

CONTENTS

| | |
|---|------|
| Preface | xiii |
| 1 ANALYSIS OF TWO-WAY TABLES | |
| 1.1 INTRODUCTION | 1 |
| 1.1.1 The Nature of Categorical Variables | 3 |
| 1.1.2 Sampling Models | 4 |
| 1.1.3 Notation | 5 |
| 1.1.4 Marginal Tables | 6 |
| 1.2 A SHORT REVIEW OF THE CHI SQUARE TEST FOR INDEPENDENCE | 7 |
| 1.2.1 Chi Square Tests Are Approximations | 9 |
| 1.2.2 Limitations of the Chi Square Test | 10 |
| 1.3 EXAMINING RESIDUALS | 11 |
| 1.4 PARTITIONING CHI SQUARE | 13 |
| 1.4.1 A General Method of Partitioning | 13 |
| 1.4.2 Partitioning into 2×2 Tables | 15 |
| 1.5 QUASI-INDEPENDENCE | 18 |
| 1.6 THE ANALYSIS OF SQUARE TABLES | 24 |
| 1.6.1 Symmetry | 24 |
| 1.6.2 Marginal Symmetry | 25 |
| 1.7 SUMMARY | 26 |
| 2 MEASURES OF ASSOCIATION FOR NOMINAL DATA | |
| 2.1 INTRODUCTION | 27 |
| 2.1.1 The Interpretation of Measures of Association | 27 |
| 2.1.2 Confounding Factors | 30 |
| 2.1.3 Standardizing Tables | 31 |
| 2.2 NOMINAL MEASURES OF ASSOCIATION FOR 2×2 TABLES | 34 |
| 2.2.1 Percents | 34 |
| 2.2.2 The Odds Ratio | 35 |
| | vii |

| | | |
|-------|--|----|
| 2.2.3 | Yule's Q | 39 |
| 2.2.4 | Yule's Y | 40 |
| 2.2.5 | A Measure Based on Chi Square, Φ^2 | 41 |
| 2.2.6 | The Correlation Coefficient, ρ | 42 |
| 2.3 | NOMINAL MEASURES OF ASSOCIATION FOR $I \times J$ TABLES | 43 |
| 2.4 | THE ODDS RATIO IN $I \times J$ TABLES | 43 |
| 2.4.1 | Definition | 43 |
| 2.4.2 | Simultaneous Confidence Intervals | 44 |
| 2.5 | MEASURES BASED ON CHI SQUARE | 46 |
| 2.5.1 | Phi Squared | 46 |
| 2.5.2 | The Contingency Coefficient | 46 |
| 2.5.3 | Tschuprow's T | 47 |
| 2.5.4 | Cramér's V | 47 |
| 2.6 | PROPORTIONAL-REDUCTION-IN-ERROR MEASURES | 47 |
| 2.6.1 | Goodman and Kruskal's Lambda | 48 |
| 2.6.2 | Variance of Lambda | 50 |
| 2.6.3 | Symmetric Version of Lambda | 51 |
| 2.6.4 | Goodman and Kruskal's Tau | 52 |
| 2.6.5 | An Alternative Interpretation of Tau | 54 |
| 2.6.6 | Behavior of Measures of Association | 55 |
| 2.7 | MEASURES OF AGREEMENT | 58 |
| 2.7.1 | Unweighted κ | 58 |
| 2.7.2 | Weighted κ | 59 |
| 2.7.3 | The Variance of $\hat{\kappa}_w$ | 60 |
| 2.7.4 | Extensions | 61 |
| 2.8 | PREDICTION LOGIC | 61 |
| 2.8.1 | Prediction Logic Propositions | 61 |
| 2.8.2 | An Index of Prediction Success | 62 |
| 2.8.3 | The Variance of ∇ | 63 |
| 2.8.4 | Brief Discussion of Prediction Logic | 64 |
| 3 | MEASURES OF ASSOCIATION FOR ORDINAL DATA | |
| 3.1 | INTRODUCTION | 66 |
| 3.1.1 | Definitions of Perfect Correlation | 67 |
| 3.1.2 | The Concept of Pairs of Observations | 69 |
| 3.2 | ORDINAL CORRELATION COEFFICIENTS | 72 |
| 3.2.1 | Kendall's τ_a | 72 |
| 3.2.2 | Goodman and Kruskal's γ | 74 |
| 3.2.3 | Asymmetric Measures, d_{AB} and d_{BA} | 75 |
| 3.2.4 | Wilson's e | 76 |
| 3.2.5 | Kendall's τ_b | 78 |

