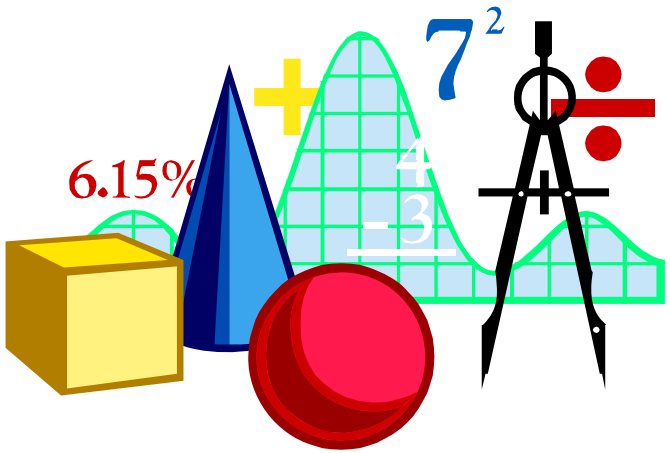

STATISTICAL LITERACY 2009

Milo Schield

Statistical Literacy 2009

Milo Schield

Fourth Edition

Printed in the U.S. by Instant Publisher

Title: Statistical Literacy 2009 SLS2008V5z.doc File size: 7,079 kb. Save Date: 9/7/2009
407 pages, 7787 paragraphs, 15442 lines, 118,363 words and 588,535 characters.

Manual update:
37 equations, 54 stories, 96 tables, 152 figures. Index: 3½ pages. Glossary: 151 terms

TABLE OF CONTENTS

Typography: **Bold** indicates a term that is being defined. Any term in bold should be found in the Glossary. *Italics* indicate an important concept, point or claim. In part-whole ratios, <u>underscoring</u> indicates a part while SMALL CAPS indicates a whole.

DEDICATION

To[1]

Florence Nightingale, Jerome Cornfield,

Dennis Haack, Victor Cohn

and Joel Best

GOAL

To help students see

the story behind the statistics

[1] Florence Nightingale (1820-1910)—the "Lady with the Lamp", the founder of modern nursing and the first female member of the Royal Statistical Society—used observational statistics to argue the need for nurses in the military. Dr. Jerome Cornfield, (1912-1979)—the creator of the Odds Ratio and Relative Risk, and a fellow of the American Statistical Society—used observational statistics to argue that the association between smoking and cancer was adequate to claim that "smoking causes cancer." Dr. Dennis Haack's 1979 textbook, "*Statistical Literacy"* is the first book to use of that phrase. He argued that statistical literacy is more about words (doublespeak) than about mathematics. Victor Cohn (1920-2000)—a former Science editor for the Washington Post, a fellow of the American Statistical Society, and author of *News and Numbers*—argued that students need to read the story behind the story when statistics are involved as evidence. Dr. Joel Best—Professor of Sociology and author of *Damned Lies and Statistics* and of *More Damned Lies and Statistics*—noted that all statistics are socially constructed by people with choices and motives. He argued that understanding this fact is most essential in evaluating statistics as evidence in arguments.

Introduction

STATISTICS are words with precision. Saying "Up to 60% of a village died during the plague" is much more precise than saying "Many people died in the plague." Saying, "Before 1930, 20% of infants died," is much more precise than saying, "Before 1930, many infants died."

Statistics can help you become aware of things you might not see. You could observe the birth of boys and girls but with statistics you could become aware that a baby is more likely to be a boy than a girl. You could visit the Central American civilizations at the time the Spanish attacked and see many natives dying, but with statistics you could become aware that up to 90% of the natives died—not from battle, but from disease. In a modern society you could attend many funerals but with statistics you could become aware that women tend to live longer than men. You could visit South Africa or Swaziland today and meet people with AIDS but with statistics you could become aware that 10% to 30% of the population has AIDS. You could visit families and notice that tall parents tend to have tall children (short parents tend to have short children), but with statistics you could become aware that tall parents are more likely to have children shorter than themselves while short parents are more likely to have children taller than themselves.

