

CONTENTS

1 Bivariate Linear Regression

1

- 1.1. Terminology, 1
- 1.2. Fitting a Least-Squares Line, 2
- 1.3. The Least-Squares Line as a Causal Model, 5
- 1.4. The Bivariate Linear Regression Model as a Statistical Model, 7
 - 1.4.1. Simplifying Assumptions, 8
- 1.5. Statistical Inference: Generalizing from Sample to Underlying Population, 10
 - 1.5.1. Hypothesis Testing, 12
 - 1.5.2. Confidence Intervals, 15
 - 1.5.3. t Values and Z Values, 16
 - 1.5.4. p Value, 17
 - 1.5.5. Importance of a Good Spread of Values of the Predictor Variable, 18
 - 1.5.6. Beware of Outliers!, 20
 - 1.5.7. Beware of Selection on the Response Variable!, 21
 - 1.5.8. Presentation of Results, 22
- 1.6. Goodness of Fit, 23
 - 1.6.1. Standard Error of the Estimate, s , 23
 - 1.6.2. Coefficient of Determination, r^2 , and Correlation Coefficient, r , 24
- 1.7. Further Reading, 28

2 Multiple Regression

- 2.1. The Problem of Bias in Bivariate Linear Regression, 30
- 2.2. Multiple Regression with Two Predictor Variables, 31
- 2.3. Multiple Regression with Three or More Predictor Variables, 34
- 2.4. Dummy Variables to Represent Categorical Variables, 34
 - 2.4.1. Categorical Variables with Two Categories, 34
 - 2.4.2. Categorical Variables with More Than Two Categories, 36
- 2.5. Multicollinearity, 38
- 2.6. Interaction, 40
 - 2.6.1. Model Specification, 40
 - 2.6.2. More Complicated Interactions, 43
 - 2.6.3. Correlation without Interaction, 45
 - 2.6.4. Interaction without Correlation, 45
- 2.7. Nonlinearities, 46
 - 2.7.1. Quadratic Specification, 46
 - 2.7.2. Dummy Variable Specification, 49
- 2.8. Goodness of Fit, 51
 - 2.8.1. Standard Error of the Estimate, s , 51
 - 2.8.2. Coefficient of Determination, R^2 , and Multiple Correlation Coefficient, R , 52
 - 2.8.3. Corrected R^2 and Corrected R , 53
 - 2.8.4. Partial Correlation Coefficient, 54
- 2.9. Statistical Inference, 55
 - 2.9.1. Hypothesis Testing, Confidence Intervals, and p Values for a Single Regression Coefficient, 55
 - 2.9.2. Testing the Difference Between Two Regression Coefficients, β_i and β_j , 56
 - 2.9.3. Testing Effects When There Is Interaction, 57
 - 2.9.4. Testing Effects When There Is a Nonlinearity, 58
 - 2.9.5. The ANOVA Table, 59
 - 2.9.6. The Omnibus F Test of the Hypothesis $\beta_1 = \beta_2 = \dots = \beta_k = 0$, 60
 - 2.9.7. Test of the Hypothesis That Some of the β_j Are Zero, 61
- 2.10. Stepwise Regression, 62
- 2.11. Illustrative Examples, 63
 - 2.11.1. Example 1, 63
 - 2.11.2. Example 2, 66
- 2.12. Further Reading, 68

3 Multiple Classification Analysis

- 3.1. The Basic MCA Table, 70
 - 3.1.1. Unadjusted Values, 70
 - 3.1.2. Adjusted Values, 73
 - 3.1.3. Unadjusted and Adjusted, R , 75
 - 3.1.4. A Numerical Example, 76
- 3.2. The MCA Table in Deviation Form, 78
 - 3.2.1. First Approach to Table Set-up, 78
 - 3.2.2. Second Approach to Table Set-up, 78
 - 3.2.3. A Numerical Example, 81
- 3.3. MCA with Interactions, 82
 - 3.3.1. Table Set-up, 82
 - 3.3.2. A Numerical Example, 84
- 3.4. MCA with Additional Quantitative Control Variables, 85
 - 3.4.1. Table Set-up, 86
 - 3.4.2. A Numerical Example, 87
- 3.5. Expressing Results from Ordinary Multiple Regression in an MCA Format (all Variables Quantitative), 88
 - 3.5.1. Table Set-up, 88
 - 3.5.2. A Numerical Example, 89
 - 3.5.3. Why Present Ordinary Regression Results in an MCA Format?, 89
- 3.6. Presenting MCA Results Graphically, 90
- 3.7. Further Reading, 92

