

Categorical Data Analysis

Second Edition

ALAN AGRESTI

University of Florida
Gainesville, Florida

Technische Universität Darmstadt FACHBEREICH INFORMATIK B I B L I O T H E K	
Inventar-Nr.:	<u>M 10-00024</u>
Sachgebiete:	<u>Mathematik</u>
Standort:	<u>M.5 / Agre</u>

Technische Universität Darmstadt
Fachbereich Informatik
Intelligente Systeme
Hochschulstraße 10
D-64289 Darmstadt

 **WILEY-
INTERSCIENCE**

A JOHN WILEY & SONS, INC., PUBLICATION

~~Inv.-Nr. 341~~

Contents

Preface	xiii
1. Introduction: Distributions and Inference for Categorical Data	1
1.1 Categorical Response Data, 1	
1.2 Distributions for Categorical Data, 5	
1.3 Statistical Inference for Categorical Data, 9	
1.4 Statistical Inference for Binomial Parameters, 14	
1.5 Statistical Inference for Multinomial Parameters, 21	
Notes, 26	
Problems, 28	
2. Describing Contingency Tables	36
2.1 Probability Structure for Contingency Tables, 36	
2.2 Comparing Two Proportions, 43	
2.3 Partial Association in Stratified 2×2 Tables, 47	
2.4 Extensions for $I \times J$ Tables, 54	
Notes, 59	
Problems, 60	
3. Inference for Contingency Tables	70
3.1 Confidence Intervals for Association Parameters, 70	
3.2 Testing Independence in Two-Way Contingency Tables, 78	
3.3 Following-Up Chi-Squared Tests, 80	
3.4 Two-Way Tables with Ordered Classifications, 86	
3.5 Small-Sample Tests of Independence, 91	

3.6	Small-Sample Confidence Intervals for 2×2 Tables,*	98
3.7	Extensions for Multiway Tables and Nontabulated Responses,	101
	Notes,	102
	Problems,	104
4.	Introduction to Generalized Linear Models	115
4.1	Generalized Linear Model,	116
4.2	Generalized Linear Models for Binary Data,	120
4.3	Generalized Linear Models for Counts,	125
4.4	Moments and Likelihood for Generalized Linear Models,*	132
4.5	Inference for Generalized Linear Models,	139
4.6	Fitting Generalized Linear Models,	143
4.7	Quasi-likelihood and Generalized Linear Models,*	149
4.8	Generalized Additive Models,*	153
	Notes,	155
	Problems,	156
5.	Logistic Regression	165
5.1	Interpreting Parameters in Logistic Regression,	166
5.2	Inference for Logistic Regression,	172
5.3	Logit Models with Categorical Predictors,	177
5.4	Multiple Logistic Regression,	182
5.5	Fitting Logistic Regression Models,	192
	Notes,	196
	Problems,	197
6.	Building and Applying Logistic Regression Models	211
6.1	Strategies in Model Selection,	211
6.2	Logistic Regression Diagnostics,	219
6.3	Inference About Conditional Associations in $2 \times 2 \times K$ Tables,	230
6.4	Using Models to Improve Inferential Power,	236
6.5	Sample Size and Power Considerations,*	240
6.6	Probit and Complementary Log-Log Models,*	245

*Sections marked with an asterisk are less important for an overview.

6.7	Conditional Logistic Regression and Exact Distributions,*	250
	Notes,	257
	Problems,	259
7.	Logit Models for Multinomial Responses	267
7.1	Nominal Responses: Baseline-Category Logit Models,	267
7.2	Ordinal Responses: Cumulative Logit Models,	274
7.3	Ordinal Responses: Cumulative Link Models,	282
7.4	Alternative Models for Ordinal Responses,*	286
7.5	Testing Conditional Independence in $I \times J \times K$ Tables,*	293
7.6	Discrete-Choice Multinomial Logit Models,*	298
	Notes,	302
	Problems,	302
8.	Loglinear Models for Contingency Tables	314
8.1	Loglinear Models for Two-Way Tables,	314
8.2	Loglinear Models for Independence and Interaction in Three-Way Tables,	318
8.3	Inference for Loglinear Models,	324
8.4	Loglinear Models for Higher Dimensions,	326
8.5	The Loglinear-Logit Model Connection,	330
8.6	Loglinear Model Fitting: Likelihood Equations and Asymptotic Distributions,*	333
8.7	Loglinear Model Fitting: Iterative Methods and their Application,*	342
	Notes,	346
	Problems,	347
9.	Building and Extending Loglinear/Logit Models	357
9.1	Association Graphs and Collapsibility,	357
9.2	Model Selection and Comparison,	360
9.3	Diagnostics for Checking Models,	366
9.4	Modeling Ordinal Associations,	367
9.5	Association Models,*	373
9.6	Association Models, Correlation Models, and Correspondence Analysis,*	379