Statistics raise interesting questions. Why are babies more likely to be boys than girls? In the 1500s, why were the Central American natives more likely to die of disease than were the Spanish invaders? In modern society, why do women live longer than men? Why do sub-Saharan Africans have the highest rate of AIDS in the world? Why do children of tall parents tend to be shorter than their parents?

A primary goal of this book is to help you make better decisions when using numbers as evidence. A secondary goal is to help you improve your critical thinking—your ability to reason about argument—to help you in making sense out of your world. But to achieve either goal, you must learn to view statistics as your friends: friends that can give you advice and counsel, but friends that must always be monitored and evaluated since they have their own strengths and weaknesses.

Statistical Literacy: Seeing the Story behind the Statistics

Not all statistics are your friends. Some statistics are false, while others are ambiguous, ill-defined, misrepresented or spurious. As a reader you have three choices: (1) treat every statistic as a fact, (2) ignore every statistic since it might be an error or an opportunistic misrepresentation, or (3) learn how to distinguish good and bad statistics. Some people get overwhelmed or feel trapped when they see a statistic. If they accept the statistic as strong evidence, they may be hoodwinked. If they reject the statistics as evidence, they may make a bad choice.

The primary goal of statistical literacy is to help you evaluate the credibility of a statistic—to be able to read everyday news stories that contain statistics: to go beyond simply reading the statistics to reading 'between the lines'—to see the story behind the statistics.

It takes training and practice to read people or to see the theme of a movie or the direction of play in a sport. It takes training and practice to read an article or untangle an argument. And it takes training and practice to read a news story that uses statistics as evidence. The goal of this book is to give you that training and practice.

Statistical literacy is critical thinking about everyday arguments that use statistics as evidence. The math—mainly arithmetic—is not as central as the words. Words carry more of the action than the numbers. If a company has a 60% market share in the Eastern U.S. and a 70% market share in the Western U.S., do they have a 130% market share in the entire U.S.? No! The phrase, "market share," is the key. Or if 40% of Republicans always vote Republican and if 40% of Democrats always vote Democratic, then does that leave 20% of all voters to determine an election? No! The phrase, "of all voters," is the key.

Statistical literacy is closer to critical thinking than to mathematics. Math is deduction: black-white, right wrong. Critical thinking involves induction: shades of grey, strength of evidence supporting a conclusion.

As human beings, our primary method of thought involves generalizing from some to all, from observed to unobserved. Since we are not omniscient; we need to evaluate the strength of our reasons. Thus, our method of thinking (generalizing under uncertainty) is more fundamental than the content of our thinking (words vs. numbers). That is why this book maintains a strong focus on critical thinking. This is why statistical literacy is a

bit different from quantitative literacy. They both focus on similar topics, but statistical literacy typically focuses on arguments in which the conclusion is not a number: arguments where the conclusion is disputable, where there is no answer at the back of the book.

Statistical literacy is fast becoming a functional requirement for survival in our computer-based society. The invention of the printing press (~1450) made the reading of text a functional requirement by producing books and newspapers. The invention of the modern computer is making statistical literacy a functional requirement by generating data at an increasing rate. Journalists are also helping to make statistical literacy a necessity for citizens in a democracy. While statisticians thrive on numbers; journalists thrive on words. While statisticians try to avoid controversy and ambiguity; journalists live on controversy and ambiguity.

Statistical literacy—the ability to read and interpret data—is a requirement to understanding issues and making intelligent decisions in modern society where anyone can find a statistic to support their view.

Statistical Literacy is an emerging discipline. Welcome aboard.

Secondary Goal

Communicating is the secondary goal of this book. This involves two activities. First to help people communicate their evaluation of statistical claims as presented in graphs, tables and statements. Second, to help people use present statistical claims in their papers at school and at work, and in their presentations. Communicating involves writing and speaking, both of which are more demanding than reading – especially for those with weak English skills. This text has a strong focus on the English grammar required to communication arithmetic comparisons, ratios (using percentage, rate and change grammar) and comparisons of ratios (using likely grammar).

References:

Popular books that illustrate statistical or quantitative literacy include:

- *Mathematics and Democracy: Quantitative Literacy* by Steen (2001)
- *Damned Lies and Statistics* by Joel Best (2002), and
- *More Damned Lies and Statistics* by Joel Best (2004).