| | | |
|----------|---|-----|
| 3.2.6 | τ_c | 78 |
| 3.2.7 | Relationships among the Measures | 79 |
| 3.3 | ALTERNATIVE INTERPRETATIONS | 81 |
| 3.3.1 | A Parametric Interpretation of Ordinal Correlation | 81 |
| 3.3.2 | Proportional-Reduction-in-Error Interpretations | 84 |
| 3.4 | STANDARD ERRORS AND STATISTICAL TESTS | 85 |
| 4 | MULTIVARIATE ANALYSIS OF CROSS-CLASSIFICATIONS | |
| 4.1 | INTRODUCTION | 89 |
| 4.1.1 | Interaction | 90 |
| 4.1.2 | Partial Associations | 91 |
| 4.1.3 | Zero Partial Association | 92 |
| 4.1.4 | Multiple Relationships | 93 |
| 4.2 | TEST-FACTOR STRATIFICATION | 93 |
| 4.2.1 | The Limitations of Test-Factor Stratification | 96 |
| 4.2.2 | Test-Factor Standardization | 96 |
| 4.3 | COEFFICIENTS OF PARTIAL ASSOCIATION FOR NOMINAL AND ORDINAL DATA | 98 |
| 4.3.1 | Treating Ordinal as Interval Scales | 98 |
| 4.3.2 | Coefficients of Partial Association and Correlation | 100 |
| 4.3.3 | A Shortcoming of Measures of Nominal and Ordinal Partial Association | 107 |
| 5 | LOG-LINEAR MODELS | |
| 5.1 | INTRODUCTION | 109 |
| 5.2 | LOG-LINEAR MODELS FOR TWO-WAY TABLES | 111 |
| 5.2.1 | No Effects Model | 112 |
| 5.2.2 | Model Containing only Row Effects | 113 |
| 5.2.3 | Independence Model | 114 |
| 5.2.4 | Saturated Model | 115 |
| 5.2.5 | The Hierarchy Principle | 116 |
| 5.3 | MODELS FOR THREE-WAY TABLES | 117 |
| 5.3.1 | Mutual Independence Model | 117 |
| 5.3.2 | Model with One Two-Factor Interaction | 118 |
| 5.3.3 | A Note on Calculating Parameters | 120 |
| 5.3.4 | Model with Two Two-Factor Interactions | 121 |
| 5.3.5 | Model with All Pairwise Interactions | 121 |
| 5.3.6 | Saturated Model | 122 |
| 5.3.7 | Independent Parameters | 124 |
| 5.4 | MODELS FOR HIGHER-DIMENSIONAL TABLES | 125 |
| 5.4.1 | Mutual Independence Model | 125 |
| 5.4.2 | Model with All Two-Factor Interactions | 126 |
| 5.4.3 | A Model with One Three-Factor Interaction | 126 |

