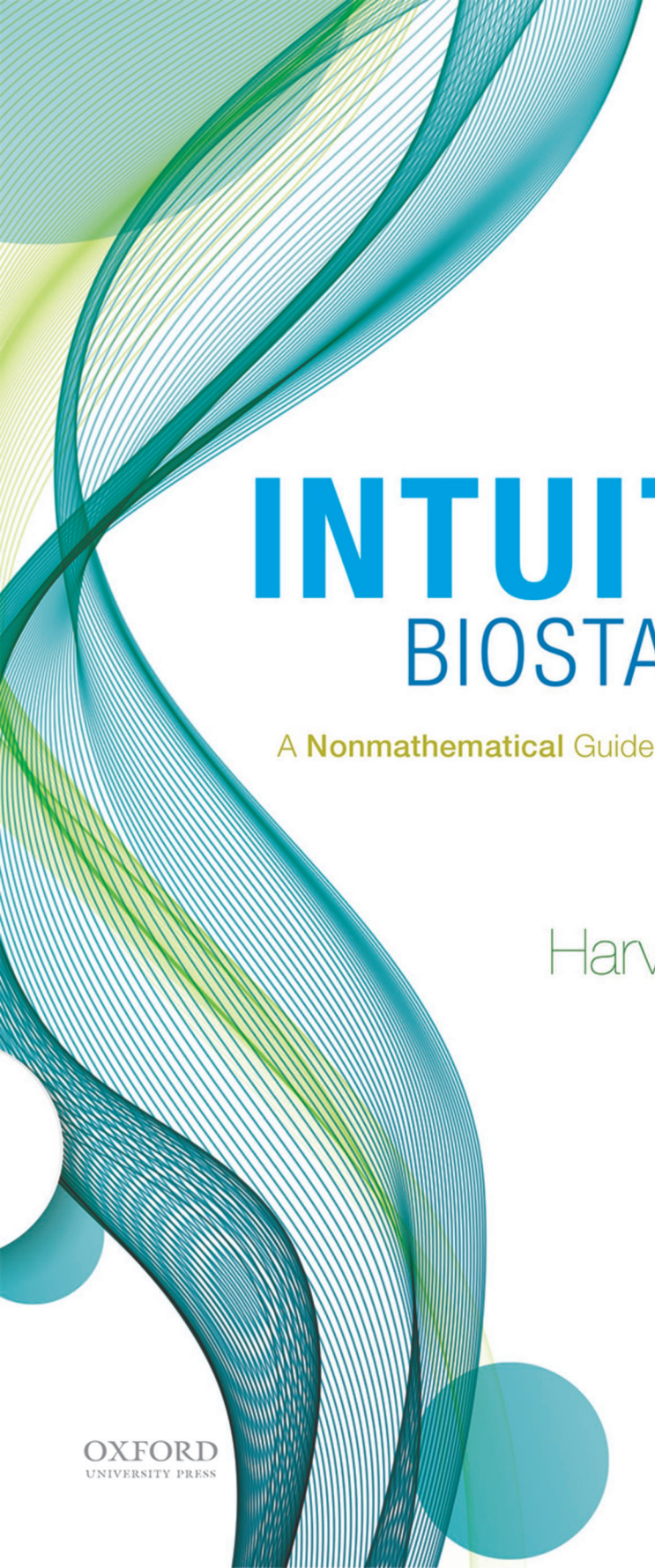

FOURTH EDITION

INTUITIVE BIOSTATISTICS

A **Nonmathematical** Guide to Statistical Thinking

Harvey Motulsky

OXFORD
UNIVERSITY PRESS

Intuitive Biostatistics

Intuitive Biostatistics

A Nonmathematical Guide to Statistical Thinking

HARVEY MOTULSKY, M.D.
GraphPad Software, Inc.

FOURTH EDITION

New York Oxford
OXFORD UNIVERSITY PRESS

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research,
scholarship, and education by publishing worldwide.

Oxford New York
Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in
Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Copyright © 2018 by Oxford University Press.

For titles covered by Section 112 of the US Higher Education Opportunity Act, please visit www.oup.com/us/he for the latest information about pricing and alternate formats.

Published by Oxford University Press
198 Madison Avenue, New York, New York 10016
www.oup.com

Oxford is a registered trademark of Oxford University Press.

All rights reserved. No part of this publication may be reproduced,
stored in a retrieval system, or transmitted, in any form or by any means,
electronic, mechanical, photocopying, recording, or otherwise,
without the prior permission of Oxford University Press.

CIP data is on file at the Library of Congress
978-0-19-064356-0

987654321
Printed by LSC Communications, Inc. United States of America

I dedicate this book to my wife, Lisa, to my kids (Wendy, Nat, Joey, and Ruby), to readers who encouraged me to continue with a fourth edition, and to future scientists who I hope will avoid common mistakes in biostatistics.

PRAISE FOR INTUITIVE BIOSTATISTICS

Intuitive Biostatistics is a beautiful book that has much to teach experimental biologists of all stripes. Unlike other statistics texts I have seen, it includes extensive and carefully crafted discussions of the perils of multiple comparisons, warnings about common and avoidable mistakes in data analysis, a review of the assumptions that apply to various tests, an emphasis on confidence intervals rather than P values, explanations as to why the concept of statistical significance is rarely needed in scientific work, and a clear explanation of nonlinear regression (commonly used in labs; rarely explained in statistics books).

In fact, I am so pleased with *Intuitive Biostatistics* that I decided to make it the reference of choice for my postdoctoral associates and graduate students, all of whom depend on statistics and most of whom need a closer awareness of precisely why. Motulsky has written thoughtfully, with compelling logic and wit. He teaches by example what one may expect of statistical methods and, perhaps just as important, what one may not expect of them. He is to be congratulated for this work, which will surely be valuable and perhaps even transformative for many of the scientists who read it.

—Bruce Beutler, 2011 Nobel Laureate, Physiology or Medicine
Director, Center for the Genetics of Host Defense
UT Southwestern Medical Center

GREAT FOR SCIENTISTS

This splendid book meets a major need in public health, medicine, and biomedical research training—a user-friendly biostatistics text for non-mathematicians that clearly explains how to make sense of statistical results, how to avoid common mistakes in data analysis, how to avoid being confused by statistical nonsense, and (new in this edition) how to make research more reproducible.
You may enjoy statistics for the first time!

