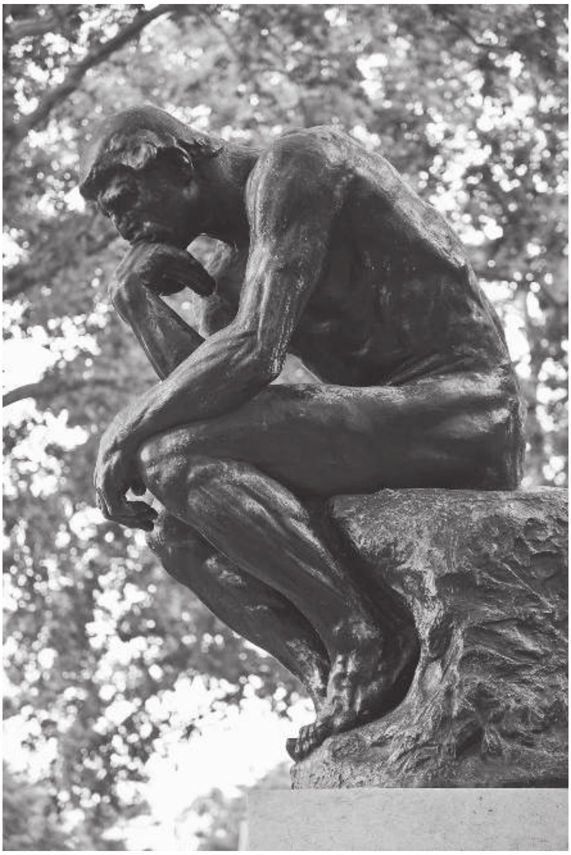
A Course in Morphometrics for Biologists

This book builds a much-needed bridge between biostatistics and organismal biology by linking the arithmetic of statistical studies of organismal form to the biological inferences that may follow from them. It incorporates a cascade of new explanations of regression, correlation, covariance analysis, and principal components analysis, before applying these techniques to an increasingly common data resource: the description of organismal forms by sets of landmark point configurations. For each data set, multiple analyses are interpreted and compared for insight into the relation between the arithmetic of the measurements and the rhetoric of the subsequent biological explanations. The text includes examples that range broadly over growth, evolution, and disease. For graduate students and researchers alike, this book offers a unique consideration of the scientific context surrounding the analysis of form in today's biosciences.

FRED L. BOOKSTEIN is generally considered the founder of modern morphometrics, an interdisciplinary field bridging computer vision, statistical science, and organismal biology. His most lasting contribution to the field is probably his 1989 invention of the thin-plate spline for depiction and decomposition of changes in landmark configurations, a method that has appeared in countless scientific publications, several courtroom proceedings, and even a dance concert. A Fellow of the Institute of Mathematical Statistics, he was the first winner (2011) of the Rohlf Medal for Excellence in Morphometrics. This is his eighth book.



Here, one of the iconic images of Western art, Rodin's *Thinker (Le Penseur)*, is photographed against sunsplashed urban greenery at the Rodin Museum in Philadelphia, Pennsylvania. In its radically different original setting, the same figure, at smaller scale, sits high on the sculptor's *La Porte de l'Enfer* (Gates of Hell), an assemblage of nearly 200 subordinate human forms exemplifying all of the paradoxes of humanity out of Dante's *Inferno.* I take this creation as an emblem of the reflexivity of human biology, humankind studying humankind. There is superb Michelangelesque detail in the heroic musculature here, but it is the pose, not the detail, that lets us infer a specific behavioral/emotional state: meditative study. And even the pedestal on which the creation sits conveys a metaphor for our science, sculpted as it is to project the Thinker off the edge of the secure toward the unknown. We would do well to imitate this level of disinterested engagement in the course of our own studies of this most interesting evolutionary lineage, from the first eukaryotes through the Cambrian explosion of metazoa and so down to our own sudden species.

© Peter Gridley/Stockbite/Getty Images

A Course in Morphometrics for Biologists

Geometry and Statistics for Studies of Organismal Form

FRED L. BOOKSTEIN
University of Washington, Seattle
University of Vienna, Austria

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE
UNIVERSITY PRESS

Shaftesbury Road, Cambridge CB2 8EA, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

314–321, 3rd Floor, Plot 3, Splendor Forum, Jasola District Centre, New Delhi – 110025, India

103 Penang Road, #05–06/07, Visioncrest Commercial, Singapore 238467

Cambridge University Press is part of Cambridge University Press & Assessment,
a department of the University of Cambridge.

