

Contents

1	Introduction	1
1.1	Introductory examples and types of outcome	3
1.1.1	Introductory examples	3
1.1.2	Types of outcome	7
1.2	Covariates	9
1.2.1	Categorical covariates	9
1.2.2	Quantitative covariates	11
1.3	Link functions	13
1.4	Building a regression model	22
1.4.1	The linear predictor and the link function	24
1.4.2	Regression models and their interpretation	28
1.5	Further examples	30
1.6	The scope of this book and how to read it	37
2	Statistical models	43
2.1	Random variables and probability	44
2.1.1	The Bernoulli distribution	47
2.1.2	The Binomial distribution	48
2.1.3	The Poisson distribution	50
2.1.4	The Normal distribution	51
2.1.5	Other common distributions	55
2.1.6	Conditional probability	56
2.2	Descriptive statistics	57
2.2.1	Binary outcome	57
2.2.2	Quantitative outcome	59
2.2.3	Survival time outcome	63
2.3	Statistical inference	64
2.3.1	Estimation	65
2.3.2	Model checking	73
2.3.3	Hypothesis testing	79
2.3.4	The likelihood function	84

2.4	Exercises	92
3	One categorical covariate	95
3.1	Binary covariate	96
3.1.1	Quantitative outcome: t -tests	97
3.1.2	Binary outcome: (2×2) -tables and the chi-square test ..	110
3.1.3	Survival time outcome: the 2-sample logrank test	123
3.2	Categorical covariate with more than two levels	137
3.2.1	Quantitative outcome: One-way analysis of variance ...	142
3.2.2	Binary outcome: The $2 \times (k + 1)$ -table	157
3.2.3	Survival time outcome: The $(k + 1)$ -sample logrank test	161
3.3	Exercises	166
4	One quantitative covariate	173
4.1	Linear effect	175
4.1.1	Quantitative outcome: Simple linear regression	176
4.1.2	Binary outcome: Simple logistic regression	194
4.1.3	Survival time outcome: Simple Cox regression	201
4.2	Nonlinear effect	210
4.2.1	Dividing the covariate range into intervals	214
4.2.2	Polynomials	221
4.2.3	Other nonlinear models with a linear predictor	223
4.3	Exercises	228
5	Multiple regression, the linear predictor	231
5.1	Two covariates: Models without interaction	234
5.1.1	Two categorical covariates	234
5.1.2	One categorical and one quantitative covariate	245
5.1.3	Two quantitative covariates	254
5.2	Two covariates: Models with interaction	263
5.2.1	Two categorical covariates	264
5.2.2	One categorical and one quantitative covariate	273
5.2.3	Two quantitative covariates	278
5.2.4	Saving degrees-of-freedom	285
5.3	Several covariates	287
5.3.1	Models without higher-order interactions	287
5.3.2	Models with higher-order interactions	289
5.4	Matched studies	294
5.5	Exercises	300
6	Model building: From purpose to conclusion	303
6.1	General principles for model selection	304
6.1.1	Identification of covariates	305
6.1.2	Model diagrams	308
6.1.3	Initial model building	311

6.1.4	Strategy of analysis	314
6.1.5	Model checks and diagnostics	317
6.1.6	Collinearity	318
6.1.7	Interactions	319
6.2	Examples	320
6.2.1	The vitamin D example	321
6.2.2	The surgery example	331
6.2.3	The PBC-3 trial	342
6.3	Sample size determination	354
6.4	Exercises	365
7	Alternative outcome types and link functions	367
7.1	Multinomial outcome	367
7.1.1	Ordinal outcome	368
7.1.2	Nominal outcome	383
7.2	Count outcome	387
7.3	Quantitative outcome	394
7.4	Binary outcome	403
7.4.1	Alternatives to the logit link	404
7.4.2	Case-control studies	409
7.5	Survival time outcome	416
7.5.1	Multiplicative hazard models	416
7.5.2	Additive hazard models	422
7.5.3	Accelerated failure time models	427
7.6	Exercises	429
8	Further topics	431
8.1	Multivariate outcome	431
8.1.1	Random effects models	433
8.1.2	Marginal models	438
8.1.3	Longitudinal and life history data	440
8.2	Errors in covariates	447
8.2.1	Regression dilution	448
8.2.2	Correction for measurement error in covariates	452
A	Appendix: Notation	457
B	Appendix: Use of logarithms	463
C	Appendix: Some recommendations	473
D	Programming in R, SAS and STATA	477
	References	483
	Index	487