

Cambridge University Press

0521791537 - Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference - Bill Shipley

Frontmatter

[More information](#)

## **Cause and Correlation in Biology**

A User's Guide to Path Analysis, Structural Equations  
and Causal Inference

This book goes beyond the truism that 'correlation does not imply causation' and explores the logical and methodological relationships between correlation and causation. It presents a series of statistical methods that can test, and potentially discover, cause–effect relationships between variables in situations in which it is not possible to conduct randomised or experimentally controlled experiments. Many of these methods are quite new and most are generally unknown to biologists. In addition to describing how to conduct these statistical tests, the book also puts the methods into historical context and explains when they can and cannot justifiably be used to test or discover causal claims. Written in a conversational style that minimises technical jargon, the book is aimed at practising biologists and advanced students, and assumes only a very basic knowledge of introductory statistics.

BILL SHIPLEY teaches plant ecology and biometry in the Department of Biology at the Université de Sherbrooke, Quebec, Canada. His present ecological research concentrates on comparative ecophysiology and the ways in which plant attributes interact to produce ecological outcomes. He has also contributed significantly to research in topics including plant competition, species richness and plant community ecology. His statistical research is equally diverse, covering such areas as permutation and bootstrap methods, path analysis, dynamic game theory and non-parametric regression smoothers. This rare combination of practical experience in both experimental science and statistical research makes him well positioned to communicate statistical methods to practising biologists in a meaningful way.

Cambridge University Press

0521791537 - Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference - Bill Shipley

Frontmatter

[More information](#)

# Cause and Correlation in Biology

A User's Guide to Path Analysis, Structural Equations  
and Causal Inference

BILL SHIPLEY

Université de Sherbrooke,  
Sherbrooke (Qc) Canada



CAMBRIDGE  
UNIVERSITY PRESS

Cambridge University Press

0521791537 - Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference - Bill Shipley

Frontmatter

[More information](#)

PUBLISHED BY THE PRESS SYNDICATE OF THE UNIVERSITY OF CAMBRIDGE  
The Pitt Building, Trumpington Street, Cambridge, United Kingdom

CAMBRIDGE UNIVERSITY PRESS  
The Edinburgh Building, Cambridge CB2 2RU, UK  
40 West 20th Street, New York, NY 10011-4211, USA  
477 Williamstown Road, Port Melbourne, VIC 3207, Australia  
Ruiz de Alarcón 13, 28014 Madrid, Spain  
Dock House, The Waterfront, Cape Town 8001, South Africa

<http://www.cup.org>

© Cambridge University Press 2000

This book is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2000

First paperback edition, with corrections 2002

Printed in the United Kingdom at the University Press, Cambridge

Typeface: Univers and Bembo 11/13pt. System: QuarkXpress® [SE]

*A catalogue record for this book is available from the British Library*

*Library of Congress cataloguing in publication data*

Shipley, Bill, 1960–

Cause and correlation in biology: a user's guide to path analysis, structural equations and causal inference / Bill Shipley.

p. cm.

ISBN 0 521 79153 7 (hb)

1. Biometry. I. Title.

QH323.5.S477 2001

570'.1'5195 – dc21 00-037918

ISBN 0 521 79153 7 hardback

Cambridge University Press

0521791537 - Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference - Bill Shipley

Frontmatter

[More information](#)

---

*À ma petite Rhinanthé, David et Élyse.*

# Contents

<i>Preface</i>	xi
<b>1 Preliminaries</b>	1
1.1 The shadow's cause	1
1.2 Fisher's genius and the randomised experiment	7
1.3 The controlled experiment	14
1.4 Physical controls and observational controls	16
<b>2 From cause to correlation and back</b>	21
2.1 Translating from causal to statistical models	21
2.2 Directed graphs	25
2.3 Causal conditioning	28
2.4 d-separation	29
2.5 Probability distributions	32
2.6 Probabilistic independence	33
2.7 Markov condition	35
2.8 The translation from causal models to observational models	36
2.9 Counterintuitive consequences and limitations of d-separation: conditioning on a causal child	37
2.10 Counterintuitive consequences and limitations of d-separation: conditioning due to selection bias	41
2.11 Counterintuitive consequences and limitations of d-separation: feedback loops and cyclic causal graphs	42
2.12 Counterintuitive consequences and limitations of d-separation: imposed conservation relationships	43
2.13 Counterintuitive consequences and limitations of d-separation: unfaithfulness	45
2.14 Counterintuitive consequences and limitations of d-separation: context-sensitive independence	47
2.15 The logic of causal inference	48
2.16 Statistical control is not always the same as physical control	55
2.17 A taste of things to come	63
	vii

