

# Contents

Figures and Tables	xi
Preface	xiii
<b>1. Introduction</b>	<b>1</b>
The Idea and Importance of Multivariate Models	1
More Detail on Multivariate Models	4
Classification of Data Used in Multivariate Models	5
Statistical Techniques and Their Data Requirements	6
Three Main Uses of Multiple Regression and Some Simplified Examples	9
Conclusions	12
Summary of Chapter 1	12
<b>2. The Bivariate or Simple Regression Model</b>	<b>13</b>
The Meaning of the Term Bivariate Regression	13
Deterministic Versus Stochastic Relationships	17
Populations Versus Samples	18
Estimators for the Bivariate Regression Model	19
The Coefficient of Determination or $R^2$	24
Conclusions	25
Summary of Chapter 2	26
<b>3. The Multiple Regression Model, Part I: Estimators, Statistical     Properties, and Significance Tests</b>	<b>27</b>

An Example of the Need for Multivariate Relationships	27
The Three Variable Regression Model	29
Statistical Characteristics of the Least Squares Estimators	30
Significance Tests and Confidence Intervals	33
Using the Computer to Obtain Regression Statistics	42
Conclusions	46
Summary of Chapter 3	46
<b>4. The Multiple Regression Model, Part II: Importance of Variables, Model Building, and Forecasting</b>	<b>47</b>
Measuring the Importance of Individual Variables	47
Techniques to Help with Model Building	51
Forecasting with a Regression Model	55
An Example of Model Building and Stepwise Regression Using the Computer	60
Conclusions	62
Summary of Chapter 4	62
<b>5. Dummy Variables and Nonlinear and Nonadditive Relationships</b>	<b>63</b>
Dummy Variables	63
Testing for the Significance of Dummy Variables	68
Nonlinear Models	71
Nonadditive Models	81
Examples of Computer Output for Dummy Variables and for Nonlinear Relationships	87
A Summary Example	90
Conclusions	92
Summary of Chapter 5	93
<b>6. Basic Assumptions and Common Problems in Regression Models</b>	<b>95</b>
The Assumption of Constant Variance or Homoskedasticity	96
The Assumption of Zero Covariance Between Errors or Zero Autocorrelation	102
The Assumption That the Error Term and the Independent Variables are Uncorrelated	110
Multicollinearity	114
An Example of Computer Output Showing Autocorrelation	118
Conclusions	122
Summary of Chapter 6	123
<b>7. Qualitative Dependent Variables</b>	<b>125</b>
The Linear Probability Model	126
The Logit and Probit Forms for Dichotomous Qualitative Choice Models	128

Estimating the Logit Model	133
Estimating the Probit Model	137
Conclusions	140
Summary of Chapter 7	140
<b>8. Some Advanced Topics: Pooled Time-Series and Cross-Section Analysis, Lagged Variables, Missing Data, Time-Series Analysis, and Multiequation Systems</b>	<b>143</b>
Using Pooled Cross-Section and Time-Series Data	144
Lagged Variables	148
Missing Data	149
Time-Series Analysis	151
Multiequation Models	156
Conclusions	166
Summary of Chapter 8	168
<b>9. An Overview</b>	<b>169</b>
A Management Use of Multiple Regression: Adjusted Productivity Measures	170
A Policy Use of Multiple Regression: Adjusted Hospital Mortality Rates	171
Many Uses, Not So Many Distinctions	173
Appendices	
<b>A. Simple Summation Operators and Expected Values</b>	<b>175</b>
Algebra of Summations	175
Expected Values	176
<b>B. The Pearson Correlation Coefficient</b>	<b>179</b>
<b>C. Review of Hypothesis Testing for the Sample Mean</b>	<b>183</b>
Forming the Null and Alternative Hypotheses	183
The Logic of Hypothesis Testing	184
The Test Statistic	184
Setting Significance Levels and Determining Cutoff Points	185
Confidence Intervals	186
The Alternative Hypothesis Again	186
<b>D. Statistical Tables</b>	<b>187</b>
Notes	191
Bibliographical Essay	203
Index	205

# Figures and Tables

## FIGURES

2.1	Scattergram for a bivariate regression	14
2.2	Regression line and errors	20
2.3	Explained, unexplained and total variation for one observation, $X(1)$	25
3.1	Biased and unbiased slope estimators	31
3.2	Best (minimum variance) and not best slope estimators	32
3.3	Cutoff points for a two-sided hypothesis test	36
3.4	Descriptive statistics, correlation matrix, and regression results	44
4.1	Confidence intervals for known population coefficients	57
4.2	Confidence intervals for unknown population coefficients	59
4.3	Stepwise regression: first variable (UUNEMPR) and last variable (SCPIUS) to be entered	61
5.1	Earnings equation with a sex dummy	67
5.2	Optimal city size example using a quadratic polynomial	80
5.3	Earnings equation with an interaction effect	82
5.4	CYTC regression results with dummy variable, DUMDEP	88
5.5	Descriptive statistics, correlation matrix, and regression results: log sales tax revenue (LTXSALE)	89
6.1	Homoskedastic distribution of errors	96
6.2	Heteroskedastic distribution of errors	97
6.3	Extreme heteroskedasticity: zero variances for two observations	98
6.4	Heteroskedastic residuals	100
6.5	Positive autocorrelation	103

6.6	CALLY equation before adjustment for autocorrelation	119
6.7	Plot of residuals and Durbin-Watson statistic: CALLY equation before adjustment for autocorrelation	120
7.1	The linear probability model	127
7.2	The probit versus logit cumulative distributions	131
8.1	Spending per pupil and dropout rates	163
8.2	Spending per pupil and dropout rates with exogenous variable	164
B.1	Scattergrams for several possible Pearson correlation coefficients	180

**TABLES**

1.1	Example of Type of Data for Chi-Square Analysis	7
1.2	Example of Type of Data for ANOVA	8
5.1	Regression Results for Preoperative Length of Stay (LOS)	72
5.2	Regression Results for Operating Expenditures per Participant in Municipal Retirement Systems	78
5.3	Regression Results for Research Unit Productivity	86
5.4	Regression Results for Earnings Equations: American and Japanese Male Workers	91
6.1	OLS Regression Results for Total City Income	108
6.2	Differenced Regression Results for Total City Income	109
6.3	Regression Results for Return on Assets for New York State Voluntary Hospitals	117
7.1	OLS Regression Results for Probability of Physician Visit	130
7.2	Logit Regression Results for Probability of Hospital Having a Mammography Unit	136
7.3	Probit Regression Results for Probability of Exceeding Mental Health Insurance Coverage Limit	139
8.1	Random Effects and Fixed Effects Regression Results for Black Children's Literacy Rates	146
8.2	Percentage Changes in Literacy Gap for Black Children	147
8.3	OLS Regression Results for per Pupil School Expenditures and Average Annual Earnings	161
8.4	OLS and 2SLS Regression Results for Number of Visits to Physician	167
D.1	Percentage Points of Student's <i>t</i> Distribution	187
D.2	<i>F</i> Distribution, 5 Percent Significance	188
D.3	Critical Points of the Durbin-Watson Test, 5 Percent Significance ( $d_1$ and $d_u$ )	190