—Gilbert S. Omenn, Professor of Medicine,
Genetics, Public Health, and Computational Medicine &
Bioinformatics, University of Michigan

I am entranced by the book. Statistics is a topic that is often difficult for many scientists to fully appreciate. The writing style and explanations of *Intuitive Biostatistics* makes the concepts accessible. I recommend this text to all researchers. Thank you for writing it.

—Tim Bushnell, Director of Shared Resource Laboratories,
University of Rochester Medical Center

GREAT FOR STUDENTS

After struggling with books that weren't right for my class, I was delighted to find *Intuitive Biostatistics*. It is the best starting point for undergraduate students seeking to learn the fundamental principles of statistics because of its unique presentation of the important concepts behind statistics. Lots of books give you the "recipe" approach, but only *Intuitive Biostatistics* explains what it all means. It meticulously goes through common mistakes and shows how to correctly choose, perform, and interpret the proper statistical test. It is accessible to new learners without being condescending.

—Beth Dawson,
The University of Texas at Austin

This textbook emphasizes the thinking needed to interpret statistical analysis in published research over knowledge of the mathematical underpinnings. The basics of choosing tests and doing simpler analyses are covered very clearly and simply. The language is easy to understand yet accurate. It brings in the higher level of intuitive understanding that we hope students will have at the end of an honors undergraduate or MSc program, skipping over the mathematical details that are now handled by software anyway. It is the prefect approach and level for undergraduates beginning research.

—Janet E. Kübler, Biology Department,
California State University at Northridge

I read many statistics textbooks and have come across very few that actually explain statistical concepts well. Yours is a stand-out exception. In particular, I think you've done an outstanding job of helping readers understand P values and confidence intervals, and yours is one of the very first introductory textbooks to discuss the crucial concept of false discovery rates. I have already recommended your text to postgraduate students and postdoctoral researchers at my own institute.

—Rob Herbert
Neuroscience Research Australia

GREAT FOR EVERYONE

I've read several statistics books but found that some concepts I was interested in were not mentioned and other concepts were hard to understand. You can ignore the "bio" in *Intuitive Biostatistics,* as it is the best applied statistics books I have come across, period. Its clear, straightforward explanations have allowed me to better understand research papers and select appropriate statistical tests. Highly recommended.

—Ariel H. Collis, Economist,
Georgetown Economic Services

BRIEF CONTENTS

PREFACE xxv

PART A **Introducing Statistics**
1. Statistics and Probability are not Intuitive 3
2. The Complexities of Probability 14
3. From Sample to Population 24

PART B **Introducing Confidence Intervals**
4. Confidence Interval of a Proportion 31
5. Confidence Interval of Survival Data 46
6. Confidence Interval of Counted Data (Poisson Distribution) 55

PART C **Continuous Variables**
7. Graphing Continuous Data 63
8. Types of Variables 75
9. Quantifying Scatter 80
10. The Gaussian Distribution 89
11. The Lognormal Distribution and Geometric Mean 95
12. Confidence Interval of a Mean 101
13. The Theory of Confidence Intervals 110
14. Error Bars 118

PART D **P Values and Statistical Significance**
15. Introducing P Values 129
16. Statistical Significance and Hypothesis Testing 145
17. Comparing Groups with Confidence Intervals and P Values 157
18. Interpreting a Result that is Statistically Significant 165
19. Interpreting a Result that is Not Statistically Significant 179

20. Statistical Power 186
21. Testing for Equivalence or Noninferiority 193

PART E Challenges in Statistics
22. Multiple Comparisons Concepts 203
23. The Ubiquity of Multiple Comparisons 214
24. Normality Tests 224
25. Outliers 232
26. Choosing a Sample Size 239

PART F Statistical Tests
27. Comparing Proportions 263
28. Case-Control Studies 273
29. Comparing Survival Curves 284
30. Comparing Two Means: Unpaired t Test 294
31. Comparing Two Paired Groups 306
32. Correlation 318

PART G Fitting Models to Data
33. Simple Linear Regression 331
34. Introducing Models 350
35. Comparing Models 357
36. Nonlinear Regression 366
37. Multiple Regression 378
38. Logistic and Proportional Hazards Regression 395

PART H The Rest of Statistics
39. Analysis of Variance 407
40. Multiple Comparison Tests After Anova 418
41. Nonparametric Methods 431
42. Sensitivity, Specificity, and Receiver Operating Characteristic Curves 442
43. Meta-Analysis 452

PART I Putting It All Together
44. The Key Concepts of Statistics 463
45. Statistical Traps To Avoid 468
46. Capstone Example 487
47. Statistics and Reproducibility 502
48. Checklists for Reporting Statistical Methods and Results 511

PART J APPENDICES 517
REFERENCES 533
INDEX 548

CONTENTS

PREFACE xxv
Who is this Book for?
What makes the Book Unique?
What's New?
Which Chapters are Essential?
Who Helped?
Who Am I?

PART A **Introducing Statistics**

1. Statistics and Probability are not Intuitive 3
We Tend to Jump to Conclusions
We Tend to Be Overconfident
We See Patterns in Random Data
We Don't Realize that Coincidences are Common
We Don't Expect Variability to Depend on Sample Size
We Have Incorrect Intuitive Feelings about Probability
We Find It Hard to Combine Probabilities
We Don't Do Bayesian Calculations Intuitively
We are Fooled by Multiple Comparisons
We Tend to Ignore Alternative Explanations
We are Fooled By Regression to the Mean
We Let Our Biases Determine How We Interpret Data
We Crave Certainty, but Statistics Offers Probabilities
Chapter Summary
Term Introduced in this Chapter

2. The Complexities of Probability 14
Basics of Probability
Probability as Long-Term Frequency

Probability as Strength of Belief
Calculations With Probabilities Can be Easier If You Switch to Calculating with Whole Numbers
Common Mistakes: Probability
Lingo
Probability in Statistics
Q & A
Chapter Summary
Terms Introduced in this Chapter

3. From Sample to Population 24
Sampling from a Population
Sampling Error and Bias
Models and Parameters
Multiple Levels of Sampling
What if Your Sample is the Entire Population?
Chapter Summary
Terms Introduced in this Chapter

