Springer Texts in Statistics

Series editors

R. DeVeaux
S.E. Fienberg
I. Olkin

Springer Texts in Statistics (STS) includes advanced textbooks from 3rd- to 4th-year undergraduate courses to 1st- to 2nd-year graduate courses. Exercise sets should be included. The series editors are currently Stephen Fienberg and Richard D. De Veaux. George Casella and Ingram Olkin were editors of the series for many years.

More information about this series at http://www.springer.com/series/417

Douglas A. Wolfe • Grant Schneider

Intuitive Introductory Statistics

Springer

Douglas A. Wolfe
Department of Statistics
The Ohio State University
Columbus, OH, USA

Grant Schneider
Upstart Network
San Carlos, CA, USA

ISSN 1431-875X ISSN 2197-4136 (eBook)
Springer Texts in Statistics
ISBN 978-3-319-56070-0 ISBN 978-3-319-56072-4 (eBook)
DOI 10.1007/978-3-319-56072-4

Library of Congress Control Number: 2017950163

Printed on acid-free paper

This Springer imprint is published by Springer Nature
The registered company is Springer International Publishing AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

To our wives, Marilyn and Jingjing, for their patience and support through the lengthy preparation of this text

Preface

Understanding and interpreting data, both experimental data and the vast quantity of observational data that is now routinely available in the public domain, need not be linked solely to comprehension of a prescribed set of mathematical expressions. Introducing students to the beauty of statistics can be both motivational and interesting without being encumbered by equations. In fact, it is our view that the goal of an introductory statistics course should be to foster an appreciation for the role of statistics and associated data analysis approaches in our everyday lives rather than to prepare the students to be statistical analysts—if necessary, there will be time enough for that secondary emphasis as they begin to concentrate on their chosen fields of study. An introductory course should foster appreciation for the relevance and importance of using statistical methods for summarizing and interpreting data, but not at the expense of enjoying the process. While it is, of course, necessary to include common statistics such as the sample mean, standard deviation, correlation, etc. in an introductory statistics course, they do little to motivate and "grab" the students and can be introduced *after* a student has come to appreciate the nature of the basic information that is provided in collections of data. This desire leads us in this book to emphasize counting

and ranking approaches as initial tools for eliciting information from data collections rather than as a fallback only when sample means and standard deviations are not effective. Not only can students easily understand such counting and ranking techniques, but they also find them to be quite valuable as they explain their analyses to others.

The second point of emphasis in this text is that examples should not be chosen simply to illustrate how a statistical procedure can be applied. Dull, contrived examples can make even interesting statistical analyses uninteresting. On the other hand, an example to which students easily relate can go a long way in both piquing and sustaining their interest in the associated statistical analysis. While they learn a statistical technique, they also unearth some information of interest to them in its own right. We all learn best when we see the relevance of a topic in our own lives. We have worked hard to motivate the statistical discussions in our text through experimental settings and data collections that we ourselves find interesting and that raise questions that we can use statistical tools to address. While we are the first to confess that our view of the world may not match up completely with that of high school juniors and seniors or college freshmen and sophomores, we have tried to include data sets in both the examples and the exercises that are both real (not just realistic) and of general relevance for the students.

A third new feature for this text is the way in which we have chosen to present the chapter exercises. In addition to the usual set of exercises at the end of each section (and there are many), we have included a substantial number of comprehensive exercises at the end of each chapter that include conceptual exercises, data analysis/computational exercises, exercises involving hands-on student activities, and exercises associated with a student's use of the Internet to access interesting and relevant data sets and statistical analyses. This provides instructors with a wide variety of exercises to challenge the individual interests of students in their classes.

The fourth and final unique feature of this introductory text is the inclusion of the necessary **R** functions to enable instructors and students to analyze data sets without the tedium often accompanying many statistical computations. Examples throughout the text still provide all of the details of the associated statistical calculations so that students are fully aware of how the various statistics elicit information from a data set(s). However, we also provide the appropriate **R** procedures that can be used to make the same calculations. It is easy enough for instructors to bypass the **R** procedures if they choose, but including them as part of the basic course would permit their students to apply the associated statistical procedures to large data sets where the direct numerical calculations by hand would be prohibitive. The use of these **R** programs also eliminates the need to include the normal, t, and χ^2 tables, as well as the relevant nonparametric null distribution tables as part of the text. We have also organized all of the R programs used in the text into a documented collection that is formally registered as an **R** package IIS specifically linked to this text.

The text is specifically designed for use in an AP statistics course for high school juniors or seniors or in a one-semester or two-quarter introductory precalculus statistics course for college freshmen or sophomores. How well we have succeeded in reaching these groups of students will clearly determine the impact of our not-just-one-more introductory statistics textbook!

Many friends and colleagues have helped with both the initial development and improvement of this text over the years. We owe a particular debt of gratitude to Brad Hartlaub for his invaluable help in initiating this project in the first place and his dedicated effort to move it forward over a number of succeeding years. We also appreciated his input from teaching much of this material in statistics classes at Kenyon College. Similarly, we owe thanks to Deborah Rumsey and Elizabeth Stasny for their feedback from using early drafts of the project in introductory statistics courses at The Ohio State

University. We also owe thanks to Jungwon Byun, Ben Chang, Neha Hebbar, Cindy Smith, and all of the attendees of the so-called "data party" at Upstart for providing a fertile testing ground for the material and concepts. Finally, we owe a special thank you to the computer support group in the Department of Statistics at The Ohio State University for their patience in helping us work through the various versions of MSWord that were confronted over the many years in preparation of the text.

Our editors Michael Penn and Hannah Bracken, who originally signed us to publish with Springer, and Rebekah McClure, who assumed our project when Hannah returned to school, were dedicated from the start of the project and provided tremendous support to see it through to publication. Our production manager Christina Oliver skillfully guided the manuscript through the production process.

To everyone who helped over the many years, our heartfelt thanks.

Columbus, OH, USA Douglas A. Wolfe
San Carlos, CA, USA Grant Schneider

Contents