## CONTENTS

<b>3 Sewall Wright, path analysis and d-separation</b>	65
3.1 A bit of history	65
3.2 Why Wright's method of path analysis was ignored	66
3.3 d-sep tests	71
3.4 Independence of d-separation statements	72
3.5 Testing for probabilistic independence	74
3.6 Permutation tests of independence	79
3.7 Form-free regression	80
3.8 Conditional independence	83
3.9 Spearman partial correlations	88
3.10 Seed production in St Lucie's Cherry	90
3.11 Specific leaf area and leaf gas exchange	94
<b>4 Path analysis and maximum likelihood</b>	100
4.1 Testing path models using maximum likelihood	103
4.2 Decomposing effects in path diagrams	123
4.3 Multiple regression expressed as a path model	126
4.4 Maximum likelihood estimation of the gas-exchange model	130
<b>5 Measurement error and latent variables</b>	136
5.1 Measurement error and the inferential tests	138
5.2 Measurement error and the estimation of path coefficients	140
5.3 A measurement model	143
5.4 The nature of latent variables	152
5.5 Horn dimensions in Bighorn Sheep	157
5.6 Body size in Bighorn Sheep	158
5.7 Name calling	161
<b>6 The structural equations model</b>	162
6.1 Parameter identification	163
6.2 Structural underidentification with measurement models	164
6.3 Structural underidentification with structural models	171
6.4 Behaviour of the maximum likelihood chi-squared statistic with small sample sizes	173
6.5 Behaviour of the maximum likelihood chi-squared statistic with data that do not follow a multivariate normal distribution	179
6.6 Solutions for modelling non-normally distributed variables	185
6.7 Alternative measures of 'approximate' fit	188
6.8 Bentler's comparative fit index	192

Cambridge University Press

0521791537 - Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference - Bill Shipley

Frontmatter

[More information](#)

## CONTENTS

6.9	Approximate fit measured by the root mean square error of approximation	193
6.10	An SEM analysis of the Bumpus House Sparrow data	195
<b>7</b>	<b>Nested models and multilevel models</b>	199
7.1	Nested models	200
7.2	Multigroup models	202
7.3	The dangers of hierarchically structured data	209
7.4	Multilevel SEM	221
<b>8</b>	<b>Exploration, discovery and equivalence</b>	237
8.1	Hypothesis generation	237
8.2	Exploring hypothesis space	238
8.3	The shadow's cause revisited	241
8.4	Obtaining the undirected dependency graph	243
8.5	The undirected dependency graph algorithm	246
8.6	Interpreting the undirected dependency graph	250
8.7	Orienting edges in the undirected dependency graph using unshielded colliders assuming an acyclic causal structure	254
8.8	Orientation algorithm using unshielded colliders	256
8.9	Orienting edges in the undirected dependency graph using definite discriminating paths	260
8.10	The Causal Inference algorithm	262
8.11	Equivalent models	264
8.12	Detecting latent variables	266
8.13	Vanishing Tetrad algorithm	271
8.14	Separating the message from the noise	272
8.15	The Causal Inference algorithm and sampling error	278
8.16	The Vanishing Tetrad algorithm and sampling variation	284
8.17	Empirical examples	287
8.18	Orienting edges in the undirected dependency graph without assuming an acyclic causal structure	294
8.19	The Cyclic Causal Discovery algorithm	299
8.20	In conclusion . . .	304
	<i>Appendix</i>	305
	<i>References</i>	308
	<i>Index</i>	316

Cambridge University Press

0521791537 - Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference - Bill Shipley

Frontmatter

[More information](#)

## Preface

This book describes a series of statistical methods for testing causal hypotheses using observational data – but it is not a statistics book. It describes a series of algorithms, derived from research in Artificial Intelligence, that can discover causal relationships from observational data – but it is not a book about Artificial Intelligence. It describes the logical and philosophical relationships between causality and probability distributions – but it is certainly not a book about the philosophy of statistics. Rather it is a *user's guide*, written for biologists, whose purpose is to allow the practising biologist to make use of these important new developments when causal questions can't be answered with randomised experiments.