PART B Introducing Confidence Intervals

4. Confidence Interval of a Proportion 31
Data Expressed as Proportions
The Binomial Distribution: From Population to Sample
Example: Free Throws in Basketball
Example: Polling Voters
Assumptions: Confidence Interval of a Proportion
What Does 95% Confidence Really Mean?
Are You Quantifying the Event You Care About?
Lingo
Calculating the CI of a Proportion
Ambiguity if the Proportion is 0% or 100%
An Alternative Approach: Bayesian Credible Intervals
Common Mistakes: CI of a Proportion
Q & A
Chapter Summary
Terms Introduced in this Chapter

5. Confidence Interval of Survival Data 46
Survival Data
Censored Survival Data
Calculating Percentage Survival at Various Times
Graphing Survival Curves with Confidence Bands
Summarizing Survival Curves

Assumptions: Survival Analysis
Q & A
Chapter Summary
Terms Introduced in this Chapter

6. Confidence Interval of Counted Data (Poisson Distribution) 55
The Poisson Distribution
Assumptions: Poisson Distribution
Confidence Intervals Based on Poisson Distributions
How to Calculate the Poisson CI
The Advantage of Counting for Longer Time Intervals (Or in Larger Volumes)
Q & A
Chapter Summary
Term Introduced in this Chapter

PART C **Continuous Variables**

7. Graphing Continuous Data 63
Continuous Data
The Mean and Median
Lingo: Terms Used to Explain Variability
Percentiles
Graphing Data to Show Variation
Graphing Distributions
Beware of Data Massage
Q & A
Chapter Summary
Terms Introduced in this Chapter

8. Types of Variables 75
Continuous Variables
Discrete Variables
Why It Matters?
Not Quite as Distinct as They Seem
Q & A
Chapter Summary
Terms Introduced in this Chapter

9. Quantifying Scatter 80
Interpreting a Standard Deviation
How it Works: Calculating SD
Why n – 1?
Situations in Which n Can Seem Ambiguous

SD and Sample Size
Other Ways to Quantify and Display Variability
Q & A
Chapter Summary
Terms Introduced in this Chapter

10. The Gaussian Distribution 89
The Nature of The Gaussian Distribution
SD and the Gaussian Distribution
The Standard Normal Distribution
The Normal Distribution does not Define Normal Limits
Why The Gaussian Distribution is so Central to Statistical Theory
Q & A
Chapter Summary
Terms Introduced in this Chapter

11. The Lognormal Distribution and Geometric Mean 95
The Origin of a Lognormal Distribution
Logarithms?
Geometric Mean
Geometric SD
Common Mistakes: Lognormal Distributions
Q & A
Chapter Summary
Terms Introduced in this Chapter

12. Confidence Interval of a Mean 101
Interpreting A CI of a Mean
What Values Determine the CI of a Mean?
Assumptions: CI of a Mean
How to Calculate the CI of a Mean
More about Confidence Intervals
Q & A
Chapter Summary
Terms Introduced in this Chapter

13. The Theory of Confidence Intervals 110
CI of a Mean Via the t Distribution
CI of a Mean Via Resampling
CI of a Proportion Via Resampling
CI of a Proportion Via Binomial Distribution
Q & A
Chapter Summary
Terms Introduced in this Chapter

14. Error Bars 118
SD Versus Sem
Which Kind of Error Bar Should I Plot?
The Appearance of Error Bars
How are SD and Sem Related to Sample Size?
Geometric SD Error Bars
Common Mistakes: Error Bars
Q & A
Chapter Summary
Terms Introduced in this Chapter

PART D **P Values and Statistical Significance**

15. Introducing P Values 129
Introducing P Values
Example 1: Coin Flipping
Example 2: Antibiotics on Surgical Wounds
Example 3: Angioplasty and Myocardial Infarction
Lingo: Null Hypothesis
Why P Values are Confusing
One- Or Two-Tailed P Value?
P Values are Not Very Reproducible
There is much more to Statistics than P Values
Common Mistakes: P Values
Q & A
Chapter Summary
Terms Introduced in this Chapter

16. Statistical Significance and Hypothesis Testing 145
Statistical Hypothesis Testing
Analogy: Innocent Until Proven Guilty
Extremely Significant? Borderline Significant?
Lingo: Type I and Type II Errors
Tradeoffs When Choosing a Significance Level
What Significance Level Should You Choose?
Interpreting A CI, A P Value, and A Hypothesis Test
Statistical Significance vs. Scientific Significance
Common Mistakes: Statistical Hypothesis Testing
Q & A
Chapter Summary
Terms Defined in this Chapter

17. Comparing Groups with Confidence Intervals and P Values 157
CIS and Statistical Hypothesis Testing are Closely Related

Four Examples with CIS, P Values, and Conclusion about Statistical Significance
Q & A
Chapter Summary

18. Interpreting a Result that is Statistically Significant 165
Seven Explanations for Results that are "Statistically Significant"
How Frequently do Type I Errors (False Positives) Occur?
The Prior Probability Influences the FPRP (A Bit of Bayes)
Bayesian Analysis
Accounting for Prior Probability Informally
The Relationship Between Sample Size and P Values
Common Mistakes
Q & A
Chapter Summary
Terms Introduced in this Chapter

19. Interpreting a Result that is not Statistically Significant 179
Five Explanations For "Not Statistically Significant" Results
"Not Significantly Different" does not Mean "No Difference"
Example: α_2-Adrenergic Receptors on Platelets
Example: Fetal Ultrasounds
How to Get Narrower CIS
What if the P Value is Really High?
Q & A
Chapter Summary

20. Statistical Power 186
What is Statistical Power?
Distinguishing Power From Beta and the False Discovery Rate
An Analogy to Understand Statistical Power
Power of the Two Example Studies
When does It Make Sense to Compute Power?
Common Mistakes: Power
Q & A
Chapter Summary
Terms Introduced in this Chapter

21. Testing For Equivalence or Noninferiority 193
Equivalence must be Defined Scientifically, not Statistically
If the Mean is Within the Equivalence Zone

