

GENERALIZED LINEAR MODELS FOR INSURANCE DATA

PIET DE JONG

Department of Actuarial Studies, Macquarie University, Sydney

GILLIAN Z. HELLER

Department of Statistics, Macquarie University, Sydney

CAMBRIDGE
UNIVERSITY PRESS

Contents

<i>Preface</i>	<i>page i x</i>
1 Insurance data	1
1.1 Introduction	2
1.2 Types of variables	3
1.3 Data transformations	4
1.4 Data exploration	6
1.5 Grouping and runoff triangles	10
1.6 Assessing distributions	12
1.7 Data issues and biases	13
1.8 Data sets used	14
1.9 Outline of rest of book	19
2 Response distributions	20
2.1 Discrete and continuous random variables	20
2.2 Bernoulli	21
2.3 Binomial	22
2.4 Poisson	23
2.5 Negative binomial	24
2.6 Normal	26
2.7 Chi-square and gamma	27
2.8 Inverse.Gaussian	29
2.9 Overdispersion	30
Exercises	33
3 Exponential family responses and estimation	35
3.1 Exponential family	35
3.2 The variance function	36
3.3 Proof of the mean and variance expressions	37
3.4 Standard distributions in the exponential family form	37
3.5 Fitting probability functions to data	39
Exercises	41

4	Linear modeling	42
4.1	History and terminology of linear modeling	42
4.2	What does "linear" in linear model mean?	43
4.3	Simple linear modeling	43
4.4	Multiple linear modeling	44
4.5	The classical linear model	46
4.6	Least squares properties under the classical linear model	47
4.7	Weighted least squares	47
4.8	Grouped and ungrouped data	48
4.9	Transformations to normality and linearity	49
4.10	Categorical explanatory variables	51
4.11	Polynomial regression	53
4.12	Banding continuous explanatory variables	54
4.13	Interaction	55
4.14	Collinearity	55
4.15	Hypothesis testing	56
4.16	Checks using the residuals	58
4.17	Checking explanatory variable specifications	60
4.18	Outliers	61
4.19	Model selection	62
5	Generalized linear models	64
5.1	The generalized linear model	64
5.2	Steps in generalized linear modeling	65
5.3	Links and canonical links	66
5.4	Offsets	66
5.5	Maximum likelihood estimation	67
5.6	Confidence intervals and prediction	70
5.7	Assessing fits and the deviance	71
5.8	Testing the significance of explanatory variables	74
5.9	Residuals	77
5.10	Further diagnostic tools	79
5.11	Model selection	80
	Exercises	80
6	Models for count data	81
6.1	Poisson regression	81
6.2	Poisson overdispersion and negative binomial regression	89
6.3	Quasi-likelihood	94
6.4	Counts and frequencies	96
	Exercises	96
7	Categorical responses	97
7.1	Binary responses	97
7.2	Logistic regression	98

7.3	Application of logistic regression to vehicle insurance	99
7.4	Correcting for exposure	102
7.5	Grouped binary data	105
7.6	Goodness of fit for logistic regression	107
7.7	Categorical responses with more than two categories	110
7.8	Ordinal responses	111
7.9	Nominal responses	116
	Exercises	119
8	Continuous responses	120
8.1	Gamma regression	120
8.2	Inverse Gaussian regression	125
8.3	Tweedie regression	127
	Exercises	128
9	Correlated data	129
9.1	Random effects	131
9.2	Specification of within-cluster correlation	136
9.3	Generalized estimating equations	137
	Exercise	140
10	Extensions to the generalized linear model	141
10.1	Generalized additive models	141
10.2	Double generalized linear models	143
10.3	Generalized additive models for location, scale and shape	143
10.4	Zero-adjusted inverse Gaussian regression	145
10.5	A mean and dispersion model for total claim size	148
	Exercises	149
Appendix 1	Computer code and output	150
A1.1	Poisson regression	150
A1.2	Negative binomial regression	156
A1.3	Quasi-likelihood regression	159
A1.4	Logistic regression	160
A1.5	Ordinal regression	169
A1.6	Nominal regression	175
A1.7	Gamma regression	178
A1.8	Inverse Gaussian regression	181
A1.9	Logistic regression GLMM	183
A1.10	Logistic regression GEE	185
A1.11	Logistic regression GAM	187
A1.12	GAMLSS	189
A1.13	Zero-adjusted inverse Gaussian regression	190
<i>Bibliography</i>		192
<i>Index</i>		195