For more information on statistical literacy, visit www.StatLit.org.

Audience for this book

This book is designed primarily for college students in majors that do not require a math course but who need a quantitative reasoning course for graduation. Majors in political science, journalism and the liberal arts are prime candidates. A secondary audience involves college students in majors that require a statistical-inference course (e.g., sociology, psychology or business) but who need a bridging course to prepare them for that course.

This text is not a substitute for a traditional statistical inference course. It is not sufficient for students in majors that require statistics as a prerequisite for another course (e.g., majors such as in Sociology, Psychology, Economics, Finance or Marketing) or for students taking an MBA. It may be a helpful reference for students in majors that require statistics but do not require it as a prerequisite for other courses in their major (e.g., Management, Accounting or MIS).

This book is suitable for both two-year and four-year colleges. It is being used in a traditional teaching format and in an on-line course. This book can serve as a textbook for a half-semester, quarter or full-semester course. It can serve as a supplement or reference book for students in other courses that satisfy a quantitative reasoning requirement.

Design and Use of this book

This book gives all the important topics in the first two chapters and then gives supporting detail in the subsequent chapters whereas most statistics textbooks leave the difficult topics toward the end. While students may overload initially, they have an entire course to learn that material.

In teaching a half-semester course, one option is to cover all of chapters 1 and 2, stop at Spread in chapter 3, stop at Missing Margins in chapter 4, omit chapters 5 and 7 entirely, and stop at Predicting Influence in Chapter 6.

Level of this book

This book is at the basic principles level. An intermediate text would focus much more directly on how different disciplines use statistical association as evidence of causal connections and treat this material as a prerequisite. An intermediate text would focus more on the choice of an association (relative risk, attributable fraction, odds ratio, Phi, etc.), on the

size of the association and on its sensitivity to confounder influence (e.g., the size confounder needed to nullify or reverse an observed association).

Web-based Support for this textbook

There are over 100 different kinds of Moodle exercises with immediate feedback. For an overview, see www.StatLit.org/pdf/2006SchieldIASE.pdf. For a list, see www.StatLit.org/pdf/2009SchieldMoodle.pdf.

There is an interactive *reading program* that presents percentages and rates in statements, pie charts and tables. Each element of the description is listed. For each element, users are asked to identify whether it is part or whole. Users receive immediate feedback. This helps develop their sensitivity in reading and decoding various presentations of part-whole relationships. This program is at www.StatLit.org/GC.

There is an interactive *writing program* that presents percentages and rates in tables, graphs and statements. The goal is to help users develop their skills in using ordinary English to describe and compare rates and percentages as presented in pie charts, statements, tables and graphs. The program decodes the user's response and gives immediate feedback. This program and documentation are available at www.StatLit.org/GC

The www.StatLit.org web site contains an interactive program that helps users see how standardizing works. Standardizing is a new graphical technique for taking into account the influence of a related factor. This site contains several articles on confounding, standardization and Simpson's Paradox and contains two web-based interactive Excel programs that demonstrate this technique. Go to www.StatLit.org/Tools.htm.

Background on Statistical Literacy

For more background on statistical literacy and the relation to quantitative reasoning, quantitative literacy, numeracy and statistical thinking, see the statistical literacy website: www.StatLit.org.

For an overview of statistical literacy, see Schield's article, "*Statistical Literacy and Liberal Education at Augsburg College*" in Peer Review, a publication of the AAC&U at www.StatLit.org/pdf/2004SchieldAACU.pdf.

See www.StatLit.org/StatLitHistory.htm for a history of the phrase, 'statistical literacy.' For articles directly related to this approach to statistical literacy, see www.StatLit.org/articles.htm.

OUTCOMES:

Statistically literate adults should be able to: (1) read and interpret the use of statistics in everyday life; (2) communicate their analysis of news reports, press releases and magazine articles that use statistics as evidence; and, in their writing, (3) use statistics found in tables and graphs. This involves:

Critical thinking:

- Identify whether or not a story involves an argument
- If a story does involve an argument, identify the point and the evidence
- Identify whether claims assert association, causation, or in between

Statistical literacy:

Understand and apply your knowledge of the four influences on a statistic and how it is perceived as represented in the "Take CARE" acronym explained below. These four influences are *Context*, *Assembly*, *Randomness* and *Error* (bias). Here are specific outcomes for each of these.

