

James K. Lindsey

Applying Generalized Linear Models

With 35 Illustrations



Springer

James K. Lindsey
Department of Biostatistics
Limburgs Universitair Centrum
3590 Diepenbeek
Belgium

Editorial Board

George Casella
Biometrics Unit
Cornell University
Ithaca, NY 14853
USA

Stephen Fienberg
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213
USA

Ingram Olkin
Department of Statistics
Stanford University
Stanford, CA 94305
USA

Library of Congress Cataloging-in-Publication Data

Lindsey, James K.

Applying generalized linear models / J.K. Lindsey

p. cm. — (Springer texts in statistics)

Includes bibliographical references (p. —) and index.

ISBN 0-387-98218-3 (hardcover : alk. paper)

1. Linear models (Statistics) I. Title. II. Series

QA279.L594 1997

97-6926

519.5'3—dc21

© 1997 Springer-Verlag New York, Inc.

All rights reserved. This work may not be translated or copied in whole or in part without the written permission of the publisher (Springer-Verlag New York, Inc., 175 Fifth Avenue, New York, NY 10010, USA), except for brief excerpts in connection with reviews or scholarly analysis. Use in connection with any form of information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed is forbidden.

The use of general descriptive names, trade names, trademarks, etc., in this publication, even if the former are not especially identified, is not to be taken as a sign that such names, as understood by the Trade Marks and Merchandise Marks Act, may accordingly be used freely by anyone.

Contents

Preface	v
1 Generalized Linear Modelling	1
1.1 Statistical Modelling	1
1.1.1 A Motivating Example	1
1.1.2 History	4
1.1.3 Data Generating Mechanisms and Models	6
1.1.4 Distributions	6
1.1.5 Regression Models	8
1.2 Exponential Dispersion Models	9
1.2.1 Exponential Family	10
1.2.2 Exponential Dispersion Family	11
1.2.3 Mean and Variance	11
1.3 Linear Structure	13
1.3.1 Possible Models	14
1.3.2 Notation for Model Formulae	15
1.3.3 Aliasing	16
1.4 Three Components of a GLM	18
1.4.1 Response Distribution or “Error Structure”	18
1.4.2 Linear Predictor	18
1.4.3 Link Function	18
1.5 Possible Models	20
1.5.1 Standard Models	20
1.5.2 Extensions	21

1.6	Inference	23
1.7	Exercises	25
2	Discrete Data	27
2.1	Log Linear Models	27
2.1.1	Simple Models	28
2.1.2	Poisson Representation	30
2.2	Models of Change	31
2.2.1	Mover–Stayer Model	32
2.2.2	Symmetry	33
2.2.3	Diagonal Symmetry	35
2.2.4	Long-term Dependence	36
2.2.5	Explanatory Variables	36
2.3	Overdispersion	37
2.3.1	Heterogeneity Factor	38
2.3.2	Random Effects	38
2.3.3	Rasch Model	39
2.4	Exercises	44
3	Fitting and Comparing Probability Distributions	49
3.1	Fitting Distributions	49
3.1.1	Poisson Regression Models	49
3.1.2	Exponential Family	52
3.2	Setting Up the Model	54
3.2.1	Likelihood Function for Grouped Data	54
3.2.2	Comparing Models	55
3.3	Special Cases	57
3.3.1	Truncated Distributions	57
3.3.2	Overdispersion	58
3.3.3	Mixture Distributions	60
3.3.4	Multivariate Distributions	63
3.4	Exercises	64
4	Growth Curves	69
4.1	Exponential Growth Curves	70
4.1.1	Continuous Response	70
4.1.2	Count Data	71
4.2	Logistic Growth Curve	72
4.3	Gomperz Growth Curve	74
4.4	More Complex Models	76
4.5	Exercises	82
5	Time Series	87
5.1	Poisson Processes	88
5.1.1	Point Processes	88

5.1.2	Homogeneous Processes	88
5.1.3	Nonhomogeneous Processes	88
5.1.4	Birth Processes	90
5.2	Markov Processes	91
5.2.1	Autoregression	93
5.2.2	Other Distributions	96
5.2.3	Markov Chains	101
5.3	Repeated Measurements	102
5.4	Exercises	103
6	Survival Data	109
6.1	General Concepts	109
6.1.1	Skewed Distributions	109
6.1.2	Censoring	109
6.1.3	Probability Functions	111
6.2	“Nonparametric” Estimation	111
6.3	Parametric Models	113
6.3.1	Proportional Hazards Models	113
6.3.2	Poisson Representation	113
6.3.3	Exponential Distribution	114
6.3.4	Weibull Distribution	115
6.4	“Semiparametric” Models	116
6.4.1	Piecewise Exponential Distribution	116
6.4.2	Cox Model	116
6.5	Exercises	117
7	Event Histories	121
7.1	Event Histories and Survival Distributions	122
7.2	Counting Processes	123
7.3	Modelling Event Histories	123
7.3.1	Censoring	124
7.3.2	Time Dependence	124
7.4	Generalizations	127
7.4.1	Geometric Process	128
7.4.2	Gamma Process	132
7.5	Exercises	136
8	Spatial Data	141
8.1	Spatial Interaction	141
8.1.1	Directional Dependence	141
8.1.2	Clustering	145
8.1.3	One Cluster Centre	147
8.1.4	Association	147
8.2	Spatial Patterns	149
8.2.1	Response Contours	149

8.2.2	Distribution About a Point	152
8.3	Exercises	154
9	Normal Models	159
9.1	Linear Regression	160
9.2	Analysis of Variance	161
9.3	Nonlinear Regression	164
9.3.1	Empirical Models	164
9.3.2	Theoretical Models	165
9.4	Exercises	167
10	Dynamic Models	173
10.1	Dynamic Generalized Linear Models	173
10.1.1	Components of the Model	173
10.1.2	Special Cases	174
10.1.3	Filtering and Prediction	174
10.2	Normal Models	175
10.2.1	Linear Models	176
10.2.2	Nonlinear Curves	181
10.3	Count Data	186
10.4	Positive Response Data	189
10.5	Continuous Time Nonlinear Models	191
Appendices		
A	Inference	197
A.1	Direct Likelihood Inference	197
A.1.1	Likelihood Function	197
A.1.2	Maximum Likelihood Estimate	199
A.1.3	Parameter Precision	202
A.1.4	Model Selection	205
A.1.5	Goodness of Fit	210
A.2	Frequentist Decision-making	212
A.2.1	Distribution of the Deviance Statistic	212
A.2.2	Analysis of Deviance	214
A.2.3	Estimation of the Scale Parameter	215
A.3	Bayesian Decision-making	215
A.3.1	Bayes' Formula	216
A.3.2	Conjugate Distributions	216
B	Diagnostics	221
B.1	Model Checking	221
B.2	Residuals	222
B.2.1	Hat Matrix	222
B.2.2	Kinds of Residuals	223

B.2.3	Residual Plots	225
B.3	Isolated Departures	226
B.3.1	Outliers	227
B.3.2	Influence and Leverage	227
B.4	Systematic Departures	228
References		231
Index		243