If the Mean is Outside of the Equivalence Zone
Applying the Usual Approach of Statistical Hypothesis Testing to Testing for Equivalence
Noninferiority Tests
Common Mistakes: Testing for Equivalence
Q & A
Chapter Summary
Terms Introduced in this Chpater

PART E Challenges in Statistics

22. Multiple Comparisons Concepts 203
The Problem of Multiple Comparisons
Correcting for Multiple Comparisons is not Always Needed
The Traditional Approach to Correcting for Multiple Comparisons
Correcting for Multiple Comparisons with the False Discovery Rate
Comparing the Two Methods of Correcting for Multiple Comparisons
Q & A
Chapter Summary
Terms Introduced in this Chapter

23. The Ubiquity of Multiple Comparisons 214
Overview
Multiple Comparisons in Many Contexts
When are Multiple Comparisons Data Torture or P-Hacking?
How to Cope with Multiple Comparisons
Q & A
Chapter Summary
Terms Introduced in this Chapter

24. Normality Tests 224
The Gaussian Distribution is an Unreachable Ideal
What A Gaussian Distribution Really Looks Like
QQ Plots
Testing for Normality
Alternatives to Assuming a Gaussian Distribution
Common Mistakes: Normality Tests
Q & A
Chapter Summary
Terms Introduced in this Chapter

25. Outliers 232
How do Outliers Arise?
The Need for Outlier Tests
Five Questions to Ask before Testing for Outliers
Outlier Tests
Is It Legitimate to Remove Outliers?
An Alternative: Robust Statistical Tests
Lingo: Outlier
Common Mistakes: Outlier Tests
Q & A
Chapter Summary
Terms Introduced in this Chapter

26. Choosing a Sample Size 239
Sample Size Principles
An Alternative Way to think about Sample Size Calculations
Interpreting a Sample Size Statement
Lingo: Power
Calculating the Predicted FPRP as Part of Interpreting a Sample Size Statement
Complexities when Computing Sample Size
Examples
Other Approaches to Choosing Sample Size
Common Mistakes: Sample Size
Q & A
Chapter Summary
Terms Introduced in this Chapter

PART F **Statistical Tests**

27. Comparing Proportions 263
Example: Apixaban for Treatment of Thromboembolism
Assumptions
Comparing Observed and Expected Proportions
Common Mistakes: Comparing Proportions
Q & A
Chapter Summary
Terms Introduced in this Chapter

28. Case-Control Studies 273
Example: Does a Cholera Vaccine Work?
Example: Isotretinoin and Bowel Disease
Example: Genome-Wide Association Studies

How are Controls Defined?
How are Cases Defined?
Epidemiology Lingo
Common Mistakes: Case-Control Studies
Q & A
Chapter Summary
Terms Introduced in this Chapter

29. Comparing Survival Curves 284
Example Survival Data
Assumptions when Comparing Survival Curves
Comparing Two Survival Curves
Why Not Just Compare Mean or Median Survival Times or Five-Year Survival?
Intention to Treat
Q & A
Chapter Summary
Terms Introduced in this Chapter

30. Comparing Two Means: Unpaired t Test 294
Interpreting Results from an Unpaired t Test
Assumptions: Unpaired t Test
The Assumption of Equal Variances
Overlapping Error Bars and the t Test
How It Works: Unpaired t Test
Common Mistakes: Unpaired t Test
Q & A
Chapter Summary
Terms Introduced in this Chapter

31. Comparing Two Paired Groups 306
When to Use Special Tests for Paired Data
Example of Paired t Test
Interpreting Results from a Paired t Test
The Ratio Paired t Test
McNemar's Test for a Paired Case-Control Study
Common Mistakes: Paired t Test
Q & A
Chapter Summary
Terms Introduced in this Chapter

32. Correlation 318
Introducing the Correlation Coefficient
Assumptions: Correlation
Lingo: Correlation

How It Works: Calculating the Correlation Coefficient
Common Mistakes: Correlation
Q & A
Chapter Summary
Terms Introduced in this Chapter

PART G Fitting Models to Data

33. Simple Linear Regression 331
The Goals of Linear Regression
Linear Regression Results
Assumptions: Linear Regression
Comparison of Linear Regression and Correlation
Lingo: Linear Regression
Common Mistakes: Linear Regression
Q & A
Chapter Summary
Terms Introduced in this Chapter

34. Introducing Models 350
Lingo: Models, Parameters, and Variables
The Simplest Model
The Linear Regression Model
Why Least Squares?
Other Models and other Kinds of Regression
Common Mistakes: Models
Chapter Summary
Terms Introduced in this Chapter

35. Comparing Models 357
Comparing Models is a Major Part of Statistics
Linear Regression as a Comparison of Models
Unpaired t Test Recast as Comparing the Fit of Two Models
Common Mistakes: Comparing Models
Q & A
Chapter Summary
Terms Introduced in this Chapter

36. Nonlinear Regression 366
Introducing Nonlinear Regression
An Example of Nonlinear Regression
Nonlinear Regression Results
How Nonlinear Regression Works
Assumptions: Nonlinear Regression
Comparing Two Models

Tips for Understanding Models
Learn More About Nonlinear Regression
Common Mistakes: Nonlinear Regression
Q & A
Chapter Summary
Terms Introduced in this Chapter

37. Multiple Regression 378
Goals of Multivariable Regression
Lingo
An Example of Multiple Linear Regression
Assumptions
Automatic Variable Selection
Sample Size for Multiple Regression
More Advanced Issues with Multiple Regression
Common Mistakes: Multiple Regression
Q & A
Chapter Summary
Terms Introduced in this Chapter

38. Logistic and Proportional Hazards Regression 395
Logistic Regression
Proportional Hazards Regression
Assumptions: Logistic Regression
Common Mistakes: Logistic Regression
Q & A
Chapter Summary
Terms Introduced in this Chapter

PART H **The Rest of Statistics**

39. Analysis of Variance 407
Comparing the Means of Three or More Groups
Assumptions: One-Way Anova
How It Works: One-Way Anova
Repeated-Measures One Way Anova
An Example of Two-Way Anova
How Two-Way Anova Works
Repeated Measures Two-way Anova
Common Mistakes: Anova
Q & A
Chapter Summary
Terms Introduced in this Chapter