We share the University's mission to contribute to society through the pursuit of
education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9781107190948

DOI: 10.1017/9781108120418

First published 2018

A catalogue record for this publication is available from the British Library

Library of Congress Cataloging-in-Publication data
Names: Bookstein, Fred L., 1947– author.
Title: A course in morphometrics for biologists : Geometry and Statistics for Studies
of Organismal Form/ Fred L. Bookstein, University of Washington, Seattle.
Description: Cambridge, United Kingdom ; New York, NY :
Cambridge University Press, 2018. | Includes bibliographical references and index.
Identifiers: LCCN 2018012338 | ISBN 9781107190948 (hardback : alk. paper)
Subjects: LCSH: Morphology–Statistical methods. | Morphology–Mathematics. |
Organisms–Size–Measurement.
Classification: LCC QH351 .B64 2018 | DDC 571.3–dc23
LC record available at https://lccn.loc.gov/2018012338

ISBN 978-1-107-19094-8 Hardback

Cambridge University Press & Assessment has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

Every effort has been made in preparing this book to provide accurate and up-to-date information which is in accord with accepted standards and practice at the time of publication. Although case histories are drawn from actual cases, every effort has been made to disguise the identities of the individuals involved. Nevertheless, the authors, editors and publishers can make no warranties that the information contained herein is totally free from error, not least because clinical standards are constantly changing through research and regulation. The authors, editors and publishers therefore disclaim all liability for direct or consequential damages resulting from the use of material contained in this book. Readers are strongly advised to pay careful attention to information provided by the manufacturer of any drugs or equipment that they plan to use.

Epigraphs

When you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind.

Sir William Thompson, Lord Kelvin,
as quoted in Kuhn, 1961

The most notable achievement of modern science . . . so far as its influence on intellectual culture is concerned, is the change that it has brought about in the standard of reasoning, in precision of thought and grasp of fundamental notions; this it has accomplished by creating a new type of rigorous thinking, more accurate and penetrating than the argumentation of an earlier age.

Edmund Whittaker, 1948

Ah, the nuanced interplay between narrative and number.

John Allen Paulos, 2015

A large acquaintance with particulars often makes us wiser than the possession of abstract formulas, however deep.

William James, 1902 (Preface)

Dedication

As ever, I am blessed by the generosity with which my wife Ede and my daughters Victoria Bookstein and Amelia Bookstein Kyazze granted me the solitude needed to construct the arguments here, and revise the text and endlessly redesign all the supporting diagrams, instead of answering their e-mails, admiring the YouTubes of the grandchildren, or participating in the city life of Seattle, Berkeley, or London. To these three Bookstein women, my enduring gratitude, now and forever.

Contents

Appendices to this book, including thirteen of the data sets used for examples, along with `Splus` code for the main software tools, some useful auxiliary routines, and nearly half of the book’s figures, can be found on the book’s website, `http://www.cambridge.org/9781107190948`.

Preface

The book you are holding is the companion to an earlier volume, *Measuring and Reasoning*, concerned with the shared logical structure of numerical inferences over a wide range of basic and applied scientific disciplines. The arguments there are broader than the specific needs of biology or any other single field of study. I needed to complement that treatise with a more accessible offering better suited to my actual teaching assignment, which mainly involved explaining methods for analyzing measurements of shape and form to advanced undergraduates and beginning graduate students in the organismal biosciences at the University of Vienna. Between 2010 and 2015, over a series of drafts and worked examples, a curriculum took shape along these lines. This book assembles the lectures and worked examples from those Vienna courses along with a variety of reflections and extensions.

By "morphometrics," the central noun of this book's title, I do not mean just the subdomain of "geometric morphometrics" that is the concern of most of my scientific papers and essays along with a considerable variety of others' textbooks, software packages, and short courses and workshops these days. Such techniques, specific to the analysis of Cartesian coordinate data from corresponding points or curves of a sample of organisms, are indeed part of our subject, especially in Chapter 5. But in their statistical algebra the analyses driving geometric morphometric studies all rely on a much more fundamental toolkit for the summary of patterns pertaining to one or more numerical measurements per se: patterns that apply to coherently measured suites of variables regardless of their empirical discipline of origin.

