

# Table of Contents

Preface ix

## 1 Observational Studies and Experiments

- 1.1 Introduction 1
- 1.2 The HIP trial 4
- 1.3 Snow on cholera 6
- 1.4 Yule on the causes of poverty 9
  - Exercise set A 13
- 1.5 End notes 14

## 2 The Regression Line

- 2.1 Introduction 18
- 2.2 The regression line 18
- 2.3 Hooke's law 22
  - Exercise set A 23
- 2.4 Complexities 23
- 2.5 Simple vs multiple regression 25
  - Exercise set B 26
- 2.6 End notes 28

## 3 Matrix Algebra

- 3.1 Introduction 29
  - Exercise set A 30
- 3.2 Determinants and inverses 31
  - Exercise set B 33
- 3.3 Random vectors 35
  - Exercise set C 35
- 3.4 Positive definite matrices 36
  - Exercise set D 37
- 3.5 The normal distribution 38
  - Exercise set E 39
- 3.6 If you want a book on matrix algebra 40

## 4 Multiple Regression

- 4.1 Introduction 41
  - Exercise set A 44
- 4.2 Standard errors 45
  - Things we don't need 48
  - Exercise set B 49
- 4.3 Explained variance in multiple regression 50
  - Association or causation? 52
- 4.4 Generalized least squares 52
- 4.5 Examples on GLS 55
  - Exercise set C 56
- 4.6 What happens to OLS if the assumptions break down? 57
- 4.7 Normal theory 57
  - Statistical significance 60
  - Exercise set D 60
- 4.8 The  $F$ -test 61
  - "The"  $F$ -test in applied work 63
  - Exercise set E 63
- 4.9 Data snooping 64
  - Exercise set F 65
- 4.10 Discussion questions 65
- 4.11 End notes 72

## 5 Path Models

- 5.1 Stratification 75
  - Exercise set A 80
- 5.2 Hooke's law revisited 81
  - Exercise set B 82
- 5.3 Political repression during the McCarthy era 82
  - Exercise set C 84
- 5.4 Inferring causation by regression 85
  - Exercise set D 87
- 5.5 Response schedules for path diagrams 88
  - Selection vs intervention 95
  - Structural equations and stable parameters 95
  - Ambiguity in notation 96
  - Exercise set E 96
- 5.6 Dummy variables 97
  - Types of variables 98

## TABLE OF CONTENTS

5.7 Discussion questions	99
5.8 End notes	106
<b>6 Maximum Likelihood</b>	
6.1 Introduction	109
Exercise set A	113
6.2 Probit models	114
Why not regression?	117
The latent-variable formulation	117
Exercise set B	118
Identification vs estimation	119
What if the $U_i$ are $N(\mu, \sigma^2)$ ?	120
Exercise set C	120
6.3 Logit models	121
Exercise set D	122
6.4 The effect of Catholic schools	123
More on table 3	126
Latent variables	126
Response schedules	127
The second equation	128
Mechanics: bivariate probit	130
Why a model rather than a cross-tab?	132
Interactions	132
More on the second equation	133
Exercise set E	133
6.5 Discussion questions	135
6.6 End notes	142
<b>7 The Bootstrap</b>	
7.1 Introduction	148
Exercise set A	159
7.2 Bootstrapping a model for energy demand	160
Exercise set B	166
7.3 End notes	167
<b>8 Simultaneous Equations</b>	
8.1 Introduction	169
Exercise set A	174
8.2 Instrumental variables	174
Exercise set B	177

8.3	Estimating the butter model	177
	Exercise set C	178
8.4	What are the two stages?	178
	Invariance assumptions	179
8.5	A social-science example: education and fertility	180
	More on Rindfuss et al	184
8.6	Covariates	184
8.7	Linear probability models	185
	The assumptions	186
	The questions	188
	Exercise set D	188
8.8	More on IVLS	189
	Some technical issues	189
	Exercise set E	191
	Simulations to illustrate IVLS	191
	Further reading on econometric technique	192
8.9	Issues in statistical modeling	192
8.10	Critical literature	195
	Response schedules	199
8.11	Evaluating the models in chapters 6–8	200
8.12	Summing up	200

References 201

Answers to Exercises 216

The Computer Labs 267

Appendix: Sample MATLAB Code 283

Reprints

Gibson on McCarthy 288

Evans and Schwab on Catholic Schools 316

Rindfuss et al on Education and Fertility 350

Schneider et al on Social Capital 375

Index 404