40. Multiple Comparison Tests after Anova 418
Multiple Comparison Tests for the Example Data
The Logic Of Multiple Comparisons Tests
Other Multiple Comparisons Tests
How It Works: Multiple Comparisons Tests
When Are Multiple Comparisons Tests Not Needed?
Common Mistakes: Multiple Comparisons
Q & A
Chapter Summary
Terms Introduced in this Chapter

41 Nonparametric Methods 431
Nonparametric Tests Based on Ranks
The Advantages and Disadvantages of Nonparametric Tests
Choosing Between Parametric and Nonparametric Tests: Does It Matter?
Sample Size for Nonparametric Tests
Nonparametric Tests that Analyze Values (Not Ranks)
Common Mistakes: Nonparametric Tests
Q & A
Chapter Summary
Terms Introduced in this Chapter

42 Sensitivity, Specificity, and Receiver Operating Characteristic Curves 442
Definitions of Sensitivity and Specificity
The Predictive Value of a Test
Receiver-Operating Characteristic (ROC) Curves
Bayes Revisited
Common Mistakes
Q & A
Chapter Summary
Terms Introduced in this Chapter

43. Meta-Analysis 452
Introducing Meta-Analysis
Publication Bias
Results from a Meta-Analysis
Meta-Analysis of Individual Participant Data
Assumptions of Meta-Analysis
Common Mistakes: Meta-Analysis
Q & A
Chapter Summary
Terms Introduced in this Chapter

PART I **Putting It All Together**

44. The Key Concepts of Statistics 463
Term Introduced in this Chapter

45. Statistical Traps to Avoid 468
Trap #1: Focusing on P Values and Statistical Significance Rather than Effect Size
Trap #2: Testing Hypotheses Suggested by the Data
Trap #3: Analyzing Without a Plan—"P-Hacking"
Trap #4: Making a Conclusion about Causation When the Data Only Show Correlation
Trap #5: Overinterpreting Studies that Measure a Proxy or Surrogate Outcome
Trap #6: Overinterpreting Data from an Observational Study
Trap #7: Making Conclusions about Individuals when the Data Were only Collected for Groups
Trap #8: Focusing Only on Means Without asking about Variability or Unusual Values
Trap #9: Comparing Statistically Significant with Not Statistically Significant
Trap #10: Missing Important Findings Because Data Combine Populations
Trap #11: Invalid Multiple Regression Analyses as a Result of an Omitted Variable
Trap #12: Overfitting
Trap #13: Mixing Up the Significance Level with the FPRP
Trap #14: Not Recognizing How Common False Positive Findings are
Trap #15: Not Realizing How Likely it is that a "Significant" Conclusion From a Speculative Experiment is a False Positive
Trap #16: Not Realizing That Many Published Studies have Little Statistical Power
Trap #17: Trying to Detect Small Signals When there is Lots of Noise
Trap #18: Unnecessary Dichotomizing
Trap #19: Inflating Sample Size by Pseudoreplication
Chapter Summary
Terms Introduced in this Chapter

46. Capstone Example 487
The Case of the Eight Naked IC_{50}S
Look Behind the Data

Statistical Significance by Cheating
Using a t Test That Doesn't Assume Equal SDs
Unpaired t Test as Linear or Nonlinear Regression
Nonparametric Mann–Whitney Test
Just Report the Last Confirmatory Experiment?
Increase Sample Size?
Comparing the Logarithms of IC_{50} Values
Sample Size Calculations Revisited
Is it Ok to Switch Analysis Methods?
The Usefulness of Simulations
Chapter Summary

47. Statistics and Reproducibility 502
The Repoducibility Crisis
Many Analyses are Biased to Inflate the Effect Size
Even Perfectly Performed Experiments are Less Reproducible than Most Expect
Summary

48. Checklists for Reporting Statistical Methods and Results 511
Reporting Methods Used for Data Analysis
Graphing Data
Reporting Statistical Results

PART J APPENDICES 517
Appendix A: Statistics with Graphpad
Appendix B: Statistics with Excel
Appendix C: Statistics with R
Appendix D: Values of the t Distribution Needed to Compute CIs
Appendix E: A Review of Logarithms
Appendix F: Choosing a Statistical Test
Appendix G: Problems and Answers

REFERENCES 533

INDEX 548

PREFACE

> My approach in this book is informal and brisk (at least I hope it is), not ceremonious and plodding (at least I hope it isn't).
>
> John Allen Paulos (2008)

Intuitive Biostatistics provides a comprehensive overview of statistics without getting bogged down in the mathematical details. I've been gratified to learn that many people have found my approach refreshing and useful. Some scientists have told me that statistics had always been baffling until they read *Intuitive Biostatistics*. This enthusiasm encouraged me to write this fourth edition.

WHO IS THIS BOOK FOR?

I wrote *Intuitive Biostatistics* for three main audiences:

- Medical (and other) professionals who want to understand the statistical portions of journals they read. These readers don't need to analyze any data, but they do need to understand analyses published by others and beware of common statistical mistakes. I've tried to explain the big picture without getting bogged down in too many details.
- Undergraduate and graduate students, postdocs, and researchers who analyze data. This book explains general principles of data analysis, but it won't teach you how to do statistical calculations or how to use any particular statistical program. It makes a great companion to the more traditional statistics texts and to the documentation of statistical software.
- Scientists who consult with statisticians. Statistics often seems like a foreign language, and this text can serve as a phrase book to bridge the gap between scientists and statisticians. Sprinkled throughout the book are "Lingo" sections that explain statistical terminology and point out when ordinary words are given very specialized meanings (the source of much confusion).

I wrote *Intuitive Biostatistics* to be a guidebook, not a cookbook. The focus is on how to interpret statistical results, rather than how to analyze data. This book presents few details of statistical methods and only a few tables required to complete the calculations.

If you think this book is too long, check out my other book, *Essential Biostatistics*, which is about one-third the size and price of this one (Motulsky, 2015).

WHAT MAKES THE BOOK UNIQUE?

Nonmathematical

Statistics is a branch of mathematics, so you can't truly understand statistics without studying many equations. In fact, some aspects of statistics cannot be understood unless you first master calculus and matrix algebra. Don't despair. This book is for the many students and scientists who may find math confusing and prefer verbal explanations. I only include equations in a few places for two reasons. One reason is that in a few places seeing an equation really helps explain a concept. The other reason is that a few simple calculations can easily be done by hand, and in these cases I present the necessary equation. Compared with most statistics books, this one uses very little math.