A pedagogy for morphometrics must necessarily unify and balance these two source streams, both the "geometry" and the "statistics" of my subtitle. The specifically *geo*metric component of the pedagogy had been laid out recently enough in one of the central chapters of a 2011 textbook with a Vienna colleague, Gerhard Weber, and also in the last two chapters of *Measuring and Reasoning*. But the *bio*metric part, the general language of patterning for studies of evolution or development – explications of the constructs like regressions and principal components that logically precede all the morphometrics – had to be drafted *de novo*. Such a presentation does not go well if it is limited to the standard statistics-textbook terminology of least-squares fits, significance tests, and bell curves. Instead, the pedagogy must anticipate the particular needs of the morphometrics to come – the passage from arithmetic to understanding as it concerns today's best and most carefully collated systems of multiple measurements of biological form. That passage emphasizes some specialized versions of pattern analysis, such as the spatial analysis of multiple measurements, and the corresponding statistical machinery needs to wrestle with models of whole covariance matrices, like the Wishart distribution, not just the simpler, more familiar Gaussian models of observed vector data. Ordinary textbooks of statistics, even advanced texts of multivariate analysis, do not teach the specific pattern languages that today's organismal biologist needs, and so a transition from prediction analysis to pattern analysis is mandatory for this applications domain.

Another central arena for this effort of reimagination was a senior capstone course, Statistics 423, "Regression and allied methods," that I taught at the University of Washington in the winter quarter of 2011. The forty students in that experimental setting were spectators at an improvisation on themes from most of the sample data sets in Chapters 2 and 3, themes that emphasized not the formulas but the justification of the scientific inferences that the formulas sometimes drive. The main examples of these two chapters – didactic data sets from ornithologist Hermon Bumpus and biostatisticians Karl Pearson and Sewall Wright, along with more recent exercises in regression by Secher et al., Tuddenham and Snyder, and Hellung-Larsen – were compiled for that course, and the exegeses here were worked out over the twelve weeks of lectures and student homework assignments in their light.

At the same time, over at the University of Vienna, most of my teaching in the Department of Anthropology was likewise modeled on the content of this accreting manuscript in one way or another. Principles for the appropriate reporting of regressions and correlations in the natural sciences, Chapters 2 and 3, lie at the core of the undergraduate course I taught called Wissenschaftliche Schreiben auf Englisch (Science Writing in English), and the appropriate reporting of covariance structures for shape coordinates, Chapter 5, was the major concern of my graduate course on geometric morphometrics. Five years of these Vienna students struggled through multiple earlier versions of the explanations of eigenanalysis here, and other students and postdocs had to intermittently but repeatedly update their morphometrics dissertations or derived publications whenever I would change the rhetoric of the techniques on which their work had hitherto been based. These stalwart, much-put-upon pioneers include Michael Coquerelle, Sascha Senck, Sonja Windhager, Stefano Benazzi, Jacqueline Domjanić, and Cinzia Fornai. Thanks go to all of you for staying the course. I am particularly grateful to the students in the Introduction to Morphometrics course for the past two years, who, in light of my threat to put any of my figures and any of my equations on the final exam, were obliged to master every detail of Chapters 2 and 3.

For engrossing conversations on these and countless related matters outside the classroom I am grateful to Philipp Mitteroecker (University of Vienna), Verena Winiwarter (University for Natural Resources and Life Sciences, Vienna), Kanti Mardia (University of Leeds), and Joe Felsenstein (University of Washington). Chapter 4, which attempts to ground the study of covariance structures in the classical topic of Wishart distributions, arose from a suggestion Mitteroecker tossed out in connection with the symmetry arguments driving our joint *Evolution* paper of 2009: On a spherical Gaussian hypothesis, why shouldn't the distribution of covariance distance (Figure 5.55) be nearly spherical in the Wishart distribution? (Why not, indeed? According to the elementary sketch of a proof on page 304, it actually is.) All this investment in Viennese pedagogy was under the oversight and continually generous support of Horst Seidler, first as my chairman in the Department of Anthropology and more recently (2008–2014) as Dean of the Faculty of Life Sciences, University of Vienna.

Other audiences for the pre-Broadway tryout period of this pedagogy include those seated for my talks or standing before my posters at the annual conferences of the American Association for Physical Anthropology in Albuquerque, Minneapolis, Portland, Knoxville, Calgary, and St. Louis, between 2010 and 2015; Philipp Gunz and the other thirty-odd predoctoral and postdoctoral fellows of the Marie Curie Labor Mobility Project "EVAN" (European Virtual Anthropology Network), Gerhard Weber (Vienna), director, 2006–2009; and Vienna faculty colleagues Katrin Schaefer, Martin Fieder, Hermann Prossinger, and Karl Grammer. An early formalization of this approach met with some success at a workshop "Measuring Biology" at the Konrad Lorenz Institute for Evolution and Cognition Research, Altenberg, Austria, in September 2008, and a later version at a follow-up conference, "Quality and Quantity," at the same institute (now relocated to Klosterneuburg) in June 2014. The talks at these two meetings were published as Bookstein

