

Contents

	Preface	<i>page</i> xiii
1	Introduction	1
	1.1 Why should I do multivariable analysis?	1
	1.2 What are confounders and how does multivariable analysis help me to deal with them?	6
	1.3 What are suppressers and how does multivariable analysis help me to deal with them?	9
	1.4 What are interactions and how does multivariable analysis help me to deal with them?	11
2	Common uses of multivariable models	14
	2.1 What are the most common uses of multivariable models in clinical research?	14
	2.2 How do I choose what type of multivariable analysis to use?	23
3	Outcome variables in multivariable analysis	24
	3.1 How does the nature of my outcome variable influence my choice of which type of multivariable analysis to do?	24
	3.2 What type of multivariable analysis should I use with an interval outcome?	24
	3.3 What are the different types of analysis of variance and when are they used?	25
	3.4 What should I do if my outcome variable is ordinal or nominal?	27
	3.5 What type of multivariable analysis should I use with a dichotomous outcome?	28

- 3.6 What type of multivariable analysis should I use with a time-to-outcome variable? 28
- 3.7 What type of multivariable analysis should I use with a rare outcome or a count? 32

4**Type of independent variables in multivariable analysis 35**

- 4.1 What type of independent variables can I use with multivariable analysis? 35
- 4.2 What should I do with my ordinal and nominal independent variables? 35

5**Assumptions of multiple linear regression, multiple logistic regression, and proportional hazards analysis 38**

- 5.1 What are the assumptions of multiple linear regression, multiple logistic regression, and proportional hazards analysis? 38
- 5.2 What is being modeled in multiple linear regression, multiple logistic regression, and proportional hazards analysis? 38
- 5.3 What is the relationship of multiple independent variables to outcome in multiple linear regression, multiple logistic regression, and proportional hazards analysis? 42
- 5.4 What is the relationship of an interval-independent variable to the outcome in multiple linear regression, multiple logistic regression, and proportional hazards analysis? 43
- 5.5 What if my interval-independent variable does not have a linear relationship with my outcome? 46
- 5.6 Assuming that my interval-independent variable fits a linear assumption, is there any reason to group it into interval categories or create multiple dichotomous variables? 51
- 5.7 What are the assumptions about the distribution of the outcome and the variance? 52
- 5.8 What should I do if I find significant violations of the assumptions of normal distribution and equal variance in my multiple linear regression analysis? 55
- 5.9 What are the assumptions of censoring? 56
- 5.10 How likely is it that the censoring assumption is valid in my study? 59

- 5.11 How can I test the validity of the censoring assumption for my data? 64

6 Relationship of independent variables to one another 68

- 6.1 Does it matter if my independent variables are related to each other? 68
- 6.2 How do I assess whether my variables are multi collinear? 69
- 6.3 What should I do with multicollinear variables? 71

7 Setting up a multivariable analysis 73

- 7.1 What independent variables should I include in my multivariable model? 73
- 7.2 How do I decide what confounders to include in my model? 73
- 7.3 What independent variables should I exclude from my multivariable model? 74
- 7.4 How many subjects do I need to do multivariable analysis? 77
- 7.5 What if I have too many independent variables given my sample size? 81
- 7.6 What should I do about missing data on my independent variables? 87
- 7.7 What should I do about missing data on my outcome variable? 94

8 Performing the analysis 96

- 8.1 What numbers should I assign for dichotomous or ordinal variables in my analysis? 96
- 8.2 Does it matter what I choose as my reference category for multiple dichotomous (“dummied”) variables? 97
- 8.3 How do I enter interaction terms into my analysis? 98
- 8.4 How do I enter time into my proportional hazards or other survival analysis? 101
- 8.5 What about subjects who experience their outcome on their start date? 106
- 8.6 What about subjects who have a survival time shorter than physiologically possible? 107
- 8.7 What are variable selection techniques? 109
- 8.8 What value should I specify for tolerance in my logistic regression or proportional hazards model? 114

- 8.9 How many iterations (attempts to solve) should I specify for my logistic regression or proportional hazards model? 114
- 8.10 What value should I specify for the convergence criteria for my logistic regression or proportional hazards model? 115
- 8.11 My model won't converge. What should I do? 115

9

Interpreting the analysis 117

- 9.1 What information will the printout from my analysis provide? 117
- 9.2 How do I assess how well my model accounts for the outcome? 117
- 9.3 What do the coefficients tell me about the relationship between each variable and the outcome? 124
- 9.4 How do I get odds ratios and relative hazards from the multivariable analysis? What do they mean? 126
- 9.5 How do I interpret the odds ratio and relative hazard when the independent variable is interval? 129
- 9.6 How do I compute the confidence intervals for the odds ratios and relative hazards? 130
- 9.7 What are standardized coefficients and should I use them? 131
- 9.8 How do I test the statistical significance of my coefficients? 131
- 9.9 How do I interpret the results of interaction terms? 134
- 9.10 Do I have to adjust my multivariable regression coefficients for multiple comparisons? 134

10

Checking the assumptions of the analysis 137

- 10.1 How do I know if my data fit the assumptions of my multivariable model? 137
- 10.2 How do I assess the linearity, normal distribution, and equal variance assumptions of multiple linear regression? 138
- 10.3 How do I assess the linearity assumption of multiple logistic regression and proportional hazards analysis? 139
- 10.4 What are outliers and how do I detect them in my multiple linear regression model? 139
- 10.5 How do I detect outliers in my multiple logistic regression model? 141
- 10.6 What about analysis of residuals with proportional hazards analysis? 142
- 10.7 What should I do when I detect outliers? 142

10.8	What is the additive assumption and how do I assess whether my multiple independent variables fit this assumption?	143
10.9	What does the additive assumption mean for interval-independent variables?	145
10.10	What is the proportionality assumption?	146
10.11	How do I test the proportionality assumption?	148
10.12	What if the proportionality assumption does not hold for my data?	150
11	Propensity scores	153
11.1	What are propensity scores? Why are they used?	153
12	Correlated observations	158
12.1	How do I analyze correlated observations?	158
12.2	How do I calculate the needed sample size for studies with correlated observations?	177
13	Validation of models	179
13.1	How can I validate my models?	179
14	Special topics	184
14.1	What if the independent variable changes value during the course of the study?	184
14.2	What are the advantages and disadvantages of time-dependent covariates?	185
14.3	What are classification and regression trees (CART) and should I use them?	187
14.4	How can I get best use of my biostatistician?	190
14.5	How do I choose which software package to use?	190
15	Publishing your study	192
15.1	How much information about how I constructed my multivariable models should I put in the Methods section?	192
15.2	Do I need to cite a statistical reference for my choice of multivariable model?	194

15.3 Which parts of my multivariable analysis should I report in the Results section? 194

Summary: Steps for constructing a multivariable model 197

Index 199