



*Essentials of*  
**Biostatistics in  
Public Health**  
*Second Edition*

Lisa M. Sullivan, PhD

Professor and Chair, Department of Biostatistics  
Associate Dean for Education  
Boston University School of Public Health  
Boston, Massachusetts



JONES & BARTLETT  
LEARNING

*World Headquarters*

Jones & Bartlett Learning  
40 Tall Pine Drive  
Sudbury, MA 01776  
978-443-5000  
info@jblearning.com  
www.jblearning.com

Jones & Bartlett Learning Canada  
6339 Ormindale Way  
Mississauga, Ontario L5V 1J2  
Canada

Jones & Bartlett Learning International  
Barb House, Barb Mews  
London W6 7PA  
United Kingdom

Jones & Bartlett Learning books and products are available through most bookstores and online booksellers. To contact Jones & Bartlett Learning directly, call 800-832-0034, fax 978-443-8000, or visit our website, [www.jblearning.com](http://www.jblearning.com).

Substantial discounts on bulk quantities of Jones & Bartlett Learning publications are available to corporations, professional associations, and other qualified organizations. For details and specific discount information, contact the special sales department at Jones & Bartlett Learning via the above contact information or send an email to [specialsales@jblearning.com](mailto:specialsales@jblearning.com).

Copyright © 2012 by Jones & Bartlett Learning, LLC

All rights reserved. No part of the material protected by this copyright may be reproduced or utilized in any form, electronic or mechanical, including photocopying, recording, or by any information storage and retrieval system, without written permission from the copyright owner.

This publication is designed to provide accurate and authoritative information in regard to the Subject Matter covered. It is sold with the understanding that the publisher is not engaged in rendering legal, accounting, or other professional service. If legal advice or other expert assistance is required, the service of a competent professional person should be sought.

**Production Credits**

Publisher: Michael Brown  
Associate Editor: Maro Gartside  
Editorial Assistant: Teresa Reilly  
Production Assistant: Rebekah Linga  
Senior Marketing Manager: Sophie Fleck  
Manufacturing and Inventory Control Supervisor: Amy Bacus  
Cover Design: Kate Ternullo  
Cover Image: © Kheng Guan Toh/Shutterstock, Inc.  
Printing and Binding: Malloy, Inc.  
Cover Printing: John Pow Company

To order this product, use ISBN: 978-1-4496-2394-4

**Library of Congress Cataloging-in-Publication Data**

Sullivan, Lisa M. (Lisa Marie), 1961-  
Essentials of biostatistics in public health / Lisa Sullivan. — 2nd ed.  
p. cm.  
Rev. ed. of: Essentials of biostatistics / Lisa M. Sullivan. 1st ed. c2008.  
Includes bibliographical references and index.  
ISBN-13: 978-0-7637-9531-3 (pbk.)  
ISBN-10: 0-7637-9531-3 (ibid.)  
1. Public health—Statistical methods. 2. Medical statistics. I. Sullivan, Lisa M. (Lisa Marie), 1961—  
Essentials of biostatistics. II. Title.  
RA409.S73 2012  
610.72—dc22

2010048845

6048

Printed in the United States of America

15 14 13 12 11 10 9 8 7 6 5 4 3 2 1

# Contents

Acknowledgments	vii
Preface	ix
Prologue	xi
About the Author	xiii
<b>Chapter 1 Introduction</b>	<b>1</b>
1.1 What Is Biostatistics?	2
1.2 What Are the Issues?	2
1.3 Summary	4
<b>Chapter 2 Study Designs</b>	<b>7</b>
2.1 Vocabulary	8
2.2 Observational Study Designs	8
2.3 Randomized Study Designs	13
2.4 The Framingham Heart Study	17
2.5 More on Clinical Trials	17
2.6 Sample Size Implications	20
2.7 Summary	20
2.8 Practice Problems	20
<b>Chapter 3 Quantifying the Extent of Disease</b>	<b>23</b>
3.1 Prevalence	23
3.2 Incidence	24
3.3 Relationships Between Prevalence and Incidence	26
3.4 Comparing the Extent of Disease Between Groups	27
3.5 Summary	30
3.6 Practice Problems	30