Context: *Influence of factors that were or were not controlled for*

Chapter 1:

- Understand how an association can have alternate explanations
- Distinguish a *confounder* from a common cause
- Distinguish alternate explanations from mechanisms

Chapter 2:

- Distinguish study designs: experiment vs. observational study, longitudinal vs. cross-sectional, controlled and random assignment
- Read and communicate numeric comparisons: difference, ratio or relative difference; distinguish percent and percentage points
- Understand how different study designs and comparisons block or control for different kinds of related factors
- Identify plausible confounders that could affect an association

Chapter 3:

- Understand distributions, frequency, ranks and percentiles.
- Understand means, medians and modes and their relationships
- Calculate influence of a *confounder* on an average
- Understand best-fit model, range, interquartile range, standard deviation, Z-scores, coefficient of variation and effect size

Chapter 4:

- Understand how ratios control for factors such as size of group
- Identify part and whole for ratios in tables and graphs.
- Read and write descriptions of ratios using percent, percentage, rate and chance grammars, and convert among different grammars

Chapter 5:

- Calculate percentage attributed and cases attributed
- Read and write comparisons of ratios using percentage, rate, chance and likely/prevalent grammar

Chapter 6:

- Understand the confusion of the inverse, the base-rate fallacy (prosecutor's fallacy) and accuracy and error in medical tests
- Calculate the influence of a *confounder* on a ratio.

Assembly: *Influence of choices in defining, measuring or presenting*

- All chapters: Understand how the size of a statistic and a comparison of two statistics can be influenced by choices in definition, grouping, measurement and presentation

Randomness: *Influence of chance*

- Ch 1: Understand Law of Very Large Numbers and Margin of Error
- Ch 7: Distinguish empirical, analytical and subjective measures of probability; understand importance of independence
- Ch 7: Understand practical applications of chance in sensitive surveys, in adjusting for guessing and in estimating population sizes
- Ch 7: Understand confidence levels, margin of error, confidence intervals, statistical significance and the impact of sample size
- Ch 7: Distinguish statistical significance and practical importance
- Ch 7: Understand how statistical significance can be changed by taking into account the influence of a confounder

Error: *Influence of bias*

- Ch 1: Distinguish respondent , measurement and sampling bias
- Ch 2: Understand Hawthorne, halo, pessimism and safety effect
- Ch 2: Understand benefits of single blind and double blind studies

The Author

Milo Schield is a Professor of Business Administration at Augsburg College in Minneapolis and has a Ph.D. in space physics from Rice University and a CMA Certificate in Management Accounting. He has been a Senior Consultant with a national CPA firm, a Senior Operations Research Analyst in the Actuarial department for a large property-casualty insurance company, and the President of a small computer business. During his last 25 years as a college teacher he has taught a variety of subjects including operations research, accounting, finance, management, marketing and micro-economics. His primary focus was traditional statistics and critical thinking. He has taught statistical literacy (GST 200) at Augsburg College since 1998.

Dr. Schield is the director of the W. M. Keck Statistical Literacy project at Augsburg College. The aim of this project is "*to support the development of statistical literacy as an interdisciplinary curriculum in the liberal arts*." This book is a result of that grant.

In 1995 he was a visiting scholar at the Royal Statistical Societies' Centre for Statistical Education at the University of Nottingham where he studied with Peter Holmes. Since then he has given papers on statistical literacy in England, Australia, China, Japan, Singapore, South Africa, Brazil, Sweden, Scotland, Wales, Canada and the U.S.. He has organized sessions on statistical literacy at every annual meeting of the American Statistical Association since 1998. He has given talks at the U.S. Bureau of the Census, the U.S. Bureau of Labor Statistics, and the South African Statistical Organization. He has presented papers at conferences sponsored by Mathematics Association of America (MAA) the International Statistical Institute (ISI: 2005 and 2008) and the International Association of Statistical Educators (IASE): ICOTS 1998, 2002 and 2006. Milo is the vice president of the National Numeracy Network (NNN) and web master of www.StatLit.org.