4 Path Analysis

93

- 4.1. Path Diagrams and Path Coefficients, 93
- 4.2. Path Models with More Than One Exogenous Variable, 100
- 4.3. Path Models with Control Variables, 105
- 4.4. Saturated and Unsaturated Path Models, 107
- 4.5. Path Analysis with Standardized Variables, 109
 - 4.5.1. Standardized Variables and Standardized Path Coefficients, 109
 - 4.5.2. Path Models in Standardized Form, 110
 - 4.5.3. Standardized Versus Unstandardized Coefficients, 113
- 4.6. Path Models with Interactions and Nonlinearities, 114
- 4.7. Further Reading, 118

5 Logit Regression

119

- 5.1. The Linear Probability Model, 119

- 5.2. The Logit Regression Model, 121
 - 5.2.1. The Logistic Function, 122
 - 5.2.2. The Multivariate Logistic Function, 125
 - 5.2.3. The Odds and the Logit of P , 126
 - 5.2.4. Logit Regression Coefficients as Measures of Effect on Logit P , 128
 - 5.2.5. Odds Ratios as Measures of Effect on the Odds, 129
 - 5.2.6. The Effect on the Odds When the Predictor Variable Is Categorical with More Than Two Categories, 132
 - 5.2.7. The Effect of the Predictor Variables on the Risk P Itself, 134
 - 5.2.8. Interactions, 135
 - 5.2.9. Nonlinearities, 136
- 5.3. Statistical Inference, 137
 - 5.3.1. Tests of Coefficients, 137
 - 5.3.2. Testing the Difference Between Two Models, 138
- 5.4. Goodness of Fit, 140
- 5.5. MCA Adapted to Logit Regression, 142
 - 5.5.1. An Illustrative Example, 142
 - 5.5.2. Table Set-up and Numerical Results, 144
- 5.6. Fitting the Logit Regression Model, 147
 - 5.6.1. The Maximum Likelihood Method, 147
 - 5.6.2. Derivation of the Likelihood Function, 147
- 5.7. Some Limitations of the Logit Regression Model, 150
- 5.8. Further Reading, 150

6 Multinomial Logit Regression

151

- 6.1. From Logit to Multinomial Logit, 151
 - 6.1.1. The Basic Form of the Multinomial Logit Model, 151
 - 6.1.2. Interpretation of Coefficients, 153
 - 6.1.3. Presentation of Results in a Multiple Classification Analysis (MCA) Format, 153
 - 6.1.4. Statistical Inference, 157
 - 6.1.5. Goodness of Fit, 157
 - 6.1.6. Changing the Reference Category of the Response Variable, 157
- 6.2. Multinomial Logit Models with Interactions and Nonlinearities, 159

6.3.	A More General Formulation of the Multinomial Logit Model,	160
6.4.	Reconceptualizing Contraceptive Method Choice as a Two-Step Process,	163
6.5.	Further Reading,	165
7	Survival Models, Part 1: Life Tables	167
7.1.	Actuarial Life Table,	168
7.1.1.	Statistical Inference,	173
7.2.	Product-Limit Life Table,	175
7.2.1.	Testing the Difference Between Two Survival Curves,	177
7.3.	The Life Table in Continuous Form,	178
7.4.	Further Reading,	180
8	Survival Models, Part 2: Proportional Hazard Models	181
8.1.	Basic Form of the Proportional Hazard Model,	182
8.1.1.	A Simple Example,	183
8.1.2.	Addition of a Second Predictor Variable,	184
8.1.3.	Redefinition of the Baseline Hazard,	185
8.1.4.	Relative Risks as Measures of Effect on the Hazard,	187
8.1.5.	Hazard Regression Coefficients as Measures of Effect on the Log Hazard,	188
8.1.6.	The Effect on the Hazard When the Predictor Variable Is Categorical with More Than Two Categories,	190
8.1.7.	Interactions,	191
8.1.8.	Nonlinearities,	192
8.1.9.	Changing the Reference Category of a Categorical Predictor Variable,	194
8.2.	Calculation of Life Tables from the Proportional Hazard Model,	194
8.2.1.	The Hazard Model in Survivorship Form,	194
8.2.2.	The Problem of Unobserved Heterogeneity,	195
8.2.3.	Test for Proportionality,	196
8.3.	Statistical Inference and Goodness of Fit,	198
8.4.	A Numerical Example,	198
8.5.	Multiple Classification Analysis (MCA) Adapted to Proportional Hazard Regression,	205
8.6.	Further Reading,	205

