

Contents

<i>Acknowledgements</i>	ix
PART 1	
The basics	1
1 What is regression analysis?	3
1.1 <i>Introduction: Sizes and prices of houses</i>	3
1.2 <i>What this book is about and why it has been written</i>	5
1.3 <i>Some terminology and a word about statistical software</i>	5
1.4 <i>Research questions in the social sciences and economics: The link to regression analysis</i>	7
1.5 <i>The book's intended audience, pedagogical approach and what lies ahead in Part 1</i>	7
Notes	8
Appendix: <i>A short primer on variables and their measurement levels and how this relates to regression analysis</i>	8
2 Linear regression with a single independent variable	11
2.1 <i>Prices of houses once more</i>	11
2.2 <i>Prices of houses: The regression coefficient (b_1) and the constant (b_0)</i>	19
2.3 <i>Prices of houses: An independent variable with two values (a dummy variable)</i>	20
2.4 <i>From data on houses to data on soccer players: On performance and pay in football</i>	22
2.5 <i>Data from questionnaires: Female students' exercise hours</i>	25
2.6 <i>(Optional) calculation of the regression coefficient (b_1) and the constant (b_0) by Ordinary Least Squares (OLS)</i>	27
2.7 <i>A quick word about the error term (e)</i>	30
2.8 <i>Summary, a few words on data types and further reading</i>	31
Notes	32

Appendix A: A short introduction to Stata's point-and-click routine 34
Appendix B: A short introduction to the point-and-click routine in SPSS 38

- 3 Linear regression with several independent variables: Multiple regression 43
- 3.1 *Introduction 1: Prices of houses again 43*
 - 3.2 *Introduction 2: Social science research questions and the need to control for a third variable 46*
 - 3.3 *Two motivations for multiple regression 50*
 - 3.4 *A special case: Sets of independent dummy variables 50*
 - 3.5 *A short summary and the regression model's explanatory power (R^2) 54*
 - 3.6 *Summary and further reading 61*
 - Notes 61*
 - Exercises with data and solutions 62*
 - Informal glossary to Part 1 64*

PART 2

The foundations 67

- 4 Samples and populations, random variability and testing of statistical significance 69
- 4.1 *Introduction 69*
 - 4.2 *Populations, samples and representative samples: The magic of random sampling 70*
 - 4.3 *The normal distribution and repeated sampling 71*
 - 4.4 *Confidence intervals for means and regression coefficients 72*
 - 4.5 *Testing for the significance of a regression coefficient: Putting the confidence interval on its head 74*
 - 4.6 *Other critical issues regarding significance testing 77*
 - 4.7 *Summary and further reading 79*
 - Notes 79*
- 5 The assumptions of regression analysis 81
- 5.1 *Introduction 81*
 - 5.2 *The assumptions you should think about before doing your regression and, quite often, before you choose your data (A) 81*
 - 5.3 *The testable assumptions not requiring a random sample from a population-framework (B) 82*
 - 5.4 *The additional testable assumptions for a random sample from a population-framework (C) 87*
 - 5.5 *Summing up the regression assumptions, some other issues and further reading 92*
 - Notes 93*

PART 3

The extensions

99

- 6 Beyond linear regression: Non-additivity, non-linearity and mediation 101
- 6.1 *Introduction* 101
 - 6.2 *Non-additivity (assumption B1): Interaction/moderator regression models* 101
 - 6.3 *Non-linearity 1 (assumption B2): Quadratic regression models* 105
 - 6.4 *Non-linearity 2 (assumption B2): Sets of dummy variables* 107
 - 6.5 *Non-linearity 3 (assumption B2): Use of logarithms* 108
 - 6.6 *Mediation models* 113
 - 6.7 *Summary and further reading* 116
- Notes 116
- 7 A categorical dependent variable: Logistic (logit) regression and related methods 120
- 7.1 *Introduction* 120
 - 7.2 *The linear probability model (LPM)* 121
 - 7.3 *The standard multiple logistic (logit) regression model* 123
 - 7.4 *Interpreting and presenting the results of the logistic (logit) regression model* 124
 - 7.5 *Interaction models, non-linearity and model fit* 127
 - 7.6 *Extension: When y has three or more unordered alternatives (multinomial logistic regression)* 130
 - 7.7 *Other related issues and further reading* 132
- Notes 133
- Exercises with data and solutions* 134
- 8 An ordered (ordinal) dependent variable: Logistic (logit) regression 136
- 8.1 *Introduction* 136
 - 8.2 *The multiple ordinal (logit) regression model* 137
 - 8.3 *Interpreting and presenting the results of the ordinal (logit) regression model* 139
 - 8.4 *The proportional odds assumption, non-linearity and interaction effects* 141
 - 8.5 *Alternative modeling strategies* 144
 - 8.6 *Using linear regression on ordered (ordinal) y-variables: Is it justified?* 146

8.7	<i>Final thoughts and further reading</i>	147
	<i>Notes</i>	148
	<i>Exercises with data and solutions</i>	148
9	The quest for a causal effect: Instrumental variable (IV) regression	150
9.1	<i>Introduction</i>	150
9.2	<i>A first take on the relationship between exercising (x_1) and health (y)</i>	151
9.3	<i>The assumptions of IV regression: Instrument relevance and instrument exogeneity</i>	153
9.4	<i>When x_1 and/or y are dummies</i>	156
9.5	<i>Final thoughts and further reading</i>	158
	<i>Notes</i>	159
	PART 4	
	Regression purposes, academic regression projects and the way ahead	161
10	Regression purposes in various academic settings and how to perform them	163
10.1	<i>Introduction</i>	163
10.2	<i>Four purposes of regression analysis: (1) prediction, (2) description, (3) making inferences beyond the data and (4) causal analysis</i>	163
10.3	<i>The structure of an academic regression-based study</i>	171
10.4	<i>Some practical advice</i>	173
	<i>Notes</i>	174
11	The way ahead: Related techniques	176
11.1	<i>Introduction</i>	176
11.2	<i>Related technique 1: Median and quantile regression</i>	176
11.3	<i>Related technique 2: Count data regression</i>	177
11.4	<i>Related technique 3: Panel data regression and time-series regression (longitudinal data)</i>	180
11.5	<i>Related technique 4: Multi-level regression models</i>	181
11.6	<i>Related technique 5: Matching</i>	182
11.7	<i>Other, related techniques and further reading</i>	182
	<i>Notes</i>	182
	<i>References</i>	184
	<i>Index</i>	188