For copies of his other publications, see www.StatLit.org, www.augsburg.edu/statlit or www.augnet.augsburg.edu/~schield.

E-mail questions, suggestions or comments to schield@augsburg.edu.

Statistical Literacy and Quantitative Literacy

In Current Practices in Quantitative Literacy, editor Gillman noted, "There is consensus that the mathematical skills necessary to be quantitatively literate include elementary logic, the basic mathematics of financial interest, descriptive statistics, finite probability, an elementary understanding of chance, the ability to model problems with linear and exponential models, estimations and approximation, and general problem solving."

A goal of this textbook is to introduce some of these mathematical skills. Of these skills, this text is stronger on descriptive statistics, finite probability and chance. This text includes some mathematical skills not included in Gillman's list such as weighted averages, standardizing, reading tables and graphs and communicating ratios and comparisons in English.

Statistical Literacy and GAISE

In 2005 the American Statistical Association approved Guidelines for Assessment and Instruction in Statistics Education (GAISE). The GAISE College report recommended that *"introductory courses in statistics should, as much as possible, strive to (1) emphasize statistical literacy and develop statistical thinking, (2) use real data, (3) stress conceptual understanding rather than mere knowledge of procedures, (4) foster active learning in the classroom, (5) use technology for developing conceptual understanding and analyzing data and (6) integrate assessments that are aligned with course goals to improve as well as evaluate student learning."*

The GAISE College report stated, "*We define statistical literacy as understanding the basic language of statistics (e.g., knowing what statistical terms and symbols mean and being able to read statistical graphs), and understanding some fundamental ideas of statistics*." The college report suggested that teachers assess statistical literacy by students *"interpreting or critiquing articles in the news and graphs in media*."

This textbook is arguably GAISE compliant on those concepts needed to read and interpret statistics in the everyday media. It emphasizes statistical literacy and encourages hypothetical thinking. It uses data that is both real and relevant. By avoiding formulas and related computational problems, it stresses conceptual understanding. It uses innovative web-based technology to teach the grammar of conditional probability. It supports an essay evaluation form that helps students "critique articles in the news."

To Teachers of Statistics

This book is different—very different! But that is because statistical literacy is very different from traditional inferential statistics. Statistical literacy is more about literacy than about the statistical tests, more about regression and confounding than about chance. Statistical literacy:

- focuses more on how groups are defined and statistics are measured. .
- focuses more on the type of study than on the details of the data.
- focuses more on what is not in the data than what is in the data.
- focuses more on alternate explanations (hypothetical thinking) than on deductive thinking.
- focuses more on large observational studies and confounding (e.g., Simpson's Paradox) than on small randomized experiments and chance.
- focuses more on systematic sources of variation than on randomness.

This book classifies the influences on a statistical association into four categories using the admonition: Take C.A.R.E. The "C" indicates *Context* or *Confounding*: an omnipresent source of systematic variation. The "A" indicates *Assembly*: the choice of what to measure, how to measure it, and how to present it. The "R" indicates *Randomness.* The "E" indicates *Error*.

Statistical literacy is very different! To prepare yourself, read Schield's 2004 AACU paper, *Statistical Literacy and Liberal Education at Augsburg College* and Isaacson's 2005 ASA paper: *Statistical Literacy—An Online Course at Capella University*. Study the first two chapters of this book before deciding to use it. Use the essay evaluation template to evaluate some arguments in the everyday news. Since describing and comparing percentages and rates is central to this text, read Burnham and Schield's 2006 IASSIST paper *Introduction to an Online Ratio Statement Validator* and then go on-line[2] to take the survey and use the grammar checker to test your skills. Investigate the availability of on-line training.

Finally, focus relentlessly on helping students read and interpret the statistics found in the everyday media, in press releases or in studies, surveys and journal articles.

