.2	<i>Pearson & Spearman Correlation Coefficients</i>	248
8.2.1	<i>Pearson's Correlation Coefficient</i>	249
8.2.2	<i>Spearman's Correlation Coefficient</i>	251
8.3	<i>Kendall's Tau</i>	253
8.3.1	<i>Computing Kendall's Tau in the Absence of Ties</i>	254
8.3.2	<i>Computing Kendall's Tau in the Presence of Ties</i>	256
8.3.3	<i>p-Value for Kendall's Tau</i>	258
8.4	<i>Kendall's Coefficient of Concordance (KCC)</i>	258
8.5	<i>Lin's Concordance Correlation Coefficient (LCCC)</i>	262
8.6	<i>Concluding Remarks</i>	262
9	Analysis of Quantitative Inter-Rater Reliability Data	264
9.1	<i>Overview</i>	265
9.2	<i>Influence Analysis</i>	265
9.2.1	<i>Benchmarking Intraclass Correlation Coefficients</i>	265
9.2.2	<i>Identification of Problem Raters</i>	265
9.3	<i>A Multivariate Approach to Inter-Rater Reliability</i>	265
9.3.1	<i>Review of Existing Multivariate Approaches</i>	265
9.3.2	<i>Introduction to Principal Components Analysis</i>	265
9.3.3	<i>Multivariate Coefficients with Quantitative Ratings</i>	266
	Part III: Appendices	267
	A Data Tables	268

Bibliography	275
List of Notations	285
Author Index	287
Subject Index	289
