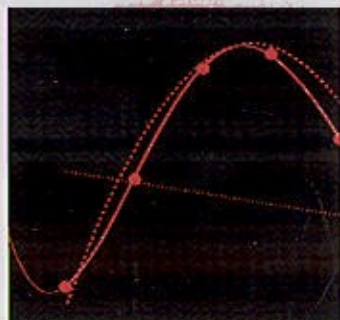
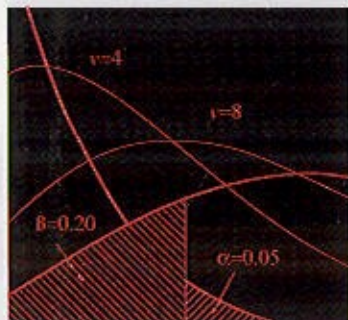


MULTIVARIATE METHODS IN EPIDEMIOLOGY



Theodore R. Holford

Contents

Part I Concepts and Definitions

1. Associations between Exposure and Disease, 3

Strategies for Studying Disease in a Population, 4

Individuals Followed for Equal Times, 4

Individuals Followed for Varying Times, 5

Studies of Cases and Controls, 6

Vital Statistics and Disease Registries, 8

Rationale for Multivariate Methods, 9

Statistical Approaches to Data Analysis, 10

How to Use This Book, 11

2. Models for Disease, 15

Models for the Disease Process, 15

Stochastic Models, 17

Basic Disease Model, 19

Outcomes for Analysis, 20

Theoretical Relationship among Outcomes, 21

Models for the Effect of Factors, 22

Models for Rates, 23

Models for Risks, 31

Summary, 34

Part II Non-Regression Methods

3. Analysis of Proportions, 39

Studies That Yield Proportion Responses, 39

Cross-Sectional Studies, 40

Cohort Studies, 41

Case-Control Studies, 42

A Single Binary Risk Factor, 43

Likelihood-Based Inference, 47

Wald Test, 48

Score Test: Pearson Chi-Square, 48

Likelihood Ratio Test, 51

Exact Test, 51

Interval Estimates of the Odds Ratio, 54

Logit Method, 54

Cornfield's Method, 55

Exact Method, 56

More Than Two Levels of Exposure, 58

Nominal Categories, 58

Ordered Categories, 62

Stratified Analysis, 63

Cochran and Mantel-Haenszel Tests, 65

Mantel-Haenszel Estimator for the Odds Ratio, 68

Testing for Homogeneity of the Odds Ratio, 70

Stratified Analysis of Trends, 73

Summary, 74

Exercises, 75

4. Analysis of Rates, 81

Rates as an Outcome, 82

Censored Observations with Constant Hazard, 82

Population Rates, 85

Standardized Ratios, 86

Comparing Nominal Groups, 88

Score Test for Trend, 94

Stratification, 95

Summary Rates, 97

Summary, 103

Exercises, 104

5. Analysis of Time to Failure, 109

*Estimating Survival Curves, 109**Actuarial Estimate, 110**Piecewise Exponential Estimate, 113**Product Limit Estimate, 115**Graphical Displays, 119**Diagnostic for Constant Hazard, 119**Diagnostic for Weibull Hazard, 121**Diagnostic for Proportional Hazards, 124**Tests for Comparing Hazards, 126**Piecewise Constant Hazards, 127**Log-Rank Test, 129**Stratified Log-Rank Test, 132**Types of Incomplete Data, 133**Right Censoring, 133**Left Censoring, 134**Interval Censoring, 135**Summary, 135**Exercises, 136*

Part III Regression Methods

6. Regression Models for Proportions, 141

*Generalized Linear Model for Proportions, 142**Linear Logistic Model, 144**Log-Linear Hazard Model, 145**Probit Model, 146**Fitting Binary Response Models, 147**Fitting the Linear Logistic Model, 147**Fitting Alternatives to the Logistic Model, 151**Problems in Model Fitting, 156**Summary, 159**Exercises, 160*

7. Defining Regressor Variables, 163

*Categorical Variables, 163**0 1 Coding, 165**-1 0 1 Coding, 169*

<i>Deviations from the Model</i>	172
<i>Threshold and Other Models for Trend</i>	176
<i>Interactions and Parallelism</i>	182
<i>Testing Linear Hypotheses</i>	186
<i>Wald Tests for Linear Hypotheses</i>	189
<i>Likelihood Ratio Tests for Linear Hypotheses</i>	192
<i>Power Transformations of Continuous Regressor Variables</i>	195
<i>Summary</i>	200
<i>Exercises</i>	202