---

[2] Survey: www.StatLit.org/Survey Grammar checker: www.StatLit.org/RSVP

Acknowledgements

Thanks to my academic guides (Bernie Folz, Bruce Reichenbach and Mark Engebretson), my intellectual guides (Eileen Schield, Ken Atkinson, Sandy Schield-Carey and Douglas Rasmussen), my conceptual guides (Ayn Rand, Nathaniel Branden, Richard Connell, Leonard Peikoff and David Kelly), my entrepreneurial guides (Wilbur, Vern and Marshall Schield, George Hock, Signe Schield and John Cerrito) and my statistical guides (Gerald Kaminski, Peter Holmes and Don Macnaughton).

Thanks to my project design guides (Magda Paleczny-Zapp and John Knight), my project co-participants (Donald Rubin and Judea Pearl) and my project supporters (Lynn Steen, Peter Holmes, David Kelley, Phillip Shively and Allan Rossman) in preparing a proposal to the W. M. Keck Foundation.

Thanks to the Augsburg project observers who reviewed these materials in class (Linda Schield, Lena Zakharova, Bill Jasperson, Julie Naylor, Michael O'Neal, John Stein, Boyd Koehler, Marc Isaacson, Peggy Cerrito, Bob Korn, Catherine Roach and Cynthia Schield).

Thanks to my editor (Larry Copes), the external reviewers (Peter Holmes and Joel Best) and the cover designer (Elliot Schield). Thanks to my colleague and the course designer of Statistical Literacy on-line for Capella University (Marc Isaacson). Thanks to my friend, colleague and co-author (Tom Burnham) for clear thinking and creative programming.

Thanks to my statistical literacy guides: to those who argue using statistics as evidence (Charles Murray, Thomas Sowell, Julian Simon, John Lott and Gerald Bracey) and to those who analyzed this activity (Freedman et al., Vic Cohn, Hans Zeisel, Jessica Utts, J. H. Abramson, Stanley Lieberson, Robyn Dawes, Howard Wainer and John Brignell).

Thanks to Joel Best (author of *Damned Lies and Statistics*) for his quintessential contribution—that all statistics are socially constructed—and for his support of this project in his book, *More Damned Lies and Statistics*.

Thanks to Augsburg's Academic Affairs Committee for approving Statistical Literacy as a catalog course in 1997 and to President Frame for his support of Augsburg's Statistical Literacy grant proposal.

Finally, a very special thanks to the W. M. Keck Foundation for their grant to Augsburg College in 2001 "*to support the development of statistical literacy as an interdisciplinary curriculum in the liberal arts*." Their entrepreneurial grant made this project viable and this book a reality.

What is Statistical Literacy?

"*What is statistical literacy? What every educated person should know*." David Moore, past-President of the American Statistical Association.

"*Statistical literacy goes beyond numeracy by focusing on reading and communicating those topics studied in numeracy*." Peter Holmes, Royal Statistical Society Centre for Statistical Education.

"*I see statistical literacy as standing in relation to traditional statistics as quantitative literacy is related to mathematics: they serve different purposes, but in each case the former is typically more useful than the latter for citizens and decision-makers*." Lynn Steen, past-President of the MAA.

Comments on Statistical Literacy:

"*There are few tasks in education today as urgent as improving the quality of statistical literacy. It is not necessary that every student learn the techniques of a professional statistician, but it is important that every student know enough to become an intelligent and critical consumer of statistical information*." Dr. David Kelley, author of *The Art of Reasoning*.

"*Many universities now have statistical or numerical literacy courses in addition to the traditional introductory statistics course. One lecture explaining the difference between an observational study and a randomized experiment, and the role of confounding variables in the interpretation of observational studies would do more to prepare students for reading the news than a dozen lectures on statistical inference procedures.*" Jessica Utts, The American Statistician, May 2003. P. 74.

"*From my perspective, this teaching of causal inference is the most interesting topic today in statistical education, certainly so at the undergraduate level*" Dr. Donald Rubin, Professor of Statistics, Harvard University.

See also Joel Best's Chronicle article, "*Telling the Truth about Dammed Lies and Statistics,*" at http://chronicle.com/free/v47/i34/34b00701.htm.

"*Misuse of the language of statistics is statistical doublespeak.*" Dennis Haack, author of *Statistical Literacy: A Guide to Interpretation*.

"*Widespread statistical illiteracy among the gifted is cause for immediate concern...*" Charles Murray, Real Education, p. 118.