<b>Chapter 4</b>	<b>Summarizing Data Collected in the Sample</b>	<b>35</b>
4.1	Dichotomous Variables	36
4.2	Ordinal and Categorical Variables	40
4.3	Continuous Variables	49
4.4	Summary	60
4.5	Practice Problems	62
<b>Chapter 5</b>	<b>The Role of Probability</b>	<b>67</b>
5.1	Sampling	67
5.2	Basic Concepts	69
5.3	Conditional Probability	70
5.4	Independence	73
5.5	Bayes' Theorem	74
5.6	Probability Models	75
5.7	Summary	96
5.8	Practice Problems	96
<b>Chapter 6</b>	<b>Confidence Interval Estimates</b>	<b>101</b>
6.1	Introduction to Estimation	101
6.2	Confidence Intervals for One Sample, Continuous Outcome	103
6.3	Confidence Intervals for One Sample, Dichotomous Outcome	104
6.4	Confidence Intervals for Two Independent Samples, Continuous Outcome	106
6.5	Confidence Intervals for Matched Samples, Continuous Outcome	110
6.6	Confidence Intervals for Two Independent Samples, Dichotomous Outcome	113
6.7	Summary	117
6.8	Practice Problems	118
<b>Chapter 7</b>	<b>Hypothesis Testing Procedures</b>	<b>123</b>
7.1	Introduction to Hypothesis Testing	123
7.2	Tests with One Sample, Continuous Outcome	130
7.3	Tests with One Sample, Dichotomous Outcome	133
7.4	Tests with One Sample, Categorical and Ordinal Outcomes	134
7.5	Tests with Two Independent Samples, Continuous Outcome	138
7.6	Tests with Matched Samples, Continuous Outcome	141
7.7	Tests with Two Independent Samples, Dichotomous Outcome	144
7.8	Tests with More Than Two Independent Samples, Continuous Outcome	147
7.9	Tests for Two or More Independent Samples, Categorical and Ordinal Outcomes	156
7.10	Summary	160
7.11	Practice Problems	162

<b>Chapter 8</b>	<b>Power and Sample Size Determination</b>	<b>169</b>
8.1	Issues in Estimating Sample Size for Confidence Intervals Estimates	170
8.2	Issues in Estimating Sample Size for Hypothesis Testing	176
8.3	Summary	185
8.4	Practice Problems	185
<b>Chapter 9</b>	<b>Multivariable Methods</b>	<b>189</b>
9.1	Confounding and Effect Modification	190
9.2	The Cochran–Mantel–Haenszel Method	197
9.3	Introduction to Correlation and Regression Analysis	198
9.4	Multiple Linear Regression Analysis	206
9.5	Multiple Logistic Regression Analysis	210
9.6	Summary	213
9.7	Practice Problems	214
<b>Chapter 10</b>	<b>Nonparametric Tests</b>	<b>219</b>
10.1	Introduction to Nonparametric Testing	221
10.2	Tests with Two Independent Samples	223
10.3	Tests with Matched Samples	229
10.4	Tests with More Than Two Independent Samples	236
10.5	Summary	240
10.6	Practice Problems	241
<b>Chapter 11</b>	<b>Survival Analysis</b>	<b>243</b>
11.1	Introduction to Survival Data	245
11.2	Estimating the Survival Function	247
11.3	Comparing Survival Curves	252
11.4	Cox Proportional Hazards Regression Analysis	260
11.5	Extensions	264
11.6	Summary	265
11.7	Practice Problems	267
<b>Appendix</b>		<b>271</b>
<b>Glossary</b>		<b>289</b>
<b>Index</b>		<b>299</b>





# Acknowledgments

I am very grateful to Dr. Richard Riegelman for his unending energy and enthusiasm for this project. I am also thankful for the support and guidance from my friend and colleague, Professor Alexa Beiser. Special thanks to Vanessa Xanthakis and Carlee Moser for their careful editing and reviewing. I thank my family for their support and, most especially, I am thankful to Kimberly Dukes and Kevin Green for their unending love and support.

This textbook is dedicated to the memory of my cousin Catherine Render. Catherine was a remarkable woman. She was a loving daughter, sister, wife, mother, cousin, and friend who excelled in every role. She lost a long battle with breast cancer in January 2006. She fought through many setbacks, always pushing forward optimistically, taking care of everyone around her, never asking why. She was and always will be an inspiration to me and to so many others who were fortunate enough to know her.

Biostatisticians play an extremely important role in addressing important medical and public health problems. Unfortunately, there are many more problems than solutions. We must never lose sight of the fact that our work is important in improving health and well-being. We need qualified biostatisticians to work in research teams to address problems like breast cancer, cardiovascular disease, diabetes, and so many others.







