

PREFACE

Causal Statistics is my name for a ^{young, but} rapidly growing field whose purpose it is to draw causal inferences from non-experimental data. The applicability of causal statistics is broad indeed; it may be utilized in varying degrees to obtain ^{causal} "knowledge" in the following disciplines: sociology, psychology, economics, political science, anthropology, history, ^{social systems and futures modelling} education, ^{and} biomedical science, ^{and} organizational behavior. In addition causal statistics will--by its very nature--facilitate and encourage the bridging of disciplinary boundaries and lead toward an eclectic approach to theory construction.

With this in mind, it seems odd that such a broadly useful inquiring system has been applied so infrequently. I believe that there are two reasons for this underutilization; one is philosophical and the other is pedagogical.

Most researchers took Philosophy 101 where they learned Hume said that causality cannot be proved, hence they go through their scientific lives avoiding the word "cause". But, since the concept of causation is crucial to theory construction, these researchers must--as Michael Scriven pointed out--replace the word with other more acceptable words which have the same meaning. Consider the following commonly used replacements: brings about, yields, leads to, results in. This is clearly erroneous behavior and this tangled situation will be straightened out in Chapter 4. Suffice it to say here that there is no philosophical reason to avoid ^{the concept of causation or} causal statistics.

The pedagogical problem with causal statistics is that, until now, there has been no single source of complete information about causal statistics. Information about causal inquiring systems was spread thinly throughout the scientific literature, the notation varied, the presentations were often mathematically very complex, and gaps in the fabric of knowledge abounded.

Due to these pedagogical difficulties, this valuable inquiring system has been available to only those who were (1) mathematically gifted and (2) willing to spend a huge investment of time to dig out the information available in the literature and fill in the gaps with their own original thinking.

It is the purpose of this book to make causal statistics readily available to anyone who has need of it. High levels of mathematical ability or sophistication are not required. Basic algebra and the careful reading of the text will suffice. The amount of statistics required is about one course in basic statistics, i.e. probability theory, probability distributions, confidence intervals, and hypothesis testing.

This book is intended for use with students at various levels, from early undergraduate to graduate. At the most elementary level the book could be used in a beginning statistics class along with a regular statistics text. Instead of teaching regression analysis in a standard way, one could use this book to teach regression analysis and causal statistics simultaneously. The book could also be used for a course in causal statistics, requiring only one standard statistics course as a prerequisite. It is also possible to teach a one-year application oriented causal statistics course from this book. The instructor could make use of outside readings and empirical studies could be carried out by the students. In addition research methods courses at either the undergraduate or graduate level-- would profit from the adoption of this book as one of the basic texts.

CAUSAL STATISTICS

Preface
Table of Contents

Part I Introduction to Causal Statistics

Chapter 1 Introduction and Summary

- 1.1 Preliminary Remarks
- 1.2 The Applicability of Causal Statistics
- 1.3 The Application of Causal Statistics
- 1.4 Summary and Discussion of Part I
- 1.5 Summary and Discussion of Part II
- 1.6 Summary and Discussion of Part III
- 1.7 Summary and Discussion of Part IV

Chapter 2 The Setting and Significance of Causal Statistics

- 2.1 Preliminary Remarks
- 2.2 The Research Process
- 2.3 The Setting of Causal Statistics
- 2.4 The Need for Causal Statistics
- 2.5 Fields in which Causal Statistics is Applicable

Chapter 3 What is Causal Statistics? (A Historical Approach)

- 3.1 Preliminary Remarks
- 3.2 History
- 3.3 Present
- 3.4 Future

Chapter 4 Philosophical Issues

- 4.1 Preliminary Remarks
- 4.2 The Definition of Cause
- 4.3 Analysis of the Causal Philosophy of David Hume
- 4.4 Assumptions of Causal Statistics

Chapter 5 Summary of Probability and Statistics

Part II Single Equation Models

Chapter 6 Two Variable Models

- 6.1 Preliminary Remarks
- 6.2 Curve Fitting
 - 6.2.1 Scatter Diagrams
 - 6.2.2 Possible Criteria for Curve Fitting
 - 6.2.3 The Least Squares Criteria
- 6.3 Statistical Inference for a Fixed Independent Variable
 - 6.3.1 The Fundamental Assumptions of Regression Analysis
 - 6.3.2 The Fundamental Assumptions of Causal Statistics