Statistical methods are tools, and scientists use lots of tools they don't always fully understand. For example, one can effectively use a radioactively labeled compound in an experiment without really understanding the nuclear physics of radioactive decay. You can do experiments with chemical reagents that you don't know how to synthesize. You can use scientific instruments without understanding exactly how they work. Similarly, you can interpret the results calculated by statistical software without understanding how those tests were derived and without having studied mathematical proofs that they work. What you need to know is the list of assumptions the test is based upon and the list of common conceptual traps to avoid.

Statistical lingo

In Lewis Carroll's *Through the Looking Glass*, Humpty Dumpty says, "When I use a word, it means exactly what I say it means—neither more nor less" (Carroll, 1871). Lewis Carroll (the pseudonym for Charles Dodgson) was a mathematician, and it almost seems he was thinking of statisticians when he wrote that line. But that can't be true because little statistical terminology had been invented by 1871.

Statistics books can get especially confusing when they use words and phrases that have both ordinary meanings and technical meanings. The problem is that you may think the author is using the ordinary meaning of a word or phrase, when in fact the author is using that word or phrase as a technical term with a very different meaning. I try hard to point out these potential ambiguities when I use potentially confusing terms such as:

- Significant
- Error
- Hypothesis
- Model
- Power
- Variance

- Residual
- Normal
- Independent
- Sample
- Population
- Fit
- Confidence
- Distribution
- Control

Includes topics omitted from many texts

I include many topics often omitted from short introductory texts, including:

- How common sense can mislead. Chapter 1 is a fun chapter that explains how common sense can lead you astray and why we therefore need to understand statistical principles.
- Multiple comparisons. It is simply impossible to understand statistical results without a deep understanding of how to think about multiple comparisons. Chapters 22, 23, and 40 explain several approaches used to deal with multiple comparisons, including the false discovery rate (FDR).
- Nonlinear regression. In many fields of science, nonlinear regression is used more often than linear regression, but most introductory statistics books ignore nonlinear regression completely. This book gives them equal weight. Chapters 34 and 35 set the stage by explaining the concept of fitting models to data and comparing alternative models. Chapter 36 then discusses nonlinear regression.
- Bayesian logic. Bayesian thinking is briefly mentioned in Chapter 2 and is then explored in Chapter 18 as a way to interpret a finding that a comparison is statistically significant. This topic returns in Chapter 42, which compares interpreting statistical significance to interpreting the results of clinical laboratory tests. These are only brief introductions to Bayesian thinking. However, a summary chapter about Bayesian approaches to data analysis, written with Daniel Lakens (Eindhoven University of Technology), will be posted at www.intuitivebiostatistics.com
- Lognormal distributions. These are commonly found in scientific data, but are rarely found in statistics books. They are explained in Chapter 11 and are touched upon again in several examples that appear in later chapters. Logarithms and antilogarithms are reviewed in Appendix E.
- Testing for equivalence. Sometimes the goal is not to prove that two groups differ but rather to prove that they are the same. This requires a different mindset, as explained in Chapter 21.
- Normality tests. Many statistical tests assume data are sampled from a Gaussian (also called normal) distribution, and normality tests are used to test this assumption. Chapter 24 explains why these tests are less useful than many hope.

- Outliers. Values far from the other values in a set are called outliers. Chapter 25 explains how to think about outliers.
- Comparing the fit of alternative models. Statistical hypothesis testing is usually viewed as a way to test a null hypothesis. Chapter 35 explains an alternative way to view statistical hypothesis testing as a way to compare the fits of alternative models.
- Meta-analysis as a way to reach conclusions by combining data from several studies (Chapter 43).
- Detailed review of assumptions. All analyses are based on a set of assumptions, and many chapters discuss these assumptions in depth.
- Lengthy discussion of common mistakes in data analysis. Most chapters include lists (with explanations) of common mistakes and misunderstandings.

Omits topics covered in most texts

To make space for the topics listed in the prior section, I have left out many topics that are traditionally included in introductory texts:

Probability. I assume that you have at least a vague familiarity with the ideas of probability, and this book does not explain these principles in much depth. Chapter 2 explains why probability can seem confusing. But you can still understand the rest of the book even if you skip this chapter.

Equations needed to compute statistical tests. I assume that you will be either interpreting data analyzed by others or using statistical software to run statistical tests. In only a few places do I give enough details to compute the tests by hand.

Statistical tables. If you aren't going to be analyzing data by hand, there is very little need for statistical tables. I include only a few tables in places where it might be useful to do simple calculations by hand.

Statistical distributions. You can choose statistical tests and interpret the results without knowing much about z, t, and F distributions. This book mentions them but goes into very little depth.

A unique organization

The organization of the book is unique.

Part A has three introductory chapters. Chapter 1 explains how common sense can mislead us when thinking about probability and statistics. Chapter 2 briefly explains some of the complexities of dealing with probability. Chapter 3 explains the basic idea of statistics—to make general conclusions from limited data, to extrapolate from sample to population.

Part B explains confidence intervals (CIs) in three contexts. Chapter 4 introduces the concept of a CI in the context of CIs of a proportion. I think this is the simplest example of a CI, since it requires no background information. Most books would start with the CI of the mean, but this would require first explaining the Gaussian distribution, the standard deviation, and the difference between the standard deviation and the standard error of the mean. CIs of proportions are much easier to understand. Chapters 5 and 6 are short chapters that explain CIs of survival data and Poisson (counted) data. Many instructors will choose to skip these two chapters.

Part C finally gets to continuous data, the concept with which most statistics books start. The first three chapters are fairly conventional, explaining how to graph continuous data, how to quantify variability, and the Gaussian distribution. Chapter 11 is about lognormal distributions, which are common in biology but are rarely explained in statistics texts. Chapter 12 explains the CI of the mean, and Chapter 13 is an optional chapter that gives a taste of the theory behind CIs. Then comes an important chapter (Chapter 14), which explains the different kinds of error bars, emphasizing the difference between the standard deviation and the standard error of the mean (which are frequently confused).