# Preface

*Essentials of Biostatistics in Public Health, Second Edition* provides a fundamental and engaging background for students learning to apply and appropriately interpret biostatistical applications in the field of public health. The examples are real, important, and represent timely public health problems. The author aims to make the material relevant, practical, and interesting for students. Throughout the textbook, the author uses data from the Framingham Heart Study and from clinical trials in a variety of major areas. The author presents example applications involving important risk factors—such as blood pressure, cholesterol, smoking, and diabetes and their relationships to incident cardiovascular and cerebrovascular disease—throughout. Clinical trials investigating new drugs to lower cholesterol, to reduce pain, and to promote healing following surgery are also considered. The author presents examples with relatively few subjects to illustrate computations while minimizing the actual computation time, as a particular focus is mastery of “by-hand” computations. All of the techniques are then applied to and illustrated on real data from the Framingham Heart Study and large clinical trials. For each topic, the author discusses methodology—including assumptions, statistical computations, and the appropriate interpretation of results. Key formulas are summarized at the end of each chapter.





# Prologue

It is essential that all educated citizens avoid being deceived by data. It is a key skill for public health practitioners, future clinicians, and health researchers. In *Essentials of Biostatistics in Public Health*, Lisa M. Sullivan, PhD, guides students through this maze. The text uses an abundance of real and relevant examples drawn from her own experience working on the Framingham Heart Study and clinical trials.

*Essentials of Biostatistics in Public Health* takes an intuitive, step-by-step, hands-on approach to walking students through statistical principles. The approach emphasizes understanding what questions to ask and the appropriate interpretation of statistical results.

The second edition builds upon the success of the first edition, providing increased emphasis on state-of-the-art methods widely used in public health and clinical research. A new chapter on survival analysis introduces important techniques in statistics that have become standard practice in randomized clinical trials as well as cohort studies. A new chapter on nonparametric statistics introduces students to flexible approaches to statistical analysis that require fewer assumptions than many traditional techniques. An expansion of the chapter on multivariable methods complements the two new chapters.

Calculations in *Essentials of Biostatistics in Public Health* are designed to enhance understanding. The accompanying workbooks utilize Microsoft Office® Excel®, a statistical computing package that is generally available to public health practitioners, making the material familiar and approachable for students. The text for the second edition provides new practice problems for each chapter. The text and Excel-based workbooks also include new problems to accompany the two new chapters. The combination of text and workbook provides a solid foundation for future course work in statistics as well as the grounding needed to read and appreciate the health research literature.

*Essentials of Biostatistics in Public Health* is a tried-and-true approach. Lisa Sullivan has over two decades of experience teaching biostatistics to undergraduates as well as graduate students. As Assistant Dean for Undergraduate Programs in Public Health at Boston University, she developed and taught undergraduate courses in biostatistics. Today she is the Chair of the Department of Biostatistics and also the Associate Dean for Education at Boston University School of Public Health, which speaks to her unique ability to combine the skills of biostatistics with the skills of education.

Dr. Sullivan has repeatedly won teaching awards for her skills and commitment to education in biostatistics, including the Association of Schools of Public Health Award for Teaching Excellence. Her national roles include serving as Chair of the Public Health Academy of Distinguished Teachers.

Dr. Sullivan possesses a unique combination of sophisticated biostatistics expertise and a clear and engaging writing style—a rare combination indeed. Even a quick look at *Essentials of Biostatistics in Public Health* will convince you of her skills in communication and education.

I am delighted that Lisa Sullivan has included her book and workbooks in our *Essential Public Health* series. There is no better book to recommend for the anxious student first confronting the field of biostatistics. Students will find the book and workbook engaging and relevant. Just take a look and see for yourself.

Richard Riegelman, MD, MPH, PhD  
Editor, *Essential Public Health* Series



# About the Author

**Lisa M. Sullivan** has a PhD in statistics and is Professor and Chair of the Department of Biostatistics at the Boston University School of Public Health. She is also Associate Dean for Education. She teaches elementary biostatistics for MPH students and lectures in biostatistical methods for clinical researchers. Lisa is the principal investigator of the National Heart, Lung, and Blood Institute's *Summer Institute for Training in Biostatistics*, which is designed to promote interest in the field of biostatistics and to expose students to the many exciting career opportunities available to them. Lisa is the recipient of numerous teaching awards, including the Norman A. Scotch Award and the prestigious Metcalf Award, both for excellence in teaching at Boston University. In 2008 she won the Association of Schools of Public Health/Pfizer Excellence in Teaching Award. Lisa is also a biostatistician on the Framingham Heart Study, working primarily on developing and disseminating cardiovascular risk functions. She is active in several large-scale epidemiological studies for adverse pregnancy outcomes and in multidisciplinary research projects in prenatal diagnosis, obstetrics, autism, cardiovascular disease, and emergency medicine. Her work has resulted in over 160 peer-reviewed publications.

