

Thomas J. Santner Diane E. Duffy

The Statistical Analysis of Discrete Data

With 30 Illustrations



Springer-Verlag
New York Berlin Heidelberg
London Paris Tokyo Hong Kong

Contents

Preface	xi
1 Introduction	1
1.1. Classes of Statistical Problems	1
1.2. Examples	3
1.3. Review of Discrete Distributions	14
2 Univariate Discrete Responses	21
2.1. Binomial Responses	21
A. Point Estimation of p , ω , and λ	22
The Estimator $\hat{p} = Y/n$	22
Bayes Estimators of p	24
Gamma Minimax and Hierarchical Bayes Estimators of p	28
Empirical Bayes and Pseudo Bayes Estimators of p	29
Estimating ω and λ	32
B. Interval Estimation	33
Small Sample Intervals	33
Large Sample Intervals	41
2.2. Multinomial Responses	43
A. Point Estimation of $\mathbf{p}$	44
Maximum Likelihood Estimation	44
Loss Functions	45
Bayes and Related Estimators	50
Smoothing Estimators	57
Comparison of Estimators	59
B. Hypothesis Tests about $\mathbf{p}$	64
Tests of Simple Null versus the Global Alternative	64
Tests of Simple Null versus Restricted Alternatives	67
Tests of Composite Null versus the Global Alternative	72
C. Interval Estimation	75
D. Selection and Ranking	79
2.3. Poisson Responses	80
A. Point Estimation of λ	81

Maximum Likelihood Estimation	81
Loss Functions	82
Alternative Estimators Under Squared Error Loss	84
Alternative Estimators Under Other Loss Functions	89
B. Hypothesis Tests	93
Testing Fit to the Poisson Model	93
Tests of Homogeneity	94
C. Interval Estimation	98
D. Selection and Ranking	99
Problems	100
 3 Loglinear Models	 113
3.1. Introduction	113
3.2. Maximum Likelihood Estimation for Loglinear Models Under Poisson Sampling	119
3.3. Maximum Likelihood Estimation for Loglinear Models Under (Product) Multinomial Sampling	125
3.4. Other Approaches	129
A. Alternatives to Maximum Likelihood Estimation for Loglinear Models	129
B. Alternatives to Loglinear Models	133
Problems	136
 4 Cross-Classified Data	 142
4.1. Introduction	142
4.2. Two-Dimensional Tables	146
Models and Interpretation	146
Assessing Fit	147
4.3. Hierarchical, Graphical, and Direct Loglinear Models for Arbitrary Dimensional Tables	153
Models and Interpretation for Three-Dimensional Tables	153
Interaction Graphs and Graphical Models	157
Direct Models	161
4.4. Numerical Evaluation of Maximum Likelihood Estimators and Residual Analysis	163
A. Calculation of Maximum Likelihood Estimators	163
Newton–Raphson Algorithm	163
Iterative Proportional Fitting Algorithm	164
B. Residual Analysis of Contingency Tables	168
4.5. Model Selection and a Case Study	170

General Notions	170
Tests for Order	173
Tests for Interactions	174
Starting Models and Stepping Rules	176
Case Study	178
4.6. Collapsing Tables	188
Simpson's Paradox	188
Collapsibility	189
Problems	195
 5 Univariate Discrete Data with Covariates	 204
5.1. Introduction	204
5.2. Two by T Tables	211
A. Introduction	211
B. Hypothesis Tests	213
The Two-Sample Problem	213
The T -Sample Problem	218
C. Interval Estimation	219
The Two-Sample Problem	219
The T -Sample Problem	225
D. Selection and Ranking	228
5.3. Logistic Regression: Basic Techniques	230
A. Introduction	230
B. Likelihood-Based Inference	232
Weighted Least Squares Estimation	232
Maximum Likelihood Estimation	233
Tests of Fit to the Logistic Model	239
Case Studies	240
5.4. Logistic Regression: Recent Developments	246
A. Alternative Estimation Methods	246
B. Graphical Assessment of Fit to the Logistic Model	252
Residuals	252
Leverages	253
Partial Residual Plots	254
Cook's Distance	255
5.5. Two by Two by S Tables: Matched Analyses	261
A. Introduction	261
B. Tests for Homogeneity of Odds Ratios	262
C. Tests of Zero Partial Association	264
D. Estimation of a Common Odds Ratio	266
The Empiric Logit (Woolf) Estimator	267
The (Unconditional) Maximum Likelihood Estimator	267
The Mantel-Haenszel Estimator	267
The Conditional Maximum Likelihood Estimator	268

Bayesian Estimators	268
Problems	269
Appendix 1. Some Results from Linear Algebra	287
Appendix 2. Maximization of Concave Functions	289
Appendix 3. Proof of Proposition 3.3.1 (ii) and (iii)	291
Appendix 4. Elements of Large Sample Theory	296
Problems	308
References	310
List of Notation	340
Index to Data Sets	345
Author Index	353
Subject Index	359