9 Survival Models, Part 3: Hazard Models with Time Dependence

207

- 9.1. Time-Dependent Predictor Variables, 207
 - 9.1.1. Coefficients and Effects, 208
 - 9.1.2. Choosing a Baseline Hazard Function, 209
 - 9.1.3. MCA Adapted to Hazard Regression with Time-Dependent Predictor Variables, 210
- 9.2. Time-Dependent Coefficients, 214
 - 9.2.1. Coefficients and Effects, 214
 - 9.2.2. Survival Functions, 216
 - 9.2.3. MCA Adapted to Hazard Regression with Time-Dependent Coefficients, 218
- 9.3. Further Reading, 220

Appendix A Sample Computer Programs

221

- A.1. Description of the FIJIDATA File, 221
- A.2. SAS Mainframe Programs, 223
 - Program 1. SAS Program for Multiple Regression (Program Used for Table 2.2), 223
 - Program 2. SAS Program for Partial Correlation Coefficient (Program Used for Table 3.4, Adjusted R Column), 224
 - Program 3. SAS Program for Logit Regression (Program Used for Table 5.1), 225
 - Program 4. SAS Program for Multinomial Logit Regression (Program Used for Table 6.4), 226
- A.3. Preparing Data for Survival Analysis, 227
 - Program 5. SAS Program to Create Input Data File for BMDP and LIMDEP Programs for Life Table and Hazard Model Analysis of Progression from Fifth Birth to Sixth Birth, 228
- A.4. BMDP Mainframe Programs, 230
 - Program 6. BMDP1L Program for Actuarial Life Table: Progression from Fifth Birth to Sixth Birth

- for Fijians and Indians (Program Used for Table 7.1 and the Actuarial Life Tables in Table 7.2), 230
- Program 7. BMDP1L Program for Product-Limit Life Table: Progression from Fifth Birth to Sixth Birth for Fijians and Indians (Programs Used for Product-Limit Life Tables in Table 7.2), 232
- Program 8. BMDP2L Program for Proportional Hazard Model: Progression from Fifth to Sixth Birth for Indians (Program Used for Table 8.1), 233
- Program 9. BMDP2L Program for Hazard Model with Time-Dependent Covariate (Program Used for Table 9.2), 235
- Program 10. BMDP2L Program for Hazard Model with Time-Dependent Covariate and Time-Dependent Effect (Program Used for Table 9.3), 237
- A.5. LIMDEP Programs for IBM-Compatible Personal Computers, 238
- Program 11. LIMDEP Program for Multiple Regression (Program used for Table 2.2), 238
- Program 12. LIMDEP Program for Partial Correlation Coefficient (Program Used for Table 3.4, Adjusted R column), 239
- Program 13. LIMDEP Program for Logit Regression (Program Used for Table 5.1), 240
- Program 14. LIMDEP Program for Multinomial Logit Regression (Program Used for Table 6.4), 241

- Program 15. LIMDEP Program
for Life Table (Actuarial and
Product-Limit): Progression
from Fifth Birth
to Sixth Birth, Indians (Program
Used for Table 7.2), 242
- Problem 16. LIMDEP Program
for Proportional Hazard Model:
Progression from Fifth
Birth to Sixth Birth,
Indians (Programs Used
for Table 8.1), 243
- Problem 17. LIMDEP Program
for Hazard Model with Time-
Dependent Covariate: Progression
from Fifth Birth to Sixth
Birth, Indians (Program
Used for Table 9.2), 243
- Problem 18. LIMDEP Program for
Hazard Model with Time-Dependent
Covariate and Time-Dependent
Effect: Progression from Fifth Birth
to Sixth Birth, Indians (Program
Used for Table 9.3), 244

Appendix B	Statistical Reference Tables	245
References		251
Index		254