

Contents

1	Introducing GLIM3	1
1.1	Statistical packages and modelling	1
1.2	Getting started in GLIM3	2
1.3	Reading data into GLIM	6
1.4	Displaying data using \$PRINT and \$LOOK	12
1.5	Graphical facilities in GLIM	17
1.6	Vectors, scalars and the workspace	23
1.7	Introduction to calculation and data modification in GLIM	28
1.8	Function evaluation and suffixed vectors	34
1.9	Files and file handling	46
1.10	Macros and text handling	53
1.11	Sorting and tabulation	61
2	Statistical modelling and statistical inference	67
2.1	Statistical models	67
2.2	Types of variables	69
2.3	An example	70
2.4	Definition of a statistical model	76
2.5	The likelihood function and maximum likelihood estimation	77
2.6	The accuracy of parameter estimation	81
2.7	The likelihood ratio test and the choice among regressions	85
2.8	Strategies for model simplification	91
2.9	Stratified, weighted and clustered samples	94
2.10	Model criticism	97
2.11	The modelling of binary response data	105
2.12	The Bernoulli distribution for binary data	107
2.13	Maximum likelihood for the Bernoulli and binomial distributions	108
2.14	Maximum likelihood fitting of the regression model	109
2.15	Likelihood-based confidence intervals	112
3	Normal regression and analysis of variance	119
3.1	The normal distribution and the Box-Cox transformation family	119

3.2	Modelling and background information	122
3.3	Link functions and transformations	128
3.4	Regression models for prediction	130
3.5	Model choice and mean square prediction error	132
3.6	Prediction model selection through cross-validation	134
3.7	The reduction of complex regression models	136
3.8	Sensitivity of the Box–Cox transformation to outliers	141
3.9	The use of regression models for calibration	143
3.10	Measurement error in the explanatory variables	145
3.11	Factorial designs	147
3.12	Cross-classifications	152
3.13	Missing data	164
3.14	Approximate methods for missing data	166
4	Binomial response data	167
4.1	Binary responses	167
4.2	Transformation and link functions	168
4.3	Model criticism for binary data	171
4.4	Contingency table construction from binary data	179
4.5	The prediction of binary outcomes	185
4.6	Profile and conditional likelihoods in 2×2 tables	195
4.7	Multidimensional contingency tables with a binary response	200
4.8	The analysis of complex contingency tables with a binary response	205
4.9	Overdispersion, omitted variables and model mis-specification	213
5	Multinomial and Poisson response data	217
5.1	The Poisson distribution	217
5.2	Overdispersion	223
5.3	Multicategory responses	225
5.4	Multinomial logit model	228
5.5	Relation between the multinomial and Poisson distributions	231
5.6	Fitting the multinomial logit model	235
5.7	Ordered response categories	236
5.8	An example	242
5.9	Structured multinomial responses	250
6	Survival data	256
6.1	Introduction	256
6.2	The exponential distribution	256
6.3	Fitting the exponential distribution	258

6.4	Model criticism	261
6.5	Comparison with the normal family	265
6.6	Censoring	268
6.7	Likelihood function for censored observations	269
6.8	Probability plotting with censored data: the Kaplan–Meier estimator	271
6.9	The gamma distribution	277
6.10	The Weibull distribution	280
6.11	Maximum likelihood fitting of the Weibull distribution	281
6.12	The extreme value distribution	283
6.13	The reversed extreme value distribution	285
6.14	Survivor function plotting for the Weibull and extreme value distributions	286
6.15	The Cox proportional hazards model and the piecewise exponential distribution	287
6.16	Maximum likelihood fitting of the piecewise exponential distribution	289
6.17	Examples	292
6.18	The logistic and log–logistic distributions	295
6.19	The normal and lognormal distributions	297
6.20	The choice between different survival distributions	300
6.21	Competing risks	304
6.22	Time-dependent explanatory variables	309
6.23	Discrete time models	312
References		316
Appendix 1. Discussion		321
Appendix 2. GLIM directives		332
Appendix 3. System defined structures in GLIM		339
Appendix 4. Datasets and macros		343
Author Index		367
Subject Index		369