

Contents

Lists of Figures, Tables and Abbreviations	viii
Preface	xii
1 Introduction	1
2 The Use of Statistical Methods in Social Research	7
3 Cases and Variables	21
4 The Logic of Sampling	34
5 Estimating and Measuring One Important Thing	51
6 Studying the Relationship between Two Variables	71
7 Linear Regression Models and Their Generalizations	91
8 Time Matters	113
9 Statistical Case-oriented Methods	132
10 Methods for Analysing Latent Variables	153
11 Causal Analysis	173
Bibliography	190
Index	195

Lists of Figures, Tables and Abbreviations

Figures

Figure 4.1	Distinguishing cluster sampling from stratified sampling	47
Figure 5.1	Distribution of respondents across religions in the UK, 2006	59
Figure 5.2	Religiousness in the UK, 2006	60
Figure 5.3	Statistic for sample as parameter for population	62
Figure 6.1	Relationship between r and scatter plot	73
Figure 6.2	Correlation in four sections of a two-dimensional space	74
Figure 6.3	Anscombe's Quartet: Same regression line, different datasets	85
Figure 6.4	Difference between males and females in terms of the number of hours chosen to work, UK, 2006	87
Figure 7.1	Ideal situation for a multiple linear regression model	95
Figure 7.2	Realistic situation for a multiple linear regression model	95
Figure 7.3	Effect of correlated error terms on classical linear regression model	106
Figure 7.4	Effect of ignoring grouping on statistical significance	106
Figure 8.1	Different types of cases in an event history study	124
Figure 9.1	An illustrated dendrogram on social trust	138
Figure 9.2	Mapping out the relative positions of three cases in one dimension	146
Figure 9.3	A biplot of access to the Internet and opinions on homosexuality	151
Figure 10.1	An illustration of variances of four observed variables	157
Figure 10.2	An illustration of scree plot	159
Figure 10.3	A confirmatory factor model with four observed variables	163
Figure 10.4	A two-factor measurement model	163
Figure 10.5	A random intercept and random slope latent growth curve model	170
Figure 11.1	Ideal situation for inferring causal effect	183
Figure 11.2	Observation available for treatment but not for control	184

Figure 11.3	Observation available for control but not for treatment	184
Figure 11.4	Association of the error term with assignment	187
Figure 11.5	Use of an instrumental variable	188

Tables

Table 3.1	Case-by-variable matrix	22
Table 3.2	Levels of measurement and their relations	29
Table 4.1	US population structure by age, gender and the number of races, 2000	39
Table 4.2	Scenarios of bias and precision in probability sampling	42
Table 5.1	Effect of k on D and IQV	61
Table 6.1	A generic 2×2 table	76
Table 6.2	British identity and ethnicity	79
Table 6.3	A generic 3×3 table of two ordinal variables	81
Table 6.4	Components of the F -test	87
Table 7.1	Regression output for the supervisor performance example	96
Table 7.2	Data matrix without grouping	104
Table 7.3	Data matrix with grouping	104
Table 7.4	Lower level data matrix	105
Table 7.5	Higher level data matrix	105
Table 7.6	Scenarios of random intercept and coefficients in multilevel modelling	107
Table 8.1	An example of the long form data matrix in a longitudinal study	118
Table 8.2	Wide form of event history data	125
Table 8.3	A protocol of the life table	126
Table 8.4	Long form of event history data matrix	126
Table 9.1	A classification matrix for two groups	142
Table 9.2	An illustration of similarity matrix with four objects	147
Table 9.3	Personal use of Internet/email/WWW and opinions on homosexuality	150
Table 10.1	A hypothetical correlation matrix	156
Table 10.2	Re-organization of Table 10.1	156
Table 10.3	An ideal factor loadings table	160
Table 10.4	Correlations of the indicators measured with a four-point scale	162

Table 10.5	Correlations of the indicators measured with an 11-point scale	162
Table 10.6	A generic covariance matrix of five variables	167
Table 10.7	A generic correlation matrix of five variables	167
Table 11.1	Scenarios of cause and outcome	175
Table 11.2	Four scenarios of an instrumental variable and assignment	189

Abbreviations

ANOVA	analysis of variance
BHPS	British Household Panel Study
CFA	confirmatory factor analysis
CFI	comparative fit index
CLA	cluster analysis
COA	correspondence analysis
CPI	consumer price index
DA	discriminant analysis
DE	design effect
<i>df</i>	degrees of freedom
EFA	exploratory factor analysis
ESS	European Social Survey
GSS	General Social Survey
HLM	hierarchical linear models (also the name of software)
ICC	intra-class correlation coefficient
IQR	inter-quartile range
IQV	index of qualitative variation
LATE	local average treatment effect
LGCM	latent growth curve models
MANCOVA	multiple analysis of covariance
MANOVA	multiple analysis of variance
MDS	multidimensional scaling
PCA	principal component analysis
PSID	Panel Study of Income Dynamics
PSU	primary sampling unit
QCA	qualitative comparative analysis
RMSEA	root mean square error of approximation
RMSR	root means square residual
s.d.	standard deviation
s.e.	standard error
SEM	structural equation models
SES	socio-economic status