- 6.3.3 Causal Parameter Estimation
- 6.3.4 The Means, Variances, and Distributions of Causal
Parameter Estimators
- 6.3.5 Confidence Intervals for Causal Parameters
- 6.3.6 Hypothesis Tests for Causal Parameters
- 6.3.7 Confidence Intervals for Causal Predictions
- 6.4 Statistical Inference for Random Independent Variables
- 6.5 Application to a Bivariate Normal Population
- 6.6 Interpretation of Results

Chapter 7 Multivariable Models

- 7.1 Preliminary Remarks
- 7.2 Graphical Representation of the Three Dimensional Case
- 7.3 Assumptions
- 7.4 Causal Parameter Estimation
- 7.5 Confidence Intervals and Hypothesis Testing
- 7.6 Multicollinearity
- 7.7 Stepwise Parameter Estimation
- 7.8 Computer Analysis
- 7.9 Interpretation of Results

Chapter 8 Correlation

- 8.1 Preliminary Remarks
- 8.2 Simple Correlation
 - 8.2.1 The Population Correlation Coefficient
 - 8.2.2 The Sample Correlation Coefficient
 - 8.2.3 Interpretation of the Correlation Coefficient
- 8.3 Relationship between the Correlation Coefficient and the
Causal Parameter
- 8.4 Partial Correlation
- 8.5 Relationship between the Partial Correlation Coefficient and the
Partial Causal Parameter
- 8.6 Multiple Correlation

Chapter 9 Nonlinear Models

- 9.1 Preliminary Remarks
- 9.2 Nonlinearity in the Variables
- 9.3 Nonlinearity in the Causal Parameters
- 9.4 Differential Models
- 9.5 Intractable Nonlinear Models

Chapter 10 Relaxation of Various Assumptions

- 10.1 Preliminary Remarks
- 10.2 Heteroschedasticity
- 10.3 Specification Error
- 10.4 Data Snooping
- 10.5 Omitted Relevant Variables
- 10.6 Included Irrelevant Variables
- 10.7 Superfluous Variables
- 10.8 Serial correlation in the Error Term
- 10.9 Lagged Variables

- 10.10 Correlation between the Error Term and an Independent Variable
- 10.11 Measurement Error
- 10.12 Transcription Error
- 10.13 Causal Relationships between Independent Variables
- 10.14 Reciprocal Causation

Chapter 11 Special Topics and Techniques in Causal Inference

- 11.1 Preliminary Remarks
- 11.2 Operational Variables
- 11.3 Proxy Variables
- 11.4 Ordinal Variables
- 11.5 Dichotomous Variables
- 11.6 Qualitative Variables
- 11.7 Moderator Variables
- 11.8 Scanning
- 11.9 Dummy Variables Used as a Jackknife
- 11.10 Time as an Independent Variable
- 11.11 Inclusion of Associative Variable
- 11.12 Dummy Variables for Seasonal Variations
- 11.13 Clustered Residuals

Chapter 12 Performing a Causal Analysis Study

- 12.1 Preliminary Remarks
- 12.2 State the Problems
- 12.3 Construct a Hypothesized Model
- 12.4 State the Assumption Set
- 12.5 Identify the Equations
- 12.6 Operationalize Variables
- 12.7 Collect Data
- 12.8 Estimate Causal Parameters
- 12.9 Interpret Results
- 12.10 Write-up Study
- 12.11 Criticize and Redo Study

Part III Simultaneous Equation Models

Chapter 13 Limited Information Estimation Techniques

- 13.1 Preliminary Remarks
- 13.2 Ordinary Least Squares
- 13.3 Indirect Least Squares
- 13.4 Instrumental Variables
- 13.5 Two Stage Least Squares
- 13.6 Least-Variance Ratio

Chapter 14 Identification

- 14.1 Preliminary Remarks
- 14.2 Under-identification
- 14.3 Identification by Addition of Exogenous Variables to the Model
- 14.4 Identification Using A Priori Information
- 14.5 Necessary Conditions for Identification
- 14.6 Necessary and Sufficient Conditions for Identification
- 14.7 Over-identification

Chapter 15 Full Information Estimation Techniques

- 15.1 Preliminary Remarks
- 15.2 Three Stage Least Squares
- 15.3 Iterated Three Stage Least Squares
- 15.4 Full Information Maximum Likelihood

Chapter 16 Comparison of Estimation Techniques

- 16.1 Preliminary Remarks
- 16.2 Small Sample Distributions
- 16.3 Summary of Overall Comparisons