8. Parametric Models for Hazard Functions, 205

<i>Constant Hazard Model</i>	206
<i>Log-Linear Model for Rates</i>	206
<i>Generalized Linear Model</i>	207
<i>Alternatives to Log-Linear Models</i>	212
<i>Models for Standardized Ratios</i>	215
<i>Weibull Hazard Model</i>	218
<i>Other Parametric Models</i>	220
<i>Gamma Failure Times</i>	220
<i>Log-Normal Failure Time</i>	221
<i>Gompertz-Makeham Model</i>	222
<i>Extreme-Value Distribution</i>	222
<i>Extra Poisson Variation</i>	223
<i>Summary</i>	224
<i>Exercises</i>	225

9. Proportional Hazards Regression, 227

<i>Piecewise Constant Hazards Model</i>	228
<i>Proportional Piecewise Constant Hazards</i>	228
<i>Testing for Constant Hazards</i>	230
<i>Alternatives to Proportional Hazards</i>	233
<i>Nonparametric Proportional Hazards</i>	235
<i>Evaluating the Fit of a Proportional Hazards Model</i>	239
<i>Time-Dependent Covariates</i>	240
<i>Effects That Vary with Time</i>	241
<i>Exposures That Vary with Time</i>	245
<i>Summary</i>	247
<i>Exercises</i>	249

Part IV Study Design and New Directions

10. Analysis of Matched Studies, 253

Designing Matched Studies, 254

Matched Pairs, 254

Many-to-One Matching, 255

Frequency Matching, 256

Caliper Matching, 256

Strategies for Analysis, 257

Case-Control Studies with Matched Pairs, 258

McNemar's Test and Its Extensions, 258

Logistic Regression Method I, 262

Logistic Regression Method II, 265

Case-Control Studies with More Than One Control per Case, 268

Cohort Studies with Matched Pairs, 272

Summary, 276

Exercises, 277

11. Power and Sample Size Requirements, 281

Estimation, 281

Proportions, 282

Rates, 283

Association Measure, 285

Two-Group Hypothesis Tests, 286

Proportion Comparisons, 288

Rate Comparisons, 289

Survival Curves, 291

Matched Pairs, 292

General Hypothesis Tests, 295

Steps for Calculating the Noncentrality Parameter, 298

Testing for Trend, 298

Covariate Adjustment, 305

Interactions, 308

Simulation, 309

Summary, 310

Exercises, 311

12. Extending Regression Models, 315

Classification and Regression Trees (CART), 316

Growing Trees, 317

Trees for Binary Responses, 320

CART for Survival Data, 325

Splines, 326

Linear Splines, 326

Polynomial Splines, 327

GAMS and MARS, 329

Missing Observations, 330

Variance Components, 332

Errors in Variables, 333

Collinearity, 335

Summary, 339

Appendix 1. Theory on Models for Disease, 343

Constant Rates in Time, 345

Idealized Model for Rates Changing over Time, 346

Observed Effects of Age on Rates, 347

Relationship between Models for Rates and Proportions, 347

Appendix 2. Theory on Analysis of Proportions, 351

Likelihood for the Linear Logistic Model, 351

Wald Statistic for a 2×2 Table, 353

Likelihood Ratio Statistic for a 2×2 Table, 354

Score Statistics for an 1×2 Table, 355

Score Statistics for Combining 1×2 Tables, 357

Appendix 3. Theory on Analysis of Rates, 361

Likelihood Formation, 361

Time to Failure Models, 361

Counts of Failures, 362

Comparisons to Standard Rates, 363

Estimation, 363

Score Test for Nominal Categories, 364

Score Test Controlling for Strata, 366

Appendix 4. Theory on Analysis of Time to Failure, 369

Actuarial Estimate, 369

Product Limit Estimate, 372

Two-Sample Score Test for Piecewise Constant Hazards, 372
Log-Rank Test, 374

Appendix 5. Theory on Regression Models for Proportions, 377

Distribution for Binary Responses, 377
Functions of the Linear Predictor, 379
Using Results to Conduct Inference, 380

Appendix 6. Theory on Parametric Models for Hazard Functions, 385

Poisson Regression, 385
Weibull Hazards, 387
Alternatives to Log-Linear Hazard Models, 390

Appendix 7. Theory on Proportional Hazards Regression, 393

Appendix 8. Theory on Analysis of Matched Studies, 397

Conditional Likelihood for Case-Control Study, 397
Matched Pairs for Case-Control Studies, 398
N-to-One Matching in Case-Control Studies, 399
Conditional Likelihood for Cohort Studies, 399

Index, 403