I have written the book assuming that you have no previous training in these methods. If you have taken an introductory statistics course – even if it was longer ago than you want to acknowledge – and have managed to hold on to some of the basic notions of sampling and hypothesis testing using statistics, then you should be able to understand the material in this book. I recommend that you read each chapter through in its entirety, even if you don't feel that you have mastered all of the notions. This will at least give you a general feeling for the goals and vocabulary of each chapter. You can then go back and pay closer attention to the details.

The book is addressed to biologists, mostly because I am a practising biologist myself, but I hope that it will also be of interest to statisticians, scientists in other fields and even philosophers of science. I have not written the book as a textbook simply because the discipline to which the material in this book naturally belongs does not yet exist. Whatever the name eventually given to this new discipline, I firmly believe that it will exist, and be generally recognised as a distinct discipline, in the future. The questions that this new discipline addresses, and the elegance of its results, are too important. None the less, the chapters follow a logical progression that would be well suited to an upper level undergraduate, or graduate, course. I have used the manuscript of this book for such a purpose and every one of my students is still alive.

Cambridge University Press

0521791537 - Cause and Correlation in Biology: A User's Guide to Path Analysis, Structural Equations and Causal Inference - Bill Shipley

Frontmatter

[More information](#)

## PREFACE

It is a pleasure and an honour to acknowledge the many people who have contributed to this project. First, Jim and Marg Shipley started everything. Robert van Hulst supplied much of the initial impulse through our conversations about science and causality while I was still an undergraduate. He has also read every one of the manuscript chapters and suggested many useful changes. Paul Keddy kept my interest burning during my Ph.D. studies and also commented on the first two chapters. As usual, his comments went to the heart of the matter.

The late Robert Peters had a large impact on my thoughts about causality and even convinced me, for a number of years, that ecologists are best to give up on the concept – not because he viewed the notion of causality as meaningless (he never believed this despite his empiricist reputation) but because it was simply too slippery a notion to demonstrate without randomised experiments. His constant prodding must have caused me to stop while wandering through the library one day when, almost subconsciously, I saw a book with the following provocative title: *Discovering causal structure. Artificial intelligence, philosophy of science, and statistical modeling* (Glymour *et al.* 1987). That book was my introduction to a more sophisticated understanding of causality. Rob Peters was much too young when he passed away and I am sorry that he never read the book that you are about to begin. I am not sure that he would have approved of everything in it but I know that he would have appreciated the effort.

Martin Lechowicz introduced me to the notion of path analysis at a time when this method had been mostly forgotten by biologists. He and I have collaborated for a number of years on this topic and he read the entire manuscript of this book, providing many insightful comments. Steve Côté and Jim Grace also read parts of this book. Jim, in particular, provided some important counterpoint to my thoughts on latent variable models. Marco Festa-Bianchet provided the unpublished data that is reported in Chapter 5. I must also acknowledge my graduate students, Margaret McKenna, Driss Meziane, Jarceline Almeida-Cortez, Luc St-Pierre and Muhaymina Sari, as well as the many members of the SEMNET Internet discussion group.

Finally, I want to thank Judea Pearl for kindly responding to my many emails about d-separation and basis sets and to Clark Glymour, Richard Scheines and Peter Spirtes of Carnegie–Mellon University for their generosity in extending an invitation to visit with them and for patiently answering my many questions about their discovery algorithms. Clark Glymour read and commented on some of the manuscript chapters.

I hope that you find this book to be useful, interesting and readable. I welcome your comments and feedback. Especially, if you don't agree with me.

Bill Shipley