Part D is unconventional, as it explains the ideas of P values, statistical hypothesis testing, and statistical power without explaining any statistical tests. I think it is easier to learn the concepts of a P value and statistical significance apart from the details of a particular test. This section also includes a chapter (unusual for introductory books) on testing for equivalence.

Part E explains challenges in statistics. The first two chapters of this section explain the problem of multiple comparisons. This is a huge challenge in data analysis but a topic that is not covered by most introductory statistics books. The next two chapters briefly explain the principles of normality and outlier tests, topics that most statistics texts omit. Finally, Chapter 26 is an overview of determining necessary sample size.

Part F explains the basic statistical tests, including those that compare survival curves (an issue omitted from many introductory texts).

Part G is about fitting models to data. It begins, of course, with linear regression. Later chapters in the section explain the ideas of creating models and briefly explain the ideas of nonlinear regression (a method used commonly in biological research but omitted from most introductory texts), multiple regression, logistic regression, and proportional hazards regression.

Part H contains miscellaneous chapters briefly introducing analysis of variance (ANOVA), nonparametric methods, and sensitivity and specificity. It ends with a chapter (Chapter 43) on meta-analysis, a topic covered by few introductory texts.

Part I tries to put it all together. Chapter 44 is a brief summary of the key ideas of statistics. Chapter 45 is a much longer chapter explaining common traps in data analysis, a reality check missing from most statistics texts. Chapter 46 goes through one example in detail as a review. Chapter 47 is a new (to this edition) chapter on statistical concepts you need to understand to follow the current controversies about the lack of reproducibility of scientific works. Chapter 48, also new, is a checklist of things to think about when publishing (or reviewing) statistical results.

If you don't like the order of the chapters, read (or teach) them in a different order. It is not essential that you read the chapters in order. Realistically, statistics covers a lot of topics, and there is no ideal order. Every topic will be easier to understand if you had learned something else first. There is no ideal linear path through the material, and many of my chapters refer to later chapters. Some teachers have told me that they have successfully presented the chapters in a very different order than I present them.

WHAT'S NEW?

What was new in the second and third editions?

The second edition (published in 2010, 15 years after the first edition) was a complete rewrite with new chapters, expanded coverage of some topics that were only touched upon in the first edition, and a complete reorganization.

I substantially edited every chapter of the third edition and added new chapters on probability, meta-analysis, and statistical traps to avoid. The third edition introduced new sections in almost all chapters on common mistakes to avoid, statistical terms introduced in that chapter, and a chapter summary.

Overview of the fourth edition

In this fourth edition, I edited every chapter for clarity, to introduce new material, and to improve the Q&A and Common Mistakes sections. I substantially rewrote two chapters, Chapter 26 on sample size calculations and Chapter 28 about case-control studies. I also added two new chapters. Chapter 47 discusses statistical concepts regarding the reproducibility of scientific data. Chapter 48 is a set of checklists to use when publishing or reviewing scientific papers.

List of new topics in the fourth edition

- Chapter 1. Two new sections were added to the list of ways that statistics is not intuitive. One section points out that we don't expect variability to depend on sample size. The other points out that we let our biases determine how we interpret data.
- Chapter 2. New sections on conditional probability and likelihood. Updated examples.
- Chapter 4. Begins with a new section to explain different kinds of variables. New example (basketball) to replace a dated example about premature babies. Added section on Bayesian credible intervals. Improved discussion of "95% of what?" Took out rules of five and seven. Pie and stacked bar graphs to display a proportion.
- Chapter 7. New Q&As. Violin plot.
- Chapter 9. How to interpret a SD when data are not Gaussian. Different ways to report a mean and SD. How to handle data where you collect data from both eyes (or ears, elbows, etc.) in each person.
- Chapter 11. Geometric SD factor. Mentions (in Q&As) that lognormal distributions are common (e.g., dB for sound, Richter scale for earthquakes). Transforming to logs turns lognormal into Gaussian.
- Chapter 14. Error bars with lognormal data (geometric SD; CI of geometric mean). How to abbreviate the standard error of mean (SEM and SE are both used). Error bars with n = 2.
- Chapter 15. Stopped using the term *assume* with null hypothesis and instead talk about "what if the null hypothesis were true?" Defines null versus nil hypothesis. Manhattan plot. Advantage of CI over P. Cites the 2016 report about P values from the American Statistical Association.

- Chapter 16. Type S errors. What questions are answered by P values and CIs?
- Chapter 18. Added two examples and removed an outdated one (prednisone and hepatitis). Major rewrite.
- Chapter 19. Rewrote section on very high P values. Points out that a study result can be consistent both with an effect existing and with it not existing.
- Chapter 20. Distinguishing power from beta and the false discovery rate. When it makes sense to compute power.
- Chapter 21. Fixed 90% versus 95% confidence intervals. Two one-sided t tests.
- Chapter 22. Introduces the phrase (used in physics) *look elsewhere effect.*
- Chapter 23. Two new ways to get trapped by multiple comparisons, the garden of forking paths, and dichotomizing in multiple ways.
- Chapter 24. QQ plots. Corrected the explanation of kurtosis.
- Chapter 25. Points out that *outlier* has two meanings.
- Chapter 26. This chapter on sample size calculations has been entirely rewritten to clarify many topics.
- Chapter 28. This chapter on case-control studies has been substantially rewritten to clarify core concepts.
- Chapter 29. Improved definition of hazard ratio.
- Chapter 31. Added discussion of pros and cons of adjusting for pairing or matching.
- Chapter 32. New common mistake pointed out that if you correlate a variable A with another A-B, you expect r to be 0.7 even if data are totally random. Points out that r is not a percentage.
- Chapter 33. Which variable is X, and which is Y? Misleading results if you do one regression from data collected from two groups.
- Chapter 34. Defines the terms *response variable* and *explanatory variable*. Discusses three distinct goals of regression.
- Chapter 39. Expanded discussion of two-way ANOVA with an example.
- Chapter 42. Removed discussion of LOD score. Added example for HIV testing.
- Chapter 43. Added a discussion of meta-analyses using individual participant data, enlarged the discussion of funnel plots, added more Q&As.
- Chapter 45. New statistical traps: dichotomizing, confusing FDR with significance level, finding small differences with lots of noise, overfitting, pseudoreplication.
- Chapter 47. New chapter on reproducibility.
- Chapter 48. New chapter with checklists for reporting statistical methods.

