

Contents

1	The Role of Categorical Data Analysis	1
1.1	Levels of Measurement	2
1.1.1	Categorical or Nominal Level of Measurement	2
1.1.2	Ordinal Level of Measurement	4
1.1.3	Interval Scale	5
1.1.4	Ratio Scale	5
1.1.5	Changing the Level of Measurement	6
1.2	Categorical or Continuous Data	7
1.3	Interaction in Statistical Analysis	10
1.3.1	Joint Effects in a Regression-Type Problem Under Joint Normality	10
1.3.2	Joint Effects in a Regression-Type Problem with Categorical Variables	12
1.4	Structure and Stochastics in Statistics	15
1.5	Things to Do	16
2	Sampling Distributions	19
2.1	The Binomial Distribution	19
2.1.1	Sampling Without Replacement	20
2.1.2	Sampling with Replacement	21
2.1.3	Properties of the Binomial Distribution	22
2.2	The Multinomial Distribution	23
2.2.1	Stratified Sampling: The Product Multinomial Distribution	26
2.3	Contingency Tables	27
2.4	Conditional and Marginal Distributions	29
2.5	The Poisson Distribution	32
2.6	Sampling with Unequal Selection Probabilities	38
2.7	Things to Do	39

3	Normal Approximations	41
3.1	Convergence in Distribution	41
3.2	Normal Approximation to the Binomial	44
3.3	Normal Approximation to the Multinomial	46
3.3.1	Normal Approximation to the Product Multinomial	49
3.3.2	Interaction Does Not Disappear Asymptotically	50
3.4	The δ -Method	50
3.5	Things to Do	55
4	Simple Estimation for Categorical Data	57
4.1	Introduction to Maximum Likelihood Estimation	57
4.1.1	Maximum Likelihood Estimation for the Multinomial Distribution	60
4.1.2	Maximum Likelihood Estimation for the Poisson Distribution	69
4.1.3	Maximum Likelihood Estimation in Parametric Models	71
4.1.4	Unbiased Estimation with Unequal Selection Probabilities	72
4.2	Standard Errors of Quantities Reported in Surveys	74
4.2.1	Standard Errors of More Complex Quantities	76
4.2.2	Standard Errors Under Stratified Sampling	78
4.3	Things to Do	81
5	Basic Testing for Categorical Data	83
5.1	Tests of $p = p_0$	83
5.2	Basic Properties of the Pearson and the Likelihood Ratio Statistics	87
5.3	Test of $p \leq p_0$	94
5.4	Test of Independence in Two-Way Tables	96
5.4.1	The Concept of Independence	97
5.4.2	Maximum Likelihood Estimation Under Independence	100
5.4.3	Tests of the Hypothesis of Independence	104
5.5	Things to Do	107
6	Association	109
6.1	The Odds Ratio	109
6.1.1	Maximum Likelihood Estimation of the Odds Ratio	111
6.1.2	Variation Independence of the Odds Ratio and the Marginal Distributions	114
6.1.3	Association Models for Two-Way Tables	121
6.2	Conditional and Higher-Order Odds Ratios	124
6.3	Independence and Conditional Independence	129
6.3.1	Maximum Likelihood Estimation Under Conditional Independence	133
6.4	Things to Do	135

7	Latent Classes and Exponential Families	137
7.1	The Latent Class Approach	137
7.1.1	Convergence of the EM Algorithm	141
7.2	Exponential Families of Probability Distributions	144
7.3	Things to Do	155
8	Effects and Associations	157
8.1	Association and Causation	158
8.2	Different Concepts of Association	160
8.3	Experiments and Observational Studies	162
8.3.1	Designed Experiments	163
8.3.2	Observational Studies	170
8.4	Some Theories of Causal Analysis Based on Observational Data	172
8.5	Things to Do	181
9	Simpson's Paradox	183
9.1	Simpson's Paradox	183
9.2	Consistent Treatment Selection	189
9.3	General Aspects of Measuring Effects and Associations	194
9.4	Things to Do	196
10	Log-Linear Models: Definition	197
10.1	Parameterizations of Multidimensional Discrete Distributions	198
10.2	Generalizations of Independence	208
10.3	Log-Linear Models and Parameters	213
10.4	Things to Do	222
11	Log-Linear Models: Interpretation	225
11.1	The Regression Problem for Categorical Variables	226
11.2	Interpretation in Regression Type and Non-regression Type Problems	230
11.3	Decomposable and Graphical Log-Linear Models	239
11.4	Canonical Representation of Log-Linear Models	248
11.5	Things to Do	252
12	Log-Linear Models: Estimation	255
12.1	Maximum Likelihood Estimation	255
12.2	Iterative Proportional Fitting	259
12.3	Things to Do	266
13	What's Next?	267
	References	277
	Index	281