(2009a,b) and Bookstein (2015a), respectively. I am grateful to the KLI, to its director Gerd Müller, and to the late Werner Callebaut, its managing intellectual, for their repeated challenge that I turn my intuitions into comprehensible, reviewable presentations in this venue. From 2005 through 2015, the annual meetings of the Leeds Applied Statistics Research (LASR) workshops organized by Mardia and John Kent in Leeds, England, offered a comparable soapbox atop which to promulgate these ideas to a quite different professional audience – applied statisticians. Others who have engaged me at length on these matters include Norm MacLeod, The Natural History Museum, London; Clive Bowman, independent scholar, London; Paul O'Higgins, University of York, England; Daniel Cook, University of Washington; and Benedikt Hallgrimsson, University of Calgary. The specific idea of a relatively basic textbook spanning these topics was hatched over a supper of burgers and beer in 2009 with bioanthropology professor Matt Cartmill (then at Duke, now at Boston University).

The running examples here have several different origins. The Vilmann rodent neurocranial data were originally brought to my attention by the late Melvin Moss more than 35 years ago under a grant supported by the National Institute for Dental Research (a component of the U.S. National Institutes of Health). The data pertaining to fetal alcohol syndrome arose from research projects on adults (1997–2002) and then on newborn infants (2004–2007), led by Ann Streissguth, Fetal Alcohol and Drug Unit, Department of Psychiatry, University of Washington, with support from the National Institute on Alcohol Abuse and Alcoholism (likewise part of NIH). The translation of these studies into the context of discrimination that is their setting in Chapter 3 benefited greatly from conversations with Kathryn Kelly (FADU), William (Billy) Edwards (Office of the Public Defender, Los Angeles), and Arthur Kowell (Neurology, UCLA). Most of the other examples were drawn from one of the two textbooks that I exploited when I taught Statistics 423 in 2011: Per Andersen and Lene Skovgaard, *Regression with Linear Predictors,* 2010, and Sanford Weisberg, *Applied Linear Regression* (I used the 2005 edition). In particular, it was Weisberg who unearthed the Pearson and Tuddenham examples (although the treatment of BMI in Figure 2.40 is new to this volume), and Andersen and Skovgaard who realized the pedagogical value of the *Tetrahymena* and the fetal ultrasound data sets. Some other examples here (Bumpus, Perrin, Hackshaw) are borrowed from my earlier treatise (Bookstein, 2014) but have been rewritten for this disciplinarily focused context. The exemplary data on human skull growth were originally digitized by Philipp Gunz for exploitation in Bookstein, Gunz et al. (2003), with support from a grant to Horst Seidler from the Austrian Ministry of Science, and the Sewall Wright data on leghorn chickens are from the electronic version of Rohlf and Bookstein (1990). Several recent extensions of this work to random walk, to phylogenetic questions, to the assessment of morphological integration, and to aspects of the biomechanical analysis of strain were supported by U.S. National Science Foundation grant DEB–1019583 to Joe Felsenstein and me, as was the pedagogic experiment in expositing the Wishart distribution that concludes Chapter 4.

Since the 1990s I have depended for all my systems programming and computer network support, for the construction of more than a dozen desktop and laptop computer systems, and for two very large bespoke software packages (both named `Edgewarp`) on Dr. William D. K. Green (wdkg@wdkg.org), nowadays an independent consultant out of Bellingham, Washington, subsequent to intensive productive collaborations with me in Ann Arbor, Seattle, Vienna, and elsewhere. It was Green who guided me through the world of imaging data resources without which geometric morphometrics could not have been brought into the twenty-first century. The adult and baby fetal alcohol brain image data sets could not have been gathered without the `Edgewarp` program package he designed and wrote, and the dance video (Figure 5.83) fusing grids with the original moving imagery is all his doing. It is also Green who arranges things so that I can carry out computations on computers at three different addresses on two continents without ever needing to know what time zone I am in. Thanks, Bill, for this quarter-century of endlessly patient and creative support.

For their many useful comments I thank Lauren Cowles, my editor at Cambridge University Press, along with several anonymous reviewers. Philipp Mitteroecker and Clive Bowman read every word of an earlier draft in search of unsound generalizations to soften or hopelessly obscure mathematical notation that could be replaced by more accessible visualizations. The errors that remain are not their fault, of course, but mine alone. Please send your comments, complaints, and suggested inserts to flb@stat.washington.edu and fred.bookstein@univie.ac.at (using both addresses will optimize the chances of a comment getting through).

Fred Bookstein
Seattle, Washington, and Vienna, Austria
December, 2017