| | | |
|-------|---|-----|
| 5.4.4 | A Model with Four Three-Factor Interactions | 128 |
| 5.4.5 | Saturated Four-Variable Model | 130 |
| 5.4.6 | Additional Models for Four or More Variables | 130 |
| 5.5 | AN ALTERNATIVE VIEW OF LOG-LINEAR MODELS | 131 |
| 5.5.1 | Path Analysis and Diagrams | 133 |
| 5.5.2 | Limitations of Path Models | 134 |
| 5.6 | THE EFFECTS OF COLLAPSING VARIABLES | 135 |
| 5.6.1 | An Algorithm for Determining Collapsibility | 136 |
| 5.6.2 | Summary | 137 |
| 5.7 | A NOTE ON THE CALCULATION OF μ TERMS | 139 |
| 5.7.1 | Saturated Models | 139 |
| 5.7.2 | Calculating μ Terms in Unsaturated Models | 140 |
| 5.7.3 | Variances of μ Terms | 140 |
| 6 | ESTIMATING AND PARTITIONING LOG-LINEAR MODELS | |
| 6.1 | INTRODUCTION | 142 |
| 6.1.1 | Maximum Likelihood Estimation of Log-Linear Models | 142 |
| 6.1.2 | Direct versus Indirect Estimates | 144 |
| 6.1.3 | How to Tell if Explicit Formulas for Expected Frequencies Exist | 145 |
| 6.2 | FORMULAS FOR CALCULATING EXPECTED FREQUENCIES | 146 |
| 6.2.1 | Notation | 146 |
| 6.2.2 | Goodman's Algorithm for Excluded Variables | 147 |
| 6.2.3 | Goodman's Algorithm for Included Variables | 148 |
| 6.2.4 | Examples | 149 |
| 6.2.5 | Summary of Goodman's Method for Obtaining Direct Estimates | 152 |
| 6.3 | ESTIMATION BY ITERATION | 152 |
| 6.3.1 | Iterative Proportional Fitting for Three Variables | 153 |
| 6.3.2 | Iterative Proportional Fitting in General | 156 |
| 6.3.3 | Iteration for Incomplete Tables | 158 |
| 6.4 | TEST STATISTICS | 158 |
| 6.5 | THE PROBLEM OF SAMPLING ZEROES | 159 |
| 6.5.1 | Estimating Parameters in Saturated Models | 160 |
| 6.5.2 | Testing Unsaturated Models | 160 |
| 6.5.3 | Estimating Parameters in Unsaturated Models | 162 |
| 6.6 | PARTITIONING MODELS | 163 |
| 6.6.1 | Nested Models | 164 |
| 6.6.2 | Tests on Marginal Tables | 166 |
| 6.6.3 | A General Method of Partitioning | 167 |
| 6.6.4 | Partitioning Two Common Models | 173 |

| | | |
|-------|--|-----|
| 6.7 | CRITERIA FOR ACCEPTING MODELS | 175 |
| 6.7.1 | Problems in Defining Acceptance Criteria | 175 |
| 6.7.2 | Congruence with Substantive Theory | 176 |
| 6.7.3 | Simplicity | 176 |
| 6.7.4 | Alternative Measures of Fit | 177 |
| 7 | REGRESSION ANALYSIS OF CROSS-CLASSIFICATIONS | |
| 7.1 | INTRODUCTION | 180 |
| 7.2 | MULTIPLE REGRESSION WITH CATEGORICAL VARIABLES | 181 |
| 7.2.1 | Coding Categorical Independent Variables | 182 |
| 7.2.2 | Dichotomous Dependent Variables | 185 |
| 7.2.3 | Heteroscedasticity | 186 |
| 7.3 | WEIGHTED LEAST SQUARES: THE GSK APPROACH | 187 |
| 7.3.1 | An Overview of GSK | 187 |
| 7.3.2 | The Arrangement of the Data in GSK Analysis | 188 |
| 7.3.3 | Defining the Dependent Variable: Response Functions | 193 |
| 7.3.4 | Defining Models | 198 |
| 7.3.5 | Interpretation of Model Parameters | 200 |
| 7.3.6 | Significance Tests | 205 |
| 7.3.7 | Examples | 207 |
| 7.3.8 | Conclusion: A Test for Marginal Symmetry | 211 |
| | Appendix A: VECTORS AND MATRICES | 214 |
| | Appendix B: DISTRIBUTION OF CHI SQUARE | 218 |
| | Appendix C: DATA SOURCES | 220 |
| | References | 221 |
| | Index | 231 |