What happened to the problems and answers?

The first three editions contained a chapter of problems and another chapter with extensive discussion of the answers. These have not been updated for the fourth edition, but the problems and answers for the third edition are available online at www.oup.com/us/motulsky and at intuitivebiostatistics.com .

WHICH CHAPTERS ARE ESSENTIAL?

If you don't have time to read this entire book, read these 16 chapters to learn the essential concepts of statistics:

1. Statistics and Probability Are Not Intuitive
3. From Sample to Population
4. Confidence Interval of a Proportion
9. Quantifying Scatter
10. The Gaussian Distribution
12. Confidence Interval of a Mean
14. Error Bars
15. Introducing P Values
16. Statistical Significance and Hypothesis Testing
18. Interpreting a Result That Is Statistically Significant
19. Interpreting a Result That Is Not Statistically Significant
22. Multiple Comparisons Concepts
23. The Ubiquity of Multiple Comparisons
33. Simple Linear Regression
44. The Key Concepts of Statistics
45. Statistical Traps to Avoid

WHO HELPED?

A huge thanks to the many people listed herein who reviewed draft chapters of the fourth edition. Their contributions were huge and immensely improved this book:

Reviewers:

John D. Bonagura, The Ohio State University
Hwanseok Choi, University of Southern Mississippi
John D. Chovan, Otterbein University
Stacey S. Cofield, The University of Alabama at Birmingham
Jesse Dallery, University of Florida
Vincent A. DeBari, Seton Hall University
Heather J. Hoffman, George Washington University
Stefan Judex, Stony Brook University
Janet E. Kübler, California State University, Northridge
Huaizhen Qin, Tulane University
Zaina Qureshi, University of South Carolina
Emily Rollinson, Stony Brook University
Walter E. Schargel, The University of Texas at Arlington
Evelyn Schlenker, University of South Dakota Sanford School of Medicine
Guogen Shan, University of Nevada Las Vegas
William C. Wimley, Tulane University School of Medicine
Jun-Yen Yeh, Long Island University
Louis G. Zachos, University of Mississippi

Thanks also to the reviewers of past editions, as many of their ideas and corrections survived into this edition: Raid Amin, Timothy Bell, Arthur Berg, Patrick Breheny, Michael F. Cassidy, William M. Cook, Beth Dawson, Vincent DeBari, Kathleen Engelmann, Lisa A. Gardner, William Greco, Dennis A. Johnston, Martin Jones, Janet E. Kubler, Janet E. Kübler, Lee Limbird, Leonard C. Onyiah, Nancy Ostiguy, Carol Paronis, Ann Schwartz, Manfred Stommel, Liansheng Tang, William C. Wimley, Gary Yellen Jan Agosti, David Airey, William (Matt) Briggs, Peter Chen, Cynthia J Coffman, Jacek Dmochowski, Jim Ebersole, Gregory Fant, Joe Felsenstein, Harry Frank, Joshua French, Phillip Ganter, Cedric Garland, Steven Grambow, John Hayes, Ed Jackson, Lawrence Kamin, Eliot Krause, James Leeper, Yulan Liang, Longjian Liu, Lloyd Mancl, Sheniz Moonie, Arno Motulsky, Lawrence "Doc" Muhlbaier, Pamela Ohman-Strickland, Lynn Price, Jeanette Ruby, Soma Roychowdhury, Andrew Schaffner, Paige Searle, Christopher Sempos, Arti Shankar, Patricia A Shewokis, Jennifer Shook, Sumihiro Suzuki, Jimmy Walker, Paul Weiss, and Dustin White.

I would also like to express appreciation to everyone at Oxford University Press: Jason Noe, Senior Editor; Andrew Heaton and Nina Rodriguez-Marty, Editorial Assistants; Patrick Lynch, Editorial Director; John Challice, Publisher and Vice President; Frank Mortimer, Director of Marketing; Lisa Grzan, Manager In-House Production; Shelby Peak, Senior Production Editor; Michele Laseau, Art Director; Bonni Leon-Berman, Senior Designer; Sarah Vogelsong, Copy Editor; and Pamela Hanley, Production Editor.

WHO AM I?

After graduating from medical school and completing an internship in internal medicine, I switched to research in receptor pharmacology (and published over 50 peer-reviewed articles). While I was on the faculty of the Department of Pharmacology at the University of California, San Diego, I was given the job of teaching statistics to first-year medical students and (later) graduate students. The syllabi for those courses grew into the first edition of this book.

I hated creating graphs by hand, so I created some programs to do it for me! I also created some simple statistics programs after realizing that the existing statistical software, while great for statisticians, was overkill for most scientists. These efforts constituted the beginnings of GraphPad Software, Inc., which has been my full-time endeavor for many years (see Appendix A). In this role, I exchange emails with students and scientists almost daily, which makes me acutely aware of the many ways that statistical concepts can be confusing or misunderstood.

I have organized this book in a unique way and have chosen an unusual set of topics to include in an introductory text. However, none of the ideas are particularly original. All the statistical concepts are standard and have been discussed in many texts. I include references for some concepts that are not widely known, but I don't provide citations for methods that are in common usage.

Please email me with your comments, corrections, and suggestions for the next edition. I'll post errata at www.intuitivebiostatistics.com.

Harvey Motulsky
hmotulsky@graphpad.com
November 2016

ABBREVIATIONS

Abbreviation	Definition	Chapter where defined
α (alpha)	Significance level	16
ANOVA	Analysis of variance	39
CI	Confidence interval	4
CV	Coefficient of variation	9
df	Degrees of freedom	9
FDR	False discovery rate	18
FPR	False positive rate	18
FPRP	False positive reporting probability	18
n	Sample size	4
OR	Odds ratio	28
SD or s	Standard deviation	9
SE	Standard error	14
SEM	Standard error of the mean	14
p (lower case)	Proportion	4
P (upper case)	P value	15
r	Correlation coefficient	32
ROC	Receiver operating characteristic curve	42
RR	Relative risk	27
W	Margin of error	4

PART A

Introducing Statistics