

Updated for
Version 9

A Handbook of
Statistical
Analyses
Using **Stata**

Fourth Edition

Sophia Rabe-Hesketh
Brian S. Everitt



Chapman & Hall/CRC
Taylor & Francis Group

Contents

1	A Brief Introduction to Stata.....	1
1.1	Getting help and information	1
1.2	Running Stata	2
1.3	Conventions used in this book	9
1.4	Datasets in Stata	9
1.5	Stata commands	13
1.6	Data management	19
1.7	Estimation	22
1.8	Graphics	24
1.9	Stata as a calculator	30
1.10	Matrix calculations using Mata	32
1.11	Brief introduction to programming	34
1.12	Keeping Stata up to date	39
1.13	Exercises	40
2	Data Description and Simple Inference: Female Psychiatric Patients.....	43
2.1	Description of data	43
2.2	Group comparison and correlations	46
2.3	Analysis using Stata	47
2.4	Exercises	57
3	Multiple Regression: Determinants of Pollution in U.S. Cities	61
3.1	Description of data	61
3.2	The multiple regression model	63
3.3	Analysis using Stata	64
3.4	Exercises	82

4	Analysis of Variance I: Treating Hypertension	85
4.1	Description of data	85
4.2	Analysis of variance model	85
4.3	Analysis using Stata	87
4.4	Exercises	96
5	Analysis of Variance II: Effectiveness of Slimming Clinics	101
5.1	Description of data	101
5.2	Analysis of variance model	102
5.3	Analysis using Stata	104
5.4	Exercises	108
6	Logistic Regression: Treatment of Lung Cancer and Diagnosis of Heart Attacks	111
6.1	Description of data	111
6.2	The logistic regression model	112
6.3	Analysis using Stata	116
6.4	Exercises	129
7	Generalized Linear Models: Australian School Children	133
7.1	Description of data	133
7.2	Generalized linear models	134
7.3	Analysis using Stata	139
7.4	Exercises	153
8	Summary Measure Analysis of Longitudinal Data: Treatment of Post-Natal Depression.....	157
8.1	Description of data	157
8.2	The analysis of longitudinal data	159
8.3	Analysis using Stata	159
8.4	Exercises	170
9	Random Effects Models: Thought Disorder and Schizophrenia	173
9.1	Description of data	173
9.2	Random effects models	173
9.3	Analysis using Stata	178
9.4	Thought disorder data	190
9.5	Exercises	199
10	Generalized Estimating Equations: Epileptic Seizures and Chemotherapy	201
10.1	Description of data	201

10.2	Generalized estimating equations	203
10.3	Analysis using Stata	205
10.4	Exercises	218
11	Some Epidemiology	221
11.1	Description of data	221
11.2	Introduction to epidemiology	222
11.3	Analysis using Stata	228
11.4	Exercises	236
12	Survival Analysis: Retention of Heroin Addicts in Methadone Maintenance Treatment	239
12.1	Description of data	239
12.2	Survival analysis	242
12.3	Analysis using Stata	245
12.4	Exercises	258
13	Maximum Likelihood Estimation: Age of Onset of Schizophrenia	263
13.1	Description of data	263
13.2	Finite mixture distributions	263
13.3	Analysis using Stata	264
13.4	Exercises	277
14	Principal Components Analysis: Hearing Measurement Using an Audiometer	281
14.1	Description of data	281
14.2	Principal component analysis	283
14.3	Analysis using Stata	284
14.4	Exercises	291
15	Cluster Analysis: Tibetan Skulls and Determinants of Pollution in U.S. Cities	295
15.1	Description of data	295
15.2	Cluster analysis	297
15.3	Analysis using Stata	298
15.4	Exercises	311
	Appendix: Answers to Selected Exercises.....	315
	References.....	327
	Index.....	335