9.7	Poisson Regression for Rates, 385	
9.8	Empty Cells and Sparseness in Modeling Contingency Tables, 391	
	Notes, 398	
	Problems, 400	
10.	Models for Matched Pairs	409
10.1	Comparing Dependent Proportions, 410	
10.2	Conditional Logistic Regression for Binary Matched Pairs, 414	
10.3	Marginal Models for Square Contingency Tables, 420	
10.4	Symmetry, Quasi-symmetry, and Quasi-independence, 423	
10.5	Measuring Agreement Between Observers, 431	
10.6	Bradley–Terry Model for Paired Preferences, 436	
10.7	Marginal Models and Quasi-symmetry Models for Matched Sets,* 439	
	Notes, 442	
	Problems, 444	
11.	Analyzing Repeated Categorical Response Data	455
11.1	Comparing Marginal Distributions: Multiple Responses, 456	
11.2	Marginal Modeling: Maximum Likelihood Approach, 459	
11.3	Marginal Modeling: Generalized Estimating Equations Approach, 466	
11.4	Quasi-likelihood and Its GEE Multivariate Extension: Details,* 470	
11.5	Markov Chains: Transitional Modeling, 476	
	Notes, 481	
	Problems, 482	
12.	Random Effects: Generalized Linear Mixed Models for Categorical Responses	491
12.1	Random Effects Modeling of Clustered Categorical Data, 492	
12.2	Binary Responses: Logistic-Normal Model, 496	
12.3	Examples of Random Effects Models for Binary Data, 502	
12.4	Random Effects Models for Multinomial Data, 513	

12.5	Multivariate Random Effects Models for Binary Data,	
	516	
12.6	GLMM Fitting, Inference, and Prediction,	520
	Notes,	526
	Problems,	527
13.	Other Mixture Models for Categorical Data*	538
13.1	Latent Class Models,	538
13.2	Nonparametric Random Effects Models,	545
13.3	Beta-Binomial Models,	553
13.4	Negative Binomial Regression,	559
13.5	Poisson Regression with Random Effects,	563
	Notes,	565
	Problems,	566
14.	Asymptotic Theory for Parametric Models	576
14.1	Delta Method,	577
14.2	Asymptotic Distributions of Estimators of Model Parameters and Cell Probabilities,	582
14.3	Asymptotic Distributions of Residuals and Goodness- of-Fit Statistics,	587
14.4	Asymptotic Distributions for Logit/Loglinear Models,	592
	Notes,	594
	Problems,	595
15.	Alternative Estimation Theory for Parametric Models	600
15.1	Weighted Least Squares for Categorical Data,	600
15.2	Bayesian Inference for Categorical Data,	604
15.3	Other Methods of Estimation,	611
	Notes,	615
	Problems,	616
16.	Historical Tour of Categorical Data Analysis*	619
16.1	Pearson–Yule Association Controversy,	619
16.2	R. A. Fisher’s Contributions,	622

16.3	Logistic Regression, 624	
16.4	Multiway Contingency Tables and Loglinear Models, 625	
16.5	Recent (and Future?) Developments, 629	
Appendix A.	Using Computer Software to Analyze Categorical Data	632
A.1	Software for Categorical Data Analysis, 632	
A.2	Examples of SAS Code by Chapter, 634	
Appendix B.	Chi-Squared Distribution Values	654
References		655
Examples Index		689
Author Index		693
Subject Index		701